Reviewer's report

Title: Planning dietary improvements without additional costs for low-income individuals in Brazil: linear programming optimization as a tool for public policy in nutrition and health

Version: 0 Date: 21 Jan 2019

Reviewer: Geraldine Lo Siou

Reviewer's report:

This is an interesting paper about using linear programming optimization to find realistic and nutritious dietary patterns for low income Brazilian individuals without increasing diet cost. However, the reviewer encountered several typos and was confused with the design of the study and with the methods (i.e., seasonality?). In details, the reviewer has the following comments:

Lines 81-83: If the selection of census tracts was done at the first stage and the selection of households was done at the second stage, the reviewer is confused about the wording of "census tracts (n=1,280) were grouped into 550 household strata". Shouldn't it be something like "x households were grouped into y census tracts"? Same question with the wording "the number of tracts in each stratum": shouldn't it be something like "the number of households in each census tracts"? Throughout the paper, the reviewer was confused with the study design. In addition, using similar words like "household strata", "sampling strata", "macro-strata", "macro-stratum", added confusion as to what the authors were referring to. Thus the reviewer suggests adding consistency in the choice of words and adding a flowchart for clarity. Lastly, it is only after reading the discussion that the reviewer was clearly informed that food consumption (from NDS) and price (from HBS) information was obtained from the same household during the same week of collection (see Line 360-361). This information is important and should have been clearly mentioned in the materials and methods section.

Lines 83-85: As dietary intake was collected from two non-consecutive food records, did the authors use both food records? Did everyone in the study complete both food records (e.g., lost-to-follow-up)? If not, what were the response rates? Also, please add the number of pregnant and breastfeeding women who got excluded from the study. Lastly, the reviewer looked at the reference paper (#21) to get details on the data collection procedures and doesn't think this is the best reference to use. Indeed, reference #21 is about the use of the NCI (National Cancer Institute) method to assess usual dietary intake using data from the Household Budget Survey (HBS). In reference #21, there is only a brief mention of HBS with no details on the data collection procedure.
Line 87: Did the authors used both food records for the determination of the mean observed diet?

Lines 88-90: Please explain what "sampling strata" refers to? Again, the reviewer is confused about the study design, please add a flowchart for clarity.

Lines 92-94: As the study used a two-stage sampling process, with the selection of census tracts being the first stage and the selection of households being the second stage, the reviewer is not sure where the "26 Brazilian states and one Federal District, totaling 89 aggregated strata" comes from. Again, please add a flowchart for clarity. In addition, please confirm that the 89 aggregated strata was obtained from a combination of the 26 Brazilian states, one Federal District and the four higher geographic areas.

Lines 104-108: Please add details regarding how the authors came up with 21 to 74 food items from a total of 305 food items. Please add the list of food items which were excluded (see Lines 105-107). Lastly, please write "fruits and vegetables (FV)" instead of just the acronym "FV".

Lines 119-120: Regarding seasonality, please could you comment as to how seasonality was dealt with in the analyses?

Lines 138-139: Please remove the "m" placed between "macro-stratum" and "we ran".

Lines 159-161: Please add "be" between "was constrained to not" and "higher than the content".

Table 1: Table 1 presents a list of nutritional constraints. Should the authors have selected a shorter list of nutritional constraints (i.e., the more important/relevant nutritional constraints) at step 1, how would that decision affect the results?
Lines 187-188: It seems both food records were used in the analyses (see "partially removes daily variance in the dietary intake approaching usual intakes"). To get usual intake, did the authors only use the mean intake of the two food records? Or did they used the NCI method (as mentioned in reference #21)? Please explain the rationale behind the decision of using either the mean intake or the NCI method to estimate usual intake.

Lines 190-191: Just to confirm that the reviewer understands correctly: were "rigorous" corresponding to the 70th percentile, "moderate" corresponding to the 80th percentile, and "flexible" corresponding to the 90th percentile?

Lines 200-201: Where do the "five" macro-regions come from?

Lines 231-232: Since 534 optimization models were performed, how did the authors present the results over the 89 macro-strata in tables 3 and 4? Were they the means and standard deviations over the 89 macro-strata?

Line 236: Why not presenting both graphs for the cost-free models in addition to the cost-constrained models?

Table 2: Is there a reason why the authors chose to present means, 5th and 95th percentiles? Usually, means and standard deviations are presented together. In the case of non-normally distributed variables, usually the medians and interquartile ranges are presented. Also, please place in footnote the letter "a" next to "mean (p5; p95)".

Lines 298-300: As the authors concluded it was not mathematically possible to meet simultaneously all the nutrient recommendations in the optimized diets, how about starting (step 1) with a shorter list of nutrient recommendations (i.e., the more important/relevant nutrient recommendations)?

Lines 306-309: Please change "an elasticity study" into "a study using regression analysis to estimate elasticity coefficients".
Lines 310-311: Please add "of" between "regardless" and "cost".

Lines 354-356: The reviewer does not understand how the seasonality was dealt with in the linear programming optimization? The equations, as written in the paper, do not take the seasonality into account. For example, would the cost-constrained and cost-free models be different depending on the season?

Lines 358-361: This is an important information and should have been clearly mentioned in the materials and methods section. Indeed, throughout the paper, it was not clear to the reviewer whether the food consumption (from NDS) and price (from HBS) information was obtained from the same household during the same week of collection.

Discussion section: Food information was collected from 2 non-consecutive self-reported food records. Therefore, there exists challenges related to mis-reporting or other types of measurement error. Please add this topic into the discussion section.

Lines 384 & 387: There is a typo, it should be "Dietary" instead of "Diatery" or "Diatary".

Line 389: There is a typo, it should be "Organization" instead of "Organizatio".

Line 390: There is a typo, it should be "States" instead of "State".

Table 3: There are typos on the last row of Table 3: "15 16 17 18".

Figures 1 & 2: What is the rationale for not presenting the same figures for cost free models? The reviewer suggests adding such figures in the paper or as supplementary materials.
Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal