Author's response to reviews

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

Authors:

Doug Brugge (dbrugge@aol.com)
Kevin Lane (k.lanejr@gmail.com)
Luz T Padró-Martínez (lpadro@gatech.edu)
Andrea Stewart (Andrea.Steward@tufts.edu)
Kyle Hoesterey (kyle.hoesterey@tufts.edu)
David Weiss (david.weiss@tufts.edu)
Ding Ding Wang (deenawang@gmail.com)
Jonathan I Levy (jonlevy@bu.edu)
Allison P Patton (allison.patton@tufts.edu)
Wig Zamore (wigzamore@gmail.com)
Mkaya Mwamburi (mkaya.mwamburi@tufts.edu)

Version: 3 Date: 22 July 2013

Author's response to reviews: see over
Dear Editors-in-Chief:

Please see the attached revised manuscript and comments to reviewers for the study titled “Highway proximity associated with cardiovascular disease risk: the influence of individual-level confounders and exposure misclassification.” (MS: 1369630084906517). We have revised this manuscript based on reviewer comments and have provided detailed responses to each comment.

Thank you for your time and consideration.
Sincerely,
Doug Brugge

Edits and responses to reviewer comments have been colored red and formatted into Arial 12point font.
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.

Thank you. We have addressed remaining concerns below.

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.

Supplemental table 4 contains the results from the fully adjusted model. We have tried to make this clearer where that table is referenced in the manuscript.

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.

Thank you for catching this. We have clarified that this was for biomarkers.

- 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).

We have redefined confounding to not be a main focus of the paper in response to the strongly stated opinion of the other reviewer. But we agree that even so the variables in the model are of likely interest to readers and are not easy to find since they are only in a footnote to the table. Therefore, we have added them into the text as well, as suggested.

- 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?

We have added:
"Distance to highway was examined as a continuous linear variable in adjusted models and while not significant had an inverse relationship with LN IL-6 and Ln hsCRP (data not present here)."

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? Please see our response to the other reviewer.

-Page 7: For the UFP measurements, was some method used to adjust for temporal trends?
We report the UFP data in much greater detail for Somerville in another paper (reference #18). The data presented here were not adjusted for season or other temporal considerations, but were collected across times of day, days of the week, and all four seasons over a full year. We have now explicitly stated the factors considered in the manuscript as part of this resubmission. In the discussion section the last paragraph on page 14 has now been edited to include these factors and we have made adjustments to improve clarity and accuracy based on our own rereading of this paragraph of the manuscript.

“UFP decay patterns were similar to the relationship between hsCRP and IL-6 using categorical distance to highway. The biomarker associations we found for distance from the highway were relatively flat across distance categories, except for the 50-150 m category for hsCRP. Associations of hsCRP and IL-6 with distance were lower on the west side of the highway (Figure 2), where UFP concentrations were lower and gradients were less pronounced (Figure 3). UFP gradients in both neighborhoods were steeper east of the highway (usually downwind; right side of Figure 3) than west, perhaps due to busy local roadways and wind direction. In a detailed analysis reported elsewhere, this UFP difference between west (upwind) and east (downwind) highway sides held for analysis by categories including season, time of day, day of week, wind speed and wind direction [18]. These factors may account in part for the substantial differences in distance associations for hsCRP and IL-6 between Somerville and Dorchester. In particular local street traffic may contribute to UFP exposures especially in the urban background area in Dorchester where participants resided much closer to a major roadway.”

-Page 8: Which potential effect modifiers were explored?
We ran a standard check for effect modification within each finalized model. This means that every variable was individually checked for effect modification with the corresponding model covariates.

The following text has been edited to explain this for readers.
“Effect modification was explored post-stepwise selection as part of the multivariate model building process for all included variables and did not yield any significant interactions.”

-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.
We have clarified that the fully adjusted model included the exposure variables.

-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.
Stratification by age was already in Supplemental Figure 1 (now listed as Supplemental Figure 2). We have added stratification by sex and diabetes status to the manuscript as Supplemental Figure 3.

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

Discretionary Revisions
-First sentence of abstract: I recommend changing this to read “Elevated cardiovascular disease risk has been reported…”
We have made this change.

-Conclusion of abstract: I recommend indicating which specific “rarely included” confounders were important.
In accordance with removing confounding as a main emphasis of the paper, we have removed this sentence.

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

-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.
We have added:

“Another possibility is that we had a vulnerable population with high prevalence of obesity and diabetes relative to the comparison study.”
Reviewer's report
Title: Highway proximity associated with cardiovascular disease risk: the influence of individual-level confounders and exposure misclassification.
Version: 2 Date: 14 May 2013
Reviewer: Sandrah Eckel
Reviewer's report:
Thank you for your responses to my initial comments, your revisions, and the opportunity to re-read this manuscript. Note that the updated version of the paper I received did not have changes highlighted — which limited my ability to detect changes from the previous version.

We are not sure what happened to the track changes we carefully put into a single color and meant to have available to you. It appears that the PDF accepted them so they were no longer visible. We mark our revisions this time using colored text rather than track changes.

I still have several major issues which remain from my initial review.
Major comments:
1. Distance variable:
The authors have rich continuous distance data, but coarsen to a categorical variable. Due to the lack of data between 450-1000m, I agree that it is sensible to perform primary analyses with their proposed categorical grouping. However, when trying to identify the “exposure-response” relationship between distance and their CVD outcomes, it would be much more interesting to subset the data to those participants with distance <450m (near highway group, approximately N=207) and fit a generalized additive model (GAM) which allows for a smooth effect of the continuous distance variable (GAM does not require the effect to be linear). GAM will avoid the “problems with maintaining sufficient sample size” in distance categories. Given the “steep pollution gradients within 200m of a highway”, one would expect to see an interesting shape of the exposure-response relationship in the 0 to 450m range of exposures. The authors have gone through a lot of trouble to reduce exposure misclassification by using refined geocoding methods and they are throwing away information by coarsening the continuous distance variable. The authors might also see the most interesting impacts of the “rarely included” confounders on the distance effect within the steep exposure gradient very close to the freeway.

We appreciate and understand the importance and value of the reviewers’ comments on using continuous data and examining the shape of the exposure-response function, which could provide insight beyond what is available from categorical variables. To address this comment, we have conducted and fit generalized additive models (GAM) for LN hsCRP and LN IL-6, restricted to individuals within 450 meters of the highway, and we have added these results to Figure S1 and to the manuscript.

Text added was:

“Adjusted GAM models for the relationship between LN IL-6 and LN hsCRP and distance to
However, we retain our primary emphasis on the categorical distance variables in the manuscript. We do this for a few reasons. First, in comparing the GAM results to the categorical results, the overall conclusions do not change. That is, there were non-monotonic patterns with distance that potentially reflected the influence of local roadways and residual confounding, but evidence of elevated IL-6 within 150 meters of the highway. Discussing the shape of the GAMs beyond this level would be potentially misleading given the relatively small sample size and the overarching concerns about distance as an exposure proxy. Second, our discussion regarding exposure misclassification is more easily understood when using a categorical exposure variable. While there are clearly approaches to quantify the effects of exposure misclassification in the context of GAMs, this would unnecessarily complicate what is an already complicated paper (also noted by reviewer 1). Finally, focusing on categorical distance variables allows us to present possible confounding variables in Table 1 in a manner that would be far more challenging with GAM results.

2. Confounding:
I still have two major issues with the author’s emphasis on and treatment of confounding, though I agree with the authors that it is nice to have information on “rarely included” confounders (e.g., time activity and time spent on highways, residential window opening, recent combustion exposures, and air conditioning). Unless the authors can show very compelling evidence of novel/dramatic impacts of adjusting for these “rarely included” confounders, I think the paper would be much stronger if the authors included these potential confounders along with the usual set of confounders (and listed this as a strength of the study in the discussion), but made the improved geocoding methods the main emphasis of the paper. We accept this reframing of the paper and have removed from the abstract, introduction, discussion and conclusion the text that we feel gives the paper an emphasis on confounding, but left in the essential findings from confounding in other places. The deleted text should be visible in in the manuscript.

In this case, the more interesting geocoding results should be moved from the supplement (e.g., misclassification table) and, for example, the figure of unadjusted results (Figure 2) could be cut.
Figure 2 is the PNC data by neighborhood. If the reviewer means removing Figure 3, we are confused. Figure 3 is the only presentation of stratified results. Perhaps the reviewer is suggesting using supplemental Table 4 in place of Table 3?

Issue 1: The authors’ model building procedure is ill-suited to their purpose. I agree with the initial screening of variables to identify potential confounders from a long list of covariates, and forcing inclusion of age, sex and smoking status. I do not agree with using the forward stepwise selection procedure (p of 0.15 as
entry/exit criteria) to decide whether to include/exclude the potential confounders. Potential confounders should be included if they change the effect estimate of interest (not if the potential confounder variable has a significant slope). It is reasonable to implement a quasi-“step-wise” procedure to investigate a long list of potential confounders where one defines sets of potential confounders and includes/excludes each of these sets based on whether the effect estimate of interest changes. For easy interpretation, I prefer a parsimonious model that includes only the essential confounders (usually choose same set for all related outcomes being considered), but I will sometimes include other covariates that are strongly significant if they explain a large proportion of variation in Y and hence reduce the standard error of the estimate of the effect of interest.

We would like to state that beyond the forward stepwise selection procedure (p of 0.15 as entry/exit criteria) to decide whether to include/exclude the potential confounders, we tested the impact of each individual variable on the beta coefficients of the distance variables and if any were changed by at least 15% (a usual convention for defining confounding) we included the variable in the model. This has been included in the manuscript text accordingly – see the revised model building process in the manuscript and added language below:

"We performed an additional manual selection process where variables were retained if they had an impact on the beta coefficients of the distance variables."

“We also fit generalized additive models (GAM) which allowed for a smooth effect of the continuous distance variables and generated corresponding spline plots.”

Issue 2. The authors’ have not addressed how the inclusion of the “rarely included” confounders (in addition to the “usual” confounders) changed the effect estimates of interest when compared to a model that controlled only for usual confounders. For example, Table 2 compares an unadjusted model, a model adjusted for “rarely included”, and a model adjusted for “rarely included” and “usual”. It would be more interesting to see a model with “usual” only than a model with “rarely included” only. Which of the rarely included confounders were most important? Which were least important? Currently, the authors have not convincingly shown “the importance of controlling for confounders rarely included in roadway proximity analyses” and haven’t provided information on whether these variables should be collected in future studies and which should have high priority. Supplemental Table 7 only addresses the association between Y and the “rarely included” covariates, not the confounding effect of the “rarely included” covariates.

See above. We have removed the emphasis on confounding.

Minor issues:
• Manuscript still says: “All statistical tests, with the exception of Wilcoxon, were two-sided” even though that authors had said in their first response to review that “We conducted a two-sided test and have changed one-sided to two-sided in the text.” The text should be changed to: “All hypothesis tests were two-sided.”
We have made this change.

- Table S1 column heading should be “>1000m” not “<1000m”? We have corrected this.

- In Supplemental Table 4, which set of adjustments are used? (“exposure adjusted” or “fully” adjusted)? Clarify in table legend. We have indicated that this table is fully adjusted.