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: 5 December 2010

Reviewer: Ly-Mee Yu

Reviewer's report:

This manuscript reports the frequency of type specific incident HPV-infection in a cohort recruited in the Finnish Family HPV Study, and the risk factors associated to the incidence. The authors have revised the manuscript based on comments from previous reviewers but I would recommend further clarification to various points I raised below before accepting it for publication. The authors should fully appreciate the advantage of online publication, which word count restriction is no longer an obstacle. Although, many issues have been addressed in other publications related to this study cohort, it would benefit readers from countries, who may not have access to those journals, to understand this manuscript fully if more clarification were provided.

Comments:

1. Page 4 – the method section stated that 329 women were recruited with 161 had an extended 6 years follow up. The authors need to explain how these 161 women were obtained, i.e. whether they remained to be followed up for 6 years or whether an invitation for a 6 years follow up was sent out and 161 responded?
2. The numbers stated in the text do not seem to correspond to the number reported in figures. For example, Page 4 stated that 329 were recruited but the flow chart reported 325. Similarly, 161 had extended follow up but 171 was reported in Figure 1. The authors need to make sure all numbers are consistently reported throughout the manuscript.
3. Page 4 – again, structured questionnaires were sent out at baseline, 36 months, and 6 years. According to the flow-chart, they were sent out at visit 2 (2 months after baseline), and at 6 years only.
4. Page 5 – Please clarify which visits the pap smears were done during the follow-up. This again does not correspond with Figure 1.
5. Page 6 – Please clarify when the serology were done during the follow-up.
6. Please describe what information was collected from the questionnaire.
7. Description of the viral events at each visit requires further clarification. Figure 2 gave an impression that all events are mutually exclusive but they did not seem to be the case when reading the text. For example, the explanation in the text gave an impression that fluctuation is a subset of clearance, and genotype specific persistence is a subset of incident HPV infection.
8. Where did the number of all baseline HPV-negative women (n=252) included
9. If survival time was calculated from baseline to visit from first incident event then why did the authors censored the first incident event? Also, since the time that an event occurred was based on the fixed time of the visit, the survival analysis should take into account of interval censoring and the authors should state the approach to handle this.

10. The authors need to be careful with the terminology used. Actuarial time is usually referred to survival time derived from Life-table analysis, which is different from the Kaplan-Meier analysis per se.

11. Can the authors explain why only 203 women were subsequently included in the calculation of the crude incidence time?

12. Table 1 – What was the rationale to present the mean time to first incident infection when its distribution is highly skewed? The number seems to be very different between the actuarial and the crude given that there were only about 50 women excluded in the crude rate calculation. Does the variation in the time to first incident infection between HPV genotypes relate to the lost to follow up during visits? Perhaps the authors can clarify this further.

13. Please make sure the footnote indicators in Table 1 match with their correct footnote description.

14. Results section and Figure 3 - why did the authors compared species 7, 9 and 10 only?

15. The authors should state how many subjects were included in the Poisson model and how were the incidences of other HPV species treated in the Poisson model? In one of the responses to a reviewer’s comments earlier, the authors did specify that 203 women who had an incident infection with a type from species 7 and, or, 9 were included in the analysis. But reading the results in Table 1, only 133 cases of species 7 and 9 were identified. The authors need to clarify this in more detail.

16. Although, the authors did explain, in the response to reviewers’ comments earlier, the difficulty of obtaining outputs in the form that contains all the different categories using STATA, but the problem should be easily overcome by fitting dummy variables to the model.

17. Selection of covariates included in the model should be described. The authors did mention in the response that over 60 baseline covariates with other additional follow-up covariates were tested. Why then only 23 were reported in Table 2 and did the authors consider the issue of Type I error regarding the amount of covariates being considered.

18. How the covariates collected during follow-up were specified in the model? Were they treated as time-dependent covariates?

**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:
None