Reviewer's report

Title: Evaluation of using composite HPV genotyping assay results to monitor human papillomavirus infection burden through simulation

Version: 3
Date: 16 December 2014

Reviewer: David Regan

Reviewer's report:

This manuscript describes a well-conceived study to evaluate the accuracy of composite HPV genotyping assays. The two key conclusions are that these assays will generally overestimate the true burden of HPV infection (and could thus underestimate the population-level impact of vaccination) and that assay specificity is as or more important than sensitivity in this regard.

Major compulsory revisions:

1) More detail is required in the methods section on how the datasets are generated, how the simulations are run, and how correlation in applied. How the figure of 0.4 for correlation between types was arrived at also needs to be explained and, if necessary, referenced.

2) 500 data sets were generated for each scenario yet only point-values are given in the results tables. It should be specified what the point-values represent (i.e., mean, median) and a measure of spread (such as standard deviation or interquartile range) for each value should be provided.

3) The estimated-to-true ratios are missing for the 37/reduced scenario in Table 1b.

4) On lines 187-193, the importance of correlation is discussed but results are not shown. This is an important sensitivity analysis and the results should be given in the results section.

Minor essential revisions:

5) The manuscript would benefit from a thorough review of the language and grammar. While the manuscript is generally clear, I found the numerous grammatical errors a distraction.

6) It is not clear what the reductions of 8%/5%, 8%/4%, 10%/8% are in the tables. This is not described in the methods where only the 50% reduction is mentioned and is not referred to in the Results. These should be described, including what these numbers are based on, or omitted.

Discretionary revisions:

7) There are numerous references to “we” in the paper (e.g., “we perform” on line
79) but only a single author is listed on the manuscript. If there were others involved in the study they should be acknowledged, otherwise different language should be used.

8) I’m not sure I agree with the author’s contention that individuals with weaker immune systems are more likely to get infected” (lines 66-67 and 188-189). In my opinion, the probability of becoming infected is related primarily to the individual’s level of sexual activity and the prevalence of infection in the population. The rate of clearance, on the other hand, or the probability of infection persisting is likely related to immunity. Furthermore, as far as I can tell, the model used here for the simulations makes no distinction between individuals in terms of risk factors for infection and the degree of correlation between types is applied as an average measure.

9) It seems strange to take US (or Canadian) reported infection burden as the true prevalence when these estimates were obtained using the very assays that are being evaluated in this study. Perhaps it would have been better to use a true prevalence that produces the observed US or Canadian estimates when an assay with baseline characteristics of 0.95/0.95 for sensitivity/specificity is applied in the simulation.

10) On lines 206-221, the issue of sensitivity versus specificity is discussed. At a population level, it is infectiousness that is important for onward transmission and ultimately the burden of infection in the population. PCR, being a very sensitive methodology, is also likely to overestimate prevalence from the point of view of infectiousness. A very tiny amount of DNA detected is likely not infectious or may represent a past infection that is no longer viable or may simply be evidence of recent sexual activity. I think it is perhaps worth noting this point in your discussion.

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

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

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests