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:

_____ _____

Raymond Puerini Date

DETECTION AND QUANTIFICATION OF *STREPTOCOCCUS PNEUMONIAE* SEROTYPES IN THE NASOPHARYNX OF HEALTHY CHILDREN IN PERU.

By

Raymond Puerini

Master of Public Health

Epidemiology

Dr. Keith Klugman

Committee Chair

Dr. Jorge Vidal

Committee Member

DETECTION AND QUANTIFICATION OF *STREPTOCOCCUS PNEUMONIAE* SEROTYPES IN THE NASOPHARYNX OF HEALTHY CHILDREN IN PERU.

By

Raymond A. Puerini
B.A., Wesleyan University, 2006

Thesis Committee Chair: Dr. Keith Klugman, PHD

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
2011

Abstract

DETECTION AND QUANTIFICATION OF *STREPTOCOCCUS PNEUMONIAE* SEROTYPES IN THE NASOPHARYNX OF HEALTHY CHILDREN IN PERU.

Raymond Puerini

Background

S. pneumoniae colonizes the nasopharynx of healthy children during the first months of life. From previous studies using time-consuming culture – based methodologies, multiple *S. pneumoniae* serotypes were detected in ~15% of healthy children in developed countries. However, improved methodologies to quantify all possible serotypes are needed. Studies within this thesis standardized quantitative PCR (qPCR) assays to quantify the most prevalent serotypes carried by Peruvian children in order to improve detection of all possible serotypes associated with carriage in an individual child.

Methods

Our parent project collected nasopharyngeal (NP) swabs (N=500) from Andean children <5 years of age in Cajamarca, Peru. Isolate serotypes were identified using multiplex PCR by conventional culture. 149 NP samples containing the most prevalent serotypes were then chosen and DNA was extracted from the nasopharyngeal swab. Serotype load (CFU/ml) and total *S. pneumoniae* load (*lytA* DNA) were quantified from these swabs using qPCR. Multiple serotype prediction was based upon numeric differences between total load and serotype load. Samples predicted to contain multiple serotypes (n=15) and 42 controls (predicted to contain one serotype) were further analyzed by multiplex PCR to identify all possible serotypes.

Results

The most prevalent serotypes detected were 6A/B, 23F, 15B/C, 19F, 19A, 9V/A. Samples having at least a 66% difference between total load and serotype load were more likely to contain multiple serotypes (60.0% vs. 21.4%, respectively, p=0.0058). Samples with less than 10^6 CFU/ml difference were less likely to contain multiple serotypes compared to samples having more than 10^6 CFU/ml difference (23% vs. 50%, respectively, p=0.0421). Neither sample serotype nor total bacterial load were associated with multiple serotype carriage. Finally, the cut-point of 20% difference between *lytA* and serotype-specific DNA amount offered the best combination of sensitivity (0.78) and specificity (0.69) for predicting multiple serotypes.

Conclusion

A difference of > 66% between serotype load and *lytA* DNA was associated with an increase in frequency of detection of multiple carriage. This study represents a first step towards developing quantification assays of *S. pneumoniae* serotypes carried by healthy children. By identifying all possible serotypes within NP samples, more effective vaccine strategies can ultimately be formulated.

DETECTION AND QUANTIFICATION OF *STREPTOCOCCUS PNEUMONIAE* SEROTYPES IN THE NASOPHARYNX OF HEALTHY CHILDREN IN PERU.

By

Raymond A. Puerini
B.A., Wesleyan University, 2006

Thesis Committee Chair: Dr. Keith Klugman, PHD

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
2011

Acknowledgements

I would first like to thank Dr. Jorge Vidal, who served as my thesis field advisor. Your assistance, guidance, and expertise throughout this process were instrumental in my navigation through data collection, analysis, and editing. I also appreciated your patience, support, encouragement, and understanding throughout the entire process of data collection and editing. Thank you so much for all of your help!

I would also like to thank Dr. Keith Klugman for his help and guidance in this paper. His extensive knowledge on *Streptococcus pneumoniae* and his attention to detail tremendously added to the quality of ideas and analysis.

I would also like to thank my friends and family for all the support they have given me during my two years at Emory.

Finally, I would like to thank myself (if that is even possible) for enduring the stresses and hurdles that have come up throughout this process. Go Me!

Table of Contents

List of Tables

Background

Streptococcus pneumoniae

Streptococcus pneumoniae was first seen in 1881 by Leo Escolar and later isolated by both Louis Pasteur and George Sternberg (1). From the time of its discovery, *S. pneumoniae* has been known to be the main cause of pneumococcal pneumonia. The original name of *S. pneumoniae* was established in 1926 as *Diplococcus pneumoniae*, but later it was given its current name in 1974 because of its chain-like growth when in liquid media. Since its isolation during disease much has been learned about *S. pneumoniae*. *S. pneumoniae* is a lancet-shaped, Gram positive, aerobic, alpha-hemolytic, bile soluble, coccus–shaped bacterium that is part of the normal flora of the human nasopharynx during childhood. The major route of transmission of *S. pneumoniae* is through respiratory droplets. Certain strains of *S. pneumoniae* can produce invasive pneumococcal disease (IPD) and incidence of IPD has been shown to be disproportionately higher in the very young and very old (2). Rates of nasopharyngeal carriage vary across age groups and geographic areas. Regardless of geographic location, most people will carry *S. pneumoniae* at some point in their lives (3).

Burden of disease

S. pneumoniae is a major cause of disease worldwide and accounts for about 800,000 child deaths per year (4, 5). In children under the age of five specifically, pneumonia is the leading cause of death; accounting for 19% of the deaths globally (6, 7). Additionally, the global burden of IPD has been estimated to be as much as 14.5 million episodes per year (5). Importantly, about 90% of pneumonia-related deaths occur in the

developing world where disease surveillance and access to adequate healthcare are limited. Within the literature, regional IPD and death rates have been found to vary markedly based upon the subjects studied and the study design. In general IPD rates tend to remain between 10 and 50 cases per 100,000 person-years (8). Rates as high as 224 cases per 100,000 person years have been seen in children aged 2 to 11 months in developing countries such as Gambia (8). Based upon Active Bacterial Core (ABC) surveillance, United States IPD rates vary between 3.5 and 38.7 cases per 100,000 people and IPD death rates vary between 0.7 and 6.56 deaths per 100,000 people (both rates depending upon age group) (9). In Peru the country-wide incidence of IPD is largely unknown as there is no established surveillance system in place to measure the incidence of pneumonia or acute respiratory infections. Most data on respiratory infections in Peru originates from selected hospitals focused on only a very specific population, so the actual burden of disease is only partially known. Isolated populations in the Andean Mountains are thought to have the greatest burden of disease but are very often excluded from studies because of their isolation.

Carriage

S. pneumoniae colonizes the mucosal surface of the upper respiratory tract, the nasopharynx, in early childhood. Depending on age, social economic status, and overall health, the rates of *S. pneumoniae* nasopharyngeal colonization vary markedly in populations from 2% to 70% (10, 11). All individuals are likely to be colonized by *S. pneumoniae* at some point in their lives and individuals can become colonized as early as the first 1-2 months of life (11, 12). Although limited data exists, it has been

demonstrated that people can be colonized by more than one strain of *S. pneumoniae* at a time (13). Previous research has shown that rates of co-colonization vary between about 1.3% to as high as 20% of study participants (14-16). After colonization, pneumococci can cause disease and rapidly spread to other people by coughing and may persist for months; thereby retaining non-invasive strains in the human population. In the absence of disease, carriage provides *S. pneumoniae* with a more stable environment and offers a suitable site from which to spread to other hosts. Thus, colonization of the nasopharynx is an important risk factor for developing pneumococcal disease and pneumococcal transmission.

Pneumococcal Diseases

S. pneumoniae is the source of many kinds of severe infections including pneumonia, meningitis, otitus media, and bacteremia. Clinical presentation of symptoms can include cough, difficulty breathing, fever, yellow or bloody sputum, fatigue, chest pains in those with pneumonia and headache, neck stiffness, confusion, dizziness, fatigue, nausea, and sensitivity to light in those with meningitis. The 93 serotypes of *S. pneumoniae* differ in virulence, level of antibiotic resistance, and prevalence. Once colonized some serotypes might never cause disease while others may cause disease rapidly.(17). Across the world the serotypes that are common in one geographic region may vary when compared to regions that are distant. Risk factors that contribute towards increased rates of *S. pneumoniae* infection have been demonstrated to include (but are not limited to): concurrent/preceding viral respiratory infection, lack of healthcare services, overcrowding, exposure to indoor pollution, lack of pneumococcal immunization, poor

sanitation, and inadequate nutrition (18-20).

Viral and Bacterial Co-infections

The burden of pneumococcal disease can be greatly influenced by the presence of other viral or bacterial infections in an affected individual. For example, in individuals suffering from influenza virus infections, the severity of disease can be amplified when patients have *S. pneumoniae* co-infections (21). Animal models of disease transmission suggest that previous infection by influenza virus increases the susceptibility of contracting IPD as well as the overall transmissibility of pneumococcal infection amongst individuals (22). In a study of prevalence in adults, carriage rates of *S. pneumoniae* have been shown to double when adults suffer from upper respiratory tract infections (23). The likelihood of opportunistic pneumococcal infections has also been shown to increase in those infected with HIV (24, 25). Bacterial pneumonia (especially pneumococcal pneumonia*)* is also the leading cause of mortality in HIV infected children (24, 25).

The pathogenic relationship between *S. pneumoniae* and *Haemophilus influenzae* bacteria has also been studied. Both bacterial species can inhabit the upper respiratory tract (26). When both species try to cohabitate the nasopharynx, hydrogen peroxide produced by *S. pneumoniae* is thought to inhibit growth of *H. influenzae* (27). Furthermore, an increased immune response to *S. pneumoniae* when *H. influenzae* has co-colonized the nasopharynx has also been shown (28). Conversely, in some instances *H. influenzae* has also been shown to provide passive protection for *S. pneumoniae* from β-lactam antibiotics by production of β-lactamase and formation of biofilms (29). Additionally, other studies show that the presence of non-capsulate *H. influenzae* in the

nasopharynx is associated with *S. pneumoniae* colonization (30).

Vaccine

Within the last 10 years vaccination has become an important public health intervention to prevent pneumococcal infection. Initially, a heptavalent vaccine was introduced which targeted the seven most prevalent serotypes of *S. pneumoniae* in developed countries (4, 6B, 9V, 14, 18C, 19F, and 23F). In the US there appeared to be a reduction in IPD in both those aged <2 years and those aged >65 years (31). When children are carrying fewer invasive strains of *S. pneumoniae* as a result of vaccination, it follows that fewer elderly people (such as grandparents) should become infected when they interact with common carriers of disease (grandchildren). In addition, the reduction of disease occurred in both antimicrobial resistant and susceptible strains (31).

With the reduction in the targeted virulent serotypes after mass vaccination, shifts in the common serotypes found within certain geographic areas resulted. Serotypes that were once uncommon (some of which were a cause for concern due to their resistance to antibiotics) have become widespread throughout vaccinated individuals. This is particularly true in the case of serotype 19A which has a high level of multidrug resistance and whose prevalence has increased after vaccination campaigns using the heptavalent formulation (32). Carriage studies examining vaccine effects in the developing world also displayed similar trends in of post-vaccine serotype prevalence shifts as well (33, 34). In response to the changing epidemiology, vaccines that target 13 different serotypes have been developed. The 13-valent pneumococcal conjugate vaccine targets all the serotypes in the 7-valent vaccine as well as serotypes 1, 3, 5, 6A, 7F, and

19A. With recommendations by the Advisory Committee on Immunization Practices (ACIP) changing in 2010 to encourage routine vaccination using the 13-valent formulation of children aged 2-59 months, it is likely that the prevalence of the most common serotypes will eventually shift in areas where 13-valent vaccine use is common (35).

Carriage and Pneumococcal Diseases in Latin America and Peru

Although all *S. pneumoniae* serotypes can potentially induce disease, only specific serotypes have been associated with invasive phenotypes (36). Serotype prevalence often vary by geographic location, highlighting the importance of investigating the distribution of serotypes in a particular geographic region (37). For example, Brazil reported ~36% carriage rates with serotypes most frequently found in the nasopharynx of children being 14, 6B and 6A (38). Venezuela has estimated an overall carriage rate of ~49% (most frequent serotypes 23F, 15B, 6B and 19F) (39, 40). Using a standard protocol for pneumococcus culture and identification, coupled to quantitative PCR, we recently detected ~78% carriage rate in healthy children from Peru (41). Since the burden of pneumococcal disease is particularly high in children (see below) and few studies have been conducted regarding carriage rates in Peru, this thesis will further identify and, develop the technology to, quantify *S. pneumoniae* serotypes in healthy children.

As mentioned earlier, in developing countries such as Latin American countries, pneumonia continues to be a leading cause of mortality in children. Furthermore, community-acquired pneumonia (CAP) caused by the pneumococcus is also an important

morbidity risk for adults in Latin American countries (42). For example, ~35% of CAP cases in adults from Mexico, Argentina, Brazil, Chile or Uruguay are caused by *S. pneumoniae* (42). Importantly, a very recent study conducted in Peru demonstrated that *S. pneumoniae* accounts for ~90% of all cases of bacterial pneumonia in children (43).

Despite the importance of the pneumococcus, Peru (and many other countries) does not currently have a surveillance system to monitor the distribution of *S. pneumoniae* serotypes among its population or to detect the serotypes implicated in pneumococcal disease cases. Studies involving 101 cases of IPD in hospitalized children, identified in Peru serotypes 14, 6B, 19F, 23F, 5, 6A, and 19A (from higher to lower prevalence) to be among the most prevalent (44). This study does little to elucidate what the serotypes might be common in more rural areas outside of Lima, in communities who might not have access to care, nor does the study help determine the serotypes in healthy Peruvian children who are also colonized. As Peru begins to roll-out a pneumococcal vaccination program, it will be necessary to develop a more permanent surveillance system to monitor serotype distribution to effectively allocate vaccines in this resource limited country.

Detection of *Streptococcus pneumoniae* serotypes:

The most common technique used to identify *S. pneumoniae* serotypes is the Quelling reaction (capsular swelling reaction). This reaction targets specific capsular proteins of *S. pneumoniae* which vary amongst each of the 93 serotypes. The Quelling reaction is very useful, accurate and rapid to identify the serotype of *S. pneumoniae* isolated (in pure culture) from disease cases. Other assays are also available to serotype *S.*

pneumoniae, using similar culture approaches, include the 12 pooled antisera "Chessboard" method and multiplex sequential PCR approach (45, 46). In addition, there are non-culture serotyping techniques that have been emerging over the past decade for the purposes of serotyping *S. pneumoniae* including latex agglutination, antibody microarray testing, radioimmunoassay, countercurrent immuno-electrophoresis (47, 48).

For carriage studies, where all possible *S. pneumoniae* serotypes present in the nasopharynx should be detected, the Quelling reaction and the other above mentioned assays may require laborious, time- and resource-consuming protocols. One of the first modern longitudinal studies of pneumococcal carriage conducted in 1980 by Gray and colleagues in the USA described that ~5% of children appeared to be colonized with two or more different serotypes (3). Thus, to have an idea whether multiple serotypes may be present in the nasopharynx of healthy children, current culture methodologies must isolate up to 100 colonies from the nasopharynx and then, using the Quelling reaction, identify the serotype of every single colony. A ballpark estimate of 9,300 Quelling reactions (100 colonies x 93 serotype panels) need to be performed to identify all possible serotypes in the nasopharynx of a "single" child.

Given the problem that is represented by applying classic methodologies for carriage studies, molecular approaches are currently being utilized to reduce both time and cost of those studies but also to improve accuracy. For example, a multiplex PCR approach has been developed to detect most of the pneumococcus serotypes (46). This molecular approach utilizes a series of 8 sequential reactions targeting regions within the capsular locus that are specific of each serotype. To detect multiple serotypes, however, it is necessary to inoculate the nasopharyngeal sample into an enrichment broth which is

incubated for 6 h to multiply the number of pneumococcus bacteria (49). After this period, DNA is extracted and utilized as template in those multiplex reactions. Using this approach a recent study identified multiple serotypes in ~12% of nasopharyngeal samples (49). Another group in Italy detected, using a similar approach, ~16% of nasopharyngeal samples (N=19) contained multiple serotypes (50). However, this multiplex PCR approach requires of high amount of DNA to be able to amplify, and then visualize in an agarose gel, the specific PCR product. Thus, when utilizing this approach, negative reactions do not completely rule out the possibility of carrying more than one serotype.

A current alternative to circumvent the limitation of conventional PCR approaches is the use of real time PCR also known as quantitative PCR (qPCR). In turn, qPCR is more sensitive than multiplex PCR and detects, if present, <10 bacteria (51). Despite this high sensitivity, only few carriage studies have utilized qPCR to detect *S. pneumoniae* serotypes in healthy children or adults (50). Even though DNA extracted from the nasopharyngeal sample is concentrated enough to be detected by qPCR, all available set of primers and probes need to be utilized in order to identify all serotypes.

Accurate identification of *S. pneumoniae* serotypes is necessary for proper control measures to be implemented and disease burden to be correctly assessed. This is particularly true in countries where resources are limited and pneumonia is a significant cause of death, which is the case of Peru.

Purpose of Study and Aims

This thesis aims to detect and develop the technology to quantify *S. pneumoniae* serotypes present in the nasopharynx of healthy children from Peru. Given the limitations

of current methodologies, this study designed a molecular approach to improve the detection and reduce the number of reactions required to quantify all possible *S. pneumoniae* present in nasopharyngeal samples. Since the standard most commonly used typing methodology does not typically identify multiple *S. pneumoniae* serotypes per nasopharyngeal sample, treatment and vaccine strategies may fall short in targeting all appropriate disease-causing strains of interest. With new insights into the sero-epidemiology of the *S. pneumoniae*, questions regarding vaccine effectiveness, post-vaccine repopulation, serotype virulence, and multi-serotype carriage effects can be answered with increased clarity.

The main aims of this thesis are as follows:

- The development of a molecular approach utilizing qPCR and multiplex PCR assays to quantify *Streptococcus pneumoniae* serotypes in healthy children.
- To examine the patterns of colonization of *Streptococcus pneumoniae* serotypes in populations in healthy Andean Children.

Methods

Hypotheses

Primary Hypothesis: The difference in amount of DNA between the *lytA* qPCR runs and the Serotype-Specific (Ser-Spc) qPCR runs is a viable means to predict the presence of multiple serotypes in nasopharyngeal samples collected from healthy Peruvian Children.

Secondary Hypothesis: Additionally, the particular serotype and total amount of pneumococcal DNA are associated with multiple serotypes and number of different serotypes within a single nasopharyngeal sample.

IRB Approval

This is a child study of parent study: IRB00033905 which has received approval. Since my study does not contain any human subject identifiers, a separate IRB application was not required. Official e-mail correspondence for IRB-exempt status was received on November 10, 2010.

Parent Study

This thesis will be conducted under a parent study currently being conducted by Jorge E. Vidal, Rollins School of Public Health Global Health Research Assistant Professor, and Keith P. Klugman, Rollins School of Public Health Epidemiology/Global Health Professor. This parent project is being jointly executed with collaborators at Vanderbilt University and within the country of Peru. The primary research interests

include the examination of the epidemiology of *Streptococcus pneumoniae,* its serotypes and the risk factors associated with carriage rates in children of Peru.

Sample Source

Nasopharyngeal samples were provided from a parent study examining episodes of Acute Respiratory Infection (ARI) in Peruvian children aged from 0 to 36 months. The samples were collected from children in the District of San Marcos, Cajamarca, Peru on a monthly basis and also during new episodes of ARI. The original study was a prospective cohort design nested within a cluster randomized community trial. The sample collection period occurred between November 2008 and October 2009.

Collection of the nasopharyngeal samples in the parent study followed the WHO recommendations for colonization studies. Deep nasopharyngeal swab samples were taken using Darcon polyester swabs. After collection, swabs were placed in separate tubes containing 0.5-1 mL of skim milk-trypyone-glucose-glycerin, STGG transport medium. Swabs were then clipped to allow closure of the tube and were stored on wet ice while in transit to the local headquarters in Peru (within 8 hours). Upon arrival, tubes were vortexed for 10-20 s to free up bacteria from the swab and then samples were frozen at -70ºC with swabs still in the tube. Samples were shipped to Lima, Peru before they were prepared to be shipped to Emory for analysis. In keeping with the WHO guidelines mentioned earlier, samples were kept at refrigerated conditions no more than 5 days and at -20ºC no more than 6 weeks.

Study Design

A total of 149 nasopharyngeal samples obtained from healthy children were used in this study. Samples for this study were selected from the first 500 Peru samples based upon previously run single colony multiplex serotype results (Figure 1). Samples (N=149) that were identified to have the most-prevalent serotypes were chosen: 6A/B/C, 23F, 15B/C, 19F, 19A, and 9V/A (from most to least prevalent). Quantitative PCR's using both *lytA* primers and probe and Ser-Spc primers and probes were run using DNA extracted from these nasopharyngeal sample. For samples predicted to have multiple serotypes (as described later), additional qPCR analyses were executed using the remaining 5 most prevalent serotype primers/probes. All samples predicted to have multiple serotypes by qPCR and a set of 42 controls (predicted to have one serotype by qPCR) were then analyzed by Multiplex PCRs using DNA extracted from NP samples that had been inoculated into THY enrichment broth and incubated for 6 h. The methodological algorithm for this thesis is shown in Figure 2. Additionally, Appendix A provides a step-by-step walkthrough of qPCR methodology with software screenshots (where relevant).

Bacterial DNA Extraction for qPCR Standards

Isolates were thawed from the -80ºC freezer and allowed to soften. Several loops of bacteria were streaked onto blood agar plates and stored in a 35ºC/5% CO2 incubator overnight (>16 hours). Plates were swabbed to extract as many *S. pneumoniae* bacteria as possible and suspensions were made in 200 µl of sterile PBS. 100µl of TE (Tris-EDTA) buffer containing 75 U/ml of mutanolysin and 0.04 g/ml of lysozyme was added to bacterial suspensions followed by incubation in a 37ºC water bath for 1 hour. DNA was

then extracted using the Qiagen DNA extraction kit (according to the manufacturer's instructions). DNA was eluted in 100µl of DNase-, RNase-free water and samples were to be stored at -80ºC. DNA concentration was initially measured using the Nanodrop system (Nanodrop Technologies, Wilmington, DE).

qPCR Standard Preparation

Based upon sample DNA concentration, qPCR standards were prepared. The target DNA concentration of the baseline standard was 400 pg/µl. 1:10 serial dilutions were performed using PCR grade water to create 100 ul standards with the following DNA concentrations: 40 pg/µl, 4 pg/µl, 0.4 pg/µl, and 0.04 pg/µl. From there 5 ul aliquots of each standard were made so that entire standard batch did not need to be thawed for every run (done to prevent breakdown of DNA).

The sensitivity and limit of detection was evaluated for each set of standards for each serotype using the CFX96 real time instrument (Bio-Rad Laboratories, Hercules, CA). The run conditions for the qPCRs were as follows: 50ºC for 2 min, 95ºC for 2 min, followed by 40 cycles of 95ºC for 15 sec and 60ºC for 1 min. Data was analyzed using the Bio-Rad CFX Manager software version 1.6.541.1028 (Bio-Rad Laboratories).

This process was executed for the standards for the suspected 6 most prevalent serotypes in Peru: 6A, 19F, 23F, 9V/A, 19A, and 15B/C.

Quantitative Real-time PCR

The presence of Ser-Spc *S. pneumoniae* DNA from nasopharyngeal samples was examined using quantitative PCR (qPCR). Bacterial DNA was extracted from

nasopharyngeal sample using the same technique as used in the standards mentioned earlier. Bacterial load in those 149 NP samples found to contain the most prevalent serotypes (6A/B/C, 23F, 15B/C, 19F, 19A, and 9V/A) were quantified using the qPCR method. Primers and probes for the *lytA* gene were selected based upon previously published sequences described by Carvalho (51). qPCR probe and primer sequences for serotypes were the same as previously described by Azzari (50). Each run contained 2 sets of 4-5 serial diluted standards; one set being run with *lytA* primers and probes and the other set run with Ser-Spc primers and probes. Each sample was run twice, once with *lytA* primers/probes and again with Ser-Spc primers/probes (based upon the previous single colony Chelex multiplex results.) Negative controls were also run along with both sets of standards to monitor for contamination. The Bio-Rad system, analytical software, and run conditions used were identical to that standard assay run conditions mentioned above.

Multiple Serotype Prediction

From the qPCR results, presence of multiple serotypes was determined from sample colony forming unit (CFU) values using a simple algorithm. If the *lytA* CFU value for any particular sample was at least 66% (2/3) greater than the (non-zero) CFU value for the Ser-Spc run of that same sample, then multiple serotypes for that sample was predicted. If both the *lytA* run and the Ser-Spc run CFU values were less than 66% disparate, then only one serotype was predicted for that sample. 66% was chosen as break point because it was the highest percentage difference above which there was reasonable sensitivity to predict multiple serotypes for all the strains. Any runs where CT values fell

below 35 units were considered negative runs (i.e. DNA target for that run's primers/probes was not present within the sample) and were coded as zero. Table 1 provides examples of the possible prediction scenarios for these qPCR runs. Any sample with a prediction of multiple serotypes was further analyzed by additional qPCR runs for the presence of the 5 other remaining serotypes of interest in this study. For example: if a sample containing 6A/B was predicted to have multiple serotypes by qPCR, subsequent qPCRs testing for the presence of serotypes 23F, 15B/C, 19F, 19A, and 9V/A were also completed on that same sample. Since only 6 sets of Ser-Spc primers and probes were used in this study, it is not possible to determine the identity of what other serotypes outside this limited set of 6 may be present in the nasopharyngeal samples by the qPCR method alone.

Multiplex PCR Confirmation

All samples whose qPCR prediction was for multiple serotypes and also a subset of 42 samples predicted to contain one serotype were examined using sequential-multiplex PCR. The 42 control samples were selected from the set of all samples predicted to contain one serotype with a *lytA* DNA amount of least 1,000,000 CFU/ml (10^6) (to ensure DNA concentrations would be high enough to be detectable by gel electrophoresis.) Multiplex PCR was done to confirm qPCR prediction of both identity of and quantity of serotype(s). Multiplex PCR was done in a 25 µl volume containing 1X PCR reaction mix (Qiagen Multiplex PCR kit without MgCl2, PCR grade water, and primers as indicated in Table 2), 2.5µl standards with 5 µl PCR grade water, and 5 µl of DNA extracted from the THY amplification broth (Qiagen DNA extraction). Thirty-nine

primer pairs were used to target serotypes/serogroups: 1, 2, 3, 4, 5, 6A/B/C, 7F/A, 7C/7B/40, 8, 9V/A, 9N/L, 10A, 10F/10C/33C 11A/D, 12F/12A/44/46, 13, 14, 15A/F, 15B/C, 16F, 17F, 18A/B/C, 19A, 19F, 20, 21, 22F/A, 23A, 23B, 23F, 24A/B/F, 31, 33F/33A/37, 34, 35A/35C/42, 35B, 35F/47, 38/25F, and 39. Primers pairs for *cpsA*-f and *cpsA*-r were included as an internal control targeting the *cpsA* locus found in all *S. pneumoniae*. Primer designation, product size, target gene, nucleotide position, and sequences can be found in Appendix B. Thermocycling was done on a Bio-Rad MyCycler thermocycler apparatus (Bio-Rad Laboratories, Hercules, CA) with the following conditions: 95°C for 15 min; 35 amplification cycles of 94°C for 30 sec, 54°C for 90 sec, and 72°C for 90 s; and a final extension step at 72°C for 10 min.

Standards (5 µl) and 20 µl volumes of those PCR reactions were analyzed by gel electrophoresis on 1.5% agarose gels (Bio-Rad Laboratories, Hercules, CA) in 1X TBE buffer (108g TRIS Base, 55g Boric acid, 9.3g EDTA, pH 8.0) at 100V/cm for 1.5 hours. Gels were stained in a 0.5 µg/ml ethidium bromide solution and photographed by using the Bio-Rad Universal Hood II equipment and Quantity One software version 4.6.3 (Bio-Rad Laboratories). Multiplex PCR results were visually analyzed and sizes of the PCR products were estimated by comparison with a molecular size standard (100-bp ladder, 500 µg/ml; New England BioLabs, Ipswich, MA).

PCR disagreement

Samples where THY broth multiplex PCR runs indicated multiple serotypes involving the 6 most common serotypes (6A/B/C, 23F, 15B/C, 19F, 19A, 9V/A) and where qPCR

results predicted only one serotype were examined for the presence of 2nd serotype by qPCR re-examination.

Data Analysis

Definition of Analytic Variables

> 66% Difference: If the quantitative *lytA* DNA amount is at least 66% greater than the Ser-Spc DNA amount within a sample, then > 66% Difference=1. If the *lytA* and Ser-Spc DNA results were less than 66% different, then > 66% Difference=0. (SAS Coding: Log_Dif)

Serotype: Based on the single colony multiplex results. Serotype was coded using a series of dummy variable for each serotype of interest. For example all 6A/B/C samples were coded as SixA=1 and all other samples were SixA=0. (SAS Coding: SixA, NinF, TweF, NinA, NiVA, FifB)

Multiple Serotypes: Primary outcome variable of interest. The results of the THY broth multiplex PCR were used to determine whether or not a sample had multiple serotypes. If a sample had 2 or more serotypes, then Multiple Serotypes=1. If less than 2 serotypes were found by multiplex PCR then Multiple Serotypes=0. (SAS Coding: VarA)

Number of Serotypes: Secondary outcome variable of interest. The results of the THY broth multiplex PCR were used to determine quantity of serotypes. This was an ordinal variable coded as 0, 1, 2, 3, or 4, depending on multiplex results. (SAS Coding: VarF)

Total Sample DNA: The total amount of *lytA* DNA (CFU/ml) in a sample was secondary predictor variable of interest. (SAS Coding: TotCFUML, Tot_lyt)

Quantitative Difference: Two coding strategies were employed for the difference in DNA quantity between total (*lytA*) DNA in a sample and Ser-Spc DNA. The variable "Absolute difference" allowed for negative quantities to exist when Ser-Spc DNA amount was subtracted from total *lytA* DNA while the "Underzero" variable set all differences less than zero to 0. Theoretically the Ser-Spc DNA amount should never be higher than the *lytA* amount as *lytA* DNA amount represents the amount of DNA in a sample for all *S. pneumoniae* serotypes. (SAS Coding: Abs_CFU, UnderCFU)

Statistical Methods

The data was managed and analyzed using SAS software, version 9.2 (SAS Institute Inc., Cary, NC). To examine the primary hypothesis, a Chi-Square analysis and two sample t-tests were used. To explore the secondary hypothesis several procedures were carried out including two-sample t-tests, chi-square tests, fisher exact tests, univariate logistic regressions, ordinal logistic regressions, and multivariate logistic regressions. As a basis to establish a quantitative cut-off value for the difference between total DNA amount and Ser-Spc amount in predicting multiple serotypes, a ROC analysis of sensitivities and specificities of different sample subsets was performed. A separate ROC analysis also examined the appropriate cut-off level for percent differences between lytA and Ser-Spc DNA to examine predictive power over a range of options.

Agreement - Kappa Statistic:

Agreement between estimates of multiple serotypes by both the qPCR method and THY (Qiagen) multiplex PCR method will be computed. For the purpose of the

analysis multiple serotypes will be predicted by qPCR method when the > 66% Difference variable is equal to 1. This analysis will only compare the 57 samples that have results by both qPCR and THY multiplex methods. Agreement was classified as follows: < 0 as indicating no agreement, 0–.20 as slight, .21–.40 as fair, .41–.60 as moderate, .61–.80 as substantial, and .81–1 as almost perfect agreement.

Univariate Analyses:

Univariate analyses were conducted to examine the relationship between the dichotomous variable, multiple serotypes, and the following variables individually: > 66% difference, serotype, total DNA amount, absolute DNA difference, and underzero DNA difference. Total DNA amount and its relationship to number of serotypes per nasopharyngeal sample were examined using an ordinal logistic regression analysis. Exploratory analyses were also conducted between sample serotype and both total *lytA* amount as well as > 66% difference, absolute DNA difference, and underzero DNA difference.

Multivariate Analyses:

A multivariate logistic model was constructed to examine both whether serotype and > 66% difference were associated with multiple serotypes (as determined from the THY multiplex PCR results). The two predictor variables were examined for the presence of interaction and collinearity. The model was also assessed for goodness of fit and confounding and an alternative reduced model was compared to the full model to determine the most precise estimate of multiple serotypes.

Sensitivity, Specificity, and Receiver Operating Characteristic (ROC) Analyses:

The overall sensitivity and specificity between the qPCR method and THY broth multiplex PCR method was conducted on all samples having results for both methods. In all analyses conducted, the THY broth multiplex PCR was considered the gold standard in determining the presence of multiple serotypes. Additionally, a ROC analysis was used to help determine the most appropriate cut-off value using the "underzero" variable for predicting the presence of multiple serotypes within a nasopharyngeal sample. Three cut-off levels were used, representing the following differences between the total amount of DNA in a sample and the Ser-Spc DNA: <100,000 CFU/ml, 100,000-1,000,000 CFU/ml, >1,000,000 CFU/ml. The ROC analysis estimates the sensitivity and specificity for each level, allowing for the examination of the predictive power of the qPCR for multiple serotypes at increasing levels of disparities between the total amount of DNA and the Ser-Spc DNA within a sample.

The sensitivities and specificities using different cut-off percentages for the percent difference between the *lytA* amount and Ser-Spc amount to predict multiple serotypes as obtained from the THY multiplex PCR. Cut-off percentages ranged from >0% to >90% difference and sensitivities and specificities were assessed at 10% increments within that range. The predictive power for multiple serotypes will be assessed to determine which cut-off levels offer the best combinations of sensitivities and specificities.

Comparison of Serotyping Methods

Theoretical comparisons of the Quelling method, Multiplex PCR, Quantitative PCR, and the Rollins School of Public Health Method ("RSPH Method", as described above) were made. Estimated number of reactions needed, time needed, and cost for each method were compared.

Results

S. pneumoniae serotypes identified from single isolates by multiplex PCR

As mentioned, our parent project collected nasopharyngeal swabs from healthy children from Peru (N=500). *S. pneumoniae* strains were isolated and identified in 311 NP samples (62.2%) by conventional bacteriological cultures. To begin evaluating the serotypes, DNA was extracted from *S. pneumoniae* strains and molecular serotypes were further identified by a multiplex PCR approach as previously described (46, 49). Results in Table 3, and below, show that the most prevalent *S. pneumoniae* serotypes identified in those 311 strains isolated from children from Peru were, 6A/B/C (20.6%), 23F (8.7%), 15B/C (7.4%), 19F (6.1%), 19A (4.2%), and 9V/A (2.9%). In summary, ~54.7% of healthy children from this community from Peru will carry, at least, one of the above mentioned *S. pneumoniae* serogroups/types (Table 3).

Quantification of *S. pneumoniae* serotypes in healthy children using qPCR assays

Nasopharyngeal samples (NP) containing the most prevalent serotypes (most to least prevalent: 6A/B/C, 23F, 15B/C, 19F, 19A, and 9V/A) (N=149) were examined by qPCR to quantify the bacterial load of the particular serotype. For this analysis, DNA was extracted from the NP samples and two qPCR assays were run in parallel: 1) a *lytA*-based assay, to quantify the total amount of pneumococcal bacteria and 2) a Ser-Spc assay, to quantify a particular amount of single serotype. All NP swabs were positive for *lytA* and contained a range of DNA amounts (Appendix C). This indicates that healthy children from this community carry different pneumococcal loads. The range of *S. pneumoniae* load detected was ~7600 CFU/ml of *S. pneumoniae* to ~30,000,000 CFU/ml (Table 4).

A total of 135/149 (~90.6%) NP swabs were found to have detectable Ser-Spc DNA. The fact that 14 samples were negative may indicate that, 1) those samples did not contain high enough amounts of Ser-Spc DNA to be detected our qPCR assays, 2) the Chelex (singe colony) multiplex PCR results were not accurate or 3) our qPCR assay did not detect all possible serotypes that are detectable by the multiplex PCR approach (e.g. multiplex PCR primers detect serotypes 6A/B/C while qPCR primers only detect serotypes 6A/B). All samples with no detectable Ser-Spc DNA, however, had detectable *lytA* DNA and *S. pneumoniae* positive cultures therefore those children were colonized by one of the 93 *S. pneumoniae* serotypes.

Results of the quantification of *S. pneumoniae* serotypes, in 135 NP swabs are shown in Appendix C. Colonization by the most prevalent serotype, 6A/B, varies between $1.6x10^4$ to $1.6x10^7$ CFU/ml (Appendix C with a mean of $3.4x10^6$ CFU/ml) (Table 4). A similar trend was found for the rest of most common serotypes (Appendix C), indicating that carriage rates of specific *S. pneumoniae* serotypes in healthy children also vary within this South American community.

Basis of simple molecular assays to quantify multiple serotypes in the nasopharynx of healthy children.

When we compared the total amount (CFU/ml) of pneumococci (*lytA*-based qPCR assay) against total amount (CFU/ml) of the specific serotype (Ser-Spc qPCR assays), we noticed that those samples had either similar or different bacterial loads (CFU/ml) (Appendix C). These results indicate that some NP swabs may contain only one serotype and others samples could potentially contain more than one serotype. For

example, NP swab #189 probably contains only one serotype since qPCR assays quantified $9.25x10^7$ CFU/ml of *S. pneumoniae* cells and $9.3x10^7$ CFU/ml of serotype 6A/B DNA.

It is important to point out that one of the aims of this thesis intended to set up, at the RSPH, the methodology and logistics to be able to detect and quantify (by qPCR) *S. pneumoniae* serotypes in NP swabs collected from healthy children. However, our results mentioned above indicate that we may be able to discriminate, based on numbers of the qPCR assays, between samples that may contain only one serotype to those containing more than one serotype. Therefore, our results presented below represented a pilot study to set the foundations for the development of a molecular approach to quantify all possible serotypes in clinical samples without the need to run all 93 qPCR assays.

Only 15 (10.1%) out of the 149 NP swabs were found to have different S. pneumoniae load and Ser-Spc bacteria loads, as defined by > 66% difference variable. DNA extracted from these 15 NP swabs were further evaluated by qPCR utilizing primers and probes that quantified bacterial loads, if present, of the other five most prevalent serotypes in this region (i.e. 6A/B/C, 23F, 15B/C, 19F, 19A, 9V/A). For example, sample 11 mentioned above that contains serotype 6A/B was now run with sets of primers and probes to additionally quantify serotypes 15B/C, 23F, 19F, 19A, and 9V/A.

Table 5 summarizes the qPCR results of samples that had at least a 66% difference between the *lytA* DNA and Ser-Spc DNA in their initial runs (n=15/149). As seen in Table 5, ~27% of these NP swabs (N=4/15) identified two serotypes included in our panel of qPCR assays (i.e. 6A/B, 23F, 15B/C, 19F, 19A, 9V/A). In the remaining

11/15 NP samples in Table 5, multiple serotypes were not found by the qPCR assay using the set of 6 Ser-Spc primers and probes, although they may have contained serotypes outside of the set of 6 serotypes available in our qPCR assays. Additionally, 3/15 (20%) were found to contain only one serotype by both methods.

When examining multiple serotypes by EITHER the qPCR method or THY multiplex PCR method, 8/15 (~53%) of NP swabs were indicated to contain multiple serotypes *involving the 6 serotypes of interest* (Table 5). Also for one NP sample (#292), multiplex PCR indentified serotype 15B/C while the qPCR technique did not detect the 15B/C serotype in that sample. Explanations for these findings could be that, if present, an additional serotype might not belong to any of those 6 targeted by our qPCR assays, the primer and probe effectiveness might vary between qPCR and multiplex techniques, and/or the amount of those missing serotypes are not enough to be detectable by our PCR approaches (even after the enrichment step in our Multiplex method). Supporting these theories, in NP swab 28 we detected by qPCR serotype 19F ($3.5x10^4$ CFU/ml) while this serotype (19F) could not be detected by our multiplex PCR approach likely because the low amount of bacteria (thus would therefore have a low amount of DNA as well). In this particular sample serotypes 17F, 4 and 31 were detected by multiplex PCR that were not included in our qPCR assays. There were 3/15 (20%) such samples where multiple serotypes were detected by multiplex PCR, but none of the multiplex serotypes were found to be within our set of interest (6A/B/C, 23F, 15B/C, 19F, 19A, or 9V/A).

A control group of 42 NP swabs containing similar amounts (as defined by having < 66% difference between *lytA* and Ser-Spc DNA amounts) of pneumococcal DNA and Ser-Spc DNA was included in our analysis (Table 6). The average percent difference

between the amount of *lytA* DNA compared to the amount Ser-Spc DNA found within control samples was -4.98% (n=42, Table 6), while the average percent differences in the samples with > 66% difference was 85.6% (n=15, Table 5).

Characteristics of qPCR results

The main goal of this thesis was to quantify of *S. pneumoniae* serotypes within nasopharyngeal samples of healthy Peruvian children through the quantitative PCR method. Samples that were examined were selected on the basis of having been identified as containing one of the 6 most prevalent serotypes found within the Peruvian population. The distribution of serotypes for the initial 149 samples is shown in Table 3. Of those 149 samples, 135 samples were found to have detectable Ser-Spc DNA. All samples with no detectable Ser-Spc DNA had detectable *lytA* DNA to at least the 3rd log power.

The results for the average *lytA* DNA amount and standard deviation per serotype result group is summarized in Table 4. Additionally, the average difference between the *lytA* DNA and Ser-Spc DNA and corresponding 95% confidence intervals for each serotype group is also included in Table 4. Minimum and maximum DNA amounts for each group are also indicated as well. Of the 6 main serogroups examined, 23F and 15B/C have the highest and lowest average amount of overall *S. pneumoniae* DNA per sample, 5.69×10^6 CFU/ml and 8.1×10^5 CFU/ml, respectively. Of the 6 main serogroups examined 23F and 9V/A have the largest and smallest absolute average disparity between the *lytA* DNA amount and serotype specific DNA amount, 4.44×10^6 CFU/ml and 2.75×10^3 CFU/ml respectively. The average DNA amounts and differences vary even more markedly in samples with multiple serotypes found by qPCR, but there are also only 1-2

observations in these result groups. It is important to note the values listed as negative in the DNA difference columns occur as the result of the Ser-Spc DNA amount being higher than the *lytA* DNA amount.

While conducting the qPCRs, abnormal peaks occurred in the chromatograms of 28 out of 43 (65%) 23F samples examined. (See Appendix D for screenshots of normal vs. abnormal 23F peaks). When run with *lytA* primers/probes, sample chromatograms appeared normal for all 23F samples examined. Furthermore, chromatograms for standards using both *lytA* primers/probes and 23F primers/probes appeared to be normal. Since it was unclear what factors differentiated those samples with abnormal vs. normal chromatograms, all samples with serotype 23F were excluded for further analysis (n=43); Thus reducing the sample size to 106.

Characteristics of THY Broth Multiplex PCR results

In order to identify all possible serotypes within NP samples, Multiplex PCR assays were performed using as template DNA extracted from NP samples that had been inoculated into THY enrichment broth and incubated for 6 h. A total of 57 (n=15 w/ > 66% difference between *lytA* and Ser-Spc amount + 42 controls w/ <66% difference) samples were examined. Overall 56/57 (98.2%) of the multiplex PCR assays identified at least 1 serotype per sample. In 38/57 (66.7%) one serotype/serogroup was identified. In 13/57 (22.8%) two serotypes/serogroups were indentified. In 4/57 (7.0%) three serotypes/serogroups were found and in 1/57 sample (1.8%) four serotypes/serogroups were identified. The distribution of results for the serotypes found for these multiplex reactions can be found in Table 7. Of the 5 serotypes also examined by qPCR, Serotypes

6A/B/C, 19F, 15B/C, 19A, and 9V/A were present in 35/57 (61.4%), 10/57 (17.5%), 9/57 (15.8%), 4/57 (7.0%), and 2/57 (3.5%) samples, respectively. For 3/56 (5.4%) samples with results, none of the 6 serotypes of interest (6A/B/C, 23F, 15B/C, 19F, 19A, 9V/A) were identified. Overall there were 18/57 (~32%) samples with multiple serotypes found and the average number of serotypes identified per sample was 1.40 (s.d. 0.73).

Multiple Serotypes by qPCR and Multiplex PCR

For 3/149 (2.0%) samples, there were multiple serotypes identified by the initial set of single colony multiplex PCRs before qPCR analysis began. The data for the comparison of multiple serotypes by both the qPCR method and multiplex PCR method are shown in Table 8. Of the samples examined (excluding serotype 23F samples, n=43), 15/106 (14.15%) samples had at least a 66% difference in the quantitative results for the total *lytA* DNA and the Ser-Spc DNA. With the exception of samples 9V/A group, all other serotype groupings had at least 1 sample with a > 66% difference detected. By the THY multiplex method, 18/57 (31.58%) samples were identified as having multiple serotypes. In contrast to the qPCR > 66 % difference findings, all serotype groupings examined by THY multiplex PCR had at least one sample with multiple serotypes identified (Table 8). 9/18 (50%) samples found to have multiple serotypes by multiplex PCR were also found to have a > 66% difference in their quantitative qPCR results. Among the 57 samples examined by both the qPCR method and THY multiplex method, agreement between the > 66% difference variable and multiple serotypes by multiplex method was determined to be fair (0.30). Agreement varied from a high of 0.61 (substantial) for 19F samples to a low of 0.00 (none) for 19A and 9V/A samples.

Statistical analysis

Univariate Analyses

A. > 66% Difference and Multiple Serotypes

In the analyses that follow, THY multiplex PCR results will be considered the definitive test for determining the presence of multiple serotypes. To examine if > 66% difference in quantitative results for the *lytA* DNA and Ser-Spc DNA was an appropriate proxy for nasopharyngeal samples containing multiple serotypes, a chi-square analysis was conducted (Table 9). The chi-square analysis was conducted using samples that had both qPCR and THY multiplex PCR results (n=57). For samples having > 66% difference in their qPCR results, 9/15 (60%) were found to harbor multiple serotypes by THY multiplex PCR. For samples having < 66% difference in the qPCR results, only 9/42 (~21%) samples were found to contain multiple serotypes by THY multiplex PCR. The association between the > 66% difference variable and multiple serotypes by multiplex PCR variable was found to be significant at the 5% level (p=0.0058).

B. Total *lytA* DNA and Multiple Serotypes

lytA DNA amounts quantified by the qPCR method represent the total amount of pneumococcal DNA (for all serotypes) within a sample. Two separate logistic models were run to examine the relationship that total amount of DNA within a sample had with multiple serotypes (Table 10). When the outcome of interest (multiple serotypes) was coded as a dichotomous variable (of either containing multiple serotypes [1] or not [0]), total sample *lytA* DNA was not found to be significantly associated (p=0.6551). Using an

ordinal logistic model, the number of serotypes found by multiplex PCR and their relationship with total sample *lytA* DNA was examined. Again, the total amount of *lytA* DNA was not found to be significantly associated with number of serotypes within a sample (p=0.8735).

The relationship of total pneumococcal DNA in a sample and its relationship to multiple serotypes was further examined by looking at differences between amount of DNA in samples that either contained or did not contain multiple serotypes. A two independent sample t-test was conducted on all 57 samples having THY multiplex results to examine difference in *lytA* DNA in samples either containing a single serotype or containing multiple serotypes. No significant difference was found in the Total *lytA* DNA amount in samples either containing or not contain multiple serotypes (p=0.6609) (Table 11.)

Finally, an analysis of the association of multiple serotypes and relative amount of total *lytA* DNA within samples was conducted. In each of these analyses, the 57 samples having THY multiplex results were separated into two groups based upon the total amount of *lytA* DNA contained in the sample. Cut-points of 1,000,000 (10^6) CFU/ml and 10,000,000 (10^7) CFU/ml were the two levels of DNA amount that were examined. Lower values of 10,000 (10^4) CFU/ml and 100,000 (10^5) CFU/ml were also examined, but there was inadequate distribution across the multiple serotype category and DNA cut-point category to conduct a statistical analysis. Fisher exact tests were performed to ascertain the level of significance across groups. Overall there was not a significant association found between presence of multiple serotypes and relative level in DNA

across both cut-points, 10^6 CFU/ml or 10^7 CFU/ml, that were examined (p=0.1908 and p=0.2467, respectively) (Table 12).

C. Sample Serotype and Multiple Serotypes

Using the original single colony (Chelex) multiplex PCR results for the coding of serotype identity, dummy variables for each of the 5 serotypes of interest were created. Chi-Square analyses and Fisher Exact Test analyses were executed to determine if samples containing specific serotypes were associated with the presence of multiple serotypes by THY multiplex PCR. Because of sparse of data across the multiple serotype variable and serotype dummy variables, Fisher exact tests statistics were used as the definitive measure of significance for the analyses of 15B/C, 19F, 19A, and 9V/A. The Chi-square p-value was used for 6A/B, as data was not sparse across cells. Overall none of the 5 serotypes of interest were found to be significantly associated with the presence of multiple serotypes within nasopharyngeal samples (Table 13).

D. Difference in *lytA* DNA amounts and Ser-Spc DNA amounts and Multiple Serotypes

The difference in the nasopharyngeal sample DNA amounts (CFU/ml) between the *lytA* DNA total and the Ser-Spc DNA amount was examined by two sample t-tests for its relationship to multiple serotypes. This difference was coded 2 ways: either allowing for negative values or setting all negative values to zero. When negative differences were permitted, difference between *lytA* DNA amount and Ser-Spc DNA amount was found to just non-significant as a predictor of multiple serotypes at the 5% level of confidence

(p=0.0535). When all negative differences were set to zero, this difference was still not a significant predictor of multiple serotypes (p=0.1885).

Amount of difference between *lytA* DNA and Ser-Spc DNA and presence of multiple serotypes was further explored by grouping level of difference into categories of absolute difference. Cut-points of 10,000 (10^4) CFU/ml, 100,000 (10^5) CFU/ml, and 1,000,000 (10^6) CFU/ml were the levels of difference between *lytA* DNA and Ser-Spc DNA that were examined. Chi-square and Fisher Exact Tests were performed to determine if an association between presence of multiple serotypes and amount of difference was present. There was a significant association between multiple serotypes and level of difference found when the group having less than 10^6 CFU/ml difference between *lytA* DNA and Ser-Spc DNA was compared to the group having more than 10^6 CFU/ml difference (p=0.0421) (Table 14). The other grouping DNA difference levels of 10^4 CFU/ml and 10^5 CFU/ml did not display a significant association (p=0.2144 and 0.5792, respectively) (Table 14).

E. Sample Serotype and Total *lytA* DNA

Serotype dummy variables identical to those used in part C were used to examine if sample serotype was associated with the total amount of DNA within a NP sample. Two sample t-tests were conducted for each serotype of interest to see if average *lytA* DNA amounts for one serotype differed from average of the *lytA* DNA amounts for the other remaining serotypes. Serotype 23F was again excluded from this analysis and overall 106 samples were used total. Results are summarized in Table 15. Satterthwaite p-values were used when the equality of variance tests were significant, otherwise the

pooled p-values and statistics were used as the definite measures. Overall, only samples containing serotype 15B/C were found to contain a significantly smaller average amount of *lytA* DNA when compared to all other serogroups ($1.02x10^6$ CFU/ml vs. $3.35x10^6$ CFU/ml, respectively, $p<0.0001$). Serotype groups 6A/B/C, 19F, 19A, and 9V/A were found to have average *lytA* DNA amounts that were not significantly different from the other remaining serotype groups (p=0.0689, 0.3740, 0.2999, 0.7625, respectively). These results might indicate that serotype 15B/C is less adapted to NP carriage than the other serotypes that were quantified or that serotype 15B/C growth might be differentially affected by other factors that influence carriage density.

F. Sample Serotype and Difference in *lytA* DNA and Serotype-Specific DNA

The relationship between sample serotype and DNA difference as coded by the > 66% difference variable, absolute difference, and difference not allowing for negative values was conducted through Fisher Exact Tests, Chi Square analyses, and two sample independent t-tests. Table 16 summarizes the findings of these analyses. Overall when DNA difference is coded as in terms of > 66% difference variable, there is no association shown between > 66% difference variable and any of the serotypes. When difference is coded in terms of the actual CFU/ml qPCR differences, compared to other serotypes, serotypes 6A/B and 19A have significantly different amounts of difference between *lytA* and Ser-Spc DNA (allowing for negative values) (p=0.0003 and 0.0165, respectively). When difference is coded in terms of the actual CFU/ml qPCR differences (but converting all negative values to 0), compared to other serotypes, serotypes 6A/B, 19A,

and 9V/A have significantly different amounts of difference between *lytA* and Ser-Spc DNA (p=0.0055, <0.0001 and 0.0292, respectively).

Multivariate Analysis

Sample Serotype and > 66% Difference as Predictors for Multiple Serotypes

> 66% difference and sample serotype were further examined as predictors for multiple serotypes (as defined by the THY multiplex results). The following model was initially constructed contain the > 66% difference variable, serotype dummy variables, and their corresponding interaction terms: logit(p=Multiple Serotypes)=β0 +β1(Log_dif) +β2(SixA)+β3(NinF)+β4(NinA)+β5(NiVA)+β6(FifB)+γ1(Log_dif*SixA)+γ2(Log_dif* NinF)+γ3(Log_dif*NinA)+γ4(Log_dif*NiVA)+γ5(Log_dif*FifB). The model was then assessed for interaction. None of the interaction terms appeared to be significant (See Table 17 for estimates and p-values). A chunk test for the significance of the group of interaction terms was performed and they collectively were also found to be non-significant (p=0.5035). After dropping interaction terms from the model, collinearity was assessed. The highest condition index was 11.6552. Thus, collinearity did not appear to be a problem as the typical condition index threshold for problematic collinearity in logistic models is 30 and above.

Goodness of fit and confounding was then assessed on the model without interaction to get the most concise model possible (output for this model in Table 18). Only the > 66% difference variable appeared to be significant (p=0.0063), while the set of serotype dummy variables had non-significant p-values (p>>0.05). A chunk test to examine the goodness of fit for the model with and without the group of serotype

variables was performed. The goodness of fit test comparing the logistic models with and without the serotype variables together showed a non-significant difference between the two models (p=0.4845). Before serotype variables were dropped from the model, a confounding/precision assessment was performed by comparing the effect estimate of the > 66% difference variable in the logistic models with and without the serotype variables included. The model with both the > 66% difference and serotype variables was slightly more precise than the model without serotype variables (CI width: 0.534 vs. 0.596). In addition, the effect estimates of the > 66% difference variable between the two models were more than 10% different (37.8% difference) (Tables 18 & 19). Taking all this information together, the final model from this analysis for predicting multiple serotypes was the gold standard model: logit(p=Multiple Serotypes)=β0 +β1(Log_dif) +β2(SixA)+β3(NinF)+β4(NinA) +β5(NiVA)+β6(FifB).

Sensitivity, Specificity, and ROC Analysis

The sensitivity and specificity of the qPCR method to predict multiple serotypes was computed using the THY multiplex PCR results as the "gold standard." The qPCR prediction for multiple serotypes was based upon a > 66% difference between the *lytA* DNA amount and Ser-Spc DNA amount. Using the 57 samples that had results for both methods and the selected cut-off value of 66%; the qPCR method had an overall sensitivity of 0.50 and an overall specificity of 0.846 for predicting multiple serotypes.

In order to gain more insight on the predictive power of the qPCR method across differing magnitudes of differences between the total *lytA* DNA and the Ser-Spc DNA, a ROC analysis was performed. For this analysis all negative differences in DNA amounts

were set to zero. The ranges for the DNA amount differences (in CFU/ml) were 0-100,000, 100,000-1,000,000, and 1,000,000+. For the ranges of 0-100,000, 100,000-1,000,000, and 1,000,000+, the sensitivities for the qPCR method were 0.5, 0.3333, and 0.5555, respectively, while the specificities were 0.9375, 0.714, and 0.8888, respectively. This ROC curve is shown in Figure 3.

Using an ROC curve analysis, the optimal cut-off level for percent difference between lytA DNA and Ser-Spc DNA to predict multiple serotypes was assessed in Figure 4. Points closest to the inflection point of the curve are considered to have the best combination of sensitivities and specificities to predict multiple serotypes. The top three options for optimal break-points appear to be: >20%, >30% & >40% (these two break-points have the same specificities and sensitivities), and >70%. The sensitivities and specificities for the top three choices are as follows: >20%: 0.78, 0.69; >30% & >40%: 0.67, 0.74; >70%: 0.50, 0.87 (Format is Cut-off Percentage: Sensitivity, Specificity;).

Methodological Comparisons

Theoretical comparisons of the Quelling method, Multiplex PCR, Quantitative PCR, and the RSPH Method comparing the number of reactions needed to determine both a single serotype and all serotypes within in a sample are shown in Table 20. When the diagnostic objective was to find one serotype per sample, the number of reactions needed per sample is 1-6, 1-8, 1-93, and 93 for the RSPH method, Multiplex PCR, qPCR, and Quelling methods, respectively. Similarly, when the objective is finding all serotypes within a sample, the number of reactions needed per sample is 6 (~average), 8, 93, and 9300 for the RSPH method, Multiplex PCR, qPCR, and Quelling methods, respectively.

These inter-technique differences are further amplified when larger sets of samples are serotyped.

The relative speed, cost, and predictive power of the 4 methods are shown in Table 20. When the diagnostic objective method is to identify a single serotype per sample, the qPCR method is more costly comparable to the other three methods. Both the qPCR and RSPH methods would in theory be a faster alternative to the multiplex and quelling methods. Finally, the quelling method and qPCR method possesses the greatest amount of predictive power when compared to the other two methods.

Results when the diagnostic objective is identifying all serotypes within a sample are similar to the single serotype objective comparisons with the exception of the cost of the quelling method (Table 20). 9300 reactions (100 panels) are needed per sample to detect all possible serotypes within a sample. Accordingly, the relative cost per sample reflects this substantial jump in number of reactions needed per sample when compared to other methods.

Discussion

Prevalence of *S. pneumoniae* in healthy children

As mentioned earlier, pneumococcal diseases are major cause of children mortality around the world. Mortality is highest amongst the very old and very young especially in areas where access to medical care is limited. It is estimated that *S. pneumoniae* accounts for 9% of mortality in children under the age of 2 years of age in the developing world (52). Only a fraction of those developing pneumococcal disease will die from complications. The actual prevalence of *S. pneumoniae* carriage and IPD is much higher with almost every child carrying *S. pneumoniae* at some point in their lives; 15% of which will eventually develop IPD (3).

At any particular time the prevalence of *S. pneumoniae* carriage in healthy children (those without obvious IPD symptoms) varies drastically around the world. In developed countries the prevalence of *S. pneumoniae* carriage in populations of children less than 5 years of age has been found to be 9.0%, 18.3%, 29.0%, and 51.9% in Finland, Italy, the United States, and Sweden, respectively (53). In less developed countries such as Argentina, Bangladesh, Indonesia, Brazil, India, Uganda, Ecuador, and Gambia, *S. pneumoniae* carriage rates in healthy children has been found to be higher on average, 14%, 47%, 48%, 55%, 53.9-70.2%, 62%, 66%, and 87.2%, respectively (53). In rural communities of Peru, access to medical resources that help to reduce the burden of disease are somewhat limited and conditions are similar to many countries with higher carriage rates in healthy children. The current study determined the *S. pneumoniae* carriage prevalence in healthy Peruvian children in the study population to be 62.2% (311/500) by positive culture and 78.6% (393/500) by positive qPCR.

Serotype Distribution in Latin America and Peru

Determining the correct serotypes that are prevalent within a country is essential to estimate the probable efficacy of particular interventions with antibiotics and usefulness of vaccination strategies. Carriage studies of *S. pneumoniae* serotypes in Latin America are sparse, while studies of IPD are more common. Both types of studies are useful when estimating which serotypes are likely to be the most prevalent. In Latin America estimates of the most common serotypes based on IPD rates have been conducted on both large scales across countries and on smaller scales within cities and communities. Based on a study across 10 Latin American countries the 13 most common invasive serotypes (from most to least prevalent) are: 14, 6B, 1, 5, 18C, 19F, 23F, 6A, 19A, 7F, 9V, 3, and 4 (54). The results of the most common invasive serotypes found in Peru vary from study to study, but some general trends are observable (Table 21) (14, 38, 40, 44, 54-56). Furthermore, the most prevalent serotypes identified in previous Latin American carriage studies are similar to those in found the IPD studies examined and also to those identified in this study (Table 21). Cross study distributions can be summarized as follows:

- Amongst the Peru IPD results and our carriage results, serogroups 6, 19, and 23 are generally amongst the most prevalent serogroups found.
- While serogroup 15 is found in high numbers in some studies it is not in high numbers in all the Latin America and Peru studies examined.
- Serogroup 14 appears in high numbers across most Table 21 studies but it not amongst the most prevalent serotypes in the Bello Gonzalez study or our present

study.

- The most common serotypes found in Peru are in high numbers throughout Latin America countries, but there are some highly prevalent Latin American serotypes (14 and 33F) that are not found in significant numbers within our Peru Study.

The current study focused upon a somewhat isolated population in the Cajamarca region in Peru which is over 500 km from the city of Lima, Peru. Regional differences in serotypes are likely to vary based upon the frequency of travel in and out of the population. Differences in serotype distribution between the studies can likely be explained as the result of variation of the serotype distributions over large geographical areas or as shifts of serotype prevalence over time. Latin America is a vast area encompassing Central America and South America. The countries in Latin America are very different with respect to their resources to combat to IPD and their capacity to administer mass vaccination campaigns. Because some Latin America countries have rolled out the hepta-valent pneumococcal vaccine (serotypes targeted: 4, 6B, 9V, 14, 18C, 19F, and 23F) it follows that non-vaccine serotypes will eventually become more prevalent in areas with extensive vaccination coverage. Many of the most common serotypes carry by children from this rural community in Peru will be targeted by the hepta-valent vaccine and the burden of IPD is likely to fall if and when vaccination becomes common practice in Peru. Given the isolated location of Cajamarca population and the somewhat restricted base populations examined in the available studies, local spikes in certain serotypes like 15B/C or the absence of others (like 14 in the RSPH study population) certainly seem plausible if there is little travel within a region or the study population differs with respect to a larger outside population.

Missing qPCR Results

There were 14 samples that were examined by quantitative qPCR whose results for the Ser-Spc DNA amount value were missing. To explain the missing results there are several possible causes. For samples with a 6A/B/C result by multiplex PCR, a possible explanation for a zero reading in the Ser-Spc DNA amount could be that the qPCR primers/probes are only specific to serotypes 6A and 6B. If a sample contained serotype 6C, it would be detected by the multiplex PCR method but using the qPCR assay, the DNA would not be replicated using the qPCR 6A/B primers. There are 3/14 such samples with a 6A/B/C Chelex multiplex result that also have missing qPCR Ser-Spc DNA values. Other possible explanations for these 14 missing results could be: 1) human error in interpreting the somewhat subjective multiplex PCR gel results or in executing either of the PCR assays, 2) non-optimal primer/probe design, 3) errors with sample DNA extraction or sample contamination, or 4) possible transcription (or other) human errors.

Number of serotypes and Quantitative DNA Data

Using the quantitative PCR to identify and quantify *S. pneumoniae* serotypes is a novel technique that is explored by this thesis. With no previous literature on using the qPCR technique in this function, exploratory analysis of how to predict more than one serotype per sample based on quantitative DNA amount disparities was important to this investigation. This research represents the groundwork for a new and promising technological advance that could potentially reduce the time needed to identify serotypes in NP samples and has the potential to be much more specific than the current multiplex

PCR assay. Additionally, knowing that multiple serotypes are present within a sample using a small set of qPCR runs may modify approaches to treatment and vaccine strategies in the future.

In this study, the total *lytA* DNA amount was compared to the Ser-Spc amount to predict whether there were multiple serotypes contained within one nasopharyngeal sample. When DNA amounts were similar (< 66% different), a single serotype per sample was predicted and when DNA amounts were different (> 66% different) then multiple serotypes per sample were predicted by the qPCR method. The rational behind using similar and different amounts is based upon the assumption that if there are different amounts of *lytA* DNA and Ser-Spc DNA, then it should follow that there must be at least one other additional serotype in the sample that accounts for the DNA difference between the *lytA* DNA and the single Ser-Spc DNA. Additional assumptions include that the *lytA* represents the total *S. pneumoniae* in the sample and that the *lytA* DNA amount is greater than or equal to the Ser-Spc DNA amount. The estimate of 14.15% (15/106) of samples that contain multiple serotypes (when similar and different DNA amounts from the qPCR assay are used) is similar to previous estimates in the literature of 12%, 16%, and 20% (14, 49, 50).

Differences with using DNA from Isolates, NP Samples, and THY Broth for Multiplex PCR

The particular base sample type used for multiplex PCR assay is important for both sensitivity detection and can be different based upon the overall sample analysis goal. Traditionally, a single colony approach is taken where only a single isolate is used

and only a single serotype is able to be detected per sample. In theory, the multiplex PCR could use DNA extracted directly from the NP sample to detect all serotypes within a sample, but generally the concentration of DNA would not be in sufficient amounts to reliably produce observable bands on the agarose gel. An enrichment step utilizing THY broth added with 20% of rabbit serum addresses this limitation by increasing the number of bacteria and therefore the amount of DNA within the sample. Thus, THY Broth Multiplex PCR techniques give the most accurate picture of serotypes contained within a sample in terms of accuracy and completeness of serotype results.

Importance of THY Enrichment and Qiagen Reagents for Our Multiplex PCR Assays

Following the qPCR assays, multiplex PCR assays were run on the 15 samples containing different amounts of *lytA* and Ser-Spc DNA and 40 samples containing similar amounts of *lytA* and Ser-Spc DNA. These multiplex PCR assays used DNA taken from the THY enrichment broth to help to ensure that DNA concentrations were high enough to be detected by gel electrophoresis as recommended by da Gloria Carvalho et al (49). Using THY enrichment broth samples was shown to enhance serotype detection and increase discernment for multiple serotypes within a sample (49). Additionally to further increase the sensitivity of the multiplex PCR assay, Qiagen reagents were used which allows lower levels of DNA to be detectable when compared to standard Chelex multiplex PCR reagents.

Ultimately, the THY Multiplex PCR was the definitive assay to determine multiple serotypes in our study. Of the samples examined by THY Multiplex, we were able to identify 9/15 non-controls (60%) and 9/42 (~21%) controls that contained

multiple serotypes. A recent study of carriage of multiple serotypes in healthy Venezuelan children using a similar Multiplex PCR method determined that 10/50 (20%) contained multiple serotypes (14). Unlike our study, this study did not use THY amplification broth and instead extracted DNA from a swath of colonies plated from NP samples. This study also allocated different PCR reagents than our study. While our study found a greater proportion of multiple serotypes, selection of our non-controls for analysis by multiplex PCR was determined on the basis of prediction of multiple of serotypes based on the qPCR results. When compared to the controls which were not selected for further multiplex PCR analysis on the basis of having multiple serotypes, the proportion of samples with multiple serotypes from the Venezuelan carriage study were not significantly different from the proportion in our controls (20% vs. 22.5%, p=0.7769). This finding may suggest that there multiple ways to effectively screen for multiple serotypes using the Multiplex method. Future validation studies would be useful to optimize multiplex PCR assay sensitivity for the purpose maximizing discriminative power for determining multiple serotype status in NP samples.

Multiple Serotypes and DNA Disparities in *lytA* and Ser-Spc DNA

As the qPCR technology advances for the purposes of determining *S. pneumoniae* serotypes within NP samples, techniques for analyzing the quantitative data for a variety of epidemiological purposes will evolve. This thesis, being a pilot study, was designed to elucidate some possible ways of determining if multiple serotypes were present in samples. Differences in between *lytA* DNA and Ser-Spc DNA within samples were explored two main ways, in both relative and absolute terms, to examine if there was an

association with multiple serotypes within a NP sample.

In terms of absolute difference between *lytA* DNA and Ser-Spc DNA (i.e. continuous coding of differences), most of the evidence suggests that there is no direct association between amount of difference and multiple serotypes. Regardless of whether negative values were permitted or coded as zero, the logistic models with absolute difference as a predictor for multiple serotypes were non-significant. From these results, we are led to conclude that absolute amount of difference between *lytA* DNA and Ser-Spc DNA is not associated with multiple serotypes. Although upon further investigation of these differences by separating DNA difference amounts into levels of differences, groups of samples having greater than 10^7 CFU/ml difference compared to samples having less than 10^7 CFU/ml difference show a statistically significant association. Perhaps the relationship of multiple serotypes to *lytA*/Ser-Spc difference is not as clear when differences are small but as the disparity reaches a certain threshold level (i.e. 10^7); multiple serotypes within a NP sample become more likely.

The non-controls (those with < 66% differences) had a much higher overall percent difference disparity that the controls did (Tables 5 & 6). When using the > 66% difference variable to predict multiple serotypes, a significant association was found; meaning that those samples with a > 66% difference were more statistically more likely to contain multiple serotypes. While these findings could have been strengthened through repeat testing and a larger sample size, they offer a promising first step towards developing a qPCR-based assay that can estimate whether multiple serotypes are present in a NP sample. These results, when taken together with the absolute coding of the difference variable, show that associations do exist between multiple serotypes and

disparities between the *lytA*/Ser-Spc DNA. Despite these results, further testing is needed to definitively reject (or fail to reject) our primary study hypothesis (multiple serotypes status is associated with difference in *lytA* and Ser-Spc DNA).

Total *lytA* DNA Amount and Multiple Serotypes in NP Samples

None of the analyses in this study showed that the samples had an association between multiple serotypes and the total amount of *lytA* DNA found by qPCR. While this does seem like an initially promising result, it is also important to note that the set of controls had a statically greater amount of *lytA* DNA compared to the non-control group ($6.26x10^6$ CFU/ml vs. $2.22x10^6$, respectively, p=0.0006). This is most likely due to the controls having been selected on the basis of containing *lytA* DNA at least 10^6 CFU/ml or higher to ensure that detectable levels of DNA would be present in the THY samples examined by Multiplex PCR. In order to accurately determine if *lytA* amount is definitively significantly associated with multiple serotypes, controls would need to be comparable to non-controls with respect to the total amount of DNA present within samples (hence removing this particular type of selection bias.)

Sample Serotype, DNA amount, and Multiple Serotypes in NP Samples

While sample serotype was not found to be a significant predictor of multiple serotypes, some serotypes were found to be associated with significantly lower amounts of overall amount *lytA* DNA (15B/C) in NP samples. Some serotypes were also found to be associated with the amount of difference between the *lytA* DNA and Ser-Spc DNA in samples (6A/B, 19A, 9V/A). Further investigation of serotype using univariate and

multivariate logistic regression also lead to non-significant findings for association of sample serotype and multiple serotypes. The significant associations between sample serotype and quantitative total DNA amounts for some serotypes may indicate that some serotypes may be more likely to be found in lower or higher densities than others. This may have important implications on whether those particular serotypes are likely to be found by an ethidium bromide-stained gel dependent method such as Multiplex PCR. Furthermore knowing which serotypes have significant differences between *lytA* and Ser-Spc DNA (Table 16) might help to determine which serotypes show greater variability in the Ser-Spc assays or which serotypes might be more likely to be cohabitating the nasopharynx with other *S. pneumoniae* serotypes. The latter finding would need to be further examined with more testing as this pilot study did not show that sample serotype was associated multiple serotypes (as determined by THY multiplex).

Prediction Methods for Multiple Serotypes

Using the > 66% difference variable to predict multiple serotypes was also further explored in this thesis. When using > 66% difference as a proxy for predicting multiple serotypes and comparing it to the results of multiple serotypes by THY Multiplex, varying levels of agreement were found depending on the sample serotype. Serotypes with higher agreements tended to have greater correlations between > 66 % difference variable and multiple serotype status (Table 8). Given the relatively small sample size that was allocated, especially when divided across 5 major serotype groupings, these agreements do not hold much statistical weight. Overall the agreements provide a foundation on whether using > 66 % difference to indicate multiple serotypes is

appropriate or not, but it might be useful for future analyses to further explore which serotypes might be more likely to be found to co-habituating by this > 66 % difference prediction method for estimating multiple serotype status .

Using > 66 % difference to predict which samples contained multiple serotypes was then examined over ranges of DNA difference between *lytA* and Ser-Spc DNA. No general trend was seen in the sensitivities and specificities of the > 66% difference variable to correctly identify multiple serotypes as the amount of difference between the *lytA* and Ser-Spc DNA increased. The results of the ROC analysis may suggest that amount of difference between the *lytA* DNA and Ser-Spc DNA is unimportant in determining the accuracy of the > 66% difference variable to predict multiple serotypes. It may also suggest that for some unknown reason(s) or by other unknown mechanisms, > 66% difference is more appropriate to predict multiple serotype status when DNA differences are very high and very small.

Finally, there was an ROC analysis performed to examine if the > 66% difference was an appropriate cut-off percentage for the most accurate prediction of whether samples contained multiple serotypes or not. From the curve in Figure 4, cut-offs selected in 60-70% range have about a 0.50 sensitivity and ~0.80 specificity. The other optimal points selected at 20%-40% have higher overall sensitivities (~0.70) but lower overall specificities (~0.70). Thus it appears as though our chosen cut-point of 66% difference would be more accurate at correctly identifying samples with single serotypes while it would be less accurate for correctly identifying samples that contained multiple serotypes than other optimal cut-points in the 20-40% difference range. For the purpose of this analysis the 20% cut-off appears to offer the best combination of sensitivity and

specificity for predicting multiple serotypes. Depending of the goal of the investigator to avoid false negatives or avoid false positives, one option for cut-point would be preferred over the other. Because of the relatively small sample size used for these predictions would need future analyses with larger sample sizes to definitively determine the optimal cut off percentage.

Strengths/Limitations

Our study had some strengths. Use of the quantitative PCR instrument to detect bacterial DNA is very sensitive and can be accurate in identifying very low concentrations of DNA. We also used Qiagen reagents as opposed to the standard Chelex agents for our multiplex PCR runs to increase the likelihood that all possible bands could be identified. Use of the THY broth amplification step also increased DNA concentrations from NP samples to levels that would be detectable by gel electrophoresis. Overall our exploratory analysis provided many useful ways to look at investigating the quantitative data to tease out relationships among sample variables that might be otherwise un-examinable by other PCR methods with limited sensitivity or by other methods that do not produce similar quantitative data.

There were also limitations. In terms of the direct quantitative comparison of the Ser-Spc DNA being subtracted from the *lytA* DNA there was a theoretical complication. Many samples had Ser-Spc DNA amounts that were greater than the *lytA* amounts. This is theoretically impossible as the *lytA* DNA value is supposed to represent the total amount of pneumococcal DNA in a sample for all serotypes. This theoretical complication was corrected in some analyses by setting differences in these samples to

zero, but ultimately highlights that quantitative data produced from method can display variability. Also, the qPCR primers and probes for 6A/B/C were only specific to serotypes 6A/B while the multiplex primers and probes targeted 6A/B/C. Thus a possible reason for negative qPCR results involving 6A/B/C samples could be that the samples contained serotype 6C which is not detectable by current 6A/B qPCR primers/probes.

Problems with identifying 23F samples by qPCR also indicated either sub-optimal primer/probe design or problems with the handling of samples for that particular serotype. The efficiency of the qPCR reactions runs was low in some instances which could indicate impurities in samples or reagents or might also be indicative of suboptimal primers/probes or reaction temperatures. Low efficiency could also have been the apparent cause of qPCR data variability (although that was not examined in this thesis). Additionally, control selection may have been problematic as they were not randomly selected over the set of available samples with > 66% differences between *lytA* load and Ser-Spc load. Thus they may not have been exchangeable to the samples having > 66% difference with respect to overall amount of DNA and other unmeasured factors that might also be associated with total DNA amount. Investigators were also not blinded to quantitative results when performing multiplex PCR assays that followed the qPCR runs. Finally the power of these analyses were constrained by both the limited sample size that were examined by both methods and the lack of validation testing among the same samples to examine the variability of the qPCR method.

Contributions of this Research to Epidemiology and Microbiology

There is very little previous work having utilized the qPCR for epidemiological

studies involving quantification of *S. pneumoniae* serotypes. As determined from the theoretical comparison discussed earlier, the qPCR technology has the potential to identify multiple serotypes much faster than is otherwise possible, especially when dealing with many samples (Table 20). Simply using results from methods that can only indentify one serotype at a time are flawed in that they might lead to ineffective intervention strategies by not fully targeting all invasive serotypes of interest. Knowing all serotypes that are present within a NP sample will ultimately lead to more effective treatment and vaccine strategies and a more accurate picture of the serotype prevalence within an area. This study represents an important step towards developing ways to examine the quantitative PCR assay data that might aid in characterizing the serotypes that are carried in healthy children.

Future Directions

Future projects to the quantitative approach should go towards expanding the sample size used in future studies. Also studies examining method variability and further validation studies with known serotypes would also be useful in determining the true sensitivity and specificity of this assay in both indentifying particular serotypes and also indentifying multiple serotypes. This quantitative method could also be useful for examining the relationship of pneumococcal DNA amount and other epidemiological risk factors, co-infections, demographic characteristics, and treatment/vaccine effectiveness. Finally, in-depth comparability studies assessing multiple factors of cost, time, and accuracy between the qPCR method and other serotyping methods will be needed before the case can be definitively made for the qPCR method as the best available serotyping

method.

References

1. Watson DA, Musher DM, Jacobson JW, et al. A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. *Clin Infect Dis* 1993;17(5):913-24.
2. Robinson KA, Baughman W, Rothrock G, et al. Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era. *JAMA* 2001;285(13):1729-35.
3. Gray BM, Converse GM, 3rd, Dillon HC, Jr. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J Infect Dis* 1980;142(6):923-33.
4. Mulholland K. Childhood pneumonia mortality--a permanent global emergency. *Lancet* 2007;370(9583):285-9.
5. O'Brien KL, Wolfson LJ, Watt JP, et al. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet* 2009;374(9693):893-902.
6. Wardlaw T, Salama P, Johansson EW, et al. Pneumonia: the leading killer of children. *Lancet* 2006;368(9541):1048-50.
7. Bryce J, Boschi-Pinto C, Shibuya K, et al. WHO estimates of the causes of death in children. *Lancet* 2005;365(9465):1147-52.
8. Usen S, Adegbola R, Mulholland K, et al. Epidemiology of invasive pneumococcal disease in the Western Region, The Gambia. *Pediatr Infect Dis J* 1998;17(1):23-8.

9. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Streptococcus pneumoniae, 2009. . Centers for Disease Control and Prevention.; 2010. (http://www.cdc.gov/abcs/reports-findings/survreports/spneu09.html). (Accessed January 7, 2010).

10. Quintero B, Araque M, van der Gaast-de Jongh C, et al. Epidemiology of Streptococcus pneumoniae and Staphylococcus aureus colonization in healthy Venezuelan children. *Eur J Clin Microbiol Infect Dis* 2011;30(1):7-19.

11. Bogaert D, De Groot R, Hermans PW. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis* 2004;4(3):144-54.

12. Hill PC, Cheung YB, Akisanya A, et al. Nasopharyngeal carriage of Streptococcus pneumoniae in Gambian infants: a longitudinal study. *Clin Infect Dis* 2008;46(6):807-14.

13. Charalambous BM, Oriyo NM, Gillespie SH. How valid is single-colony isolation for surveillance of Streptococcus pneumoniae carriage? *J Clin Microbiol* 2008;46(7):2467-8.

14. Rivera-Olivero IA, Blommaart M, Bogaert D, et al. Multiplex PCR reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuela. *J Med Microbiol* 2009;58(Pt 5):584-7.

15. Huebner RE, Dagan R, Porath N, et al. Lack of utility of serotyping multiple colonies for detection of simultaneous nasopharyngeal carriage of different pneumococcal serotypes. *Pediatr Infect Dis J* 2000;19(10):1017-20.

16. Auranen K, Mehtala J, Tanskanen A, et al. Between-strain competition in acquisition and clearance of pneumococcal carriage--epidemiologic evidence from a longitudinal study of day-care children. *Am J Epidemiol* 2010;171(2):169-76.

17. Hausdorff WP, Bryant J, Paradiso PR, et al. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* 2000;30(1):100-21.

18. Lynch JP, 3rd, Zhanel GG. Streptococcus pneumoniae: epidemiology, risk factors, and strategies for prevention. *Semin Respir Crit Care Med* 2009;30(2):189-209.

19. Greenberg D. The shifting dynamics of pneumococcal invasive disease after the introduction of the pneumococcal 7-valent conjugated vaccine: toward the new pneumococcal conjugated vaccines. *Clin Infect Dis* 2009;49(2):213-5.

20. Rudan I, Boschi-Pinto C, Biloglav Z, et al. Epidemiology and etiology of childhood pneumonia. *Bull World Health Organ* 2008;86(5):408-16.

21. Okada T, Morozumi M, Matsubara K, et al. Characteristic findings of pediatric inpatients with pandemic (H1N1) 2009 virus infection among severe and nonsevere illnesses. *J Infect Chemother* 2010.

22. McCullers JA, McAuley JL, Browall S, et al. Influenza enhances susceptibility to natural acquisition of and disease due to Streptococcus pneumoniae in ferrets. *J Infect Dis* 2010;202(8):1287-95.

23. Regev-Yochay G, Raz M, Dagan R, et al. Nasopharyngeal carriage of Streptococcus pneumoniae by adults and children in community and family settings. *Clin Infect Dis* 2004;38(5):632-9.

24. Lopez-Palomo C, Martin-Zamorano M, Benitez E, et al. Pneumonia in HIV-infected patients in the HAART era: incidence, risk, and impact of the pneumococcal vaccination. *J Med Virol* 2004;72(4):517-24.

25. Gray DM, Zar HJ. Community-acquired pneumonia in HIV-infected children: a global perspective. *Curr Opin Pulm Med* 2010;16(3):208-16.

26. Aarts JW, van den Aardweg MT, Rovers MM, et al. Alterations in the nasopharyngeal bacterial flora after adenoidectomy in children: a systematic review. *Otolaryngol Head Neck Surg* 2010;142(1):15-20 e1.

27. Pericone CD, Overweg K, Hermans PW, et al. Inhibitory and bactericidal effects of hydrogen peroxide production by Streptococcus pneumoniae on other inhabitants of the upper respiratory tract. *Infect Immun* 2000;68(7):3990-7.

28. Lysenko ES, Ratner AJ, Nelson AL, et al. The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. *PLoS Pathog* 2005;1(1):e1.

29. Weimer KE, Juneau RA, Murrah KA, et al. Divergent Mechanisms for Passive Pneumococcal Resistance to {beta}-Lactam Antibiotics in the Presence of Haemophilus influenzae. *J Infect Dis* 2011;203(4):549-55.

30. Abdullahi O, Nyiro J, Lewa P, et al. The descriptive epidemiology of Streptococcus pneumoniae and Haemophilus influenzae nasopharyngeal carriage

in children and adults in Kilifi district, Kenya. *Pediatr Infect Dis J* 2008;27(1):59-64.

31. Kyaw MH, Lynfield R, Schaffner W, et al. Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. *N Engl J Med* 2006;354(14):1455-63.

32. Reinert R, Jacobs MR, Kaplan SL. Pneumococcal disease caused by serotype 19A: review of the literature and implications for future vaccine development. *Vaccine* 2010;28(26):4249-59.

33. Obaro SK, Adegbola RA, Banya WA, et al. Carriage of pneumococci after pneumococcal vaccination. *Lancet* 1996;348(9022):271-2.

34. Mbelle N, Huebner RE, Wasas AD, et al. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. *J Infect Dis* 1999;180(4):1171-6.

35. Nuorti JP, Whitney CG. Prevention of pneumococcal disease among infants and children - use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine - recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep* 2010;59(RR-11):1-18.

36. Weiser JN. The pneumococcus: why a commensal misbehaves. *J Mol Med* 2010;88(2):97-102.

37. Obaro S, Adegbola R. The pneumococcus: carriage, disease and conjugate vaccines. *J Med Microbiol* 2002;51(2):98-104.

38. Laval CB, de Andrade AL, Pimenta FC, et al. Serotypes of carriage and invasive isolates of Streptococcus pneumoniae in Brazilian children in the era of pneumococcal vaccines. *Clin Microbiol Infect* 2006;12(1):50-5.

39. Rivera-Olivero IA, Bogaert D, Bello T, et al. Pneumococcal carriage among indigenous Warao children in Venezuela: serotypes, susceptibility patterns, and molecular epidemiology. *Clin Infect Dis* 2007;45(11):1427-34.

40. Bello Gonzalez T, Rivera-Olivero IA, Pocaterra L, et al. [Pneumococcal carriage in mothers and children of the Panare Amerindians from the State of Bolivar, Venezuela]. *Rev Argent Microbiol* 2010;42(1):30-4.

41. Bozio K. *Analysis of Risk Factors for Pneumococcal Nasopharyngeal Carriage in Peru.* . Atlanta, GA: Emory University, Rollins School of Public Health; 2011.

42. Isturiz RE, Luna CM, Ramirez J. Clinical and economic burden of pneumonia among adults in Latin America. *Int J Infect Dis* 2010;14(10):e852-6.

43. Padilla Ygreda J, Lindo Perez F, Rojas Galarza R, et al. [Etiology of community acquired pneumonia in children 2-59 months old in two ecologically different communities from Peru]. *Arch Argent Pediatr* 2010;108(6):516-23.

44. Ochoa TJ, Egoavil M, Castillo ME, et al. Invasive pneumococcal diseases among hospitalized children in Lima, Peru. *Rev Panam Salud Publica* 2010;28(2):121-7.

45. Lalitha MK, Thomas K, Kumar RS, et al. Serotyping of Streptococcus pneumoniae by coagglutination with 12 pooled antisera. *J Clin Microbiol* 1999;37(1):263-5.

46. Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of Streptococcus pneumoniae isolates. *J Clin Microbiol* 2006;44(1):124-31.

47. Findlow H, Laher G, Balmer P, et al. Competitive inhibition flow analysis assay for the non-culture-based detection and serotyping of pneumococcal capsular polysaccharide. *Clin Vaccine Immunol* 2009;16(2):222-9.

48. Marimon JM, Monasterio A, Ercibengoa M, et al. Antibody microarray typing, a novel technique for Streptococcus pneumoniae serotyping. *J Microbiol Methods* 2010;80(3):274-80.

49. da Gloria Carvalho M, Pimenta FC, Jackson D, et al. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. *J Clin Microbiol* 2010;48(5):1611-8.

50. Azzari C, Moriondo M, Indolfi G, et al. Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. *PLoS One* 2010;5(2):e9282.

51. Carvalho Mda G, Tondella ML, McCaustland K, et al. Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. *J Clin Microbiol* 2007;45(8):2460-6.

52. Report of the meeting of the Scientific Group of Experts (SAGE) of the Children's Vaccine Initiative and the Global Programme for Vaccines and Immunization. Geneva, Switzerland: Global Programme for Vaccines and Immunization, 1996.

53. Cardozo DM, Nascimento-Carvalho CM, Souza FR, et al. Nasopharyngeal colonization and penicillin resistance among pneumococcal strains: a worldwide 2004 update. *Braz J Infect Dis* 2006;10(4):293-304.

54. Castaneda E, Agudelo CI, Regueira M, et al. Laboratory-based surveillance of Streptococcus pneumoniae invasive disease in children in 10 Latin American countries: a SIREVA II project, 2000-2005. *Pediatr Infect Dis J* 2009;28(9):e265-70.

55. Cullotta AR, Kalter HD, Delgado J, et al. Antimicrobial susceptibilities and serotype distribution of Streptococcus pneumoniae isolates from a Low socioeconomic area in Lima, Peru. *Clin Diagn Lab Immunol* 2002;9(6):1328-31.

56. Ochoa TJ, Rupa R, Guerra H, et al. Penicillin resistance and serotypes/serogroups of Streptococcus pneumoniae in nasopharyngeal carrier children younger than 2 years in Lima, Peru. *Diagn Microbiol Infect Dis* 2005;52(1):59-64.

Figures

Figure 1. Preliminary Single Colony Multiplex PCR Algorithm

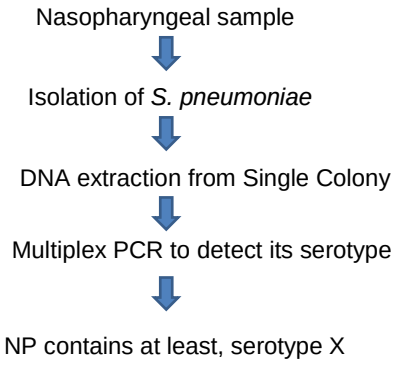
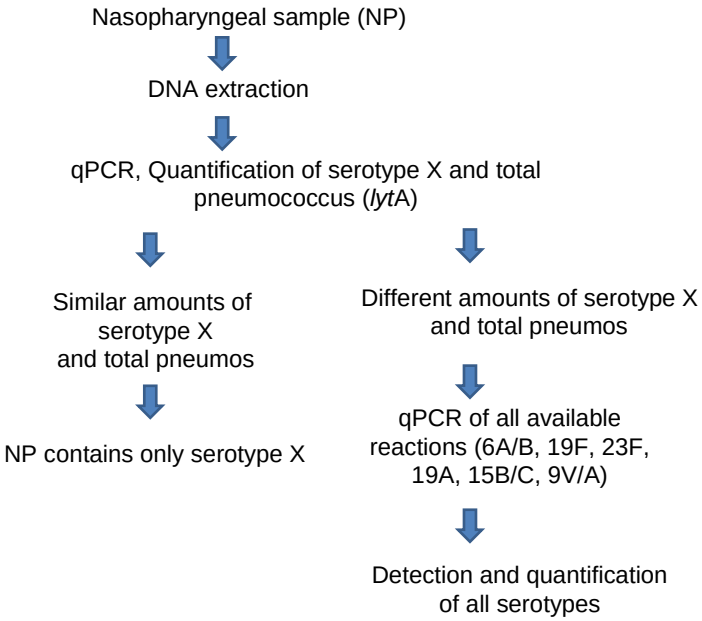


Figure 2. "Rollins School of Public Health" Serotyping Algorithm



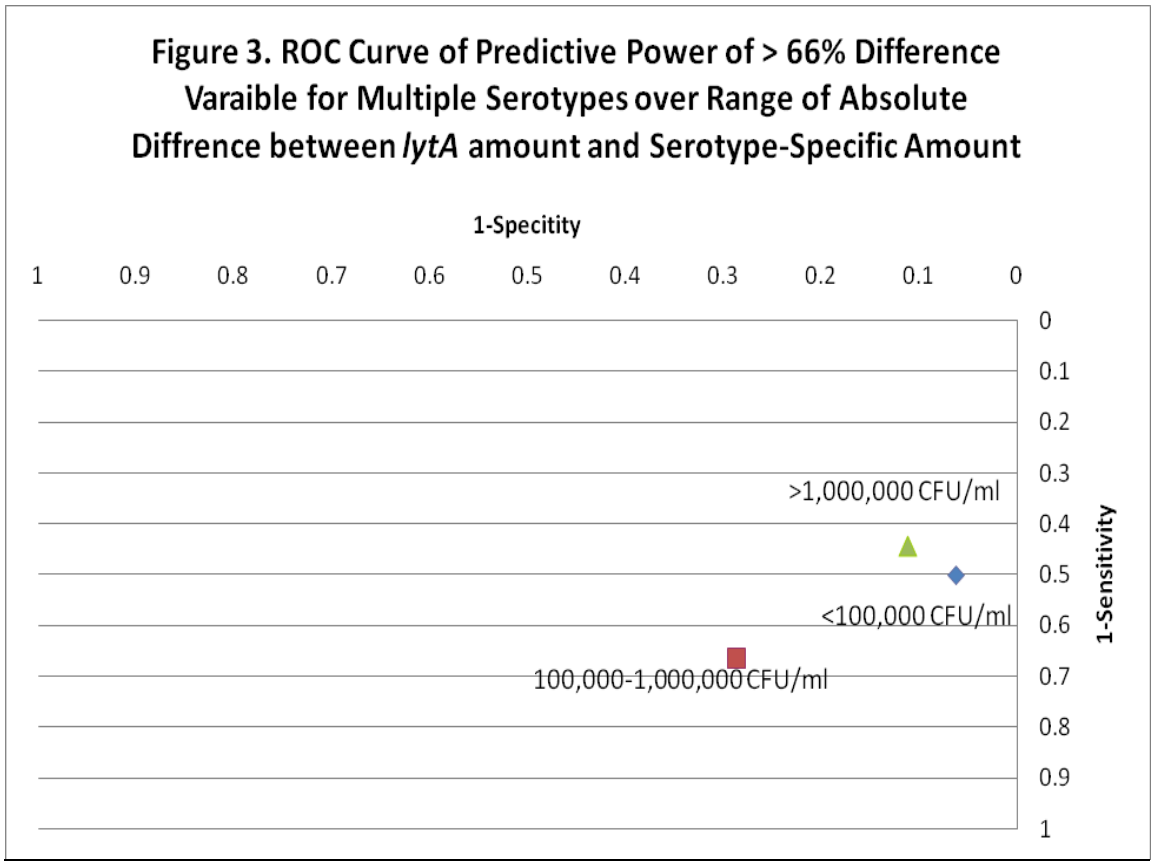


Figure 3. ROC Curve of Predictive Power of > 66% Difference Varaible for Multiple Serotypes over Range of Absolute Diffrence between lytA amount and Serotype-Specific Amount
1-Specitity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1-Sensitivity
>1,000,000 CFU/ml
<100,000 CFU/ml
100,000-1,000,000 CFU/ml

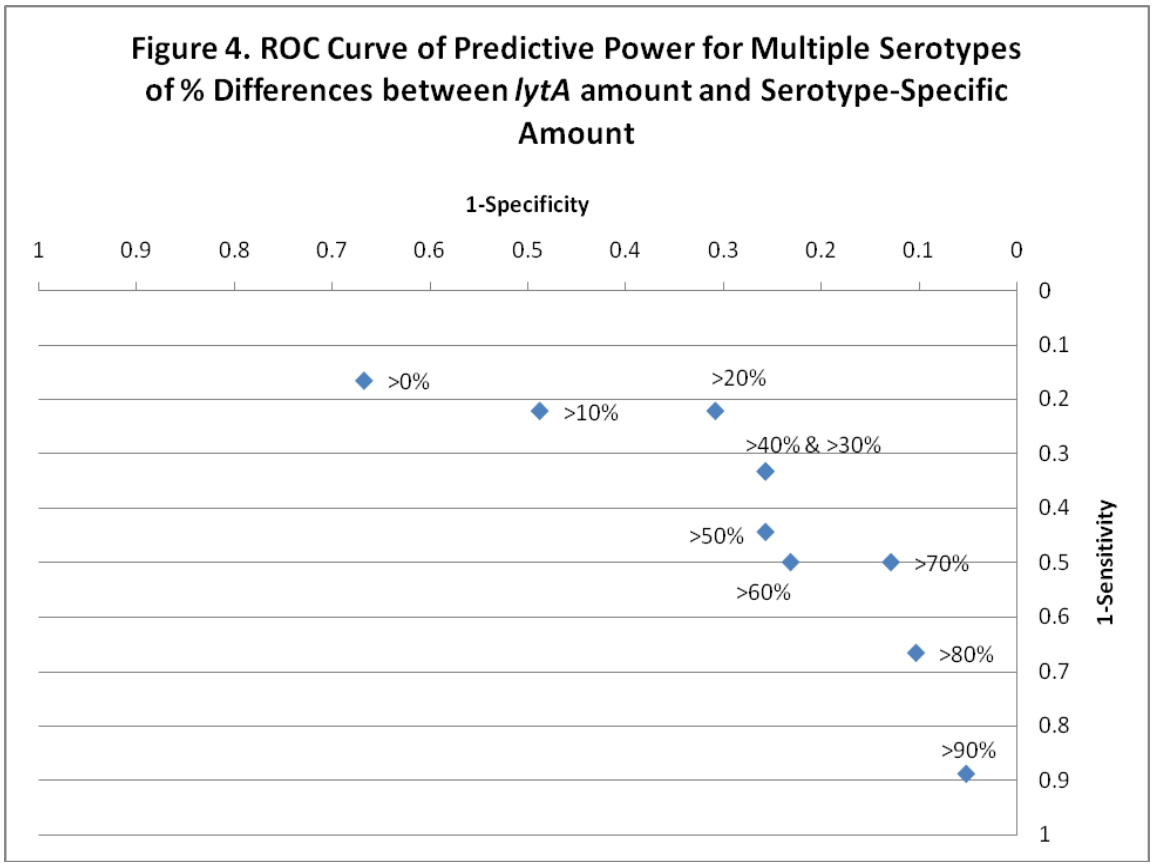


Figure 4. ROC Curve of Predictive Power for Multiple Serotypes of % Differences between lytA amount and Serotype-Specific Amount
1-Specificity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1-Sensitivity
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
>0%
>10%
>20%
>40% & >30%
>50%
>60%
>70%
>80%
>90%

Tables

Table 1. Theoretical examples of multiple serotype determination based upon DNA colony forming units (CFU/ml) values using lytA and serotype specific quantitative PCR assays.

Sample Number	*lytA* DNA (CFU/ml)	9V/A DNA (CFU/ml)	Conclusion
1	2.34E+06	1.98E+06	9V/A serotype only
2	5.67E+06	2.30E+05	Multiple serotypes (including 9V/A)
3	8.23E+05	0	No 9V/A DNA present (other serotype present)
4	0	0	No S. pneumo DNA present

Table 2. Primer Concentration for Multiplex PCR assays (1 of 2)

Reaction	Primer (Primer ID)	Primer Conc. (µM)[a]
1	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	14-F (914845)	0.3
	14-R (914846)	0.3
	6A/B/C-F (914837)	0.3
	6A/B/C-R (914838)	0.3
	23F-F (905349)	0.5
	23F-R (905350)	0.5
	19A-F (914855)	0.3
	19A-R (914856)	0.3
	9V/9A-F (913426)	0.5
	9V/9A-R (913427)	0.5
2	CPSA-F	0.1
	CPSA-R	0.1
	19F-F (903759)	0.5
	19F-R (903760)	0.5
	3-F (914831)	0.3
	3-R (914832)	0.3
	15B/15C-F (914849)	0.3
	15B/15C-R (914850)	0.3
	18/(18A/18B/18C)-F (914853)	0.3
	18/(18A/18B/18C)-R (914854)	0.3
	17F-F (903755)	0.5
	17F-R (903756)	0.5
3	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	1-F (914827)	0.3
	1-R (914828)	0.3
	5-F (914835)	0.3
	5-R (914836)	0.3
	9N/9L-F (906187)	0.5
	9N/9L-R (906188)	0.5
	7F/7A-F (918151)	0.4
	7F/7A-R (918152)	0.4
	16F-F (918138)	0.4
	16F-R (918139)	0.4
4	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	8-F (911983)	0.2
	8-R (911984)	0.2
	2-F (914829)	0.3
	2-R (914830)	0.3
	4-F (914833)	0.3
	4-R (914834)	0.3
	20-F (903761)	0.3
	20-R (903762)	0.3
	22F/22A-F (903763)	0.5
	22F/22A-R (903764)	0.5

[a]Final concentration in reaction mixture for each primer.

Table 2. Primer Concentration for Multiplex PCR assays (2 of 2)

Reaction	Primer (Primer ID)	Primer Conc. (µM)[a]
5	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	7C/7B/40-F (807802)	0.3
	7C/7B/40-R (807803)	0.3
	12F/(12A/44/46)-F (914843)	0.5
	12F/(12A/44/46)-R (914844)	0.5
	11A/11D-F (914841)	0.3
	11A/11D-R (914842)	0.3
	10A-F (906191)	0.5
	10A-R (906192)	0.5
	23A-F (903765)	0.5
	23A-R (903766)	0.5
6	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	21-F (914857)	0.2
	21-R (914829)	0.2
	33F/33A/37-F (807842)	0.3
	33F/33A/37-R (807843)	0.3
	15A/15F-F (914847)	0.3
	15A/15F-R (914848)	0.3
	35F/47F-F (807848)	0.3
	35F/47-R (807849)	0.3
	13-F (911987)	0.4
	13-R (911988)	0.4
7	39-F (911999)	0.2
	39-R (912000)	0.2
	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	23B-F (914859)	0.2
	23B-R (914860)	0.2
	35A/35C/42-F (911997)	0.3
	35A/35C/42-R (911998)	0.3
	38/25F-F (807850)	0.3
	38/25F-R (804851)	0.3
	35B-F (914863)	0.5
	35B-R (918606)	0.5
8	24(24A/24B/24F)-F (911995)	0.2
	24(24A/24B/24F)-R (911996)	0.2
	CPSA-F (913424)	0.1
	CPSA-R (913425)	0.1
	10F/10C/33C-F (911985)	0.3
	10F/10C/33C-R (911986)	0.3
	34-F (807844)	0.3
	34-R (807845)	0.3
	31-F (906193)	0.5
	31-R (906194)	0.5

[a]Final concentration in reaction mixture for each primer.

Table 3. Serotypes of *S. pneumoniae* Isolates Identified by Multiplex PCR.

Isolate Serotype(s)	Distribution of Serotypes in Culture Positive Samples (n)[a]	%	Distribution of Samples Analyzed by qPCR (n)	%
15B/C	23	7.4	23	15.4
19A	13	4.2	14	9.4
19F	19	6.1	12	8.1
23F	27	8.7	43	28.9
6A/B/C	64	20.6	45	30.2
9V/A	9	2.9	9	6.0
18A/B/C	8	2.6	-	-
4	6	1.9	-	-
6A/B/C & 19F	-	-	3	2.0
Other Serotypes	94	30.2	-	-
NT by multiplex PCR	48	15.4	-	-
Total	**311**	**100.0**	**149**	**100.0**

[a]Out of the first 500 nasopharyngeal samples from healthy children from Peru.

Table 4. Quantification of *S. pneumoniae* serotypes by Quantitative PCR in Nasopharyngeal Samples n=135[a]

Serotype(s)	Distribution (n)	%	Mean *lytA* DNA (CFU/ml)	*lytA* DNA Std dev (CFU/ml)	*ltyA* Min (CFU/ml)	*lytA* Max (CFU/ml)	Mean DNA Difference (CFU/ml)[b, c]	% Difference [d]	Diff Min (CFU/ml)	Diff Max (CFU/ml)	Mean Difference 95% CI L.B.*	Mean Difference 95% CI U.B.*
15B/C	19	14.07	8.10E+05	1.12E+06	8.23E+03	5.01E+06	2.32E+05	28.62%	2.19E+03	2.55E+06	-2.90E+04	4.93E+05
19A	13	9.63	1.82E+06	3.68E+06	7.64E+03	1.18E+07	-7.38E+05	-40.49%	-3.16E+06	1.50E+04	-1.30E+06	-1.77E+05
19F	11	8.15	4.12E+06	4.32E+06	2.48E+05	1.32E+07	-2.82E+06	-68.45%	-1.22E+07	2.68E+06	-5.57E+06	-6.48E+04
23F	37	27.41	5.69E+06	6.30E+06	1.28E+05	2.97E+07	4.44E+06	78.07%	-1.45E+06	2.63E+07	2.56E+06	6.32E+06
6A/B	42	31.11	3.39E+06	4.78E+06	2.30E+04	1.83E+07	7.16E+05	21.14%	-5.94E+05	4.63E+06	3.74E+05	1.06E+06
9V/A	7	5.19	2.66E+06	4.40E+06	5.72E+04	1.20E+07	2.75E+03	0.10%	-5.26E+05	1.02E+06	-3.54E+05	3.60E+05
6A/B & 19F	2	1.48	4.14E+06	8.46E+05	4.73E+06	3.54E+06	3.76E+06	90.96%	3.04E+06	4.48E+06	2.35E+06	5.17E+06
6A/B & 15B/C	2	1.48	1.30E+07	1.46E+07	2.67E+06	2.33E+07	6.15E+06	47.35%	1.73E+06	1.06E+07	-2.50E+06	1.48E+07
6A/B & 19A	1	0.74	1.36E+05	-	1.36E+05	1.36E+05	1.16E+05	84.85%	1.16E+05	1.16E+05	-	-
9V/A & 6A/B	1	0.74	7.54E+06	-	7.54E+06	7.54E+06	5.24E+05	6.94%	5.24E+05	5.24E+05	-	-
Total	**135**	**100**	**3.69E+06**	**5.21E+06**	**7.64E+03**	**2.97E+07**	**1.32E+06**	**35.91%**	**-1.22E+07**	**2.63E+07**	**6.29E+05**	**2.02E+06**

[a]There were 14 missing results due to no detectable serotype-specific DNA

[b]Difference when serotype-specific DNA amount is subtracted from total *lytA* DNA amount (allowing for negative values).

[c]Negative values indicate that serotype-specific amount of DNA is higher that the *lytA* DNA for all serotypes

[d]Percentage of difference between *lytA* DNA and serotype-specific DNA out of the total *lytA* DNA amount.

*Grey shaded cells indicate significant differences between total *lytA* DNA amount and serotype-DNA amount.

Table 5. Quantitative PCR Summary for Samples with Suspected Multiple Serotypes of the 6 Most Prevalent Serotypes in Peru.

NP Sample Number	Serotype[#]	*S. pneumoniae* (*lytA*) [a,*]	6A/B*	19F*	23F*	9V/A*	15B/C*	19A*	Serotypes detected by qPCR[b]	Serotypes detected by multiplex PCR[c]	% Difference between *lytA* DNA and Serotype-Specific DNA (1st run)
11	6A/B/C, 19F	6.60E+06	2.52E+05	6.16E+06	-	-	-	-	6A/B & 19F	19F	96.2
28	19F	1.81E+06	-	3.54E+04	-	-	-	-	19F	17F, 4, 31	98.0
45	15B	1.24E+07	1.90E+06	-	-	-	9.36E+05	-	6A/B & 15B	6A/B/C, 15B/C, 13	92.5
156	6A/B/C	8.15E+04	1.67E+04	-	-	-	-	-	6A/B	15A/F, 10F/C/33C	79.5
206	6A/B/C, 19F	4.40E+06	4.95E+05	3.14E+06	-	-	-	-	6A/B & 19F	6A/B/C, 19F, 13, 31	88.8
207	6A/B/C	1.63E+06	1.70E+05	-	-	-	-	-	6A/B	6A/B/C, 13	89.6
229	15B	3.73E+05	-	-	-	-	7.21E+04	-	15B	15B/C, 22F/A	80.7
268	6A/B/C	7.17E+05	2.24E+05	-	-	-	-	-	6A/B	6A/B/C	68.8
282	19A	3.86E+05	3.00E+05	-	-	-	-	2.07E+04	6A/B & 19A	6A/B/C	94.6
284	19A	4.61E+04	-	-	-	-	-	6.98E+03	19A	22F/A, 35F/47F	84.9
288	6A/B/C	9.45E+05	2.37E+05	-	-	-	-	-	6A/B	6A/B/C, 10F/10C/33C	74.9
292	6A/B/C	1.97E+05	3.96E+04	-	-	-	-	-	6A/B	6A/B/C, 15B/C, 23A	79.9
311	6A/B/C	5.50E+05	6.33E+04	-	-	-	-	-	6A/B	6A/B/C	88.5
429	15B	4.98E+06	-	-	-	-	6.17E+05	-	15B	Negative	87.6
469	15B	4.06E+05	-	-	-	-	8.21E+04	-	15B	15B/C	79.8
Averages of Non-controls (n=15)	**-**	**2.37E+06**	**3.70E+05**	**3.11E+06**	**-**	**-**	**4.27E+05**	**1.38E+04**	**-**	**-**	**85.6**

[#]Molecular serotype of *S. pneumoniae* strain isolated from NP swabs.

*CFU/ml quantified by qPCR assays.

[a]Mean *lytA* DNA amount for each sample over all *lytA* runs for that sample.

[b]Serotypes detected in NP swabs using qPCR assays targeting serotypes 6A/B, 15B/C, 23F, 19F, 19A, and 9V/A.

[c]Serotypes detected by multiplex PCR using DNA extracted from enrichment broth.

Note: Values in grey are the serotype-specific DNA values from the first set of qPCR runs.

Table 6. Quantitative PCR Summary of Control Samples (Suspected Single Serotypes) of the 6 Most Prevalent Serotypes in Peru.

NP Sample Number	Serotype[#]	*S. pneumoniae* Run 1* (*lytA*)	*S. pneumoniae* Confirmatory Run* (*lytA*)	6A/B*	19F*	23F*	9V/A*	15B/C*	19A*	Suspected qPCR results	THY Multiplex Results	Difference between Run 1 *lytA* DNA and Serotype-Specific DNA	% Difference between Run 1 *lytA* DNA and Serotype-Specific DNA
9	6A/B/C	1.42E+06	-	5.82E+05	-	-	-	-	-	6A/B	6A/B/C	8.38E+05	59.0
10	6A/B/C	1.29E+07	-	1.05E+07	-	-	-	-	-	6A/B	6A/B/C	2.35E+06	18.2
29	6A/B/C	7.00E+06	-	7.60E+06	-	-	-	-	-	6A/B	6A/B/C	-5.94E+05	-8.5
41	6A/B/C	1.89E+06	-	1.97E+06	-	-	-	-	-	6A/B	6A/B/C	-7.20E+04	-3.8
46	19F	1.32E+07	-	-	2.54E+07	-	-	-	-	19F	19F	-1.22E+07	-92.4
50	6A/B/C	9.45E+06	-	8.36E+06	-	-	-	-	-	6A/B	6A/B/C	1.09E+06	11.6
52	6A/B/C	2.85E+06	-	2.26E+06	-	-	-	-	-	6A/B	6A/B/C, 16F	5.86E+05	20.6
60	6A/B/C	7.46E+06	-	2.83E+06	-	-	-	-	-	6A/B	6A/B/C	4.63E+06	62.0
65	19F	5.67E+06	-	-	2.99E+06	-	-	-	-	19F	19F, 11A	2.68E+06	47.3
74	6A/B/C	1.83E+07	-	1.66E+07	-	-	-	-	-	6A/B	6A/B/C	1.74E+06	9.5
75[a]	15B	5.01E+06	4.95E+06	-	-	-	-	2.45E+06	-	15B	6A/B/C, 15B/C	2.55E+06	51.0
85	19A	1.13E+06	-	-	-	-	-	-	1.91E+06	19A	19A	-7.82E+05	-69.3
92[a]	9V/A	7.54E+06	5.44E+06	1.29E+06	-	-	7.02E+06	-	-	9V/A, 6A/B	9V, 6A/B/C	5.24E+05	6.9
94	19F	1.32E+06	-	-	2.25E+06	-	-	-	-	19F	19F	-9.36E+05	-71.0
102	6A/B/C	8.11E+06	-	6.79E+06	-	-	-	-	-	6A/B	6A/B/C	1.32E+06	16.3
106	19F	2.81E+06	-	-	5.46E+06	-	-	-	-	19F	19F	-2.65E+06	-94.3
109	6A/B/C	1.53E+07	-	1.45E+07	-	-	-	-	-	6A/B	6A/B/C	7.69E+05	5.0
110	6A/B/C	1.08E+06	-	1.07E+06	-	-	-	-	-	6A/B	6A/B/C	9.10E+03	0.8
134	19A	8.03E+06	-	-	-	-	-	-	9.44E+06	19A	19A, 7F	-1.41E+06	-17.6
136	6A/B/C	1.96E+06	-	1.99E+06	-	-	-	-	-	6A/B	6A/B/C	-3.14E+04	-1.6
173	6A/B/C	3.74E+06	-	1.29E+06	-	-	-	-	-	6A/B	6A/B/C	2.45E+06	65.6
190[a]	6A/B/C	2.33E+07	5.82E+06	1.27E+07	-	-	-	1.99E+05	-	6A/B, 15B/C	6A/B/C, 15B/C, 35F	1.06E+07	45.3
192	6A/B/C	6.82E+06	-	6.34E+06	-	-	-	-	-	6A/B	6A/B/C	4.77E+05	7.0
212	19F	5.30E+06	-	-	1.10E+07	-	-	-	-	19F	19F	-5.68E+06	-107.2
218	19F	2.14E+06	-	-	4.03E+06	-	-	-	-	19F	19F	-1.89E+06	-88.4
236	6A/B/C	1.22E+06	-	1.34E+06	-	-	-	-	-	6A/B	6A/B/C	-1.20E+05	-9.8
244	6A/B/C	1.17E+06	-	1.02E+06	-	-	-	-	-	6A/B	6A/B/C	1.49E+05	12.8
249	6A/B/C	7.87E+05	-	2.93E+05	-	-	-	-	-	6A/B	6A/B/C	4.94E+05	62.8
264	19A	1.18E+07	-	-	-	-	-	-	1.49E+07	19A	19A	-3.16E+06	-26.8
266[a]	9V/A	4.59E+06	4.02E+06	-	-	-	5.12E+06	-	-	9V/A	9V/A, 6A/B/C	-5.26E+05	-11.5
274	19F	1.35E+06	-	-	2.61E+06	-	-	-	-	19F	19F	-1.26E+06	-92.9
296	6A/B/C	1.61E+07	-	1.15E+07	-	-	-	-	-	6A/B	6A/B/C	4.58E+06	28.5
297	15B	1.41E+06	-	-	-	-	-	1.05E+06	-	15B	15B/C	3.68E+05	26.0
299	19F	1.09E+07	-	-	2.10E+07	-	-	-	-	19F	19F	-1.00E+07	-91.5
335	6A/B/C	1.33E+06	-	1.16E+06	-	-	-	-	-	6A/B	6A/B/C	1.72E+05	12.9
337	6A/B/C	1.63E+06	-	1.41E+06	-	-	-	-	-	6A/B	6A/B/C	2.27E+05	13.9
341	6A/B/C	7.65E+06	-	5.86E+06	-	-	-	-	-	6A/B	6A/B/C, 35A	1.79E+06	23.4
343	15B	1.42E+06	-	-	-	-	-	1.20E+06	-	15B	15B/C	2.19E+05	15.4
371	9V/A	1.20E+07	-	-	-	-	1.09E+07	-	-	9V/A	9V/A	1.02E+06	8.5
470	6A/B/C/19F	3.13E+06	-	2.83E+06	-	-	-	-	-	6A/B	6A/B/C	2.97E+05	9.5
479	15B	1.25E+06	-	-	-	-	-	1.23E+06	-	15B	15B/C	2.20E+04	1.8
491[a]	19A	1.12E+06	9.06E+05	-	-	-	-	-	1.84E+06	19A	19A, 6A/B/C	-7.21E+05	-64.4
Average	**-**	**6.01E+06**	**4.23E+06**	**5.09E+06**	**9.33E+06**	**-**	**7.69E+06**	**1.23E+06**	**7.03E+06**	**-**	**-**	**-2.48E+03**	**-4.98**

[#]Molecular serotype of *S. pneumoniae* strain isolated from NP swabs.

*CFU/ml quantified by qPCR assays.

[a]These samples had a 2nd set of qPCRs run on them as the result of their THY Mulitplex results.

Note: Values shaded grey are the serotype-specific DNA values from the first set of qPCR runs. (for samples having more than one set of qPCR runs)

Note: Bolded samples were classified as non-controls, but did not meet cut-off of 2/3 (66.6%) between *lytA* and Ser-Spc DNA

Table 7. Serotypes Identified using Multiplex PCR and DNA Extracted from THY Broth Samples n=57[a]

Serotype(s)	Distribution (n)	%	Number of Serotypes
No Result	1	1.75	0
15B/C	4	7.02	1
19A	2	3.51	1
19F	8	14.04	1
6A/B/C	23	40.35	1
9V/A	1	1.75	1
15A/F, 10F/C/33C	1	1.75	2
15B/C, 22F/A	1	1.75	2
19A, 6A/B/C	1	1.75	2
19A, 7F	1	1.75	2
19F, 11A	1	1.75	2
22F/A, 35F/47F	1	1.75	2
6A/B/C, 10F/10C/33C	1	1.75	2
6A/B/C, 13	1	1.75	2
6A/B/C, 15B/C	1	1.75	2
6A/B/C, 16F	1	1.75	2
6A/B/C, 35A	1	1.75	2
9V/A, 6A/B/C	2	3.51	2
17F, 4, 31	1	1.75	3
6A/B/C, 15B/C, 13	1	1.75	3
6A/B/C, 15B/C, 23A	1	1.75	3
6A/B/C, 15B/C, 35F	1	1.75	3
6A/B/C, 19F, 13, 31	1	1.75	4

[a]Results from 15 samples with >66% difference (in *lytA* DNA and serotype-specific DNA) in qPCR results and set of 42 controls with < 66% difference.

Table 8. qPCR Results and Multiplex Results - Multiple Serotypes Findings

Single Colony		> 66% Difference*		Multiple Serotypes by THY Multiplex PCR?			Multiplex Multiple Serotype Samples have > 66% Difference?*		
Multiplex Serotype(s)	Distribution (n)	Yes? (n)	(%)	Number of Samples Run	Yes? (n)	(%)	Yes? (n)[a]	(%)	Agreement (Cohen's Kappa Statistic)[b]
15B/C	23	4	17.39	8	3	37.50	2	66.67	0.25
19A	14	2	14.29	6	3	50.00	1	33.33	0.00
19F	12	1	8.33	9	2	22.22	1	50.00	0.61
23F[c]	43	-	-	0	0	-	-	-	-
6A/B/C	45	6	13.33	28	7	25.00	4	57.14	0.36
6A/B/C & 19F	3	2	66.67	3	1	33.33	1	100.00	0.40
9V/A	9	0	0.00	3	2	66.67	0	0.00	0.00
Total	149	15	14.15[d]	57	18	31.58	9	50.00	0.30

*****>66% difference as defined as at least 2/3 difference between *lytA* DNA amount and Ser-Spc amount.

[a]Samples where multiple serotypes were found in THY Multiplex PCR and where there was a log difference between *lytA* DNA and serotype-specific DNA in qPCR.

[b]Where Agreement is classified as: <0 as indicating no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement.
Note: agreement is based solely on presence of multiple serotypes (not serotype identity).

[c]23F results were excluded due to abnormal chromatograms in majority of qPCR results

[d]Percentage calculated with the forty-three 23F samples deducted from the total number of samples run.

Table 9. Chi-Square Analysis of < 66% Difference and Multiple Serotypes by THY Multiplex PCR

		At least 66% Difference*		
		No	Yes	**Total**
Multiple Serotypes?	No	33	6	39
	Yes	9	9	18
	Total	42	15	57

Statistic	*DF*	*Value*	*Prob*
Chi-Square	1	7.6104	0.0058

*As defined by having at least 66% difference between *lytA* DNA and Ser-Spc DNA amounts

Table 10. Analysis of Maximum Likelihood Estimates For:

A. Total DNA amount (Tot_lyt) (predictor) and Multiple Serotypes (VarA) (outcome)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.6532	0.3865	2.8568	0.091
Tot_lyt	1	2.58E-06	5.77E-06	0.1995	**0.6551**

B. Total DNA amount (Tot_lyt) (predictor) and Number of Serotypes (VarF) (outcome)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 0	1	-4.068	1.0442	15.1781	<.0001
Intercept 1	1	0.7332	0.3816	3.6917	0.0547
Intercept 2	1	2.3024	0.5311	18.7899	<.0001
Intercept 3	1	3.9851	1.0392	14.7042	0.0001
Tot_lyt	1	8.60E-07	5.40E-06	0.0253	**0.8735**

Table 11. Total *lytA* DNA Amount across Groups of Samples Either Containing Multiple Serotypes or a Single Serotype

Grouping	Distribution (n)	Mean *lytA* DNA (CFU/ml)	95% *lytA* DNA CL Mean (CFU/ml) (L.B.)	(U.B.)	Std Dev	Two-Sample T-test p-value[a]
Single Serotype	39	5.00E+06	3.32E+06	6.67E+06	5.18E+06	
Multiple Serotypes	18	4.33E+06	1.63E+06	7.04E+06	5.43E+06	
Difference (N-Y)		6.61E+05	-2.34E+06	3.66E+06		**0.6609**

[a]Used pooled method as variances appeared to be equal according to the Folded F-test (p=0.7757).

Table 12. Multiple Serotypes and Level of Total *lytA* DNA

Grouping	Multiple Seroypes?		Total	Fisher Exact p-value
	No	*Yes*	*(n=57)*	
< 1,000,000 (1.E+06) CFU/ml	3	4	**7**	0.1908
> 1,000,000 (1.E+06) CFU/ml	36	14	**50**	
Total	**39**	**18**	**57**	
< 10,000,000 (1.E+07) CFU/ml	31	17	**48**	0.2467
> 10,000,000 (1.E+07) CFU/ml	8	1	**9**	
Total	**39**	**18**	**57**	

Table 13. Chi-Square Analyses of Sample Serotype and Presence of Multiple Serotypes

Serotype		Multiple Seroypes?		Total	p-value[a]
		No	*Yes*	*(n=57)*	
6A/B	*No*	16	10	26	0.3059
	Yes	23	8	31	
19F	*No*	30	15	45	0.5158
	Yes	9	3	12	
19A	*No*	36	15	51	0.3961
	Yes	3	3	6	
9V/A	*No*	38	16	54	0.5868
	Yes	1	2	3	
15B/C	*No*	34	15	49	0.4480
	Yes	5	3	8	

[a]All serotype grouping analyses except for that of 6A/B used Fisher's Exact Test statistic because of sparsity of data. Chi-square p-value was used for analysis of 6A/B grouping.

Table 14. Multiple Serotypes and Level of Difference between *lytA* DNA and Serotype-Specific DNA

Grouping	Multiple Serotypes?		Total	Chi-Square
	No	*Yes*	*(n=57)*	**p-value**
< 10,000 (1.E+04) CFU/ml	14	3	**17**	0.2144*
> 10,000 (1.E+04) CFU/ml	25	15	**40**	
Total	**39**	**18**	**57**	
< 100,000 (1.E+05) CFU/ml	16	6	**22**	0.5792
> 100,000 (1.E+05) CFU/ml	23	12	**35**	
Total	**39**	**18**	**57**	
< 1,000,000 (1.E+06) CFU/ml	30	9	**39**	**0.0421**
> 1,000,000 (1.E+06) CFU/ml	9	9	**18**	
Total	**39**	**18**	**57**	

*****Used Fisher exact test as definitive test as data was sparse (<5) in some cells.

Table 15. Sample Serotype and Association with Total *lytA* DNA Amount

Serotype	Distribution (n)	Mean *lytA* DNA (CFU/ml)	95% *lytA* DNA CL Mean (CFU/ml) (L.B.)	(U.B.)	Std Dev	Two-Sample T-test p-value
6A/B						
No?	58	2.10E+06	1.24E+06	2.97E+06	3.28E+06	
Yes?	48	3.73E+06	2.18E+06	5.28E+06	5.34E+06	
Difference (N-Y)		-1.63E+06	-3.39E+06	1.29E+05		0.0689*
19F						
No?	91	2.69E+06	1.75E+06	3.62E+06	4.48E+06	
Yes?	15	3.78E+06	1.67E+06	5.89E+06	3.81E+06	
Difference (N-Y)		-1.09E+06	-3.52E+06	1.34E+06	4.39E+06	0.3740
19A						
No?	92	3.01E+06	2.08E+06	3.95E+06	4.49E+06	
Yes?	14	1.70E+06	-3.57E+05	3.76E+06	3.57E+06	
Difference (N-Y)		1.31E+06	-1.18E+06	3.81E+06	4.39E+06	0.2999
9V/A						
No?	97	2.80E+06	1.91E+06	3.69E+06	4.43E+06	
Yes?	7	3.27E+06	8.47E+04	6.45E+06	4.14E+06	
Difference (N-Y)		-4.65E+05	-3.51E+06	2.58E+06	4.41E+06	0.7625
15B/C						
No?	83	3.35E+06	2.30E+06	4.40E+06	4.43E+06	
Yes?	23	1.02E+06	4.89E+05	1.55E+06	4.14E+06	
Difference (N-Y)		2.33E+06	1.16E+06	3.49E+06		**<0.0001***

*Used Satterthwaite differences and p-values, as p-values for equality of variance tests for these analyses were significant

Table 16. Sample Serotype and Difference between *lytA* DNA and Serotype-Specific DNA Amounts (p-values)

Serotype*	> 66% Difference[a]	Absolute Difference[b]		Underzero Difference[c]	
	P-Value	Mean Significantly Higher or Lower than others?	P-Value	Mean Significantly Higher or Lower than others?	P-Value
6A/B	0.4990[d]	**Higher**	**0.0003**	**Higher**	**0.0055**
19F	0.4424	-	0.1251	-	0.4905
19A	1.0000	**Lower**	**0.0165**	**Lower**	**<0.0001**
9V/A	0.2387	-	0.7844	**Lower**	**0.0292**
15B/C	0.7354	-	0.5476	-	0.0858

*N=106, Serotype 23F excluded from analyses

[a]Chi-square or fisher analyses of > 66% Difference variable and serotype dummy variable

[b]Two Sample t-test results of difference of *lytA* DNA and serotype-specific DNA (allowing for negative values) across serotype dummy varaibles

[c]Two Sample T-test Results of Difference of LytA DNA and serotype-specific DNA (**not** allowing for negative values) across serotype dummy varaibles

[d]Chi-square analysis, all others in column were fisher exact tests

Note: Bolded values indicate significant results at the 5% level of significance.

Table 17. Interaction Assessment of Initial Model

Analysis of Maximum Likelihood Estimates						
Parameter	**Model Parameter**	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	β0	1	5.605	109.8	0.0026	0.9593
LogDif	β1	1	4.6207	160.8	0.0008	0.9771
SixA	β2	1	-7.4508	109.8	0.0046	0.9459
NinF	β3	1	-7.5509	109.8	0.0047	0.9452
NinA	β4	1	-5.605	109.8	0.0026	0.9593
NiVA	β5	1	-4.9119	109.8	0.002	0.9643
FifB	β6	1	-6.7036	109.8	0.0037	0.9513
LogDif*SixA	γ1	1	-2.0817	160.8	0.0002	**0.9897**
LogDif*NinF	γ2	1	6.8578	109.8	0.0039	**0.9502**
LogDif*NinA	γ3	1	-4.6207	160.8	0.0008	**0.9771**
LogDif*NiVA	γ4	0	0	.	.	**.**
LogDif*FifB	γ5	1	-3.5221	160.8	0.0005	**0.9825**

Initial Model:
logit(p=Multiple Serotypes)=β0 +β1(Log_dif) +β2(SixA)+β3(NinF)+β4(NinA)
+β5(NiVA)+β6(FifB)+γ1(Log_dif*SixA) +γ2(Log_dif*NinF) +γ3(Log_dif*NinA)+γ4(Log_dif*NiVA)

Resulting Model After Interaction Terms Dropped:
logit(p=Multiple Serotypes)=β0 +β1(Log_dif) +β2(SixA)+β3(NinF)+β4(NinA) +β5(NiVA)+β6(FifB)

Table 18. Confounding & Precision Assessment (Gold Standard Model)

Gold Standard Model: logit(p=Mult. Ser.)=β0 +β1(Log_dif) +β2(SixA)+β3(NinF)+β4(NinA) +β5(NiVA)+β6(FifB)										
Analysis of Maximum Likelihood Estimates							Odds Ratio Estimates			
Parameter	Model Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Point Estimate	95% Wald Confidence Limits		95% CI Width
Intercept	β0	1	1.0147	1.6522	0.3772	0.5391				
LogDif	β1	1	-2.0226	0.7403	7.4639	**0.0063**	0.132	0.031	0.565	0.534
SixA	β2	1	0.6573	1.6393	0.1608	0.6885				
NinF	β3	1	0.5473	1.4449	0.1435	0.7049				
NinA	β4	1	-0.3912	1.9183	0.0416	0.8384				
NiVA	β5	1	-1.7079	2.0566	0.6896	0.4063				
FifB	β6	1	0.6438	1.9657	0.1073	0.7433				

Table 19. Confounding & Precision Assessment (Reduced Model)

Reduced Model: logit(p=Mult. Ser.)=β0 +β1(Log_dif)

Analysis of Maximum Likelihood Estimates							Odds Ratio Estimates			
Parameter	Model Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Point Estimate	95% Wald Confidence Limits		95% CI Width
Intercept	β0	1	1.2993	0.3761	11.9375	0.0006				
LogDif	β1	1	-1.7047	0.6475	6.9328	**0.0085**	0.182	0.051	0.647	0.596

Table 20. Method Comparisons for Determining Serotypes in Nasopharyngeal *S. pneumoniae* Samples*

	Quelling	Multiplex PCR[a]	Standard qPCR	RSPH Method
Objective: Single Serotype per Sample				
Number of Reactions per Sample	93 (1 panel)	1-8	1-93	1-6
Number of Reactions per 150 Samples	13950 (150 panels)	150-1200	150-13950	150-900
Relative Cost	+	+	+++	+
Relative Speed	+	+	+++	+++
Predictive Power	+++	++	+++	++
Objective: All Serotypes in Sample				
Number of Reactions per Sample	9300 (100 panels)[b]	8	93	6
Number of Reactions per 150 Samples	1395000 (15000 panels)	1200	13950	900
Reactions Used in Samples with Single Serotype[c, d]	1171800 (12600 panels)	1008	11718	756
Relative Cost (for many samples)	+++	+	+++	+
Relative Speed (for many samples)	+	++	++	+++
Predictive Power	+++	++	+++	++

* (+=low, ++=mid-range, +++=high)

[a]Does not detects all serotype because low sensitivity

[b]Requires the isolation of at least 100 colonies from the sample

[c]Based of RSPH results, 84% of samples contain single serotype

[d]For 150 samples

Table 21. Most Common Serotypes in Children Across Latin America and in Peru[a]

Type	Invasive Pneumococcal Disease Studies				Carriage Studies			
Country	Latin America	Peru			Brazil	Venezuela		Peru
Serotype Rank	Castenada et al. (2009)	Cullotta et al. (2002)	Ochea et al. (2005)	Ochoa et al. (2010)	Laval et al. (2006)	Rivera-Olivero et al. (2007)	Bello Gonzalez et al. (2010)	RSPH Current Study
1	14 (28.9)	**23F (13.7)**	**19 (19.9)**	14 (26.3)	14 (14.9)	**23F (19.7)**	**6B (48)**	**6A/B/C (20.6)**
2	**6B (9.2)**	**6A (10.3)**	**6 (19.0)**	**6B (20.2)**	**6B (13.1)**	**6A (19.7)**	33F (21.5)	**23F (8.7)**
3	1 (7.5)	**6B (10.3)**	**23 (15.2)**	**19F (11.1)**	**6A (10.8)**	**15B (10.5)**	**6A (6.0)**	**15B/C (7.4)**
4	5 (6.9)	**15B (8.9)**	**15 (8.5)**	**23F (6.1)**	**19F (8.6)**	**6B (9.2)**	**19A (3.1)**	**19F (6.1)**
5	18C (6.0)	**19F (8.2)**	14 (5.2)	5 (6.1)	10A (6.8)	**19F (7.2)**	**23F (3.1)**	**19A (4.2)**
6	**19F (6.0)**	14 (7.5)	**9 (3.8)**	**6A (4.0)**	**23F (6.3)**	11 (7.2)	**-**	**9V/A (2.9)**
7	**23F (4.7)**	34 (3.4)	33 (2.8)	**19A (4.0)**	18C (5.0)	14 (4.6)	**-**	18A/B/C (2.6)
8	**6A (3.7)**	23B (2.7)	-	**9V (1.0)**	**19A (4.5)**	**19A (3.3)**	**-**	4 (1.9)
9	**19A (3.7)**	**9V (3 2)**	-	18C (1.0)	9N (4.0)	**9V (2.6)**	**-**	-
10	7F (3.3)	**19A (2.1)**	-	3 (1.0)	18A (4.0)	10 (2.0)	**-**	-
11	**9V (2.9)**	21 (2.1)	-	4 (1.0)	**9V(3.6)**	18C (1.3)	**-**	-
12	3 (2.1)	**9A (0.7)[b]**	-	-	-	-	**-**	-
13	4 (1.5)	**15C (0.7)[b]**	-	-	-	-	**-**	-
Other	Other/NT (11.7)	Other (18.9)	-	-	-	-	**-**	Other (30.2)
NT	-	NT (8.9)	-	-	Other (18.4)	Other (12.5)	Other/NT (18.3)	NT (15.4)

[a]Table format is Serotype (Percentage of Serotype within study sample population).

[b]Rank for both these serotypes in distribution is 21 amongst other serotypes found in study.

Appendices

Appendix A. Quantitative PCR Methodology Walkthrough

Walkthrough covers sample plate set up and data analysis using Bio-Rad software that is specific to the Rollins School of Public Health "RSPH" Method Developed in this thesis. All screenshots were obtained using the Bio-Rad CFX Manager software version 1.6.541.1028 (Bio-Rad Laboratories). In this example walkthrough, 9 nasopharyngeal samples previously identified by Chelex multiplex PCR to contain serotype 9V/A are examined to confirm multiplex PCR results and to estimate if multiple serotypes are likely to be contained within the sample.

1. Reaction Mixture Preparation

Separate reaction mixtures using *lytA* primers/probes and 9V/A primers/probes are prepared in parallel to examine these 9 NP samples. A *lytA* run set and a Serotype-Specific run set are needed to quantify DNA and confirm serotype for these NP samples. Each set includes: 5 standards, 1 No Template Control, and 9 samples (Total of 15 reactions per set). 9V/A DNA was used for the preparation of both sets of standards. In the tables below reaction reagent volumes needed for each set are shown:

			lytA
Invitrogen-Platinum Quantitative PCR SuperMix-UDG	**Per Rxn**	**# of Rxn**	**17**
	12.5	µl	212.5
H2O	8.5	µl	144.5
Primer F (10uM)	0.5	µl	8.5
Primer R (10uM)	0.5	µl	8.5
Probe (10uM)	0.5	µl	8.5
Total Reagent Volume	22.5	µl	382.5
DNA/H2O Volume	2.5	µl	
Total Single Reaction Volume	**25**	**µl**	

			9V/A
Invitrogen-Platinum Quantitative PCR SuperMix-UDG	**Per Rxn**	**# of Rxn**	**17**
	12.5	µl	212.5
H2O	8.5	µl	144.5
Primer F (10uM)	0.5	µl	8.5
Primer R (10uM)	0.5	µl	8.5
Probe (10uM)	0.5	µl	8.5
Total Reagent Volume	22.5	µl	382.5
DNA/H2O Volume	2.5	µl	
Total Single Reaction Volume	**25**	**µl**	

NOTE: Reaction mixes are prepared with two extra reactions added as there is volume loss when transferring reaction mix to sample plate (15 rxn's needed + 2 extra rxn's= 17)

2. Plate Set-Up

Up to 96 reactions per run can be performed. In this example only 30 reactions total are being run. Plate set-up can vary based on number of standards needed and diagnostic goals of run. Sample Plate set up is shown below:

	LytA Primers and Probes						9V/A Primers and Probes					
	1	2	3	4	5	6	7	8	9	10	11	12
A	LytA Std 1 (400pg/ml)	LytA Sample 2					9V/A Std 1 (400pg/ml)	9V/A Sample 2				
B	LytA Std 2 (40 pg/ml)	LytA Sample 3					9V/A Std 2 (40 pg/ml)	9V/A Sample 3				
C	LytA Std 3 (4 pg/ml)	LytA Sample 4					9V/A Std 3 (4 pg/ml)	9V/A Sample 4				
D	LytA Std 4 (0.4 pg/ml)	LytA Sample 5					9V/A Std 4 (0.4 pg/ml)	9V/A Sample 5				
E	LytA Std 5 (0.04pg/ml)	LytA Sample 6					9V/A Std 5 (0.04pg/ml)	9V/A Sample 6				
F	LytA NTC	LytA Sample 7					9V/A NTC	9V/A Sample 7				
G		LytA Sample 8						9V/A Sample 8				
H	LytA Sample 1	LytA Sample 9					9V/A Sample 1	9V/A Sample 9				

Note: Std=Standard, NTC=No Template Control

3. qPCR Run Conditions

Once samples are loaded onto plate, the run conditions for the qPCRs were as follows: 50ºC for 2 min, 95ºC for 2 min, followed by 40 cycles of 95ºC for 15 sec and 60ºC for 1 min.

4. Bio-Rad Software Run Plate Set-up

Once Bio-Rad Software has been open, create new plate file. *S. pneumoniae* DNA can be detected using the FAM fluorescence settings. Click the "Select Flourophones" button and check the "FAM" button. Plate set-up in software should correspond to how samples were loaded in the actual run plate. Designate sample type using the drop down box on the left. When loading Standards, sample concentrations need to be loaded as well. For Std 1 concentration is set to 4.40×10^5, Std 2 is 4.40×10^4, Std 3 is 4.40×10^3, and so on.
An example of completed software plate is shown below:

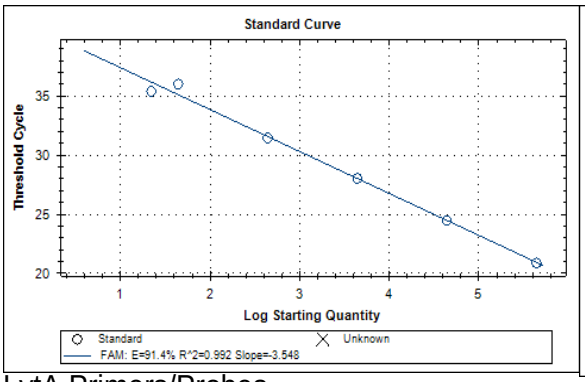
	1	2	3	4	5	6	7	8	9	10	11	12
A	Std lytA	Unk LytA 87					Std 9VA	Unk 9VA 87				
B	Std lytA	Unk LytA 90					Std 9VA	Unk 9VA 90				
C	Std lytA	Unk LytA 92					Std 9VA	Unk 9VA 92				
D	Std lytA	Unk LytA 122					Std 9VA	Unk 9VA 122				
E	Std lytA	Unk LytA 235					Std 9VA	Unk 9VA 235				
F	NTC lytA	Unk LytA 266					NTC 9VA	Unk 9VA 266				
G		Unk LytA 371						Unk 9VA 371				
H	Unk LytA 31	Unk LytA 494					Unk 9VA 31	Unk 9VA 494				

Note: Plate files should be saved with only one set of standards per file. In the example above there should be two separate plates files saved containing only the samples for the primers/probes of interest.

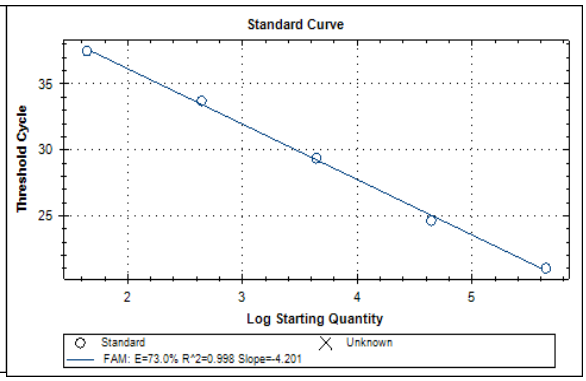
5. Standard Curve Analysis

Once qPCR run data has been retrieved, open a “Stand-alone Run” from the file menu of the Bio-Rad software menu. Select run file and plate of interest. De-select all samples, so that only curves for standards and NTC’s are shown. Repeat the process for the other primer/probe set used. Run slope, efficiency, and CT values should be checked and compared to ensure that they are fairly equivalent amongst both primer/probe sets. Ideally, run efficiency should be as close to 100% as possible and slope should be as close to 1.0 as possible. It is also important to note that any samples that had CT values above the threshold of 35, were considered negative (containing *S. pneumoniae* DNA that was below the detectable limit.) Below are screenshots used for this example:

9V/A Standards with LytA Primers/Probes



9V/A Standards with 9V/A Primers/Probes



LytA Primers/Probes

Well	Fluor	Content	Target	Sample	Threshold Cycle (Ct(t))	Ct(t) Mean	Ct(t) Std. Dev	Starting Quantity (SQ)	Log Starting Quantity	SQ Mean	SQ Std. Dev
F01	FAM	NTC	lytA		N/A	0.00	0.000	N/A	N/A	0.00E+00	0.00E+00
A01	FAM	Std	lytA		20.83	20.83	0.000	4.400E+05	5.643	4.40E+05	0.00E+00
B01	FAM	Std	lytA		24.59	24.59	0.000	4.400E+04	4.643	4.40E+04	0.00E+00
C01	FAM	Std	lytA		28.82	28.82	0.000	4.400E+03	3.643	4.40E+03	0.00E+00
D01	FAM	Std	lytA		33.16	33.16	0.000	4.400E+02	2.643	4.40E+02	0.00E+00
E01	FAM	Std	lytA		35.54	35.54	0.000	4.400E+01	1.643	4.40E+01	0.00E+00

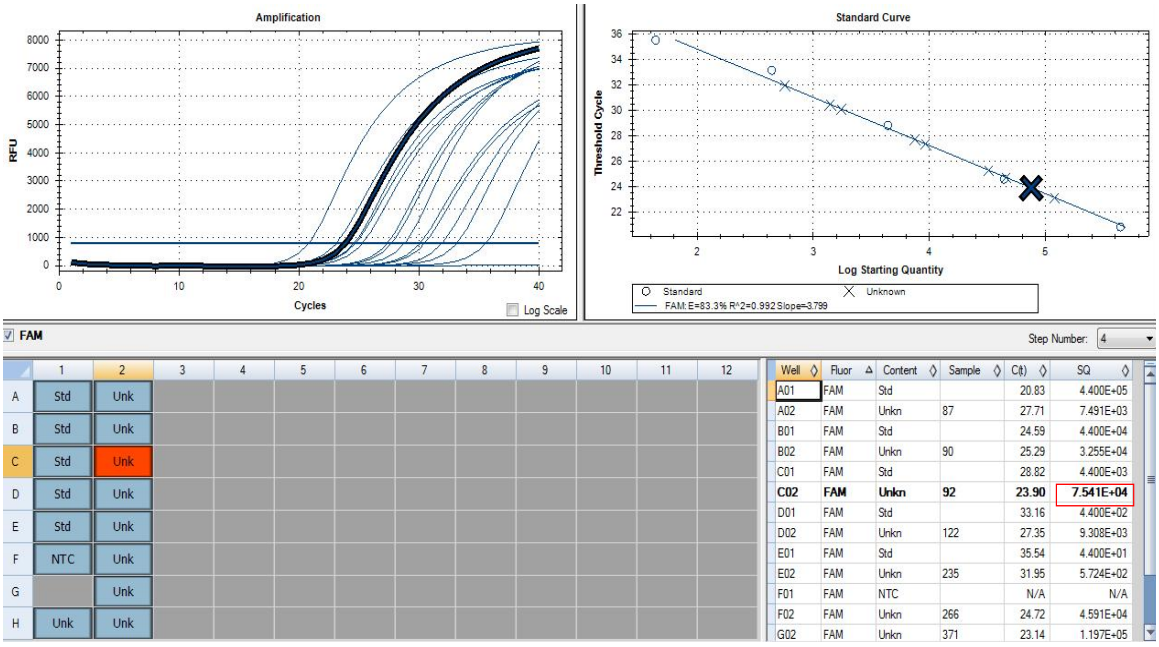
9V/A Primers/Probes

Well	Fluor	Content	Target	Sample	Threshold Cycle (Ct(t))	Ct(t) Mean	Ct(t) Std. Dev	Starting Quantity (SQ)	Log Starting Quantity	SQ Mean	SQ Std. Dev
F07	FAM	NTC	9VA		N/A	0.00	0.000	N/A	N/A	0.00E+00	0.00E+00
A07	FAM	Std	9VA		21.03	21.03	0.000	4.400E+05	5.643	4.40E+05	0.00E+00
B07	FAM	Std	9VA		24.64	24.64	0.000	4.400E+04	4.643	4.40E+04	0.00E+00
C07	FAM	Std	9VA		29.37	29.37	0.000	4.400E+03	3.643	4.40E+03	0.00E+00
D07	FAM	Std	9VA		33.72	33.72	0.000	4.400E+02	2.643	4.40E+02	0.00E+00
E07	FAM	Std	9VA		37.49	37.49	0.000	4.400E+01	1.643	4.40E+01	0.00E+00

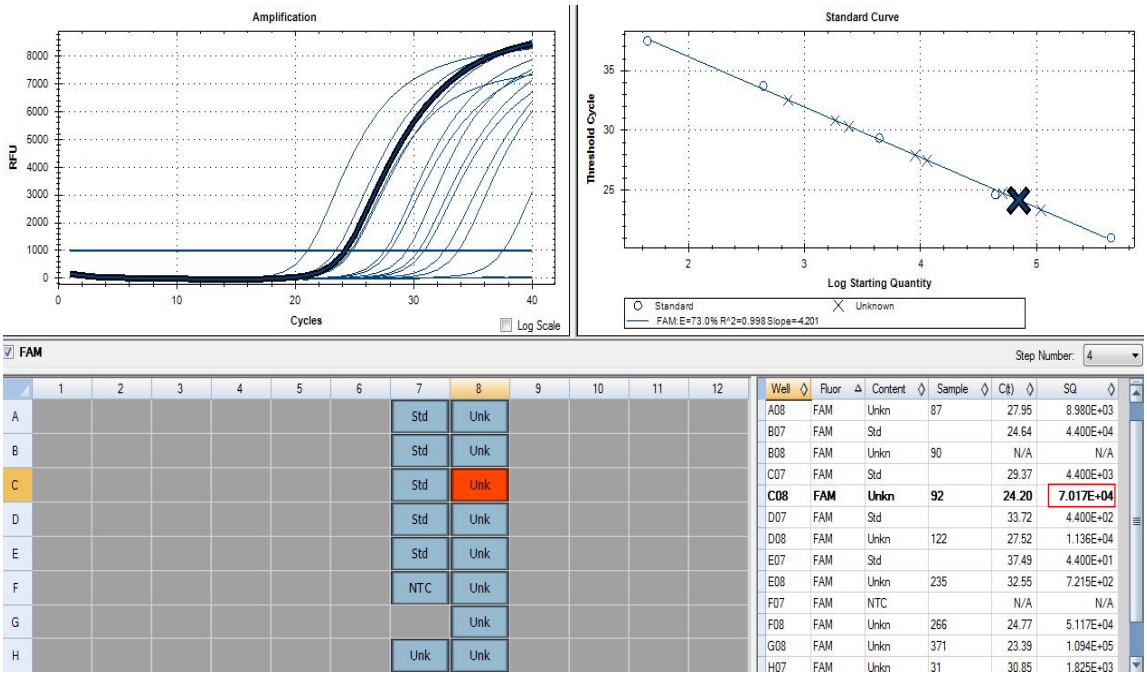
6. Multiple Serotype Prediction

For each run, reselect the samples so their corresponding curves re-appear on the screen. For each sample CFU amounts are shown in the SQ columns. The CFU value represents the number of colony forming units in the sample and is a proxy for the amount of DNA in the NP sample. By comparing the CFU values between the *lytA* runs and serotype-specific runs for each sample, an estimate of multiple serotype status can be made. For this thesis, > 66% difference between *lytA* and Ser-Spc CFU led to a prediction of multiple serotypes within a NP sample. Table 1 presents the possible prediction scenarios that can results from this method. Below are some screenshots for sample 92, which, based upon our prediction algorithm, contained one serotype as it did not have difference of > 66% between its *lytA* and Ser-Spc CFU values:

Sample Screenshot LytA



Sample Screenshot 9V/A



Appendix B: List of oligonucleotide primers used for pneumococcal serotype deduction

Primers*	GenBank accession no.	Primer sequence (5'-3')	Gene	Nucleotide position	Product size (bp)	Reference
1-f	CR931632	CTC TAT AGA ATG GAG TAT ATA AAC TAT GGT TA	*wzy*	9935	280	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
1-r		CCA AAG AAA ATA CTA ACA TTA TCA CAA TAT TGG C		10181		
2-f	CR931633	TAT CCC AGT TCA ATA TTT CTC CAC TAC ACC	*wzy*	10271	290	New
2-r		ACA CAA AAT ATA GGC AGA GAG AGA CTA CT		10531		
3-f	CR931634	ATG GTG TGA TTT CTC CTA GAT TGG AAA GTA G	*gal* U	9020	371	
3-r		CTT CTC CAA TTG CTT ACC AAG TGC AAT AAC G		9360		
4-f [a]	CR931635	CTG TTA CTT GTT CTG GAC TCT CGA TAA TTG G	*wzy*	9596	430	
4-r		GCC CAC TCC TGT TAA AAT CCT ACC CGC ATT G		9995		Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
5-f	CR931637	ATA CCT ACA CAA CTT CTG ATT ATG CCT TTG TG	*wzy*	6123	362	
5-r		GCT CGA TAA ACA TAA TCA ATA TTT GAA AAA GTA TG		6450		
6A/6B/6C-f	CR931639	AAT TTG TAT TTT ATT CAT GCC TAT ATC TGG	*wci* P	8656	250	
6A/6B/6C-r		TTA GCG GAG ATA ATT TAA AAT GAT GAC TA		8877		
6C-f	EU714777.1	CAT TTT AGT GAA GTT GGC GGT GGA GTT	*wci* Nbeta	1441	727	Carvalho *et al* . 2009, J. Clin. Microbiol. 47: 557-559
6C-r		AGC TTC GAA GCC CAT ACT CTT CAA TTA		2141		
7C/(7B/40)-f	CR931642	CTA TCT CAG TCA TCT ATT GTT AAA GTT TAC GAC GGG A	*wcw* L	9438	260	Pai et al. 2006, J. Clin. Microbiol. 44: 124-131
7C/(7B/40)-r		GAA CAT AGA TGT TGA GAC ATC TTT TGT AAT TTC		9665		
7F/7A-f	CR931643	TCC AAA CTA TTA CAG TGG GAA TTA CGG	*wzy*	14683	599	New
7F/7A-r		ATA GGA ATT GAG ATT GCC AAA GCG AC		15256		
8-f	CR931644	GAA GAA ACG AAA CTG TCA GAG CAT TTA CAT	*wzy*	11193	201	New
8-r		CTA TAG ATA CTA GTA GAG CTG TTC TAG TCT		11364		
9N/9L-f	CR931647	GAA CTG AAT AAG TCA GAT TTA ATC AGC	*wzx*	11948	516	Dias *et al* . 2007, J. Med. Microbiol. 56: 1185-1189
9N/9L-r		ACC AAG ATC TGA CGG GCT AAT CAA T		12439		
9V/9A-f	CR931648	GGG TTC AAA G TC AGA CAG TG A ATC TTA A	*wzy*	9966	816	New
9V/9A-r		CCA TGA ATG A AA TCA ACA TT G TCA GTA GC		10753		
10A- f	CR931649	GGT GTA GAT TTA CCA TTA GTG TCG GCA GAC	*wcr* G	12423	628	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
10A-r		GAA TTT CTT CTT TAA GAT TCG GAT ATT TCT C		13020		

10F/(10C/33C)- f	CR931652	GGA GTT TAT CGG TAG TGC TCA TTT TAG CA	*wzx*	12403	248	New
10F/(10C/33C)-r		CTA ACA AAT TCG CAA CAC GAG GCA ACA		12624		
11A/11D-f	CR931653	GGA CAT GTT CAG GTG ATT TCC CAA TAT AGT G	*wzy*	11640	463	
11A/11D-r		GAT TAT GAG TGT AAT TTA TTC CAA CTT CTC CC		12071		Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
12F/(12A/44/46)-f	CR931660	GCA ACA AAC GGC GTG AAA GTA GTT G	*wzx*	14407	376	
12F/(12A/44/46)-r		CAA GAT GAA TAT CAC TAC CAA TAA CAA AAC		14753		
13-f	CR931661	TAC TAA GGT AAT CTC TGG AAA TCG AAA GG	*wzx*	14005	655	New
13-r		CTC ATG CAT TTT ATT AAC CG C TTT TTG TTC		14630		
14-f	CR931662	GAA ATG TTA CTT GGC GCA GGT GTC AGA ATT	*wzy*	7959	189	Dias *et al* . 2007, J. Med. Microbiol. 56: 1185-1189
14-r		GCC AAT ACT TCT TAG TCT CTC AGA TGA AT		8119		
15A/15F-f	CR931663	ATT AGT ACA GCT GCT GGA ATA TCT CTT C	*wzy*	7804	434	
15A/15F-r		GAT CTA GTG AAC GTA CTA TTC CAA AC		8212		Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
15B/15C-f	CR931665	TTG GAA TTT TTT AAT TAG TGG CTT ACC TA	*wzy*	7314	496	
15B/15C-r		CAT CCG CTT ATT AAT TGA AGT AAT CTG AAC C		7779		
16F-f,	CR931668	GAA TTT TTC AGG CGT GGG TGT TAA AAG	*wzy*	11679	717	New
16F-r		CAG CAT ATA GCA CCG CTA AGC AAA TA		12371		
17F-f	CR931670	TTC GTG ATG ATA ATT CCA ATG ATC AAA CAA GAG	*wci* P	10484	693	
17F-r		GAT GTA ACA AAT TTG TAG CGA CTA AGG TCT GC		11145		Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
18/(18A/18B/18C/18F)-f	CR931673	CTT AAT AGC TCT CAT TAT TCT TTT TTT AAG CC	*wzy*	12687	573	
18/(18A/18B/18C/18F)-r		TTA TCT GTA AAC CAT ATC AGC ATC TGA AAC		13230		
19A-f	CR931675	GAG AGA TTC ATA ATC TTG CAC TTA GCC A	*wzy*	9603	566	Pimenta *et al* . 2009, J. Clin. Microbiol. *in press*
19A-r		CAT AAT AGC TAC AAA TGA CTC ATC GCC		10142		
19F-f	CR931678	GTT AAG ATT GCT GAT CGA TTA ATT GAT ATC C	*wzy*	11135	304	
19F-r		GTA ATA TGT CTT TAG GGC GTT TAT GGC GAT AG		11407		Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
20-f	CR931679	GAG CAA GAG TTT TTC ACC TGA CAG CGA GAA G	*wci* L	9567	514	
20-r		CTA AAT TCC TGT AAT TTA GCT AAA ACT CTT ATC		10048		
21-f	CR931680	CTA TGG TTA TTT CAA CTC AAT CGT CAC C	*wzx*	13247	192	New
21-r		GGC AAA CTC AGA CAT AGT ATA GCA TAG		13412		

22F/22A-f	CR931682	GAG TAT AGC CAG ATT ATG GCA GTT TTA TTG TC	*wcw* V	11055	643	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
22F/22A-r		CTC CAG CAC TTG CGC TGG AAA CAA CAG ACA AC		11666		
23A-f	CR931683	TAT TCT AGC AAG TGA CGA AGA TGC G	*wzy*	7739	722	New
23A-r		CCA ACA TGC TTA AAA ACG CTG CTT TAC		8434		
23B-f	CR931684	CCA CAA TTA G CG CTA TAT TCA TTC AAT CG	*wzx*	13227	199	
23B-r		GTC CAC GCT GAA TAA AAT GAA GCT CCG		13399		
23F-f [a]	CR931685	GTA ACA GTT GCT GTA GAG GGA ATT GGC TTT TC	*wzy*	8768	384	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
23F-r		CAC AAC ACC TAA CAC TCG ATG GCT ATA TGA TTC		9119		
24/(24A, 24B, 24F)-f	CR931688	GCT CCC TGC TAT TGT AAT CTT TAA AGA G	*wzy*	11701	99	New
24/(24A, 24B, 24F)-r		GTG TCT TTT ATT GAC TTT ATC ATA GGT CGG		11770		
31-f	CR931695	GGA AGT TTT CAA GGA TAT GAT AGT GGT GGT GC	*wzy*	9144	701	
31-r		CCG AAT AAT ATA TTC AAT ATA TTC CTA CTC		9815		
33F/(33A/37)-f	CR931702	GAA GGC AAT CAA TGT GAT TGT GTC GCG	*wzy*	11129	338	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
33F/(33A/37)-r		CTT CAA AAT GAA GAT TAT AGT ACC CTT CTA C		11436		
34-f	CR931703	GCT TTT GTA AGA GGA GAT TAT TTT CAC CCA AC	wzy	7350	408	
34-r		CAA TCC GAC TAA GTC TTC AGT AAA AAA CTT TAC		7725		
35A/(35C/42)-f	CR931704	ATT ACG ACT CCT TAT GTG ACG CGC ATA	*wzx*	14394	280	New
35A/(35C/42)-r		CCA ATC CCA AGA TAT ATG CAA CTA GGT T		14646		
35B-f	CR931705	GAT AAG TCT GTT GTG GAG ACT TAA AAA GAA TG	*wcr* H	10556	677	
35B-r		CTT TCC AGA TAA TTA CAG GTA TTC CTG AAG CAA G		11199		
35F/47F-f	CR931707	GAA CAT AGT CGC TAT TGT ATT TTA TTT AAA GCA A	wzy	7374	517	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
35F/47F-r		GAC TAG GAG CAT TAT TCC TAG AGC GAG TAA ACC		7858		
38/25F-f	CR931710	CGT TCT TTT ATC TCA CTG TAT AGT ATC TTT ATG	wzy	13848	574	
38/25F-r		ATG TTT GAA TTA AAG CTA ACG TAA CAA TCC		14392		
39-f	CR931711	TCA TTG TAT TAA CCC TAT GCT TTA TTG GTG	wzy	12289	98	New
39-r		GAG TAT CTC CAT TGT ATT GAA ATC TAC CAA		12357		
cps A-f	CR931662	GCA GTA CAG CAG TTT GTT GGA CTG ACC	*wzg*	1473	160	Pai *et al* . 2006, J. Clin. Microbiol. 44: 124-131
cps A-r		GAA TAT TTT CAT TAT CAG TCC CAG TC		1607		

*All serotypes that are co-detected are listed

Appendix C. Raw Data & Findings (1 of 4)

NP Sample Number	Serotype[#]	Total *lytA* DNA 1st run * (w/ 6A/B stds)	Total *lytA* DNA 2nd run * (w/ 19F stds)	Total *lytA* DNA 3rd run * (w/ 9V/A stds)	Total *lytA* DNA 4th run* (w/ 6A/B stds)	Total *lytA* DNA 5th run* (w/ 15B/C stds)	Mean *lytA* DNA Amount*	*lytA* 95% CI L.B.*	*lytA* 95% CI U.B.*
1	6A/B/C	7.08E+05	-	-	-	-	7.08E+05	.	.
4	19A	1.76E+05	-	-	-	-	1.76E+05	.	.
5	6A/B/C	9.35E+05	-	-	-	-	9.35E+05	.	.
8	23F	2.83E+06	-	-	-	-	2.83E+06	.	.
9	6A/B/C	1.15E+06	-	5.63E+05	2.53E+06	-	1.42E+06	2.72E+05	2.56E+06
10	6A/B/C	1.29E+07	-	-	-	-	1.29E+07	.	.
11	6A/B/C, 19F	4.73E+06	4.38E+06	-	1.07E+07	-	6.60E+06	1.62E+06	1.16E+07
17	6A/B/C	1.28E+06	-	-	-	-	1.28E+06	.	.
18	6A/B/C	3.24E+06	-	-	-	-	3.24E+06	.	.
24	6A/B/C	3.39E+05	-	-	-	-	3.39E+05	.	.
28	19F	1.72E+06	-	5.86E+05	3.12E+06	-	1.81E+06	2.96E+05	3.32E+06
29	6A/B/C	7.00E+06	-	-	-	-	7.00E+06	.	.
31	9V/A	1.40E+05	-	-	-	-	1.40E+05	.	.
34	23F	8.16E+06	-	-	-	-	8.16E+06	.	.
41	6A/B/C	1.89E+06	-	-	-	-	1.89E+06	.	.
44	6A/B/C	3.39E+05	-	-	-	-	3.39E+05	.	.
45	15B/C	2.67E+06	-	-	2.20E+07	-	1.24E+07	-9.87E+05	2.57E+07
46	19F	1.32E+07	-	-	-	-	1.32E+07	.	.
50	6A/B/C	9.45E+06	-	-	-	-	9.45E+06	.	.
52	6A/B/C	2.85E+06	-	-	-	-	2.85E+06	.	.
56	23F	1.24E+07	-	-	-	-	1.24E+07	.	.
60	6A/B/C	7.46E+06	-	-	-	-	7.46E+06	.	.
61	15B/C	7.14E+04	-	-	-	-	7.14E+04	.	.
65	19F	5.67E+06	-	-	-	-	5.67E+06	.	.
67	23F	3.09E+06	-	-	-	-	3.09E+06	.	.
72	15B/C	8.23E+05	-	-	-	-	8.23E+05	.	.
74	6A/B/C	1.83E+07	-	-	-	-	1.83E+07	.	.
75	15B/C	5.01E+06	-	-	-	4.95E+06	4.98E+06	1.20E+06	8.75E+06
82	23F	4.52E+06	-	-	-	-	4.52E+06	.	.
83	23F	1.28E+05	-	-	-	-	1.28E+05	.	.
85	19A	1.13E+06	-	-	-	-	1.13E+06	.	.
86	19A	2.60E+04	-	-	-	-	2.60E+04	.	.
87	9V/A	7.49E+05	-	-	-	-	7.49E+05	.	.
90	9V/A	3.26E+06	-	-	-	-	3.26E+06	.	.
92	9V/A	7.54E+06	-	-	-	-	7.54E+06	.	.
94	19F	1.32E+06	-	-	-	-	1.32E+06	.	.
96	23F	4.80E+05	-	-	-	-	4.80E+05	.	.
98	23F	8.08E+06	-	-	-	-	8.08E+06	.	.
99	15B/C	7.15E+05	-	-	-	-	7.15E+05	.	.
102	6A/B/C	8.11E+06	-	-	-	9.06E+05	4.51E+06	-4.08E+05	9.42E+06
106	19F	2.81E+06	-	-	-	-	2.81E+06	.	.
109	6A/B/C	1.53E+07	-	-	-	-	1.53E+07	.	.
110	6A/B/C	1.08E+06	-	-	-	-	1.08E+06	.	.
122	9V/A	9.31E+05	-	-	-	-	9.31E+05	.	.
123	19A	7.64E+03	-	-	-	-	7.64E+03	.	.
131	6A/B/C	2.30E+04	-	-	-	-	2.30E+04	.	.
133	23F	1.19E+07	-	-	-	-	1.19E+07	.	.
134	19A	8.03E+06	-	-	-	-	8.03E+06	.	.
136	6A/B/C	1.96E+06	-	-	-	-	1.96E+06	.	.
142	23F	1.80E+06	-	-	-	-	1.80E+06	.	.
145	15B/C	9.23E+05	-	-	-	-	9.23E+05	.	.
156	6A/B/C	1.14E+05	-	4.89E+04	-	-	8.15E+04	1.19E+04	1.51E+05
170	23F	2.68E+06	-	-	-	-	2.68E+06	.	.
173	6A/B/C	3.74E+06	-	-	-	-	3.74E+06	.	.
181	15B/C	6.27E+05	-	-	-	-	6.27E+05	.	.
182	6A/B/C	1.20E+05	-	-	-	-	1.20E+05	.	.
189	6A/B/C	9.25E+05	-	-	-	-	9.25E+05	.	.
190	6A/B/C	2.33E+07	-	-	-	5.44E+06	1.44E+07	3.84E+05	2.83E+07
192	6A/B/C	6.82E+06	-	-	-	-	6.82E+06	.	.
194	6A/B/C	2.27E+05	-	-	-	-	2.27E+05	.	.
206	6A/B/C, 19F	3.54E+06	-	-	5.26E+06	-	4.40E+06	9.55E+05	7.84E+06
207	6A/B/C	2.32E+06	-	9.39E+05	-	-	1.63E+06	2.19E+05	3.04E+06
208	15B/C	2.25E+05	-	-	-	-	2.25E+05	.	.
212	19F	5.30E+06	-	-	-	-	5.30E+06	.	.
218	19F	2.14E+06	-	-	-	-	2.14E+06	.	.
219	23F	3.55E+05	-	-	-	-	3.55E+05	.	.
220	6A/B/C	2.76E+04	-	-	-	-	2.76E+04	.	.
221	19A	1.50E+05	-	-	-	-	1.50E+05	.	.
225	19F	6.05E+05	-	-	-	-	6.05E+05	.	.
227	23F	2.84E+06	-	-	-	-	2.84E+06	.	.
229	15B/C	1.06E+05	-	-	6.40E+05	-	3.73E+05	-1.24E+04	7.58E+05
234	15B/C	7.85E+05	-	-	-	-	7.85E+05	.	.
235	9V/A	5.72E+04	-	-	-	-	5.72E+04	.	.
236	6A/B/C	1.22E+06	-	-	-	-	1.22E+06	.	.
244	6A/B/C	1.17E+06	-	-	-	-	1.17E+06	.	.
246	23F	2.97E+07	-	-	-	-	2.97E+07	.	.
249	6A/B/C	1.16E+06	-	4.13E+05	-	-	7.87E+05	8.62E+04	1.49E+06
250	15B/C	2.29E+05	-	-	-	-	2.29E+05	.	.

[#]Molecular serotype of *S. pneumoniae* strain isolated from NP swabs.

*CFU/ml quantified by qPCR assays.

[a]Only samples with log difference=1 were examined for the complete set of primers and probes {6A/B, 19F, 9V/A, 15B/C, 19A} by qPCR.

[b]Serotypes detected in NP swabs using qPCR assays targeting serotypes 6A/B, 19F, 23F, 9V/A, 15B/C and 19A, (missing values indicate no serotype-specific DNA was found in qPCR Run 1)

[c]Serotypes detected by multiplex PCR using DNA extracted from enrichment broth (only 57 samples were examined by this method).

Appendix C. Raw Data & Findings (2 of 4)

NP Sample Number	Serotype[#]	Total *lytA* DNA 1st run * (w/ 6A/B stds)	Total *lytA* DNA 2nd run * (w/ 19F stds)	Total *lytA* DNA 3rd run * (w/ 9V/A stds)	Total *lytA* DNA 4th run* (w/ 6A/B stds)	Total *lytA* DNA 5th run* (w/ 15B/C stds)	Mean *lytA* DNA Amount*	*lytA* 95% CI L.B.*	*lytA* 95% CI U.B.*
252	19F	1.88E+04	-	-	-	-	1.88E+04	.	.
255	6A/B/C	2.47E+05	-	-	-	-	2.47E+05	.	.
256	6A/B/C	3.37E+05	-	-	-	-	3.37E+05	.	.
257	23F	6.22E+06	-	-	-	-	6.22E+06	.	.
258	6A/B/C	1.86E+05	-	-	-	-	1.86E+05	.	.
261	19F	2.48E+05	-	-	-	-	2.48E+05	.	.
264	19A	1.18E+07	-	-	-	-	1.18E+07	.	.
266	9V/A	4.59E+06	-	-	-	5.82E+06	5.21E+06	1.21E+06	9.20E+06
268	6A/B/C	1.08E+06	-	3.50E+05	-	-	7.17E+05	6.48E+04	1.37E+06
269	15B/C	8.23E+03	-	-	-	-	8.23E+03	.	.
270	15B/C	3.63E+06	-	-	-	-	3.63E+06	.	.
271	19A	5.22E+05	-	-	-	-	5.22E+05	.	.
274	19F	1.35E+06	-	-	-	-	1.35E+06	.	.
279	23F	1.82E+06	-	-	-	-	1.82E+06	.	.
282	19A	1.36E+05	-	-	6.36E+05	-	3.86E+05	4.25E+03	7.68E+05
283	6A/B/C	8.15E+05	-	-	-	-	8.15E+05	.	.
284	19A	2.20E+04	-	-	7.02E+04	-	4.61E+04	3.88E+03	8.83E+04
288	6A/B/C	1.37E+06	-	5.18E+05	-	-	9.45E+05	1.14E+05	1.78E+06
291	6A/B/C	4.29E+05	-	-	-	-	4.29E+05	.	.
292	6A/B/C	4.07E+05	-	1.83E+05	-	-	2.95E+05	4.57E+04	5.45E+05
296	6A/B/C	1.61E+07	-	-	-	-	1.61E+07	.	.
297	15B/C	1.41E+06	-	-	-	-	1.41E+06	.	.
299	19F	1.09E+07	-	-	-	-	1.09E+07	.	.
305	23F	1.65E+06	-	-	-	-	1.65E+06	.	.
311	6A/B/C	7.56E+05	-	3.44E+05	-	-	5.50E+05	8.62E+04	1.01E+06
318	15B/C	5.13E+04	-	-	-	-	5.13E+04	.	.
319	15B/C	9.60E+05	-	-	-	-	9.60E+05	.	.
328	6A/B/C	7.17E+04	-	-	-	-	7.17E+04	.	.
335	6A/B/C	1.33E+06	-	-	-	-	1.33E+06	.	.
337	6A/B/C	1.63E+06	-	-	-	-	1.63E+06	.	.
341	6A/B/C	7.65E+06	-	-	-	-	7.65E+06	.	.
343	15B/C	1.42E+06	-	-	-	-	1.42E+06	.	.
351	23F	4.37E+06	-	-	-	-	4.37E+06	.	.
352	23F	1.43E+05	-	-	-	-	1.43E+05	.	.
371	9V/A	1.20E+07	-	-	-	-	1.20E+07	.	.
385	15B/C	4.24E+05	-	-	-	-	4.24E+05	.	.
397	23F	8.90E+05	-	-	-	-	8.90E+05	.	.
398	19A	2.79E+05	-	-	-	-	2.79E+05	.	.
400	15B/C	4.28E+05	-	-	-	-	4.28E+05	.	.
417	19A	1.61E+05	-	-	-	-	1.61E+05	.	.
419	19A	3.20E+05	-	-	-	-	3.20E+05	.	.
429	15B/C	1.19E+06	-	-	8.78E+06	-	4.98E+06	-3.23E+05	1.03E+07
434	23F	1.51E+05	-	-	-	-	1.51E+05	.	.
435	15B/C	3.76E+05	-	-	-	-	3.76E+05	.	.
438	23F	4.16E+04	-	-	-	-	4.16E+04	.	.
456	23F	4.16E+05	-	-	-	-	4.16E+05	.	.
460	23F	1.33E+05	-	-	-	-	1.33E+05	.	.
469	15B/C	1.11E+05	-	-	7.01E+05	-	4.06E+05	-1.65E+04	8.29E+05
470	6A/B/C, 19F	3.13E+06	-	-	-	-	3.13E+06	.	.
471	23F	8.80E+05	-	-	-	-	8.80E+05	.	.
479	15B/C	1.25E+06	-	-	-	-	1.25E+06	.	.
481	23F	4.11E+06	-	-	-	-	4.11E+06	.	.
491	19A	1.12E+06	-	-	-	9.06E+05	1.01E+06	2.37E+05	1.79E+06
494	9V/A	1.74E+05	-	-	-	-	1.74E+05	.	.
497	23F	9.01E+06	-	-	-	-	9.01E+06	.	.
514	23F	7.68E+06	-	-	-	-	7.68E+06	.	.
516	23F	7.25E+06	-	-	-	-	7.25E+06	.	.
528	23F	1.82E+06	-	-	-	-	1.82E+06	.	.
532	23F	6.32E+05	-	-	-	-	6.32E+05	.	.
548	23F	5.48E+06	-	-	-	-	5.48E+06	.	.
602	23F	4.27E+05	-	-	-	-	4.27E+05	.	.
665	23F	1.84E+07	-	-	-	-	1.84E+07	.	.
668	23F	2.81E+06	-	-	-	-	2.81E+06	.	.
669	23F	1.61E+07	-	-	-	-	1.61E+07	.	.
722	23F	5.24E+05	-	-	-	-	5.24E+05	.	.
724	23F	4.05E+06	-	-	-	-	4.05E+06	.	.
730	23F	6.43E+06	-	-	-	-	6.43E+06	.	.
735	23F	1.00E+07	-	-	-	-	1.00E+07	.	.
739	23F	1.41E+07	-	-	-	-	1.41E+07	.	.
740	23F	1.18E+06	-	-	-	-	1.18E+06	.	.
742	23F	9.88E+06	-	-	-	-	9.88E+06	-	-

[#]Molecular serotype of *S. pneumoniae* strain isolated from NP swabs.

*CFU/ml quantified by qPCR assays.

[a]Only samples with log difference=1 were examined for the complete set of primers and probes {6A/B, 19F, 9V/A, 15B/C, 19A} by qPCR.

[b]Serotypes detected in NP swabs using qPCR assays targeting serotypes 6A/B, 19F, 23F, 9V/A, 15B/C and 19A, (missing values indicate no serotype-specific DNA was found in qPCR Run 1)

[c]Serotypes detected by multiplex PCR using DNA extracted from enrichment broth (only 57 samples were examined by this method).

Appendix C. Raw Data & Findings (3 of 4)

NP Sample Number	6A/B*	19F*	23F*	9V/A*	15B/C*	19A*	> 66% Difference in Run 1 (1=yes, 0=no)[a]	Serotypes detected by qPCR[b]	Serotypes detected by multiplex PCR[c]
1	2.98E+05	-	-	-	-	-	0	6A/B	-
4	-	-	-	-	-	2.73E+06	0	19A	-
5	1.76E+05	-	-	-	-	-	0	6A/B	-
8	-	-	3.78E+05	-	-	-	0	23F	-
9	5.82E+05	8.56E+03	-	-	-	3.88E+03	0	6A/B	6A/B/C
10	1.05E+07	-	-	-	-	-	0	6A/B	6A/B/C
11	2.52E+05	6.86E+06	-	-	3.74E+03	1.32E+04	1	6A/B, 19F	19F
17	-	-	-	-	-	-	0	-	-
18	-	-	-	-	-	-	0	-	-
24	3.22E+05	-	-	-	-	-	0	6A/B	-
28	4.72E+03	3.54E+04	-	-	-	2.47E+03	1	19F	17F, 4, 31
29	7.60E+06	-	-	-	-	-	0	6A/B	6A/B/C
31	-	-	-	1.83E+05	-	-	0	9V/A	-
34	-	-	2.94E+05	-	-	-	0	23F	-
41	1.97E+06	-	-	-	-	-	0	6A/B	6A/B/C
44	2.36E+05	-	-	-	-	-	0	6A/B	-
45	1.90E+06	2.21E+04	2.59E+05	3.94E+03	9.36E+05	1.42E+05	1	6A/B, 15B/C	6A/B/C, 15B/C, 13
46	-	2.54E+07	-	-	-	-	0	19F	19F
50	8.36E+06	-	-	-	-	-	0	6A/B	6A/B/C
52	2.26E+06	-	-	-	-	-	0	6A/B	6A/B/C, 16F
56	-	-	1.19E+07	-	-	-	0	23F	-
60	2.83E+06	-	-	-	-	-	0	6A/B	6A/B/C
61	-	-	-	-	2.17E+04	-	0	15B/C	-
65	-	2.99E+06	-	-	-	-	0	19F	19F, 11A
67	-	-	4.50E+05	-	-	-	1	23F	-
72	-	-	-	-	8.20E+05	-	0	15B/C	-
74	1.66E+07	-	-	-	-	-	0	6A/B	6A/B/C
75	-	-	-	-	2.45E+06	-	0	15B/C	6A/B/C, 15B/C
82	-	-	2.08E+05	-	-	-	0	23F	-
83	-	-	3.69E+04	-	-	-	1	23F	-
85	-	-	-	-	-	1.91E+06	0	19A	19A
86	-	-	-	-	-	2.32E+04	0	19A	-
87	-	-	-	8.98E+05	-	-	0	9V/A	-
90	-	-	-	-	-	-	0	-	-
92	-	-	-	7.02E+06	-	-	0	9V/A, 6A/B	9V/A, 6A/B/C
94	-	2.25E+06	-	-	-	-	0	19F	19F
96	-	-	8.28E+05	-	-	-	0	23F	-
98	-	-	-	-	-	-	0	-	-
99	-	-	-	-	5.95E+05	-	0	15B/C	-
102	6.79E+06	-	-	-	-	-	0	6A/B	6A/B/C
106	-	5.46E+06	-	-	-	-	0	19F	19F
109	1.45E+07	-	-	-	-	-	0	6A/B	6A/B/C
110	1.07E+06	-	-	-	-	-	0	6A/B	6A/B/C
122	-	-	-	1.14E+06	-	-	0	9V/A	-
123	-	-	-	-	-	3.48E+04	0	19A	-
131	1.71E+04	-	-	-	-	-	0	6A/B	-
133	-	-	2.04E+06	-	-	-	1	23F	-
134	-	-	-	-	-	9.44E+06	0	19A	19A, 7F
136	1.99E+06	-	-	-	-	-	0	6A/B	6A/B/C
142	-	-	2.74E+06	-	-	-	0	23F	-
145	-	-	-	-	8.91E+05	-	0	15B/C	-
156	1.67E+04	-	-	-	1.86E+04	2.41E+03	1	6A/B	15A/F, 10F/C/33C
170	-	-	1.95E+05	-	-	-	1	23F	-
173	1.29E+06	-	-	-	-	-	0	6A/B	6A/B/C
181	-	-	-	-	5.55E+05	-	0	15B/C	-
182	8.68E+04	-	-	-	-	-	0	6A/B	-
189	9.30E+05	-	-	-	-	-	0	6A/B	-
190	1.27E+07	-	-	-	-	-	0	6A/B, 15B/C	6A/B/C, 15B/C, 35F
192	6.34E+06	-	-	-	-	-	0	6A/B	6A/B/C
194	3.08E+05	-	-	-	-	-	0	6A/B	-
206	4.95E+05	3.14E+06	2.17E+05	-	3.00E+03	5.73E+03	1	6A/B, 19F	6A/B/C, 19F, 13, 31
207	1.70E+05	-	-	-	-	3.98E+03	1	6A/B	6A/B/C, 13
208	-	-	-	-	1.83E+05	-	0	15B/C	-
212	-	1.10E+07	-	-	-	-	0	19F	19F
218	-	4.03E+06	-	-	-	-	0	19F	19F
219	-	-	6.07E+05	-	-	-	0	23F	-
220	3.95E+04	-	-	-	-	-	0	6A/B	-
221	-	-	-	-	-	2.67E+05	0	19A	-
225	-	1.19E+06	-	-	-	-	0	19F	-
227	-	-	6.93E+05	-	-	-	0	23F	-
229	3.58E+03	-	-	-	7.21E+04	-	1	15B/C	15B/C, 22F/A
234	-	-	-	-	-	-	0	-	-
235	-	-	-	7.21E+04	-	-	0	9V/A	-
236	1.34E+06	-	-	-	-	-	0	6A/B	6A/B/C
244	1.02E+06	-	-	-	-	-	0	6A/B	6A/B/C
246	-	-	3.41E+06	-	-	-	1	23F	-
249	2.93E+05	-	-	-	-	-	0	6A/B	6A/B/C
250	-	-	-	-	1.79E+05	-	0	15B/C	-

[#]Molecular serotype of *S. pneumoniae* strain isolated from NP swabs.

*CFU/ml quantified by qPCR assays.

[a]Only samples with log difference=1 were examined for the complete set of primers and probes {6A/B, 19F, 9V/A, 15B/C, 19A} by qPCR.

[b]Serotypes detected in NP swabs using qPCR assays targeting serotypes 6A/B, 19F, 23F, 9V/A, 15B/C and 19A, (missing values indicate no serotype-specific DNA was found in qPCR Run 1)

[c]Serotypes detected by multiplex PCR using DNA extracted from enrichment broth (only 57 samples were examined by this method).

Appendix C. Raw Data & Findings (4 of 4)

NP Sample Number	6A/B*	19F*	23F*	9V/A*	15B/C*	19A*	> 66% Difference in Run 1 (1=yes, 0=no)[a]	Serotypes detected by qPCR[b]	Serotypes detected by multiplex PCR[c]
252	-	-	-	-	-	-	0	-	-
255	2.03E+05	-	-	-	-	-	0	6A/B	-
256	3.29E+05	-	-	-	-	-	0	6A/B	-
257	-	-	6.12E+05	-	-	-	1	23F	-
258	1.89E+05	-	-	-	-	-	0	6A/B	-
261	-	4.15E+05	-	-	-	-	0	19F	-
264	-	-	-	-	-	1.49E+07	0	19A	19A
266	-	-	-	5.12E+06	-	-	0	9V/A	9V/A, 6A/B/C
268	2.24E+05	-	-	-	-	-	1	6A/B	6A/B/C
269	-	-	-	-	3.87E+03	-	0	15B/C	-
270	-	-	-	-	-	-	0	-	-
271	-	-	-	-	-	8.81E+05	0	19A	-
274	-	2.61E+06	-	-	-	-	0	19F	19F
279	-	-	3.64E+04	-	-	-	1	23F	-
282	3.00E+05	-	-	-	-	2.07E+04	1	6A/B, 19A	6A/B/C
283	-	-	-	-	-	-	0	-	-
284	-	9.10E+03	-	-	-	6.98E+03	1	19A	22F/A, 35F/47F
288	2.37E+05	-	-	-	-	5.48E+03	1	6A/B	6A/B/C, 10F/10C/33C
291	1.87E+05	-	-	-	-	-	0	6A/B	-
292	3.96E+04	-	-	-	-	4.64E+03	1	6A/B	6A/B/C, 15B/C, 23A
296	1.15E+07	-	-	-	-	-	0	6A/B	6A/B/C
297	-	-	-	-	1.05E+06	-	0	15B/C	15B/C
299	-	2.10E+07	-	-	-	-	0	19F	19F
305	-	-	1.53E+05	-	-	-	1	23F	-
311	6.33E+04	-	-	-	-	-	1	6A/B	6A/B/C
318	-	-	-	-	4.58E+04	-	0	15B/C	-
319	-	-	-	-	-	-	0	-	-
328	7.55E+04	-	-	-	-	-	0	6A/B	-
335	1.16E+06	-	-	-	-	-	0	6A/B	6A/B/C
337	1.41E+06	-	-	-	-	-	0	6A/B	6A/B/C
341	5.86E+06	-	-	-	-	-	0	6A/B	6A/B/C, 35A
343	-	-	-	-	1.20E+06	-	0	15B/C	15B/C
351	-	-	5.82E+06	-	-	-	0	23F	-
352	-	-	-	-	-	-	0	-	-
371	-	-	-	1.09E+07	-	-	0	9V/A	9V/A
385	-	-	-	-	3.71E+05	-	0	15B/C	-
397	-	-	1.39E+06	-	-	-	0	23F	-
398	-	-	-	-	-	5.84E+05	0	19A	-
400	-	-	-	-	3.75E+05	-	0	15B/C	-
417	-	-	-	-	-	2.21E+05	0	19A	-
419	-	-	-	-	-	4.38E+05	0	19A	-
429	7.17E+03	6.58E+03	-	5.57E+03	6.17E+05	5.39E+03	1	15B/C	No Result
434	-	-	-	-	-	-	0	-	-
435	-	-	-	-	2.51E+05	-	0	15B/C	-
438	-	-	-	-	-	-	0	-	-
456	-	-	1.06E+05	-	-	-	0	23F	-
460	-	-	-	-	-	-	0	-	-
469	6.35E+03	4.09E+04	-	-	8.21E+04	5.38E+03	1	15B/C	15B/C
470	2.83E+06	-	-	-	-	-	0	6A/B	6A/B/C
471	-	-	1.29E+06	-	-	-	0	23F	-
479	-	-	-	-	1.23E+06	-	0	15B/C	15B/C
481	-	-	2.96E+05	-	-	-	1	23F	-
491	-	-	-	-	-	1.84E+06	0	19A	19A, 6A/B/C
494	-	-	-	2.40E+05	-	-	0	9V/A	-
497	-	-	5.63E+05	-	-	-	1	23F	-
514	-	-	4.21E+05	-	-	-	1	23F	-
516	-	-	6.42E+05	-	-	-	1	23F	-
528	-	-	1.75E+06	-	-	-	0	23F	-
532	-	-	7.16E+04	-	-	-	1	23F	-
548	-	-	6.34E+05	-	-	-	1	23F	-
602	-	-	9.03E+04	-	-	-	1	23F	-
665	-	-	2.06E+06	-	-	-	1	23F	-
668	-	-	4.01E+05	-	-	-	1	23F	-
669	-	-	1.58E+06	-	-	-	1	23F	-
722	-	-	4.68E+05	-	-	-	0	23F	-
724	-	-	5.41E+05	-	-	-	1	23F	-
730	-	-	-	-	-	-	0	-	-
735	-	-	9.97E+05	-	-	-	1	23F	-
739	-	-	1.39E+06	-	-	-	1	23F	-
740	-	-	1.16E+05	-	-	-	1	23F	-
742	-	-	9.51E+05	-	-	-	1	23F	-

[#]Molecular serotype of *S. pneumoniae* strain isolated from NP swabs.

*CFU/ml quantified by qPCR assays.

[a]Only samples with log difference=1 were examined for the complete set of primers and probes {6A/B, 19F, 9V/A, 15B/C, 19A} by qPCR.

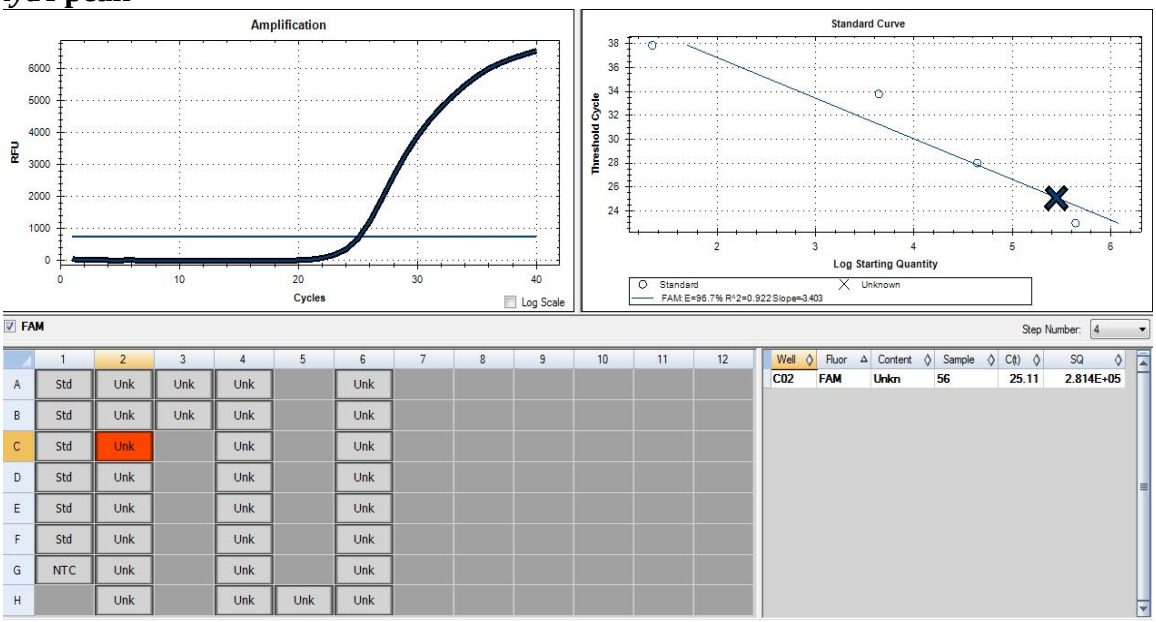
[b]Serotypes detected in NP swabs using qPCR assays targeting serotypes 6A/B, 19F, 23F, 9V/A, 15B/C and 19A, (missing values indicate no serotype-specific DNA was found in qPCR Run 1)

[c]Serotypes detected by multiplex PCR using DNA extracted from enrichment broth (only 57 samples were examined by this method).

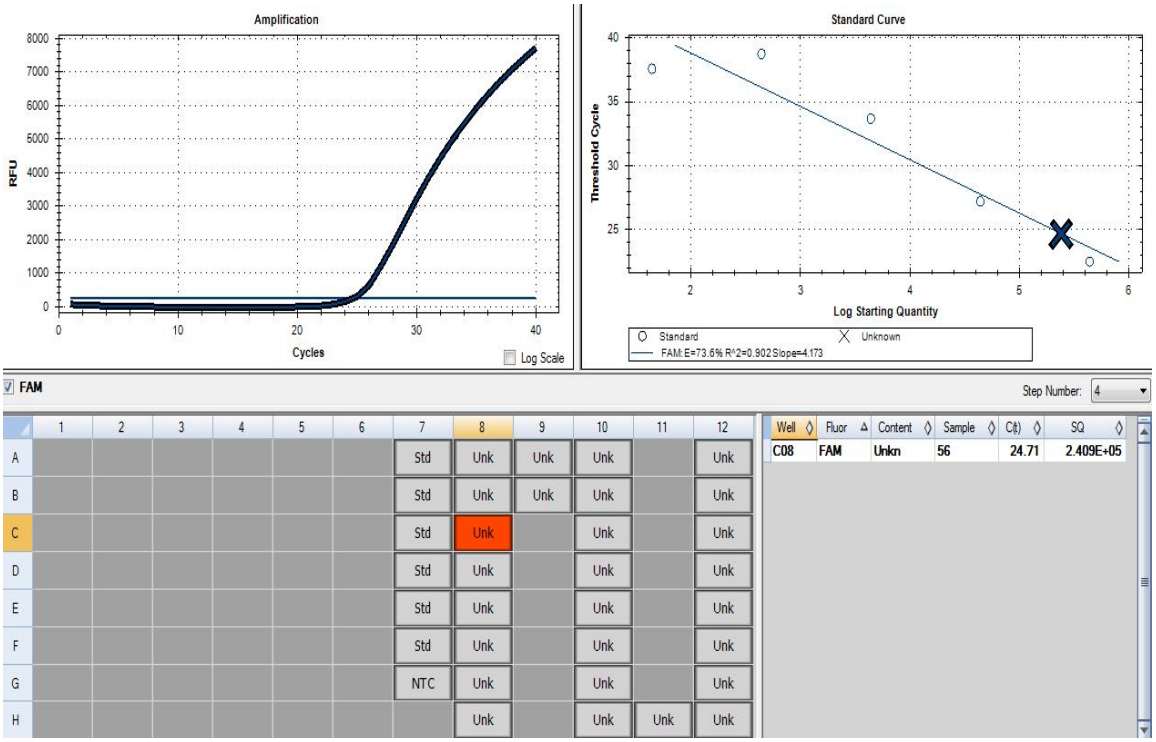
Appendix D. Typical vs. Atypical Serotype 23F qPCR Screenshots

Below are the screenshots from the Bio-Rad CFX Manager software (Bio-Rad Laboratories) for a sample with typical *lytA* and serotype-specific qPCR peaks (sample 56):

lytA peak

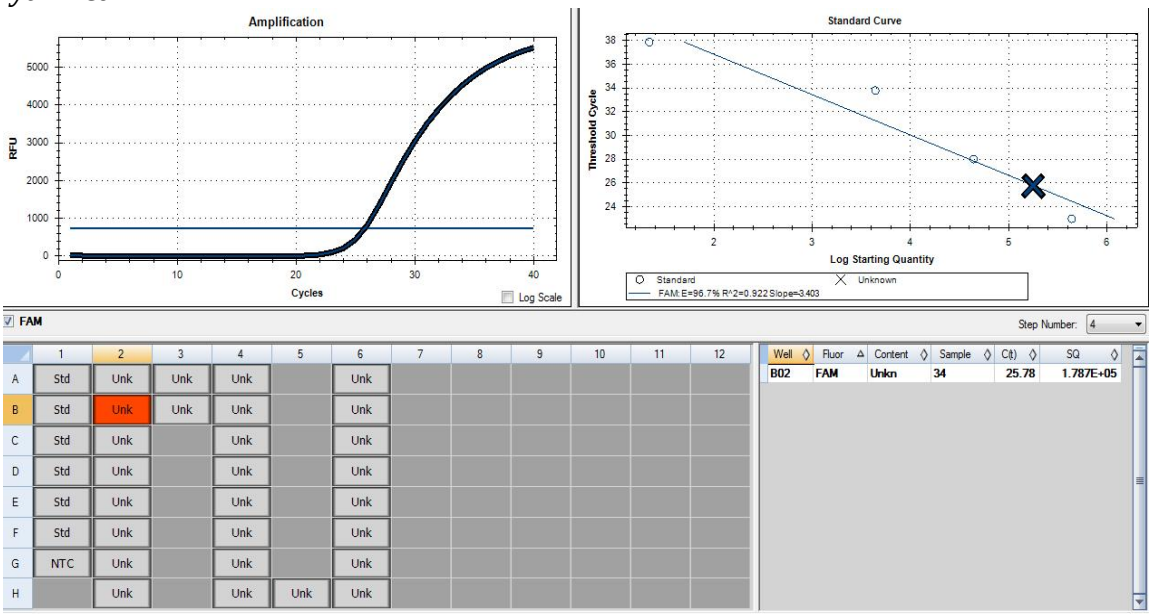


23F Peak

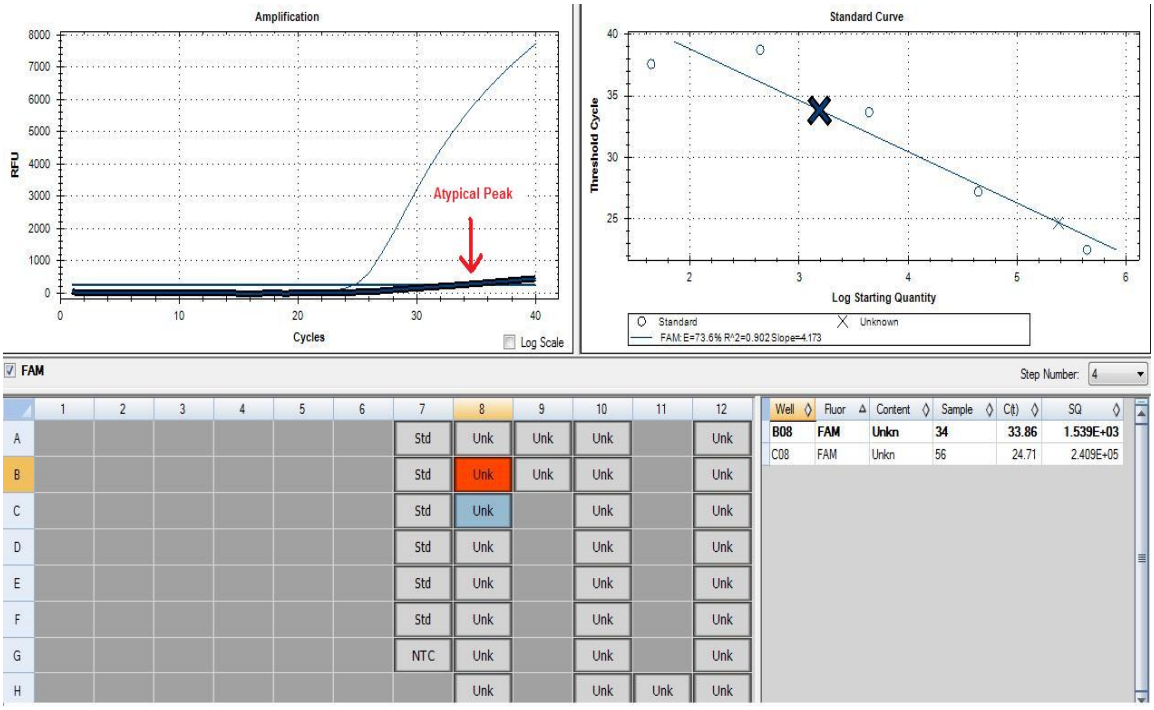


Below are the screenshots from the Bio-Rad CFX Manager software (Bio-Rad Laboratories) for a sample with a typical *lytA* peak but an atypical serotype-specific qPCR peak (sample 34):

lytA Peak



23F Peak



The 23F peak appears to be below the CT threshold of 35 for negative results (so it would have shown up as positive for containing serotype 23F). When contrasted to a typical 23F positive peak (shown above), we can see that this atypical 23F peak is not behaving

normally. When contrasted to a typical negative result (shown below), the atypical 23F peak also does not appear to correspond to a negative fluorescence pattern.
Below is the screenshot for a typical negative qPCR result: