Reviewer's report

Title: Evaluation of contaminated drinking water and preterm birth, small for gestational age, and birth weight at Marine Corps Base Camp Lejeune, North Carolina: A cross-sectional study

Version: 1 Date: 1 September 2014

Reviewer: Lucas Salas

Reviewer's report:

The paper by Perri Z. Ruckart, Frank J. Bove, & Morris Maslia entitled “Evaluation of contaminated drinking water and preterm birth, small for gestational age, and birth weight at Marine Corps Base Camp Lejeune, North Carolina: A cross-sectional study” is a continuation of their previous work in evaluating a historical leaking of volatile organic compounds (mainly TCE, PCE and benzene) into the groundwater provided to an US Marine Corps Base in North Carolina between years 1968 and 1985. They have previously evaluated other birth outcomes in the same setting.

The following are my comments for the authors, that I hope will help to improve the paper quality:

• Major Compulsory Revisions

Comment 1: I disagree about the design stated by the authors. A cross-sectional design assumes that prevalence of the outcomes and the exposures occurred at the same timeframe just like a snapshot (cross sectional). In your case, your dataset is a historical registry based cohort and the exposure assessment provides information of a timing of exposure which is assumed to be previous to the outcome appearance. In my opinion that corresponds to a historical (retrospective) cohort not to a cross sectional design. Please correct accordingly or if you disagree provide specific references to support the design.

Methods:

Comment 2: At study population subtitle paragraph (page 6) “All singletons 28-47 weeks of gestation weighting #500 grams”. This period is non-plausible, as human pregnancy is normally 39 to 40 weeks. Maximum postterm delivery is 42 weeks under very tightly controlled conditions. At this sentence also weighting has the “t” missing. Please correct accordingly.

Comment 3: You have used unconditional logistic regression to estimate your relative risks-RR using odds ratios to approximate your RR (Miettinen, 1982). You are dealing with subjects in a historical dynamic cohort whose you censor once the delivery occurred (the subject was excluded from further prenatal exposure). This class of design usually uses other regression approaches to deal with this censoring to estimate the RR (usually a Poisson or a semiparametric Cox survival analysis). However, you decide to use of OR and then you should
be careful. When you use OR to estimate your risks, it requires the “rare disease” assumption for being an adequate proxy of the RR in this kind of cohorts, otherwise your estimations will be inflated given than both the exposed and unexposed subjects had higher than 10% of “cases”. As stated in table 1A SGA and preterm births do not hold the classic “rare diseases” label (as both are higher than 10%). In consequence your estimates may be inflated for these outcomes. Have you considered this potential inflation on your risk estimates? Have you used an alternative regression to compare the results? Please state this issue clearly on the paper.

Comment 4: In your case you do not have information about specific risk factors during pregnancy. Neither have you had available population exposure to other environmental or lifestyle factors in a very long time frame (1968 to 1985). On the other hand, your design includes the year of delivery and the mother age, both of them influencing the birthweight. In my opinion, this makes your data suitable for an age-period-cohort analysis given that other unmeasured factors in the period of pregnancy may be related to the outcome. A recent approach for attaining this has been recently published (Margerison-Zilko C. The contribution of maternal birth cohort to term small for gestational age in the United States 1989-2010: an age, period, and cohort analysis. Paediatr Perinat Epidemiol. 2014 Jul;28(4):312-21. doi: 10.1111/ppe.12127. Epub 2014 May 7.), which is accompanied by and an editorial comment (Keyes KM, Ananth CV. Age, period, and cohort effects in perinatal epidemiology: implications and considerations. Paediatr Perinat Epidemiol. 2014 Jul;28(4):277-9. doi: 10.1111/ppe.12129.). Please support your approach and state the age-period (cohort) effect that may affect your estimations.

• Minor Essential Revisions

Tables and figures:

Comment 5: Usually when presenting cubic splines or GAMs you add a scatter of the number of subjects just above the x axis (exposure variable). This guides the reader about potential flat areas related to insufficient sample size. Other graphical approaches are also acceptable, but the estimates alone are difficult to interpret without this information. Please correct accordingly.

Comment 6: On table 3, the star marks PCE as it was adjusted by mother race. If I am right the star should mark the OR column, not the row, or it could be interpreted as if the percentages were adjusted and not the risk estimates. Please correct accordingly.

• Discretionary Revisions

Comment 7: When defining the three outcomes: Have you compared your term low birth weight (TLBW) with your term Small for gestational age (SGA)? I guess they agree but just out of curiosity given the different populations, how much they agree?

Comment 8: Have you any estimation of the number of pregnant women who had their delivery off-base? If so, this information may be interesting to add.

Comment 9: On table 2, given the TCE findings a linear p-trend would be
interesting to be added. The same comment for Table 5.

**Level of interest:** An article of outstanding merit and interest in its field

**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