Author's response to reviews

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

Authors:

Wen Liu (wliu_1999@yahoo.com)
Jill A MacKinnon (jmackin@med.miami.edu)
Recinda Sherman (rsherman@med.miami.edu)
Lora E Fleming (lfleming@med.miami.edu)
Caio Rocha-Lima (crocha@med.miami.edu)
Jennifer J Hu (jhu@med.miami.edu)
David J Lee (dlee@med.miami.edu)

Version: 6 Date: 14 February 2013

Author's response to reviews: see over
February 11th, 2013

Christna Chap, PhD
Senior Executive Editor
BMC-series
BioMed Central
236 Gray's Inn Road
London WC1X 8HB
Email: editorial@biomedcentral.com

RE: MS 1032211416636929 - Pancreatic Cancer Clusters and Arsenic-Contaminated Drinking Water Wells in Florida

Dear Dr. Chap,

Thank you for your email in which you sent us the comments of Reviewer 2. Please find the following response to Reviewer 2. Your consideration is deeply appreciated and we look forward to hearing from you soon.

Sincerely,

Wen Liu, MD, PhD
Response to the Reviewer 2

Reviewer’s report
Title: Pancreatic Cancer Clusters and Arsenic-Contaminated Drinking Water Wells in Florida
Date: 9 December 2012
Reviewer: How-Ran Guo

Reviewer’s report:

From the cover letter, I believe the authors have tried to address my major concerns that they failed to cover. My comments are as following:

Major Compulsory Revisions

1. The figure provided in the response is helpful and thus should be included if the Editor decides to publish the manuscript. The formula I was looking for is for the logistic regression model, not for the likelihood ratio test, nor for the SaTScan program. There is no harm to include those descriptions, but they can be omitted if the Editor feels that the size of the manuscript should be cut down. While the authors still failed to give the formula of the logistic regression, the description in the revision is sufficient for the readers to understand the models. Nonetheless, they stated in their response that “The dependent variable was the binary outcome that whether the block groups categorized as having either an excess or expected pancreatic cancer incidence by the cluster detection software described above.”; but in the text they stated “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.” These two are different. To begin with, one used a “block” as the unit of analysis, but the other used a patient as the unit. Furthermore, even using a patient as the unit, as I pointed out in the first review, 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 (Why the association between arsenic exposure and being diagnosed in a cluster of pancreatic cancer can be explained as an association between arsenic exposure and the occurrence of pancreatic cancer itself? Were there no alternative explanations?).

As stated in our response in July, the authors have used the similar methodology examining disease clusters in other types of cancer and cancer related risk factors in the published papers (MacKinnon et al 2007, Nieder et al, 2009, and Dietz et al, 2011). The method used in our study including the figure in our response in July has been described in details previously (MacKinnon et al 2007). We have included the figure (Figure 2) in the modified version of the manuscript if the Editor wishes to include it.

The formula for logistic regression is standard:
Logit (p)=\beta_0+\beta_1X_1+\beta_2X_2+…+\beta_mX_m

As you indicated, there is sufficient description of logistic regression in our July response:
“Logistic regression analysis, which then can be used to model binary outcomes, was used to document the probability of a block group having a higher-than-expected incidence of pancreatic cancer. Of note, the words “neighborhoods” and “area” used throughout this paper are intended to be synonymous with the term “block group(s).” The logistic regression analysis was utilized to model the probability of pancreatic cancer cases falling
within and outside of these geographic clusters at the time of diagnosis. The dependent variable was the binary outcome that whether the block groups categorized as having either an excess or expected pancreatic cancer incidence by the cluster detection software described above. The independent variables included: (a) race/ethnicity, (b) census-derived SES, (c) reported tobacco use, and (d) proximity to known arsenic-contaminated wells.”

In the context of our manuscript, it is stated:

In these models, the dependent variable was a patient with pancreatic cancer living in a neighborhood (“block groups”) with a higher than expected pancreatic cancer incidence (a “cluster”) versus being diagnosed not in a cluster.

In a summary, a “block” was used as the unit in the analyses. As indicated in the chart in Figure 2, the data of each patient were aggregated at block group and level and used in the analyses as the unit.

It is stated in the Discussion section: “Furthermore, we found an association between relatively close proximity to arsenic-contaminated drinking water wells and clusters of pancreatic cancer. Specifically, we found that living within 1 mile of the known arsenic-contaminated drinking water well might be a threshold distance for an increased risk of being diagnosed within a cluster.” Based on the findings of this study, the clusters of pancreatic cancer were found associated with proximity with arsenic-contaminated wells. The associations with other risk factors for pancreatic cancer have also been discussed.

Arsenic is a known carcinogen for the development of several cancer types. Our findings indicated a possible association with pancreatic cancer.

1) To our knowledge, this is the first study to find that more pancreatic cancers were observed among those living closer (within one mile) to arsenic contaminated wells than those living farther from these wells;

2) If arsenic exposures alone do not cause pancreatic cancer, arsenic may promote/enhance the potential existing carcinogenesis such as smoking, diet, etc.;

3) This study also could provide clues for next step evaluation of the interactions between arsenic exposure with smoking, and nutrition;

With the findings of the association of pancreatic cancer clusters with proximity to arsenic contaminated well water, we believe that publication of the findings of our study are hypothesis generating so that the potential association could be explored, replicated and validated by other research groups in the future studies.

2. My previous comment was that 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. In their second reply, the authors explained that it might due to an interaction between arsenic and smoking. If so, they should demonstrate it by including an interaction term in the model. As an alternative explanation, the authors admitted that their “data on smoking was [were] both incomplete and subject to potential misclassification.” If so, they should not include the variable in the model.
Smoking is an independent variable and it was included in the multivariate logistic model with variable proximity to arsenic contaminated well and other variables such as race/ethnicity, and social economic status (SES) as described in the above response to item 1. Therefore, our finding indicated the relationship between pancreatic cancer clusters and proximity to arsenic contaminated wells after controlling smoking status (the confounding effect of smoking) and other variables (please see Table 2).

In animal models, the synergistic effects of arsenic and other carcinogens (such as smoking) are suggested to enhance the tumorigenicity (38). But this interaction has not been studied in human pancreatic cancer. In this study, we focused on the analyses of the main effect of proximity of arsenic contaminated wells with pancreatic cancer. In future studies, when sample size and power are permitted, the interaction terms could be explored in the multivariate analyses.

3. While the authors tried to explain why there seemed to be a protecting effect of smoking in the cover letter (again, they offer the fact that data on smoking were both incomplete and subject to potential misclassification as an explanation) but did not discuss this issue in the revised manuscript.

We would like to emphasize again that the findings of our study did NOT indicate “a protecting effect of smoking”. As indicated in our response in July:

“In our study, smoking cessation would reduce the risk of being in the pancreatic cancer clusters (OR=0.9, 95% CI=0.8-1.0). That is, if a participant was a smoker previously but stopped smoking at the time of study, he had a lower risk of being in a pancreatic cancer cluster. On the contrary, if a participant was a current smoker at the time of study, then he had a higher risk of being in a pancreatic cancer cluster (OR=1.1, 95% CI=1.0-1.2).”

As in any epidemiological study, there are missing data in this study. As indicated in the manuscript, we have incomplete information on smoking status in the cancer registry record as noted in our limitations section: “Smoking status was abstracted from the medical record at the time of cancer diagnosis, which may be inaccurate. Furthermore, over one fourth of patients (26.9%) did not report their smoking status. Therefore, misclassification of this important risk factor is likely present.”

4. The revised Figure 1 still shows a discrepancy between the distribution of arsenic contaminated wells and distribution of higher than expected pancreatic cancer clusters: many green spots were not in a red area, and many red areas had no green spots. Can the authors calculate the proportion of green spots in red areas (blocks) and the proportion of red areas with green spots?

If we were to create a similar figure to show the distribution of smoking and lung cancer as in this manuscript, it could still indicate such “discrepancy” as you described between the two. Other factors influence why not everyone who smokes has lung cancer, and even among those who have lung cancer, not everyone smokes. Similarly, in this study not every clusters of pancreatic cancer (red areas) overlap with arsenic-contaminated wells (green spots). Such relationship is appropriately and adequately examined using the standard multivariate logistic regression described above.