Associations of cereal grains intake with cardiovascular disease and mortality across 21 countries in the Prospective Urban and Rural Epidemiological Study: A prospective cohort study

We thank the reviewers and the journal editors for the detailed review of our manuscript.

The ORCID account of the corresponding author, Dr Mahshid Dehghan (https://orcid.org/0000-0003-2149-847X), has been updated now.

Detailed comments from the meeting:

1. Nutritional epidemiology is a popular topic with our readers and there is a lot of interest in the respective merits of whole or refined grains. Thank you for thinking of us.

2. We thought confounding by poverty could account for some of the associations observed. Refined grain consumption might be associated with deprivation in low income countries. Might you consider separating your results by country income level (such as low, medium and high) to explore confounding and different effects?

Response:
We agree that poverty might confound for some of the observations. We conducted the analysis adjusted by wealth index (in addition to education) which captures the construct similar to “deprivation” and may in fact be a stronger descriptor. The results are unchanged. Apart from the base model, for all models in all tables and figures showing associations of exposures with clinical outcomes, wealth index has been included as a covariate.

3. In the abstract please give some sense of what "high intake" means. Can you quantify it in terms of servings or frequency or other suitable measure?

Response:
The quantity considered as high intake is ≥350 g/d or approximately 7 servings per day. The serving size has been included in the abstract and methods section (Page 2, and page 10 line 8-10).

4. Please include a declaration of public and patient involvement. How were the members of the community at various sites involved in helping with the study? Please share this information. Please also include a statement about your plans for disseminating the findings from your study.

Response:
The patient and public involvement section has been included in the manuscript. The following text has been added:

Page 14 lines 12-16: “The individuals participating in the study are those chosen from 630 communities and are not patients but are instead members of the public. They volunteered to be part of the study and provided their time for free. We intend to publish the information in a high
impact peer review journal, followed by press releases issued from each of the participating institutions in each of the 21 countries. In addition, we intend to use social media.”

5. Please also revise your paper to respond to all of the comments by the reviewers. Their reports are available at the end of this letter, below.

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how you have dealt with them in the paper.

The point by point response to the reviewer’s is appended below in bold font. Revisions in the manuscript are in red font. The values in all tables have been updated after adjustment for household wealth. Only additional tables or figure legends are in red font in the manuscript and supplementary appendix.

** Comments from the external peer reviewers**
Reviewer: 1

Recommendation:

Comments:
Originality - does the work add enough to what is already in the published literature? If so, what does it add? If not, please cite relevant references.
It depends a bit on how the study is interpreted

Importance of work to general readers - does this work matter to clinicians, patients, teachers, or policymakers? Is a general journal the right place for it?
Studies of diet are always popular

• Scientific reliability
This is a difficult question.
Observational studies of diet are notoriously difficult to interpret, because diet is very strongly associated with socio-economic position which is a key determinant of health. As such, it is always difficult to know whether studies of diet give attributes of the people who eat a specific diet or the actual effects of diet.

Response
The reviewer makes a very important point. As with any observational study, residual confounding is a possibility, and no degree of statistical adjustments can completely account for such effects.
However, we investigated the influence of SES/poverty using 3 different measures of SES to adjust in the analysis of the associations between different grain intakes and clinical outcomes. Three measures of SES were household wealth, education, economic level of the country, and within that urban and rural locations. We evaluated several of these in the INTERHEART (Rosengren, 2009, BMJ) study as well as in the PURE study (Rosegren 2019, Lancet Global Health) and found that education was the strongest marker of SES associated with risk of cardiovascular disease. Additionally, we adjusted for study center as a random effect which takes into account the clustering of socioeconomic factors by community. When, we included household wealth, or economic level of the country our results were unchanged.

We have therefore done an extensive analyses by SES as is possible by using education, wealth index, centre, urban v rural and economic level of the countries as a comprehensive approach to adjust for SES. The results remain unchanged.

What is unusual about this paper is that the minimally adjusted (age, sex and centre) results are pretty much null but the association is much stronger after adjustment. In addition, that adjustment is for factors that are not clearly justified as confounders, when it is well known that adjusting for mediators or “colliders” can generate biased associations.

Response:
We agree with the reviewers that some of the variables such as history of diabetes and waist-hip ratio, considered in the model might be mediators. But considering the biology and the causal pathway there aren’t any colliders included. Removing these two variables from the model did not change the results of the analysis apart from minor changes in the hazard ratio. Therefore, we chose to retain both variables in the final fully adjusted model. The choice of the confounding variables was made based on the approaches used in previous publications of similar topics. The analysis done with and without the use of history of diabetes as a covariate in the model is
presented in the supplement and the results remain unchanged (Supplementary table S5, S6, and S7).

• Research Question - clearly defined and appropriately answered?
The results are presented in the abstract, “What this study adds” as showing an association for ‘refined grains’. However, page 7 line 36/7 says “refined grains included only refined wheat products”. It would be better if the Methods had a section entitled “exposures” where it was explained why this categorization was chosen, how it corresponds to other similar studies and exactly what it represents. It then should be made clear throughout the paper that the key exposure is “refined wheat”.

Response:
The term “refined grains” has been retained in the section on “What this study adds” and in the rest of the sections as apart from wheat and wheat products it included food items with ready to eat cereals with corn.
A section entitled “Exposures” has been added in the methods section. Some of the existing text in the methods section in the previous version has been shifted to this section. The justification for the categories is also provided (Pages 8-10).

• Overall design of study - adequate?
To be fair, it is hard to think of a better design to answer this specific question. However, to date, observational studies of dietary items, such as vitamin supplements, have not turned out to be reliable guide to effective interventions, while it has been difficult to check other observed associations. So, it raises the question of whether it would be better to work out which particular dietary constituents (proteins, fatty acids, etc), are harmful or protective and then deduce from that the effect of particular types of food. It would also provide a framework for addressing other questions about dietary items.

Response:
We previously assessed association between fats, carbohydrates, and proteins with lipid markers and clinical outcomes. Higher carbohydrate intake was significantly associated with higher total mortality (Dehghan et al. 2017 Lancet). In this paper, we demonstrate that the harms are mainly from refined grains.
We agree that nutrition epidemiology is extremely complex and so we have analyzed the data by macronutrient content (Dehghan et al, 2017, Lancet and Mente et al, 2017, Diabetology) and also by specific food categories in the present manuscript. These three papers together are helpful in guiding what the effects of different dietary components on outcomes are.

• Participants studied - adequately described and their conditions defined?
This is a well-known study.
It would be helpful to have a table showing the exposures by confounders, so that we can see what factors are contributing the change in estimates after adjusting, or maybe it could be stated in the text, as it would help the reader interpret the different models.

Response:
The covariates for the models were chosen a-priori based on previous literature on factors associated with these outcomes. We have now statistically examined the association of these covariates with exposures namely grain intake. To check for these associations the one-way analysis of variance was used for continuous variables of wealth index, age, waist-hip ratio, intakes of total energy, fruits and vegetables, and red meat for whole grain and refined grain as appropriate. For the categorical variables of sex, location, education, smoking, physical activity, and history of
diabetes, Pearson’s chi square test was used to test the association before inclusion as a covariate in the model. All covariates were associated with grain intake and we have now added this analysis as a statement indicating the significant association in the results section as “All co-variates considered were significantly associated with grain intake and were considered in the fully adjusted models (data not presented, all p<0.001).” In addition we have added the results of the associations as supplementary tables (S2 and S3, S4).

• Methods - adequately described? Complies with relevant reporting standard - Eg CONSORT for randomised trials? Ethical?

The methods would be better with more clearly headed sub-sections.

Response:
Subheadings have now been added to the methods section.

Please provide more detail about the 200kcal white bread exposure (Table 4)

Response:
We have revised the manuscript to include the following information. Table 4 which is now Table 6 describes the macronutrient values contributed by white bread, whole wheat bread, cooked rice and other intact grains which are commonly consumed by the PURE participants. The main source of energy from carbohydrates is from white bread followed by cooked white rice. Thus, the following text has been included

Page 18 lines 1-4: “Mean daily energy intake of PURE participants was about 2100Kcal/d of which white bread is the main source. Because three servings of white bread contributes to about 10% of total daily energy intake, we estimated the effect of a 10% increase in refined carbohydrate consumption on risk of mortality and other outcomes”.

Please explain in more detail about the “Cox frailty” model. What was the rationale for using it? What is the underlying time variable? Does it require the Cox proportional hazard assumption? If so, please explain how it was tested.

Response:
The data for the study was obtained from 630 centres across 21 countries, it was expected that there would be a correlation between observations within a cluster which violates the assumption of independent observations, which is required for analyses using the Cox proportional hazard model. Hence the Cox frailty model was used where the frailty variable (cluster in this case) accounts for the clustering at each centre. The Cox proportional frailty model with centre as a random effect was therefore considered more appropriate for analysis compared to the standard Cox proportional model.

We evaluated the assumption of the Cox and added as stated in the text “The proportional hazards assumption was evaluated by visual inspections of log-log plots, which were consistent with proportional hazards. We have now added a few more details on the frailty model in the methods section. Thus, the following sentence has been included;

Page 11 lines 22-23: “The frailty was assumed to follow a Gamma distribution with correlated observations within cluster”

Please justify in detail each variable included in the model as a confounder, i.e., a common cause of both the relevant dietary intake and the outcomes.
The covariates for the models were chosen a-priori based on previous literature on factors associated with these outcomes. We have now statistically examined the association of these covariates with exposures namely grain intake. To check for these associations the One-way analysis of variance was used for continuous variables of wealth index, age, waist-hip ratio, intakes of total energy, fruits and vegetable, red meat as well as for whole grain and refined grain as appropriate. For the categorical variables of sex, location, education, smoking, physical activity and history of diabetes, Pearson’s chi square test was used to test the association before inclusion as a covariate in the model. All covariates were associated with grain intake and we have now added this analysis as a statement indicating the significant association in the results section (Page 15 lines 6–11).

Tables of covariates and their associations with the 3 grains are also added to the supplementary appendix (Tables S2, S3 and S4)

Please give the rationale for any subgroup analysis and the methods used to test for differences by subgroup

Response:

We conducted sub-group analysis by regions for refined grains (Figure 1) and white rice (Figure 2) as regions of high and low intakes.
The median intake of refined grains was lower in Africa and South Asia and were considered as regions with low intakes of refined grains as opposed to the rest of the regions and a stratified analysis was conducted (Figure 1). The reason for selecting Asian vs Non-Asian for white rice was (1) because diet (especially the consumption of carbohydrates) varies markedly between these 2 regions, with Asian countries consuming higher carbohydrates (>65% energy) and non-Asian lower carbohydrate intake (~55% energy) (2) the intake of rice was higher in the Asian regions compared to the rest of the regions. (Figure 2)

We also did an analysis by BMI (Supplementary figures) because previous studies showed the association between refined grains or whole grains and CVD is stronger among obese individuals (Newby et al, 2007, AJCN; Gaesser, 2019, Advances in Nutrition). In addition, BMI is the indicator used for public health purposes.
Both analyses demonstrate that the associations between refined grains and events were similar in the above subgroups.

• Results - answer the research question? Credible? Well presented?
Please give the follow-up rate for the study.

Response:
. The rate of follow up was 94% over the median duration of follow-up of 9·4 years (IQR 5·3-9·3).

Please give the model when quoting results in the text.

Response:
The text in the results section has been revised to indicate the models in the results section. The following texts were changed:
Page 15 Lines 13-18: “In the fully adjusted model, compared with low intake, a higher level of intake (<50g/d vs >350g/d) was significantly associated with higher risk of composite outcome (HR 1.29; 95% CI 1.16 to 1.43 p trend <0.001), total mortality (HR 1.28; 95% CI 1.11 to 1.47 p trend= 0.003), non-CV mortality (HR 1.31; 95% CI 1.11 to 1.56 p trend= 0.004), major CVD events (HR 1.34; 95% CI 1.17, 1.53 p trend <0.001) and stroke (HR 1.47; 95% CI 1.22 to 1.78 p trend <0.001).”

Page 16 lines 1-5: “No significant associations of whole grains were observed with mortality or cardiovascular events in the fully adjusted models. In the base model, a higher risk for composite events, total mortality, non-CV mortality, CV mortality, major CVD and stroke was observed in those with no whole grain intake.

Please note that the “base model” showed whole grains to be protective.

Response:
Text has been added to indicate the results of the base model. In response to the reviewer’s comment, the reference group has been changed to the highest category (≥100 g/d) of whole grain intake based on comments from other reviewers. The following sentence has been added:

Page 16 lines 3-5: “In the base model, a higher risk for composite events, total mortality, non-CV mortality, CV mortality, major CVD and stroke was observed in those with no whole grain intake.”

Please check the text for white rice. Supplementary Table 1 appears to show higher white rice intake to be protective in the base and minimally adjusted models.

Response:
The following sentence to indicate this has been added in the results:

Page 16 Lines 12-14: “In both the base and minimally adjusted models, higher intakes (≥450 g/d) had a lower risk of the composite events, total mortality, non-CV mortality and major CVD.” However, in the final fully adjusted model this association was not significant.

• Interpretation and conclusions - warranted by and sufficiently derived from/focused on the data?
Message clear?
Please make it clear throughout that “refined grains” means “refined wheat products”, and that any comparisons are drawn with a similar exposure

Response:
The term “refined grains” has been retained throughout the document as it includes not only refined wheat products but also ready-to-eat corn cereals. The comparisons made in the discussion also make these distinctions clear.

Please revise the interpretation and conclusions after reviewing whether the models interpreted were adjusted correctly or explain clearly the rationale for preferring one model rather than another.

Response:
The final interpretation of the analysis is based on the fully adjusted frailty model as this model takes into account all confounders identified from the literature. This model also accounted for the intake of other grains, socio-demographic, economic and dietary confounders. The fully adjusted
model with medications is an additional model to examine the confounding effect of medication on outcome. However, the results of the two models (fully adjusted and additional adjustment for medications) did not differ. There were several instances where the base model (whole grain for instance) showed statistically significant associations, but these were not significant on further adjustments. A sentence has been added in the discussion to indicate this:

Page 23 Lines 5-6: “Despite the addition of medications in the model, these results did not change indicating that our findings were robust.”

• References - up to date and relevant? Any glaring omissions?
Maybe the study recently published in the BMJ on refined carbohydrates is also relevant

Response:
The recent publication in the BMJ (Ho et al, BMJ 2020) has been added to the references.

• Abstract/summary/key messages/What this paper adds - reflect accurately what the paper says?
Maybe these could be revised when the interpretation of the study is more nuanced and reflects the finding about refined wheat grain.

Response:
As the results have remained the same, apart from changes in hazard ratio values (due to an additional adjustment for household wealth) no changes have been made in the Abstract/summary/key messages/What this paper adds.

Please enter your name: CM Schooling

Job Title: Associate Professor
Institution: The University of Hong Kong
Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No
Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No
Reviewer: 2

Recommendation:

Comments:
In this manuscript, the authors analysed data from 21 countries in the PURE study, and reported on the associations between refined grains, whole grains and white rice with risks of mortality and cardiovascular outcomes, as well as the cross-sectional associations of these foods with blood lipids and blood pressure at baseline. The higher risk associated with refined grains is consistent with existing literature, while the null association for whole grains is inconsistent with inverse associations reported in previous studies; and there had been few previous studies which examined white rice. This is a large study, and so the data presented is valuable since there has been limited evidence previously from lower to middle income countries, however, I have some queries regarding the analyses, and some suggestions for the authors to consider.

1) In my view, the major advantage of the PURE study over that of other large existing cohorts mostly based in Europe and the US is the range of countries included, which allows a unique opportunity to explore possible heterogeneity by region/income level. The countries included are also likely at different stages of the nutrition transition model, which means consumption of the same food groups might mean different things in different countries, for example, refined grains and other processed foods might be consumed more by high socio-economic status participants in the lower income countries who could afford the purchase, but the same foods are more likely to be consumed by lower socio-economic status groups in high income countries. Therefore, I think it would both be interesting and of value to present stratified analyses by region/country income level, alongside the main analyses.

Response:
As suggested by the reviewer, we present the analysis by country income level in the supplement (Table S17 to S19). We agree that looking at association by individual countries or additional regions could be of interest, but it is unlikely to give robust results due to fewer events in several countries and some regions (e.g. the Middle East or Africa). Therefore, we elected to present our findings based on country income level (low-, middle- and high-income countries). Whole grain intake was not associated with outcomes in any of the income categories (Table S18). We have added a meta-analysis of region-wise on association of grains with outcome and the results are consistent.

2) The authors presented analyses in 4 or 5 categories of grain consumption, using the lowest category as the reference group. While this allows easy interpretation of the findings, in the case of whole grains, this meant using zero consumption as the reference group, which might not be appropriate since total non-consumers are likely to be different from consumers in many other ways, so residual confounding might be present. Therefore, alternative (using a different reference group) or complementary presentations (modelling the association continuously, or explore possible non-linear effects) could be considered to ensure the robustness of the findings. In particular, as shown in Table 1, there were substantial variations in whole grain consumption by regions, and so there might be some clustering by region in categorisation of the whole grain variable, which may not completely accounted for in the statistical modelling. It would therefore seem important to explore possible heterogeneity in whole grain intake by region as well.

Response:
We agree with the point. Therefore, we now have the highest intake category of grains as the reference and described whether there was any increased risk of any of the outcomes with lower
intake of whole grains, which we did not observe. All the tables on whole grains both in the main document and in the supplementary tables, and text (Page 16) have been changed to reflect this. To assess the shape of associations between whole grains intake and primary outcomes we used restricted cubic splines with knots at 50 g intervals. However, the earlier survival model remains the primary analysis as splines are “descriptive analyses used to support” and help describe the shape of associations (Figures 3). In addition, we have fitted cubic splines for intakes of refined grain and white rice with knots at the same 50 g intervals. In order to address the clustering of intakes within region we have now performed the cox-proportional hazards models with frailty separately within high and low intake regions. For refined grain, low intake consisted of Africa and South Asia whose median intakes were below 100 g and the remaining regions constituted those with high intake. These stratified analyses are presented in Figure 2. The whole grain intake of Africa was much higher compared to the rest of the regions, but the sample size and the numbers of events were not large enough to perform separate analysis of the African region. Hence, we performed a sensitivity analysis where in addition to the current analysis based on the whole data, we excluded data from African countries and explored the association of whole grain on outcomes in the rest of the regions with lower intake (Supplementary Table S10). The results remained consistent.

3)  For the categorisation of the main exposures, the authors used grammes of intake of each food, and secondarily adjusted for energy in the statistical modelling. By doing this, a participant who eats 100g of refined grains on a 1200kcal diet would be put in the same category as another participant who eats 100g of refined grains on a 2500kcal diet, despite the fact that refined grains would make up a much larger proportion of the diet in the first participant. Adjustment for total energy in the modelling is likely not sufficient to remove the confounding here when the exposure is analysed categorically, since participants are not allowed to move across the exposure categories, and this could sometimes attenuate associations to the null or in the opposite direction. Alternative methods of energy adjustment, such as nutrient residuals method (PMID: 9094926, 10084242) which accounts for energy intake when categorising the exposure, should be considered instead.

Response:
We understand the reviewers concern, however nutrient residuals modeling is mainly used when assessing association between nutrients and clinical outcomes and previous studies on association between food intake and clinical outcomes did not apply this method (Tong et al, 2019; Farvid et al, 2016).

The following sentence to indicate this has been added in the statistical analysis section:

Page 12 lines 1-3: “For all analyses on associations, total energy intake was adjusted using a standard multivariable model as the study examined the association between food intake and clinical outcomes19,20.

4) Statins and blood pressure medication were adjusted for in secondary analyses, but the results were not presented in the tables. Since these are likely important confounders, was there any reason for not including these variables in the fully adjusted model? Alcohol intake and household wealth would also appear to be possible confounders, but these do not appear to be adjusted for?

Response:
We updated results including blood pressure medication and statin in the tables as “Fully adjusted with medications” with statins and blood pressure medications included as covariates. The results were unchanged.
We did not adjust for alcohol intake because in five Muslim countries (Bangladesh, Iran, Pakistan, Saudi Arabia, and UAE), alcohol intake was presumed to be low and so was not recorded. We have updated our result using wealth index as an additional covariate and this did not change the results. All tables have been updated.

5) On a stylistic note, the results for refined grains and whole grains were presented as tables while the results for white rice was presented as a figure in the main results, but I think unless there is a clear reason for doing so readers might find it easier to interpret the findings if they were all in a consistent format. In the main results tables/figures, footnotes defining composite events and major CVD would also be helpful.

Response: We thank the reviewer for the comment, and we have now presented results for rice in a table. (Page 40 Table 4)

6) Some additional details in text about the outcome ascertainment process would also helpful, without the need to check previous publications. Specifically, it is not clear to me from the current text whether the participant or their physician filled out the case report form, the format of the questions asked, or how the end of follow-up was determined (was it possible to get a precise date of diagnosis?). Proportion loss to follow up should also be reported.

Response: We have added the following statement into the text:

Page 7 lines 6-11: “Data were collected at the community, household, and individual levels with standardized questionnaires. Trained staff interview-administered the questionnaires. Standard case-report forms were used to record data on major cardiovascular events and mortality (classified by cause) during follow-up, which were adjudicated centrally in each country by trained physicians using common definitions. The events definition is detailed in the methods section of the supplementary appendix.”

Page 14 lines 19-21: The rate of follow up was 94% over the median duration of follow-up of 9.4 years (IQR 5.3-9.3).

7) Proportion of missing covariates and method of dealing with this should be reported. It would also be informative to report the treatment of covariates in the modelling, whether they are adjusted as continuous or categorical variables, and if categorical, what are the categories.

Response:

In our study, the data were >99.5% complete for all variables, except for wait-to-hip ratio (WHR) and physical activity (PA), which are control variables and therefore not central to the message of the paper. For WHR, the completeness was 94%, and for physical activity, 93%. We recognize that because missingness may introduce bias, reduce power, and/or affect the representativeness of the results. This is unlikely to be the case as the percent with missing data is low. We undertook a series of sensitivity analyses to check whether the missing data on these two covariates is informative and therefore possibly non-random. First, we excluded WHR and physical activity from the multivariable models and found that the estimates between refined and whole grains and rice with outcomes were similar. Additionally, we excluded the 6 countries that had missing data exceeding
10% on either of the two covariates. The estimates were again similar. This indicates that the missingness in our study likely had little effect on our results.

The above has been included in the discussion section (Page 22, Lines 15-24).

8) Some summary information of the results of outcome adjudication might be useful to report.

Response: We have added the following statement to the text: “

Page 7 lines 8-11: Standard case-report forms were used to record data on major cardiovascular events and mortality (classified by cause) during follow-up, which were adjudicated centrally in each country by trained physicians using specific definitions. The events definition is detailed in the methods section of the supplementary appendix.” The details of the adjudication will require an extensive methods paper on its own and will be published separately at a later time in a specialist epidemiology journal.

9) In the abstract, it would be helpful to clarify what the highest and lowest category of consumption refers to.

Response: The highest and lowest category of consumption has now been included in the abstract. The revised text is as follows:

Page 2: “The highest category of intake of refined grains (≥350 g/d or about 7 servings/d) was associated with higher risk of total mortality (HR 1.27; 95% CI 1.11 to 1.46 p trend=0.004), major CVD events (HR 1.33; 95% CI 1.16, 1.52 p trend <0.001), and the composite of the above outcomes (HR 1.28; 95% CI 1.15, 1.42 p trend <0.001) as compared to the lowest category of intake (<50g/d)”.

10) In the what’s already known on the topic box, I suggest removing the statement that high carbohydrate intake is linked with high mortality, since previous studies have in fact suggested whole grains and refined grains drive the associations differently, so I think it is misleading to report on total carbohydrates.

Response: The statement that high carbohydrate intake is linked with high mortality has been removed from the box on “What is already known on this topic”.

Please enter your name: Tammy Tong
Job Title: Nutritional Epidemiologist
Institution: University of Oxford
Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No
Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No
Reviewer: 3

Recommendation:

Comments:
I read, with interest, the paper by Swaminathan entitled "Associations of cereal grain intake with cardiovascular disease and mortality across 21 countries in the Prospective Urban and Rural Epidemiological study: a prospective cohort study". This large prospective study, enrolling over 137,130 participants, covers a wide range of urban and rural locations, and captures wide variation in carbohydrate intakes, specifically with respect to whole and refined grains. Major limitations of this study center on (1) the methodology applied to estimate whole-grain intake and (2) the fact that only one estimate of intake was captured at baseline. Both of these limitations may lead to considerable misclassification of dietary exposures. In addition, the discussion could be improved and additional supplementary tables included.

Response:
We interview-administered country specific validated food frequency questionnaire to measure habitual intake of PURE participants. We agree with reviewer that FFQ is not a measure of absolute intake and although the FFQs we used were validated, some measurement error is inevitable. However, FFQ is the most commonly used method by large epidemiological studies. FFQs are useful in categorizing individuals by groups with high, moderate or low intakes and so comparisons across categories of intakes are valid.

We acknowledge the limitation that we only measured diet comprehensively at baseline and changes in diet during the follow up may have occurred, however, if those changes occurred, they would tend to attenuate the estimates of any associations between an exposure and outcomes.

We agree that repeated measures may improve precision of estimates, but we do not have repeated dietary measurements in PURE. It has been shown in large observational studies that four different approaches for analyzing the association of dietary fats with CHD using repeated dietary measurements (baseline diet only, the most recent diet, and two different algorithms for calculating cumulative average diets) yield substantively similar results (Hu et al, 1999, Am J Epidem). However, based on these data by the Harvard group, we are confident that over a relatively short follow-up (<10 y), the substantive estimates would not differ with repeated measures. This has been added in the strengths and limitations section of the manuscript.

Major comments:
In contrast to other previously published studies (summarized in several meta-analyses to date - Tang et al 2015 Am J Cardiol; Aune et al. 2016 BMJ; Bechthold et al. 2019 Crit Rev in Food Sci & Nutr; Ye et al. 2012 J Nutr; Reynolds et al. 2019 The Lancet; Chen et al 2016 Am J Clin Nutr; Li et al 2016 Medicine; Zong et al. 2016 Circulation; Johnsen et al. 2015 Br J Nutr; Schwingshackl et al. 2017 Am J Clin Nutr), no association was observed between higher whole-grain intake and reduced risk of cardiovascular disease (CVD) and total mortality. It is possible that disparate findings in this study may be attributed to difference in the estimation of whole-grain intake. In this study, whole-grain intake is expressed as grams of whole-grain product or intact whole grain, and I believe this is leading to considerable misclassification. For example, in Table 4, whole grains that are intact and cooked were translated into 1 cup (i.e. 100 grams). If the authors did not estimate the whole grain content on a dry-weight basis and, instead, estimated the whole grain content in the prepared product/cooked basis (which includes water),
this would lead to extreme misclassification. Using data from this publication (Ross et al. 2015
&d=DwIFaQ&c=QbTp4bRXkXCIGN-AaFr7IN7w2igwuzjAb7os8TvEi8&r=TLFe9xullRz2GowGliAq0qpT6qbdmlXVluteIRTO0X0&m=hI7B
WCtdT_qFJre-AQ4r2AmfVD1CaEqH1vHRRo8111r0<saw=MaPDe4tIZ0e7igDYma4d3bE16S-bEPJtw_Xs_o65Np_A6ce4 ), for example, 200 grams of cooked oatmeal (prepared/wet weight) should not be estimated as 200 grams of whole grain. Rather the whole grain amount should be calculated on the dry basis, which translates into 34 grams of whole-grain food (before cooking). This is because oatmeal is 100% whole grain and, thus, has almost equivalent grams of whole-grain in the dry portion of the food. In contrast, for example, a 40 gram portion of ready-to-eat breakfast cereal that is classified as a ‘whole-grain food’ should not be calculated as 40 grams of whole grain because the actual amount of whole grain in the food could range from 6-38 grams. If the estimated whole-grain intake is calculated for whole-grain food products and intact whole grains as they are prepared (such as cooked barley, oat and maize porridges, bulgur, cracked wheat, etc.), and if the calculations are not based on the dry weight, this will lead to considerable misclassification. Typically, this type of misclassification is considered ‘non-differential’ and can attenuate associations; however, given the fact that this study combines the data from different countries and region, the type of error introduced may be systematic error. Could the authors confirm how whole-grain intake was estimated on dry weight? For example, in the meta-analysis by Chen and colleagues 2016 (Chen GC, Tong X, Xu JY, Han SF, Wan ZX, Qin JB, et al. Whole-grain intake and total, cardiovascular, and cancer mortality: a systematic review and meta-analysis of prospective studies. Am J Clin Nutr. 2016;104(1):164-72) whole-grain products were reported in grams, the authors multiplied product weights by 0.57 (16/28=0.57; 28 g of whole-grain products approximates 16 g of whole grain) to estimate whole-grain intake.

Response:
We agree with the reviewer that the amount of whole grain food is affected by moisture content and we acknowledge this as a limitation of our analyses. We stress this point in the limitation section of our manuscript. (Page 24 lines 6-9) However, to address reviewer’s concern we computed dry matter for whole grain foods at the region level. The justification for the region analysis is that in PURE study we used 30 different FFQs and the number of food items in the refined grains group is 566, in whole grains group is 207, and in rice group is 176. Indeed, the whole grain group in North America/Europe, South America, and Middle East is mainly comprised of whole wheat bread. Similarly, in South Asia and South East Asia breads (Chapati, paratha, puree, and roti) are main source of whole grain. But in Africa whole grain is mainly consumed as porridge (Sadza, mahindi, and samp). Therefore, we estimated the amount of whole grain consumption by taking into account the percentage moisture. We used 30-40% moist for various types of bread (McKeown 2015, AJCN) and 40-50% for different types of porridge and then assessed the association between whole grain dry matter and outcomes and results were unchanged. The results are presented in the Supplementary appendix Table S11 and reported in the results in Page 16 line 6-7.

Page 16 line 6-7: “When examined on a dry weight basis, similar results were obtained (Supplementary Table S11)”.

With respect to the exposures, it would be helpful if the authors provided the following information:
1. What are the whole and refined grain line items on each of the country-specific questionnaires, and what are the corresponding portion sizes? This information could be provided in a supplementary table.

Response:
The food items included for each of the 7 regions as refined grains, whole grains and white rice has been added in the supplementary appendix as text under the methods section on “PURE food frequency questionnaires- Regional list of food items and portion sizes for refined grains, whole grains, and white rice” (Page 25-27). We felt it was better as text rather than as a table.

2. In a supplementary table, please provide data on the individual food groups with respect to the increasing categories of whole and refined grain intake. For example, in the refined grain group, did higher intake of sweet/dessert contribute more in that category?

Response
We have included a table of intake of food groups of fruits, vegetables, dairy, red meat, nuts, sweets, egg, fish and legumes by categories of refined grain and whole grain intake in the supplementary appendix (Table S26 and Table S27). In the PURE study, sweets from various countries included not just pastries or those made from refined or whole grains but included candies, chocolates, those made with dairy, nuts etc.

3. Methods section should clearly define whole grain and refined grain intake (i.e. the foods included), rather than presenting this as a supplementary table.

Response:
The definition of refined grain, whole grains and white rice has been shifted to the methods section of the main document (Page 8-9)

4. If the authors run the analysis based on whole and refined grain estimated as ‘servings per day’ (as derived from the country-specific FFQ), are results consistent with what is observed in the current analysis (grams of whole-grains)?

Response:
The analysis in servings per day has been done and presented as a table in the supplement Table S13 and the results reported in the main document. Similar results were obtained when consumption of refined grain and whole grain, when the refined grains or whole grains were described using servings per day or grams of whole grain.

Page 16 lines 20-22: When intakes of refined and whole grains were examined as servings per day (Supplementary Table S14), similar results for refined grains and whole grains were observed.

5. What is the justification for excluding popcorn as a whole grain?

Response:
Consumption of popcorn as a snack was only recorded in Brazil and Canada. Therefore, it was not included under the whole grain group.

6. Why did refined grains only contain ‘wheat’ products? For example, with ready-to-eat-breakfast cereals, puffed rice or cornflakes are considered refined grains.

Response:
Ready-to-eat breakfast made of corn was included in the refined grains group and we have clarified this point in our revised definition. In fact, all refined grains product excluding those with
white rice were included in the refined grain group. Also, ready to eat breakfast made of rice was included in the rice group.

7. Most observational studies examining the relationship between refined grain and disease outcomes or mortality have included white rice as a refined grain, and, conversely, brown rice as a whole grain. Could the authors comment on the justification for not including white rice with refined grains, and, conversely, not including brown rice in the whole grain category?

Response:
Rice was included as a separate category from other refined grains, as about 60% of the PURE population were from Asian countries where rice was the major staple food consumed. Therefore, when we examine the association between rice and health outcomes separately, the impact of our paper would be more profound and more applicable among Asian countries. For these reasons, we were interested to separately examine the association between rice and health outcomes.

The main reason was that brown rice intake was only recorded in Argentina and Brazil and even in those countries, the consumption of brown rice was very low (in Argentina 2g/d and Brazil 16g/d). Therefore, including brown rice in the rice classification would not influence the results.

8. Refined grains include desserts and pastry. When thinking about refined grain equivalents, to whole-grains, these foods are not typically included in the refined grain category. Suggest creating two refined grain exposure categories, with and without the desserts and pastry.

Response:
The refined grains considered from the recipes without considering their sugar content. We have included the following text in the discussion “Although we recognize that desserts and pastries include both sugar and cereal grains, in mixed dishes, the sugar content was not accounted for. Sweets however, included other food groups in the recipe and therefore, the effect of added sugars in cereal foods alone cannot be assessed. The intake of the various food groups by the categories of refined and whole grain intake (Supplementary Tables S26 and S27) does not indicate an increase in added sugar intake with increasing categories of intake. A similar pattern persisted for intake of the other food groups too.

9. What is the justification for grouping North America with Europe?

Response:
We assumed that participants in Europe (Sweden and Poland) and North America (Canada) consume western diet, and these individuals are likely to be of similar ethnicity as the majority of North Americans are originally from Europe and therefore, we grouped them together.

Minor comments:
Introduction
-Page 5, lines 33-40: Split this into two sentences. Also, consider adding the following references:

Response:
The sentence has been revised and split in the introduction. Both references have been added in the introduction.

Methods
Page 6, Line 54: Were the blood samples collected fasting?

Response: we have added the following statements:

Page 7 lines 2-4: Of the 148,858 individuals, 129,740 (87%) had fasting blood sample collected, 3,632 had non fasting blood samples collected and in the rest we were not able to ascertain the fasting status when the bloods were drawn. A statement to indicate this has been added to the text in the methods section.

Page 9, line 53: Type error: “? 350 g”

Response:
Changes have been made in the manuscript.

Statistical Analysis
-Is the refined grain intake association independent of saturated fat or sodium intake?

Response:
We included an analysis with saturated fat and dietary sodium as covariates in the model and found similar results (Supplementary table S20 to S22).

-Refined grains include sweet-baked desserts—does exclusion of these foods from the refined grain category change the association?
-Did you test for interaction by sex?

Response:
We did not exclude sweet baked desserts. However, the sugar content was not considered for analysis. We have added the following text under strengths and limitations:

Page 23 lines 11-17: “Although we recognize that desserts and pastries include both sugar and cereal grains, in mixed dishes, the sugar content was not accounted for. Sweets however, included other food groups in the recipe and therefore, the effect of added sugars in cereal foods alone cannot be assessed. The intake of the various food groups by the categories of refined and whole grain intake (Supplementary Tables S27 and S28) does not indicate an increase in added sugar intake with increasing categories of intake. A similar pattern persisted for intake of the other food groups too.”

There was no interaction by sex and a sentence indicating this has been added to the results (Page 17, Line 17-18):

“Associations were similar for both men and women (p interaction>0.05).”
-Did the authors look at the percent energy from carbohydrates as the exposure variable of interest rather than absolute grams of carbohydrate?

Response:
We have previously reported the association between percent energy from macro-nutrients and clinical outcome and reported that higher consumption of carbohydrate was associated with higher risk of mortality and major CVD (Dehghan et al. 2017, The Lancet). These results are not included in this manuscript, but they are referenced in the introduction and in the discussion.

Discussion
-Page 11, lines 48-50: “…data were reanalyzed by comparing the risks…” can you clarify what this means?

Response:
This was an error and has been removed.

-Page 11, lines 52-57: Why did you choose to analyze by 200kcal increase when others, such as the reference in the line just before this, analyze based on gram increases?

Response:
The common source of energy obtained from carbohydrates is from white bread followed by cooked white rice. Thus, the following text has been included in the results section;

Page 18 lines 1-4: “Mean daily energy intake of PURE participants was about 2100Kcal/d of which white bread is the main source. Because three servings of white bread contains about 10% of total daily energy intake, we estimated the effect of a 10% increase in refined carbohydrate consumption on risk of mortality and other outcomes.”

-Page 12, lines 17-27: Can you say the observed result only because of the reduction in high GI foods when there was also an increase in low GI/high fiber (more WG) foods?

Response:
We agree with the reviewer’s point that lower consumption of refined grains might have resulted in higher consumption of whole grain (low GI). This may happen with replacement of any unhealthy food with healthy food. For example, lower consumption of starchy foods and higher consumption of fruits and vegetables.
The following text has been added in the discussion:

“Thus, these foods may result in replacement of these foods probably by lower consumption of starchy foods and higher consumption of fruits and vegetables”. (Page 19, Lines 16-18)

-Page 12, line 43: The inverse associations of WG intake with total and CVD mortality has been reported in more than “a few” meta-analyses. Consider also adding the following references:
Reynolds et al. 2019 The Lancet
Chen et al 2016 Am J Clin Nutr
Li et al 2016 Medicine
Zong et al. 2016 Circulation
Response:
All the above references have now been cited.

-Page 13, line 6-8: You mention that there is no standard definition for WG and that it varies between countries. Can you comment on any notable differences in how WG foods were captured on the different FFQs?

Response:
The FFQs were not specially designed to capture whole grain foods. The foods reported are as defined by the country specific questionnaires. However, foods such as whole wheat bread that was captured across North America/Europe, South America and the Middle East have different compositions of whole grain in the product, although within a specific range. In the supplementary material, list of foods in each categories of whole and refined grains are given (Supplementary appendix, Page 25-27).

-Page 13, lines 15-20: Can you provide a reference supporting the claim that nutrients such as fiber, vitamins, minerals, and antioxidants are removed during processing of WG products?

Response:
The reference has been added.

Page 14, lines 22-24: Why would you expect that rice and refined wheat have opposing effects on the risk of CHD?

Response:
Refined grain and white rice may have different chemical compositions, different physical properties, etc and this is suggested by our data that the adverse effects are from refined wheat products and not with rice.
We have added the following text in our discussion

Page 22 lines 9-15: “The reason for the different associations observed with refined grains and white rice is not clear. However, refined wheat is uniformly and rapidly digested and raises the blood glucose, with no health advantages of fibre. In India on the other hand increased long grain rice and especially parboiled white rice may have both a definite glycemic advantage and an overall nutritional advantage over refined wheat products. Thus, depending on the culture and the nature of the rice eaten, rice may have displacing less desirable foods. There could be other differences in chemical composition and physical properties which could contribute to the associations seen.”

Please enter your name: Nicola McKeown
Job Title: Scientist
Institution: JM USDA HNRCA at Tufts University
Reimbursement for attending a symposium?: Yes
A fee for speaking?: Yes
A fee for organising education?: No
Funds for research?: Yes
Funds for a member of staff?: No
Fees for consulting?: Yes
Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No
Reviewer: 4 BMJ Statistician

Recommendation:

Comments:
BMJ.55688: Associations of cereal grains intake with cardiovascular disease and mortality across 21 countries in the Prospective Urban and Rural Epidemiological Study: A prospective cohort study

Stats Report:

The present manuscript presents results from the PURE study on the association between white flour (referred to as refined grains), whole grains, and white rice with cardiovascular outcomes and all cause mortality. The data collected is impressive and unique, potentially providing high level evidence of this association.

Given that the data were collected across 21 countries, there are some issues regarding the analysis and the presentation of these data that I consider will need addressing to be able to properly evaluate these findings.

Major issue: Substantial heterogeneity of exposure and possibly of outcomes across the different countries/regions.
According to Table 1, there are substantial differences in the levels of exposure (for the three exposures) across the regions. It is unclear if within the presented regions, there is some homogeneity within their countries. The creates a critical problem. The use of frailty models, although in principle accounts for potential differences in hazards by centre, has as a basic assumption the independence between frailty and the included covariates (exposures). This is untestable, however, given that for some of the observed categories of the exposure, most of the data come from a single region (if not country), this is likely to represent a substantial confounding issue. If this is the case, the analysis as presented would be invalid.

I can think of two potential alternatives (the authors are likely to imagine others): a) to carry out stratified analyses by country using quantiles or exposure as continuous and obtain a pooled estimate of the HR's (based on meta-analysis ) or b) to standardise intake within region/centre so that it is possible to carry out a one-step multivariable analysis based on similar (relative) exposures. It might be possible from this to then back-transform to specific (absolute) exposure levels.

Response:
We too were concerned regarding the assumption of independence between the frailty variable (center) and exposure of interest for the three exposures. We observed that there was broad similarity of intakes between centers within each geographic region.
We performed meta-analysis of region-wise estimates of hazard ratios adjusting for all covariates in the full model, for each grain type and the results were consistent with the Cox-frailty full model. The forest plot of the analysis is given below.
Africa region could not be included in this analysis as the sample size was insufficient to estimate the hazard ratio within the region. Due to the consistency in results and the inability to include the entire sample in the meta analysis, we hope the reviewer would agree to us retaining the original Cox-frailty model in the paper as primary analysis. We have mentioned this analysis in the methods section and the consistency in results between the two analysis within the results section as follows “The meta-analysis of region specific effects were consistent with the findings of the overall Cox frailty model described above”. (Page 16, Line 19)

We also created quantiles of refined grain intake by country and analysed the association with events, the association was similar to that of the results obtained from the current analyses using the frailty model (Supplementary Table 9).

The distribution of intake of refined grains was similar in all regions except Africa and South Asia which had lower intake. We have now performed stratified analyses separately in regions with higher and lower intakes of refined grains to examine the association of intake and events and the results were consistent (Figure 1, Supplementary tables 8a,8b). These analyses reassures us that our results are not distorted by the correlations between frailty variable and intake.

Regardless of approach used for the analysis, an exploration of country and regional differences is required given the substantial differences in exposure. Please provide these either as a Tables or as Figures (forest plots). The variation observed by country is likely to be of substantial interest. This is critical for our understanding of the perceived differences in socio-economic and health development issues in the countries studied.

Response:
The distribution of intakes and outcomes by region is provided in Table 1 of the manuscript. The participant countries are classified into regions with reasonable homogeneity of exposures. For rice
intake, which was the most region-specific food, we have reported the results by 2 strata, Asian countries (combining the regions south Asia, south east Asia and China with high consumption) and non-Asian countries (lower consumption) (Figure 2, Supplementary table S12a, S12b). There was no association of events with rice intake in any of the two strata, similar to the whole sample analysis (Table 4). Similarly, we performed a stratified analysis of high and low intake regions for refined grain (Figure 1). The high intake regions were combined for all regions other than Africa and South Asia. These data are also presented in Supplementary table (Tables S8a and S8b). The results were comparable to that of the whole sample (Table 2). For whole grains, we performed an analysis without the Africa region which was the only region with high intake, as the sample size in this region alone was not sufficient for a separate stratified analysis. The result of this sensitivity analysis is presented in Supplementary Table S10 and the results were comparable to the whole sample analysis.

Major issue: Choice of categories for exposures
Related to the point above, the current exposure categories used cover the range of response across all participating countries. However, there appears to be in some instances countries/regions with minimal numbers in these categories. For example, based on Tables 1 and 2, it would appear to me that most (all?) of the >=350g/d group comes from China. The use of different number of categories for the different exposures (four for whole grains and five for the other two) also makes comparisons across exposures difficult.

Response:
We have now added cubic spline figures (Figure 3) with knots at 50 grams increased intakes for all foods. This facilitates the comparison between groups. These figures are also in response to a comment by another reviewer regarding the possible non-linear nature of the association. However, we have chosen to retain the purposive grouping of the foods as these translate to serving sizes that clearly differentiate high and low intake levels.

Major issue: Full reporting of regional differences for all outcomes
In the current version of the manuscript, stratified analyses by region are provided only for white rice exposure. Given that the manuscript explores the association for three exposures, please report (in the supplementary material) the same stratified analyses for the white flour and whole grains.

Response:
We have now included the stratified analysis of refined grain has been included as refined grain intakes in high and low intake regions (Figure 1, Supplementary table S8a and S8b). For whole grain, as the African region was the only region with high intake. With sample size in this region alone was not sufficient for a separate stratified analysis, a sensitivity analysis was done (Supplementary table S10).

Please also provide information on the total number of events per region in Table 1.

Response:
This has been added to Table 1 as instructed by the reviewer.

Major issue: choice of covariates in the model
Based on the reported results, the choice of covariates has a substantial impact on the findings. In particular, please clarify why the ‘Minimally adjusted model’ includes fruit and vegetable intake and
diary intake. There could be issues of collider bias created by the inclusion of some of the covariates into the model. Clarify which model was regarded as the primary analysis and why.

Response:
The minimally adjusted model includes location (urban/rural), wealth index, education, smoking, waist-hip ratio, physical activity, history of diabetes, daily intakes of energy (kcal), vegetable and fruits, red meats with centre as a random effect in addition to age and sex. Dairy was not included although mentioned in the earlier manuscript submitted. Fruits and vegetables have been included as they have been associated with CVD mortality in the PURE study and because fruits and vegetables contain fibre. These are covariates that have been previously identified as risk factors with at least one of the events listed. These are potential confounders in the association of the foods with events. The primary analysis is based on the fully adjusted model which has age and sex with centre as a random effect in addition to the food of interest. Age and sex are definite covariates for the outcome events and a model normally includes these covariates as CVD mortality increases with age and sex differences in mortality exist.

Minor issue: Definition of refined grains
The definition of refined grains used in the study is at odds with common understanding of this term. For example, from the US Department of Agriculture website:
‘Some examples of refined grain products are white flour, de-germed cornmeal, white bread, and white rice.’ Therefore, the term used needs to be justified or clarified to focus on white flour only.

Response:
We prefer to continue using the term refined grain as ready-to-eat corn was also included. White rice and rice products alone were not included.

Please enter your name: Rafael Perera
Job Title: Professor of Medical Statistics
Institution: University of Oxford
Reimbursement for attending a symposium?: No
A fee for speaking?: No
A fee for organising education?: No
Funds for research?: No
Funds for a member of staff?: No
Fees for consulting?: No
Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No
Do you hold any stocks or shares in an organization that may in any way gain or lose financially from the publication of this paper?: No