Pneumococcal Regional Serotype Distribution for Pneumococcal AMC TPP

Codebook to assess whether a pneumococcal vaccine meets the Pneumococcal AMC Target Product Profile for regional vaccine serotype coverage

Prepared by GAVI's PneumoADIP
Department of International Health
Johns Hopkins Bloomberg School of Public Health

11/30/2008

Contact: Kate O'Brien, MD, MPH
GAVI's PneumoADIP
Johns Hopkins Bloomberg School of Public Health
621 N. Washington St.,
Baltimore, MD 21205
410-955-6931
klobrien@jhsph.edu

Table of Contents

1. Abbreviations 3
2. Executive Summary 4
3. Introduction 7
4. Methods 9
5. Results 12
   Table 1. Summary of serotype data in the analysis. 14
   Figure 1. African serotypes by rank order and cumulative serotype distribution. 15
   Figure 2. Asian serotypes by rank order and cumulative serotype distribution. 16
   Figure 3. European serotypes by rank order and cumulative serotype distribution. 17
   Figure 4. LAC serotypes by rank order and cumulative serotype distribution. 18
   Figure 5. North American serotypes by rank order and cumulative serotype distribution. 19
   Figure 6. Oceania serotypes by rank order and cumulative serotype distribution. 20
   Figure 7. Cumulative regional serotype distribution. 21
   Table 2. Serotype distribution by region. 22
   Table 3. Comparison of serotypes with ICC <0.3, by region. 23
6. Discussion 24
7. Acknowledgements 28

Appendix 1. Countries by World Bank region. 33
Appendix 2. Description of serotype data in analysis. 36

1. Abbreviations

AMC – Advance Market Commitment
CDC – US Centers for Disease Control and Prevention
GAVI – GAVI Alliance
GSP – Global Serotype Project
LAC – Latin America and the Caribbean
PCV – Pneumococcal Conjugate Vaccine
TPP – Target Product Profile
WHO – World Health Organization

2. Executive Summary

Despite the availability of a safe and effective pneumococcal conjugate vaccine (PCV), financial barriers limit access to these life-saving vaccines for children in the poorest countries---places where the risk of pneumococcal disease is highest. An Advance Market Commitment (AMC) of $1.5 billion for advancement of pneumococcal vaccines was announced in 2007 as a novel funding mechanism for accelerating the introduction of pneumococcal vaccines into the world's poorest countries. A Target Product Profile (TPP) describing minimum characteristics required for a pneumococcal vaccine to be eligible for AMC financing was developed by WHO in support of the AMC; the serotype distribution of disease causing strains was a part of the TPP.

At the request of WHO, GAVI's PneumoADIP at Johns Hopkins University undertook to estimate the serotype distribution of invasive pneumococcal disease among children <5 years of age globally and by region. An External Expert Committee (EEC) and ad hoc consultations with other relevant experts provided technical review and advice on the statistical methods, model design, data collection, literature review, and interpretation of analyses. Version 1 of the analysis was presented in September 2007 and Version 2, a pre-planned analysis designed to include additional data and apply refined analytic methods, is presented here. As stipulated by the TPP Expert Committee, the Version 2 GSP output serves as the definitive vaccine serotype analysis against which pneumococcal conjugate vaccines will be assessed for AMC eligibility throughout the course of the AMC.

The primary purpose of this analysis of serotypes is to guide the evaluation of vaccine formulations meeting the TPP. Serotype analyses and projections of the expected health impact of vaccination with specific formulations are also important for decisions pertaining to country adoption of pneumococcal vaccines. For this reason, the GSP may be updated in the future to incorporate additional serotype data and/or further refinements to the analytic methods when these become available. Furthermore, while regional estimates from meta-analyses will help to inform local decisions, the actual health impact in a particular country for a particular year may vary from projections based on regional estimates.

Published and unpublished data from 1980 through June 2007 were systematically reviewed to identify studies with at least 20 serotyped invasive pneumococcal isolates; data from 169 studies, 70 countries and 60,090 isolates made up the Version 2 analysis dataset. Serotype

data from individual countries were combined according to World Bank regions; there was a sizeable number of isolates included in the analysis from every region (range: 3,649-18,788), with the fewest number of isolates from Oceania (3,649) and Asia (4,752).

The methods used in this analysis were determined in consultation with an External Expert Committee including expertise in biostatistics, epidemiology, pediatrics, infectious diseases, and microbiology and in consultation with the TPP Committee as the modeling approach that best addressed the data and analytic requirements. A marginal random effects meta-analysis model was determined to be the best choice for estimating the proportion of pneumococcal invasive disease among young children attributable to a given serotype (with 95% confidence intervals). Unlike a basic weighted average (a.k.a. fixed-effects) model that reflects the distributions observed in larger studies, the random-effects assigns weighting based on the variance within and between the studies. Other important methodological attributes of the model include the inclusion of studies with relatively small sample size to allow for the greatest regional representation within the analysis; the need for a Bayesian estimation to allow for the inclusion of observations with a zero value; normalization of the estimated proportions within a region to assure they summed to 100%; and the decision to model the 21 most relevant serotypes out of the 90 possible types.

It is important to understand these methodological features because they result in a directional bias that systematically underestimates the most common serotypes and overestimates the less common serotypes. For any individual serotype the degree of bias is relatively small, but this effect becomes more pronounced when the proportions of invasive disease estimated for the most common serotypes are summed (e.g., the cumulative proportion of disease due to the top 8-13 serotypes). Therefore the estimates of cumulative proportions presented in Version 2 should be interpreted as minimum point estimates with the degree of underestimation increasing as the number of serotypes included in the cumulative proportion increases. This conservative bias was accepted because it increases the likelihood that AMC eligible vaccines will meet or exceed the expected levels of health impact.

The key findings from the analysis are as follows:

- Serotype 14 is the most common serotype in each region of the world among children < 5 years.

- Serotypes 1 and 5 are among the top 3 ranked serotypes occurring in the GAVI-eligible countries and are among the top 6 ranked serotypes occurring among children < 5 years in regions with the highest pneumococcal disease burden (Africa, Asia, and LAC).

- The seven most common disease causing serotypes in Africa are identical to those in Asia. The rank order of these seven serotypes differs between the two regions and the cumulative proportion of all serotypes included in these top seven is somewhat higher in Africa than in Asia (65% and 61% respectively).

- A limited set of seven serotypes (1, 5, 6A, 6B, 14, 19F, 23F) are common and important in all regions with a substantial number of GAVI eligible countries. Together these 7 serotypes account for ~58%-66% of all invasive pneumococcal disease in each region.

- With as few as 6-7 serotypes (including serotypes 1, 5, and 14, and assuming that 6B conjugate vaccine provides cross protection against 6A disease), it should be possible to develop pneumococcal conjugate vaccines that include serotypes causing at least 60% of all invasive pneumococcal disease in each region.

In summary, to ensure that pneumococcal vaccines eligible for the AMC financing have characteristics that are both desirable by countries and achieve a minimum level of public health impact, the TPP stipulates that vaccines must include:

1) serotypes 1, 5, and 14, the most frequently occurring serotypes in GAVI-eligible countries as a whole; and
2) serotypes that account for at least 60% of invasive pneumococcal disease isolates among children in the region for which the proposed pneumococcal vaccine will be used.

3. Introduction

Streptococcus pneumoniae is a leading cause of bacterial pneumonia and meningitis, and a leading cause of childhood mortality worldwide. The World Health Organization (WHO) estimates that more than 800,000 and up to nearly one million children die of serious pneumococcal infections each year, and ~98% of these deaths occur in developing countries. Despite the availability of a safe and effective pneumococcal conjugate vaccine (PCV), financial barriers limit access to these life-saving vaccines for children in countries with the risk of highest pneumococcal disease.

On February 9, 2007 an Advance Market Commitment (AMC) of $1.5 billion for advancement of pneumococcal vaccines was announced in Rome, Italy. This AMC is a novel funding mechanism for accelerating the introduction of pneumococcal vaccines into the world's poorest countries. This AMC creates a market for vaccine manufacturers and predictable, defined financing and pricing for countries who are considering introduction of pneumococcal vaccine. The AMC is supported by the governments of Italy, the United Kingdom, Canada, Norway, Russia, and the Bill and Melinda Gates Foundation, and its Secretariat is provided by the GAVI Alliance and the World Bank.

As part of the AMC process, the WHO was charged with establishing a Target Product Profile (TPP) Expert Committee to develop the TPP for pneumococcal vaccines. The TPP describes minimum characteristics required for a pneumococcal vaccine to be eligible for AMC financing. Characteristics stipulated in the TPP include public health impact and suitability of the product, taking into consideration measures of vaccine efficacy, safety, dose-scheduling, presentation and packaging. Vaccines that do not meet the TPP are not eligible for financing through the AMC.

The WHO requested GAVI's PneumoADIP to undertake a systematic literature review and meta-analysis of existing pneumococcal serotype data to support the TPP. The aim of this project – the Pneumococcal Global Serotype Project (GSP) - was to estimate the serotype distribution of invasive pneumococcal disease among children <5 years of age globally and by region. Results of Version 1 of the GSP analysis were used by the TPP Expert Committee when they established the minimum vaccine serotype characteristics for AMC-eligible pneumococcal conjugate vaccines. To ensure that pneumococcal vaccines eligible for the AMC financing are both desirable by countries and achieve a minimum level of public health impact, the TPP stipulates that vaccines must include:

1) serotypes 1, 5, and 14, the most frequently occurring serotypes in GAVI-eligible countries as a whole;
2) serotypes that account for at least 60% of invasive pneumococcal disease isolates among children in the region for which the proposed pneumococcal vaccine will be used.

During the TPP Expert Committee deliberations (September 2007) the Committee asked that the final serotype distribution codebook be based on the final GSP analysis rather than on the Version 1 analysis. This current document presents the final analysis (Version 2) of the pneumococcal Global Serotype Project. Version 2 was a pre-planned analysis that includes all data available in Version 1, and a significant amount of additional data from countries that were represented in Version 1 as well as from countries previously not represented. Version 2 also incorporates refined analytic methods that maximized the use of all available data. As stipulated by the TPP Expert Committee, the Version 2 GSP output serves as the definitive vaccine serotype analysis against which pneumococcal conjugate vaccines will be assessed for AMC eligibility throughout the course of the AMC.

The primary purpose of this analysis of serotypes is to guide the evaluation of vaccine formulations meeting the TPP. Serotype analyses and projections of the expected health impact of vaccination with specific formulations are also important for decisions pertaining to country adoption of pneumococcal vaccines. For this reason, the GSP may be updated in the future to incorporate additional serotype data and/or further refinements to the analytic methods when these become available. Furthermore, while regional estimates from meta-analyses will help to inform local decisions, the actual health impact in a particular country for a particular year may vary from projections based on regional estimates.

4. Methods

Systematic literature review. We conducted a systematic literature review to identify publications with data on the serotype distribution of invasive pneumococcal disease isolates among children less than 5 years of age from 1980 through June 2007. Four comprehensive literature searches had already been conducted by other investigators for similar pneumococcal serotype data but with limited years of the search or limitations in time or geographic scopes. We leveraged these existing literature searches by using the list of articles included in their analyses; we added to the output of those literature searches by conducting a comprehensive literature search of time-periods or regions not otherwise covered by those existing searches. We used modified keywords and search terms from the World Health Organization Hib and Pneumococcal Global Disease Burden Project (estimating year 2000 disease burden in children under 5 years of age) in 14 literature databases and included publications in the following languages: English, Spanish, French, Chinese, and Russian.

We supplemented the published literature with unpublished data by contacting key researchers around the world who were known to conduct surveillance for invasive pneumococcal disease. We included in the analysis studies with a minimum of 20 reported serotyped pneumococcal isolates, at least 12 months of surveillance, and a study initiation date of 1980 or later, among other characteristics. To assure accuracy, the data from all studies were double-abstracted and, where discordance was observed, these discrepancies were adjudicated by a third abstractor.

Our aim was to estimate the distribution of serotypes causing disease among children <5 years of age (i.e. ages 0 to 59 months) because this is the childhood age group with the highest risk of pneumococcal disease. Many studies reported data on serotypes that were limited to this age group, but others reported aggregated data that included children above and below 5 years of age. To avoid excluding studies that contained valuable data on children <5 years old simply because it was mixed with serotype data among children older than 5 years of age, we relaxed the age-specific inclusion criteria to allow such studies into the analysis under the assumption (based on knowledge of the age-specific pneumococcal incidence rates) that the majority of isolates in these studies are actually from children <5 years of age. In North America and Europe where data were plentiful, we allowed abstraction of aggregated data for children age <7 years (i.e. up through 83 months). For all other countries and regions, serotype data were abstracted for all studies that aggregated data for children up to <18 years of age (i.e. up through 215 months of age) under the above mentioned assumption.

Another effort to avoid excluding studies that contained valuable data was the inclusion of studies that reported serogroup but not serotype data. For example, many studies from the 1980s, when serotyping factor kits were in limited supply and expensive, reported their data as serogroup 6 or 19 for example, rather than as serotypes 6A, 6B, 19A and 19F. Studies presenting serogroup data not further sub-typed were included in this analysis. The distribution of serotypes within a serogroup was assigned using the region-specific relative proportions of serotypes within each specific serogroup. These relative proportions were estimated from regional analyses of studies which reported fully serotyped data. For example, if in a given region, 70% of serogroup 6 isolates were serotype 6A, and 30% were serotype 6B, then this distribution was applied to the number of isolates indicated as "serogroup 6" in the abstracted data from serogroup only studies from that region.

Primary Analytic Outcome: The primary outcome of the estimation process was the proportion (with 95% confidence intervals) of isolates causing invasive pneumococcal disease due to a particular serotype within a geographic region. The serotype data were abstracted from each study and then combined using a marginal meta-analysis random-effects model to estimate the proportion of invasive pneumococcal disease due to each serotype with 95% confidence intervals, within each region (World Bank region definitions--see Appendix 1). Confidence interval lower bounds for serotype distribution meta-analyses were set to zero if the lower bound was a negative value. Using the random effects model, the proportion of isolates due to a specific serotype for each study in a given region is weighted by a factor taking into account both the inverse of the variance for that study and the between-study variance; these weighted estimates are used to estimate the regional proportion of isolates due to each specific serotype.

Simulation exercises of the analytic method showed that reducing the number of serotypes estimated improved the accuracy of the estimates. This is because the bias introduced by substituting a small non-zero value for all 0% observations required by the random-effects model weighting procedure was magnified as the number of rare serotypes estimated increased. Therefore, we chose to estimate the serotype-specific proportions for the fewest possible serotypes meeting certain specifications (outlined below) This resulted in limiting results of serotype proportion within a region to only 21 serotypes; serotypes of approximately rank 20 or greater individually account for a very small proportion (generally less than 1%) of total invasive pneumococcal disease in a region. Individual serotypes beyond the 21 most common are included as an aggregated "other serotypes".

Serotypes had to meet at least one of the following criteria for inclusion in the model:

1) One of the 20 most common serotypes in the global serotype distribution, adjusted for regional disease burden;

2) One of the 13 most common serotypes in each region;
3) Included in the PCV products that are licensed or in late stage development i.e. 7-valent, 10-valent, and 13-valent PCV formulations.

Concordance of the serotype proportions observed among individual studies used within a regional analysis were examined by calculating the Intra-Class Correlation (ICC) for each serotype. When the ICC is high (i.e., informally defined as between 0.7 and 1.0) the variability in serotype proportion between studies is considered small compared to the natural statistical variation within a study (i.e. within-study variance). By contrast, when the ICC is low (i.e., below 0.5) different studies are as, or more, variable in their reporting of a given serotype proportion than the natural statistical variation within a study. In geographic regions with high ICCs for the reported serotypes, the removal of some studies and the addition of others does not substantially impact the overall meta-analytic estimate of the serotype proportion for that region.

We assumed serotype 6A/B cross-protection in discussion of results for potential vaccine formulations. Evidence supporting this assumption includes declines in invasive pneumococcal disease due to serotype 6A with introduction of the 7-valent Prevnar vaccine observed in the United States Centers for Disease Control and Prevention Active Bacterial Core surveillance,(1;2) and reduction in incidence of acute otitis media due to serotype 6A with vaccination using the heptavalent pneumococcal polysaccharide-CRM197 conjugate vaccine in a randomized, controlled efficacy trial.(3)

External Expert Review: An External Expert Committee (EEC) was established at the outset of the project to provide technical review and advice to the GSP project team on the model approach, data collection, literature review, and analysis interpretation. The External Expert Committee formally convened on four occasions through the life of the project including at key junctures in the development of both Version 1 and Version 2 analyses. When specific supplementary input on statistical and modelling methodologies was needed, additional individual consultations with EEC members and other external experts took place to solicit advice.

1. Whitney CG, Farley MM, Hadler J et al. Decline in Invasive Pneumococcal Disease after Introduction of Protein-Polysaccharide Conjugate Vaccine. *NEJM* 2003; 348(18):1737-1746.
2. Whitney CG, Pilishvili T, Farley MM et al. Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: a matched case-control study. *Lancet* 2006; 368:1495-1502.
3. Eskola J, Kilpi T, Palmu A et al. Efficacy of pneumococcal conjugate vaccine against acute otitis media. *NEJM* 2001; 344(6):403-409.

5. Results

From 1230 published studies reviewed and 60 datasets/supplemental data provided by investigators, 169 were included in this analysis. The 169 studies represent 70 countries and had data on over 60,000 pneumococcal isolates (Table 1). This is a substantial increase in the 61 studies, 34 countries, and 21,810 isolates incorporated in a previous analysis by Hausdorff, et. al. in 2000 (4). There was a sizeable number of isolates included in the analysis from every region (range: 3,649-18,788), with the fewest number of isolates from Oceania (3,649) and Asia (4,752). When the size of population under 5 years within the region is considered there are substantial differences in the amount of available data. The regional number of isolates in the analysis per 100,000 children age <5 years ranges from 1-136, with Asia and Africa having the least.

Regional Serotype Distribution

Regional serotype distributions of the serotype-specific proportions of invasive pneumococcal disease among young children, by region, along with the 95% Confidence Intervals for those point estimates are presented in Figures 1-6 and in Table 2. Figure 7 shows the cumulative distribution of the top ranked serotypes by region. The primary findings are as follows:

- Serotype 14 is the most common isolate among children <5 years old in all regions.
- Serotypes 1 and 5 are among the top 3 ranked serotypes occurring in the GAVI-eligible countries (data not shown) and are among the top 6 ranked serotypes occurring among children < 5 years in regions with the highest pneumococcal disease burden (Africa, Asia, and LAC) whereas they are not included in the top 6 in North America, Europe or Oceania.
- By contrast, serotypes 18C, 4, and 9V are ranked much higher in North America and Oceania than in the other regions. For example, these serotypes are not included in the top 7 in either Africa or Asia, whereas they are among the top 7 in North America and Oceania.
- Among GAVI-eligible countries in the analysis, serotypes 1, 5, and 14 are the three most common serotypes, each causing a similar proportion of disease (data not shown).
- In each region, three serotypes (1, 5, and 14) account for between one-quarter and almost one-half (28 – 43%) of invasive pneumococcal disease in children <5 years old.
- In each region, except Asia, among children <5 years old
    - 4 to 7 serotypes account for >60% of isolates.
    - 6 to 9 serotypes account for >70% of isolates.
    - 8 to 15 serotypes account for >80% of isolates.

- In Asia, it takes 7, 11, and 18 serotypes to account for >60%, >70%, and >80% of isolates causing invasive pneumococcal disease among young children, respectively.
- The seven most common disease causing serotypes in Africa are identical to those in Asia. The rank order of these seven serotypes differs between the two regions and the cumulative proportion of all serotypes included in these top seven is somewhat higher in Africa than in Asia (65% and 61% respectively).
- The same seven serotypes that are most common in Africa and Asia are among the most common in other regions also. They include 6 of the top 7 in LAC, 5 of the top 7 in Europe, and 4 of the top 7 in North America and Oceania.

Table 3 compares serotypes with ICC <0.3 by region. The top ranked serotypes in regions with the highest pneumococcal disease burden (Africa, Asia, and LAC) have low ICC indicating high variability across studies in the reported proportion of invasive pneumococcal disease due to these serotypes. In geographic regions with low ICCs for the reported serotypes, the addition or removal of studies has a large impact on the overall regional meta-estimate for those serotypes. In Africa, 10 serotypes had a ICC <0.2, thus the meta-estimates for these 10 serotypes are particularly sensitive to the addition or removal of serotype data in this region. Asia and LAC also had a large number of serotypes with ICC <0.2 (6 and 5 serotypes, respectively). By contrast, in Oceania, only serotype 14 had an ICC <0.2, indicating relative stability or homogeneity in the reported serotype proportion data for this region.

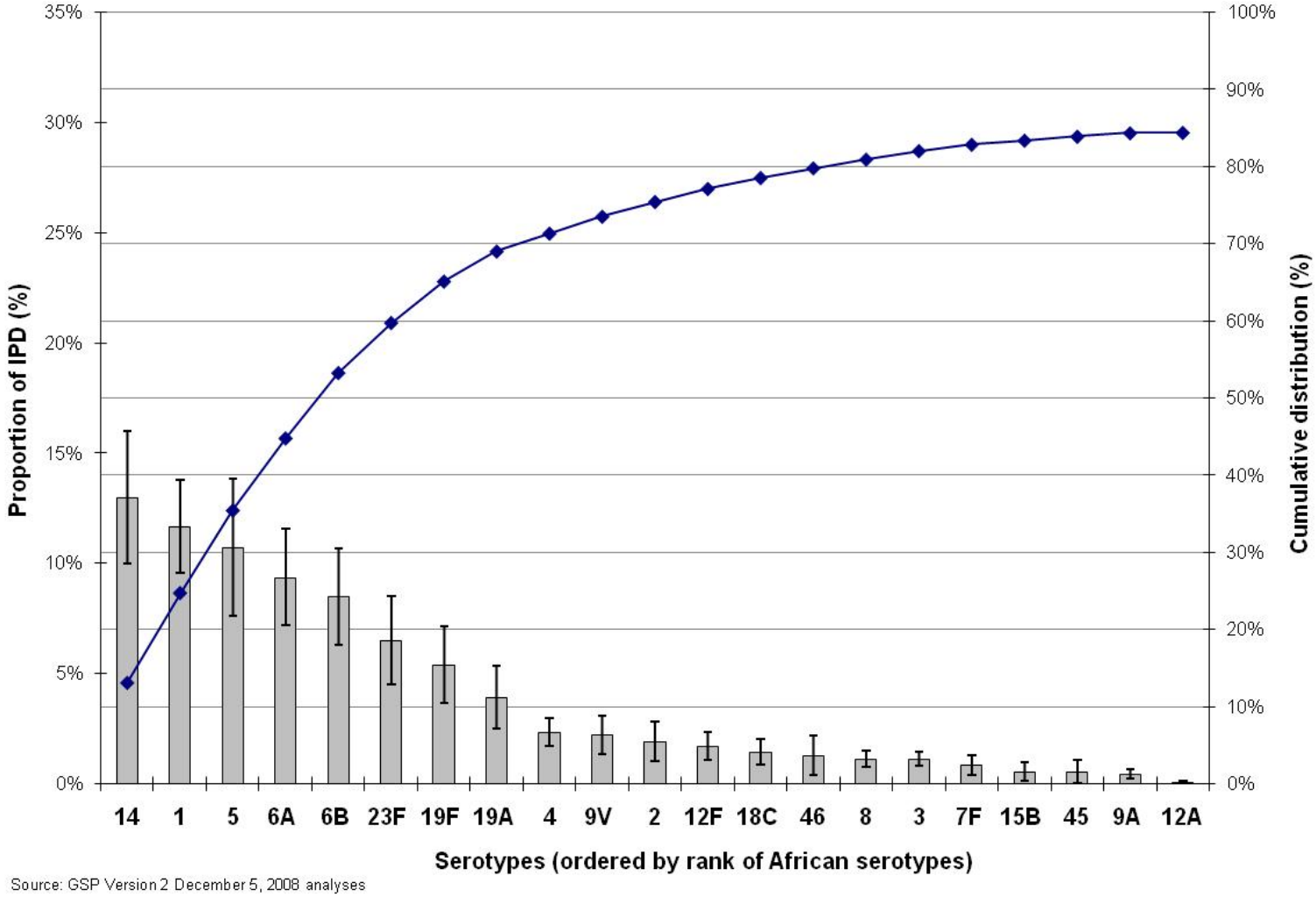
4. Hausdorff WP, Bryant J, Paradiso PR et al. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* 2000; 30:100-121.

Table 1. Summary of serotype data in the analysis.

	Africa	Asia	Europe	LAC	North America	Oceania	TOTAL
No. of Countries	13	18	16	17	2	4	70
No. GAVI-eligible countries	10	6	0	3	0	1	20
No. of Studies	22	33	39	42	17	16	169
No. of Isolates	11,181	4,752	10,279	18,788	11,441	3,649	60,090
No. of Isolates per 100,000 children <5 years	9	1	28	34	53	136	9

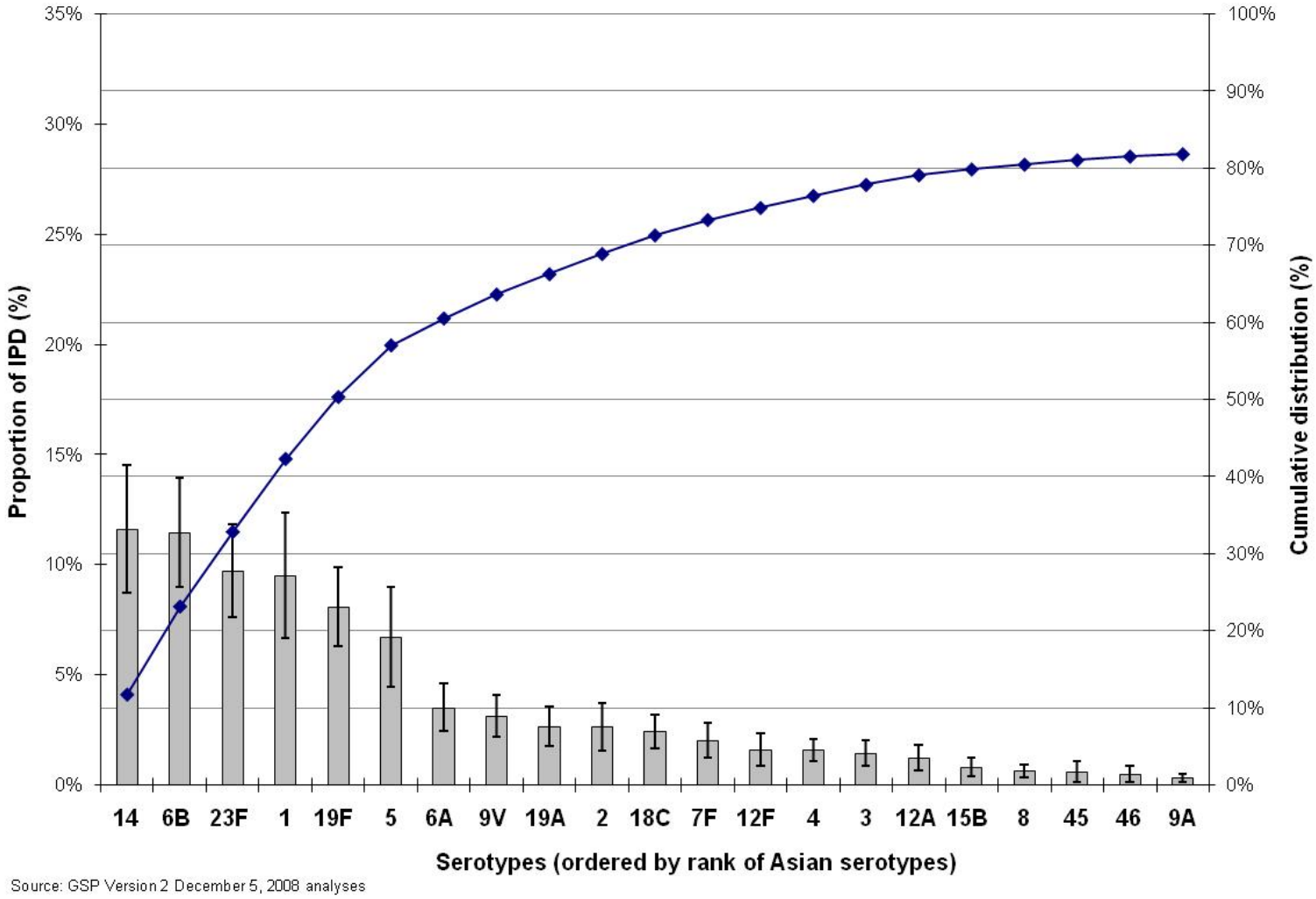
LAC = Latin America and the Caribbean; No. = Number

Figure 1. African serotypes by rank order and cumulative serotype distribution.



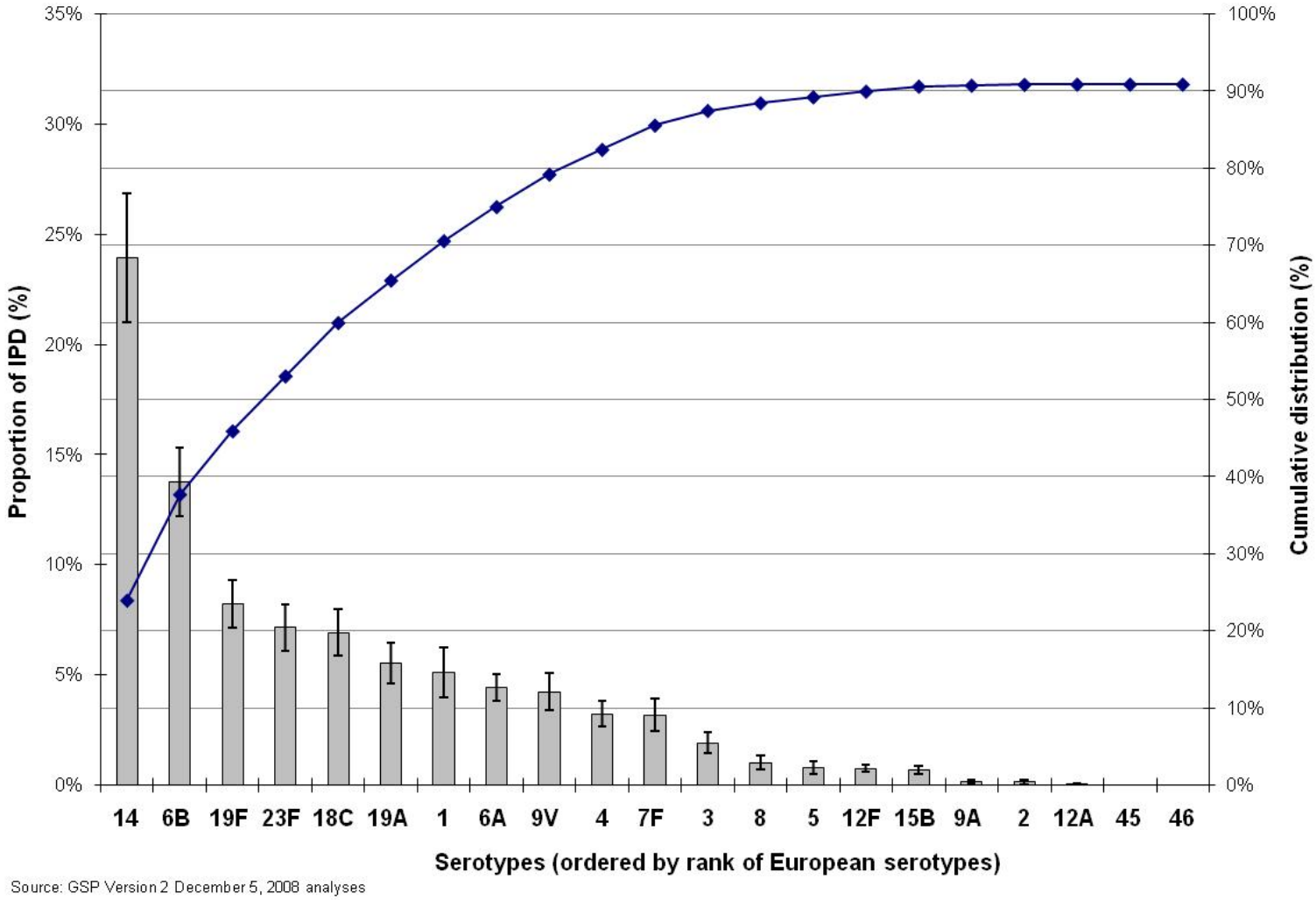
Note: Error bars represent 95% Confidence Intervals.

Figure 2. Asian serotypes by rank order and cumulative serotype distribution.



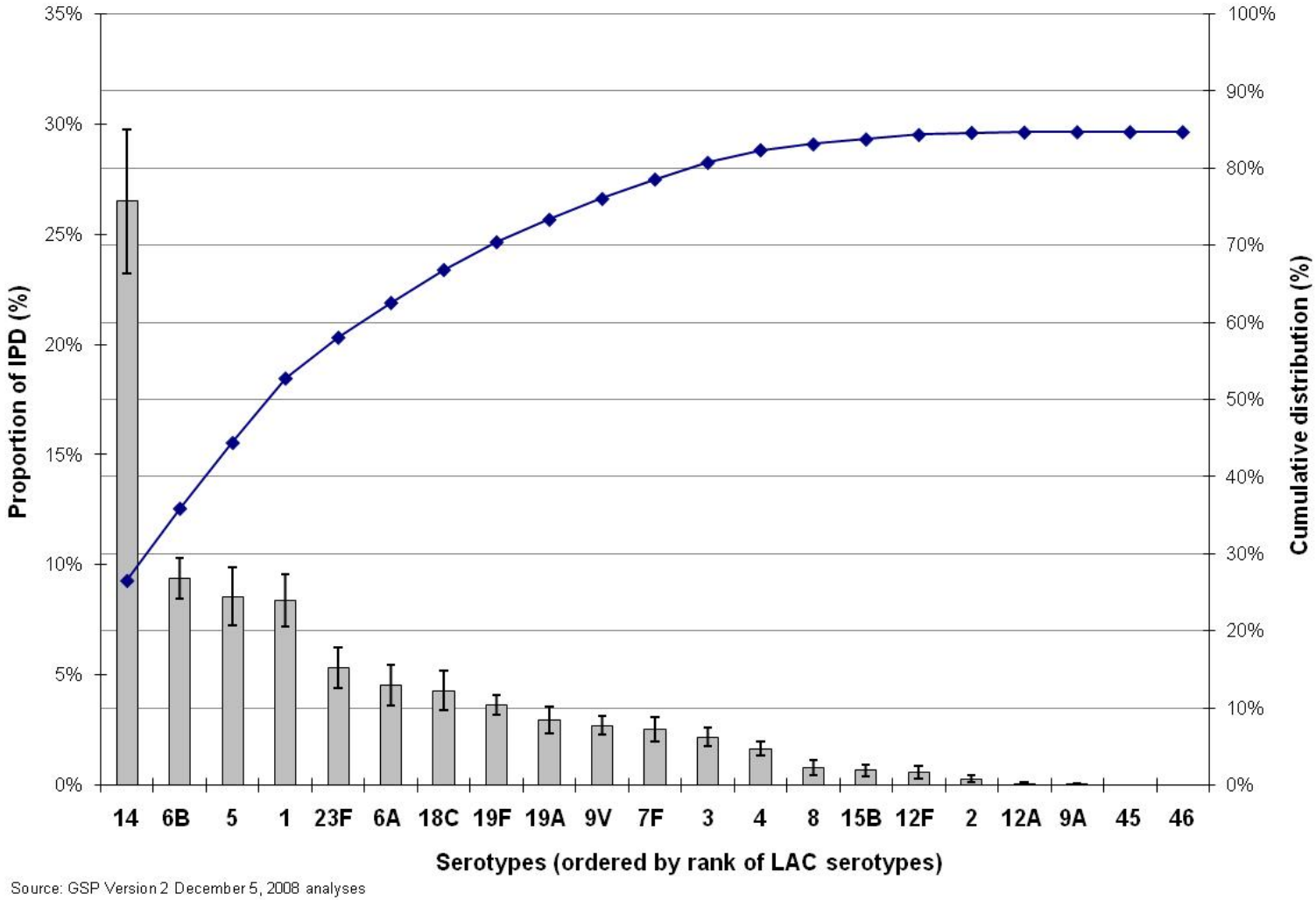
Note: Error bars represent 95% Confidence Intervals.

Figure 3. European serotypes by rank order and cumulative serotype distribution.



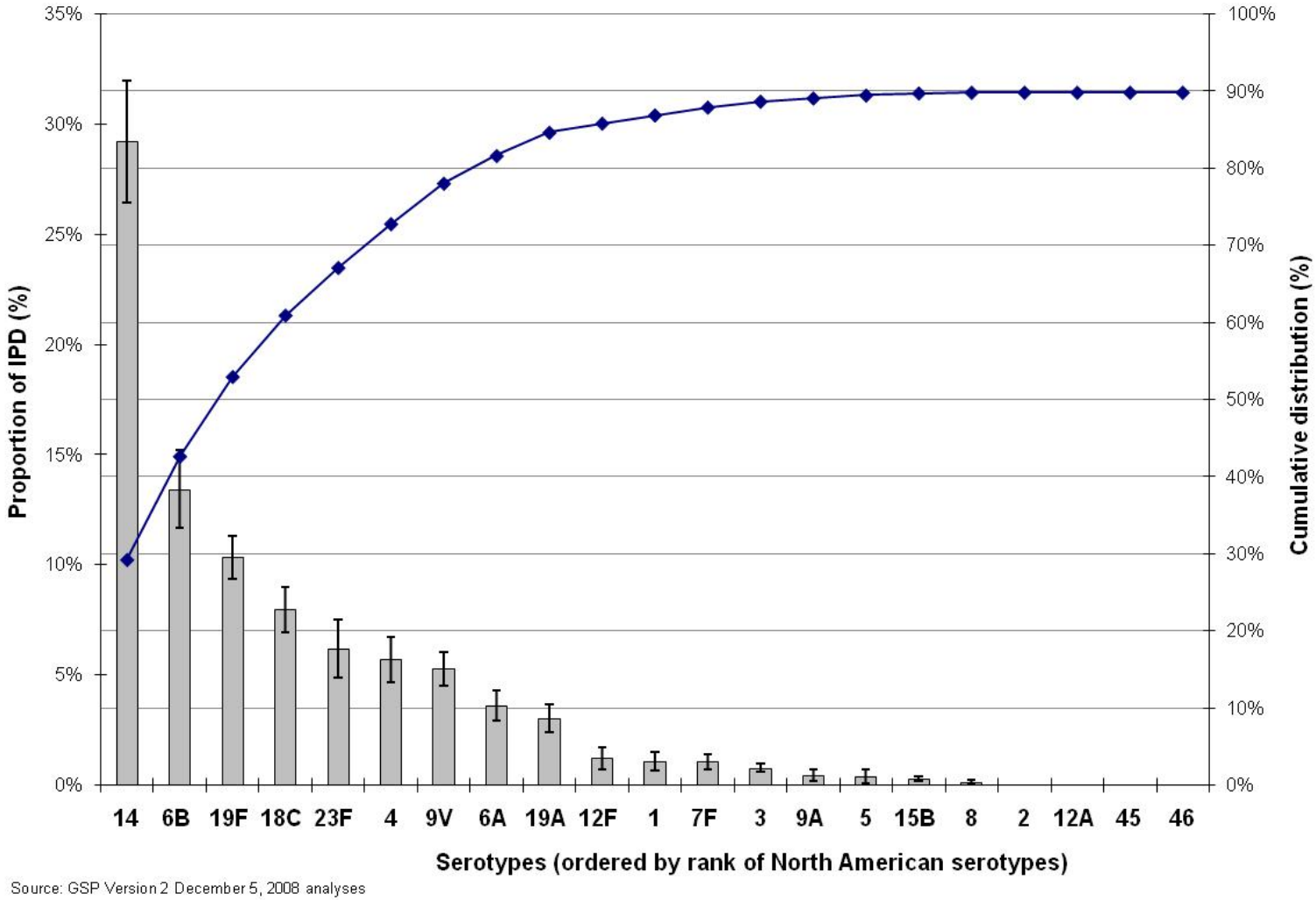
Note: Error bars represent 95% Confidence Intervals.

Figure 4. LAC serotypes by rank order and cumulative serotype distribution.



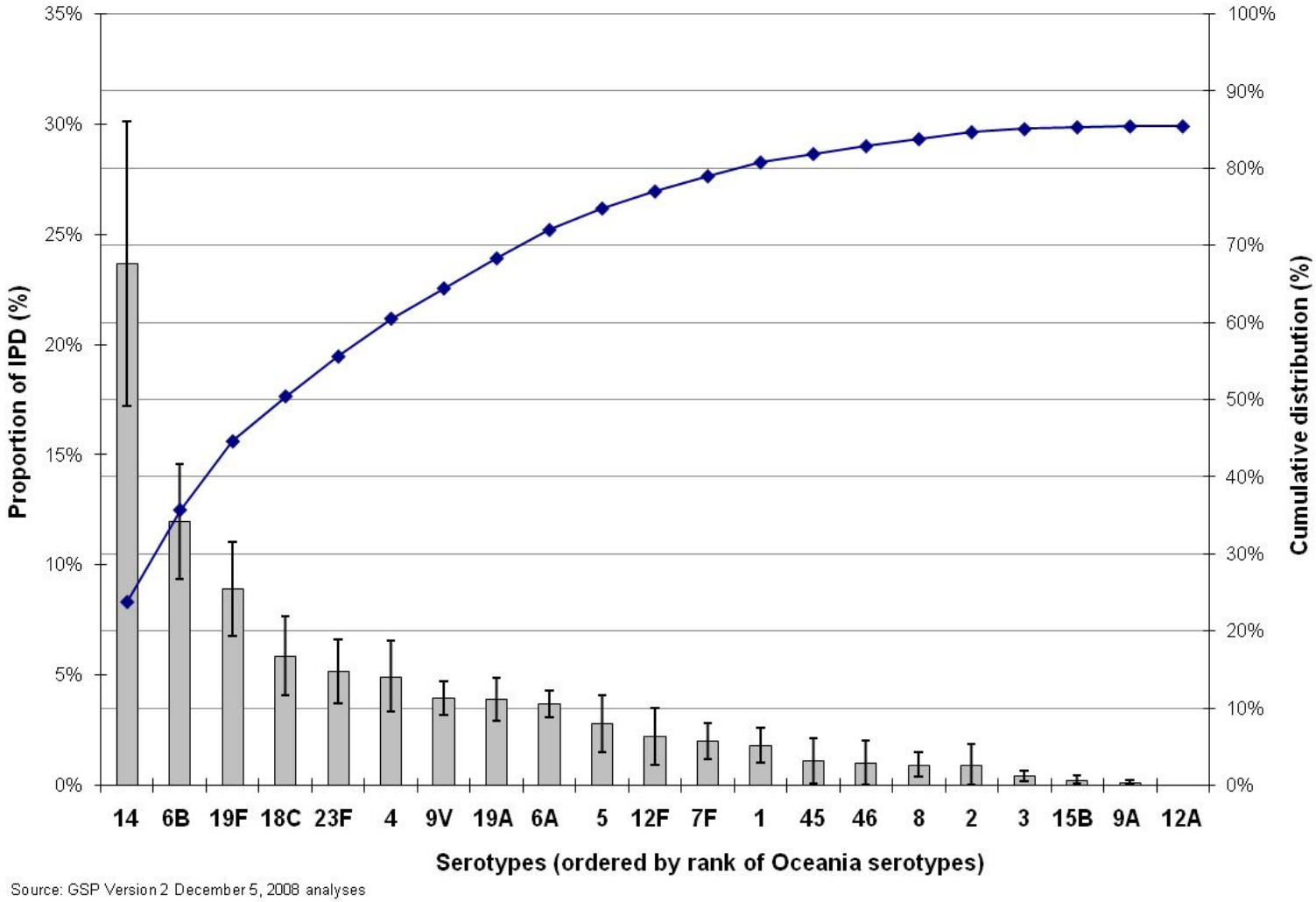
Note: Error bars represent 95% Confidence Intervals.

Figure 5. North American serotypes by rank order and cumulative serotype distribution.



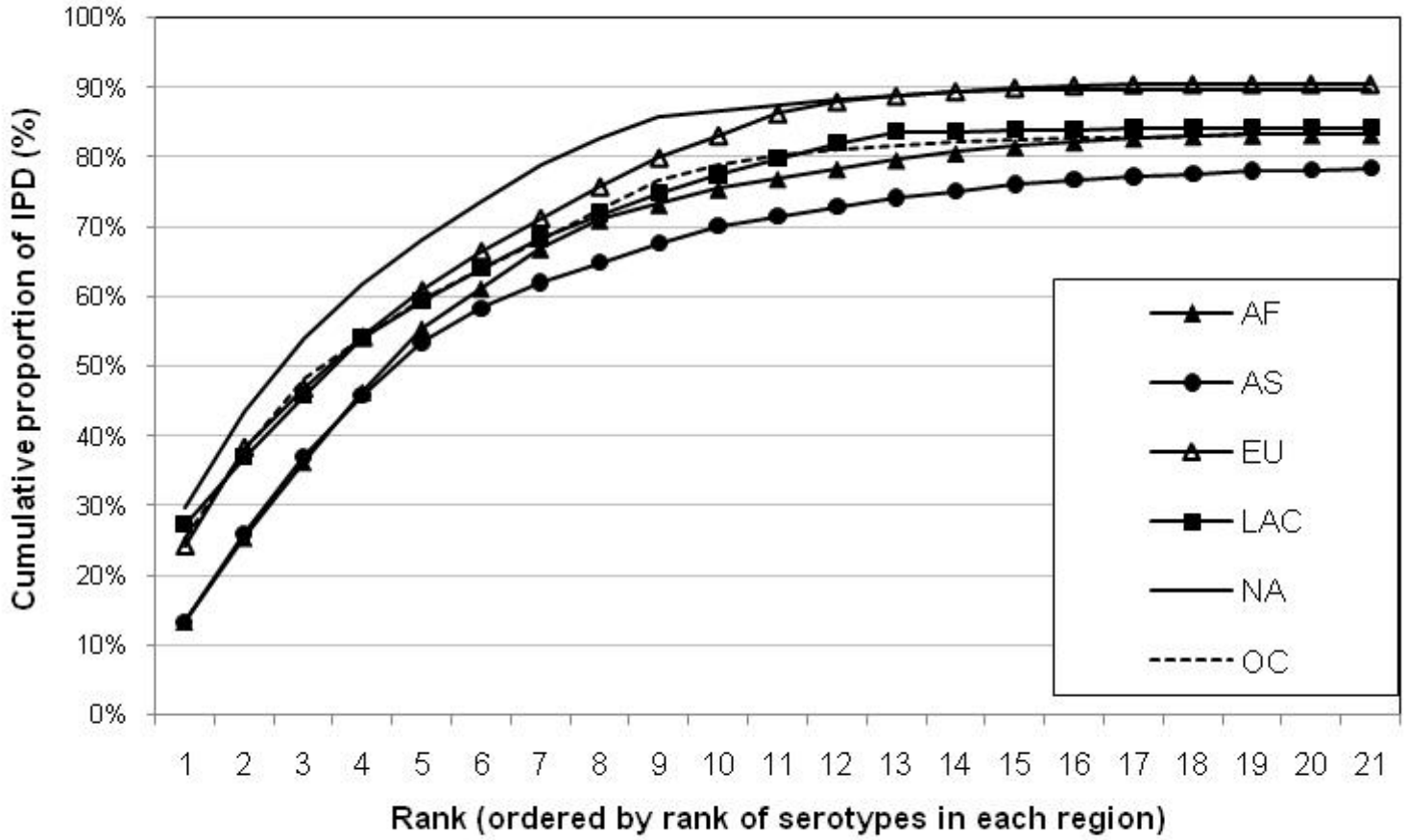
Note: Error bars represent 95% Confidence Intervals.

Figure 6. Oceania serotypes by rank order and cumulative serotype distribution.



Note: Error bars represent 95% Confidence Intervals.

Figure 7. Cumulative regional serotype distribution.



Source: GSP Version 2 December 5, 2008 analyses

Table 2. Serotype distribution by region.

Serotype	Africa (N=11,181)		Asia (N=4,752)		Europe (N=10,279)		LAC (N=18,788)		North America (N=11,441)		Oceania (N=3,649)	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI
1	11.7%	9.5, 13.8	9.5%	6.6, 12.3	5.1%	4.0, 6.2	8.4%	7.2, 9.6	1.1%	0.6, 1.5	1.8%	1.0, 2.6
2	1.9%	1.0, 2.8	2.6%	1.5, 3.7	0.1%	0.0, 0.2	0.3%	0.1, 0.4	0.0%	0.0, 0.0	0.9%	0.0, 1.8
3	1.1%	0.8, 1.5	1.4%	0.8, 2.0	1.9%	1.5, 2.4	2.2%	1.8, 2.6	0.8%	0.6, 1.0	0.4%	0.2, 0.7
4	2.3%	1.7, 3.0	1.6%	1.0, 2.1	3.2%	2.6, 3.8	1.6%	1.3, 1.9	5.7%	4.7, 6.7	4.9%	3.3, 6.6
5	10.7%	7.6, 13.8	6.7%	4.5, 9.0	0.8%	0.5, 1.1	8.5%	7.2, 9.8	0.4%	0.1, 0.7	2.8%	1.5, 4.1
6A	9.4%	7.2, 11.5	3.5%	2.4, 4.6	4.4%	3.8, 5.0	4.5%	3.6, 5.4	3.6%	2.9, 4.3	3.7%	3.1, 4.3
6B	8.5%	6.3, 10.7	11.5%	9.0, 14.0	13.7%	12.2, 15.3	9.4%	8.4, 10.3	13.4%	11.7, 15.1	12.0%	9.3, 14.6
7F	0.8%	0.4, 1.3	2.0%	1.2, 2.8	3.2%	2.4, 3.9	2.5%	2.0, 3.1	1.0%	0.7, 1.4	2.0%	1.1, 2.8
8	1.1%	0.8, 1.5	0.6%	0.3, 0.9	1.0%	0.7, 1.3	0.8%	0.4, 1.1	0.1%	0.0, 0.2	0.9%	0.4, 1.5
9A	0.4%	0.2, 0.7	0.3%	0.1, 0.5	0.1%	0.1, 0.2	0.0%	0.0, 0.1	0.4%	0.2, 0.7	0.1%	0.0, 0.2
9V	2.2%	1.3, 3.1	3.1%	2.2, 4.1	4.2%	3.4, 5.1	2.7%	2.3, 3.1	5.3%	4.5, 6.0	3.9%	3.1, 4.7
12A	0.1%	0.0, 0.1	1.2%	0.7, 1.8	0.0%	0.0, 0.1	0.1%	0.0, 0.1	0.0%	0.0, 0.0	0.0%	0.0, 0.0
12F	1.7%	1.1, 2.3	1.6%	0.8, 2.3	0.7%	0.6, 0.9	0.6%	0.3, 0.9	1.2%	0.7, 1.7	2.2%	0.9, 3.5
14	13.0%	10.0, 16.0	11.6%	8.7, 14.5	23.9%	21.0, 26.8	26.5%	23.2, 29.7	29.2%	26.4, 31.9	23.7%	17.2, 30.1
15B	0.5%	0.1, 0.9	0.8%	0.4, 1.2	0.7%	0.5, 0.8	0.7%	0.4, 0.9	0.3%	0.2, 0.4	0.2%	0.0, 0.4
18C	1.4%	0.9, 2.0	2.4%	1.7, 3.2	6.9%	5.9, 8.0	4.3%	3.4, 5.2	8.0%	6.9, 9.0	5.9%	4.1, 7.7
19A	3.9%	2.5, 5.3	2.6%	1.7, 3.5	5.5%	4.6, 6.4	2.9%	2.3, 3.5	3.0%	2.4, 3.7	3.9%	2.9, 4.9
19F	5.4%	3.6, 7.1	8.1%	6.3, 9.8	8.2%	7.1, 9.3	3.6%	3.2, 4.1	10.3%	9.3, 11.3	8.9%	6.8, 11.0
23F	6.5%	4.5, 8.5	9.7%	7.6, 11.8	7.1%	6.1, 8.2	5.3%	4.4, 6.2	6.2%	4.9, 7.5	5.2%	3.7, 6.6
45	0.5%	0.0, 1.0	0.6%	0.1, 1.0	0.0%	0.0, 0.0	0.0%	0.0, 0.0	0.0%	0.0, 0.0	1.1%	0.1, 2.1
46	1.3%	0.4, 2.1	0.5%	0.1, 0.9	0.0%	0.0, 0.0	0.0%	0.0, 0.0	0.0%	0.0, 0.0	1.0%	0.0, 2.0
All Others	15.7%	12.7, 18.6	18.2%	14.7, 21.6	9.2%	7.9, 10.4	15.3%	12.5, 18.1	10.2%	7.0, 13.4	14.6%	11.1, 18.1
TOTAL	100.0%		100.0%		100.0%		100.0%		100.0%		100.0%	

CI = Confidence Interval; N= Number

Table 3. Comparison of serotypes with ICC <0.3, by region.

	Serotype ICC <0.2	Serotype ICC=0.2-<0.3
Africa	1, 2, 5, 6A, 6B, 14, 15B, 19A, 19F, 23F	9V, 12F, 18C,
Asia	1, 5, 6B, 14, 19F, 23F	6A ,19A
Europe	14, 6B	1, 6A, 7F, 9V, 18C, 19A, 19F
LAC	1, 5, 14, 18C, 23F	6B, 7F
North America	14, 16B, 23F,	4
Oceania	14	-

ICC = Intra-Class Correlation

6. Discussion

This Version 2 of the GSP provides the most updated and comprehensive regional serotype analysis of invasive pneumococcal isolates among children around the world. This analysis allows the evaluation of whether candidate pneumococcal vaccines meet the minimum vaccine serotype characteristics as stipulated in the Pneumococcal AMC TPP. Consistent with findings in the Version 1 analysis, below is a summary of key findings in the GSP analysis as they pertain to the TPP vaccine serotype characteristics:

- **Serotype 14 is the most common serotype in all regions of the world among children < 5 years**

- **Serotypes 1 and 5 are among the top 3 ranked serotypes occurring in the GAVI-eligible countries and are among the top 6 ranked serotypes occurring among children < 5 years in regions with the highest pneumococcal disease burden (Africa, Asia, and LAC).**

- **A limited set of seven serotypes (1, 5, 6A, 6B, 14, 19F, 23F) are common and important in all regions with a substantial number of GAVI eligible countries. Together these 7 serotypes account for ~58%-66% of all invasive pneumococcal disease in each region.**

- **With as few as 6-7 serotypes (including serotypes 1, 5, and 14, and assuming that 6B conjugate vaccine provides cross protection against 6A disease), it should be possible to develop pneumococcal conjugate vaccines that include serotypes causing at least 60% of all invasive pneumococcal disease in each region.**

The serotype analysis presented here draws from a large body of observational studies of invasive disease causing pneumococcal strains from around the world. Data from many countries are represented in each region and each region has a substantial number of isolates in the analysis. The inclusion of a large number of isolates in the analysis from three decades results in smoothing of secular trends in the serotype data. The data in the GSP represent the most up-to-date serotype data from Africa and Asia, the two regions with the highest serious pneumococcal disease burden and greatest number of GAVI eligible countries.

The serotype distribution of disease causing pneumococcal strains is known to vary between populations. The most obvious example here is the regional variation of serotype 1 and 5 prevalence between some regions. These serotypes are observed as more common causes of disease in Asia and Africa compared to North America. In previous analyses, regional

estimates were generated by calculating averages weighted by number of isolates in the study. That method assumes homogeneity of serotype distributions within a region and produces estimates dominated by the serotype proportions observed in the largest studies within the region because variability in results of small studies is assumed to be the result of statistical random variability, not plausible biologic variability. But because serotype distribution may also vary within a region and not only between regions, particularly in large, diverse regions such as Asia, we preferred to use a method that incorporates diversity of serotype proportions rather than suppress it. Furthermore, there were a large number of countries that only had serotype data from studies of small sample size. We wanted a statistical approach that would allow for the serotype distributions in those countries to be fully represented and not overcome by large studies from other countries within the region. For these reasons we used a **random effects meta-analytic model** that enhanced the contributions of small studies (compared with a fixed-effect model) and reflected diversity when there was evidence of heterogeneity of results within a region. In a random effects model, studies of smaller sample size are weighted more heavily than otherwise would be possible in a fixed effects model. However, this characteristic of the random effect model also introduces a limitation when many small studies are included in the analysis: it may enhance the observed variability of serotype proportions within a region.

Study Limitations:

While the data included in this report represent a major improvement over the past, the results of the GSP analytic effort should be considered in light of several important limitations of the data and the methods used.

This analysis included a **substantial amount of unpublished data** coming from either unpublished studies, or as supplements to studies whose publication did not include sufficient detail for the GSP analysis. Some might suggest that unpublished data is of lesser quality than that of published data. However, in general, the unpublished data used in the GSP analyses were usually too recent to have the opportunity for publication, or consisted of the raw data used to generate published results, and nearly all came from surveillance projects with a record of publishing their results.

Pneumococcal **serogroup, not serotype,** was sometimes the only data available from studies. Given the intent to include as much data as possible in the analysis, serogroup data from these studies were apportioned into serotypes using the region-specific relative proportions of serotypes within each specific serogroup estimated from regional analyses of studies with fully serotyped data.

It was sometimes unclear whether the absence of a serotype among a strain collection represented failure to test for that serotype or true lack of the serotype causing disease in the collection of isolates. Several studies only tested for or only reported serotypes included in

the PCV products which are licensed or in late stage development i.e. 7-valent, 10-valent, and 13-valent serotypes; because results for non-vaccine serotypes were missing, the proportion of pneumococcal disease due to non-vaccine type serotypes is based only on the subset of studies that included serotype specific details of these non-vaccine serotype distributions.

In spite of specific efforts to seek serotype data from countries with large pediatric populations, there remain **important countries with limited or no serotype data**, including China (236 isolates), India (216 isolates), Indonesia (0 isolates), and Nigeria (0 isolates). On the other hand, each of these countries comes from a region or sub-region where there were data from neighboring countries. For example, a large number of isolates were available from Bangladesh, Nepal, Pakistan, and Sri Lanka, which surround India and share many epidemiologic characteristics.

The data sparsity and incompleteness of serotype reporting within a study posed some unique analytic challenges. One of the weighting factors used in modeling the serotype distribution is the inverse of the variance, which has in the denominator the number of isolates observed for a particular serotype. Therefore for serotypes reported as 0%, the variance goes to infinity and it becomes impossible to establish a weighting factor for that serotype in that study. To resolve this, we used **Bayesian estimation** to make all proportions effectively non-zero by adding 0.5 isolates to the numerator and 1.0 isolates to the denominator in the calculation of all serotype-specific proportions for each study. This adjustment, although required by the model, likely resulted in a small overestimation of less common serotypes. This effect is more pronounced in studies of small sample size (N<100) but probably has relatively little impact in very large (N>500 isolates) studies. Because the random-effects method weights small studies more heavily than in a fixed-effect analysis, the benefit of reflecting diversity in serotype distribution that the random-effects method affords is at the expense of exaggerating heterogeneity across serotypes.

When meta-analytic estimates of each serotype proportion are generated, the sum of the serotype proportions does not equal 100% exactly. This results from the weighting in the random effects model and the fact that each serotype proportion is estimated independent of the proportions for any other serotype within a study, We correct the failure of the cumulative proportions to sum exactly to 100% by **normalizing the data** for each regional analysis so that the sum of all serotype proportions within a region is exactly 100%. Normalizing the data (i.e. by applying a constant factor to all calculated serotype proportions within a regional analysis) may bias the serotype distribution for regions with a large number of small studies because the overestimation of uncommon serotypes as described above makes the sum greater than 100%. To make all serotypes sum to 100%, all serotypes must be reduced, including those of the more common serotypes, thereby underestimating the proportions of the most common serotypes.

To increase country representativeness in the dataset, a specific objective of the GSP, we included a large number of studies with small sample size (<100 isolates). Studies with small sample size have less chance of observing a large number of different serotypes, thus 0% is observed for most serotypes in these studies. The bias introduced with the Bayesian estimation (i.e. substituting a small non-zero value for all 0% observations required by the random-effects model weighting procedure) was magnified as the number of rare serotypes estimated increased. The inclusion of studies of small sample size, although desirable from the perspective of representativeness and completeness of the dataset, affects the meta-analysis point estimates of serotype proportions; the biases described above are more pronounced when a greater number of small studies are included in the dataset compared with results excluding these smaller datasets. Thus the decision to include studies with fewer than 100 isolates represents a tradeoff balancing the desire for widest geographical inclusion against directional bias of the meta-analysis point estimates. One approach to minimize the biases exacerbated by the inclusion of small datasets is to limit the number of serotypes modeled. Therefore we estimated the serotype-specific proportions for the fewest possible serotypes meeting certain specifications outlined in the methods section of this document, or the 21 most important serotypes instead of all 90. Providing results for only 21 serotypes instead of all 90 serotypes is not expected to have a meaningful impact on the ability of this document to serve the interests of the AMC or the use by manufacturers because serotypes beyond rank 20 individually comprise a very small proportion of invasive pneumococcal disease.

These methodological attributes (use of random effects meta-analytic model, Bayesian estimation, inclusion of studies with small sample sizes, and normalization of the serotype data) result in a directional bias making the contribution of the most common serotype somewhat underestimated and the less common serotypes somewhat overestimated. Therefore the **estimations presented here of serotype-specific proportions for the more commonly observed serotypes (those above 2%) should be interpreted as minimum estimates**. Also, consequently, cumulative estimates (e.g., vaccine formulations containing several serotypes) should also be considered minimum point estimates because this bias will tend to be compounded in cumulative distributions.

Although age (and perhaps other characteristics such as syndrome, body fluid, HIV prevalence, population size, and disease burden) may be important confounders of the regional estimates of serotype proportions, the regional estimates have not been adjusted/corrected for these potential biases.

7. Acknowledgements

The Pneumococcal GSP gratefully acknowledges the collaboration of a large number of investigators worldwide. It is the generous, open and willing collaboration of these investigators which has contributed to the broad representativeness and timeliness of the data in the GSP analysis. We also want to thank members of the GSP External Expert Committee and Technical Experts for their time and consideration in providing guidance in the design and interpretation of the findings in the GSP. Also, we thank members of the GSP Team at Johns Hopkins Bloomberg School of Public Health and WESTAT, Inc. for their dedicated work to complete the GSP. Finally, we thank the AMC Secretariat, including Tania Cernuschi, Nina Schwalbe, and Andrew Jones for their efforts and support to this project.

Groups who shared unpublished serotype data:

1. **Bangladesh**: Samir Saha; Steve Luby; Abdullah Brooks; Aliya Naheed; Shams el-Arifeen; Maksuda Islam; Hassan Al Emran.
2. **Belgium**: Jan Verhaegen.
3. **Burkina Faso**: Bradford D. Gessner; Judith Mueller; Jean-Louis Koeck; Lassana Sangare; Berthe-Marie Njanpop-Lafourcade; Mathilde Lourd; Seydou Yaro.
4. **Canada**: James Kellner (CASPER-Calgary Area S. Pneumoniae Epidemiology Research Group); Allison McGeer, Donald E. Low (Toronto Invasive Bacterial Diseases Network).
5. **Dominican Republic**: Jesus M. Feris Iglesias; Jacqueline Sanchez; Hilma Coradin; Josefina Fernandez.
6. **Guatemala**: Edwin Asturias; Neal Halsey; Erica Dueger (Johns Hopkins Bloomberg School of Public Health).
7. **Israel**: Ron Dagan, Noga Givon-Lavi (Pediatric Infectious Disease Unit, Soroka University Medical Center and the Faculty of Health Sciences, Be-Gurion University, Beer-Sheva, Israel).
8. **Japan**: Nahoko Chiba, Kimiko Ubukta, Keisuke Sanakawa (Laboratory of Molecular Epidemiology for Infectious Agents, Kitasato Institute for Life Sciences, Kitasato University, Tokyo, Japan).
9. **Kenya**: Anthony Scott, Brett Lowe, Salim Mwarumba, Joyce Nyiro, Salome Njenga, Angela Karani, Eva Wanjiru, Stella Ngugi, Daisy Mugo, Moses Ndiritu, Maranga Wamae, Sandra Mudhune (netSPEAR).
10. **Lebanon**: Ghassan Dbaibo, American, George Araj, Omar Salam, Sahar Koubar, American University of Beirut, Lebanon; Guillermo Pimentel, (NAMRU-3, Cairo, Egypt).
11. **Mali**: Souleymane Diallo; Uma Onwuchekwa; Mahamadou M. Keita; Mama N. Doumbia; Myron M. Levine; Karen Kotloff; Milagritos D. Tapia; Samba O. Sow.

12. **Mozambique**: Anna Roca.
13. **Nepal**: Andrew J. Pollard; Neelam Adhiraki; Stephen Thorson; Dominic Kelly; David Murdoch.
14. **Nigeria**: AG Falade; RA Bakare; AA Odekanmi; RA Adegbola.
15. **Oman**: Suleiman El-Busaidy; Guillermo Pimentel(NAMRU-3, Cairo, Egypt); Kevin Russell and lab group (Naval Health Research Center in San Diego, California); Global Emerging Infections Surveillance and Response System (DOD-GEIS).
16. **Pakistan**: Anita K.M. Zaidi (Pediatrics and Microbiology, Aga Khan University, Karachi, Pakistan); Guillermo Pimentel(NAMRU-3, Cairo, Egypt); Kevin Russell and lab group (Naval Health Research Center in San Diego, California); Global Emerging Infections Surveillance and Response System (DOD-GEIS).
17. **India, Nepal, Sri Lanka**: Mark Steinhoff, Kurien Thomas, Poorva Kulkarni (SAPNA).
18. **Russia**: Alina Martynova (Vladivostok Medical University).
19. **South Africa**: Anne von Gottberg for the Group for Enteric, Respiratory and Meningeal Surveillance (GERMS-South Africa); Members of GERMS-South Africa; Keith Klugman (Emory University).
20. **Tanzania**: Anthony Scott, Brett Lowe, Salim Mwarumba, Joyce Nyiro, Salome Njenga, Angela Karani, Eva Wanjiru, Stella Ngugi, Daisy Mugo, Moses Ndiritu, Maranga Wamae, Sandra Mudhune (netSPEAR).
21. **Thailand**: Thailand Ministry of Public Health; Surang Dejsiri; Leelawadee Sangsuk; Sonja J. Olsen, Leonard Peruski, Possawat Jorakate, Anek Kaewpan, Prasert Salika, Henry C. Baggett, Susan Maloney (International Emerging Infections Program).
22. **Uganda**: Anthony Scott, Brett Lowe, Salim Mwarumba, Joyce Nyiro, Salome Njenga, Angela Karani, Eva Wanjiru, Stella Ngugi, Daisy Mugo, Moses Ndiritu, Maranga Wamae, Sandra Mudhune (netSPEAR).
23. **United Kingdom**: Robert George, Pauline Kaye, Liz Miller (Health Protection Agency).
24. **United States of America (Navajo)**: Kate O'Brien, Mathuram Santosham, Robert Weatherholtz, Ray Reid (Johns Hopkins Bloomberg School of Public Health).

Groups who provided additional information for published serotype data:

1. **Algeria**: Nadjia Ramdani.
2. **Australia**: Vicki Krause, Heather Cook (Center for Disease Control, NT Department of Health and Community Services, Australia); Paul Roche (Australian Government Department of Health and Aging); Geoff Hogg (Microbiological Diagnostic Unit, University of Melbourne); Lyn Gilbert (Centre for Infectious Diseases and Microbiology, Institute of Clinical Pathology and Medical Research); Denise Murphy (Queensland Health and Pathology Scientific Services); Marion Easton (University of Melbourne).
3. **Belgium**: Jan Verhaegen.

4. **Brazil**: Eitan Berezin (Pediatrics Infectious Diseases Department of Brazilian Pediatrics Society); Maria Cristina C. Brandileone (Adolfo Lutz Institute, São Paulo, Brazil, and the National Coordination of the Public Health Laboratories CGLAB, SVS, Ministry of Health of Brazil); Luiza Helena F Carvalho; Marco Aurelio Safadi; Claudia Lopes; Cristiana Maria Nascimento-Carvalho; Otávio Augusto Moreno-Carvalho; Noraney Nunes Alves; Renilza Menezes Caldas; Maria Goreth Barberino; Jussara Duarte; Maria Angélica Brandão; Maria Luiza Guerra; José Luis Di Fabio; Orlando Mantese.
5. **Canada**: James Kellner (CASPER-Calgary Area S. Pneumoniae Epidemiology Research Group); Allison McGeer, Donald E. Low (Toronto Invasive Bacterial Diseases Network).
6. **Egypt**: Guillermo Pimentel(NAMRU-3, Cairo, Egypt); Kevin Russell and lab group (Naval Health Research Center in San Diego, California); Global Emerging Infections Surveillance and Response System (DOD-GEIS).
7. **France**: Catherine Doit; Edouard Bingen.
8. **Gambia**: Richard A. Adegbola; Philip Hill; Brian M Greenwood; Felicity Cutts.
9. **Israel**: Ron Dagan, Noga Givon-Lavi (Pediatric Infectious Disease Unit, Soroka University Medical Center and the Faculty of Health Sciences, Be-Gurion University, Beer-Sheva, Israel).
10. **Japan**: Nahoko Chiba, Kimiko Ubukta, Keisuke Sanakawa (Laboratory of Molecular Epidemiology for Infectious Agents, Kitasato Institute for Life Sciences, Kitasato University, Tokyo, Japan).
11. **Kenya**: Anthony Scott, Brett Lowe, Salim Mwarumba, Joyce Nyiro, Salome Njenga, Angela Karani, Eva Wanjiru, Stella Ngugi, Daisy Mugo, Moses Ndiritu, Maranga Wamae, Sandra Mudhune (netSPEAR).
12. **Mali**: Souleymane Diallo; Uma Onwuchekwa; Mahamadou M. Keita; Mama N. Doumbia; Myron M. Levine; Karen Kotloff; Milagritos D. Tapia; Samba O. Sow.
13. **Mozambique**: Anna Roca.
14. **New Caledonia**: Benoit Garin.
15. **Portugal**: Mario Ramirez, J. Melo-Cristino (Faculty of Medicine, Lisbon, Portugal).
16. **South Africa**: Anne von Gottberg for the Group for Enteric, Respiratory and Meningeal Surveillance (GERMS-South Africa); Members of GERMS-South Africa; Keith Klugman (Emory University).
17. **Sweden**: Karl Ekdahl (European Center for Disease Prevention and Control); Jonas Hedlund.
18. **Thailand**: Thailand Ministry of Public Health; Surang Dejsiri; Leelawadee Sangsuk; Sonja J. Olsen, Leonard Peruski, Possawat Jorakate, Anek Kaewpan, Prasert Salika, Henry C. Baggett, Susan Maloney (International Emerging Infections Program).
19. **Turkey**: Burcin Sener (Department of Clinical Microbiology, Hacettepe Univ. Medical Faculty).

20. **United Kingdom**: Gary J.A. Eltringham (Health Protection Agency-Newcastle Laboratory); John P. Leeming (Regional Health Protection Agency Laboratory-Bristol); David Spencer (Freeman Hospital); Angel M. Kearns (Health Protection Agency-Laboratory of Healthcare Associated Infection); Katherine Eastham; Julia Clark; Robert George, Pauline Kaye, Liz Miller (Health Protection Agency).
21. **United States of America (CDC Active Bacterial Core Surveillance)**: Matthew Moore, Tamara Pilishvili, Bernard Beall, Cynthia Whitney, Thomas Adamkiewicz (US Centers for Disease Control and Prevention); Catherine Lexau, Lindsey Lesher (Minnesota Department of Health).
22. **United States of America (Navajo)**: Kate O'Brien; Mathuram Santosham; Ray Reid; Robert Weatherholtz (Johns Hopkins Bloomberg School of Public Health).
23. **United States of America (West Virginia)**: Maurice Mufson, Ronald Stanek.
24. **Uruguay**: Maria Hortal (National Child Health Program, Ministry of Health, Uruguay).

Groups who shared literature reviews:

1. Global serotype distribution meta-analysis: William Hausdorff, Peter Paradiso, Greg Yothers. Hausdorff WP, Bryant J, Paradiso PR, Siber GR.

   Citation: Which pneumococcal serogroups cause the most invasive disease: Implications for conjugate vaccine formulation and use, Part I. *Clin. Infect Dis.* 2000; 30: 100-121.
2. Sub-Saharan Africa serotype distribution meta-analysis: Anushua Sinha, Jackson Wan, Dharmistha Kaul.

   Citation: Sinha A, Wan J, Kaul D. Serotype distribution of invasive pneumococcal disease in sub-Saharan Africa. (personal communication).
3. Latin America and Caribbean serotype distribution meta-analysis: Rosalyn O'Loughlin (Respiratory Diseases Branch, National Center for Immunizations and Respiratory Diseases CDC); Maria Teresa Valenzuela (Departamento de Salud Pública y Epidemiología Universidad de los Andes, Facultad de Medicina, Santiago, Chile); Elizabeth Gomez (Dirección General de Epidemiología, Secretaria de Estado de Salud Publica, Santo Domingo, Republica Dominicana); Fernando Pio de la Hoz (Universidad Nacional de Colombia-Facultad de Medicina, Bogota, Colombia); The Albert B. Sabin Vaccine Institute (SVI); Pan American Health Organization (PAHO); GAVI's Pneumococcal Accelerated Development and Introduction Plan at Johns Hopkins (PneumoADIP); Centers for Disease Control and Prevention (CDC).

   Citation: Collaborative Project of the Sabin Vaccine Institute, Pan American Health Organization, GAVI's PneumoADIP, U.S. Centers for Disease Prevention and Control. The burden of pneumococcal disease and cost effectiveness of a

pneumococcal vaccine for Latin American and the Caribbean. Draft version April 27, 2007 (personal communication).

4. Asia-Pacific serotype distribution meta-analysis: Paul Kilgore/IVI.

   Citation: Nyambat B, Kilgore PE, Wang XY, Lee SH, Grandy M, Chun LY, Jodar L. Systematic review of invasive pneumococcal disease in Asia-Pacific countries: Implications for estimating the burden of disease. Draft version November 27, 2006 (personal communication).

Individuals who assisted in identifying and abstracting literature from Eastern Europe
Liudmyla Chernyshova, PATH ; Anton Luchitsky, PATH; Eileen Quinn, PATH; Elena Kononova, PATH; Halyna Romaniyk, PATH.

GSP External Expert Committee:
Steve Goodman, James Kellner, Keith Klugman, Samir Saha , Anthony Scott, Cynthia Whitney.

GSP Technical Experts:
Scott Zeger, Anushua Sinha, Natalie McCall, Steve Black

GSP Team:
Orin Levine, Kate O'Brien, Maria Knoll, Hope Johnson, Zunera Gilani, Emily Henkle (PneumoADIP); Jim Korelitz, Richard Reithinger, Laura Freimanis, Larry Muenz, Kellar Wilson, Sonia Stoszek, Judith Chamberlin, David Chang, Bhumika Patel, Rebekah Zanditon, Jeannine Sachar, Maureen Stawick, Wen-Chau Haung, Johns Edmonds, Katie Hubbell, Bethany Baer, Andrea Sutherland, Moussa Sarr, Karen Megazzini, , Brian Stout, Margot Krauss (WESTAT).

Funding:
This work and the PneumoADIP were funded in full by the GAVI Alliance and The GAVI Fund.

Appendix 1. Countries by World Bank region.

World Bank Region	Country	
Africa	Algeria	Libya
	Angola	Madagascar
	Benin	Malawi
	Botswana	Mali
	Burkina Faso	Mauritania
	Burundi	Mauritius
	Cameroon	Morocco
	Cape Verde	Mozambique
	Central African Republic	Namibia
	Chad	Niger
	Comoros	Nigeria
	Congo	Rwanda
	Cote d´Ivoire	Sao Tome and Principe
	Democratic Republic of the Congo	Senegal
	Djibouti	Seychelles
	Egypt	Sierra Leone
	Equatorial Guinea	Somalia
	Eritrea	South Africa
	Ethiopia	Sudan
	Gabon	Swaziland
	Gambia	Togo
	Ghana	Tunisia
	Guinea	Uganda
	Guinea-Bissau	Tanzania
	Kenya	Zambia
	Lesotho	Zimbabwe
	Liberia	
Asia	Afghanistan	Lebanon
	Armenia	Malaysia
	Azerbaijan	Maldives
	Bahrain	Mongolia
	Bangladesh	Myanmar
	Bhutan	Nepal
	Brunei	Oman
	Cambodia	Pakistan
	China	Philippines
	Cyprus	Qatar
	Democratic People's Republic of Korea	Republic of Korea
	Timor-Leste	Saudi Arabia
	Georgia	Singapore
	India	Sri Lanka

World Bank Region	Country	
	Indonesia	Syrian Arab Republic
	Iran	Tajikistan
	Iraq	Thailand
	Israel	Turkey
	Japan	Turkmenistan
	Jordan	United Arab Emirates
	Kazakhstan	Uzbekistan
	Kuwait	Viet Nam
	Kyrgyzstan	Yemen
	Lao	
Europe	Albania	Malta
	Andorra	Monaco
	Austria	Montenegro
	Belarus	Netherlands
	Belgium	Norway
	Bosnia and Herzegovina	Poland
	Bulgaria	Portugal
	Croatia	Moldova
	Czech Republic	Romania
	Denmark	Russian Federation
	Estonia	San Marino
	Finland	Slovakia
	France	Slovenia
	Germany	Spain
	Greece	Sweden
	Hungary	Switzerland
	Iceland	Macedonia
	Ireland	Ukraine
	Italy	United Kingdom
	Latvia	Yugoslavia
	Lithuania	
	Luxembourg	
Latin America and the Caribbean	Antigua and Barbuda	Guyana
	Argentina	Haiti
	Bahamas	Honduras
	Barbados	Jamaica
	Belize	Mexico
	Bolivia	Nicaragua
	Brazil	Panama
	Chile	Paraguay
	Colombia	Peru
	Costa Rica	Saint Kitts and Nevis
	Cuba	Saint Lucia

World Bank Region	Country	
	Dominica	Saint Vincent and The Grenadines
	Dominican Republic	Suriname
	Ecuador	Trinidad and Tobago
	El Salvador	Uruguay
	Grenada	Venezuela
	Guatemala	
Northern America	Canada	
	United States of America	
Oceania	Australia	Niue
	Cook Islands	Palau
	Federated States of Micronesia	Papua New Guinea
	Fiji	Samoa
	Kiribati	Solomon Islands
	Marshall Islands	Tonga
	Nauru	Tuvalu
	New Zealand	Vanuatu

Appendix 2. Description of serotype data in analysis.

World Bank Region	Country	No. of studies in analysis	No. isolates in analysis
Africa	**N=13**	**N=22**	**N=11,181**
	Algeria	1	45
	Burkina Faso (G)	1	22
	Egypt	1	113
	Ethiopia (G)	1	46
	Gambia (G)	5	539
	Kenya (G)	2	641
	Malawi (G)	1	122
	Mali (G)	2	624
	Mozambique (G)	1	259
	Rwanda (G)	1	130
	South Africa	4	8566
	Tanzania (G)	1	27
	Uganda (G)	1	47
Asia	**N=18**	**N=33**	**N=4,752**
	Bangladesh (G)	5	462
	China	1	236
	China, Hong Kong	1	88
	China, Taiwan	4	584
	India (G)	2	216
	Israel	5	2125
	Japan	3	408
	Lebanon	1	24
	Malaysia	1	20
	Nepal (G)	1	49
	Pakistan (G)	1	87
	Philippines	1	65
	Saudi Arabia	1	51
	South Korea	1	56
	Sri Lanka (G)	1	21
	Thailand	2	173
	Turkey	1	60

World Bank Region	Country	No. of studies in analysis	No. isolates in analysis
	Viet Nam (G)	1	27
Europe	**N=16**	**N=39**	**N=10,279**
	Austria	1	56
	Belgium	3	421
	Czech Republic	1	108
	Denmark	1	1123
	Finland	2	459
	France	3	375
	Greece	1	46
	Italy	3	118
	Norway	2	349
	Portugal	1	80
	Russia	1	37
	Slovenia	1	195
	Spain	7	465
	Sweden	3	177
	Switzerland	1	204
	United Kingdom	8	6066
Latin America and the Caribbean	**N=17**	**N=42**	**N=18,788**
	Argentina	3	1489
	Bolivia (G)	1	149
	Brazil	12	6693
	Chile	5	2279
	Colombia	3	3168
	Cuba (G)	1	841
	Dominican Republic	2	814
	Ecuador	1	55
	El Salvador	1	23
	Guatemala	2	153
	Mexico	2	691
	Nicaragua (G)	1	39
	Panama	1	101
	Paraguay	1	478

World Bank Region	Country	No. of studies in analysis	No. isolates in analysis
	Peru	1	132
	Uruguay	4	1275
	Venezuela	1	408
North America	**N=2**	**N=17**	**N=11,441**
	Canada	4	1057
	United States of America	13	10384
Oceania	**N=4**	**N=16**	**N=3,649**
	Australia	10	2547
	New Caledonia	1	20
	New Zealand	2	752
	Papua New Guinea (G)	3	330

(G) – GAVI-eligible country; N = Number; No. = Number