Reviewer’s report

**Title:** HIV prevalence among high school learners - Opportunities for schools-based HIV testing programmes and sexual reproductive health services

**Version:** 2  **Date:** 9 November 2011

**Reviewer:** Meade Morgan

**Reviewer’s report:**

***General Comments***

This manuscript provides results from a linked anonymous cross-sectional study of HIV prevalence among 12-25 year olds in two schools randomly selected from among 24 secondary schools in Vulindlela district in rural KwaZulu-Natal, South Africa.

The paper is very well written and the results concisely presented. The tables and figures are clear and help in communicating the findings. While I have a major concern with some of the statistical methods used as noted under “major compulsory reviews”, I think that these may be corrected and that the paper should be reconsidered for publication once that is done.

***Major Compulsory Revisions***

Results, page 5; table 1, page 13 (overall column for HIV prevalence portion of the table); figure 1 – In any study is it important that the statistical analyses that are done match the study design as nearly as possible. In this study the data come from two separate schools which were selected at random from among 24 schools. While it is appropriate to report results for individual schools separately (as the authors have done), it is not appropriate to simply pool the data when looking at results across schools. When calculating CIs and performing tests of significance, the data need to be treated as arising from a cluster sample which each school being a separate cluster. This concern is reinforced by the results showing significant differences between school A and school B for overall HIV prevalence and for HIV prevalence in girls.

To illustrate what difference this might make, I reanalyzed the raw data shown in table using SAS PROC SURVEYLOGISTIC (version 9.2, SAS Institute, Cary, NC USA) as follows. First, I recalculated the overall confidence bounds based on data from both schools together, and second I recalculated the p-values testing for a difference in prevalence between boys and girls. For all analyses I treated school as a “cluster” and used a sampling weight of 1 for each individual observation (this assumes that the two schools were selected from among the 24 with equal probability and that within each school each learner had an equal probability of selection, i.e. that all learners in each school were invited to participate.) I did not try to include non-response weights as the data provided in the manuscript did not allow that to be done.
For overall prevalence I obtained a two-tailed 95% CI from 1.8% to 6.0% (as opposed to 2.4% to 4.2% in the manuscript) and for prevalence in girls a CI from 2.0% to 11.4% (as opposed to 3.3% to 6.3%). The authors should either remove the CIs for the pooled data in the results section and in table 1 and in figure 1, and report just the sample prevalence, or they should use appropriate statistical techniques for the analysis of cluster survey data if they wish to report overall CIs for secondary school learners in Vulindlela district.

For a difference in prevalence between boys and girls, I calculated a p-value of 0.19 (as opposed to p<0.0001) as reported in the manuscript. Please note that, at least by my re-analysis if the data, the difference is no longer significant! In this case the authors might wish to report p-values for differences between boys and girls for the two schools separately rather than for the pooled data. For school A the difference is significant (p<.01) but for school B it is not (p=.11).

***Minor Essential Revisions***

Figure 1 – please indicate whether the whiskers represent standard errors or 95% confidence bounds. I believe they represent 95% bounds. Note that per my earlier comment, these will need to be re-calculated to account for the cluster sampling.

***Discretionary Revisions***

Results, page 5 – please consider reporting all percentage the text to the same number of significant digits. Most often the results are reported to one significant digit, but in some cases they are not (for example a lower bound of 1.99 on the prevalence for girls in school B).

Discussion, page 5 – in the first paragraph of this section the authors comment that HIV prevalence in young women attending ANC clinics in Vulindlela district (16.6% in 2006 and 20.8% in 2008) is lower than what was found in the same age group in among the teen girls and young adult women in this study. While there does appear to be a difference with respect to the point estimates, the previous discussion regarding confidence bounds coupled with the fact that the CIs from the ANC clinics are not available, suggests that it is possible that the differences noted might not actually be statistically significant. Given the differences in the point estimates, I’m not certain that this actually requires any revision to the manuscript, but as the authors consider the impact of their findings on prevention programs in these schools, they should realize that the differences may not be as substantial as their limited (two schools only) data suggest.

Discussion – it would be helpful to add a section on limitations or discuss some of the limitations somewhere in the paper. In particular I’m concerned about the possible non-response bias, particular in school A where the difference between HIV prevalence between boys and girls was greatest. While I’m not making it a recommendation for either a major or minor revision, it might actually be helpful to test for differences in response rates by sex and age in each school.
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.