Reviewer's report

Title: Socioeconomic status, urbanization and risk behaviours in Mexican youth: an analysis of three cross-sectional surveys

Version: 4 Date: 25 July 2011

Reviewer: Joanna Stewart

Reviewer's report:

Major
1. The old Table 1 should not be included as additional material. As stated last time, given the different ways the samples were chosen within the different types of localities, stratified on poverty score in different ways, any analysis across type of locality must include a measure of poverty score as well as the cluster effect. Also it is invalid to do multiple pairwise comparisons in separate analyses.

2. The reporting in the results of differences observed in the original table 1 needs to be removed from the results and conclusions. Ie all of the second paragraph of results bar the 1st sentence. Table 1 simply presents the data to show the characteristics of those surveyed. It is incorrect to state that rural communities have more in lower SES groups when the surveys differently selected according to poverty score or to comment on %’s living in poor households. Any comments on differences need to come from the multivariable analysis.

3. Table 2. Both interactions included in the analysis need to have their p values reported for all outcomes. It is not stated anywhere whether in fact there was statistical evidence of an interaction of SES and locality type for these outcomes although the text reports it as if there was. The results should only be split on SES for those outcomes where there is some evidence of an SES, locality type interaction (and similarly for gender and marital status). The text interpreting this table needs to be rewritten accordingly. At present there is also ambiguity as to what test conclusions are based on (eg no sig association found for locality size and current smoking or the p<.05 in the following paragraph).

4. The data from table 2 has been over interpreted in the paragraph referring to the figure. It is implying real difference within the semi and urban localities for which no evidence has been given (eg reference to U shaped relationships. Eyeballing the odds ratios there would seem to be little evidence of differences within the urban groups

5. Methods. You state that the additional rural localities added in 2003 were ‘matched’ to those in the original sample. On what variables? The original sample was stated to be ‘the most marginalised’. Was the additional sample therefore by definition less marginalised? How was ‘marginalised’ defined? What was its connection to poverty score which is stated in the data management section to be used to determine program eligibility? Was the cut point the same for
programme eligibility in all locality types? It is important that the SES variable included in the analysis which is a categorisation of the poverty score incorporates divisions that tie in with those used in the eligibility for the program and hence in selection of the survey samples.

6. Discussion. There are several places where conclusions being drawn are not backed up by results of analyses.

7. In the limitations there is reference to the population from which the sample is selected. However it does not point out the differences in the populations in the different types of localities in relation to eligibility to the programme. It needs to be made clear to the reader exactly who was included and therefore who these conclusions can be generalised to. At present it is very difficult to understand exactly who was included. My understanding is that for small urban it was people in locations eligible for the programme (which needs a clearer definition) , for rural it was those eligible originally (although most of those are excluded as they had received the programme) with the addition of those also from poor localities but presumably not as extreme as the original locations, and for large urban everybody. However I certainly could be wrong as it is very difficult to understand this at present. It needs to be clearly summarised in the methods as well as in the limitations as it is very important in the interpretation of the findings. Some indication also needs to be given in the abstract as to who these results are for – at present there is nothing to convey that the surveys are not random samples of all Mexican households.

Minor essential

1. Stats methods – remove the sentence describing the old table 1 analysis. It would be clearer if all the independent variables were included in one sentence rather the scattered over 2 paragraphs. What level is used as the reference category needs to be clear for all variables, either by stating in the methods or as a footnote in the table.

2. Remove the footnote to table 1 about the adjustment for PSU

3. If you are going to use *’s in what will need to be a new table 2 they need to be explained.

4. Abstract Methods – The second to last sentence says localities were classified according to the number of inhabitants in each residence. This sounds like a household crowding measure and is not referred to in the paper. Is this supposed to be the population size measure?

5. Abstract Results Odds ratio, not rates

6. Background 6th paragraph. Quoting the number of adolescents that have been pregnant is of little use without a denominator.

Discretionary

1. If the additional figure is included it would seem more useful to have a separate plot for each outcome rather than for SES group to graphically display the interaction of SES and locality type (if indeed it exists)
**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