Reviewer's report

Title: Human papillomavirus infection and cervical dysplasia in female sex workers in Northeast China: an observational study

Version: 2  Date: 16 April 2015

Reviewer: Francesca Lantieri

Reviewer's report:

This study report on the prevalence and risk factors of HPV infection in Chinese female sex workers (FSWs), finding that age of first sexual intercourse and menopause are two independent risk factors.

My major comments are about the necessity of more details on the recruitment and on some of the risk factors measured by the questionnaire.

In particular:

1) What is the period of recruitment of FSWs sample and controls? Also, the authors report that 1000 women who underwent regular gynecologic examinations were considered as a representative sample of the general population, however I suggest to provide more details, such as, in addition to the period of recruitment, if the recruitment was consecutive, and if any criteria of exclusion or inclusion were applied. Please, although the recruitment has been carried out in Northwest China, specify if sample and controls were of Northwest China ethnicity.

2) I was surprised that age at first sexual intercourse was a significant risk factor, while the time as sex worker was not. How was the time as sex worker measured? There was any measure about the rate of working activity? Some details about the questionnaire on this might be helpful

3) What do the author mean with menopause: age at menopause or if the women are already in menopause?

4) Since the HPV prevalence in FSWs in this study is especially higher among women aged less than 20 or more than 51, why age was not added in the multivariate analysis?

Other major comments are about the necessity of more details in the statistical analysis section and the revision of the tables. In detail:

5) “Putative risk factors… were analyzed by univariate and multivariate analyses” (line 145), please specify which test was used (logistic regression?)

6) Table 1: which test was used to calculate the statistical significance? Pearson’s chisquare? Comparing what? I assume that for HPV infection the four cells were 210, 191, 790 and 118, but I wonder how it was calculated for ASCUS
and CIN. Please, explain better.

7) Table 2: please, add N and HPV prevalence for controls, and discuss these data with respect to the literature about Chinese population (if any). Add also 95% confidence intervals in both tables 1 and 2.

8) Table 3: the title reports “linear regression”. I guess it’s a typo and the correct analysis is logistic regression, otherwise explain how the analysis was performed. In addition, the p-value reported on the first line for the risk factors with more than two categories (i.e. 0.36 in the line that refers to the marital status single) does not have any sense, since that line refer to the reference category by definition (OR=1). I guess the authors meant the significance in general; in this case please, move the p-value to the heading line (i.e. marital status). Besides, when the overall test is not significance, it is of small use to report the significances for the distinct categories (this last comment is discretionary).

9) Why the result for the univariate analysis on Menopause is not reported in Table 3 and Use of vaginal medications and Regular Ob/Gyn exams (and only those) are not reported in the multivariate analysis (Table 4)?

Minor comment:

10) In the statistical analysis section, “data on the prevalence rates of HPV and CIN were analyzed by Pearson’s chi-square test” (line 144): I guess the authors mean “data on the prevalence rates of HPV and CIN in FSWs in comparison to controls and also considering age stratification were analyzed by Pearson’s chi-square test”. Please, specify.

11) Refs 9 and 10 do not seem to be about FSW (lines 97-98)

12) “HPA infection” in Table 1: is it a typo?

Discretionary comment:

13) is it available any data about age at first sex intercourse and menopause among controls? It would be interesting to know if the results obtained on the FSWs sample is valid also for the general population.

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

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