PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diet quality indices, genetic risk and risk of cardiovascular disease and mortality: a longitudinal analysis of 77,004 UK Biobank participants
AUTHORS	Livingstone, Katherine; Abbott, Gavin; Bowe, Steven; Ward, Joey; Milte, CM; McNaughton, Sarah

VERSION 1 – REVIEW

REVIEWER	Jimikaye Courtney Pennsylvania State University, United States of America
REVIEW RETURNED	05-Nov-2020

GENERAL COMMENTS	Abstract: In the methods, place the information about the polygenic profile prior to the HR models, and make it clear that genetic profile was included in the model with diet quality - this is necessary for readers to understand the conclusion that associations with diet quality were independent of genetic risk. In the results, prior to listing HRs, indicate what they are for - example: 'one-point higher RFS for all-cause mortality was 0.96 (95% CI: 0.94, 0.98)." Please also indicate that the information in parentheses is the 95% CI. Main manuscript Introduction: The introduction is well-written, concisely states prior research and the need for additional studies, as well as how this study fills those gaps. Methods: Supplemental Tables 1-3 are difficult to read because of a lack of horizontal lines separating the different food groups. For example, supplemental Table 2, it looks as though 'Mixed vegetables' are part of the 'dietary fibre' category. Suggest adding horizontal lines are where the text for the reference row is so that it is clear how rows align. There are also some typos in the tables. Results: Table 2 states in the footer that all associations were p<.001. this makes sense given the extremely large sample size; however, the meaningfulness of this is questionable given that the actual values for the scores in different tertiles do not appear to be meaningfully different. Restate the sentences reporting HRs to say the HR is for prior to reporting it, similar to the suggestion for the abstract: "The adjusted HR associated with a one point higher RFS for all cause mortality was ..., for CVD-related mortality was ..." - this will make it much easier for the reader, as the current sentence structure requires readers to look back at the HR after finding out what the HR is for.

	Given the 'evidence of an association between RFS and all-cause mortality and stroke in females" the reader wonders why the researchers did not run interactions with sex in the models? Suggest not reported potential sex differences based on stratification. Rather, the researchers would be better-suited to run interactions, as different results based on stratification cannot be compared statistically, which causes the reader to question the decision to not run interactions in the models for sex. Regardless of the final decision, the authors should indicate in the statistical analysis section that they conducted additional analyses with interactions by sex (preferred) or stratified by sex (not preferred). This is recommended for all of the diet quality indices examined, particularly given that the stratified analyses suggest that a significant interaction term will exist for sex in many cases. Conclusions: The authors conclusion regarding different pathways underlying 'genetics of both MI and death' for males and females is somewhat inappropriate given the use of stratification rather than interaction by sex. The reviewer strongly recommends running interaction analyses or softening language regarding conclusions about sex differences. Given the ease of running an interaction model, it is a disservice that the authors are relying on stratification results instead of interactions. The paragraph suggesting the need for more diverse populations is well-stated, but could be improved by acknowledging that genetic risk profiles may be different in more diverse populations. Similarly, the associations between genetic risk profiles and risk for CVD, MI, etc., may be different in other populations (e.g. Asians, Hispanics, Blacks, etc.). The paragraph starting on line 370 claiming 'genetic CVD risk in males confirm previous research' is, in the reviewer's opinion, overstated based on that statistical approach used to examine sex effects. This conclusion cannot be drawn based on stratification. Please consider running interactions - this will greatly strengthen the manuscript and the inferences that can be drawn regarding sex differences. Overall - this is a well-designed study and well-written manuscript. The reviewer's primary concern/complaint is the use of stratification analyses by sex, rather than interactions. This weakens the paper and the ability to draw inferences regarding sex differences. The authors should either run interaction analyses by sex or remove language from the manuscript drawing inferences based on supposed sex differences in the stratification analyses. The reviewer acknowledges that these sex differences do appear to exist, but that they should be tested with interaction terms. Alternatively, if the authors have strong methodological or statistical reasons for using stratification rather than interaction analyses, please state that in the manuscript to justify the statistical approach used. Thank you for your manuscript - the results are interesting and add to the literature. The use of three different diet quality indices is a huge strength of this study, as is the addition of genetic profiles.

REVIEWER	Sheila Barrett Northern Illinois University
REVIEW RETURNED	19-Nov-2020

GENERAL COMMENTS	Novel research on using 3 different dietary indices for this research as well as being able to use genetic risks in the analyses. Statistics chosen were appropriate, very in-depth analyses. Clear explanations given for each variable and what they measured. Best wishes in your publication.

REVIEWER	Steen Stender University of Copenhagen Department of Nutrition, Exercise and Sports Rolighedsvej 26, DK-1958 Frederiksberg Denmark Denmark
REVIEW RETURNED	22-Nov-2020

GENERAL COMMENTS	This study appears to me to be very well conducted and I have only very few tecnical comments and then a major concern. Tecnical comments: Page 3 dot 3:: for each person. Page 3 dot 4: I find the logic in the long sentence difficult to follow. Apparently it tells the reader that self-selection bias among the 77.004 participants out of 502.536 UK-biobank is of minor importance because associations between demographic and behavioural risk factors and mortality in the UK-bibank are comparable to those from from national health surveys. Page 8 line 164 add NDI in Australia according to the tittle of ref 39. Page 8 line 166: CVD related mortality what is it? Page 8 line 177 Polygenic risk score. I assume that the PRS is calculated for each of the 77.004 participants that fulfill the incusion criteria. How many were included? Page 9 line 191 What is plink score? Page 9 line 204-207: why is iii missing and why is there two v. The acronyms for the various educations should be spelled out some where. Maybe in the foot notes to table 1. In this table PA shoul be spelled out. General comment My problem with this study is that it has created the base for something that is even more interesting: how does different classes of foods associate with total mortaliy or CVD. In my mind the first question is how do foods from the plant kingdom associate with total mortality and CVD compared to foods from the animal kingdom? If it turns out that plant food is the driver for the combined result, then a subdivision of plant food in healthy plant food and unhealthy plant food would be interesting. And similar subdivision of foods from the animal kingdom. Inclusion of the polygenic risk factor may reveal further insight. The Predimed study had shown in a randomized setting that 30 gram of tree nuts daily and 50 ml of extra virgin olive oil each significantly reduced CVD events compared with a group that did not receive extra amounts of tree nuts or olive oils.. It would be interesting to see if the intake of the two foods associate with total mortality or CVD in a UK-population.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author
Abstract: In the methods, place the information about the polygenic profile prior to the HR models, and make it clear that genetic profile was included in the model with diet quality - this is necessary for readers to understand the conclusion that associations with diet quality were independent of genetic risk.

Response: As requested, information about the risk score has been moved earlier and the word "independent" has been added to indicate that both diet quality and genetic risk were included in the same model:

> Abstract, line 8: "A polygenic risk score was created from 300 single nucleotide polymorphisms (SNP) associated with CVD. Cox proportional hazard ratios (HR) were used to estimate independent effects of diet quality and genetic risk on all-cause mortality, CVD mortality, MI and stroke risk. Dietary intake (Oxford WebQ) was used to calculate Recommended Food Score (RFS), Healthy Diet Indicator (HDI) and Mediterranean Diet Score (MDS).."

In the results, prior to listing HRs, indicate what they are for - example: 'one-point higher RFS for all-cause mortality was 0.96 (95% CI: 0.94, 0.98)." Please also indicate that the information in parentheses is the 95% CI.

Response: This has been revised as requested:

> Abstract, line 16: "The adjusted HR associated with one-point higher RFS for all-cause mortality was 0.96 (95% CI: 0.94 to 0.98), CVD mortality was 0.94 (95% CI: 0.90 to 0.98), MI was 0.97 (95% CI: 0.95 to 1.00) and stroke was 0.94 (95% CI: 0.91 to 0.98). The adjusted HR for all-cause mortality associated with one-point higher HDI and MDS was 0.97 (95% CI: 0.93 to 0.99) and 0.95 (95% CI: 0.91 to 0.98), respectively. The adjusted HR associated with one-point higher MDS for stroke was 0.93 (95% CI: 0.87 to 1.00)."

Main manuscript
Introduction: The introduction is well-written, concisely states prior research and the need for additional studies, as well as how this study fills those gaps.

Response: Thank you.

Methods: Supplemental Tables 1-3 are difficult to read because of a lack of horizontal lines separating the different food groups. For example, supplemental Table 2, it looks as though 'Mixed vegetables' are part of the 'dietary fibre' category. Suggest adding horizontal lines are where the text for the reference row is so that it is clear how rows align. There are also some typos in the tables.

Response: Horizontal lines have been added to Supplemental Tables 1-3. The authors cannot see any typos but are happy to address any if these are brought to our attention.

Results: Table 2 states in the footer that all associations were p<.001. this makes sense given the extremely large sample size; however, the meaningfulness of this is questionable given that the actual values for the scores in different tertiles do not appear to be meaningfully different.

Response: We have included Table 2 for the following reasons:

1) To show that diet quality scores reflect differences in underlying food and nutrient intakes that show linear trends across tertiles
2) To provide a descriptive understanding of what the diet quality indices mean to assist with interpretation and translation into actual food intakes
3) To show the overall pattern of intakes, as differences across tertiles of diet quality indices reflect small difference across multiple, cumulative exposures
4) To show the absolute differences for both foods and nutrients, while the nutrient difference are small, some of the food group differences are substantial e.g. fruit, vegetables and wholegrain.

We have added additional rationale for this Table in the Methods:

> Line 234: “This descriptive analysis aimed to show that diet quality scores reflect differences in underlying food and nutrient intakes, thus assisting with interpretation and translation into actual food intakes.”

To clarify the range of values for diet quality scores across tertiles, the total possible scores for the diet quality indices are: RFS between 0 to 21, HDI between 0 to 11 and MDS between 0 to 9). We have added this to the Table 2 footnotes to help the reader interpret the differences in values for the score across the tertiles.

> Table 2 footnote: “The total possible scores for the diet quality indices were: RFS between 0 to 21, HDI between 0 to 11 and MDS between 0 to 9.”

Restate the sentences reporting HRs to say the HR is for prior to reporting it, similar to the suggestion for the abstract: "The adjusted HR associated with a one point higher RFS for all cause mortality was ..., for CVD-related mortality was ..." - this will make it much easier for the reader, as the current sentence structure requires readers to look back at the HR after finding out what the HR is for.

Response: The results have been updated accordingly.

Given the 'evidence of an association between RFS and all-cause mortality and stroke in females" the reader wonders why the researchers did not run interactions with sex in the models? Suggest not reported potential sex differences based on stratification. Rather, the researchers would be better-suited to run interactions, as different results based on stratification cannot be compared statistically, which causes the reader to question the decision to not run interactions in the models for sex. Regardless of the final decision, the authors should indicate in the statistical analysis section that they conducted additional analyses with interactions by sex (preferred) or stratified by sex (not preferred). This is recommended for all of the diet quality indices examined, particularly given that the stratified analyses suggest that a significant interaction term will exist for sex in many cases.

Conclusions: The authors conclusion regarding different pathways underlying 'genetics of both MI and death' for males and females is somewhat inappropriate given the use of stratification rather than interaction by sex. The reviewer strongly recommends running interaction analyses or softening language regarding conclusions about sex differences. Given the ease of running an interaction model, it is a disservice that the authors are relying on stratification results instead of interactions.

Response: We have now run interaction analyses and have added information on this to the Methods and Results. As detailed below, interaction analyses for sex by diet quality were not statistically significant, while the interaction by sex and genetic risk score was statistically significant for MI. Our

rationale for including the stratified results is based on recommendations from this publication (http://dx.doi.org/10.1136/heartjnl-2019-315299), where the author recommends that sex-specific associations should be presented in cardiovascular research, even in the absence of apparent sex differences (e.g., based on moderation analyses). We have cited this reference in the Methods now too. We have also revised any conclusions to soften the language regarding conclusions about sex difference:

- Methods, line 248: "The role of sex by diet quality and by PRS interactions were further tested by adding an interaction term to each model. Consistent with recommendations for sex differences in cardiovascular associations,[49] analyses were presented stratified by sex regardless of whether there were any apparent sex differences."
- Results, added to each diet quality section: "There was limited evidence (all p-values > 0.1) of sex by diet interactions." and "The interaction of sex by PRS showed evidence that the effect of higher PRS on higher risk of MI was more pronounced for males (HR=1.21 [95% CI: 1.06, 1.37], p=.004)."
- Discussion, line 322: "There was some evidence suggesting that the underlying genetics of MI may follow different pathways in males and females."
- Discussion, line 352: "Although there was limited evidence of interactions for sex by diet quality index, our stratified results showed large effect sizes for associations between diet quality and genetic CVD risk in males. This warrants further investigation as previous research shows stronger associations in males than females.[14 60] The role of sex may be partly explained by the high prevalence of diabetes and unhealthy behaviours in men,[61] and in this study may be due the lower number of events and deaths in women compared with men in the present study. Nonetheless, it is possible that the biological and behavioural pathways in which risk factors exert their effects on CVD risk are different between men and women.[60]"

The paragraph suggesting the need for more diverse populations is well-stated, but could be improved by acknowledging that genetic risk profiles may be different in more diverse populations. Similarly, the associations between genetic risk profiles and risk for CVD, MI, etc., may be different in other populations (e.g. Asians, Hispanics, Blacks, etc.).

Response: We have updated this accordingly:

> Discussion, line 347: "Moreover, the UK Biobank participants are predominately white and highly-educated and genetic risk profiles and their associations with risk of CVD may be different in more diverse populations. Thus, further research in diverse populations is needed to investigate the applicability of these diet quality methodologies for examining CVD risk independent of genetic risk.[7]"

The paragraph starting on line 370 claiming 'genetic CVD risk in males confirm previous research' is, in the reviewer's opinion, overstated based on that statistical approach used to examine sex effects. This conclusion cannot be drawn based on stratification. Please consider running interactions - this will greatly strengthen the manuscript and the inferences that can be drawn regarding sex differences.

Response: We have revised this sentence to soften the language and to be clear that we are referring to stratified analyses:

> Discussion, line 352: "Although there was limited evidence of interactions for sex by diet quality index, our stratified results showed large effect sizes for associations between diet quality and genetic CVD risk in males. This warrants further investigation as previous research shows stronger associations in males than females.[14 60] The role of sex may be partly explained by the high prevalence of diabetes and unhealthy behaviours in men,[61] and in this study may be due the lower number of events and deaths in women compared with men

in the present study. Nonetheless, it is possible that the biological and behavioural pathways in which risk factors exert their effects on CVD risk are different between men and women.[60]"

Overall - this is a well-designed study and well-written manuscript. The reviewer's primary concern/complaint is the use of stratification analyses by sex, rather than interactions. This weakens the paper and the ability to draw inferences regarding sex differences. The authors should either run interaction analyses by sex or remove language from the manuscript drawing inferences based on supposed sex differences in the stratification analyses. The reviewer acknowledges that these sex differences do appear to exist, but that they should be tested with interaction terms. Alternatively, if the authors have strong methodological or statistical reasons for using stratification rather than interaction analyses, please state that in the manuscript to justify the statistical approach used.

Response: Thank you. The issue raised regarding interactions/stratification has been addressed above.

Thank you for your manuscript - the results are interesting and add to the literature. The use of three different diet quality indices is a huge strength of this study, as is the addition of genetic profiles.

Reviewer: 2

Comments to the Author
Novel research on using 3 different dietary indices for this research as well as being able to use genetic risks in the analyses. Statistics chosen were appropriate, very in-depth analyses. Clear explanations given for each variable and what they measured. Best wishes in your publication.

Response: Thank you.

Reviewer: 3

Comments to the Author
This study appears to me to be very well conducted and I have only very few technical comments and then a major concern.

Technical comments:
Page 3 dot 3:: for each person.

Response: We have revised this to make it clearer that a score is available for each participant: "A polygenic genetic risk score was created for each participant using 300 SNPs known to be associated with CVD and all-cause mortality."

Page 3 dot 4: I find the logic in the long sentence difficult to follow. Apparently it tells the reader that self-selection bias among the 77.004 participants out of 502.536 UK-biobank is of minor importance because associations between demographic and behavioural risk factors and mortality in the UK-biobank are comparable to those from national health surveys.

Response: This has been revised to remove reference to the risk factors: "The present analysis is likely to be subject to self-selection bias associated with the number of participants who completed the dietary assessment and the low response rate."

Page 8 line 164 add NDI in Australia according to the tittle of ref 39.

Response: We have added “in Australia” to line 164.

Page 8 line 166: CVD related mortality what is it?

Response: We have removed all mention of “related” in the manuscript, so “CVD-related mortality” is now referred to as simply “CVD mortality”.

Page 8 line 177 Polygenic risk score. I assume that the PRS is calculated for each of the 77.004 participants that fulfill the inclusion criteria. How many were included?

Response: Yes, PRS were calculated for all participants included in this analysis. We have clarified in the Methods:

> Methods, line 193: “PRS were available for all participants included in the final study sample, where PRS were transformed to standardised Z scores and were treated as a continuous variable in all modelling.”

Page 9 line 191 What is plink score?

Response: PLINK is an open-source tool for genomic research. We have clarified this in the Methods ands have added a supporting reference.

- Methods, line 189: “The PRS was estimated using PLINK, an open-source tool for genomic research,[45] by generating the sum of the number of risk alleles present at each locus and weighting by the log of the odds for that locus18 estimated from the list of 300 SNPs using the PLINK “–score” command – with no-mean-imputation flag.”
- Ref 45: Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).

Page 9 line 204-207: why is iii missing and why is there two v. The acronyms for the various educations should be spelled out some where. Maybe in the foot notes to table 1. In this table PA should be spelled out.

Response: We have made the following changes:

> Methods: The typo in letter ordering has been revised.
>
> Table 2 footnotes: “A levels/AS levels, Advanced levels/Advanced Subsidiary levels; O levels/GCSE/CSEs, Ordinary levels/General Certificate of Secondary Education/General Certificate of Education; Professional qualifications include NVQ (National Vocational Qualification)/HND (Higher National Diploma)/HNC (Higher National Certificate), other professional qualifications;”

General comment
My problem with this study is that it has created the base for something that is even more interesting: how does different classes of foods associate with total mortality or CVD. In my mind the first question is how do foods from the plant kingdom associate with total mortality and CVD compared to foods from the animal kingdom? If it turns out that plant food is the driver for the combined result, then a subdivision of plant food in healthy plant food and unhealthy plant food would be interesting. And similar subdivision of foods from the animal kingdom. Inclusion of the polygenic risk factor may reveal further insight.

The Predimed study had shown in a randomized setting that 30 gram of tree nuts daily and 50 ml of extra virgin olive oil each significantly reduced CVD events compared with a group that did not receive extra amounts of tree nuts or olive oils.. It would be interesting to see if the intake of the two foods associate with total mortality or CVD in a UK-population.

Response: Thank you for this suggestion. While we agree that looking at plant foods and animal foods would indeed be interesting, this would require a new aim and approach that is outside of the scope of this manuscript and UK Biobank application, which was specifically designed to examine overall dietary patterns. It is also noteworthy that a 2019 publication using the UK Biobank dietary data highlighted that "Components of the Mediterranean diet, like nuts, fish or olive oil, may not be consumed frequently enough in the UK for a small number of 24-h dietary assessments to record reliably, which should be considered if researchers plan to assess associations of this dietary pattern with incident disease." (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6842574/)

We have added a sentence to the Discussion to suggest looking at plant and animal foods as a recommendation for future research:

> Discussion, line 346: "Future diet-disease research should extend this to better understand the role of specific plant and animal foods as part of overall dietary patterns."

VERSION 2 – REVIEW

REVIEWER	Steen Stender Department of Nutrition, Exercise and Sport University of Copenhagen Denmark
REVIEW RETURNED	23-Jan-2021

GENERAL COMMENTS	The manuscript has improved considerably in clarity. The findings will be important in future discussions concerning the diet/disease relations