Reviewer's report

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

Version: 1 Date: 5 August 2010

Reviewer: Seonaidh Cotton

Reviewer's report:

This is an interesting study, but I recommend revision of the manuscript prior to it being accepted for publication. In particular, aspects of the methods section require clarification in terms of what has been done. The discussion covers some interesting issues, but the authors do not appear to describe any potential limitations of the study.

Major compulsory revisions

In the background section it would be helpful if the authors provided some information to help justify their choice of sample. Similarly, in the discussion section some consideration of the generalisability of their findings to the wider population would be helpful.

Throughout the manuscript there is some lack of clarity in relation to the numbers of women included in the study. A flow diagram might be helpful showing that 329 were recruited; 252 were negative at baseline (this is the group under consideration in the study and which constitutes the sample), of whom 203 tested HPV positive during follow-up. The statement “part of the women were lost from the study mainly due to difficulties to attend or family reasons” should be clarified by stating how many women were lost to follow-up – again this could be included on a flow-diagram. In addition, clarification is required as to whether these women were included up to the point they were lost to follow-up or excluded. In the section on outcomes of HPV infection (page 7), if the groups are to be defined, the outcome patterns should be more clearly defined. For example does incident HPV mean that they had one positive test, or one or more positive test; do women who have type-specific persistence have a positive test at every time-point, including at baseline, or can women with incident HPV join the persistence group if they are found to have persistent infection. The groups included in this analysis (always negative, incident HPV) should also be highlighted. Again, if all groups are to be included in the description, it might also be helpful to present the numbers in each of these groups. It is also not clear whether women described as having multiple-type infections are infected with multiple types at the time of the incident infection, or have an initial single infection with a second infection at a subsequent time-point.

Some clarification of the testing undertaken in the study is required. It is not clear from the manuscript the purpose of testing samples to determine whether they
were high-risk HPV positive or negative – these results do not appear to be presented. Were all samples subject to HPV genotyping? The authors describe steps to identify potential contamination with HPV16 testing, but do not describe whether there was evidence of this. Serology is mentioned, but the text should describe what the samples were tested for and how these results were used in the analysis.

The description of calculation of actuarial and crude times and IRs should be revised for greater clarity. The discussion section does include some consideration of actuarial and crude results but could expand this point, for example giving more information as to when each of the methods would be appropriate to use. It would perhaps be helpful to introduce these concepts in more detail in the introduction.

When describing the analysis of predictors of incident HPV infection, clarification is required as to who is included in this analysis (for example, does this analysis include the 252 women who were HPV negative at baseline?). In this same section, there needs to be much greater clarity in when covariates were measured, together with justification for the choice of covariates measured and the timing of the measurement. One example of this is the variable “new partner” it is not clear whether this is a new partner before the incident infection, or a new partner at a specific follow-up point. Similarly, in the discussion, a comment is made about failure to seroconvert, but it is not clear when this was measured. There is also a comment that analysis was clustered by mother ID and run – and it is not clear what these variables relate to.

It would be helpful at the start of the results section to provide some baseline characteristics of the sample – for example age (the authors comment in the discussion that it is essential to know the age-profile of the cohort, but fail to disclose this) and any other factors which would help to describe the sample. When reporting the results, care should be taken to ensure that all results presented are defined as crude or actuarial (this is a recurring problem in the discussion). In addition, in the second and third paragraphs of the results it is not clear whether the some of the actuarial and crude times are mean times.

The mean follow-up time for women who remain HPV negative should also be reported (either separately from the HPV positive women, or combined).

It is not clear why the calculation of incident infections has the denominator of those with any infection rather than all those negative at baseline - a more precise description in the methods may help. The figures presented in table 1 should be checked, for example: (i) can the crude/actuarial mean time be the same – eg for HPV11; (ii) can the 95% CI round a figure be equivalent to the figure – eg for HPV33? Again, greater clarity in the methods section may address these issues.

The presentation of RRs in table 1 demonstrates that different conclusions would be reached on the basis of using actuarial or crude incident rates but the usefulness of each approach is unclear.
The reference groups need to be more clearly defined in table 2 (and linked to the comment made earlier, the timepoint at which these were measured made explicit). It would also be helpful to include numbers and rates of infection with species 7 and 9 infection for each of the covariates (reference and comparison group separately) in table 2.

Discretionary revisions
The section described as background in the abstract is a statement of the aim rather than background information.

Paragraph 1, second sentence – this reads as if there are only 30 HPV types and the authors should consider revising to imply that there are additional types.

It is not clear why only baseline oral HPV swabs were used in the present study, and it would be helpful to clarify this.

On the second line of page 13, the authors note that no significant differences were observed between HPV16, 18 and multiple type infections. Is this in terms of time to infection - this should be clarified. It would be interesting for them to comment as to how many of the multiple type infections overlap with the HPV16 and HPV18 infections.

The authors comment that IR is critically age-dependent (page 13) – should this point be clarified to say that the IR for HPV infection is critically age-dependent?

The authors conclude that this type of data is needed to understand the natural history of HPV infections and for designing tools for their prevention. They might expand on these issues in the discussion section.

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

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

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.