

## **Author's response to reviews**

**Title:** Hazardous Waste Sites and Stroke: A New York State Ecologic Study

### **Authors:**

Ivan Shcherbatykh ([shcheriy@mcmaster.ca](mailto:shcheriy@mcmaster.ca))

Xiaoyu Huang ([xiaoyu.h@gmail.com](mailto:xiaoyu.h@gmail.com))

Lawrence Lessner ([LLessner@uamail.albany.edu](mailto:LLessner@uamail.albany.edu))

David O. Carpenter ([carpent@uamail.albany.edu](mailto:carpent@uamail.albany.edu))

**Version:** 2 **Date:** 22 June 2005

**Author's response to reviews:** see over

## Response to the Comments

*Black italic* – original comments

**Red** – author's response to the reviewers

### Reviewer #1

1. *Since hospital discharge diagnoses is the outcome variable, a main question is what happened with the diagnoses of those who deceased after admission due to a stroke?*

**The SPARCS registry requires reports of all discharged patients, which includes anyone who is admitted and dies after admission.**

2. *Page 8: I would suggest to move part of the last paragraph in the Result section to the Discussion section: "As previously reported. Variation of stroke discharge rates". Merge it with the identical sentences in the 1st paragraph of the Discussion section.*

**Done**

3. *The conclusion is too long and large parts should be moved into the Discussion section. The Conclusion section should only include a conclusion such as the last as paragraph on page 11.*

**Done**

4. *The 2nd paragraph on page 10 addresses the problem of aggregative data. This needs to be moved to the Discussion. It needs to be stated clearly, that it is possible that individuals from an residing in an exposed ZIP code never were individually exposed (aggregative data). I do not like the term ecologic study/analyses, since aggregative data analyses are done in a variety of areas, not only ecology.*

**Done. The title is also corrected.**

5. *The 1st sentence in the Conclusion section needs to be supported by a reference. What are the previous investigations?*

**The references are included**

6. *Page 11: Give a reference (or personal communication?) for the association between low birth weight and ZIP codes with potential PCB exposure.*

**Done**

7. *Page 2: Result section: Comma after "For ischemic stroke only",*

**Corrected**

8. *Page 2: Background section: I would recommend to state that POPs may lead to. The statement conveys final evidence where there are only suggestions.*

**Done**

9. *Tables 1 and 2: Include only one decimal for percentages.*

**Done**

10. Please provide information about dispersion and goodness of fit in tables 3 and 4. The information is below. We have not included this in the text, but certainly could do so if you think it desirable. It just seems to us to be a bit long and detailed.

**Table 3**

Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	40	57.8688	1.4467
Scaled Deviance	40	57.8688	1.4467
Pearson Chi-Square	40	60.6882	1.5172
Scaled Pearson X2	40	60.6882	1.5172
Log Likelihood		174076.2286	

Algorithm converged.

Value/DF close to 1 indicates that there's no evidence of over-dispersion.

**Table 4**

Criteria for Assessing Goodness of Fit

A. Ischemic stroke

Criterion	DF	Value	Value/DF
Deviance	88	118.4691	1.3462
Scaled Deviance	88	118.4691	1.3462
Pearson Chi-Square	88	123.4681	1.4030
Scaled Pearson X2	88	123.4681	1.4030
Log Likelihood		100433.5989	

Algorithm converged.

Value/DF close to 1 indicates that there's no evidence of over-dispersion.

B. Hemorrhagic stroke

Criterion	DF	Value	Value/DF
Deviance	88	99.1554	1.1268
Scaled Deviance	88	99.1554	1.1268
Pearson Chi-Square	88	100.6710	1.1440
Scaled Pearson X2	88	100.6710	1.1440
Log Likelihood		13477.1941	

Algorithm converged.

Value/DF close to 1 indicates that there's no evidence of over-dispersion.

C. Subset of Hudson River zip codes

Criterion	DF	Value	Value/DF
Deviance	88	110.8844	1.2600
Scaled Deviance	88	110.8844	1.2600
Pearson Chi-Square	88	115.5996	1.3136
Scaled Pearson X2	88	115.5996	1.3136
Log Likelihood		123259.5285	

Algorithm converged.

Value/DF close to 1 indicates that there's no evidence of over-dispersion.

11. Page 7: What test of confounding was conducted. Confounding is not a question of a statistical test but whether the strength of the association is changed by more than 10% when the confounder is taken out of the model.

Corrected. A statement about confounding test is inserted into the “Statistical analysis” section (end of the first paragraph).

\*\*\*\*\*

## Reviewer #2

### Major Compulsory Revisions

1. The authors should provide additional qualification of their conclusions. In ecologic studies (or partially ecologic designs like this one), confounding can occur both between and within groups; control for median income at the group level does not eliminate confounding within groups. Effect measure modification can also lead to bias. Comparison with the Hudson Valley does not eliminate these potential problems. A plausible alternative hypothesis is that the small RRs in Table 4 are due to bias. The RRs for both POPs zip codes and other wastes zip codes are very similar; perhaps proximity to waste sites (rather than POPs) is correlated with a socio-economic or other risk factor.

We agree that the results of our study could be confounded or biased. The discussion section is corrected to clearly acknowledge that fact.

2. page 10: Explain why the limitations of this data set would be expected to lead to underestimation of true relationships. Ecologic bias due to confounding and effect measure modification can work in any direction.

Correct, any direction of bias can be expected in this kind of studies. The sentence in question has been removed. However, we believed given that the study used zip codes of residency as surrogate of exposure (where errors are relatively random, since there is no prior information about or/and strong dependence between exposure and health outcome) and keeping in mind a fact that environmental exposures in general tend to be of a low magnitude, that any biases introduced in this manner would be non-differential and oriented towards the null.

### Minor Essential Revisions

1. page 3: Additional information regarding inhalation exposures to POPs near hazardous waste sites would increase the plausibility of the exposure scenario. While reference 14 (cited on page 3), indicates that inhalation of total PCBs may be an important route of exposure relative to diet for some people, it does not discuss waste combustion or volatilization from soil; instead, it provides some evidence for indoor sources of PCBs such as sealants. Furthermore, when expressed as dioxin equivalents, inhalation of PCBs was small compared to diet.

Done

2. page 12: Describe contributions of other authors.

Done

1. page 1: A very brief discussion of known risk factors for the outcome might be helpful for many researchers primarily interested in POPs.

We have not chosen to add to the abstract, which already contains reference to the role of elevated serum lipids in increasing risk of stroke. There is considerable information in the discussion of the role of lipids, hypertension, diet, exercise and smoking as risk factors.

2. page 6: Was there evidence of overdispersion in the data leading to the choice of the negative binomial model?

Yes. At first we have applied the GENMOD Procedure of the SAS software to run a Poisson Regression and, as a result, the presence of extra Poisson variation in the form of a large value for the deviance and a large spread on the residual plots was detected. This extra variation (a.k.a. overdispersion) meant that Poisson regression was not adequate to describe the counts of stroke discharge. Extra Poisson variation frequently occurs with real data (1). Cameron A, Trivedi P. Regression Analysis of Count Data. New York: Cambridge University Press; 1998;4; (2). Byers AL et al. Application of negative binomial modeling for discrete outcomes: a case study in aging research. J Clin Epidemiol. 2003; 56: 559-64)

3. page 7: I suggest rewording or explaining how you tested for confounding. Non-epidemiologists might think this means a statistical test.

Done

4. pages 9-10: Is there any evidence that non-dioxin-like PCBs affect serum lipids?

This has not been investigated to the best of our knowledge. The reports of endothelial cell damage by PCB exposure is apparently limited to dioxin-like congeners. Since non-dioxin-like PCBs (especially the phenobarbital inducers) also induce P450s in the liver there is no reason to believe that the stimulation of the liver is limited to dioxin-like congeners, however.

\*\*\*\*\*

### Reviewer #3

Major compulsory revisions:

1. In the discussion, the issue of ecologic fallacy needs to be addressed. For instance, the information obtained on behavioral factors from the Behavioral Risk Factor Surveillance System may not reflect the health behaviors of those hospitalized for cerebrovascular disease. The issue of completed exposure pathways also needs to be discussed in more depth in the discussion. Living in a zip code area containing a waste site with persistent organic pollutants does not necessarily equate with a completed exposure pathway or close residential proximity (i.e. within a mile) to these sites.

Done. Pages 10-12

2. *In the analyses, race is divided into two categories: white and African American. There might be some residual confounding, however, by not accounting for other ethnic/racial groups such as Hispanics, Asians, etc.; this potential residual confounding might have led to some of the positive associations found. The choice of just two racial groups needs to be commented on as well as limitations that this choice might pose on the interpretation of the positive associations*

Race and/or ethnicity are known to be reported inaccurately and incompletely to SPARCS and the Census; for instance, SPARCS database hospital personnel assess race and ethnicity subjectively. Therefore, although stratifying by several types of race may provide a clue as to which segments of the general population are at a greater risk of stroke, limitations in classification must be recognized. In addition, our preliminary analysis (which included all races) showed zero cases in many age categories for Asian and Native American races. Restriction of race to two categories (White and African American) did not change the estimates significantly. Although this fact still leaves the possibility for residual confounding (by some unidentified factors associated with different race groups), we decided to restrict the analysis to these two categories.

Minor essential revisions:

1. *The authors refer to the comparison zip codes as clean sites and as zip codes without any contamination. It is recommended that the authors refer to the referent zip codes as zip codes without documented hazardous waste sites. Clean sites may include areas with yet-undiscovered hazardous waste sites or other environmental entities that could be a source of pollution.*

This is certainly correct, and we have added a qualifier in the text at the point where the term “clean” is introduced. We use the term as a space-saving mechanism, recognizing that “clean” may not be “clean”.

2. *There needs to be some indication in the abstract that age and race were adjusted for in the analyses and/or results. The analysis in the abstract would also be clearer if the authors indicated that the referent group was zip codes without any documented hazardous waste sites rather than clean sites.*

Done

Discretionary revisions:

1. *In the statistical model, it would be helpful if the authors indicated what values each variable contains, i.e. what are the values for AGE5?*

The age group ‘25-34’ was a reference group in the model, AGE3 represented the age group ‘35-44’, AGE4 – ‘45-54’, and AGE5 – ‘55-64’