Reviewer's report

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

Version: 1 Date: 11 February 2013

Reviewer: Sandrah Eckel

Reviewer's report:

Comments to the authors:

Thank you for the opportunity to review this interesting manuscript that quantified the associations between distance from a residence to a highway/major roadway to 4 blood markers: hsCRP, IL-6, TNF-RII, and fibrinogen, while assessing the sensitivity of the results to the method used for geocoding residential addresses. There were associations of distance to freeway with hsCRP and IL-6. These associations were stronger when using the most refined geocoding method (ortho-photo corrected parcel matching) and when restricting the analysis to downwind of highway or upwind with considerable side street traffic.

Major compulsory revisions/comments:

1. Careful consideration of potential confounders is standard in studies of air pollution health effects, particularly in prospective cohort studies where detailed information is collected expressly for this purpose at the individual-level. The authors place a heavy emphasis on controlling for individual-level confounders. Do they consider that any of their confounders are unique or novel, or that their approach is different from common practice? If not, I'd recommend reducing the emphasis on confounding (while appropriately accounting for confounders) in favor of focusing on the more novel features of this work: sensitivity to different geocoding methods and sensitivity to upwind/downwind.

2. Why was the data from 450 to 1000 m excluded?

3. I usually think of automated model selection procedures such as stepwise selection as being most appropriate for model building when the goal is prediction rather than estimation of a particular association of interest. In estimation models, one typically focuses on whether inclusion/exclusion of potential confounders changes the regression coefficient of interest (or, if an important predictor of the outcome reduces the unexplained variance, and hence reduces the standard error on the estimated regression coefficient of interest) and whether there are important interactions with the exposure of interest. I found that the exhaustive discussion and inclusion/exclusion of potential confounders distracted from the presentation of the main results. To simplify the presentation and interpretation of results, I would recommend carefully selecting a single set of covariates for use in “exposure adjusted” and in “adjusted” models across biomarker outcomes, with additional sensitivity analyses exploring, for
example, the robustness of results to also accounting for medication use.

4. Was there any important effect modification in this study?

5. Table 3- Why compare geocoding methods with no adjustment for potential confounders when the authors emphasize the importance of adjusting for confounders? Ultimately, I think most readers will be interested in the impact of refined geocoding methods on the final (adjusted) results.

6. The continuous exposure of interest (distance to highway) was only considered as a categorical variable. Was there sensitivity to the choice of categories? Perhaps using a continuous version of the exposure variable and exploring the (adjusted) non-linear exposure-response relationship using smoothing splines (for example, with a Generalized Additive Model) would provide additional insight (particularly for the 50-150m group and hsCRP). Since the sample size is relatively small, such an approach might be a more efficient use of degrees of freedom.

Minor essential/discretionary revisions/comments:

1. Since this was a prospective study where study personnel visited the residence location of each participant, would it have been possible for each study personnel to carry a GPS and manually “mark” the location of each residence? Would this approach have any advantages or disadvantages over the method of ortho-photo corrected parcel matching?

2. P.3. approx line 14. More than a “few” studies of air pollution health effects have considered distance metrics as well as pollutions levels (e.g., estimated from land-use regression models). Perhaps the authors meant to further restrict the sort of health studies they were discussing?

3. P.3. approx line 22. “this cohort” has not yet been defined

4. P.7 Was the Wilcoxon test one-sided because you used a chi-square test statistic? While the hypothesis test is – strictly speaking – one sided for a chi-squared test statistic, since the chi-square with k degrees of freedom is the sum of k-many squared independent standard normal variables, the hypotheses regarding equality of medians is typically two-sided rather than one-sided (in the case of 2 groups, the hypotheses are H0: median1=median2 vs. median1#median2).

5. Table 1 –label units for “time spent inside home”

6. Table S1 not mentioned in main text, but is very important for interpreting the results regarding sensitivity to geocoding method.

7. Tables 1&2, Figure S2, Table S2 – clarify what geocoding method is used here

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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

I declare that I have no competing interests