Reviewer's report

Title: Applying Verbal Autopsy to Determine Cause of Death in H River Basin, China

Version: 1 Date: 8 April 2011

Reviewer: Zhengming Chen

Reviewer's report:

The study took place following extensive media report in China of "tumour village" found in certain rural regions along the Huai River that was readily attributed to pollution, with the aim to establish or refute these claims made in the media. Since there were no established death registries in those regions, verbal autopsy was used to ascertain causes of death and the disease rates were then compared using SMR between affected and control areas. In my view this is an important study with major public health implications. However, there are several issues with study design and data analysis.

Major comments:

1). The study showed that the overall death rates, especially those from cancer, were much higher in the affected than in the control areas. Before it can be accepted as real, several other possibilities should be considered first. It is not clear from the calculated SMR whether the affected and control areas are compatible at all. This is particularly important in China given the well documented large regional variations in cancer rates. If possible, basic information should be provided for these regions, perhaps based on a random sample of the general population involving not only general socioeconomical status, but also known risk factors for chronic diseases (such as smoking, alcohol drinking, blood pressure, HBV infection, etc) to help assess their general comparabilities.

2). In the Methods section, it was mentioned that there were 3 regions selected for the study, each with affected and control areas. Based on sample size calculation, about 50,000 people in each area were included. It is not clear what does the figure mean? Is this the number of people investigated or the population size in the study catchment area where the death events were derived? Apart from use of VA for ascertaining causes of death, what else were done during the survey? What were the selection criteria for control and affected areas and did they follow mainly the media report or something else? For easy of communicatino, it is also important to spell out the full name of the river and study areas as well as to provide a location map. Otherwise it makes people feel that the whole thing is very secretive.

3). Since the study concerns primarily with the disease of middle or old ages (eg, cancer), it is not appropriate to mix the data from children, young adults, middle-aged and elderly population. To improve the sensitivity, all the main
analyses should be separated for those aged 15-34, 35-69 and 70+ and, where possible, separate analyses should also be done for men and women. For regional specific analysis, it would also be important to present the overall results that combine all three areas, without it the specific regional findings may be considered by many as inappropriately selective emphasis.

4). To help establish the validity of the study findings, it is also important to present the results on other non-neoplastic conditions (eg, CVD, COPD). Given the widespread media coverage on cancer, it is quite possible the results on cancer rates may be inflated in the affected areas both by informants and by the VA process (eg, medical staff who were doing event adjudication seemed not blinded to data from different study areas). Since the main source of pollutions was through water, it is somewhat surprising to see that there was significant excess risk of lung cancer, which is unlikely to be affected by the water pollution. The lack of specificity on cancer excess risk raised major concerns about the validity of the study findings, especially the effects of confounding by other known risk factors such as smoking.

5). From the reading of paper, it is not clear what is the main focus of the paper. Is it about the validity of VA or the confirmation of pollution-related cancer excess risk in the Huai river basin? From scientific and public health perspective, the latter would be more important but on current standpoint, the authors seemed to try to play down it. For example, all the discussion was related to VA rather than validity of the study findings (eg, effects of uncontrolled confounding factors, comparability of the study and control areas).

Minor comments:
The writing needs to be improved and use of certain abbreviations does not seem conventional (eg, COD for causes of death). Apart from authors, many other people were involved in the project and there were probably several funding sources for the study, however, no acknowledgements were made to these at all.

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

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests.