Reviewer's report

Title: The dynamics of nasopharyngeal streptococcus pneumoniae carriage among rural Gambian mother-infant pairs

Version: 1 Date: 3 February 2010

Reviewer: Simo Granat

Reviewer's report:

General comments

The manuscript describes some basic features in the development of nasopharyngeal carriage of pneumococci in Gambian infants, and the association of carriage with carriage in their mothers.

The study found a rapid increase in carriage starting at birth, and a stable, high prevalence from two months of age onwards. These findings are in line with previous findings from both The Gambia, as well as from other developing countries.

The pattern of a relatively small number of serotypes being responsible for the majority of all isolated serotypes, is likewise what one would expect based on previous data. The most common serotypes found in this material correspond well with reports on nasopharyngeal carriage in children in other parts of the world.

The study design is appropriate enough for its purpose, though it suffers from the fairly long intervals between samples, and from the fact that no other family members than the infants and the mothers were sampled. These shortcomings have been acknowledged by the authors. The few sampling points, and the long intervals between them, may lead to underestimation of carriage of rare serotypes with a rapid clearance of carriage.

In order to better describe the dynamics of pneumococcal acquisition and transmission within the family, it would have been worthwhile to swab also the other family members, particularly siblings. Now, only the mothers have been swabbed in addition to the infants, giving only a limited insight into the role of family exposure in acquisition and in the patterns of transmission and carriage within the family. It is, however, interesting to note the doubling in the carriage in the mothers following the birth of the infant.

The manuscript is concisely written and easy to follow. However, there are parts that need to be spelled out better, so as to the reader unversed in statistical methods to better grasp how the presented results and conclusions are actually reached (please see our detailed comments below).

The manuscripts findings are in-line with previous reports on pneumococcal
carriage in different parts of the world. The manuscript does not really provide novel or advanced knowledge on the patterns or the dynamics of natural pneumococcal carriage. The study is, however, with its limitations, well performed and the manuscript mostly well written, and adds to the epidemiological data on natural pneumococcal carriage available. There are major methodological questions regarding the analysis of the dispersion of serotypes, which needs to be sorted out, please see our detailed comments below. In general, the manuscript would be of interest to people studying pneumococcal carriage.

Major Compulsory Revisions

1. p.6, the last paragraph: In some regression analyses log(prevalence) was used as a predictor for log(odds). This applies at least the regression model to investigate seasonality in carriage prevalence. Were there other?

More importantly, this choice of predicting odds (i.e. basically prevalence) by prevalence is an odd and unorthodox choice. The motivation to avoid serotype specific categorical explanatory variables seems very plausible. However, please clarify where this assumption was actually used (i.e. for what analyses). Can you please give a more formal (and even analytical) justification and interpretation of this.

2. It is not clear from the description on page 8 how the permutation test was realised. There are several related questions that need clarification:
   a. Did you allocate serotypes to all sampling occasions of the individual, irrespective of whether the individual carried pneumococcus in the actual data? If so, would it not have been more reasonable to allocate serotypes only to those samples at which the individual carried pneumococcus?
   b. Why is it that, according to Figure 3a, some infants seem to have carried more than 4 serotypes? There were 4 samples at maximum per infant and only one colony was serotyped (according to line 7, page 6). So, shouldn’t the maximum number of serotypes be 4?
   c. The same question applies to Figure3b (the mothers).

3. The interpretation of the permutation test depends heavily on the sampling interval. For example, had all samples from the individual been taken so frequently that they represent a single episode of carriage, they would represent the same serotype and would thus exhibit obvious ‘under-dispersion’ with respect to the distribution of serotypes in the cohort at large. This would be in a way pure sampling artefact. The question of the impact of sampling interval on the finding needs to be discussed properly.

4. Subsequently, if it can be assumed that the sampling interval is long enough so that (most) samples represent a new episode, one can process with the question of ‘under-dispersion’. Then again, whether one describes the finding as under- or over-dispersion depends on the chose measure. In general, over-dispersion follows from heterogeneity across individuals in proneness to the
outcome measure (e.g. carriage of a particular serotype). Looking at the patterns of serotype-specific carriage in this way, the distribution of the times (number of samples) a child carries a given serotype is over-dispersed. Looking the same data in the way you have defined the measure (the number of difference serotypes per child) is under-dispersed. The above discussion should convince us that the statement on page 13 (lines 6-5 from the bottom) that there is a strong tendency in nature for over-dispersion still holds. The new data you report here merely adds to the same conclusion! Please clarify your argument and adjust the discussion and presentation of results accordingly.

5. Still related to ‘under-dispersion’, the most obvious reason to it, apart from continuing carriage in the same child, is transmission. Transmission of pneumococcal carriage is known to induce serotype-specific ‘micro-epidemics’ (see e.g. Leino et al, Clustering of serotypes in a longitudinal study of Streptococcus pneumoniae carriage in three day care centres, BMC Infect Dis 2008, 8:183). It is possible that also the families and larger units (compounds) in the Gambia exhibit the same type of behaviour in which a certain serotype is circulating in that mixing unit for a longer time, with possible re-acquisitions of the type in the individuals. It is true that some type of immunity may additionally modulate this phenomenon, but it is not clear that immunity should be the first culprit here.

6. What has been said above means also that we feel the conclusion (see page 3) that immunity plays a role in the dynamics of carriage somewhat unwarranted, based on the present data and their analysis.

7. Page 8, last chapter: The way the half life of the pneumococcal colonisation is calculated assumes implicitly that if a serotype is found at a given time point, carriage has not been cleared in the observed interval. This assumption excludes the possibility of a clearance and a subsequent reacquisition of the same serotype taking place at a later stage during the interval. A reacquisition of a given serotype in the interval would be plausible if e.g. a given strain was circulating in the family. Particularly, as the intervals between the samples are quite long, the possibility of reacquisition and what effects it would have on the estimated durations (half-lives) of carriage should be discussed, at the least.

8. Page 9, results, fourth sentence: There appears to be a discrepancy between the reported “All study subjects carried pneumococci at some point during the study” and what is presented in both table 1 and Fig 3 (a and b). In the table, and the figures, a proportion of the study subjects appear to have not ever encountered pneumococci (table 1, 98.0% of infants and 55.1% mothers, respectively, had ever encountered any serotype, corresponding with the figures in fig 3). This needs to be clarified.

9. Page 13, last chapter: In discussing the calculated under-dispersion of the serotype distribution, the authors speculate on the role of cross-immunity between serotypes, and refer to studies on anti-capsular and anti-protein antibody response. It should be noted, in this context of natural carriage, that children less than 2 years of age mount only low or no antibody response to the
polysaccharide capsule. The conjugate vaccines, of course, mount antibody production, due to the adjuvant included in the vaccines. However, there is some supporting evidence for serotype independent natural immunity in humans (e.g. see Granat SM et al JID 2009).

Minor Essential Revisions

10. It is not clear at what point of the actual analysis multi-level models were used and what kind of models are actually referred to. Does ‘multi-level’ model refer to a random effect model to account for heterogeneity across individuals and/or serotypes (so that in these models there are two hierarchical levels)? If so, these models were not reported after all because of the reasons given on page 6, last chapter?

11. p 7., the 2nd sentence in the 2nd paragraph: It is not clear what the actual modification to the standard PAF you need to make is. Do you refer to the use of a common (estimate) of the OR? Please clarify.

12. p. 7, the formulae of PAF (in the middle of the page) and AF (the last line): Please check what sign of multiplication is used in the Journal. Now ‘.’ is being used which might go unnoticed.

13. p. 7, line 5 from the bottom: There is a mistake in the explanation of notation I_e. It should read ‘exposed infants’ (and not ‘unexposed infants’).

14. p. 8, line 3. The overall PAF is calculated as a weighted average of serotype-specific PAF’s, the weight being equal to serotype-specific prevalence in the infants. Is this an ad-hoc choice or is there another justification for this choice? Would not prevalence in the mothers be more appropriate?

15. p. 10, line 2. It is said that “the non-vaccine serotypes reached a peak at approximately 5 months’. However, there were no samples until at the age of 12 months, so this interpretation is too bold. Taken the data, you can say that non-vaccine type prevalence decreases from that at 5 months until 12 months of age.

Discretionary Revisions

16. Page 6, last sentence in the chapter laboratory analysis: the serotypes included in the 13-valent vaccine should be spelled out, as they are not obvious to all readers.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests