Reviewer's report

Title: Tobacco use and household expenditures on food, education, and healthcare in low- and middle-income countries: A multilevel analysis

Version: 1 Date: 23 February 2015

Reviewer: Rosemary Hiscock

Reviewer's report:

This paper looks at whether tobacco use reduces spending on other vital areas such as education. The introduction and discussion are well written. The methods are inadequately presented

1. Is the question posed by the authors well defined? - yes
2. Are the methods appropriate and well described? – appropriate but not well described
3. Are the data sound? - yes
4. Do the figures appear to be genuine, i.e. without evidence of manipulation?
5. Does the manuscript adhere to the relevant standards for reporting and data deposition? NA
6. Are the discussion and conclusions well balanced and adequately supported by the data? yes
7. Are limitations of the work clearly stated? Yes
8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished? yes
9. Do the title and abstract accurately convey what has been found? yes
10. Is the writing acceptable? mostly

Major revisions

1) P7 You need to introduce your analysis by saying that it was multi level and then individuals were level one and countries were level two. It would be helpful to include the exact stata models that were used e.g. xtreg

2) How you define your three types of model (fixed effects, random intercept, random slope) is crucial to your paper but I feel it is insufficient. This is important because there are lots of competing definitions of fixed effects and random effects see e.g.

http://andrewgelman.com/2005/01/25/why_i_dont_use/

In this context, from my understanding of multilevel modelling (via MIWin) I would assume the following:

Fixed effects: the equation defining the relationship between tobacco use and expenditure is the same for all countries i.e. level 2 is not needed in the model
Random intercept: The average level of expenditure is allowed to vary for each country.

Random slope: The relationship between expenditure and tobacco use is allowed to vary by each country i.e. in some countries tobacco use will be associated with reduced expenditure, whereas in other countries there might be no relationship or expenditure might increase.

Either expand and add these definitions or correct them.

3) Tables 2 & 3
a. Tests: Is it necessary to present the coefficients/degrees of freedom of the likelihood and hausman tests or would the significance suffice? In the footnotes can you remind the reader what the tests mean? i.e. if the likelihood test is significant it means that model X is preferable, if the hausman test is significance it means that model y is preferable.

b. Why are you presenting all 3 models (random slope random intercept, fixe effects) when your tests suggest some models are inferior? Why not only present the hausman and likelihood results and then the model that your tests suggest is the best?

c. You need to define what is meant by standard deviation in tables 2 and 3 in the methodology. Output from MLWin on random variance is presented differently so you need to be clear so that the reader understands.

4) You need to explain how you calculated the marginal effects and the pooled effects presented in Fig 1.

5) Table 1
a. could you use significance tests to look for difference between groups on tobacco use prevalence e.g. chi square.

b. Can you make clear whether prevalence includes just daily or also occasional smokers?

6) There is insufficient evidence in your paper that you have dealt with the following issues:

a. I am concerned that your model might be confounded by differences in government provision between countries. In Russia, with its communist past, there is extensive government provision of health care and education unlike some African countries for example. If smoking prevalence happens to vary systematically with the extent of government provision your results may be confounded. Can you add country level government provision as a level 2 variable?

b. Does your modelling take into account that some households will not have children and will not need to spend on education?

c. non normal distribution of variables

d. multicollinearity
Minor revisions

1) Table 1

could you use significance tests to look for difference between groups on tobacco use prevalence e.g. chi square. Can you make clear whether prevalence includes just daily or also occasional smokers?

Education: What is the difference between secondary completed and high school completed? Are secondary and high school not the same institution?
Can you change the country income indicators to the following:
“Country-level income group
Middle income
Low income”
If you like you can remind the reader that high income countries were excluded in a footnote

2) Typos and clarifications
a. P3 para 1 ‘prevalence’ should be ‘prevalences’
b. P3 delete “Clearly, the extent of the current literature on smoking suggests that”
c. P4 last paragraph. “more” You need to be careful here- low SES households may have lower absolute levels of spending but a higher proportion in the decline of spending on essential items
d. P5 second paragraph I think you need to bring out more clearly here that the strength of this study is that data is available from many countries in an easily comparable format
e. P6 “Focusing on the male respondents who are the main household income providers, 22,571 observations were further excluded” Do you mean households where a woman was the main income provider were excluded? Please be more specific here. What proportion of households were headed by women?
f. P8 I think some of the information about stata in the citation should be in the bibliography
g. P8 results Could you add some percentages about education to “A relatively higher smoking prevalence was reported among respondents with a lower level of education compared with those who completed at least high school.”
h. P10 “Although previous research on the crowding-out effect of tobacco use HAS reported a negative association between tobacco use and household expenditure on healthcare”

Discretionary revisions

1) Might higher costs of food be because in LMIC the very poorest don’t as yet smoke?
Level of interest: An article of importance in its field

Quality of written English: Acceptable

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