Reviewer's report

Title: Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads

Version: 1 Date: 1 October 2006

Reviewer: Dave Stinchcomb

Reviewer's report:

General

A valuable study of the impact of positional geocoding error on traffic pollution exposure estimates in an urban environment.

-----------------------------------------------------------------------------------------------

Major Compulsory Revisions

-----------------------------------------------------------------------------------------------

Minor Essential Revisions

-----------------------------------------------------------------------------------------------

Abstract
- Methods section is incomplete - should also include description of the street proximity and traffic pollution part.
- Methods section includes results (the median error and percentile errors) - suggest moving to the Results part of the abstract.
- Results: "... bias and error in spatial analysis over short distances" - suggest changing "spatial analysis" to something more specific like traffic pollution exposure estimates. This will help readers understand the "acceptable" levels in the next sentence.

Background
- Middle of first paragraph: "The purpose of geocoding is to ..." - not really what you mean to say, I think. Suggest "Geocoding results are often used to ..."
- End of first paragraph: "... the effect this has on epidemiological studies has received limited attention." This is certainly true, but I think you need to include here reference to Ward et al Epidemiology 2005 (positional accuracy and agricultural proximity/pesticide exposure, your reference 14). You might also include Burra et al "Conceptual and practical issues in the detection of local disease clusters" in The Canadian Geographer, 2002; 42(2): 160-171.
- Page 5, top partial paragraph, last sentence. I think you mean "positional accuracy" rather than "positional error" here.

Methods
- Page 9 "all public schools in Orange County, Florida". Not everyone will know where Orange County, Florida is. Further description is needed (Orlando area). A simple map of the study area would be nice.
- Page 9. Should include specifications of geocoding parameters (spelling sensitivity, minimum match score, etc.) and whether or not any manual geocoding was done.
- Page 10 - parcel geocoding. Were matches primarily in the urban areas? It would be nice to give the urban/rural breakdown for street geocoding, parcel geocoding, and for the 104,865 records that were matched both ways.
- Page 10 - use of parcel centroid. I browsed some satellite photos of the Orlando area to see if the centroid assumption was reasonable. It looks like it is: most homes seem to be in the middle of the lot. You might try this yourself (I used Google Maps satellite view) and, if you arrive at the same conclusions, include it here as further support for the centroid assumption. In some places,
houses tend to be at the front of the lot with larger back yards which could
make the parcel centroid assumption biased towards reduced exposure to
traffic pollution.
- Page 10, bottom. "First, the cumulative distribution functions of the
proximity of children to roads with high traffic counts were compared to
determine ...". It's not clear what the comparison is here (I imagine street
gecoding versus parcel geocoding but you should be explicit). Also, I don't
see any graph like this in the results or any discussion.

Results
- Page 11, first sentence. The results given in Figure 1 (cumulative
distribution of positional errors) are not mentioned in the Methods section.

Discussion and Conclusions
- Need to include some discussion of the impact of using a street geocoding
offset that places the residence at the curb. One could argue that naturally this
will place an artificial number of children closer to the roads and it is simply a
problem with using a bad offset. Need to emphasize that most homes are not
on major roads.

Discretionary Revisions

Abstract
- Conclusions: first sentence. Repetitious from the Background section of the
abstract above. Suggest deleting.

Background
- Page 4, first full paragraph, last sentence. Suggest changing "The effect of
positional accuracy of street geocoding has received ..." to "The effect of
positional accuracy of street geocoding on traffic pollution exposure estimates
has received ...".
- Page 6, first full paragraph, last sentence. Something is amiss with "... the
lower higher exposure threshold ...". Lower or higher?
- Page 8, first full paragraph after numbered paragraphs, third sentence. "Both
thee studies ..." -> "Both these studies ..."?

Methods
- Page 9, last line. Spelling: "geocdoing".
- Page 10, third line. Suggest changing "... for a street segment without
verification if the exact street number exists of not" to "... for a street segment,
without verification whether the exact street number exists of not".

Discussion and Conclusions
- Page 14, second paragraph, third sentence starting, "Figures 3a shows ...".
Should be "Figure 2a shows ...".
- Page 15, top line. Change "off-set" to "offset".
- Page 15, middle. Change "To interpret this result, the algorithm behind
placing the locations needs to be revisited; an address ..." to "The reason for
this can be seen by looking at the algorithm behind placing the locations: an
address ...".

Figures
- Figure 2. There is a font problem with the point symbols in the maps - they
have turned into parentheses on my copy (use embed fonts option under "File
- Save As").

What next?: Accept after minor essential revisions

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

Quality of written English: Acceptable
Statistical review: No

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