Reviewer's report

Title: A Prospective Study of Absolute Risk and Determinants of Human Papillomavirus Incidence among Young Women in Costa Rica

Version: 3 Date: 8 January 2013

Reviewer: Marjan Javanbakht

Reviewer's report:

This is an interesting paper which aims to describe the incidence of HPV (overall and type-specific) among young women in Costa Rica. The paper has multiple strengths. It is generally well-written with clearly defined topics. It is also one of a few studies to evaluate the incidence of HPV among a relatively large group of women. Additional strengths include the relatively long follow-up period and HPV-typing.

Minor essential revisions

1. Abstract, Methods section, Line 41 – Instead of “enrollment questionnaire data” it may be more useful to state factors considered from enrollment, including demographics and sexual risk behavior information.

2. Abstract, Results section – are the hazards ratios presented in this section based on multivariable analysis? If so, it would be useful to either state this or indicate that the data presented are adjusted hazard ratios.

3. Methods section, description of study participants – It would be useful to describe the inclusion/exclusion criteria. In the statistical analysis section (lines 117-118) it is implied that women who were not sexually active at baseline, were not included in the analysis. If this is correct, how many of the 3,739 were excluded from this study? Also, is it known whether women were sexually active at follow-up? It seems having women who were not sexually active at follow-up would bias incidence estimates given that it is unclear if women were ‘at-risk.’

4. Methods section, description of study participants, line 133 – Partner questions were only asked of women who had one sexual partner at the time of interview or ever? Does this mean women who may have had a main partner, but other concurrent partners were excluded?

5. Results section, multivariate analysis, lines 190-194 – It would seem that being positive for HPV at enrollment would be highly correlated with the other factors included in the multivariate model and consequently may result in collinearity issues. Was this assessed?

6. Results section, line 203 – did this model adjust for any other factors?

7. Discussion, line 244-246 – It seems that this section should be removed as the data in this analysis does not support the statement. While there was an association with condom use and HR-HPV in the univariate analysis, this relationship was confounded and wasn’t seen in the multivariable analyses.
8. Table 1 – column for % HPV negative can be eliminated as this can be easily computed from data in the first column.

9. Table 4 – reference categories are missing, please also specify what is being controlled/adjusted for in the model

10. Tables 2 and 3 – please consider adding the adjusted hazard ratios (Table 4) to these tables; it would be helpful to be able to see the unadjusted and adjusted HRs in the same table.

Minor typographical/editing issues

11. Results, line 174 – “frequency sexual activity” should be changed to “frequency of sexual activity”

12. Results, line 201 – the word “common” can be eliminated from this sentence

13. Results, line 234 – this should read “have also been” and not “have been also”

Discretionary revisions

1. Background section – It may be useful to provide some information on other cohort studies assessing the incidence of HPV and the unique contribution of this study.

2. Results section, data on incidence – rather than or in addition to presenting data on cumulative incidence, it may be useful to present another measure of risk such as product-limit/Kaplan-Meier method. The survival/risk curve would help to give a distribution of time to occurrence of events and would be useful given the study population (women who are so close to the time of their sexual debut).

3. Discussion, section on limitations – given that face-to-face interviews were used to administer the enrollment questionnaire it may be useful to include a discussion of this in the limitations. During an interview, participants may be reluctant to disclose information regarding sensitive or socially stigmatized behaviors resulting in response bias and a potential underestimation of these behaviors.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests