Author’s response to reviews

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

Authors:

Eliseu Verly-Jr (eliseujunior@gmail.com)
Rosely Sichieri (rosely.sichieri@gmail.com)
Nicole Darmon (nicole.darmon@inra.fr)
Matthieu Maillot (matthieu.maillot@ms-nutrition.com)
Flavia Sarti (flamori@usp.br)

Version: 1 Date: 22 Mar 2019

Author’s response to reviews:

Reviewer #1: 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:

Thanks for the valuable reading of the manuscript, we are sorry for sending a version with such number of typos, and really appreciate your corrections and 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.

Answer: The reviewer is correct about the confusing terminologies. It was modified in the manuscript. We used “sampling strata” as synonymous to “household strata”, but now they will be referred as just as “strata”. Indeed, the 1280 census tracts sampled were grouped into 550 sampling strata, thus there are a number of census tracts in each sampling strata. This procedure was done during the sampling process by the Brazilian Institute of Geography and Statistics, and everyone who analysis this data set should consider these strata as they were defined. However, considering the high variance in the dietary intakes, we aggregated these strata into 89 larger strata (now called geographic strata) so that we could obtain a more precise estimative of the dietary intakes.

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.

Answer: Yes, we used both one day (when the participant filled only one food record) and two days of food records (when filled two days; in this survey: ~97% of the sample). The reference #21 was replaced by an article based on both HBS and NDS data.
Line 87: Did the authors use both food records for the determination of the mean observed diet?

Answer: Yes, we obtained the mean food intakes for those who filled two food records, available for ~97% of the sample.

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.

Answer: We opted for replacing this term by “strata”, that is a more common used terminology to express the stratification procedure in the complex sample design. A flowchart was included.

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.

Answer: Thanks for this suggestion, a flowchart was included. Yes, these 89 geographic strata came from the four higher geographic areas from all those 26 states and 1 federal district.

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".

Answer: The number of food items reported in the survey was 305; we then clustered them into a lower number of food items (110) once the original list comprised variations of the same food item (for ex.: chicken leg and chicken wings were clustered into chicken). However, the food list varied according to the geographic stratum, from 37 to 92 food items reported (that is, in some areas the individuals reported a higher number of foods than other).
Lines 119-120: Regarding seasonality, please could you comment as to how seasonality was dealt with in the analyses?

Answer: The data collection was performed throughout a 12-month period; thus, it comprised the variations in food intakes due to seasonality. The seasonality in the food prices was addressed by deflating the food prices collected throughout the 12-month period to a reference period (January, 2009).

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

Answer: Thanks.

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

Answer: Thanks.

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?

Answer: Depending on the nutrient. For example, if fiber were removed maybe a lower fruits and vegetables content would be obtained (we did not test it). We avoided making judgment on which nutrients are more relevant to be included in the models, but the list of nutrients was limited to those with an Estimated Average Requirement (EAR) established; nutrients with only an Adequate Intake (AI) were not considered due to the lack of evidence to define their intake recommendations. Considerations were made concerning vitamin D and E in the discussion section.
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.

Answer: Thanks for this comment. The rationale behind this procedure is that the food quantities in the optimized diet should not be higher than what is usually consumed in the high-consumer strata (i.e., 70th, 80th, and 90th percentiles) from a distribution of strata. The mean usual food intakes in each stratum are obtained by averaging the short-term intake (i.e., 1- or 2-day food record) from all individuals in the stratum. However, the small number of individuals within some strata may not be enough to provide a precise estimative of the usual food consumption, that is why it is stated in the manuscript that it “partially removes daily variance approaching usual intakes”. To avoid misunderstanding, we removed this sentence.

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?

Answer: Yes; and we made it clearer in this version.

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

Answer: The five macro geographic regions: North, Northeast, Southeast, South, Center-west.

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?

Answer: These 534 models are split into: 89 cost-constrained models*3 levels of acceptability, and 89 cost-free models*3 levels of acceptability (89*6=534). The results are presented as the mean and standard-deviation over the 89 geographic-strata for each set of models.
Line 236: Why not presenting both graphs for the cost-free models in addition to the cost-constrained models?

Answer: Thanks. We opted to not present because the main question of the study concerns the modifications in food quantity and nutritional improvement without cost increase. The results referring to the cost-free models were presented in tables only to assess the effect of the budget on the food selection and nutritional contents.

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)".

Answer: Thanks for this suggestion, we changed the summary statistics.

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)?

Answer: This question was addressed in a similar question above. Moreover, it is necessary to perform the step 1 with all constraints to be able to identify the limiting nutrients in order to proceed with the step 2.

Lines 306-309: Please change "an elasticity study" into "a study using regression analysis to estimate elasticity coefficients".

Answer: Thanks.

Lines 310-311: Please add "of" between "regardless" and "cost".

Answer: Thanks.
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?

Answer: This question was partially addressed in a similar question above. Probably the models would return different results depending on the period of the year. However, it is difficult to link this variation to the seasonality. In Brazil, the seasons are not well defined, the weather varies across the country, and the local of consumption is not necessarily the local of the production. In this sense, we believe that showing the long term 1-y period provides a good picture of the main changes that could be effective to improve the diet quality.

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.

Answer: In fact, as it can be seen in the flowchart, we opted to use the price information from all households and not only those with dietary data. We argue that this procedure increases the precision of the price estimation in each geographic-stratum.

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.

Answer: Thanks, we included a sentence addressing this limitation.

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".

Answer: Thanks, we corrected all these typos.
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.

Answer: Thanks. We opted to not present because the main question of the study concerns the modifications in food quantity and nutritional improvement without cost increase.

Reviewer #2: This paper describes the use of linear programming optimization to model the dietary changes needed among low-income individuals in Brazil to enable nutrient requirements to be met, and at minimal cost. It is an important question to address but in its present form the paper is quite challenging for the non-specialist reader to understand - and in some places there is a need for additional detail. My comments are set out below.

Answer: Thanks for your careful reading, comments and suggestions.

General comments:

1. The term 'macro-stratum' is not meaningful without reading the Methods section (eg Abstract). As these strata represent groups defined by geography (state and within-state areas) could the term be changed? This would avoid confusion with 'household' and 'sampling' strata also described in the paper.

Answer: Thanks for this suggestion; this comment is in line with the other reviewer’s comment. We tried to make this description clearer in the manuscript. Also, we replaced the term “macro-strata” by “geographic-strata”, for which we believe be easier the get its meaning when reading the manuscript.

2. The overlap between HBS and NDS needs clear explanation. Was NDS a sub-sample of HBS and if so, are the dietary data analysed for this paper restricted only to individuals in NDS? This is difficult to understand as dietary data were available for individuals aged 10 years or older (line 84) but Table 2 shows 24% below 10 years. Line 97 suggests that the linear optimization models are based on dietary data from 18,139 individuals - the number of people studied needs to be explicit throughout, stated in the Abstract, and represented in the Tables. It would be clearer if Table 2 only represented the individuals with modelled dietary data.

Answer: Thanks, we modified the abstract and the tables.
3. The dietary data are modelled against recommended cut-offs that were averaged for sex/age groups (p8). Is this justified in a group that ranges from children aged 10 to adults over 60 years of age, and includes men and women? With such a large dataset - could subgroups be considered to avoid the need for weighting (eg children and adults)? My concern is that messages regarding the dietary changes needed to meet nutrient requirements might differ between these subgroups.

Answer: We have tested it; our judgment is that the dietary changes to meet nutrient adequacy do not differ importantly between age groups, i.e., the conclusion of the study will be concerning the same food groups.

4. The separation of cost/no-cost models is straightforward for the reader but the different models considered in terms of 'levels of acceptability' are not. The definition of these levels (p11) needs clearer explanation: 10th, 70th, 80th and 90th percentiles of consumption of individual foods were used but the mapping to these levels is not clear. It would be simpler for the reader if the groups were labelled in a way that would indicate their derivation without reference to the Methods.

Answer: Thanks for this comment. We understand your concern but we think that these labels help in the description of the results and in the arguments in the discussion. We added a footnote in the tables and figures so the readers can readily have their meaning.

5. Detail of the dietary data collection should be included in the paper. It was not possible to evaluate the significance of the information presented without this - but I could not find the cited paper either through PubMed or via the journal site (ref 21). Line 84 states that food consumption data were collected from individuals; detail is needed of who they were and/or how selected from the households studied (line 81). Dietary data always have limitations - these need to be considered in the Discussion.

Answer: Thanks, we included a sentence addressing this limitation.

6. The use of individual vs household data needs to be made clear in the paper. An example is Table 2 where number of household members is reported together with per capita income - but it is not clear what the significance of this information is. Were a number of individuals per household studied (ie all had dietary assessments) and if so, is the per capita income their own or given for the head of household? This also relates to point 2. above.

Answer: Thanks, we made it clearer in this version.
7. A key point from these analyses is that it was not possible to model a diet that met all the nutrient thresholds. Although this is commented on in the Discussion, if no model achieves recommended intakes, it questions the use of the RDA thresholds to consider the nutritional value of the Brazilian diet. It needs to be explicit throughout the paper that the stage 2 models were used for the data presented and a clearer message that no model met all thresholds (Abstract), such that the focus is on dietary improvement.

Answer: We made some considerations on the use of the DRIs for the Brazilian population concerning the vitamin D and E recommendations, and did not consider those with no EAR established. We think that the fact of not meeting all nutritional constraint could be due to a potential non-specificity of the DRI to Brazil, but also because the consumption of key nutrient-source foods was low, and it would need a much higher amount than what was defined as the flexible acceptability (90th percentile).

Specific points:

* Line 71 - please amend to clarify meaning
Answer: Thanks, we modified this sentence.

* Line 73/Title - this is confusing as half the models allowed for cost increases - please amend to clarify
Answer: Thanks, we modified this sentence.

* Line 99 - explanation is needed of use of sampling weights.
Answer: This is a usual procedure to take into account different probabilities of being selected from a population-based survey.

* Line 118 - were costs per 100g then multiplied by reported consumption to assess cost per portion consumed?
Answer: Yes, that is what we did, but we modified this sentence to make it clearer.
* Line 127 - 'acceptable range of food contents' needs to be explained
Answer: Thanks, we modified this sentence.

* Line 139 - 'macrostratum m' needs explanation
Answer: Thanks, this term was replaced throughout the manuscript.

* Line 186 - why were mean food intakes derived for 550 strata?
Answer: We included a flowchart, as suggested by the other reviewer, that better explains these numbers.

* Line 200 - 'five macro-regions' needs explanation
Answer: Thanks, we included an explanation.

* Table 2 needs to specify number of individuals studied and would be easier to understand if restricted to individuals with dietary data; years of schooling may be unhelpful in a group that includes children
Answer: Thanks. We limited these results to the household with dietary data.

* Tables 3 and 4 should specify number of individuals studied
Answer: Thanks, we added this information as a footnote to not be confused with the actual number of observations which the results were derived from (the number of observations is 89 geographic-strata).
* Figure 1 - it is difficult to map the data from Table 3 to the figure - for example the cost-constrained flexible model shows mean vitamin C intakes increasing from 44.5mg (observed) to 105.8mg - but the figure shows median increase of about 60%? This suggests that median (IQR) data should be presented rather than means in Table 3? Answer: We really appreciate this comment. The figure was based on wrong numbers. In this version we corrected the box-plot and now they are in accordance with results from the table 3. Anyway, the pattern of the figure remained as before, although the magnitude of some differences is more remarkable now.

* To help the reader understand the dietary changes needed to map against the increases in nutrient intakes, it would be helpful to include a figure showing food group intakes that is equivalent to Figure 1

Answer: Thanks. That would be informative. However, as there are many foods with different amount of consumption, the y-axis scale either expressed as difference in grams or in percentage, would be highly influenced by high values and the smaller differences would be difficult to see in the figure. Thus, the authors’ judgment is that, in this case, the table is the most appropriate way to present these results.

* Discussion - it is a limitation that these food intake data were collected 10 years ago; comment is needed on the relevance of the findings now and implications for costs given in the paper.

Answer: Thanks. We addressed this point in the discussion section.