Reviewer's report

Title: Sexual behavior and awareness of Chinese university students in transition with implied risk of sexually transmitted diseases and HIV infection: A cross-sectional study

Version: 4 Date: 9 June 2006

Reviewer: Andrew Vickers

Reviewer's report:

Numerous aspects of the statistical analyses could be improved. These should all be seen as essential revisions: 1, 2, 5 and 8 are minor, the rest major

1. Use appropriate levels of precision. It is silly reporting mean age to two decimal places: are we really interested in how old students are to the nearest 4 days? Similarly, don't give p values to 3 significant figures.

2. Do not report p values of zero (e.g. top of page 9, chi of 40.8, p=0.000). There is no such thing as a zero p value, every experimental outcome has a non-zero probability. Say p<0.0005 instead.

3. Don't compare p values. For example, on page 8, the authors state that "rates were significantly lower for male students in lower rather than higher grades, but not for females". The correct approach here is a multiple regression, with grade, gender and grade by gender interaction. The p value for the interaction term tests whether grade makes a difference for one gender but not the other.

4. Percentages should add up to 100%. For example, on page 8, 85% of respondents were heterosexual, 2% homosexual and 1% bisexual. Which makes about 12% of the population rather confused.

5. Some of the numbers are clearly erroneous. For example, in table 2, a total of 9.3% of women had 2 or more sexual partners. However, no grade had less than 11% of women with 2 or more sexual partners. Clearly an average cannot be less than all of the numbers that go into the average.

6. In table 4, give a p value for the test for trend.

7. In table 4, the reference category isn't clear. For example, as regards number of sexual partners (which, btw, rather confusingly compares 1 vs. more than 2) there is an odds ratio of 0.18 for women in grade 1 compared to 4. Does this mean that women in grade 1 are more or less likely to have multiple sexual partners than women in grade 4?

8. The univariate and multivariate results should be compared. For example, if something is significant on univariate but not multivariate analysis, you could say e.g. "the apparent association between A and B was actually due to C"

9. What is a grade anyway?!?!?

Overall, the paper needs to be simplified down. I counted over 120 p values. Now remember that a p value gives you information about a specific hypothesis. Do we really want to test >120 hypotheses on this data set? I doubt it. The authors need to identify some key questions, work out the best methods to answer them, and run a limited number of analyses.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

'I declare that I have no competing interests'