Reviewer's report

Title: Is HPV-16 seropositivity a correlate of immunity? Insights from compartmental models

Version: 2 Date: 28 September 2012

Reviewer: Mark Jit

Reviewer's report:

This manuscript describes the application of mathematical modelling to understand the relationship between seroconversion and natural immunity to HPV 16 infection. It is a novel use of modelling to an important area. In general, the analytical issues are handled well, although there are some areas which need further clarification:

Major compulsory revisions

1. p. 6: It’s not clear to me why excluding over 60s reduces the dimensionality of the model since the age-dependent sexual matrix is already data defined, and it hence would not require any additional variables to incorporate this age group. On the other hand, excluding this group removes an additional data point to which the model could be fitted to so would appear to increase rather than reduce overspecification.

2. p. 7: Is the Australian population structure between 12 and 60 also uniform (in addition to age-specific mortality)? If not, then the birth and death process described would not adequately capture it.

3. p. 7: If sexual activity commences at age 15, then why is there a need to model 12-14 year olds at all? Although to be honest I am sceptical that 15 really is the minimum age of sexual debut for the whole population; I suspect there is a small but still important proportion of children who commence sexual activity before 15.

4. p. 8: The age-independence of infection clearance is a problem because the detailed natural history of pre-cancerous neoplasias is not modelled. Older women are less likely to clear infections because their lesions are further progressed and hence less likely to clear. Models that capture lesion progression do not need age-dependent clearance rates but a simplified model like this would appear to require this age dependence to be built in explicitly.

5. p. 12: The posterior probability of seroconversion for females and males is 10-40% and 3-20% respectively; this is much lower than previous empirical estimates. For example, Dillner et al. Semin Cancer Biol 1999; 9:423 suggests >60% of females seroconvert, and Desai et al. Sex Transm Dis 2011; 38:5 suggests the figures for males might be about 1/3 of that.
6. p. 15: While the point about the need to consider SIS models may be valid, it should be noted that the "SIS models" in this paper are not true SIS models. In the second susceptible state individuals still have a degree of protection which they would not have in a classic SIS structure.

7. Table 3: I assume that the betas are the probabilities per partnership rather than per sex act? And if so, shouldn't they be dependent on risk group? (High risk individuals with many sexual partners may have a smaller probability of transmission per partnership since the majority of these partnerships will be casual.)

8. It would be helpful to show the posterior distributions of the fitted parameters (at least the marginals; the pairwise joint distributions would be even better). It would also be good to compare results to other model-based estimates of natural history parameters (eg. Van de Velde Am J Epidemiol 2007; 165:762, Bogaards et al. Am J Epidemiol 2010; 171:817, Jit et al. Med Decis Making 2010; 30:84, Insinga et al. BMC Inf Dis 2009; 9:119 to name just a few).

Minor essential revisions

1. p. 10: It would be good to show the preliminary analyses that indicate that having seroreversion in state R+ only produces an inferior fit.

2. The actual data (serology and DNA) do not seem to be described anywhere. It would be useful to know how it was collected, what the source population is, what the sample sizes are etc.

3. p. 29: In the model equations, shouldn't lambda be a function of [I-] and [I+] as well?

4. p. 9: "we associate seroreversion with state S+". This is confusing since it is not clear whether people serorevert FROM state S+ or TO state S+. Better would be "seroreversion moves individuals from state S+ to state S-".

5. Figure 4: The actual data points (serology and DNA) are not described in the legend; I assume they are the points with error bars around them.

Level of interest: An article of importance in its field

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