Reviewer's report

Title: Food Environment and Fruit and Vegetable Intake in a Urban Population: a Multilevel Analysis

Version: 2

Date: 23 February 2015

Reviewer: Yan Kestens

Reviewer's report:

Review BMC Public Health

Overall assessment: This manuscript presents interesting data on food consumption and food environments from a Brazilian city. There is a lack of studies on food environments from non US/UK countries, where most of the work is currently being done. In that sense, this study is much welcomed. However, my opinion is that the current state of the manuscript does not make it publishable in its present form. The text is difficult to read, and holds many imprecisions. I raise a series of methodological concerns below. Overall, I feel the data is interesting and needs to be reported. I suggest major revisions to make the text more coherent and focused. I would also suggest having an anglophone person edit the text for clarity purposes.

Major Compulsory Revisions

1. Abstract: The terms ‘probably’ and ‘potentially’ used in the results are a little bit strange… would suggest changing this (see below further elaboration about this point).

2. Introduction: A better job could be done to present the topic, review of literature is only partial in my eyes, and the framing of the problem could be improved. Particularly, other aspects than food environments are being tested in the models (HVI socio-economic index vs. Education and Income) and this question comes somehow as a surprise in the methods and results section (i.e. is not well introduced).

3. Methods:
   The question of areas that were used is unclear to me. You mention the health units (n=148 if I understand well?) but you also refer to Census Tracts (lines 149-150). Clarify. It would be good to explain the level of aggregation of the data used: for example, are businesses coded at address level? Is the Health Vulnerability Index directly available at Health Unit level or does it result from an aggregation of Census Tract data?

4. I found the description of the construction of the dependent variable very unclear. I do not understand how the original collected data became a FV consumption variable with a 1-12 scale range. What do these values mean exactly? What would be a score for somebody who would follow the Brazilian
dietary guidelines for example?

Regarding the food businesses, please explain how the retained classification was obtained. Are the mentioned categories directly available in the database? How are the itinerant food vendors georeferenced? Do all itinerant food vendors sell fruit and vegetables?

Explain that you used the ‘total number of businesses selling fruit and vegetables’ only, not the sub-categories. Along that note, did you test for significance of specific food store categories? If yes, what were the results (i.e. are some type of businesses more significantly associated with FV consumption than others?)

Finally, you seem to use a measure of ‘number of stores’ within an area. I suppose areas differ in size. Why did you not use a measure of density, which would ‘correct’ for differences in area size. Please comment (or use density!).

8. Results:

This section could be condensed and focus on the main findings of interest. Table three seems to point to two different store variables:

Number of supermarkets, hypermarkets and stores and open-air markets specializing in selling FLV in AA

and

Number of stores and open air markets specialising in selling of FLV in AA. Are these two variables different? If yes, why?

9. Overall, reference to geographical unit needs to be simplified (covered areas, health units, AA, etc.) Just mention health areas and explain how it is constructed once, and then refer to a simple term.

10. Discussion:

The explanation of what a score of 6 means is given in the discussion. It would be good to have more details on how the score works in the methods section. Particularly, it sounds as if one could obtain a similar score with quite different food consumption patterns (i.e. few but regularly vs. a lot but occasionnally) - the FV consumption measure needs to be much better explained.

11. I am not convinced by the way results re FV stores are presented in the discussion. Particularly, when using the term ‘probably’ when refering to possible causation: ‘Environment variables such as the number specialized stores selling fruits and vegetables in the neighborhood area probably increase consumption’. In this study, which is cross-sectional, the number of FV stores is associated with consumption. This is a correlation, which can not be interpreted as causation to my eyes. Furthermore, using the term ‘probably’ does not mean much. Rather stay at the correlation level in the discussion (this was associated with this). Same problem when you use the term ‘potentially’ (social variables ‘potentially’ decreased intake (l316). Other problem here is that it is not the variable itself that has the ‘decreasing’ effect.
12. You use both an overall socio-economic index (HVI) (Model 1) or the more ‘raw’ measures of Education and Income (model 2). This ‘comparison’, if relevant, should be better framed in this paper. Why do you compare both and what are the conclusions regarding this distinction?

13. You refer to other studies done in the US showing that distance was less important than ‘having’ a resource in the area. Did you test distance measures, and if not, can you explain why?

Minor Revisions

14. Minor - Wording:
Some sentences are too vague and need rewriting, for example:
‘factors associated with fruit and vegetable intake transcend issues at an individual level’
(not very clear what is meant here)
The same for sentence starting line 87 According…
etc.
Overall the text needs rewriting as some sentences are unclear. It would probably be good to have an anglophone helping with the writing
Example: l148: Georeferenced information regarding the covered areas by the basic health units was also obtained requested from the Municipal Health Departament…

16. Minor - other:
Use of acronyms. Several unexplained acronyms
l113: put MG in full letters
l138: explain CEP
IBGE
Some of the references don’t match between text and bibliography (example: ref 28 in text says ‘Mendes’ but it is 27 in the bibliography (etc.)
Formulation: Again, the way you describe the geographical areas is a little strange. I would recommend using the notion of ‘health areas’ (i.e. replace for example ‘within the covered areas by the basic health units’ with ‘within health areas’).

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

Quality of written English: Needs some language corrections before being published

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

I declare that I have no competing interests.