

Pneumococcal Global Serotype Project

Summary report of Stage 1/Version 1 analysis

Background

Pneumococcal vaccine has been chosen as the pilot vaccine for the Advanced Market Commitment (AMC), a novel funding mechanism to encourage the development of vaccine products and capacity for developing world markets. As part of the AMC process, a Target Product Profile (TPP) for pneumococcal vaccines has to be established. The TPP will describe minimum characteristics required for a pneumococcal vaccine to be eligible for AMC funding. The WHO has been charged with developing the TPP. The WHO requested GAVI's PneumoADIP to undertake a global pneumococcal serotype project (GSP) and analysis, a summary of which is provided here.

Vaccines based on capsular polysaccharides, pneumococcal protein antigens or both, may be developed during the AMC period. Both types of vaccines are potentially eligible for AMC funding. While the TPP should be structured to be equally relevant to both vaccines based on the capsular polysaccharides and protein antigens, the literature on the distribution of protein antigens is not yet well developed and therefore, the capsular serotype distribution is a better parameter on which to set health impact and effectiveness benchmarks for the TPP in the meantime.

The current "state of the art review" of the global and regional serotype distribution of disease causing strains was published in 2000 by Hausdorff et al. However, the data in that review are now mostly over 10 years old and there are key limitations to the existing published analyses. In the interim substantial new data have become available that make this the right time to conduct a new, comprehensive review of data, and to use that data for determining the AMC TPP.

The specific objectives of the GSP included:

- To identify and collect all existing published and unpublished data with information on serotypes of pneumococci causing invasive disease in children less than five years of age, globally.
- To analyze the data collected to understand what variables impact the serotype distribution
- To provide the TPP Expert Committee with an analysis and summary of the serotype burden among young children globally and by region (by sub-region if possible).

- To identify regions or sub-regions where additional data on serotypes of pneumococci causing invasive disease in children less than five years of age is needed.

Methods

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. There were four existing systematic reviews and/or meta-analyses of serotype distribution; one published review of global data by Hausdorff, and three unpublished reviews/meta-analyses for Asia (Kilgore et al.), Latin America and the Caribbean (CDC/PAHO/Sabin), and sub-Saharan African (Sinha et al.). We accepted the list of citations they identified as having relevant data and did not repeat the literature review search for those specific region/time-periods. We supplemented the published literature with unpublished data by contacting key researchers around the world conducting surveillance for invasive pneumococcal disease. We included in the analysis studies with a minimum of 20 reported isolates, 12 months of surveillance, and a study date of 1980 or later, among other characteristics.

The primary outcome was the proportion of isolates due to a particular serotype. There are up to 90 serotypes; information on each was obtained from multiple published articles and summarized using a marginal meta-analysis by the random-effects model. Weighting of the results in a given study was by the inverse of the variance. The World Bank Regions were used to group countries. The analyses were stratified by the primary study covariates, geographic region and one of: age group, syndrome, body fluid, or HIV status.

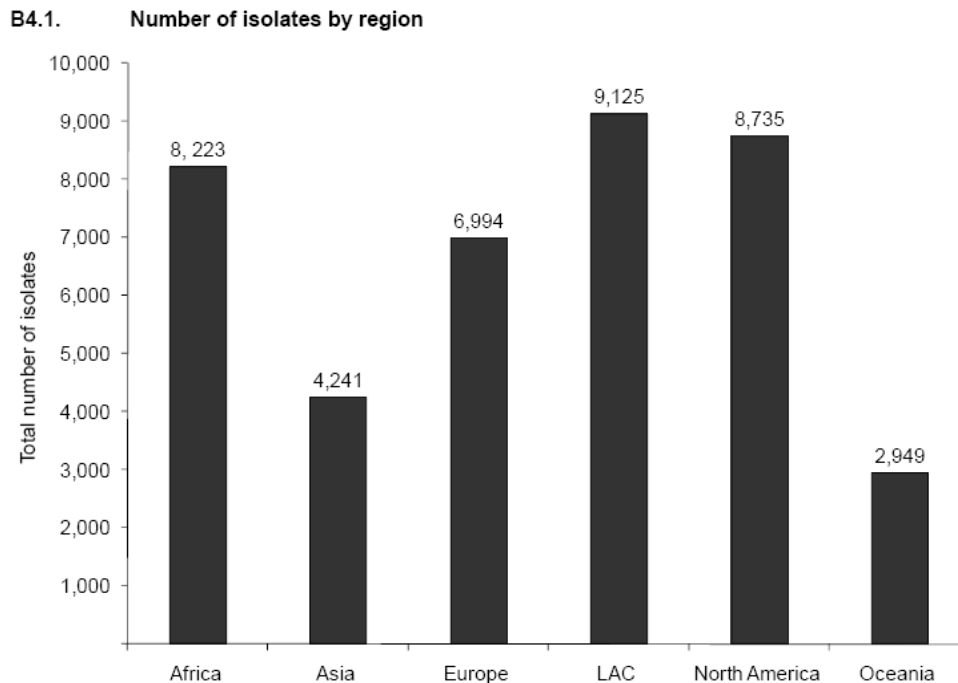
Stage 1 of the analysis is the univariate analysis of the data. Version 1 of this analysis, summarized in this report, is based on the dataset locked on July 31, 2007. Version 2 of the Stage 1 analysis will be the final version, adding data that could not be incorporated for the Version 1 data lock, and providing refined analyses of some variables. Stage 2 of the GSP will be a multivariable analysis.

Results (Stage 1/Version 1)

Analytic Dataset: From 1231 published studies reviewed and 34 datasets provided by investigators, 113 were included in this analysis. An additional 13 were fully data abstracted but excluded from the analyses because they had serogroup, not serotype data. The 126 articles were from 51 countries and had data on over 40,000 pneumococcal isolates. This is a substantial increase in the number of datasets (net

increase of 80), countries (net increase of 15), and isolates (net increase of over 25,000) over what was previously reviewed in four pre-existing meta-analyses combined. Figure 1 shows the number of isolates in the analysis by region.

Figure 1. Number of isolates in GSP analysis by region of isolates (Stage 1/Version 1)

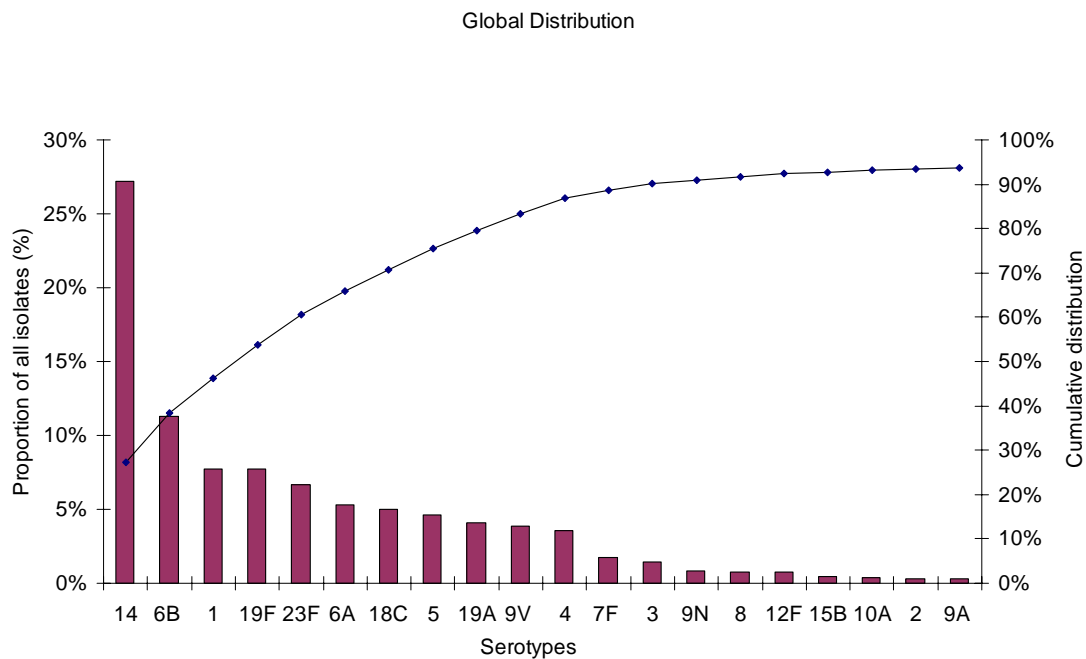


Global Serotype Distribution

Overall findings

The global level analyses cited here are those of the serotypes in the dataset without further adjustment; specifically, they are not weighted by regional pneumococcal disease burden. Therefore this global level serotype distribution is influenced by the regions or countries with the greatest amount of serotype data in the analytic dataset which are not necessarily those with the greatest number of pneumococcal cases or deaths among children less than 5 years of age. The overall global serotype distribution from the Version 1 analysis is shown in Figure 2.

Figure 2. Proportionate and cumulative serotype distribution of invasive disease causing pneumococcal isolates among children < 5 years of age, globally.
(Stage 1/Version1)



- Globally, serotype 14 is the most commonly isolated serotype in children <5 years of age.
- The five most common global serotypes (14, 6B, 1, 19F, 23F) account for approximately 60% of all serotypes causing invasive disease in children <5 years old.
- 11 serotypes account for >80% of all pneumococcal isolates reported.
- Of 90 known pneumococcal serotypes, 15 were not identified in any of the data sets as a cause of invasive pneumococcal disease among children.

Global serotype distribution by age group

- Serotype 1 is more commonly identified among children 2-<5 years of age than among those <2 years of age.

Global serotype distribution by syndrome or body fluid

- Serotype 14 is the most commonly identified isolate from both blood and CSF, and from both pneumonia and from meningitis.
- Serotype 1 is more commonly isolated from pneumonia cases than meningitis cases however, when age distribution is accounted for this finding may be reduced or eliminated.

Global serotype distribution by HIV status

The data available for Version 1 analysis were too sparse to allow analysis and meaningful interpretation.

Regional Serotype Distribution

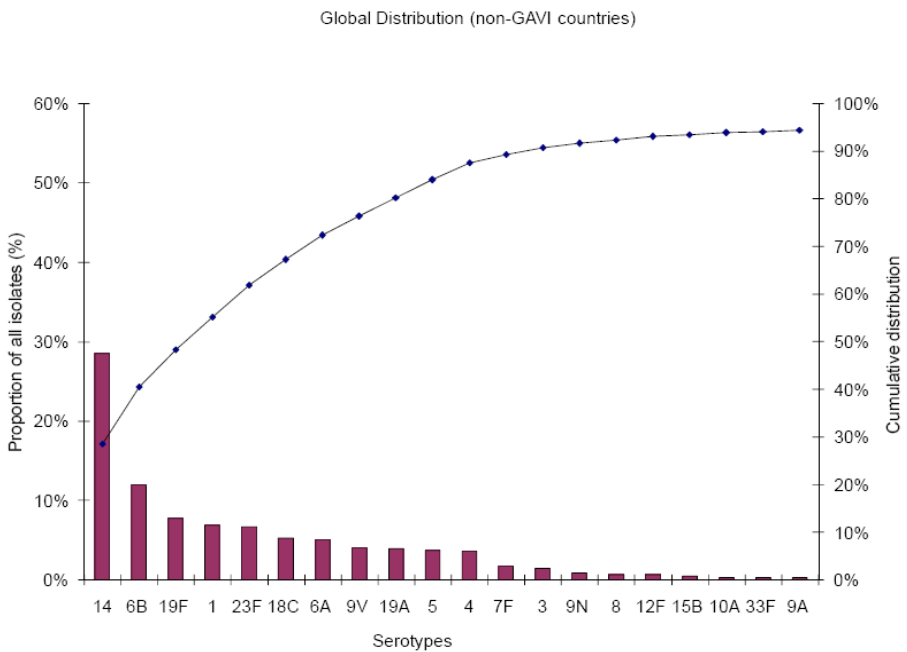
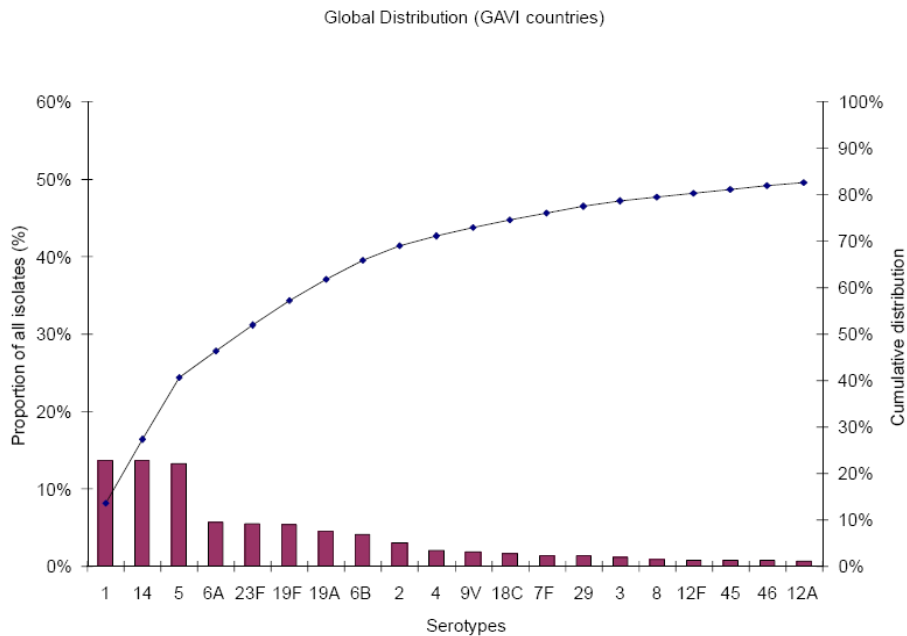
Comparisons of serotype distributions between regions must account for differences in the age distribution of isolates reported from each region. Age is a potential confounder of serotype distributions across regions because:

- age and serotype distribution are associated (i.e., some serotypes are relatively more common among older children than among younger children)
- the age distribution of the isolates reported varies by region. Specifically a greater proportion of the isolates in data sets from Asia and Africa are from 2-<5 year olds than among the datasets from other parts of the world.
- the proportion of serotyped isolates from children aged 2-<5 years of age is greatest in Asia and Africa, therefore summaries of serotype distributions among those “<5 year olds” will tend, relative to other regions, to emphasize serotypes causing disease in children aged 2-<5 years.

Distribution and rank order of serotypes, by region

- Serotype 14 is the most common isolate among children <5 in all regions except Asia where it and serotype 1 are co-ranked as the most common serotypes. [Note that this may change after adjustment for age distribution.]
- In each region, three serotypes (1, 5, and 14) account for between one-third and one-half of disease in children <5 years old.
- Serotype 1 is ranked in the top 4 serotypes in every region, except North America and Oceania.
- In each region, among children <5 years old,
 - 3 to 5 serotypes account for >50% of isolates.
 - 5 to 7 serotypes account for >65% of isolates.
 - 7 to 11 serotypes account for >80% of isolates.
- Africa and Asia share the same top 8 serotypes in the Version 1 analysis. The cumulative proportion of these 8 is somewhat higher in Africa than in Asia, with a range of approximately 70-80%. [With the addition of more data in Version 2 this may modify somewhat]
- Serotype distributions from countries in the dataset that are eligible for GAVI funding (i.e. twelve of 72 GAVI-eligible countries had data in the analytic dataset) have a somewhat different serotype distribution than countries that are not eligible for GAVI funding (Figure 3).

Figure 3. Proportionate and cumulative serotype distribution of isolates from (a) GAVI countries, and (b) non-GAVI countries. (Stage 1/Version1)



Regional serotype distribution by age group

- In all regions, among children <2 years old, serotype 14 is the most commonly isolated serotype.
- In Asia and Africa, among children 2-<5 years old, serotype 1 is the most commonly isolated serotype. In all other regions, serotype 14 is the most commonly isolated serotype.
- In all regions, serotype 1 is relatively more common or higher ranked among children 2-<5 year old as compared to children <2 years old.
- Regional differences in the apparent proportion of disease due to serotype 1 may be explained in part by differences in the age distribution of the children < 5 years of age providing isolates. To put it more precisely, while Africa and Asia appear to have a higher proportion of disease among children <5 years old due to serotype 1, they also have a higher proportion of isolates from children aged 2-<5 years old, and this may in part explain this difference.

Serotype-Specific Disease & Mortality Incidence

One metric to provide 'perspective' on the potential health impact of different vaccine formulations in different regions is to estimate the absolute burden of pneumococcal disease potentially preventable by specific formulations. In short, proportions of disease covered by specific serotypes do not alone inform the relative merit of formulations between regions, because the absolute disease burden of all-serotype pneumococcal disease varies between regions.

Serotype specific disease and mortality incidence is estimated by multiplying the proportion of pneumococcal disease caused by given serotype by the incidence of pneumococcal disease as estimated by the WHO's Pneumococcal Global Disease Burden project. The latter estimates include cases and deaths of both invasive pneumococcal disease and non-bacteremic pneumonia.

The analyses of serotype specific incidence should be interpreted with caution for a variety of methodologic reasons including: (a) the assumption that the serotype distribution of invasive pneumococcal disease cases reflects that of non-bacteremic pneumonia cases (b) the case fatality rate does not vary by serotype (c) there are broad confidence intervals on the Global Disease Burden case and death estimates.

Serotype-specific incidence by region:

- Estimated serotype specific incidence rates vary substantially by region. Serotype 1, for example, varies in incidence from a low of less than 5 cases

per 100,000 children <5 years old in North America to over 500 in Africa, a relative difference of over 100-fold.

- In regions with high pneumococcal disease incidence (Africa, Asia, and Latin America) the incidence of leading individual serotypes exceeds the overall incidence of pneumococcal disease in North America.
- In regions with high pneumococcal disease incidence (Africa, Asia, and Latin America), the top 4 serotypes combined account for an estimated incidence of disease ≥ 1000 cases per 100,000 children <5 years old.

Serotype-specific mortality by region

- Estimated serotype specific mortality rates vary substantially by region. Serotype 14, for example, varies in incidence from a low of 1 cases per 100,000 children <5 years old in North America to over 70 in Africa, a relative difference of over 70-fold.
- In high incidence regions (Africa, Asia, and Latin America) the mortality from leading individual serotypes exceeds the overall mortality of pneumococcal disease in North America.
- In Africa, the top 4 serotypes combined account for an estimated incidence of over 200 pneumococcal deaths per 100,000 children <5 years old. The next highest mortality regions, Asia and LAC, the top 4 serotypes account for approximately 20-30 deaths per 100,000 children <5 years old.

Serotype specific incidence/mortality & multivalent vaccine formulations

It can be useful to consider as “benchmarks” the potential impact of the existing and late-stage products. This can give a basis for understanding how changes in the TPP would lead to different health impacts in different regions.

7-valent vaccine (serotypes 4, 6B, 9V, 14, 18C, 19F, 23F [+ 6A under the assumption that there is cross protection from 6B])

- In every region, the existing 7-valent vaccine is estimated to include 54-75% of serotypes causing disease in children <5 years old.
- There is substantial regional variation in coverage among children <5 years old with the lowest estimated coverage for Africa and Asia .
- The 7-valent serotypes have less inter-regional variability and higher overall coverage among strains from children <2 years old than among those <5 years of age. This is explained by the relationship between age and serotype 1.

10-valent vaccine candidate (7-valent + serotypes 1, 5, 7F [+ 6A under the assumption that there is cross-protection from 6B])

- In all regions, the 10-valent formulation is estimated to include 75-90% of serotypes causing disease in children <5 years old.
- Regional variations are relatively minor, with only Asia having a coverage of <80%.

13-valent vaccine candidate (10-valent + serotypes 3, 6A, 19A [+ 6A under the assumption that there is cross-protection from 6B])

- In all regions, the 13-valent formulation is estimated to include 80-92% of serotypes causing disease in children <5 years old.
- Regional variations are minor, with Asia having the lowest coverage at approximately 80%.

Alternative formulation vaccines

Alternative formulations with fewer serotypes could provide coverage greater than or equal to the existing 7-valent vaccine and nearly as high as the candidate 10-valent and 13-valent vaccines. For example, a vaccine with 6 serotypes (1, 5, 6B, 14, 19F, 23F [+6A assuming cross-protection from 6B]) would provide ~70% coverage worldwide with little variation (from a low of 66% in North America to a high of 76% in Africa). The impact on serotype-specific incidence from this vaccine would exceed the projected impact of the existing 7-valent vaccine in Africa, Asia, and Latin America & the Caribbean. Its impact on incidence would be close to, but less than the incidence prevented by the 10-v and 13-v vaccines. In Europe, North America, and Oceania, its impact on incidence would be less than the existing 7-valent.

Formulations with 3-5 serotypes can account for >50% of disease in a given region, but these would be regional formulation vaccines as the serotypes accounting for 50% of disease are not the same serotypes in all regions.

Discussion

Study Limitations

While the data included in this report (Stage 1;Version 1) represent a substantial improvement over the past, the results here should be considered in light of several important limitations of the data and the process and methods used. First, age (and perhaps other characteristics) may be an important confounder of the regional estimates of serotype proportions. Adjusted calculations which might quantitatively illustrate and correct for these potential biases will require multivariable analyses on raw datasets, of which there are substantially fewer than the total number of datasets in the univariate analysis. In the meantime, readers should consider the direction and potential magnitude of any known bias in their interpretation of the data.

Second, interannual variations in serotype distribution are known to exist. Specific analyses establishing the magnitude of these effects are planned. However, the large volume of data in this review collected over a period of 25 years from around the world should serve to smooth out some of those expected fluctuations.

Third, global analyses of serotype distributions adjusted for disease burden, and population size would inform more clearly at the global level the quantitative contributions of individual serotypes.

Public health impact of various serotypes by region

The TPP is seeking to articulate characteristics of pneumococcal vaccines which will provide a substantial public health benefit and therefore justify the investment by key stakeholders. Key stakeholders include countries, donors, manufacturers and public health agencies. Two key elements of the serotype distribution need to be established: how many and which specific serotypes must be included in vaccines to be eligible for AMC funding. There is an *a priori* belief that serotype distributions vary by geographic region; this has important implications and has been explored in this analysis along with other variables.

The major findings and inferences from this Stage 1;Version 1 analysis are as follows.

- **Age is a potential confounder of serotype 1 distribution by region.**
- **Serotype 14 is the most common serotype in all regions of the world among children < 5 years, and among those <2 years.**
- **Existing 7-valent product provides greater than 50% coverage in all regions but with substantial variation between regions for those < 5 and less variability for those < 2 years.**
- **Planned 10- and 13-valent products provide substantial coverage in all regions with less variability by region or age strata than the 7-valent product.**
- **A limited number of serotypes are common and important across all regions (i.e. there are serotype formulations with a limited number of serotypes that are global formulations).**
- **Vaccines with fewer than 10 serotypes can provide broad protection. With as few as 6 serotypes (assuming 6A/6B cross protection) it should be possible to cover over 70% of all disease worldwide with regional variations ranging from approximately 66-76%. Sub-regional variability may be greater than this.**
- **Products with as few as 3-5 serotypes can provide >50% coverage in each region but would be prone to significant inter-regional variations in impact (and also perhaps inter-annual variation).**

These findings highlight the potential value of the AMC for bridging the needs of developing countries and the needs of industry. For industry, knowing that they can develop vaccines with as few as 6 serotypes which might provide reasonably high serotype coverage in all regions of the world, and which actually favor the distribution in Africa and Asia, could help to increase the likelihood of success in their program. For GAVI countries, this profile would assure that all vaccines developed for the AMC would meet a high basic standard of public health impact.

Implications for TPP

Proportion coverage, number of serotypes and which ones needed

Based on this analysis, it is clear that the committee could set a TPP on the basis of the proportion of serotypes covered. This would be easily understandable by industry as it uses the TPP to prepare a vaccine candidate, and it could be easily measurable against the distributions reported in the GSP, thereby helping to avoid disputes that might be counterproductive to the AMC process.

The challenge is where to set the cut-off point. Establishing the cut-off for the TPP at around 50% coverage might allow manufacturers to develop products with fewer serotypes that could have a relatively high impact in some regions but in others a substantially lower impact. On the other hand, by setting the minimum serotype coverage somewhat higher (~65-70%), the AMC virtually assures the development of “global” formulations that might have broader appeal to more countries. In order to assure that the TPP leads to product research and development that drives manufacturers to invest in vaccines that preferentially benefit GAVI countries it may be useful to require the inclusion of serotype 1 or another serotype that is relatively more common in GAVI countries or Africa than it non-GAVI countries, North America or Europe for example.

Rank order of serotypes

While it is important to understand the rank orders (or at least the sets that comprise the top 50% or 70% of isolates), selecting a relative cut-off rather than an absolute cut-off could lead to substantial inter-regional differences in proportion of disease prevented. Specifying, for example, that the vaccine must include the “Top 5” in any region, could result in a vaccine that addresses 30% of disease in one region and 80% of disease in another if the distribution of serotypes is broader in the former region than the latter. As such it would be preferable to specify an absolute cut-off in terms of the proportion of isolates covered in the vaccine.

Serotype specific incidence

This measure is conceptually satisfying as a public health metric for targeted disease prevention as it provides an absolute measure of expected public health impact. However, on further reflection and analysis of the implications of its use there are significant practical limitations described in the document.

Serotype specific incidence is however a useful analysis for the TPP committee deliberations. At the country level the expected public health impact will likely be a major component of policy decision making about vaccine introduction. Therefore, the TPP committee may find the serotype specific incidence analysis useful for providing 'reality checking' whether the expected health impact of any given formulation would likely provide a compelling case for country level introduction. It also provides a means to benchmark formulations against other vaccine interventions that are routinely used or are expected to be routinely used (e.g., Hib or meningococcal vaccination).

In summary the serotype specific incidence lacks the characteristics to be the public health metric for setting the TPP (i.e. precision and stability over time) but provides a context for understanding the likely public health impact of various formulations. It should probably be used as a guide for the TPP committee but not as a criterion in the TPP.