Author’s response to reviews

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

Authors:

Krithiga Shridhar (g.krithiga@phfi.org)
Ambika Satija (ambikasatija@gmail.com)
Preet Dhillon (preet.dhillon@phfi.org)
Sutapa Agrawal (sutapa.agrawal@phfi.org)
Ruby Gupta (ruby.gupta@phfi.org)
Liza Bowen (lizabowen@gmail.com)
Sanjay Kinra (sanjay.kinra@lshtm.ac.uk)
Bharathi AV (bharathi.avb@gmail.com)
Prabhakaran Dorairaj (dprabhakaran@phfi.org)
Srinath Reddy (ksrinath.reddy@phfi.org)
Shah Ebrahim (shah.ebrahim@lshtm.ac.uk)

Version: 1 Date: 17 Nov 2017

Author’s response to reviews:

Reviewer reports:

Reviewer #1: General comments: The manuscript addresses an interesting topic in the field of Nutrition Epidemiology that is the association between empirically derived dietary patterns and cardio-metabolic risk factors in a sample of migrants and non-migrants adults from India.

We thank the reviewers for their positive feedback and valuable suggestions. We have revised the manuscript based on reviewers’ suggestions and all changes in the manuscript are marked in red.
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 cardio-metabolic 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".

Thank you for this valid suggestion. We have corrected the title as suggested. “Association between empirically derived dietary patterns with blood lipids, fasting blood glucose and blood pressure in adults - The India Migration Study” (We have also included systolic and diastolic BP as continuous outcomes in the analyses in addition to blood lipids and fasting glucose as suggested)

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.

Thanks again for this comment. We have revised the methods to include details on blood pressure measurement (lines 170-177; pg.8); also, we have included systolic and diastolic BP as continuous outcomes in the analyses 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 II 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.

Once again many thanks for this comment. We have revised the introduction section of the manuscript as below and the changes in the manuscript are marked in red (lines 75-78; pg.4).

“Diet is a key modifiable determinant of CVD and type-II diabetes, and potentially influences risk of developing these diseases by modulating blood levels of lipids and fasting glucose as well as blood pressure, important proximal risk factors of CVD and type-II diabetes”

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 analysing dietary patterns in Nutrition Epidemiology studies.

We have revised the introduction section appropriately as below to reflect the changes suggested and the changes in the manuscript are marked in red (lines 80-87; pg.4).
“Dietary patterns” describe the overall diet; i.e. the varied combinations of foods and nutrients consumed in totality by individuals. Conventionally, nutritional epidemiological studies evaluated associations of individual food components or nutrients with several health outcomes. ‘Dietary pattern’ studies are relatively a recent approach in nutritional epidemiology studying the cumulative influence of composite diet as a ‘whole’ on different health outcomes. This approach can better account for any residual confounding by other components or nutrients of diet, and also provides useful evidence for more practical and appropriate dietary recommendations that are likely to succeed in real circumstances”.

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.

Thank you for this comment. We have re-framed the statement as below to reflect the percent increase in disability and deaths over a ten year period (2005-2016) based on recent GBD data (lines 95-101; pg.5) http://www.healthdata.org/india.

“Evidence from India regarding the role of different dietary patterns on blood lipids, fasting glucose and blood pressure levels is critical given rising death and disability due to cardiovascular disease (leading at 53% increase in deaths and 41.5% increase in pre-mature deaths recorded between years 2005 and 2016) and diabetes (9% increase in disability between years 2005 and 2016) [16]. Diet and associated high blood pressure, fasting glucose and total cholesterol are cited as the leading risk factors of NCDs [16]. However, studies based on Indian dietary patterns are limited to date.”

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

We have now included the details of study location in the objective (lines 103-107; pg.5).

“We examined the associations of empirically derived indigenous dietary patterns with cardio-metabolic risk factors i.e., blood lipids, fasting blood glucose and blood pressure levels in rural and urban Indians recruited across four geographic regions of the country [Lucknow, North India; Nagpur, Central India; Hyderabad, South-central India and Bangalore, South India] using data from the Indian Migration Study”

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.

Thank you. Of the 7,594 eligible participants identified for the India Migration Study, 7,102 (94%) agreed in principle to participate with their sibling, of whom 3,537 (50%) sib-pairs actually participated by the close of field work. The final IMS sample was 7,067 respondents as seven respondents did not complete the clinical examination and/or some section of questionnaire. We have revised the methods section to reflect these numbers (lines 120-126; pg.6).

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.

Many thanks for this observation. We empirically derived common patterns of foods in this population that were likely to be consumed together by using principal components analysis (reduced rank regression approach). We did not exclude participants based on implausible energy intakes. With a cut-off for implausible energy consumption, less than 5% of participants were estimated to be consuming implausible levels of energy [i.e., at <500kcal/day (N=2) or >5000 kcal/day (N=224)] and so we did not exclude them.

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.

FFQ responses were complete for 7067 participants. There were some food items which had a lot of missing values such as bhagar and kesari bhath and those items were removed from the food list for PCA analyses (lines 158-161; pg.7). Participants who did not complete the clinical examination and/or the FFQ were excluded from the analyses (N=7/7074) (lines 125-126; pg.7). We have mentioned these details in the methods section.

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

Thanks for the comment. We have included the accuracy statement in the methods section. A digital personal scale (Beurer Model PS16, Ulm, Germany) accurate to 0.1 kg and stadiometer accurate to 1 mm (Leicester height measure (Chasmors Ltd London UK) were used by trained personnel to record weight and height respectively of the participants in light indoor clothes without shoes. We have added the details to the text (line 166; pg.7).
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).

10 ml blood sample was collected by trained phlebotomists at the field sites, centrifuged at 2500 rpm for 15 minutes to separate serum or plasma. We have added this information to the text (lines 179-184; pg.8).

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?

After processing (within an hour of collection), samples were stored at -200C in deep freezers at field sites for two to three weeks, after which they were transported to the lab at All India Institute of Medical Sciences, (AIIMS, New Delhi) where they were stored at -800Cin ultra low deep freezers. We have added this to the methods section (lines 179-184; pg.8).

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.

LDL-C was calculated for participants having TG <400 mg/dl. 50 participants had TG levels>400 mg/dL and we have not calculated LDL-C for these participants, treating these as missing values. The details have been included in appropriate sections of the manuscript (lines 188-190; pg.8).

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.

Per day consumption in g was used as the input variable for 30 food items included in the PCA. We have included this information in the manuscript methods (line 202; pg.9).

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|).

Based on eigenvalue limit, there were seven components which met the criteria of eigenvalue >1. Then we used scree plot to retain 3 (out of 7) components. When we visually examined the scree plot, we observed a sharp drop after the first three components, after which the curve became substantially shallower. In addition, when we examined the contributions of individual food groups to the various components, the interpretability of the last four components in terms of what kind of composite dietary patterns they represented was unclear; i.e. they didn’t seem to represent cohesive dietary patterns with clear interpretability. We have included these information in results section (lines 224-232; pg.10).

As the reviewer rightly points out, choice of factor loading cutoff can be a rather arbitrary decision. In a previous publication (Satija A, Hu FB, Bowen L, Bharathi AV, Vaz M, Prabhakaran D, Reddy KS, Ben-Shlomo Y, Davey Smith G, Kinra S, and Ebrahim S, Dietary patterns in India and their association with obesity and central obesity. Public Health Nutr, 2015: 1-11), we examined the sensitivity of our results to choice of cutoff, trying a higher cutoff of 0.3, and found component interpretability was largely the same. Even with the cutoff of 0.2, the interpretation of the three included components remains largely the same.

We also ran a number of other sensitivity analyses in this previous publication (Satija et al 2015) (e.g. we randomly divided the data set into two halves and carried out the same statistical procedures to obtain dietary patterns from each, using both 0.15 and 0.3 as cut-offs; and we restricted the final factor solutions to three, four, five and six factors, again with both cut-offs). In all of these analyses, the first three components were the most consistent and reproducible, and in fact, of the three, the animal-foods pattern was most consistent. Lastly, no significant associations were found between the last four patterns and anthropometric indicators in this dataset (in previous publication).

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.

Thank you for this important observation. Data on the history of diagnosis of dyslipidemia (self-reported, previously doctor-diagnosed) was not available for this population. It could partly be due to the fact that evidence-based dietary modifications and therapeutic interventions for dyslipidemia was not as routinely practised as for diabetes during the study period and thus less likely to affect our associations. However, we have adjusted for regular use of any medication which includes a possible use of statins (if any).

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

Many thanks for your minute observation. ‘Chi-square test’ corrected (line 210; pg.9).

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.

All statistical analyses were conducted using STATA software version 10 (Stata-Corp.2009.Stata Statistical Software: Release 10. StataCorp LP). 5% significance level was adopted. Necessary information have been included in the text at appropriate sections (lines 219-221; pg.10).

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

Thanks for this observation. This statement refers to our earlier publication using the same dataset wherein the dietary patterns were assessed in quintiles.


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.

Yes, we have replaced ‘outcomes’ with ‘risk factors’ (line 336; pg.14).
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?

Several potential mechanisms through which ‘animal food’ patterns could have influenced cardio-metabolic risk factors in this population can be proposed, although further studies are required to confirm the findings. For example, imbalance in saturated and unsaturated fat (high saturated and low unsaturated fat) content and potentially lower fibre content in the ‘animal food’ pattern coupled with high salt content in cooked animal food could have influenced blood lipids and pressure levels respectively. The positive association of ‘fruits-vegetables-sweets-snacks’ with diastolic blood pressure lends some support to this theory as high salt content in snacks could have driven the observed association. A study from India’s third National Family Health Survey (NFHS-3) that showed positive association of fish consumption with diabetes alluded that the pattern of cooking such as deep frying with high amounts of cooking oils could have contributed to the positive findings. We have included these in the discussion as suggested (lines 339-350; pg.14-15).

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 cardio-metabolic 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 cardio-metabolic risk factors in India".

Thank you for this suggestion. Yes, we agree with the reviewer that the quantity of consumption of ‘animal food’ is less in India compared to Western counterparts. The positive association of the ‘animal food’ pattern with cardio-metabolic risk factors could have been partly driven by cooking pattern. We have mentioned it in our discussion and made it clearer in conclusion. ‘High adherence’ could also have been a factor driving this association, however currently there are no data from India to suggest anything about differences in terms of adherence to animal food in Indians compared to the West and our study do not have this information. So, we are unable to comment further on this aspect of diet at this moment but it could certainly be a factor for future research.
Reviewer #2: General:

* The clinical relevance of the study should be better pointed out in the Discussion, and even in the Introduction

We thank you for this suggestion. We have included statements indicating some aspects of clinical significance of the study in the discussion section (lines 399-401; pg.17).

* If the study used the CVD and Diabetes markers, better clarifications could be presented because the use of these markers

These are the well-established intermediate (proximal) markers of CVD and diabetes. We have mentioned the importance of studying these markers for associations with diet in the introduction section (lines 75-78; pg.4).

* A highlight was given to the regions at the beginning of the manuscript, but the results were not presented by region

We have presented region-wise, location-wise (urban, rural and migrants) as well as socio-economic status-wise (low, medium, high) results for ‘animal food’ pattern (the most consistent pattern across the four regions, 3 locations and 3 SES of the country) as supplemental material. The other 2 patterns were less consistent and the numbers (of participants) across tertiles of these 2 patterns in different strata were also small for any meaningful sub-analyses.

Objectives (Abstract and Introduction):

The objective of the study suggests that the study will be done comparing rural and urban regions of India, as well as in four different regions of India, however the principal results are shown only for the total population of the study (In abstract and Introduction).

Thanks for this comment. Our study population included participants who were spread across four different geographic regions (North, Central, South-Central and South) of the country. We have now revised the objective to reflect this idea as well as the principal results of the study. However, we have presented region-wise, location-wise (urban, rural and migrants) as well as socio-economic status-wise (low, medium, high) results for ‘animal food’ pattern (the most consistent pattern across the four regions, 3 locations and 3 SES of the country) as supplemental material. The other 2 patterns were less consistent and the numbers (of participants) across tertiles of these 2 patterns in different strata were also small for any meaningful sub-analyses.
Abstract
"Cross-sectional data from the Indian Migration Study (2005-2007) have been used for this study. Study participants (n=7067, mean age 40.8yrs) included urban migrants and residents from Lucknow, Nagpur, Hyderabad and Bangalore with their rural and urban dwelling siblings respectively"

This sentence is too extended and does not make clear the design of the study.

Thank you. We have made the sentence simpler and clearer as follows: “Study participants included urban migrants, their rural siblings and urban residents, their urban siblings from Lucknow, Nagpur, Hyderabad and Bangalore (n=7067, mean age 40.8yrs)”

Introduction
General comments: It's clear, but simple. A little more information should be included on the various dietary patterns (especially in India), markers used in the analysis and the importance of the study. The objective should be improved, reflecting the presented results.

Many thanks for this valuable suggestion. We have revised the introduction to include information on dietary patterns (Indian dietary patterns are explained in detail in the discussion), markers used in the analysis (lines 75-87; pg.4). We have added a statement regarding clinical importance of the study in the discussion section (lines 399-401; pg.17). We have also re-worded the objective (lines 95-107; pg.5).

Methods
* General comments: It's clear and detailed.

Thank you

* Did variable education evaluate with two categories: illiterate and literate? If yes, it's should be clear in the study.

Education variable was evaluated in 4 categories: 1.no formal education 2.primary school 3. Secondary school 4. Beyond secondary school. We have made this clearer in the text (lines 212-213; pg.9) and tables.
Statistical analysis: In the study was used the principal components analysis? Would be more appropriate to adopt the factorial analysis, being the Principal Components the extraction method?

Thank you for this important comment. PCA and factor analysis are both data reduction techniques. PCA creates a smaller set of components using a linear combination, or a weighted average, of a larger set of variables. Factor analysis, on the other hand, measures an underlying latent variable, which cannot be directly measured with a single variable (e.g., intelligence). The two methods are very similar, but the interpretation of PCA is more applicable to dietary intake data and dietary pattern creation. Moreover, PCA has been very commonly used to create empirical dietary patterns in the literature, and has been shown to have appreciable reliability and validity (ref: Hu, FB. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr 1999;69:243–9)

Results

* Line 187: correct the word "screen plot";

Apologies for the mistake. Corrected (scree plot) (line 198; pg.9)

* Why were the results presented by different regions? Is this relevant for the study? If yes, the principal variables of interest in the study (cholesterol, LDL, Triglycerides, HDL and fasting glucose) should also be presented by region.

Many thanks for this comment. We have included region-wise information of biomarkers in Table 1 of the manuscript (pg. 20).

* Why were dietary patterns not evaluated as a risk factor for diabetes, since there is this information available?

Previously, IMS investigators using latent class analysis (LCA) evaluated the association of diabetes with different dietary patterns. We are adding to that by examining the IMS dietary data in a different way, using PCA and fasting blood glucose as continuous outcome.
Discussion/ conclusion

The strengths and limitations of the study should be better defined and presented.

Thank you. We have re-worded the discussion as suggested (lines 405-410; pg.17).

“The cross-sectional study design and possible measurement error in dietary intake are the major limitations of the study with potential impact on the study results, which should be interpreted with caution. However, the large, diverse and heterogeneous nature of this study population and their diets and consistency of associations across several sensitivity analyses for ‘animal foods’ pattern, are strengths of the study”

Tables
1) Table 1
   a. The explanation for * is not clear

Thank you. We have made it clearer (lines 458-59; pg.20).

   “* p-values for difference in means or proportions are from ANOVA for continuous data and Chi-square test of significance for categorical data (p<0.0001 for all characteristics among the regions)”

   b. Include information for all variables used in adjusted models, for instance: use of regular medication for diabetes, food supplements

We have included adequate information as footnote in all tables.

   “Robust standard error adjusted for age (continuous in years), sex (male/female), migration status (rural, urban, urban migrant), site(Lucknow, Nagpur, Hyderabad, Bangalore), SLI(continuous score), education (no formal education, primary school, secondary school and beyond secondary school), BMI(continuous in kg/m2), total energy(continuous in kcal/day), physical activity(continuous in totalMETS), tobacco(never, past, current), alcohol (never, past, current), use of any regular medication for chronic conditions such as diabetes, hypertension and/or food supplements (yes/no), sib-pair”

2) Table 2 - just a suggestion: include "Ref" in the table, not "0.00 (Ref)"

Thank you for this minute observation. We have corrected the table content as suggested.