Reviewer's report

Title: Prevalence and predictors of cervical epithelial cell abnormality among women attending gynecological examination in cervical cancer screening unit at Debremarkos referral hospital, East Gojjam, Northwest Ethiopia

Version: 2
Date: 10 February 2015
Reviewer: Hannah Leslie

Reviewer’s report:

Major compulsory revisions

1. This manuscript represents an important addition to the knowledge base on prevalence and risk factors for precursors of cervical cancer in developing countries. However, it suffers from a critical shortcoming in study design and analysis, at least as currently presented. The decision to stratify the sample by HIV status is not explained or justified, and the major implications of this decision on the analysis and study findings are overlooked. Due to the deliberate oversampling of HIV+ and the absence of consideration of this decision in the analysis, the current results are not a valid attempt to depict a hospital-based sample of reproductive-age women. Sampling by HIV would be justified if the purpose of the study were to assess whether HIV contributes to epithelial cell abnormalities (ECA), but the authors note existing research suggesting this association in the background and do not pursue it as a major focus of the manuscript. Another reason for sampling by HIV would be if HIV status is considered a major confounder of the risk factors of interest. However, most of the risk factors analyzed are likely to be antecedents of HIV, making HIV status an intermediate causal variable. Because the sample is fundamentally shaped by the decision to sample by HIV status, the resulting analyses are biased relative to an overall clinic sample, let alone the general population. Aside from an etiologic investigation, even for the purpose of providing demographic traits associated with ECA to improve targeting of public health programs, the results are skewed by the sampling strategy. The authors must directly address this critical sampling decision, identify the research question(s) that motivated this choice, and take this into account in all analyses and discussion. The sample of 400 as constructed lends itself to analysis of HIV itself as a risk factor and to analysis of any factors that follow it in a causal pathway to ECA or that are specific to HIV-positive women (such as CD4+ T cell count). To obtain results pertinent to an overall clinic sample, the authors should weight the sample based on the selection probabilities or, less optimally, truncate the HIV+ sample at the number who would have been encountered in the time it took to collect 200 HIV- women. The presentation of a causal model, while not required, might help provide a clear depiction of the authors’ key research questions and ensure that these questions form the core of the introduction, results, and discussion.

2. The statistical analysis strategy is not described in sufficient detail: how did
authors determine what covariates to collect and how to categorize them? For example, how were age categories chosen – did authors assess linearity of the relationship between age and outcome and determine these categories captured a non-linear relationship? Are nulliparous women included in parity <3? Why is 5 years chosen as the breakpoint for OCP use? Were any other hormonal contraceptives assessed? Analysis of OCP in particular should be clarified: how are the women who have never used OCP, the majority of the sample based on Table 3, incorporated in this analysis? The text does not explain which factors are adjusted for in the AORs; the process of developing the adjusted model and deciding what to include (whether that decision process was causal or statistical) should be described.

3. The response (i.e., counseling, referral to HIV care, follow-up appointment?) if any, for women identified as HIV positive should be noted as part of the study protocol.

4. Population summary information (Results, paragraph 2) is not very meaningful given that HIV-positive individuals can be expected to differ systematically from the HIV-negative women in terms of both demographic and behavioral traits. As in critique #1, the authors should clarify what population they are trying to assess and how their sample (or a weighted version of the sample) represents this target population.

5. Statistical tests should be applied to any meaningful comparisons. This includes a reliability statistic such as a kappa test for consistency of examinations (results paragraph 1) and chi square or t tests as appropriate for comparisons between the HIV-positive and HIV-negative strata (Results paragraph 3).

6. The first paragraph of the discussion focuses on the utility of Pap smears and a recommendation for extending this method; this does not appear to be a focus of the article up to this point. The manuscript does not present itself as a feasibility study of this method, making the choice to emphasize this element at the start of the discussion misleading.

7. The representativeness of this hospital population to the overall population of women should be discussed, particularly when findings are put in context of existing studies of ECA (Discussion paragraph 3), since presenting at the hospital for gynecological investigation could clearly be related to disease state. It is inappropriate to compare the prevalence documented in this sample to other studies without considering the deliberate oversampling of HIV+ women.

8. The conclusion that ‘This study signifies that large numbers of these ECA positive women are at increased risk of cervical cancer’ is not correct – the study does not include future assessment of cervical cancer.

Minor essential revisions

1. Paragraph 4 of the Background overstates the evidence for risk factors such as oral contraceptive pills (OCP) – while the named variables are clearly associated with cervical cancer, causality has not yet been established due to confounding by behavior.

2. In keeping with the first major compulsory revision, the sampling strategy is
inadequately described, including number of women approached for involvement, number of women excluded due to pregnancy etc., and response rate. This information should be broken out by HIV status if possible. Other details to help contextualize the patient population would be useful, such as typical number of women presenting with gynecological problems, type of problems most often seen in this patient population. Estimated population prevalence of HIV should be presented, particularly for women in this age group if known. Did sampling take longer for the HIV+ population to achieve the desired sample size?

3. The sample size calculation is unclear (Methods, end of first paragraph) – a reference or explanation of the two-population proportion formula should be provided. If the proportion of expected ECA cases was used in this calculation, that estimate and the basis for it should be explained, since one purpose of this article is to estimate prevalence of ECA in this population.

4. The first paragraph of results includes information that should be incorporated within methods. The identity of the ‘investigator’ is unclear – is this the data collector, the laboratory technologist?

5. COR is not defined in the text, only in the table legend. OR is more typically used to refer to a crude odds ratio.

6. Citation 21 (Lippincott & Wilkins) includes errors in the authors and title. Please re-check references in case of other such errors.

Discretionary revisions

1. I was surprised to see an area of 2 million individuals described as a town (Methods, first paragraph); it would be helpful to see an estimate of the hospital catchment area by population and geographical size. There is no reference provided for the central statistical agency report.

Minor issues not for publication

1. Suggested English language revisions:
   a. Spelling (sever instead of severe, line 68; crud instead of crude, line 557; up on instead of upon, line 194; down ward line 216, line 296, line 326; counter parts line 222 and line 322; one or two sexual partner, line 222; corner stone line 241; parities instead of parity lines 319;).
   b. Missing hyphens: for example, ECA-positive women, HIV-positive women, long-term OCP use
   c. Variable spacing around numbers, including references as well as ‘100’ on line 60, ‘4 648’ on line 103, 1.0% line 261, ‘vaginal discharge were observed’), line 225 – 227, lines 235 – 238, 261 – 263, 309 – 310, 322-323
   d. Unclear sentences / phrases: line 70 – 71, 186 (‘were used condom’), 193 (‘vaginal discharge were observed’), line 225 – 227, lines 235 – 238, 261 – 263, 309 – 310, 322-323
   e. Repetition: lines 83 – 89, severity of disease informing treatment options

2. Suggested stylistic revisions: the background could be slightly reorganized to include the information on Ethiopia specifically at the end, prior to the research question addressed by the manuscript.
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: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests.