Reviewer’s report

Title: Dynamic models of pneumococcal carriage and the impact of the Heptavalent Pneumococcal Conjugate Vaccine on invasive pneumococcal disease.

Version: 1 Date: 2 November 2009

Reviewer: Kari Auranen

Major compulsory revisions

(1) It is not clear what are the steps were actually taken and in which order when estimating the model parameters from empirical data. In particular, how many parameters were estimated from data altogether? Were all parameters estimated at the same time or in sequence? How were the six best parameter combinations identified (Table 2). As a minor remark to the last of these questions, it is not clear how the deviance per se should be interpreted here (it is always the different of deviance values rather than any particular absolute value that should count in model choice).

I would suggest a better explanation of the steps in estimating the model parameters. This could be given in the Methods section or in the Appendix. Please check that the corresponding paragraph in the Results section corresponds to what is being said in the Methods.

(2) The notation is not always consistent. For example, the sub-index denoting the individual being vaccinated is sometimes “V” and sometimes “nu” (see e.g. the equations in the Appendix I). There are more comments on this below (see ‘Minor comments’).

(3) The use of the term “validation” is not clear. For example, it is not clear whether parameter estimation, as explained in the first paragraph of the Results section, should be called validation at all. I understand a set of crucial model parameters were estimated, based on the US pre- and post-vaccination data on IPD and the E&W data on carriage. But clearly this is not validation, rather “estimation” of “model calibration”.

(4) How sensitive are the results to the assumption that serotype 6A behaves like a vaccine type? Two references are given about the effect of PCV on 6A carriage (references [34] and [35], p.7, line 5). I’m not sure if the latter of these actually discusses carriage at all (but AOM).

(5) In the Discussion it is said that the cause of the difference between the model predictions and the recent data about the extent of replacement in IPD in the UK needs to be studied further. What do you mean by saying that this could be “due to vaccine introduction with or without a natural increase in NVT”? Do you refer to
calendar time trends as a possible explanation here?

(6) p. 14, line 1-2. It is well known that differences in the mixing pattern may have much more dramatic effects on model predictions than any statistical parameter uncertainty. Do you actually mean that the strength of the analysis was to estimate the one weighing parameter from empirical data (and not refer to statistical uncertainty)?

Minor revisions

Abstract

• What time frame does the decrease of 39,000 cases refer to? How do these figures relate to those given in the Results (page 11, the last paragraph)?

• p2, line -6, the last sentence is a little obscure. In what sense was ‘uncertainty reduced’ (e.g. with respect to what). Also, the claim about improved validity seems to be a comment which is more suitable to the discussion part of the article. This does not mean that I objected the idea to use empirical data to estimate model parameters!

• p3, line 4, there is something wrong in this sentence (“... and to then being able...”) 

Materials and methods

• Legend to Figure 1: notation ‘VT’ should be explained here.
• p.6, England & Wales IPD data: the total number of isolates should be mentioned

• Legend to Figure 2: notations ‘VT’ and ‘NVT’ should be explained in the legend. The same applies to other figure legends, i.e., some of the notations should perhaps be written out, despite they had been defined in the main text.

• p. 6, the last paragraph (‘Population’): What is the mortality rate, i.e. how does it depend on age (class)? What is the age structure of the population?

• Figure 4
  o Both arrows between compartments V_i and B_i point to the same direction towards B_i. The other of these should be reversed.
  o Subindex ‘i’ is missing in the clearance rates of the non-vaccine types, i.e., should it not read r_{N_i}? There is also inconsistency in this notation in the main text (check page 7)
• p. 7, line -11, ‘V’ and ‘N’ are obviously sub-indices to $\lambda$?
• p 7, line -10, there is misplaced comma between $c_N$ and the ‘is’
• p 7, line -6, the clearance rate r_i should be italicised. Similar mistakes occur also later (at least on p. 8, line -7).
• p.7, line-2: The percentage reduction in the rate of acquisition of carriage, due to the individual begin vaccinated, is called ‘degree of protection’ here. In the
legend to Figure 4, I understand the same parameter is called ‘vaccine efficacy against VT carriage acquisition’. Could either of the two used consistently?

• p. 8, line -10: Does it mean that the inverse of the rate of clearance (“the mean duration”) was 72 days in children 0-1 years of age etc?

• p. 8, line -8. I understand the work ‘infection’ here, but maybe it is clearer to use ‘carriage’ or ‘carriage episode’ also here.

• p. 8, line -4, index ‘j’ should be italicised

• p.8, the equation on the last line, should read $\lambda_{V_i}$ (i.e. subindex ‘i’ to the sub-index V ?

• I understand the parameterisation and estimation of the parameters are explained in the one section starting on page 8 (in Material and Methods) and then in the first section of the Results. While I like the approach to explain which data sets carried information about which parameters, I don’t find it easy to follow how different data sets were used in combination or successively.

  o For example, I gather that the force of infection was based on the E&W data and the mixing matrix on the US data, but I think this is not clearly stated in the main text. Were there parameters estimated simultaneously?

  o How was the E&W data used in the estimation of the case:carrier ratios? How about the US data? Was either of the two IPD data sets used for validation or both for estimation of the parameters?

  o Were the case:carrier ratios estimated simultaneously with the “beta” matrix and other carriage-related parameters?

  o In summary: How many and which parameters were estimated at a time, and what level of the others were assumed (see my respective comments to the

Results

• p 10, line -3, it should obviously read ‘Fig. 5’ (instead of Fig. 6)

• p. 10, line -2. What does it mean that the sensitivity analysis was conducted to “study uncertainty of different number of PCV/ dose effects?” How was this done, otherwise than changing the PCV/ coverage (which seems a different thing to me)?

  “Figure 6” should read “Figure 5”. In general, the enumeration of figures is shifted from this figure onwards. Please correct.

  The legend to this figure uses abbreviation ABC, which might not be obvious to all readers, despite it being defined in the main text.

  It is not clear how many and which parameters were estimated at a time, and what level of the others were assumed (see my respective comments to the
Methods section above)

• What were the actual case:carrier ratios? This would be an interesting piece of information, and also provide some information about how the model accounts for pneumococcal epidemiology.

• Related to the question above, were the model validated against the pre-vaccination IDP data from E&W or how was this piece of data used?

• p. 11, line 10, it should obviously read ‘Fig. 6’ (instead of Fig. 7)
  o Legend to Figure 6: Were there only four parameters ($\gamma, \epsilon, c_N$ and $c_V$) that define the joint likelihood for the number of IPD cases? How about the waning rate for protection (cf. Table 2), or other parameters?

• p. 11, line 11, confidence interval area? should be confidence area?

• p. 11, line 16: Do you mean that “…predictions were insensitive to the assumptions”?

• p. 11, the same paragraph: It is not clear why you say “thus the baseline parameters…”.

• p. 11, line -7, Figure 8 should be Figure 7

• p. 11, line -4, does the 63% refer to reduction in the accumulated number or the annual incidence after the “system” has settled to a new equilibrium?

• p. 11, bottom line, Figure 9 should be Figure 8
  o How is the approx. 20,000 cases of IPD

• p. 12, line 6, Figure 10 should be Figure 9

Discussion

• p. 13, line 1, It is said that the reduction in the annual IPD number if 3,700 in the long-term. How does this relate to the 2,300 prevented cases in the long-term (see p. 11, line -2)?

Appendix 1

• The subscript for the vaccinated susceptible is $\nu$. In Figure 4, $V$ is used. Please be consistent in this notation.

• The term $\pi_i(t)$ is clearly a rate, not just ‘function’

• The $\beta_{ij}$ is now explained, without reference to the force of infection. A reference to the main text would be appropriate here.

• Subscripts $\nu$ and $V$ are used interchangeably to denote vaccination. See also p. 8, the equation on the last line.

Appendix 2
• It seems that the competition parameters were estimated first. How was this done without knowledge of the beta parameters? Shouldn’t they be estimated simultaneously when applying the model to the US data?

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

I declare that I have no competing interests.