Reviewer's report

Title: Pre-Invasive And Invasive Disease In Women With A Cytological Diagnosis Of High-Grade Lesion And High Grade Lesion Can Not Exclude Microinvasion

Version: 2 Date: 7 October 2014

Reviewer: Rebecca Landy

Reviewer's report:

The question posed is an interesting one, and does not appear to have been the focus of prior work, but there are some serious issues with the paper that would need to be addressed prior to publication. It would also benefit from some improvements to the English.

Major Compulsory Revisions

1. The last sentence of the 'methods' section of the abstract is not mentioned in the article, though there is a relevant section of table 1 that is not discussed at all in the results section of the article. I'm not sure that this last sentence reflects the results in the table, but it may just be that the wording is unclear – the abstract says the 'presence of lesions', but the table seems to compare HSIL/HSIL-micro with colposcopic findings and possibility of vision of squamo-columnar junction.

2. When there is an expected value of less than 5 in a cell, a Fisher's exact test can be carried out instead of a chi-squared test, rather than not doing a test.

3. More information is needed in the Background section – for example, is there cervical screening in Rio de Janeiro? That cervical cancer is the 3rd most common cancer in Brazil (as stated in the abstract). What are the current recommendations for HSIL and HSIL-micro/what are the pathways to diagnosis? (This is especially important, as I could only find them in Portuguese online.)

4. The methods section does not contain a statistical methods section. Some of the necessary information is given in the last paragraph f the results section – this should be moved to the methods section. However more information is needed, for example how the PR was calculated and should be interpreted.

5. It would be useful to know what proportion of HSIL-micro are seen at the named collaborating secondary units – 68 cases over 6.5 years seems very low.

6. Results section – Tables 1 and 2 need explaining in the text, rather than just saying the tables exist. This is especially true for results that are picked out in the abstract, but are currently not mentioned anywhere in the text.

7. The results for Tables 3 and 4 are currently given in the Discussion section (paragraph 6 of the discussion), but are not given in the results section – this should be moved to the results section.

8. The age categories for Tables 3 and 4 are not sensible, as they do not split up the HSIL-micro women (42/47 are in the 35+ category). As a broader point, I am
concerned about the selection of controls, and the fact they have such a different age distribution. In my opinion it would be preferable to match on age (and possibly year of cytology, if you are concerned that the quality of cytology has changed during your study) rather than selecting the 2 previous and 2 following women with HSIL.

9. You must say that the sample is a symptomatic sample in the methods section, instead of only mentioning it in the discussion.

10. The discussion would benefit from following STROBE guidelines. In particular, there is no ‘strengths and limitations’ section of the discussion.

11. Paragraph 1 of the discussion should include the search terms used if they are claiming there’s no prior research on the topic.

12. In Table 1, it is not clear to me how women pap smear findings of normal were included in the study – I assume that these are not the smears that got the woman into the study, but I don’t know when they are from. Similarly, for ‘procedure made’, is this the first, or the most severe, or most recent, or diagnostic?

13. In tables 3 and 4, is age adjusted for as a categorical or continuous variable?

Minor Essential Revisions

1. In the abstract, methods section, the dates are given as June 2006-June 2012, whereas in the paper and figure it’s June 2006 – December 2012.

2. The word ‘once’ is regularly used instead of the word ‘since’.

3. I would remove ‘mass screening’ from your keywords, given that your sample is symptomatic women.

4. In paragraph 5 of the discussion, you have said that 100% of women with HSIL-micro had preinvasive or invasive disease, but according to your figure 1, there were 2 women (out of 47) with HPV/CIN1 – your earlier definition of preinvasive is CIN2+ (for Table 3).

5. Table 1 – clarify what ‘gesta’ and ‘para’ refer to.

6. Table 4 – for ages 25-34, the total for HSIL-micro should be 4, not 7.

7. The manuscript needs to be edited by an English speaker before publication

Discretionary Revisions

1. In paragraph 3 of the results section, I would give the total number as well as the observed number alongside the percentages – for example, in the second line of the third paragraph of results, I would have 63.8% (30/47).

2. In the discussion, paragraph 4, I would mention that the sample in the Massad paper was based on a screening sample, to make clear the comparison to a symptomatic sample.

3. In paragraph 4 of the discussion, I would change the order of the points regarding difference in age, and the symptomatic sample, as the symptomatic
sample is likely to be more important in explaining the difference in results.

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

**Declaration of competing interests:**

I declare that I have no competing interests