Reviewer's report

Title: Human Papillomavirus Serology and Tobacco smoking in a Community Control Group

Version: 2 Date: 4 August 2014

Reviewer: Matthew B Schabath

Reviewer's report:

This is an interesting article from a well established group of researchers. The paper is nicely written and somewhat novel. However, there are some rather pressing concerns regarding their analytical approach that must be addressed. My comments are below.

Major Compulsory Revisions

1. More detail about the exposure assessment data is required. Specifically:

1a. How was smoking status defined, i.e., what was thresholds to determine a never smoker, former smoker, and current smoker.

1b. Was family history of cancer restricted to first degree relatives or any blood relatives?

1c. How was drinking habits captured? The authors should explicitly state this relates to alcohol consumption. Was it any type of alcohol consumption or did the intake differentiate between beer, liquor/spirits, and wine? Moreover, what was the time-frame for alcohol consumption, i.e., over the last year? The year prior to the interview? Etc.?

2. Why wasn't age of first intercourse included in the multivariable models? Considering there was a significant difference based on HPV status, this may suggest a possible confounder. Moreover, why did the authors only include age, race, gender, education, and average drinks per week in the multivariable models? Certainly education and income could be risk factors and/or confounders. The authors need to better explain their model building approach.

3. Table 1. Why isn't there a p-value comparing the mean age for the two groups?

4. What were the cut-points based on in Table 1 for income, age of first intercourse, and number of sexual partners? This should be included in the text and/or a footnote in Table 1.

5. Tables 2 and 3 should be combined; no need to waste space by creating separate tables.

6. Why did the authors only examine the possibility that age was acting as an effect modifier? What about other factors such as sex, SES, sexual behavior,
etc?

Minor Essential Revisions
1. The authors should include as a footnote in Table 1 or in the text how pack-years was calculated.

2. The authors should include smoking status (never, former, current) and drinking status (ever vs. never) in Table 1.

3. A very minor point, but the p-value for number of sexual partners ($p = 0.86$) should be in the same row as “16+” to be consistent with the formatting of the other p-values.

4. Another minor point: the authors interchange sex and gender. It would be more appropriate to use sex since they were referring to biological characteristics that define men and women.

Discretionary Revisions
None.

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:
I declare that I have no competing interests