Reviewer’s report

Title: Association of socioeconomic and behavioral factors with causes of death: Analysis of data from Verbal Autopsy in Addis Ababa, Ethiopia.

Version: 2 Date: 16 May 2012

Reviewer: Georgie MacArthur

Reviewer’s report:

This paper investigates the association between lifestyle behaviours and socio-economic/demographic indicators and mortality from communicable and non-communicable diseases among adults in Ethiopia using verbal autopsy data. This is an interesting topic but several points could be addressed by the authors:

Major compulsory revisions

1. Overall, the study addresses a very broad aim in analysing the association of the chosen variables and all underlying causes of death, grouped into broad categories of communicable diseases, non-communicable diseases and injury. It would be more helpful if data were described in relation to the associations between risk factors and more specific categories of diseases.

2. Similarly, the authors could provide greater detail in relation to analysis by age group – currently the three categories include a wide range of ages in which the extent of engagement in the risk behaviours measured, and incidence of communicable/ non-communicable diseases, is likely to vary widely. As such, the data described are likely to mask important differences between groups which should be reported.

3. It is not clear the extent to which the standardized WHO and INDEPTH network questionnaires were adapted and why this was done; and justification is required for the choice of behaviours and their categorisation; please also provide details of the categorisations themselves e.g. frequency categories.

4. The categorisation of farmers and those that are self-employed in a separate category to manual labour or professional should be explained. Should these not be moved from ‘other’? A different category should be used as the reference category for these data in Tables 4 and 5.

5. Overall, the abstract is a little long and could be shortened so that only the key results are presented- for instance, the proportions drinking frequently, smoking and chewing khat are similar among the entire sample and among those deceased by communicable and non-communicable diseases; this could be made clearer and results summarised. The OR data needs to state the finding in one direction but not both.

7. The introduction would benefit from additional referencing; and there is some repetition in the content which should be removed for greater clarity.

8. The introduction would benefit from a greater focus on the Addis Ababa/
Ethiopia context, for instance in relation to the prevalence and/or patterning of the risk behaviours and diseases examined.

9. Please could the authors state how deaths were randomly selected for inclusion.

10. Greater detail is needed around the statistical methods - e.g. chi-squared test - please expand the statement relating to selected variables; and note outcomes and variables included in adjusted regression analyses.

11. In the discussion, the authors need to include a section specifically relating to the limitations of the study e.g. impact of missing data; and additional sources of bias not mentioned. The impact of the selection bias mentioned should be described; and the repetition of results could be reduced.

12. Greater justification is required for explanations of findings given in the discussion, supported by the literature.

13. Figure 1: please add labels to all of the columns, report the age range considered in the title, highlight in the legend the variables detailed e.g. ever/ frequent etc, and use the same labels for each behaviour.

14. It would be helpful if the authors could amend Table 2 so that it is clear which behaviour the n numbers refer to, e.g. by including a total in sample label in the left column – does this refer to the total number that answered the question? If just 9% answered a question relating to ‘smoking until they were sick’ it is unlikely that this will provide an accurate measure of that behaviour. In this table, please also note in the Figure/ Table legends which measures of behaviours are used; what the p values refer to; and provide n numbers for alcohol as well as tobacco/khat. Lastly, it is not clear why some of the data are highlighted in bold text, please list in the same font for clarity.

15. In Table 3, the measure of each risk behaviour used here needs to be specified – is this ever/ occasional/ frequent? In all tables, sig should be replaced with ‘p’ (and where listed as zeros changed to <0.01 or other as appropriate; and the legend should describe what confounders were included in the multivariable logistic regression. Please amend alcohol users to alcohol use so that all column titles reflect the behaviour rather than the population.

Minor essential revisions

1. It would be helpful if authors could structure the methods such that each aspect is discussed in turn e.g. burial surveillance; VA; risk behaviours.

2. Justification is required around use of Global Burden of Disease classification rather than ICD-10 codes.

3. It would be helpful to describe in greater detail the training and supervision of interviewers e.g. the nature of training; a description of supervision – was this during interviews/ a proportion of interviews?

4. In the results section, for clarity, please amend sentences with two proportions to read e.g. ‘47% of males and 10% of females had...’ rather than ‘males 47% and females 10%...’ There is also some repetition in summary of the results here.
5. Please amend the last sentence in section 2 of the results ‘the deceased who were practicing both alcohol and cigarette smoking were higher 4% than both smoking and khat chewing practitioners 2%’. Perhaps the authors could list the proportions practicing each combination of 2 behaviours?

6. It would be more appropriate if the authors replaced phrases such as ‘x% of deaths.. were drinking alcohol/ reported drinking alcohol’ to ‘x% of individuals ... were..’.

Discretionary revisions
1. Descriptive data about the sample could be included in the abstract e.g. n; age range.
2. Presentation of data would be clearer with fewer decimal places.
3. It would be of interest for the authors to assess the association of combinations of risk behaviours with the outcomes.

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

**Quality of written English:** Not suitable for publication unless extensively edited

**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