Serogroup-Specific Epidemiology of *Streptococcus pneumoniae*: Associations with Age, Sex, and Geography in 7,000 Episodes of Invasive Disease


From the Communicable Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, London, England; Pediatric Infectious Diseases Unit, Soroka Medical Center, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel; Provincial Laboratory of Public Health, University of Alberta, Edmonton, Alberta, Canada; Division of Microbiology, St. Thomas' Hospital, London, England; Servicio de Bacteriología, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain; Microbiology Unit, Central Public Health Laboratory, Montevideo, Uruguay; Laboratoire de Santé Publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada; Department of Pathology, University of Texas Health Science Center, San Antonio, Texas, USA; Service de Microbiologie et Maladies Infectieuses, Hôpital Saint-Luc, Centre Hospitalier affilié à l'Université de Montréal, Montréal, Québec, Canada; Microbiology Service, Hospital Saint Joan de Déu, Barcelona, Spain; Department of Respiratory Medicine, Nottingham City Hospital, Nottingham, England; Microbiology Unit, Veterans Administration Medical Center, Cleveland, Ohio, USA; Scottish Meningococcus and Pneumococcus Reference Laboratory, Department of Laboratory Medicine, Ruchill Hospital, Glasgow, Scotland; and Department of Bacteriology, Instituto Adolfo Lutz, São Paulo, Brazil

A study sample of 7,010 episodes of invasive *Streptococcus pneumoniae* disease was obtained by combining 13 existing datasets. Disease episodes due to each of 12 pneumococcal serogroups (1, 3–9, 14, 18, 19, and 23) were then compared with episodes in a constant internal control group to describe serogroup-specific variations in disease frequency by age, sex, and geographic origin. The results are presented as odds ratios (with 95% confidence intervals) derived by logistic regression, with adjustment for the major confounders, including dataset of origin. Variation in the male:female ratios between serogroups is small, suggesting that capsular characteristics are an unlikely explanation for the male preference of *S. pneumoniae*. Serogroups associated with higher nasopharyngeal prevalence (e.g., 19 and 23) are relatively more common in Europe and North America, while the invasive serotypes 1 and 5 are much more common in South America. The custom of reporting serogroup frequencies in two age groups, children and adults, conceals much of the variation in the age distributions across the whole span of life. The reduction of risk associated with serogroups 6, 14, 18, 19, and 23 beyond childhood follows different gradients, being most abrupt in serotype 14 and most gradual in serogroup 18. The relative risk of disease with serotype 1 declines steadily throughout life, while with serotypes 3 and 8 it increases over middle age. Serogroups 7 and 23 are found unusually frequently in the third decade of life. Because of the wide differences in the epidemiology of individual serogroups of *S. pneumoniae*, it is questionable whether pneumococcal infection should continue to be classified as a single disease entity.

The capsular serotype of *Streptococcus pneumoniae* is the most important subclassification of the species because it is the strongest known influence on human immunity [1]. Susceptibility to invasive disease is determined by the ability of the host to generate specific opsonizing antibody against capsular antigens [2, 3]. Any one serotype will be at an advantage if the homologous antibody response in the host is blunted either by lack of previous exposure or by variation in the maturation of humoral immunity with age. The maturation of children's antibody responses to vaccine antigens has been shown to vary widely with serotype [4, 5], but the age distribution of invasive disease due to different pneumococcal serotypes has not yet been examined in detail.

An important determinant of the ecological success of *S. pneumoniae* is its ability to transfer from one host to another in different environments. The density, age structure, and socioeconomic conditions of human populations all affect the indices of transmission, namely, the incidence of pneumococcal pneumonia, the number of effective contacts (i.e., contact between two individuals in which transmission occurs) per case, the

Received 25 August 1995; revised 2 January 1996.

Financial support: Dr. Scott is supported by the Wellcome Trust of Great Britain through a research training fellowship in clinical epidemiology (no. 035375).

Reprints or correspondence: Dr. J. Anthony G. Scott, Wellcome Trust/KEMRI-CRC, Kilifi Research Unit, P.O. Box 230, Kilifi, Kenya.

Clinical Infectious Diseases 1996;22:973–81
© 1996 by The University of Chicago. All rights reserved.
1058-4838/96/2206-0010$02.00
prevalence of carriers, and the rate of effective contact between carriers and uninfected individuals. Since there is marked variation among serotypes of S. pneumoniae in their propensity to colonize the nasopharynx [6], they exploit different human environments with varying degrees of success. The geographic distribution of a serotype is likely to reflect the environmental characteristics that most suit its transmission.

Pneumococcal disease has a consistent preference for males that remains unexplained [7, 8]. There is some evidence to suggest that the male:female ratio varies with serotype; for example, the ratio of male to female patients with meningitis in Dakar, Senegal, ranged from 0.75:1 (among cases due to serotype 3) to 2.0:1 (among those due to serogroup 23), although these estimates were not tested statistically [9]. If the association with males is confined to a few serotypes, then the microbiological and epidemiologic characteristics common to these types might suggest an explanation for their predominance among males.

Despite the potential insights gained from studying the distribution of pneumococcal serotypes as related to age, sex, and geographic factors, several problems have limited such study. First, as there are 90 different serotypes [10, 11], studies of small numbers of isolates have little discriminating power. Second, the distribution of pneumococcal serotypes is likely to vary significantly with each of these interassociated variables, thus rendering a univariate analysis susceptible to confounding. Third, the analysis must be undertaken from the perspective of the serotype and not from that of the hospital or public health system. Many collections of isolates have been assembled to guide the selection of antigens for pneumococcal vaccines, but these are concerned with the rank order of serotypes or with the cumulative percentage of potential vaccine serotypes.

To address these problems, we collated a large dataset of isolates from 13 existing studies in which information on age, sex, geographic origin, and anatomic source was available for each isolate; a form of analysis was selected that allows direct comparison of the distributions of each variable among the 12 most prevalent serogroups.

**Methods**

A systematic search was undertaken with use of the MEDLINE database (1987–1993), Index Medicus (1965–1987), and major textbooks of microbiology in order to find all studies involving the serotyping of S. pneumoniae that had been reported since 1940. Following this search, secondary and tertiary citations from the reports identified in the initial search were examined. Investigations of both invasive disease and asymptomatic nasopharyngeal carriage were included. Studies were excluded if it was obvious from the report that there was no information on either the age or sex of the patients or if there was no named author or address for correspondence. All of the corresponding authors were sent a letter asking them to participate in the study and to contribute their original database, specifically with regard to age, sex, and geographic factors.

Further information on the nature of each dataset was collated: the period of collection, the age restrictions of the population studied, and the geographic range of collection. To prevent duplication, contributors were asked to indicate whether the isolates had been sent to reference laboratories or epidemiologic surveillance institutions.

Each dataset was entered into a common database format and checked. Duplicate serotypes from the same patient-episode were removed from the database, leaving only the first reported isolate. If two serotypes were reported from the same specimen, these were included as two separate entries with the same epidemiologic characteristics. When referrals were reported, isolates were removed from the receiving institution’s dataset if they matched an isolate in the referring institution’s dataset with regard to age, sex, serogroup, and anatomic source.

Analyses were conducted on the dataset combined from all sources. The distributions as related to age, sex, and geographic origin were explored for each serogroup. As there was no reliable information on the structure of the denominator populations from which the specimens were drawn, the patterns for individual serogroups were compared with an internal standard. The distributions of the 12 most frequent serogroups were compared individually with the distribution of all other serogroups, which were combined into a single reference group.

Odds ratios were calculated for the occurrence of the 12 most prevalent serogroups vs. that of the reference group in each exposure stratum; exposures examined were age (in decades), sex, and country of origin. Odds ratios and 95% confidence intervals were estimated by means of logistic regression. The odds ratio in each baseline exposure stratum (e.g., the youngest age group) was therefore set at one. In the analysis of age and sex patterns, adjustment for differences between the 13 datasets was achieved by introducing a term into the regression model for each contributing dataset, after statistical interactions dataset were excluded.

**Results**

The search of the medical literature identified 261 studies reported since 1940 in which pneumococcal serotyping was performed and either the age or the sex of the infected subject was identified. Since several publications were by the same author, this list was reduced to 126 investigators; an address was available for 120. When contacted, 45 of the 120 investigators replied, of whom 13 had suitable data records and agreed to participate in the study. The combination of these 13 datasets forms the substance of this report.

The characteristics of the contributing datasets are illustrated in table 1. (Herein, datasets are referred to by the contributors’ names shown in table 1.) One dataset (Hortal) included 52 isolates from asymptomatic carriers, which have not been considered further in this analysis. All of the remaining isolates were from cases of disease. Seven specimens, in three different datasets, were found to contain two different pneumococcal serotypes. Duplicate reports of the same isolate, in different
Table 1. Characteristics of the contributing datasets.

<table>
<thead>
<tr>
<th>Dataset origin</th>
<th>Study period</th>
<th>Age restriction for participants</th>
<th>Geographic area</th>
<th>Study type\size (no. of isolates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macfarlane [16]</td>
<td>1980–81</td>
<td>15–79 y (most)</td>
<td>Nottingham, UK</td>
<td>RP/89</td>
</tr>
<tr>
<td>Smart [17]</td>
<td>1982–86</td>
<td>None</td>
<td>Scotland</td>
<td>RP/688</td>
</tr>
<tr>
<td>Lamothe [18]</td>
<td>1984–86</td>
<td>None</td>
<td>Quebec province</td>
<td>RP/468</td>
</tr>
<tr>
<td>Jeté [22]</td>
<td>1989–92</td>
<td>None</td>
<td>Quebec province</td>
<td>RS/220</td>
</tr>
<tr>
<td>Latorre [23]</td>
<td>1990–93</td>
<td>&lt;18 y (most)</td>
<td>Catalonia, Spain</td>
<td>RS/132</td>
</tr>
<tr>
<td>Shlaes [24]</td>
<td>1990–93</td>
<td>Adults</td>
<td>Cleveland, Ohio, USA</td>
<td>RS/194</td>
</tr>
</tbody>
</table>

* Publications relating to these datasets are referenced against the names of contributors to the present study. In most instances the datasets included in this study contain more isolates than in the original publications. In some instances fewer isolates were available for this analysis.

† RP = research project; RS = routine surveillance.

datasets, occurred in 199 instances; when these duplications were eliminated, the total number of isolates was 9,147 (table 1). Information regarding serogroup was available for 8,297; of these, 1,141 were from sputum and 7,010 (84%) were from normally sterile sites. The analyses described below were restricted to the group of sterile-site isolates.

For one dataset (Macfarlane), counter-immunoelectrophoresis was used to detect type-specific antigens in body fluids; in the remaining 12 datasets the serotype was obtained by culture of *S. pneumoniae*. The term serotype refers to a unique capsular antigen and serogroup refers to a collection of antigenically related serotypes [10]. Combinations of both serotypes and serogroups are described here collectively as serogroups. Four of the contributing datasets reported serotypes in addition to serogroups. Among these, some typeable isolates were reported only by serogroup. For both these reasons, an analysis by serotype was not feasible, so these, some typeable isolates were reported only by serogroup. The proportion of males between different serogroups was

The proportional frequency of each serogroup, in rank order, is shown in figure 1. The most frequent was serogroup 14 (10.5%), followed in turn by groups 6, 19, 3, 23, 1, 9, 4, 8, 18, 7, and 5. These 12 serogroups accounted for 80.9% of all isolates. There were no isolates from each of serogroups 43, 44, and 47. The reference group used in the analysis therefore consisted of the remaining 30 serogroups, comprising 1,336 isolates (19.1% of the total dataset). The numbers of isolates for which there were complete data concerning sex and age of participants were 6,875 and 4,843, respectively.

Age

More than one-half of the isolates came from patients at the extremes of life: 34% were aged <5 years and 20% were aged ≥65 years. The comparison of serogroups 1, 3–9, 14, 18, 19, and 23 against the reference group revealed marked differences in their distribution as related to age. As these patterns were thought to be confounded by variation in the demographic structure of the contributing populations and by geographic variation in the prevalence of serogroups, the odds ratios were adjusted for the dataset of origin. These results, together with their 95% confidence limits, are presented in figure 2. Although adjustment produced statistical improvements in the models, the patterns of risk with age were altered perceptibly only for serotypes 5 and 14.

For patients with pneumococcal disease, these odds ratios indicate the risk, attributable to age, that the infection may be due to the serogroup specified rather than to one of the 30 serogroups in the reference group. Serotype 1 is associated with a progressive decline in relative risk throughout adulthood ($\chi^2$ trend $= 25.4$; $P < .0001$). In contrast, the risk of serotype 3 disease increases progressively up to a peak in the seventh decade of life. Serotype 8 is the only other group examined that has a relative preference for adults. Serogroups 6, 14, and 19 are associated with an abrupt reduction in risk beyond the first decade, as compared with the risk for the reference pneumococci. The same effect is seen, although to a lesser extent, in serogroup 23. Serogroup 18, which also has a relative predilection for the first decade of life, shows a more gradual decline in risk over 3 decades. There is an increase in risk for serogroups 7 and 23, which is sustained for only 1 decade, between the ages of 20 and 29 years. Interactions between dataset and age were examined for each serogroup, but the results across each contributing dataset were relatively consistent and no statistically significant interactions were observed.

Sex

The proportion of all isolates that were recovered from male patients was 0.64 (male:female ratio = 1.8:1). The variation in the proportion of males between different serogroups was
not marked, and the overall association between sex and serogroup was not statistically significant. Logistic regression analysis of the effect of the patients’ sex upon the 12 individual serogroups (1, 3–9, 14, 18, 19, and 23), with adjustments both for age (in 10-year bands) and for the dataset of origin, revealed a small relative preference for females in two serogroups; the odds ratios (95% CIs) for serogroups 14 and 23 in comparison with the reference group were 1.30 (1.04–1.62) and 1.32 (1.01–1.71). No significant association was observed in the other 10 analyses.

Geographic Origin

The data were grouped according to their geographic origin: 2,892 came from Spain, 1,249 from Britain, 693 from Canada, 146 from the United States, 1,713 from Brazil and Uruguay combined, and 317 from Israel. The isolates from Israel were recovered from children and could not be used in a summary analysis involving the whole span of life. The proportional distributions of each serogroup between the remaining five geographic groups showed distinct differences. International variations in preferences for the types of specimens collected and the different demographic structures of the various countries may influence the association between serogroups and geographic location. Therefore, when odds ratios were calculated for the 12 most frequent serogroups in comparison with the reference group, they were adjusted for the source of the specimen (blood, CSF, or other sterile sources) and for age in 10-year bands. These odds ratios, with 95% confidence limits, are shown in figure 3.

The widest variation is seen in serotypes 1 and 5, which are associated with a marked increase in risk for those living in South America (vs. all other areas). Both types are especially uncommon in the United States and Canada. Outside South America, Spain showed a relative excess of serotype 5 and Britain a relative excess of serotype 1. The risk of disease due to serotype 14 is also slightly higher in Britain. There is some similarity in the distributions of serotypes 3 and 8, both of which are rare in Canada and the United States. Conversely, serotype 4 is more common in North America than in Europe or South America. Serogroups 19 and 23 are distinctly less common in Brazil and Uruguay than elsewhere.

With use of the likelihood ratio test, each of the regression models examining age and geography was highly significant \((P < .001)\). The fit of the model to the data was calculated by dividing the deviance of each model by the degrees of freedom. For all but one serogroup, this value was between 0.85 and 1.17, suggesting that the variation in the data had been well accounted for by these two variables. In the case of serotype 5, some residual variation existed after age and geography were accounted for (goodness of fit: 0.65).

Discussion

The analysis included only a sample of the total number of studies (published and unpublished) on serotype patterns of \textit{S. pneumoniae}; was the dataset obtained representative of pneumococcal epidemiology in general? The age distribution represented by the combined dataset, with more than one-half of the data from patients at the extremes of life, follows the established pattern for pneumococcal disease [7, 8, 25]. Simi-
larly, a predominance of males is invariable in studies of
*S. pneumoniae*, and the male:female ratio of 1.8:1 obtained
here is typical [8, 9, 26, 27]. Examination of the relative distri-
bution of pneumococcal serogroups showed that the 10 groups
encountered most frequently in this study differ by only 1 from
the list of 10 most commonly observed in the two largest series
published recently, each with >10,000 isolates [28, 29].

The aim of the study was to examine how different sero-
groups varied with regard to age, sex, and geographic factors.
The ideal approach is a comparison of serotype-specific inci-
dence rates at each exposure level. As the denominator popu-
lations for the 13 contributory studies were not defined, a compar-
isation of rates was not possible; an alternative approach was
taken, comparing the risk of disease with a given serogroup
against the risk of disease with a control or reference group.
The fact that the same control group was used for each of the
12 analyses validates the comparison of any one serogroup
pattern with those of the other 11.

For example, the relative risk of serotype 14 falls dramati-
cally between the first and second decades of life (OR, 0.09),
while the relative risk for serotype 1 remains unchanged (OR,
1.08). In general, the incidence of pneumococcal disease falls
sharply between the first and second decades, so these findings
should be interpreted as follows: the incidence of serotype 1

---

**Figure 2.** Dataset-adjusted odds ratios and 95% confidence limits for
12 serogroups, in comparison with
the reference group, per age in 10-
year bands.
Figure 3. Age- and source-adjusted odds ratios and 95% confidence limits for 12 serogroups, in comparison with the reference group, per country of origin.
disease follows the standard pattern for pneumococcal disease in general (represented in this study by the reference group), but the fall in incidence of serotype 14 is more than 10 times greater. It may be argued that the control group, comprising the 30 least common serogroups, is hardly representative of pneumococcal disease as a whole. However, one of the findings of this study is that individual serogroups vary so widely with regard to age and geographic patterns that no single type or collection of types would be likely to meet this criterion. The only essential requirement for the control group is that it should be consistent for each of the 12 analyses performed.

The choice of an internal control group is particularly advantageous in the analysis of age and sex patterns. Variations in the study populations and in sampling and laboratory methods may confound a simple analysis of different datasets. However, because each dataset contributes information to both the study group and the control group, this potentially confounding effect can be eliminated by including in the regression analysis a variable defining the source-dataset of each isolate. The absence of statistical interaction between the contributory datasets and age means that serotype-specific age patterns do not vary markedly from one dataset to another and can therefore be combined into a single summary result for the study as a whole.

The data were analyzed by serogroup rather than serotype because of inconsistency in the reporting of serotypes. Serogroups of *S. pneumoniae* are collections of serotypes that provoke cross-reactive antibody responses in rabbits [10]. What is the value of examining the epidemiologic characteristics of these groupings in humans? Of the 12 categories analyzed here, six are serotypes that cannot be subdivided (1, 3, 4, 5, 8, and 14). Of the remaining serogroups, three (7, 18, and 23) are dominated numerically by a single serotype, and a group-based analysis will largely reflect the behavior of these types (7F, 18C, and 23F).

Serogroup 6 comprises two serotypes that have similar chemical structures and are mutually cross-reactive in human experiments [28]. For exposures considered here, between-group differences are likely to be large compared to the within-group differences of serogroup 6. The poorer human cross-reactivity of the clinically important serotypes of groups 9 and 19 makes their type-differentiation more important [28]. Serogroup 19 has two and serogroup 9 has three clinically important serotypes. Failure to distinguish the type-specific patterns in serogroup 9 may explain why the age and geographic distributions of the group are those that differ least from the reference group.

Age is a strong determinant of serogroup risk, and the different serogroups that predominate in children and adults are well documented [29–32]. The reduction of age to a dichotomous variable, however, conceals the way in which serogroups distribute themselves across the whole span of life.

The excess of risk, concentrated on the first decade of life, in serogroups 6, 14, 18, 19, and 23 accords with our existing knowledge of maturing antibody responses. In children aged 5–7 years, postimmunization levels of antibody to serotypes 6A, 14, 18C, and 19F are low. By the age of 8–10 years they are higher, but there is no further elevation of titer with age; in contrast, the antibody response to serotype 23F continues to rise at least up to the age of 13–15 years [5]. Although the age distributions for these serogroups all show an excess in childhood, there is diversity in the patterns beyond the first decade. For example, in serotype 14 the change is large and abrupt, but in serogroup 18 the risk declines gradually. This difference is not reflected in the serotype-specific antibody maturation patterns. The explanation may lie in the preference of serogroup 18 for meningitis [27, 33] and in the fact that pneumococcal infections other than meningitis are relatively uncommon in adolescence.

The most dramatic age-related changes are seen in serotypes 1 and 3, which have opposite trends. The relative risk of disease due to serotype 1 falls progressively throughout the whole span of life. Although serotype 1 is rarely found to be colonizing the nasopharynx [34–36], exposure to the type—infected from preimmunization antibody prevalence—is relatively high even in infancy [4, 5]. However, serotype 1 polysaccharide is only weakly immunogenic, and the mean rise in antibody titer on stimulation with vaccine antigen is poor at any age. For example, among children 13–15 years old, only 29% have a twofold rise in titer of antibody to serotype 1, vs. >90% to serotypes 2, 4, 6A, 9N, and 18C [5]. Given these immunologic characteristics, a plausible explanation of the age distribution observed is that multiple exposures are necessary to acquire immunity and that a decreasing probability of invasive disease is associated with each successive exposure.

In 1937, Finland et al. described the association of serotype 3 with adult disease [37]. Our results demonstrate a risk that rises throughout the whole of middle age. Serotype 3 is highly immunogenic at all ages, including young children, and antibody is detectable at a high titer from <2 years of age [4, 5, 38]. In contrast to that for serogroups 6, 18, 19, and 23, therefore, the explanation for the age distribution of disease due to serogroup 3 is likely to lie outside the maturation of antibody responses.

Serotype 3 has a thick mucoid capsule that suggests it may have a physical advantage in certain anatomic foci; for example, it is a common cause of otitis media [30, 39–41]. Although it is a potent stimulus of specific antibody, serotype 3 is poorly opsonized and evades phagocytosis in vitro [3, 42]. This may help explain its wide anatomic distribution, with many extrapulmonary sites, in disease case series [43]. Another unusual feature of serotype 3 is that it is a common cause of bronchopneumonia rather than lobar pneumonia, particularly in patients with chronically damaged lungs [1, 37]. The age prevalence of chronic lung damage is similar to the pattern of risk for serotype 3 and may be an explanatory factor of its age-related distribution.

Serotype 8 also has an abundant capsule, is immunogenic in young children [4, 38], is relatively resistant to phagocytosis in vitro [3], and has a relative preference for bronchopneumonia [37, 43]. It is also the only other type in our study that is associated with an increased relative risk in adulthood.
The age-incidence curve for all pneumococcal disease remains poorly defined because of the difficulty in characterizing the age structure of denominator populations. Existing information suggests that, in addition to the excess of disease at the extremes of life, there is also a smaller peak in the decade of ages 20–29 years [44]. Young adults may have an increase in risk because of their role in child-rearing, since young children have high rates both of disease and of nasopharyngeal carriage. In our study, serogroups 7 and 23 were associated with an increase in relative risk that was confined to this decade. If children are a source of infection for their parents, they are likely to transmit a serogroup commonly carried in the nasopharynx (e.g., 6, 14, 18, 19, or 23). Because antibody responses to most of these antigens are mature by the age of 10, disease is unlikely to develop in adults on exposure. However, as observed above, the response to serotype 23F has a prolonged maturation course in adolescence, the upper limit of which has not yet been described [5].

Serogroup 7 is an invasive, epidemic group [39] that is found in the nasopharynx of children relatively infrequently [35, 41]. It is difficult to explain the nature of its pathogenic advantage in this age group; perhaps when it does infect the nasopharynx, the corya of the child produces an amplification of the inoculum required for successful transmission to adults.

The sex of an individual has little influence upon the risk of invasive disease with particular pneumococcal serogroups. Of the 12 analyses performed, only two produced associations of significance at the 0.05 probability level, and even in these the size of the effect was small. The male preference of pneumococcal disease is unlikely to be explained by mechanisms mediated by serogroup.

Geographic comparisons cannot be separated from the history of serotype distributions. When large-scale observations of serotypes were made in the 1930s, pneumococcal disease was dominated by a small number of groups, particularly 1, 2, 3, 5, 7, and 8 [1, 43]. The microbiology records of serotyped isolates from Boston City Hospital have been used to document the gradual decline of types 1, 2, and 5 throughout the period 1935–1974 [43]. In the 1980s groups 3 and 7 also became less frequent in Boston, and the breadth of serotypes isolated widened [31]. As these invasive groups declined, those associated primarily with pediatric disease and with nasopharyngeal colonization—serogroups 6, 18, 19, and 23—increased in frequency. Similar observations from the Statens Seruminstitut in Denmark between 1939 and 1969 also record a decline in the frequency of serotypes 1, 2, and 5 while serogroups 4, 8, 14, 18, and 23 increased their representation [33, 39].

Recent collections of isolates from West, South, and Central Africa reflect a distribution similar to that in Boston and Denmark in the 1930s [9, 45, 46]. Serotypes 1, 2, and 5 are still the dominant causes of disease and serogroups 19 and 23 are, by contrast, relatively rare. The data presented here further support and expand these observations. The relative risk of serotypes 1 and 5 is extremely low (and there were no observations of serotype 2) in North America; conversely, the same types are common in South America. Of the pediatric types, serogroups 19 and 23 are particularly uncommon in Brazil and Uruguay.

It is tempting to attribute the predominance of the low-order serotypes in present-day Africa, in Brazil and Uruguay, and in Denmark and Boston in the 1930s to the common socioeconomic conditions of their environments, with poverty and overcrowding leading to an increased exposure to active cases of disease. In the less crowded environment of developed countries today, where patients with active disease are treated promptly or isolated in the hospital, the types associated with prolonged nasal carriage appear to be more prevalent.

We have observed considerable variation in the basic epidemiologic characteristics of invasive disease caused by different pneumococcal serogroups. In the distributions associated with age, there is as much variation between serogroups as one might expect between different species of respiratory bacterial pathogens. Studies of S. pneumoniae that fail to distinguish the serogroups involved in individual infections will confuse the features of quite different etiological agents and hinder the elucidation of the environmental, host-, and pathogen-determined risk factors that are of critical importance to public health strategy.

References
13. Taunay AE, Austrian R, Landgraf IM, Vieira MFP, Melles CEA. Sorotipos de Streptococcus pneumoniae isolados de líquido cefalorraquidiano no


