Reviewer's report

Title: Does the local food environment around schools affect diet? Longitudinal associations in adolescents attending secondary schools in East London

Version: 4 Date: 6 August 2012

Reviewer: Shannon Zenk

Reviewer's report:

The longitudinal design is a major strength of this paper. I appreciate the removal of the cross-sectional analyses and imputation of the missing variables. Overall, I think the writing should be improved and be more precise in order for the paper to make a contribution to the literature.

Major

I still do not understand why you are using Thompsons and Yellow Pages in 2001 and Yellow Pages only in 2005. It seems you would be better off just using Yellow Pages for both time points since any change over time you observe could be due to changes in data sources rather than changes in environment. Your current approach is a major limitation.

Precisely how items were combined for the healthy and unhealthy diet scores is unclear from me from the text and Table 2. I do not see how you derived maximum scores of 13 and 20.

Please clarify the minimum distance measures. It seems that the minimum distance to an outlet would be the same for a 400 meter buffer and an 800 meter buffer. If there was not an outlet within 400 meters, then the 800 meter buffer could pick up an outlet. But if this was the case then what distance did you use in the analysis for 400 meters? In fact, it looks like they are the same for 400 meters and 800 meters in Table 3. So do you really have a minimum distance and then counts and median distances for 400 and 800 meters? Why are the minimum distances 0 for takeaway in Table 3?

The analysis section should be more precise. Please identify which statistical test you used to test bivariate relationships. I am not clear why you would not adjust for clustering in bivariate analyses between food environment and diet variables. When you indicate that the imputed diet data was included in the GLM, does that mean it was not used in the bivariate analyses? The last paragraph of the analysis section sounds as if you included distance to an outlet by type and counts of outlets by type all in the same regression model – is that correct? If so, what were the correlations between these measures and did you assess for multicollinearity? If not, then please describe the models with more precision. Did you use any statistical tests to compare to the 2001 and 2005 food environment measures? Same question for change in diet indicators over time.
If you conducted the longitudinal analysis based on 757 participants (with imputed data from 2005), then I do not understand why you are presenting baseline data for 1382 at baseline or the complete case sample of 524 in 2005. Are you trying to show that the 1382 in 2001 were similar to the 757 imputed in 2005? Or that the 757 imputed were similar to the 524 complete case in 2005? This can be described, with statistical tests as appropriate, in the text. It is confusing to report this in the table since it sounds like you are only using 757 imputed in this analysis.

How you are interpreting the study findings is not clear in the discussion as written. It looks like you found a positive association between distance to a grocer and healthy diet; this is the opposite of what I would have hypothesized. Your findings for takeaway seem consistent with what we would expect. Your explanations for these and other findings do not come through very clearly in the discussion. Your grocery/convenience/supermarket measures makes these results even more difficult to interpret.

Minor

There are some typos (e.g., missing and grammatical errors throughout.

The two new paragraphs on page 5 do not seem well integrated with the rest of the background section particularly the one beginning “When classifying the residential food environment…”

Page 7, “FSM status” – what is that? You start using FSM before you define it.

I still do not understand why you are using postcodes to geocode rather than the street address. You indicate that 15 street address locations typically belong to a postcode. But how large of an area can a postcode be, particularly in your study area? Is this more accurate than if you address-matched to a street network? Again, I recognize that I may not understand the U.K. data. But it seems there could be quite a bit of positional error depending on how large postcodes are in the study area.

The new paragraph/information included in the analysis is not well integrated.

I assume Table 3 is at the school level, that is, based on the 30 schools and not the individual participants?

Discretionary

I think the first couple of paragraphs could be improved to make it stronger and more interesting. The methodological focus (beyond cross-sectional) does not seem to serve the study well.

It seems that the text in the new paragraph on page 6 could be incorporated into the prior paragraph – on same topic of longitudinal studies.
The organization of the first part of the methods section does not seem logical to me. It seems strange to split the information on the sample, with a description of the diet measure in between. I think this section should be rearranged a bit.

Some organizational issues in food environment methods section. It seems the new text justifying outlet classification should immediately follow the identification of how outlets were classified.

The discussion is relatively long.

The point that length of exposure is unknown (new paragraph, page 6) in cross-sectional studies does not seem to be the main limitation of cross-sectional studies.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.