Reviewer's report

Title: A Prospective Cohort Study of Stroke Mortality and Arsenic in Drinking Water in Bangladeshi Adults

Version: 1 Date: 4 November 2013

Reviewer: Katherine James

Reviewer's report:

This research investigates the association between arsenic concentrations in residential well water and CBV mortality in a prospective cohort study. The research is fundamentally sound and with revisions will contribute to the body of literature suggesting that arsenic is a systemic toxicant with adverse effects even at low levels. Stroke is a top contributor to mortality and findings from this study could help in public health planning worldwide.

Major Compulsory Revisions

The authors use a prospective study design which is sound, however there are several concerns that should be addressed which are documented below:

1) The second sentence, second paragraph Introduction “is one of the leading causes of death, is responsible for 10% of deaths globally” this sentence is disjointed and state where it ranks globally verses using a vague word like ‘leading’.

2) In the third paragraph Introduction you state the major limitation of the previous studies (ecologic assessment of exposure), however you justify this study by stating that the arsenic exposure is variable. It should be stated that the exposure assessment is individual level here (you state this later in the paper)

3) The cohort inclusion and exclusion are confusing, for example in “A total of 180 811 individuals were eligible and visited their homes during the study period (January 2002-August 2003). Thus, 166 934 individuals (92 % of eligible) were interviewed and examined” It isn’t clear how you reduced the cohort by ~ 13,000.

4) The greatest concerns I have is in the statistical analysis. Specifically:

a. The exposure variable is residential baseline arsenic levels and then subjects are followed over ~7 years. Given the research location, it is expected that arsenic levels could vary substantially causing misclassification bias. It is understandable that doing continuous water sampling for 13,000 wells may not be feasible however, the 1-2 sentences in the Discussion section addressing this as a weakness is not substantial. The authors acknowledge it as a major weakness, however there is no discussion on the temporal variability of arsenic in the study area nor is there discussion addressing the potential variation in water consumption all of which contribute to the limitation in accuracy of the exposure assessment. This section of the discussion needs to be greatly expanded.
b. Along the lines of the exposure assessment, why did you categorize arsenic levels vs. running the proportional hazards model with exposure as a continuous value and separately as a categorical variable. What is the reason for categorizing?

c. A sensitivity analysis should be run comparing the results when the records with SES imputed are included and when they are removed to ensure the imputation process was appropriate.

d. How variable is SES in this area?

e. How did you handle loss to follow up? Were these right censored?

f. Did you assess for historical stroke events prior to baseline? If so did you exclude on this? What about family history, is this a concern with CVB?

g. What is the time interval in the proportional hazards model? Month? Year?

h. Given the study design with monthly data collection I am surprised you did not use time dependant covariates and exposure, this would strengthened the analysis greatly.

5) The crude mortality rate for CVB was presented throughout the second paragraph of the results section. Present the overall crude rate, but then use age-sex standardized rates. For example “The crude death rate was much higher among the participants who were more than 60 years of age” This sentence doesn’t add anything because the crude rate is typically higher in older people.

6) Since you reported differences in demographic factors for those excluded you should discuss more about how this could influence the findings. This is especially true given the disparity in the sex distribution and the much higher percent of females in the study cohort compared to the LTFU. This higher percent in the cohort may be driving the overall cohort significance.

7) Given that males have a significance level of p=0.99, it would appear that the association is non-existent in the males so it might be better to downplay the overall significance of the association and focus and stress the association in females since the high percent of females (59%) may be driving the overall significance.

8) Limit the survival curves to just women

9) Unless there is a mechanistic justification to analyze by specific diagnostic code do not include this as it doesn’t add to the findings.

10) In the third paragraph of the Discussion, you suggest that women have a higher association bc they use betal squid which can increase risk for stroke, however this argument only accounts for the outcome and not the exposure. If betal squid had high levels of arsenic in it, then it might explain why women have a stronger association. As written you are suggesting that all women are at a higher risk for stroke due to anemia/folate/squid but this would apply to all women and would not explain the higher association with arsenic unless squid is associated with arsenic. A much stronger discussion needs to be done on the gender issue as a whole. This could be a very significant finding that isn’t explained well. Is it that women are getting exposed to more arsenic? Is it that
arsenic exposure and altered folate or iron levels act synergistically to increase risk?

Minor Essential Revisions

1) Carefully edit the entire paper for language, sentence and paragraph structure, and grammatical mistakes. The paper is inhibited due to the writing. There are many throughout the whole paper, however a few are documented below:

a. Second paragraph of Introduction “Much studies supports the fact that arsenic in drinking water causes increased coronary artery disease (CAD) and death from CAD…” should read, “Many studies support the fact that arsenic in drinking water may increase coronary artery……”

b. First paragraph Introduction. “This is a major concern; ….” This is out of place, and should be removed

c. Second sentence and paragraph of intro is a fragmented sentence

d. You need to spell out all acronyms, what is icddrb

e. Second paragraph of Methods is a fragment

f. Include commas in all of your numbers especially if they go to a next line

g. The following sentence doesn’t make sense “This study used prospectively collected monthly health and demographic surveillance health and demographic surveillance system (HDSS) mortality data during the period from January 2003 to December 2010 collected through verbal autopsy (VA)” it says monthly health and demographic health were collected along with demographic surveillance system. It is assumed you meant “This study used prospectively collected health and demographic data collected monthly through the Health and Demographic Surveillance System (HDSS)”

2) The use of the verbal autopsy is impressive and even though it is detailed in other literature, 2-3 sentences explaining the methods would be appropriate.

3) Need to add the DISCUSSION subtitle

Level of interest:An article of importance in its field

Quality of written English:Needs some language corrections before being published

Statistical review:No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I have no competing interests.