Reviewer's report

Title: Modelling alcohol consumption as a distribution and determining the impact of the distribution on estimated alcohol-attributable harms

Version: 1 Date: 31 July 2011

Reviewer: Tanja Srebotnjak

Reviewer's report:

The manuscript investigates 3 continuous and one categorical distribution for modeling alcohol consumption and their impact on estimating the alcohol attributable fraction (AAF). It is well-written, concise and addresses a problem that falls within the scope of Population Health Metrics, although it is probably of main interest to those working on alcohol-related burden of disease and/or AAF

* Major Compulsory Revisions

None.

* Minor Essential Revisions

1. Aside from the log-Normal, the Gamma and the Weibull distributions, the authors examine a categorical distribution for alcohol prevalence distribution (see Methods section and Formula 2). It seems, however, that the quality of the categorical distribution to adequately capture prevalence depends among other things on how fine-grained the exposure intervals are. Could the authors discuss this aspect and how it has been dealt with in the literature. Which categorization did they choose for this analysis?

2. The Conclusion to recommend the Gamma distribution - while justifiable - appears to be a bit too strong given that there is no significant difference between the tested functions. The Gamma has the advantage to be flexible and be up- or downshifted to account for differences between survey data and sales data on alcohol consumption (the previous paper by the authors on what distribution is best suited to model alcohol consumption is relevant in this context, too). It should be made clearer if this was the deciding factor in the end to recommend the Gamma function?

3. The discussion of the AAF estimates from the 3 continuous and the categorical distribution is cut very short in the Results section and could be elaborated more.

* Discretionary Revisions

4. "it" missing in "the AAF using categorical measurements rather than modelling in a more mathematically appropriate continuous manner."

5. When introducing the 3 continuous distributions it would be good to define the random variable x (alcohol consumption) and parameters (e.g., mu=mean, sigma=variance, etc.) for consistency. In some cases they are already defined in
others they are not.

6. The authors state that alcohol consumption values above 150 g/day were truncated to 150 g/day. Could this lead to a local maximum by heaping values onto 150 g/day and create problems in estimating the AAF?

7. Alcohol consumption is often a heavily right-skewed distribution but the means of the Gamma and Weibull distributions were found to be close to the empirical means. In particular the Gamma’s mean is always equal to the empirical mean. Question: is the mean a statistic of interest when dealing with heavily skewed distributions or would other location parameters such as the median be more appropriate and hence functions that have the property of estimating it well?

8. While the title of the manuscript is very informative it is quite long. Could it be shortened, e.g., to something such as “The impact of the choice of alcohol prevalence distribution on estimated alcohol-attributable harms”?

**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.