Reviewer's report

Title: Alcohol use and misuse: What are the contributions of occupation and work organization conditions?

Version: 4 Date: 27 May 2008

Reviewer: Andrew Roddam

Reviewer's report:

This paper reports on an analysis of the Canadian QHSS data set looking at alcohol use and misuse. This review concentrates on the methodological/analytical aspects of the manuscript only.

There are a number of compulsory revisions which the author needs to address/clarify:

1. On the whole the author has data which is either Yes/No or comes from a scale. There are some instances where the scale results from collapsing a continuous variable, e.g. age is a scale collapsed from the original values. However at no point in the manuscript does the author justify using the scale variables as linear continuous variables in the analysis. For example it is assumed that age is linearly related to outcome but this is not tested – has the author assessed these assumptions with the variables used, and if not consideration/justification should be given.

2. It would appear that for the two models, the author picked a range of variables that was felt a-priori to be important, included them in the model and tested whether or not overall the model was significant without regard for the significance of the individual variables. Moreover, for multiple category variables (like occupation), the author doesn’t provide a test of the significance of the variable – for example is occupation significantly associated with alcohol use? If not, then it is not clear how – or even if – one should mention that one of the many levels of this variable is associated with risk – since the lack of an overall association suggests that individual levels are simply associated with risk due to chance. The author should include such a p-value/test for this variable. Similarly for the continuous (scale) variables an overall test of whether or not each of them are significantly associated with alcohol use/misuse across both categories of alcohol use should be included.

3. With regard to model 2 – the author then includes some further variables which change, rather dramatically, the estimated ORs for occupations from model 1. To me this suggests that there is a high correlation between occupations and one of the new variables included – for example, household income? One has to be careful with this sort of analysis to ensure that variables are not too co-linear, and that including additional variables, does in fact improve the overall model fit. The author should consider whether the approach of
including everything in the model is the best - and again should include tests of significance for each individual variable (as described above).

4. A final comment regarding the last 2 points – would it be possible/appropriate for the author to consider some other strategy in an attempt to find the most parsimonious model?

There are also some rather minor changes to improve the presentation of the data:

1. In Table 1 – rather than show percentages as 0.xx, show as xx%.
2. When reporting variables such as Age, which is a categorisation of a natural scale it is much more informative to a reader to show actual mean age, rather than mean age of the categorical scores.
3. In Table 2 – for each of the groups it might be helpful to include an n so that readers can see how many people contribute to each of the categories. In doing so it might be worth then making Model 2 into Table 3.
4. Some of the variables used by the author are on scales which are not as meaningful as they could be – for example cigarettes are coded as per cigarette. Perhaps recoding of some of these scales, say per 5 or per 10 cigarettes would provide more directly meaningful results.

**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