Reviewer's report

**Title:** Incident cervical infections with high- and low-risk human papillomavirus (HPV) infections among mothers in the prospective Finnish Family HPV Study

**Version:** 2  **Date:** 22 December 2010

**Reviewer:** Sheri Lippman

**Reviewer's report:**

The authors have responded to many of the critical issues that were raised in this revision. However, there are some concerns that remain and others that still need attention. These are discussed below.

**Compulsory Revisions**

- The sample for each analysis remains a bit confusing as the numbers don’t always match up between the text, the figures, and the tables. The following are concerns about inconsistencies:
  - The authors added 2 figures to better describe the sample. However, the text says there were 329 mothers recruited and the new figure starts with 325.
  - On page 7 the authors state that to calculate the time to infections using the actuarial method, all baseline HPV negative women (n=252) were included – this includes 203 baseline negative women with incident infection and women who never developed HPV (which in the second figure is n= 52). That would make 255 women – but there were only 252 in the sample for this analysis. Then in the first table the footnotes imply that the actuarial method included only 203 women whereas the crude rates included all baseline negative women (it is written the other way around in the text). More care needs to be paid to ensuring the various pieces of the manuscript match up.

- The authors’ response to the comments about presenting covariates in Table 2 is disconcerting. They state that the program used (Stata) does not allow them to have estimates of coefficients for each category of a categorical variable. This is untrue. The interaction expansion (xi and i.varname) creates dummies for categorical variables which can be used with essentially all regression commands in stata, including “xtgee” with “family(poisson)”, which appears to be the regression command used. If it is the case, that the “xi / i” command was not used, then it may imply that the categorical variables are being entered into the model as if they are continuous. As it stands, the categories of categorical variables are still not presented; currently only 1 category and mention of a reference group is included so it’s impossible for the reader to know what comparison is being made or how many potential categories there are in each variable.

**Minor Essential Revisions**
I would recommend revising the comparisons presented at the bottom of table 1. The authors need to make results easy to follow and lead the reader through the result that the authors believe are most important. Do the authors think the rate ratio comparisons by species are most interesting presented by actuarial or crude estimates? It is misleading to present both, not to mention that it’s inconsistent to present the RR of sp 9 compared to 7 using actuarial rates (5.04) and then turn it around to present the RR of sp 7 compared to 9 using crude rates (1.54). These numbers become more difficult to interpret for the reader as they are currently presented. Another option is to remove the RR from the paper completely, since the authors said in their response to the reviews that the interpretation of the RRs is unclear and that they spend little time discussing it. The reader can easily compare IRs between species and types by just looking at the numbers in table 1. This could also shorten the paper a little.

Discretionary Revisions

• The second sentence of the second complete paragraph on page 13 (in discussion) was confusing. “the absolute IRs was much lower” than what?
• Under “Serology” – the second sentence says: Samples were analyzed and interpreted using the technique and cut-off values as recently described.” I would add what you were testing for and at least include the diagnostic test.
• In the methods section, I would modify the sentence: “Part of the women was lost from the study mainly due to difficulties to attend or family reasons.” Assuming “family reasons” means child care and “attend” refers to study appointments, I would rewrite this as: “Some study participants were lost to follow-up, mainly due to difficulties in getting to study appointments or lack of child care”.

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

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests.