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: 9 August 2010

Reviewer: Sheri Lippman

Reviewer's report:

This paper describes type specific time to HPV infection, infection rates, and correlates among previously HPV negative pregnant women enrolled in the Finish Family HPV Study. The paper contributes some informative data on incidence rates in this population, however there is a good deal of information that is not communicated in the paper that would help the reader understand the data and a number of analyses and findings require clarification.

Major Compulsory Revisions:

1. While there are previous publications that describe this cohort, more background information would be useful to the reader. For example, in the first paragraph of methods, there is no description of the basic objective of the parent study (the Finish Family HPV Study). A brief description of the research cohort is needed, as is the years in which this study was carried out and the number of people in the total study. Is the full study comprised of the 329 mothers or is this a subset of the parent study? Additionally, there is no table presented with a socio-demographic description of the cohort. The authors make statements about this cohort being younger or older than other cohorts, but it would help the reader to see a description of this cohort as a part of this paper- particularly a description of age.

2. How was the time of an incident infection calculated if follow-up visits occurred at 2, 12, 24, 36, and 6 years? If someone was positive in year 6 but negative at 36 months, did the authors assume infection occurred at the mid-point? There is no description of how time to event was determined given long periods between testing.

3. The tables should contain all necessary information to stand alone, including the sample size, a description of who is included, the name and location of the cohort, and relevant statistical tests that are used. This information should be included in the heading.

4. Covariates in Table 2 and in the text are not well described. Categorical variables should indicate all categories, not only the referent group, in order to understand what comparison is being made. When statements are made in the text about significant predictors (i.e. 1-2 lifetime sexual partners was a predictor of infection) the reader needs to know who is being compared. Is the group with 1-2 partners being compared with three or more partners? Were people with 3
partners and 20 partners grouped together? In the tables the same is true. For example, “baseline pap smear status (ASCUS ref):” are the authors comparing infection rates among all women with pap smear results (negative, LSIL, ect…) to an ASCUS category at baseline? That seems a strange choice. If there are multiple comparisons being made, the categories must all be listed. As it stands it seems that many variables are being dichotomized. These classifications and choices should be described.

5. It is generally difficult to follow the analysis. I believe the rate ratios and time to event estimates are relevant to a single incident event per HPV infection per person – in other words, a first positive diagnosis. The analysis of predictors uses a poisson regression which implies a count variable. However, the description of the model (second paragraph under the heading “Predictors of type-specific incident HPV infection”) says the outcome variables was incident HPV infection – which is a binary variable and incidence occurs once – unless the authors are including repeat positive outcomes. In other words, is this a description of some incident infections and some persistent infections mixed together? I would think to look at risk factors or correlates of infection, one would use the first recorded positive outcome. OR if someone can be classified multiple times (pos or neg at 2, 12, 36 months, etc…) this may be better modeled with a regression model for binary data or with hazards models. (The authors do note that the results are identical to a PA GEE model – but the above confusion remains).

6. The utility of the crude IR is unclear. The authors argue that the crude IR provides a measure to compare different HPV genotypes in their speed to develop the first incident event. However, without knowledge of when a person is actually exposed to HPV, one cannot say that the virus took x months to develop or that one genotype develops more rapidly than others: such a statement makes the assumption that all people were exposed at the same time, which is certainly not the case.

Minor Essential Revisions:

1. The conclusions can expand on how this data adds to the field and how it can contribute to furthering our understanding of HPV and potentially to prevention.

2. The samples collected were cervical samples; this should be reflected in the title, which currently describes infections studied as “genital infections.”

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.