Reviewer's report

Title: Alcohol consumption and risk of incident ovarian carcinoma: a pooled analysis of 5,342 cases and 10,358 controls from the Ovarian Cancer Association Consortium

Version: 1 Date: 18 August 2012

Reviewer: Matteo Rota

Reviewer's report:

Review of manuscript 'Alcohol consumption and risk of incident ovarian carcinoma: a pooled analysis of 5,342 cases and 10,358 controls from the Ovarian Cancer Association Consortium'

This article by Dr. Kelemen et colleagues presents a pooled analysis of 12 case-control studies in the Ovarian Cancer Association Consortium investigating the relation between alcohol drinking and epithelial ovarian cancer risk. Although the relation between alcohol consumption and epithelial ovarian cancer risk has been widely investigated in the previous years, and recently by a meta-analysis (Gynecol Oncol 2012; 125(3):758-93), this pooled analysis could be considered the largest study to date to perform this evaluation quantitatively on original data, allowing also to investigate the most common histotypes of ovarian carcinoma, and also the two groups of borderline tumours.

Data and methods are well described, and conclusions are balanced. The authors concluded that alcohol intake was not associated to epithelial ovarian carcinoma. This finding is consistent with the results of the aforementioned meta-analysis.

I have only two general major comments and some minor observations that I wish authors to take into account.

Major compulsory comments:

• The main challenge of statistical analysis of pooled data is between studies heterogeneity. The authors fitted a logistic regression model including an interaction term between alcohol intake and study site. To this aim, the paper “Two-stage methods for the analysis of pooled data” of Stukel et al., published in Stat Med 2001; 20:2115-213, describes a two stage random-effects regression model that allows taking into account study specific covariates and between studies heterogeneity. The choice of the authors to use a “joint fixed-effects logistic regression model has been demonstrated to produce unbiased estimates only when studies are homogeneous with respect to exposure effects and when all important confounders are measured and uniformly defined by all studies, a situation that rarely occurs in practice”. From Supplementary Table 1 it emerges that the distribution of alcohol drinking of the HAW and HOP studies is very different from the distribution of alcohol intake of the other included studies.
Moreover, the two forest plots of Figure 1 seem to show some heterogeneity. Nevertheless, similar parameter estimates for both the fixed and random effects based models could be indeed obtained considering the large sample size of this investigation.

I wonder whether authors were aware of this technical aspect, and if they planned during the analysis the possibility of fitting to the data random-effects models, and why they decided to analyze data with fixed-effects modeling. Were results of the models similar?

• The authors “re-assigned EOC histology type according to the expected distributions of histology combined with grade observed from large population-based series” (page 8). The application of this pathology-based algorithm was seen by the authors as strength of the investigation (page 14). However, this algorithm could also introduce a bias. Thus, this could also be a limitation. I wonder whether authors evaluate the association between total alcohol and wine intake and histological types of EOC (Table 3) on the original data, i.e. without the application of the pathology-based algorithm. Are there differences in the magnitude of the association between total alcohol and wine intake and histological types of EOC without the application of the pathology-based algorithm? This point could be better described within the discussion.

Minor essential comments:

• I agree with the author’s choice of excluding from the analyses non-epithelial ovarian tumours (page 7). In fact, it is worldwide recognized that approximately 90% of ovarian cancers originates in the surface epithelium tissue. However, in the title and within the manuscript text, the authors often use only “ovarian carcinoma”. I would recommend the authors to use the term “epithelial ovarian carcinoma” or its abbreviation (EOC) when appropriate.

• Please define the abbreviation OCAC (“Ovarian Cancer Association Consortium”) within the text (page 5 row 12).

• Page 5. Reference number 43 is not informative. In fact, Table 1 of this manuscript is essentially the same as Table 1 of the referred article. So, there is no need to include this citation.

• Page 5. In the Study Subjects section (page 5), I wonder whether it is possible to add a reference to each included case-control study after the study abbreviation. This consideration also applies to Table 1. Is it possible to add a reference after the study abbreviation in Table 1?

• Page 6. Authors state that “Daily alcohol intake for all studies was calculated by summing the product of the frequency of consumption...using national estimates of alcohol content for that country”. This is correct. However, considering that the ethanol content of a specific alcohol beverage could be very different across countries, I wonder whether it is possible to add a sentence describing briefly the range of ethanol content across countries for beer, wine and liquor (for example, beer from 4% in Europe to 6% in USA, or somewhat similar).

• Page 7. Please add “consumed daily” at the sentence “Alcohol intake
categories were derived in increments of one standard drink consumed daily”.

- Page 7. The authors state that “The trend in risk were evaluated treating the category values of alcohol intake as continuous variables in the regression models.” Please specify if the mean or median value of each category was used.

- Page 7. I would remove the paragraph “In addition, among a subset of studies, primary analyses were also adjusted for total energy intake,…..again in increments of alcohol grams of one standard drink per day”. In fact, as stated by the authors within the Results section, “Further adjustment for total energy intake had little effect (data not shown)”. So, considering that data adjusted for energy intake was not shown, I think that there is no need to describe so extensively the adjustment for total energy intake within the Statistical analysis section.

- Please add within the Statistical analysis section a paragraph describing methods related to the forest plots of Figure 1, i.e. DerSimonian and Laird paper (Control Clin Trials 1986;7(3):177-88). Did the authors use random effects modeling for pooling estimates across studies?

- Page 11 row 10. I think that the P of heterogeneity was computed on 4 df (number of categories -1), and not 5. Is it true?

- Page 12. Within the second paragraph of the discussion, the authors state that “alcohol has been equivocal for ovarian cancer [17, 26-33, 36, 38]”. However, the recent meta-analysis (reference 38) provided no evidence of an association between alcohol and epithelial ovarian cancer. So, the findings of this manuscript are consistent with the results of the recent meta-analysis. The use of the term equivocal for reference [38] is not correct.

- A general issue in case-control studies is the possible presence of recall and selection bias. I think that this should be listed within the discussion as a limitation of pooling case-control studies.

- Table 1. The last column of Table 1 is not informative. In fact, authors clearly state within the Study subjects section (page 5) that all the studies used population-based ascertainment methods. Is it possible to substitute this column with a new one including the period of enrolment of each case-control study? Furthermore, could authors add a column in Table 1 with the (main) matching factors of each case-control study, along with a brief description in page 6?

- Table 2 and Table 3 (and also supplementary table 3). Please define the abbreviation DK (don’t know?), please specify in the table notes “ * 1 drink (10 g of alcohol)…”.

- Figure 1A and 1B. Please add the I2 heterogeneity statistics and the P-value for heterogeneity within the two forest plots.

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

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