Reviewer's report

Title: Highway proximity associated with cardiovascular disease risk: the influence of individual-level confounders and exposure misclassification.

Version: 2 Date: 17 May 2013

Reviewer: Ryan Allen

Reviewer's report:

The paper is much improved from the previous version and I appreciate the authors’ efforts to incorporate reviewers’ suggestions. In particular, the paper’s objectives and contributions (e.g., assessment of ‘novel’ confounders and geolocation methods) are more clearly stated. I still find the results section a little difficult to follow given all of the different models and sensitivity analyses.

Major Compulsory Revisions

- Why does the analysis of different geocoding results (Table 3) rely on unadjusted models? Results from fully adjusted models (at least for variables that are routinely included in these kinds of analyses) are much more relevant.

Minor Essential Revisions

- 4th sentence of abstract: Please clarify whether “Restricted analysis found the effect of proximity was mostly downwind from the highway...” refers to effects on concentrations or effects on health markers. I think you’re referring to health effects here.

- Please include the variables that were adjusted for in the final models in the main text of the paper (I could only find them by looking at Table 2). Under Table 2 I see that driving on a highway, cooking with oil, time spent at home, and AC type were included in the fully adjusted models. Given that these infrequently considered variables are a main focus, this would seem like an important result that should be explicitly stated in the text (but only cooking with oil is mentioned).

- Although it was not my comment, I agree with the other reviewer’s comment/question (#6) about also using a continuous distance variable. The authors’ responded that “the use of a linear distance variable would not reflect the dispersion gradients for PNC from the highway...” The authors seem to have misinterpreted the reviewer’s suggestion of a “continuous” variable with a suggestion to use a continuous and LINEAR variable; couldn’t a spline, or even a simple log-transformed distance (which correlates very closely with concentrations in many near-roads studies), be used? Also, I don’t understand the statement that “Distance to highway was explored as a continuous variable, but was found to not be appropriate since there is a gap between 450 m and 1000 m where participants were intentionally not recruited.” Why do you need participants at every distance in order to use a continuous distance variable?
Page 7: For the UFP measurements, was some method used to adjust for temporal trends?

Page 8: Which potential effect modifiers were explored?

Page 8: “two other models were developed, a model adjusted for variables that could influence exposure to air pollution (“exposure adjusted”) and a fully adjusted model”: it’s not clear from this statement whether the fully adjusted model also included the variables that could influence exposure to air pollution (i.e., were the fully adjusted models adjusted for ALL the exposure modifying variables PLUS additional variable such as BMI?). At present, one needs to study Table 2 in order to understand the different models.

In the stratified analyses (Figure 2), why were age and sex not evaluated as is common in these types of analyses? I also recommend stratifying by diabetes, which has been shown in several previous studies to modify the CV risks of air pollution.

Page 14: Please add to the list of limitations that the analysis excluded an entire neighborhood because it was too complex.

Discretionary Revisions

First sentence of abstract: I recommend changing this to read “Elevated cardiovascular disease risk has been reported...”

Conclusion of abstract: I recommend indicating which specific “rarely included” confounders were important.

Page 8: “Residuals were checked and found to be normal.” Please change this to “...found to be normally distributed.”

Page 12/13: When discussing possible reasons for the much larger effect estimates, I recommend also mentioning potential participant differences. For example this study included diabetics (did the other studies?) and had a high mean BMI near 30. This may have been a susceptible population.

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

Quality of written English: Acceptable

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

Declaration of competing interests:

No competing interests.