Reviewer's report

Title: Characteristics of Rural Cambodian Youth Sexual Reproductive Health and Risky Behaviors in two rural provinces of Cambodia

Version: 1 Date: 12 December 2014

Reviewer: Ilene Speizer

Reviewer's report:

This is a descriptive study of youth in two rural provinces of Cambodia. The paper covers numerous risk-taking behaviors and demonstrates the prevalence of these behaviors in the study population. The biggest problem I had with this study was the small sample size and the desire of the authors to examine in more depth the frequency and exposure among the small number of youth engaged in risk-taking behaviors (e.g., number of drinks among those who drink). One possible strategy to strengthen this study might be to develop a risk-taking index based on the main behaviors (sexual activity, alcohol use, drug use, exposure to pornography, etc.) and focus the multivariable examination on who is involved in more of these behaviors (and if enough sample, do this stratified by sex). This is similar to the last variable included in Tables 6 and 7. That said, in Table 7, if alcohol use is in the risk-taking behavior index, it can't be used as an independent variable of the model. If the authors focus the multivariable analysis on this index, then the more in-depth analysis of risk-taking (e.g., number of sexual partners) could be done among the medium and higher risk groups. This, however, would be a different study than the one presented here. Below, I list some ways that the authors could strengthen this or a future paper with these data.

Major Comments:

1) There are numerous grammatical problems throughout the paper. In addition, the paper jumps around and loses the reader (particularly in the introduction and discussion).

2) Methodology: The sample design was not completely clear. Did you select 15 villages (from some master list of villages in each province) or did you select one cluster of villages (from some master list of clusters of villages). As a reader, I am trying to get clarity on what the first phase was representative of. In the second stage, I think you selected 10 households (systematically) from each village, but there was no information to clarify on where the list of households came from – was there a listing and mapping exercise before selection? Or did you go to the center of the village and work systematically from there in some manner? Finally, how were youth selected at the household level? Were all youth eligible or only one eligible per household (or one female and one male)? If only one, how was this one selected? Was it a random selection with return to the household to find the selected youth if she/he was not home or was it based on presence at the
time of interview? A related matter is the low response rate – 62.6% - What was the reason for this low rate – was it refusal? Not being available (after some number of visits)? Or something else? Did you weight the data to adjust for non-response? Also, you seemed to nicely end up with 150 female and 150 male youth, were households substituted to get these exact numbers?

3) Survey methods (questionnaires) – from the write-up, it is not clear if all youth were asked all questions or if some youth were skipped. Some examples were questions on attitudes on plans for future sex – this seems only appropriate for unmarried (and never had sex) youth, but it seems all youth may have responded to this. It also wasn’t clear how things were defined. For example, there is a question on attitudes toward future expectations of sex in remaining adolescence. How was “adolescence” defined? If it is under 20, how would the youth who are 20-24 respond? Or did this include all of the ages surveyed? Finally, there are lots of hypothetical questions in this survey (and paper) which lead to lots of potential for response bias as indicated by the authors. This might be a good place to drop material and make the presentation more focused on actual risk-taking behaviors.

4) The last statement of the methods section says that only significance of $p<0.05$ are considered. But, in the results section (and discussion) many non-significant results are discussed. This seems inconsistent and the focus should be on the significant results. In addition, most of the logistic regression results are not significant which likely reflects the small cells so I think you are trying to do too much with the data.

5) In the results section (and Table 1), you need to discuss the total n’s by sex and then provide the specific ages of the sample. The mean age is useful but if there are about 25 female youth who are 20-24, it would make sense that about that many are married. Also, since the male youth are older, it is not surprising that they are more likely to be engaged in some of the behaviors so simply comparing by sex does not adjust for the age differences and can be misleading.

6) The results on relationships are confusing. First, for the question on current girlfriend/partner/boyfriend – how are married women or men meant to respond? Would they be “no” if no outside partner? Also, the low percentage that intend to get married is surprising and made me wonder if something went wrong with the coding/analysis (the n’s are smaller too).

7) In the presentation of the findings – instead of saying that 845% increase risk for males, it seems better to say that male youth have 9.45 higher odds of doing the behavior.

Minor Comments

1) There is no information on Institutional Review or ethics review in Cambodia and in U.S.

2) Page 12, I see 11% of youth have alcohol 13+ times whereas in the text is says 4%

3) Table 1 – there is something wrong with the numbers for earns income and control of income for boys – they do not sum to 100%
4) Table 5 – I see where the n’s came from but have no idea where the percentages come from.

**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