Reviewer’s report

Title: Association between empirically derived dietary patterns with blood lipids, fasting blood glucose and blood pressure in adults - The India Migration Study

Version: 0 Date: 03 Sep 2017

Reviewer: Michelle Castro

Reviewer's report:

General comments: The manuscript addresses an interesting topic in the field of Nutrition Epidemiology that is the association between empirically derived dietary patterns and cardiometabolic risk factors in a sample of migrants and non-migrants adults from India.

However, some corrections must be made in order to improve the manuscript quality, starting by the title. It should be more specific by informing the readers that the dietary patterns are empirically derived and mentioning the cardiometabolic risk factors evaluated i.e., blood lipids and fasting blood glucose. So, the title could be: "Association between empirically derived dietary patterns with blood lipids and fasting blood glucose in adults - The India Migration Study".

Moreover, as the authors have hypertension data I suggest include in the text how blood pressure was measured and how hypertension was determined and also include this risk factor in the analysis of the association with dietary patterns.

Introduction, Page 4, lines 72-74 - Please, include a reference for the statement about the role of the diet as a risk factor for CVD and type 2 diabetes mellitus. Moreover, include the role of diet as a risk factor for CVD by affecting blood pressure, the most important risk factor for CVD.

Introduction, Page 4, lines 76-84 - Please, include a definition of the dietary patterns, explain the different types of dietary patterns (e.g., data-driven, hypothesis driven) and comment the advantage of analyzing dietary patterns in Nutrition Epidemiology studies.

Introduction, Page 4, lines 81-84 - Please, describe with more details the prevalence of CVD and type 2 diabetes in population of India, as the authors stated that the prevalence of these diseases are "increasing" in the country.

Introduction, Page 4, lines 86-88 - In objective, does not mention the main study where the data were gathered from.

Methods, Page 5, lines 104-105 - It is not clear why the sample included 7067 participants if 7102 individuals agreed to complete the clinical examination with their siblings. Please, explain
and include this information in the text. Moreover, please review all the description of the sample size.

Methods, Page 6, Dietary Assessment - Did the authors investigate implausible energy intakes? If so, please, include this information and the procedures applied in the text.

Moreover, how many missing responses in FFQ and what were done with missing data criterion or methods for imputation or exclusion? Please, explain and include this information in the text.

Methods, Page 7, line 144 - Please, include the accuracy of the digital scale used.

Methods, Page 7, lines 148-150 - Please, include the amount of blood sample collected per individual, who collected the samples, where the samples were collected, and the time and speed of centrifugation (e.g., RPM).

If I understood correctly, the centrifuged blood samples were stored at 20ºC (room temperature) for a month until transportation to AIIMS. If I am correct, what is the effect of storing centrifuged blood samples at room temperature for a month?

Methods, Page 7, lines 154 - The Friedewald equation was used to estimate the LDL-C. However, the estimates of LDL-C from this equation are not reliable when individuals have triacylglycerol (TG) > 400 mg/dL. Please, include in the text how many individuals in the sample had TG >400 mg/dL and what was done in this case for estimating LDL-C.

Methods, Page 8, Statistical analysis - What information was used as the input variable for PCA (e.g., daily frequency or grams of food consumed per day)? Please, explain and include this information in the text.

Methods, Page 8, Line 162 - How do the authors analyzed the scree plot for defining the number of components to retain? Moreover, what it means "component interpretability"? I understand that it is based on the factor loadings of each food group on each component extracted. However, there is no mention about the factor loading cut-off considered for dietary patterns interpretability. Looking at the footnote of the Supplemental Table 1, the factor loadings < 0.15 were omitted for simplicity, so, it seems that this value was the cut-off applied for component interpretability. If the cut-off was settled at 0.15, this value is too low, considering that the majority of studies that derive dietary patterns by factor analysis or principal component analysis use cut-offs ≥ 0.20.

Please, explain how the screeplot was analyzed, include the information in the text about the factor loading cut-off used for component interpretability and the reasons for choosing the value
and correct the information presented in the footnote because the value 0.15 was considered in module (e.g. $<|0.15|$).

Methods, Page 8, Lines 169-173 - The authors excluded from the analysis individuals diagnosed with diabetes because of the change in fasting glucose levels and also in dietary habits. However, why do the authors did not excluded individuals diagnosed with dyslipidemia?, considering the possible changes in blood lipids owing to medication use and modifications in dietary intake? Please, justify and include this information in the text.

Methods, Page 8, Line 174 - Please, correct the "chi-square test" for "Chi-square test" (with capital letter).

Methods, Pages 8-9, Statistical Analyses - What was the statistical software used to perform the analysis? What was the significance level adopted by the authors? Please, include this information in the text.

Results, Page 10, Line 213 - Please, verify if is "quintiles" or "tertiles"

Discussion, Page 13, Lines 277-278 - Please, replace "cardiometabolic outcomes" by "cardiometabolic risk factors", considering the associations of animal food pattern with blood lipids and fasting plasma glucose.

Discussion, Pages 13-14, Lines 276-293 - Please, include the biological effects of consuming animal foods in the increase of total cholesterol, LDL-C, HDL-C, TG and blood glucose. What are the biochemical evidences that could explain the results observed by the authors?

Conclusion, Page 343-346, Page 16: I disagree with the statement that the "high intake of the animal food dietary pattern was positively associated with cardiometabolic risk factors in India". As the authors commented in lines 324-325, the average consumption of meats by Indians is low (less than 50 g/day). So, I think that the conclusion should be re-written stating that "High adherence to the animal dietary pattern was positively associated with cardiometabolic risk factors in India".

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.