Reviewer’s report

Title: Pancreatic Cancer Clusters and Arsenic-Contaminated Drinking Water Wells in Florida

Version: 2 Date: 3 June 2012

Reviewer: How-Ran Guo

Reviewer’s report:

The authors have tried to address the issues raised in the previous review, but some of them were not addressed fully.

Major Compulsory Revisions

1. The logistic regression model used to generate the major results (Table 2) should be given in mathematical formula. In their reply, the authors explained that logistic regression was used to “document the probability of a block group having a higher-than-expected incidence of pancreatic cancer,” but in the manuscript they stated that logistic regression was used to “assess potential predictor variables across groups… the dependent variable was a patient with pancreatic cancer living in a neighborhood with a higher than expected pancreatic cancer incidence (a “cluster”) versus being diagnosed not in a cluster.” What was the unit of analysis (dependent variable)—a block group or a patient? The authors have explained in their reply that it is not feasible to model the probability of a disease occurring in a participant versus not occurring. However, there is a gap between a patient occurring vs. not occurring and a patient being diagnosed in a cluster vs. not in a cluster. The authors should discuss briefly the issues involved in extrapolating the model used in this study to the model corresponding to the research question directly.

2. In their reply, the authors seemed to indicate the confounding effect of smoking is a potential explanation for the reason why an association between arsenic exposure from drinking water and pancreatic cancer was observed in this study, but not in most studies which evaluated the associations between arsenic ingestion and the whole spectrum of cancers. However, they did not address this issue directly in the manuscript. On the other hand, since the model has included smoking as an independent variable, the confounding effect of smoking cannot be used to explain the association, unless there were residual confounding effects.

3. In comparison with people who never smoke, current smokers had a higher risk (OR=1.1) whereas former smokers had a lower risk (OR=0.9), both with marginal statistical significance (95% CI with 1.0 as the upper or lower bound). In their reply, the authors interpreted the results as “smoking cessation would reduce the risk.” However, the data indicated a protecting effect, no just off-setting the increased risk.

4. In their revision, the authors indicated in the legend the meaning of green
spots in Figure 1. However, Figure 1 shows a discrepancy between the
distribution of arsenic contaminated wells and distribution of higher than
expected pancreatic cancer clusters. Can a correlation coefficient be generated
between the two distributions in terms of block group?

**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