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: 4  Date: 2 February 2015

Reviewer: Rebecca Landy

Reviewer's report:

Major Compulsory Revisions
n/a

Minor Essential Revisions

1. If you are going to select the cut-offs for age from a ROC-curve, it is necessary to make sure that they are appropriate for the desired purpose, which I don’t think they are in this situation. The selected cut-offs split the women with HSIL-micro into groups of size 1, 4 and 42 – it is difficult to place much trust in results based on 1 and 4 women in the HSIL-micro group. Age categories which split the data for both HSIL and HSIL-micro women more equally make the results more reliable. Your previous response to comment 8 ("The age categories for Table 3 and 4 are not sensible, as they do not split the HSIL-micro women (42/47 are in the 35+ category).") explains why you chose the categories, but did not address the fact that the cut-offs do not result in reasonable numbers in each category.

2. Not everything in the abstract is mentioned in the main paper (for example, cervical cancer is the 3rd most common cancer in Brazil).

3. ETZ, AIS, EJC should be spelt out before first use (in the text, not just as footnotes to a table).

4. Please name the confounders adjusted for in the statistical analysis paragraph.

5. In the last paragraph of the discussion, you say that the cPR was 5.45 for women under 25 – no cPR was calculated for women under 25, and 5.45 is the overall cPR.

Discretionary Revisions

1. In the last paragraph of the discussion, you say the cPR for women 35+ was similar to the overall aPR – this may be a result of age 35+ being the baseline category when the age variable was adjusted for – if so you would expect the cPR for women 35+ to be the most similar to the aPR, and it is an artifact of the analysis, rather than something to comment on.

Minor issues not for publication

1. In the statistical analysis paragraph, the test is a Student’s t-test (there is
currently a typo), and there are words missing from this sentence so it currently doesn't make sense – Student's t-test was used to do what to numerical variables?

2. It is more usual to present results in order of increasing age when split by age, rather than decreasing age, as in Tables 3 and 4.

**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