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: 23 December 2010

Reviewer: Seonaidh Cotton

Reviewer's report:

The manuscript is greatly improved. There remain, however, a few issues with the paper which require to be addressed before this paper can be published.

Major essential revisions

1. Presentation of results relating to the predictors of infection

Although the reference groups are now stated for most variables presented in table 2, it is still difficult to interpret these results relating to predictors of HPV infection. There are at least two ways of presenting this table in a way that is more interpretable for the reader. Either one could more clearly describe the groups; for example...

Seroconversion to HRHPV during follow-up (yes, compared to reference group no)
Seropositive to HRHPV at baseline (no, compared to reference group yes)
Age at onset of sexual activity (>=13 years, compared to reference group <13 years)

Or one could present the reference groups and comparisons on separate lines within the table:
No seroconversion to HRHPV during follow-up (reference)
Seroconversion to HR HPV during follow-up 0.81 (0.67-0.98)
Not seropositive to HRHPV at baseline (reference)
Seropositive to HR HVP at baseline 1.10 (0.85-1.42)
Onset of sexual activity <13 years (reference)
Onset of sexual activity >=13 years 1.15 (1.05-1.27)

I appreciate that these would make the table larger, but the gain in clarity would outweigh this.

It is still not clear when some of the variables presented in table 2 were measured. For example, do the variables employment status, lifetime number of sexual partners, weekly intercourse etc relate to the assessment made at baseline or at one of the follow-points? Without this information it is hard to
interpret the findings.

Other, more minor clarifications required to the table include: (i) should the reference group for life-time number of sexual partners be 1-2 rather than 0-2 – all should have at least 1 partner; (ii) in terms of the practice of oral and anal sex – is this current or ever, and are the reference categories “no” and “never” equivalent?; (iii) some women had not reached the age of 20 at baseline – how is this handled within the variable number of sexual partners until age 20 years; (iv) for smoking habits, are ex/current smokers combined?

The abstract states that later initiation of oral contraception is an independent protective factor, but the IRR suggests that it is a risk factor – this should be clarified. Similarly, in the first paragraph of the section on predictors of species-specific incident infections, this is described as protective. In this same paragraph, some of the variables listed appear to be protective, and others are risk factors. It is very difficult to follow this section and I would suggest revising the sentence to state “seven variables were significantly associated with incident infection.”, and then explicitly stating which factors are protective, and which are risk factors. In the second paragraph in this section, the late initiation of OC is described as protective – but the IR appears to indicate it is a risk factor. This requires clarification.

2. Clarity in relation to the number of participants

It is not clear how many women were included in this analysis because of inconsistencies in the manuscript. In the methods section, the number in the cohort is given as 329; but in figure 1, the number is given as 325. In the first paragraph of the section on subjects in the methods section, the authors describe extended follow-up of 161 women; while figure 1 suggests that this is 171 women. At the outset of the results section it is stated that there were 252 HPV-negative women enrolled in the cohort, but it is not clear how this relates to the numbers given in figure 2. It is my understanding that groups 1 and 2 were HPV negative at baseline – a total of 255. The number of participants in the cohort, and in the current analysis should be clarified, and the authors should ensure that this is consistent throughout the manuscript.

In one of their responses the authors indicate that women were required to have completed two follow-up visits to be included – this is not explicit in the text, and should be added. This may, in part, help to explain inconsistencies in the numbers of women included in the cohort/current analysis.

3. Outcome groups

The additional text describing the outcome groups is helpful. However, although the text has been revised, the definition of the groups are still slightly unclear. As noted above, it is my understanding that groups 1 and 2 were HPV negative at baseline, and groups 3-6 were HPV positive at baseline (though if this is not the case, then it is not clear why groups 3-6 are not subsumed by group 2). It may be helpful to explicitly state the HPV status at baseline in the text. It may also be helpful to include an additional row of boxes in figure 2 (between the existing first
and second rows of boxes) showing the baseline HPV result.

In their response to my initial comment on checking the figures in table 1, the authors note that the two cases of HPV11 infection are both incident events. It was my understanding that this paper dealt only with incident invents, but this comment suggests otherwise. Clarification is required.

4. Data collection

Figure 1 is helpful in describing the data collected at each of the time-points. However, the information given in the figure does not appear to match that given in the text. Examples of this include (i) the text indicates that a questionnaire was completed at baseline but the figure suggests this was done at the second visit, and (ii) the figure suggests that blood sample was taken at visit 2, this is not included in the text, but there are other instances of this throughout the text. Care should be taken to ensure that the information given on the figure and in the text match and reflect what was done (and what is relevant for this paper).

5. Analysis

The authors state that the calculation of crude incidence times only those 203 women with incident events were included, which suggests that these 203 women were included in all calculations of crude incidence times. If this is the case, then no amendment is required to this section; however if only those with, for example, incident HPV16 infection were included in the calculations relating to HPV16 then this should be stated.

Minor essential revisions

1. In their response to my previous comment on clustering by mother ID and run; the authors do not appear to have described what the “run” variable is. If this is related to follow-up time-point, then it would be helpful to use this term in the manuscript.

2. In the last paragraph of the methods section, the authors note that selected variables from the follow-up questionnaire were used – they should clarify whether this is the follow-up questionnaire at 3 year or 6 year follow-up, or both.

3. In the fourth paragraph of the results, the authors state that “Actuarial mean time for species 7, 9 and 10 was identical” – they should use the term similar rather than identical.

4. In the third paragraph of the discussion, (i) it is not clear which individual studies the authors are referring to – if this is a general comment rather than referring to specific individual studies, then it may be clearer to say “...comparison between individual studies difficult”; (ii) in the statement “certainly the IRs in the latter are higher”, does this refer to crude, actuarial or both types of IR?

5. In the fourth paragraph of the discussion, reference is made to IRs from other studies – are these crude or actuarial – this should be added to the manuscript, and if they are not crude, then comment should be made as to whether it is
appropriate to compare different types of IR. In this same paragraph, it would be helpful to clarify that the shorter follow-up times are in the other studies rather than in the current study.

6. In the eighth paragraph of the discussion (staring “women who started their sexual activity..”) I suggest revising the last sentence to say “...have already experienced infection, and either cleared or developed a persistent infection by the time they reach their 20s (the mean age of the current cohort).

7. In the ninth paragraph, lifetime sexual partners are likely to be a proxy for HPV exposure rather than a measure of HPV exposure.

8. In the eleventh paragraph (third sentence), the authors should include the reference to where they have recently shown in relation to women committed to the second child.

9. In the paragraph on limitations, the authors may wish to comment on whether they think the findings from the current study are likely to be generalisable to non-pregnant women.

Discretionary revisions
1. Methods/subjects, para 2; should the first sentence say “followed-up for up to 6 years”. When describing the age of subjects, it might be helpful to include the standard deviation in addition to the mean. Rather than saying “Part of the women was lost...”, it would be better to say “Some of the women were lost...”.

2. In the methods section (second paragraph) the structured questionnaires are described – I suggest that the authors add the text from one of their responses (information on demographic factors, sexual behaviour, gynaecological and obstetric history and other potential risk factors for HPV infection) to give the reader a clearer idea of what information was collected.

3. The last paragraph on the HPV genotyping section is not clear, and it might be clearer to state “As HPV16 is the most prevalent genotype we retested all HPV16 positive samples as described by Schmidt and co-workers (21) and discounted the possibility of false positive samples due to contamination.” (if this is what was done).

4. The authors are not consistent in whether they use the word “months” or abbreviate to “mo” in the manuscript.

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.