Reviewer's report

Title: Income inequality and alcohol related harm in Australia

Version: 1 Date: 21 July 2008

Reviewer: Gerhard Gmel

Reviewer's report:

The present study looks at the interplay between income inequalities and alcohol-related harm in an ecologic analysis. The authors claim to have found an uncommon relationship (U-shaped). Unfortunately, I can not follow the authors’ conclusion about a U-shaped curve, and I'm afraid that some of their findings may be related to methodological artifacts, which have to be addressed in a future version of the MS.

1) All coefficients for the Gini and Gini-squared in tables 2 and 3 are strictly positive. Therefore, there is no way that the quadratic function can take a U-shaped form (given that the independent variable is strictly positive too); the resulting function must be monotonically increasing. Thus, most of the arguments made in the discussion are hard to follow. At least the authors should explain the differences between regression models and the Loess fits in figures. May it be that Loess fits did not adjust for the same variables as the regression models?

The authors need to explain more the link between the U-shape and the coefficients from regression models that do not show a U-shape.

2) I am very much concerned about the ecologic validity of the study. The risk of running into an ecologic bias is well known. The risk of an ecologic bias increases with the heterogeneity within the ecologic units. An ecologic analysis would be best if there was large heterogeneity across units but a strong homogeneity within units. The large variation of unit sizes (LGA) in representing Australian inhabitants is striking. It ranges from 85 inhabitants in the smallest LGA to more than 880’000 (see table 1) in the largest. Hence it ranges between a street block and a metropolitan city.

3) Related to 2) is the problem that one aim of the study is partly in contradiction with the need for homogenous units. Either a unit is homogeneous, than there should be no or little income inequality, or vice versa. The authors should discuss this.

4) I’m wondering, how much of the variation in Gini coefficients is actually related to the size of the LGA. The larger the LGA the more likely are income inequalities within the LGA. Thus, much of the effects found in the paper may just be related to size of LGAs. This problem is even aggravated by the fact that the authors weighted the analysis with the size of the LGA (namely the number of deaths or hospitalizations, see page 6 “data analysis”). The rationale for this should be given. Personally I’m rather opposed to weighting by size, because this only
favors large units. I suggest as a sensitivity analysis a) an unweighted analysis
and b) an analysis that controls for the size of LGAs.

5) How reliable is an analysis with 885 deaths and more than 500 LGA, i.e. less
than 2 deaths per LGA on average. This will probably mean that in many small
LGAs there were no deaths at all!? This leads to a serious problem:

How did the authors deal with those zeros? As an SMR is the ratio of observed
deaths divided by expected deaths, the SMR will remain zero for LGAs without
deaths. The log of zeros (see also my point 8), however, is undefined. From the
visual inspection of e.g. Figure 3 with only 28 datapoints (a datapoint should
represent a LGA), I’m afraid that I have to assume that all LGAs with zero deaths
were just left out of the analysis, and thus the mortality analysis probably only
refers to large LGAs.

I do have a couple of minor comments.

6. Abstract: “chronic (primarily related to the short term consequences of
drinking)” should read long term.

7. Abstract: I can not see that there is a “strong” association given that the overall
R-squared (including control variables) are around 0.15; authors should avoid
labeling their results.

Generally, the conclusions could be more concrete. What was the direction of the
association; what is "complex" in the relationship; with which findings are the
current findings inconsistent?

8. Why were SMRs logged? The distribution of the SMRs is not important for the
assumptions of linear regression, because the dependent variable was not the
SMR directly, but a ratio of SMR (alcohol-related versus other). Was this ratio
also skewed? (I don’t want to be a nitpicker, but in fact, not the distribution of the
dependent variable is important, but the distribution of errors). Not logging may
solve some of the problems (see above).

9. Why did the author adjust for the Socio-Economic Index for Areas (SEIFA).
Wouldn’t this be in the causal path between Gini and alcohol-related harm.
Variables in the causal path should not be adjusted for (see Rothman,
Introduction to Epidemiology). Similarly the remoteness; again I thing this
variable just reflects larger (major cities) versus smaller LGAs and thus
eliminates some of the variance of the Gini (which is in turn related to size of
LGAs). At least, the authors should explain why they think both variables are
confounders. Again a sensitivity analysis (leaving out the confounders) may help
here.

10. I had sometimes difficulties to follow the different numbers of LGAs in
different analyses. For example, what does it mean that "There were around 140
more LGAs available for the mortality analyses" (page 6)? More than what? More
than the 373 LGAs for the analysis of acute alcohol-attributable hospitalizations
or more than the 349 LGAs for chronic hospitalizations (same page)? How do
these numbers match with the 580 Gini coefficients in table 1, respectively with
the 428 and 562 Gini “coefficients used” in table 1? Authors could be a bit more
explicit about the n’s of LGAs for different analyses; could a flowchart help here?

**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'