



Inserm



INRA
SCIENCE & IMPACT

le cnam



Editor, Dr Daoxin Yin,
The British Medical Journal
October 25th, 2017

Dear Dr. Yin,

Please find enclosed **the revised version** of our manuscript BMJ.2017.040538 titled **'Consumption of ultra-processed foods and cancer risk: results from the NutriNet-Santé prospective cohort'**.

We would like to sincerely thank you, the Editorial team, and all the reviewers for your valuable comments and for the opportunity that you gave us to revise this manuscript. All comments from the Editors and the Reviewers have been carefully taken into account. Changes have been highlighted throughout the text (red font). Also, please find enclosed the Editor and Reviewers' comments followed by our point-by-point responses.

We have included additional sensitivity analyses as recommended, we have also detailed several points in the manuscript and have added four appendixes (online supplemental data). Main results remain very consistent across all these further analyses, which strengthens the robustness of our findings.

Please note that, as agreed between all co-authors, we slightly modified the authorship for this paper: M. Fiolet and Dr. Srour are now co-first authors with equal contribution. Indeed, Dr Srour (corresponding author) has been largely involved since the beginning of the study and has been in charge of performing the present revision (statistical analyses, responses to the reviewers and manuscript modification), under the supervision of Dr Touvier, and with the assistant of all co-authors.

We hope that our modifications adequately address the Editor and Reviewers' comments and that our paper is now suitable for publication in The BMJ. We thank you again for your consideration.

Yours sincerely,

Dr Bernard Srour, PharmD
(Corresponding author)

Sorbonne Paris Cité, Epidemiology and Statistics Research Center, Nutritional Epidemiology Research Team (EREN), Inserm U1153; Inra U1125; Cnam; Paris 13, 7 and 5 Universities,
SMBH Paris 13, 74, rue Marcel Cachin, 93017 Bobigny Cedex, France
Tel: +33.148.388.968 ; Fax: +33.148.388.931 ; e-mail: b.srou@eren.smbh.univ-paris13.fr

04-Oct-2017

Dear Dr. Srouf

Manuscript ID BMJ.2017.040538 entitled "Consumption of ultra-processed foods and cancer risk: results from the NutriNet-Santé prospective cohort"

Thank you for sending us your paper. We sent it for external peer review and discussed it at our manuscript committee meeting. We recognise its potential importance and relevance to general medical readers, but I am afraid that we have not yet been able to reach a final decision on it because several important aspects of the work still need clarifying.

We hope very much that you will be willing and able to revise your paper as explained below in the report from the manuscript meeting, so that we will be in a better position to understand your study and decide whether the BMJ is the right journal for it. We are looking forward to reading the revised version and, we hope, reaching a decision.

Daoxin Yin
dyin@bmj.com

*** PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. ***

https://mc.manuscriptcentral.com/bmj?URL_MASK=93a76e3085eb47f58cb88f39fe54d35f

Report from The BMJ's manuscript committee meeting

These comments are an attempt to summarise the discussions at the manuscript meeting. They are not an exact transcript.

Members of the committee were: Elizabeth Loder (chair), Julie Morris (stats), Wim Weber, Jose Merino, Georg Roggla, Tiago Villanueva, Daoxin Yin, Rubin Minhas, Sophie Cook

Decision: Put points

Detailed comments from the meeting:

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

+++We would like to sincerely thank the Editors and Reviewers for their valuable comments and suggestions. All of them have been carefully taken into account. We believe that they truly contribute to improve the quality, the overall presentation and the clarity of our manuscript. Below, you will find a point by point description of how each comment was addressed in the manuscript.

Please also respond to these additional comments by the committee:

Our statistician made the following comments regarding a number of statistical issues that need to be addressed by the authors:

1. The study group is a very select cohort of subjects who are willing, and able, to provide detailed food consumption data every six months. This may introduce bias in various ways (not just to 'weaken' the observed associations, as noted by the researchers), and this should be acknowledged.

+++Large-scale population-based cohorts worldwide (including the most prestigious ones such as the Nurses' Health Study or the Health Professional Follow Up Study coordinated by Harvard, USA) involve volunteers who agreed to participate to long-term studies on lifestyle factors and health. By definition, recruited participants are not representative of the general population of the country. This is not specific to the NutriNet-Santé cohort. Although in such etiological studies representativeness is less crucial than diversity of exposure among studied populations, we agree that this may introduce some bias in HR estimates that should be discussed. Compared to the general French population, participants to the NutriNet-Santé cohort are more often women, with health-conscious behaviours and higher socio-professional and educational levels, as compared to the general French population¹. This might limit the generalizability of the findings and may have resulted in 1) a lower cancer incidence compared to national estimates (age and sex standardized incidence rate per 100,000 persons per year: 786 cases in our cohort vs 972 cases in France²) and 2) an overall lower exposure to ultra-

processed foods, with less contrast between extreme categories. These points rather tended to underestimate the strength of the associations. However, the possibility that selection bias may have led to an overestimation of some associations cannot be totally excluded. We have added some discussion about this point in the limits section of the manuscript, pages 20-21.

It should be noted that although there was a higher proportion of women than men in this cohort, the contribution of ultra-processed foods to the overall diet was very similar between men and women (18.74% for men and 18.71% for women, $p=0.7$). This information has been added to the manuscript page 11.

2. More information on the definition of ultra-processed food should be provided. For example, a table showing examples of ultra-processed vs non ultra-processed food. The mean % ultra-processed food consumption in this cohort was 18.7% (Table 1). It would be interesting to know what is the estimated average % ultra-processed food in the 'normal' population?

+++We have added an appendix (Appendix 1) that provides detailed explanation on the definition of ultra-processed foods, examples, and method of classification based on the published NOVA classification³⁻⁵. The present study is pioneer in France and the evaluation of the contribution of ultra-processed food to the diet in the framework of a nationally representative survey has not been performed yet. The only available information for the moment at the national level in France relies on the INCA3 survey⁶ (conducted by ANSES, the French Food safety Agency in 2016). This study does not quantify strictly speaking the proportion of ultra-processed foods in the diet, however, the following figures are interesting and can be cited: In the INCA3 study, "transformed" foods included sweet pastries, biscuits, dairy desserts, ice cream, fruit purée and fruit in syrup, fruit and vegetable juices, soups and broths, sandwiches, pizzas and salted pastries, as well as mixed dishes composed of egg, meat, fish, vegetable and/or starchy foods (cereals, legumes or potatoes). More than half of the "transformed" foods consumed outside catering establishments by adults aged 18-79 were manufactured industrially, and more than one-third were homemade, while the rest was handcrafted (e.g. caterer). These figures illustrate the important share of processed - and especially industrially processed - foods in the diet of French adults. These elements have been added in the discussion, page 18.

3. The mean number of dietary records per subject over the first two years of follow-up is reported to be 5.4 (Page 9). How does this relate to the fact that three 24-hour dietary records were made every six months? What was the extent of missing dietary information?

+++Participants are invited to complete three 24-hour food records every 6 months. However, to be included in the nutrition component of the NutriNet-Santé cohort, only two dietary records are mandatory. Subjects are not excluded if they do not complete all optional questionnaires. As a result, the minimum number of 24-hour records per subject in the first two years of follow-up in this study was 2 (but it only represented 7.2% of the participants), the maximum was 12, and the mean was 5.4. After the launching of the study by the end of May 2009, 49.6% of the records were filled between June and November and 50.4% between December and May. These points have been clarified in the methods, page 7 and the results page 11.

An average of 5 dietary records is greater than most of other nutritional epidemiology studies using 24h records or recalls. Of note, most cohorts are based on Food Frequency Questionnaires which provide interesting insights into usual dietary exposure but which are not very adapted in this particular case to properly discriminate foods according to their degree of processing (indeed, FFQ are often based on 100-200 food groups while >3300 food items were distinguished in the 24h dietary records used in NutriNet).

In nutritional epidemiology, it is usually a matter of finding a balance between a high number of records per subject (guaranteeing a better accuracy of the data but, in return, a higher selection bias towards a very compliant population) or conversely a smaller number of dietary records (corresponding to a lower degree of precision but also a lower selection bias compared to the general population). There is no perfect answer. For the main model, we chose to include all subjects with at least two dietary records (this minimum number being necessary to apply the classical Black criterion method for dietary underreporting identification), and to adjust all models for the number of dietary records per subject.

However, as requested by the Reviewers, we have now included sensitivity analyzes to demonstrate the stability of the results: i) if we included only participants with at least six 24h records on the one hand (overall cancer risk: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.13 (1.06-1.21)$), $P\text{-trend}=0.0003$, $n = 1494$ cases and 47 920 non-cases included) and ii) if we re-included participants with only one 24h record on the other hand (overall cancer risk: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.11 (1.06-1.16)$), $P=0.0001$, $n = 2383$ cases and 122 196 non-cases included). This stability of our results is certainly due to the fact that while a relatively high number of 24h records is needed to properly evaluate intakes of very specific micro or macro-nutrients, the tendency of an individual to consume ultra-processed foods is likely "captured" with a relatively limited number of days of dietary records. Results of these sensitivity analyzes were added in the manuscript page 17.

4. The statistical methodology used is appropriate, but it is not clear why quartiles of food consumption were selected? Why not quintiles, or the exact % of food consumption?

+++The exact % of ultra-processed food consumption was used in fact. It corresponds to the HR for a 10-point increment in the proportion of ultra-processed foods in the diet. In addition to this continuous coding, we used quartiles (to categorize the ultra-processed food variable and show the trends of the associations) because it represented the best compromise between a sufficient number of categories and a sufficient number of cases by categories (even in less represented cancer locations). However, main results were similar when quintiles were used instead of quartiles:

	cases / non-cases	HR	lower	Upper	P-trend
Overall cancer quintiles					
Q 1	557/20438	1	-	-	0.001
Q 2	541/20455	1.097	0.974	1.235	
Q 3	443/20554	1.042	0.919	1.182	
Q 4	417/20579	1.199	1.054	1.363	
Q 5	270/20726	1.253	1.078	1.456	

	cases / non-cases	HR	lower	Upper	P-trend
Breast cancer quintiles					0.03
Q 1	191/16240	1	-	-	
Q 2	174/16258	1.045	0.850	1.285	
Q 3	152/16280	1.079	0.870	1.339	
Q 4	138/16294	1.236	0.989	1.544	
Q 5	84/16348	1.248	0.958	1.625	

We have added the results of this sensitivity analysis in the manuscript, page 17.

5. Table 1. The p-value reported for ‘ultra-processed’ food should be omitted here (differences between quartiles are a consequence of their definition).

+++This p-value has been deleted in Table 1, page 13.

6. There are many differences in population characteristics between subjects in the various quartiles of consumption (Table 1). Several adjusting factors were employed in the statistical analysis, however it is very difficult to adjust for all factors related to ‘behavioural’ aspects. Thus, it would seem more appropriate to give a greater emphasis on the tentative nature of the study findings.

+++We agree with this point: this is an observational study, thus, residual confounding can never be entirely excluded. To limit this bias, a large number of potential confounders have been taken into account in the models. In addition, several additional sensitivity analyses (testing further adjustments and/or stratifications) are presented below in this review and illustrate the stability of the results. More generally, the epidemiological approach can greatly contribute to advancing knowledge about the impact of ultra-processed food intake on health. Randomized controlled trials have long been considered the only gold standard. However, this type of study does not capture consumption as it is “in real life” and, more important here, it is not ethically feasible to investigate exposures for which a deleterious effect is suspected. Our large observational cohort study is therefore particularly adapted to provide insights in this field. We have added some discussion about this limitation page 21, as well as a sentence in the conclusion, page 22.

Editors thought the research question is very interesting and the study would have public health implications, and they also made the following comments:

* Please be sure to respond to the comments made by the statistician and reviewers.

* Editors noticed the authors gave a brief description of “ultra-processed food” in “MATERIAL AND METHODS,” and they thought the authors could better describe the definition of “ultra-processed food.” What is the difference between processed food and ultra-processed food? The paper would benefit from a table listing common ultra-processed foods as examples.

+++We sincerely thank the Editors for this overall positive comment. All comments have been taken into account, as explained below. As explained above in our response to the Statistical Reviewer, point 2, we have now added an appendix (Appendix 1) that provides detailed explanation on the definition of ultra-processed foods, examples, and method of classification based on the published NOVA classification³⁻⁵.

* The gender distribution is very skewed (80% women). In “DISCUSSION,” the authors highlighted the limitations of the study, such as “overrepresented women” and some other characteristics higher than the average level of the general French population, please discuss how it would affect the interpretation of the findings.

+++Please refer to our response to the Statistical Reviewer, point 1. The manuscript has been modified accordingly (result section, page 11, and discussion section, page 20).

* Please provide explanation and justification regarding the ascertainment of cancer cases. This concern was shared by some reviewers and editors.

+++We apologize if this point was not clear in the initial version of the manuscript. We have added detailed information on the ascertainment of cancer cases in the Methods section pages 8-9. Please see below our response to Reviewer 2, “case ascertainment - point 1”.

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.

Comments from Reviewers

Reviewer: 1

Recommendation:

Comments:

Relevance - Study aims

I was a carer for my husband who died of bowel cancer in less than 2 years of being diagnosed of the disease, and I am also a breast cancer patient on remission for 10 years. We as a family have always believed in eating healthily and taking good care of our body; we never smoke, drink alcohol, nor eating ultra-processed food. However, both my husband and I found ourselves brushed with the dread word of cancer. As a result of our experiences, I have become very interested in how cancer comes about, and how to prevent it to happen.

This study aimed to assessing the prospective association between ultra-processed food and cancer risk, which is important to me as a patient on remission in trying to understand cancer. I also believe this study is important to the general public (carer or non-carer), the health care professionals, and the government who overseeing the health care budget, research and policies for the nation.

Areas that are missing - data collection

Although the participants completed a set of 5 questionnaires related to:

1. socio-demographic and lifestyle characteristics (e.g. date of birth, sex, occupation, educational level, smoking status, number of children),
2. anthropometry 36;37 (e.g. height, weight),
3. dietary intakes(every 6 months),
4. activity (validated 7-day International Physical Activity Questionnaire [IPAQ]) 38,
5. and health status (e.g. personal and family history of diseases, medication use including use of hormonal treatment for menopause, oral contraceptive, and menopausal status),

but emotional well-being has not been included.

Various research suggested stress plays an important part in our health and could lead to disease. A recent TV series broadcasted by the BBC titled ‘A doctor in the house’, showed how the doctor, a GP, able to help suffering family reverts illness by taking a holistic approach in the care plan; that is, in addition to specialist treatment, he helped the family to understand and change their diet, physical exercises, and relief stress.

+++We wanted to testify to Mrs Cheung (Reviewer 1) our deepest empathy and salute her for her courage and her involvement as a patient in the review process of biomedical research articles. This input is important for researchers.

We did not have a variable measuring strictly the emotional wellbeing in the cohort. However, we performed a sensitivity analysis by adjusting the models for depression at baseline. Results remained statistically significant after this adjustment (HR_{for a 10-point increment in the proportion of ultra-processed foods in the diet}=1.11 (1.06-1.17), P-trend<0.0001

for overall cancer and 1.11 (1.01-1.22), P-trend=0.02 for breast cancer). We have added these results in the manuscript page 18.

Measured outcome

his study concluded that a 10% increase in ultra-processed food consumption leads to 10% increase in cancer risk. This is a very significant finding not only to patients, but to the general public as a whole. It is important that this preliminary study will lead to further investigation.

Conclusion

Due to the significant finding of this study, I strongly believe this paper should be published on the BMJ. So that health care professionals, such as GP's, are aware of the important of this finding and be able to offer a more holistic approach in their patient care.

+++We sincerely thank Reviewer 1 for this positive message. Indeed, this large prospective cohort was the first to investigate and highlight an increase in cancer risk associated with ultra-processed food intake. We hope that a publication in the BMJ, a prestigious and broad-spectrum journal, could have an impact: i) on medical researchers, to stimulate the setting of similar studies to confirm these results in different populations and contexts, ii) on general practitioners and other health care professionals, who are key information providers, to improve the prevention messages they deliver to their patients, and iii) on patients and healthy citizens to increase their awareness about potential risks associated with high intakes of ultra-processed foods.

Additional Questions:

Please enter your name: Kin Cheung

Job Title: patient reviewer

Reviewer: 2

Recommendation:

Comments:

I have reviewed with great interest the manuscript; is a document mostly well written. The article is interesting and relevance to the field of this journal and is importance to general medical readers, and scientific reliability. However, the authors might consider some minor recommendations:

+++We thank Reviewer 2 for this positive comment.

INTRODUCTION

line 27-29, these sentences need correction and Language editing is required.

+++This sentence has been corrected, page 5.

Material and Methods

Data collection

1. line 7, since dietary supplements effect on overall daily nutrient intake, how did you assess use of multivitamin and selected individual dietary supplements?

Did you consider it in analyzes?

+++This study did not use any variable related to vitamin or mineral intake (from foods or dietary supplements). Thus dietary supplement use is not included in the main analysis. However, in order to comply with the request of Reviewer 2, we have performed a sensitivity analysis by adjusting for dietary supplement use (yes/no) at baseline since it may represent an adjustment for an overall profile of subjects. This further adjustment did not modify the findings: HR_{for a 10-point increment in the proportion of ultra-processed foods in the diet}=1.12 (1.06-1.17), P-trend<0.0001 for overall cancer and 1.11 (1.02-1.22), P-trend=0.02 for breast cancer. We have added these results in the manuscript page 18.

2. line 12, place space between "periodfor" and correct it to "period for"

+++We apologize for this typo. It has been corrected in the revised version, page 7.

Case ascertainment

The participants are adequately described, on the whole.

1. However, it was unclear how this physician diagnose case event?

+++We apologize if this point was not clear in the initial version of the manuscript. We have elaborated the case ascertainment and validation strategy in the Methods section, pages 8-9:

Participants self-declared health events through the yearly health status questionnaire, through a specific check-up questionnaire for health events (every three months) or at any time through a specific interface on the study website. For each incident cancer declared, participants were contacted by a physician of the team and asked to provide any relevant medical records. Whenever necessary, the study physicians contacted the physician of the patient and/or hospitals to collect additional information. Afterwards, all medical data were reviewed by a physician expert committee. Medical records were obtained for >90% of cancer cases. Besides, our research team was the first in France to obtain the authorization by Decree in the Council of State (n°2013-175) to link data from our cohorts to medico-administrative databases of the National health insurance (SNIIRAM database). Declared health events were therefore completed by the information from these databases, thereby limiting any potential bias due to people with cancer who may not report their disease to the study investigators. Last, an additional linkage to the French National cause-specific mortality registry (CépiDC) is used to detect death and potentially missed cancer cases for deceased participants.

2. No information is available on how participants were followed for example if subjects change their address?

+++Participants are followed using an online platform connected to their email address. They have the possibility to change their email address, phone number or postal address at any moment on the NutriNet-Santé website. Newsletters and alerts about new questionnaires are sent via email. In case of an “undelivered email to recipient” problem, participants are then contacted by telephone and then by regular mail. We have added this information in the Methods section page 6.

Statistical analysis

line 25, Given reference (35) is not about IPAQ, please correct this reference to (38).

+++We apologize for this typo, this has been corrected, page 7.

Results

line 5, participants among the highest quartile of ultra-processed food intake tended to be younger,... Which group was reference?

+++The group of reference was the first quartile. We have clarified this point page 11.

Discussion

1. line 14, about reference 50: is it online website or article? reference should be corrected.

+++This reference has been corrected, page 29.

2. line 28-32 need references. please provide correct references. as well as, line 32-34, experimental studies require references.

+++We have added the corresponding references as requested, pages 19-20.

3. line 43, provide references.

+++We have added the corresponding references as requested, page 20.

Reference

1. reference 24: provide correct reference of website.

+++This reference has been corrected, page 27.

2. reference 46 is not correct according to journals format.

+++This reference has been corrected, page 28.

3. about reference 50: is it online website or article? Reference should be corrected

+++This reference has been corrected, page 29.

Recommendation: This manuscript is rigorous and provides a useful contribution to its area of research and should be published.

+++We sincerely thank Reviewer 2 for this positive comment.

Additional Questions:

Please enter your name: mahdiah

Job Title: Ph.D candidate of nutrition
Institution: Faculty of Nutrition and Food Sciences, Tabriz University of Medical Sciences

Reviewer: 3

Recommendation:

Comments:

Dr Andrea Darling

Thank you for inviting me to review this interesting and well written manuscript. There is little existing data in this area so the results of this paper are original and will have an impact of public health messages in terms of diet for cancer avoidance.

I think the main limitation of the study is how the cases were ascertained. This was done through self-report measures. Although validated by a physician and patient medical records there may be bias in that some people with cancer may not report this to the study investigators. Therefore some of the controls may actually have cancer. Although classification is never perfect in any study with medical data linkage this would be less of a problem as the quality of an individual's linked data (accuracy, coverage) is not likely to be linked to characteristics of the participant, unlike self-report. I think the authors need to add a sentence to the discussion regarding the limitations of their method of ascertaining cases.

+++As stated in our response to Reviewer 2, "case ascertainment - point 1":

We apologize if this point was not clear in the initial version of the manuscript. We have elaborated the case ascertainment and validation strategy in the Methods section, pages 8-9:

Participants self-declared health events through the yearly health status questionnaire, through a specific check-up questionnaire for health events (every three months) or at any time through a specific interface on the study website. For each incident cancer declared, participants were contacted by a physician of the team and asked to provide any relevant medical records. Whenever necessary, the study physicians contacted the physician of the patient and/or hospitals to collect additional information. Afterwards, all medical data were reviewed by a physician expert committee. Medical records were obtained for >90% of cancer cases. Besides, our research team was the first in France to obtain the authorization by Decree in the Council of State (n°2013-175) to link data from our cohorts to medico-administrative databases of the National health insurance (SNIIRAM database). Declared health events were therefore completed by the information from these databases, thereby limiting any potential bias due to people with cancer who may not report their disease to the study investigators. Last, an additional linkage to the French National cause-specific mortality registry (CépiDC) is used to detect death and potentially missed cancer cases for deceased participants.

Despite this multi-source strategy, the fact that some cancer cases may have been missed can never be ruled out. We have acknowledged this in the discussion section page 21.

However, otherwise the study design and methodology appears strong, the paper comprehensively covers all the methods used and presents the results well. Details of the participants are comprehensively given. It has a clear research question, which was appropriately answered. The results and conclusions are credible, clear and well presented. Most importantly it highlights an important message about the consumption of ultra processed food and cancer risk. This will be important to public health providers and policy makers, but is not as relevant for clinicians.

+++We sincerely thank Reviewer 3 for this positive comment. We agree that this paper may have a major impact on public health providers and policy makers, but we also think that it could have an important impact on clinicians. Physicians (general practitioners and specialists) are generally trusted by their patients and therefore, are potential key health information providers for the citizens (patients but also healthy individuals and their families). The public health messages they deliver to their patients in terms of nutritional and other lifestyle factors are strategic levers for public health improvement and chronic disease risk prevention.

I am an epidemiologist (not a food scientist) so I can't comment on the science/methods surrounding analysing ultra-processed foods content of diets.

Very minor comments

P6 L5 Can you clarify here whether the higher lipidaemia is prevalence or incidence (also for obesity and hypertension).

+++The few studies performed observed that ultra-processed food intake was associated with a higher incidence of dyslipidaemia in Brazilian children ⁷, and higher risks of overweight, obesity ⁸ and hypertension ⁹ in a prospective cohort of Spanish University students. We have clarified this in the introduction page 6.

P8 L40 You say that 'for physical activity a missing class was included in the analysis'. Can you just clarify why this was included, especially why it was included for physical activity but not the other confounding variables.

+++For all covariates except physical activity, the proportion of missing values was very low (<5%) and thus, they were imputed to the modal value (for categorical variables) or to the median (for continuous variables). Concerning physical activity, the proportion of missing values was higher (14%) since the answers to all IPAQ questions were needed to calculate the score. To avoid massive imputation for a non-negligible number of subjects or exclusion of subjects with missing data and risk of selection bias, we included a missing class in the models for this variable. This has been clarified in the Methods page 9. This strategy is often used in other large cohorts such as EPIC for instance ¹⁰⁻¹³.

P8 L56 What do you mean by ‘censored’? Do you mean excluded?

+++As recommended for site-specific analyses, cancers of other locations than the one studied were censored at the date of diagnosis. This means that they were considered as non-cases for the cancer of interest and they contributed person-year until the date of diagnosis of their cancer. We have explained this in the revised manuscript page 9.

Table1 There is a typo for premenopausal status. 728% should be 72.8%?

+++We apologize for this typo. It has been corrected, page 13.

Additional Questions:

Please enter your name: Andrea Darling

Job Title: Postdoctoral Researcher

Reviewer: 4

Recommendation:

Comments:

The manuscript is well-written and presented quite interesting results. The association between ultra-processed food and cancer incidence were significant and remain there in the sensitivity analysis by removing cases that occur within the first years of follow up. The authors have described clearly the limitation and strength of the manuscript. However, there are some improvement should be made to improve the quality of the manuscript.

+++We thank Reviewer 4 for this positive comment and for showing interest in our results.

Page 7: the author stated that “Usual dietary intakes were assessed every 6-months through a series considered as baseline usual dietary intakes in this prospective analysis.” It means that in a perfect assumption that there will have 12 days 24h-dietary records. However, in the results, the mean number of dietary records per subject over their first two years of follow-up was 5.4 (SD=2.9). It seems that more than half of the participants did not accomplished the half of the 12-days surveys, isn't it? There lacks of sufficient information about the of 24h-dietary records in the four times surveys during the two years. If most of individuals answered the questions only in the first half of the years or in second half of the years, the dietary information collected might may lacks of sufficient accuracy.

+++As explained in our response to the Statistical Reviewer, point 3:

Participants are invited to complete three 24-hour food records every 6 months. However, to be included in the nutrition component of the NutriNet-Santé cohort, only two dietary records are mandatory. Subjects are not excluded if they do not complete all optional questionnaires. As a result, the minimum number of 24-hour records per subject in the first two years of follow-up in this study was 2 (but it only represented 7.2% of the participants), the maximum was 12, and the mean was 5.4. After the launching of the study by the end of May 2009, 49.6% of the records were filled between June and November and 50.4% between December and May. These points have been clarified in the methods, page 7 and the results page 11.

An average of 5 dietary records is greater than most of other nutritional epidemiology studies using 24h records or recalls. Of note, most cohorts are based on Food Frequency Questionnaires which provide interesting insights into usual dietary exposure but which are not very adapted in this particular case to properly discriminate foods according to their degree of processing (indeed, FFQ are often based on 100-200 food groups while >3300 food items were distinguished in the 24h dietary records used in NutriNet).

In nutritional epidemiology, it is usually a matter of finding a balance between a high number of records per subject (guaranteeing a better accuracy of the data but, in return, a higher selection bias towards a very compliant population) or conversely a smaller number of dietary records (corresponding to a lower degree of precision but also a lower selection bias compared to the general population). There is no perfect answer. For the main model, we chose to include all subjects with at least two dietary records (this minimum number

being necessary to apply the classical Black criterion method for dietary underreporting identification), and to adjust all models for the number of dietary records per subject.

However, as requested by the Reviewers, we have now included sensitivity analyzes to demonstrate the stability of the results: i) if we included only participants with at least six 24h records on the one hand (overall cancer risk: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.13 (1.06-1.21)$), $P\text{-trend}=0.0003$, $n = 1494$ cases and 47 920 non-cases included) and ii) if we re-included participants with only one 24h record on the other hand (overall cancer risk: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.11 (1.06-1.16)$), $P=0.0001$, $n = 2383$ cases and 122 196 non-cases included). This stability of our results is certainly due to the fact that while a relatively high number of 24h records is needed to properly evaluate intakes of very specific micro or macro-nutrients, the tendency of an individual to consume ultra-processed foods is likely "captured" with a relatively limited number of days of dietary records. Results of these sensitivity analyzes were added in the manuscript page 17.

Page 8: The author said "Participants self-declared health events through the yearly health status questionnaire". So, it means that the outcome was self-reported. The participants in the cohort was invited to send their medical records the study physicians contacted the participants' treating physician or the medical structures in any necessary time. Certainly, medical records might provide a good approach to ascertain the outcome. However, how about with those who did not report the medical records but in fact had the interested outcomes, such as cancer in this study. If there is a higher proportion about this, the credibility of the results might be discounted largely. In addition, what's proportion of the cases who provided the medical records? Did all participants finish the yearly health status questionnaire?

+++As stated in our response to Reviewer 2, "case ascertainment - point 1" and Reviewer 3, point 1:

We apologize if this point was not clear in the initial version of the manuscript. We have elaborated the case ascertainment and validation strategy in the Methods section, pages 8-9: Participants self-declared health events through the yearly health status questionnaire, through a specific check-up questionnaire for health events (every three months) or at any time through a specific interface on the study website. For each incident cancer declared, participants were contacted by a physician of the team and asked to provide any relevant medical records. Whenever necessary, the study physicians contacted the physician of the patient and/or hospitals to collect additional information. Afterwards, all medical data were reviewed by a physician expert committee. Medical records were obtained for >90% of cancer cases. Besides, our research team was the first in France to obtain the authorization by Decree in the Council of State (n°2013-175) to link data from our cohorts to medico-administrative databases of the National health insurance (SNIIRAM database). Declared health events were therefore completed by the information from these databases, thereby limiting any potential bias due to people with cancer who may not report their disease to the study investigators. Last, an additional linkage to the French National cause-specific mortality registry (CépiDC) is used to detect death and potentially missed cancer cases for deceased participants. Despite this multi-source strategy, the fact that some cancer cases may have been missed can never be ruled out. We have acknowledged this in the discussion section page 21.

Page 9: The author established four different model to estimate the effect of ultra-processed food on the cancer incidence. The author adjusted for western dietary pattern in model 3 and model 4. Because the western pattern is a new generated variable, some more detail description is needed. Previous study also reported the fresh food and nutrients might contributed to the cancer incidence. Thought the author in this study considered the confounding effect from lipid intake, sodium intake, carbohydrate intake and western dietary pattern, the dietary effects from less ultra-processed food can not be balanced.

+++An appendix (n°2) presenting the method used to derive dietary patterns by principal component analysis as well as corresponding loading factors has been added pages 33-34.

We already adjusted for several factors reflecting a lower nutritional quality of the diet (such as the Western dietary pattern). In order to comply with Reviewer's 4 recommendations, we have now tested further adjustment for the PCA-derived healthy dietary pattern, and for fruit and vegetable intake. These further adjustments did not modify the findings (adjustment for the PCA-derived healthy dietary pattern: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}}=1.11 (1.05-1.17)$, $P\text{-trend}<0.0001$ for overall cancer and 1.10 (1.00-1.21), $P\text{-trend}=0.04$ for breast cancer ; adjustment for fruit and vegetable consumption in g/d: $HR=1.10 (1.04-1.16)$, $P\text{-trend}=0.0009$ for overall cancer and 1.11 (1.01-1.22), $P\text{-trend}=0.03$ for breast cancer). We have added these results page 18.

Page 10 Ultra-processed food is the major exposure in this study, the author provided the information of the Relative contribution of each food group to ultra-processed consumption in the diet. However, more information about the distribution among the participants are needed.

+++As requested, a histogram showing the distribution of the proportion of ultra-processed food in the diet (main exposure) was added as Appendix 5.

Page 10: Participants among the highest quartile of ultra-processed food intake tended to be younger. What's the age range of participants in this study? Age is a major confounder. A stratified analysis by age may be necessary. This study overrepresented women, hence, a stratified analysis by sex may be also necessary.

+++Age range (18.0-72.8) has been added in the results section page 11. All models were adjusted for age and sex. Models corresponding to sex-specific cancer locations were already stratified by sex, by definition. In addition, for breast cancer, stratification by pre- and post-menopausal status (which indirectly accounts for age) was already presented.

In addition, in order to comply with Reviewer's 4 request, we have performed stratified analyses by age and sex for the overall cancer cases model: results were statistically significant in all strata: in men (HR_{for a 10-point increment in the proportion of ultra-processed foods in the diet}=1.12 (1.02-1.24), P=0.02, 663 cases and 22158 non-cases), in women (HR= 1.13 (1.06-1.20), P<0.0001, 1565 cases and 80594 non-cases), in younger adults (<40 years old, HR= 1.21 (1.09-1.35), P=0.0006, 287 cases and 48627 non-cases) and in older adults (≥40 years old, HR= 1.09 (1.03-1.16), P=0.03, 1941 cases and 54485 non-cases) These findings have been added to the Results section page 14.

Additional Questions:

Please enter your name: LIU, XUDONG

Job Title: Research Associate

Institution: JC School of Public Health and Primary Care, Chinese University of Hong Kong

Reviewer: 5

Recommendation:

Comments:

Reviewer: Martin Lajous

In this paper the authors evaluate the relation between ultra-processed foods and total and site-specific cancer incidence. This is a timely topic that is complex that requires thoughtful evaluation of the challenges faced by epidemiologic studies in evaluating the health impact of food processing. Unfortunately, excitement for the paper is dampened by lack of detail on the description of the design and methods for data collection, limited information on the validity of exposure and outcome assessment, and a restrained discussion on the limitations of the study. Confounding is perhaps the biggest limitation of this study. Consumption of ultra-processed foods is related to behavioral risk factors for several cancers. Nevertheless, the authors should be commended for a creative study design that undercuts the limitations of new epidemiological studies on diet and health.

+++We would like to thank Reviewer 5 for his positive and constructive comments. All the points that he pointed out were addressed below. In particular, following his recommendations and the ones formulated by other Reviewers and Editors, we have added many details on the description of the design and methods, we have clarified our strategy for exposure and outcome assessment (including linkage with medico-administrative data bases of the national health insurance), and several points have been added in the discussion section about potential limitations of the study. Regarding confounding factors, many sensitivity analyses have been performed (all leading to similar results, thereby confirming the stability of the findings). All these points are described below.

Major comments

1. Validity of the dietary assessment. Usual dietary intake was assessed using computer-based 24-hr records. Investigators report that that participants were asked to record daily dietary intake for 3 days on a 2 week period every 6 months. More information on the validity of the assessment would reassure the reader. Was seasonality taken into account? Some participants appear to have less than 3 records. Is having one record every 6 months on three occasions equal to having 3 records on the same 2 week period? Also, did the timing of dietary assessment ever coincide with that of self-reported cancer? Were the assessments averaged? Individuals with 1 assessment were excluded suggest considering sensitivity analyses to determine the impact of individuals with fewer records.

+++Several of these comments were already addressed in our response to Statistical Reviewer, point 3:

Participants are invited to complete three 24-hour food records every 6 months. However, to be included in the nutrition component of the NutriNet-Santé cohort, only two dietary records are mandatory. Subjects are not excluded if they do not complete all optional questionnaires. As a result, the minimum number of 24-hour records per subject in the first two years of follow-up in this study was 2 (but it only represented 7.2% of the participants), the maximum was 12, and the mean was 5.4. After the launching of the study by the end of May 2009, 49.6% of the records were filled between June and November and 50.4% between December and May. These points have been clarified in the methods, page 7 and the results page 11.

An average of 5 dietary records is greater than most of other nutritional epidemiology studies using 24h records or recalls. Of note, most cohorts are based on Food Frequency Questionnaires which provide interesting insights into usual dietary exposure but which are not very adapted in this particular case to properly discriminate foods according to their degree of processing (indeed, FFQ are often based on 100-200 food groups while >3300 food items were distinguished in the 24h dietary records used in NutriNet).

In nutritional epidemiology, it is usually a matter of finding a balance between a high number of records per subject (guaranteeing a better accuracy of the data but, in return, a higher selection bias towards a very compliant population) or conversely a smaller number of dietary records (corresponding to a lower degree of precision but also a lower selection bias compared to the general population). There is no perfect answer. For the main model, we chose to include all subjects with at least two dietary records (this minimum number being necessary to apply the classical Black criterion method for dietary underreporting identification), and to adjust all models for the number of dietary records per subject.

However, as requested by the Reviewers, we have now included sensitivity analyses to demonstrate the stability of the results: i) if we included only participants with at least six 24h records on the one hand (overall cancer risk: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.13 (1.06-1.21)$), $P\text{-trend}=0.0003$, $n = 1494$ cases and 47 920 non-cases included) and ii) if we re-included participants with only one 24h record on the other hand (overall cancer risk: $HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.11 (1.06-1.16)$), $P=0.0001$, $n = 2383$ cases and 122 196 non-cases included). This stability of our results is certainly due to the fact that while a relatively high number of 24h records is needed to properly evaluate intakes of very specific micro or macro-nutrients, the tendency of an individual to consume ultra-processed foods is likely "captured" with a relatively limited number of days of dietary records. Results of these sensitivity analyses were added in the manuscript page 17.

In addition, please find below complementary information:

The NutriNet-Santé web-based self-administered 24h dietary records have been tested and validated before the launching of the cohort and corresponding results were published. In one study, we compared the web-based 24h record with an interview by a trained dietitian¹⁴, and in a second study, dietary intakes from these web-based 24h records were validated against blood and urinary biomarkers¹⁵. This has been clarified in the manuscript page 7.

Participants were asked to fill dietary records every 6 months, in order to vary the season of completion. Besides, we performed a sensitivity analysis by adjusting for the season of inclusion in the cohort (corresponding to the season of baseline dietary records for each participant), which did not modify the findings ($HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.12 (1.06-1.18)$, $P<0.0001$ for overall cancer and 1.12 (1.02-1.22), $P=0.02$ for breast cancer). We have added this information in the method (page 7) and result (pages 11 and 18) sections.

As Reviewer 5 properly understood, mean dietary intakes from all the 24h-dietary records available during the first two years of each participant's follow-up were averaged. This has been clarified in the manuscript page 7. Thus having 1 record on 3 different occasions contributed equally to the mean as 3 records on the same 2 week period. However, as shown above, our results were robust when we tested variations in the number of included dietary records and when performed adjustment for the season of baseline dietary records.

During the first two years of follow-up, the timing of dietary assessment coincided with that of self-reported cancer. In order to test for potential reverse causality bias, we had previously verified that results were unchanged when we excluded cancer cases diagnosed during the first year of follow-up. We have now tested exclusion of cancer cases diagnosed during the first two years of follow-up and results were also similar ($HR_{\text{for a 10-point increment in the proportion of ultra-processed foods in the diet}} = 1.10 (1.03, 1.17)$, $P\text{-trend}=0.005$ for overall cancer risk; $HR=1.15 (1.03, 1.29)$, $P\text{-trend}=0.02$ for breast cancer risk). This has been added to the results section page 17.

2. Definition of the exposure. The exposure is the proportion of ultra-processed foods in the diet (processed foods in weight/total food intake in weight) including beverages and solid foods. More detail should be provided on what were the main food components of this composite measure. More detail should be given to the reader of what this composite exposure means. In terms of public health advice, individuals would want to know which foods to avoid. Also, I suggest evaluating the main contributors to the exposure independently to assess the presence of a food item that could be exerting a particularly important impact on cancer risk.

+++As explained in our response to Statistical Reviewer point 2, we have added an appendix (Appendix 1) that provides detailed explanation on the definition of ultra-processed foods, examples, and method of classification based on the published NOVA classification³⁻⁵.

Besides, the food groups contributing to ultra-processed food intake in this study are presented in Figure 1. Main contributors were sugary products (26%) and beverages (20%), followed by starchy foods and breakfast cereals (16%) and processed fruits and vegetables (15%). We have added this information in the results section page 11.

As requested, we have now assessed the associations between the quantity of ultra-processed food from each contributing food group and cancer risk. Results are presented in Appendix 6 and in the manuscript page 17: ultra-processed fats and sauces (P=0.002), sugary products (P=0.03), and beverages (P=0.005) were associated with increased overall cancer risk and ultra-processed sugary products were associated with increased breast cancer risk (P=0.006).

3. Outcome assessment. Participants self-reported cancer cases that were then validated via record review. Were all cancer cases validated? What happened to cases that were not validated? Is the cancer incidence in this cohort what would be expected for the age distribution? No detail is provided on mortality follow-up. How was this performed? Were there any cases identified in mortality follow-up? What was the follow-up rate? Acknowledging that some cases may have been missed and pointing out how this may affect the interpretation of results would be useful to the reader.

+++Several of these comments were already addressed in our response to Reviewer 2, “case ascertainment - point 1”:

We apologize if this point was not clear in the initial version of the manuscript. We have elaborated the case ascertainment and validation strategy in the Methods section, pages 8-9:

Participants self-declared health events through the yearly health status questionnaire, through a specific check-up questionnaire for health events (every three months) or at any time through a specific interface on the study website. For each incident cancer declared, participants were contacted by a physician of the team and asked to provide any relevant medical records. Whenever necessary, the study physicians contacted the physician of the patient and/or hospitals to collect additional information. Afterwards, all medical data were reviewed by a physician expert committee. Medical records were obtained for >90% of cancer cases. Besides, our research team was the first in France to obtain the authorization by Decree in the Council of State (n°2013-175) to link data from our cohorts to medico-administrative databases of the National health insurance (SNIIRAM database). Declared health events were therefore completed by the information from these databases, thereby limiting any potential bias due to people with cancer who may not report their disease to the study investigators. Last, an additional linkage to the French National cause-specific mortality registry (CépiDC) is used to detect death and potentially missed cancer cases for deceased participants. Despite this multi-source strategy, the fact that some cancer cases may have been missed can never be ruled out. We have acknowledged this in the discussion section page 21.

In addition, please find below complementary information:

Medical records were obtained for >90% of self-reported cancer cases so far. As done in some other large cohorts¹³ and because of the high validity of self-reports (95% of self-reported cancers for whom a medical record was obtained were confirmed by our physicians), we included all cases who self-reported incident cancers, unless they were identified as non-case subjects by a pathology report. In the latter situation, cases that were not validated due to a wrong self-report and after checking in the SNIIRAM databases of the national health insurance were classified as non-cases. We have clarified this point in the methods, pages 8-9. Results were unchanged when non-validated cancer cases were excluded (HR for a 10-point increment in the

proportion of ultra-processed foods in the diet=1.11 (1.05-1.17), P=0.0003 for overall cancer risk, n=1967 cases and 102752 non-cases included; HR=1.12 (1.02-1.23), P=0.02 for breast cancer risk, n=677 cases and 81274 non-cases included). This analysis has been added page 17.

As it is generally the case in volunteer-based cohorts, participants to the NutriNet-Santé cohort were more often women, with health-conscious behaviours and higher socio-professional and educational levels as compared to the general French population¹. This might have resulted in 1) a lower cancer incidence compared to national estimates (age and sex standardized incidence rate per 100,000 persons per year: 786 cases in our cohort vs 972 cases in France²) and 2) an overall lower exposure to ultra-processed foods, with less contrast between extreme categories. These points rather tended to underestimate the strength of the associations. However, the possibility that selection bias may have led to an overestimation of some associations cannot be totally excluded. This has been acknowledged and discussed in the manuscript page 21.

Regarding mortality follow-up, information about vital status and causes of death were obtained through the exhaustive national death registry (Inserm-CépiDC). 108 cases (4.8%) among the cancer cases of this sample were identified during mortality follow-up. This information has been added page 14.

To date, the abandon rate in the NutriNet-Santé cohort is 6.7%. This information has been added page 14.

4. Confounding. Perhaps the most important challenge for this analysis is the possibility of residual confounding. Several strong risk factors for cancer (i.e. smoking, physical inactivity) appear to be associated to consumption of ultra-processed foods. I would suggest aiming at lowering as much as possible the risk for residual confounding (e.g. estimating cigarette years and/or conducting analyses among non-smokers only). In

addition to conducting analyses, the authors should discuss at length the possibility of missmeasured confounders in the discussion section. Do the authors believe that the dietary assessment allows for a good estimation of sodium intake? Please provide details to support this.

+++We have tested further adjustment for the number of pack-years of cigarettes, this did not modify the findings: (HR_{for a 10-point increment in the proportion of ultra-processed foods in the diet} =1.13 (1.07-1.019), P<0.0001 for overall cancer and 1.13 (1.03-1.24), P=0.009 for breast cancer). This information has been added to the results section page 18. As mentioned in the same section of the revised manuscript, further adjustment was also tested for dietary supplement use at baseline, prevalent depression at baseline, PCA-derived healthy dietary pattern, fruit and vegetable consumption in g/d and season of inclusion in the cohort. None of these further adjustments modified the findings.

As requested, we have also run overall cancer risk models stratified according to baseline smoking status and level of physical activity: results were statistically significant in all strata investigated: in smokers (including adjustment for pack-years of cigarette smoked, HR =1.18 (1.04-1.33), P=0.01, 255 cases and 15355 non-cases), in non-smokers (HR=1.11 (1.05-1.17), P=0.0002, 1943 cases and 85219 non-cases), in subjects with low-to-moderate levels of physical activity (HR=1.07 (1.00-1.15), P=0.04, 1216 cases and 59546 non-cases), and in those with a high level of physical activity (HR=1.19 (1.09-1.30), P<0.0001, 744 cases and 28859 non-cases). We have added this information in the manuscript page 14.

More generally, we agree with Reviewer 5: this is an observational study, thus, residual confounding can never be entirely excluded. To limit this bias, a large number of potential confounders have been taken into account in the models. In addition, several sensitivity analyses (testing further adjustments and/or stratifications) are presented in this review and illustrate the stability of the results. Residual confounding may be related to unmeasured confounders, but also to imprecision in the measure of included confounders. For instance in breast cancer models, oral contraception was a binary variable, since the precise doses, type and duration of contraceptive use across reproductive life were not available. A paragraph acknowledging and discussing the possible residual confounding has been added in the limits section of the revised manuscript page 21.

Sodium intake is particularly complex to measure due to multiple sources (native sodium in foods, salt added during the cooking, salt added in the plate). We have developed a specific module in the NutriNet-Santé 24h dietary records that specifically aims at estimating salt intake. It has been validated against sodium urinary excretion biomarkers¹⁵. This has been added to the Methods section page 7.

5. Various issues. Please provide more detail regarding the mediation. These analyses are complex and require several sometimes untenable assumptions. The reader is left wondering if the assertions made by the investigators regarding mediation are warranted. I would suggest separating fat into different types of fat. As the authors are aware the relation of fatty acids with different health outcomes including cancer appears to differ by type. Adjusting for Western dietary pattern seems odd in that the correlation between the exposure and this cofounder that is derived from the exposure may be strong and the resulting estimate difficult to interpret. Why were quartiles chosen? Using quintiles may have provided more insight on the nature of the relation. Please comment on the average follow-up time and how this affects the interpretation of results.

+++We have added an appendix and a figure providing a detailed explanation of the mediation analysis according to the method proposed by Lange et al.¹⁶ (Appendix 3).

As requested, we performed additional mediation analyses considering the different types of fatty acids: saturated, mono-unsaturated and poly-unsaturated. None of these factors showed significant indirect effects in the mediation analysis (all p>0.8). We have added this information in the methods page 10 and in the results page 17.

The concepts of ultra-processed food and Western dietary pattern are definitely different and the correlation between the two variables was very limited (Pearson correlation coefficient = 0.06). This information has been added in the manuscript page 17. Indeed, the Western dietary pattern was derived by principal component analysis on 20 major food groups (as now detailed in Appendix 2) while the classification of ultra-processed food was performed at the food item level (for each of the 3300 item of the NutriNet-Santé nomenclature). If an individual consumes a lot of butter, red meat, etc., he will score high for the Western pattern, but he will not necessarily have a high proportion of ultra-processed foods in his diet (since staple butter or unprocessed red meat are not considered as ultra-processed foods).

As explained in our response to the Statistical Reviewer, point 4:
The exact % of ultra-processed food consumption was presented and corresponds to the HR for a 10-point increment in the proportion of ultra-processed foods in the diet. In addition to this continuous coding, we used quartiles (to categorize the ultra-processed food variable and show the trends of the associations) because it represented the best compromise between a sufficient number of categories and a sufficient number of cases by categories (even in less represented cancer locations). However, main results were similar

when quintiles were used instead of quartiles (detailed results are presented in our response above). We have added the results of this sensitivity analysis in the manuscript, page 17.

The length of follow-up was relatively limited in time, since the cohort was launched in 2009. Thus, it allowed us to study mostly mid-term associations between ultra-processed food consumption and cancer risk. As it is usually the case in nutritional epidemiology, the assumption is made that the measured exposure at baseline (especially since we averaged a two-year period of exposure) actually reflects more generally the usual eating habits of the individual during adulthood, including several years prior to his/her entry into the cohort. However, since some carcinogenic processes may take several decades, it will be important in the future to re-assess the associations between ultra-processed food and cancer risk in the cohort, in order to investigate longer-term effects. This will be one of the perspectives of the present work for the upcoming 5-10 years. We have added some discussion about this point page 21, as requested.

Minor points

1. Abstract. Please consider the use terms that may not be understandable to a general reader (ie. food contact materials). Include the median age of the population and specify that models were adjusted for risk factors.

+++These three modifications have been made in the abstract, page 3.

2. Page 6 line 29 delete the s in "details".

+++This typo has been corrected page 6.

3. Page 7 line16 check "periodfor"

+++This typo has been corrected page 7.

Additional Questions:

Please enter your name: Martin Lajous

Job Title: Faculty-Researcher

Institution: National Institute of Public Health

Information for submitting a revision

Deadline: Your revised manuscript should be returned within one month.

How to submit your revised article: Log into <http://mc.manuscriptcentral.com/bmj> and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Once the revised manuscript is prepared, you can upload it and submit it through your Author Center. When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) and Committee in the space provided. You can use this space to document any changes you make to the original manuscript and to explain your responses. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s). As well as submitting your revised manuscript, we also require a copy of the manuscript with changes highlighted. Please upload this as a supplemental file with file designation 'Revised Manuscript Marked copy'. Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission.

When you revise and return your manuscript, please take note of all the following points about revising your article. Even if an item, such as a competing interests statement, was present and correct in the original draft of your paper, please check that it has not slipped out during revision. Please include these items in the revised manuscript to comply with BMJ style (see: <http://www.bmj.com/about-bmj/resources-authors/article-submission/article-requirements> and <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists>).

Items to include with your revision (see <http://www.bmj.com/about-bmj/resources-authors/article-types/research>):

1. What this paper adds/what is already known box (as described at <http://resources.bmj.com/bmj/authors/types-of-article/research>)
2. Name of the ethics committee or IRB, ID# of the approval, and a statement that participants gave informed consent before taking part. If ethics committee approval was not required, please state so clearly and explain the reasons why (see <http://resources.bmj.com/bmj/authors/editorial-policies/guidelines>.)
3. Patient confidentiality forms when appropriate (see http://resources.bmj.com/bmj/authors/editorial-policies/copy_of_patient-confidentiality).
4. Competing interests statement (see <http://resources.bmj.com/bmj/authors/editorial-policies/competing-interests>)
5. Contributorship statement+ guarantor (see <http://resources.bmj.com/bmj/authors/article-submission/authorship-contributorship>)
6. Transparency statement: (see <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/transparency-policy>)
7. Copyright statement/licence for publication (see <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse>)
8. Data sharing statement (see <http://www.bmj.com/about-bmj/resources-authors/article-types/research>)
9. Funding statement and statement of the independence of researchers from funders (see <http://resources.bmj.com/bmj/authors/article-submission/article-requirements>).
10. Patient involvement statement (see <http://www.bmj.com/about-bmj/resources-authors/article-types/research>).
11. Please ensure the paper complies with The BMJ's style, as detailed below:
 - a. Title: this should include the study design eg "systematic review and meta-analysis."
 - b. Abstract: Please include a structured abstract with key summary statistics, as explained below (also see <http://resources.bmj.com/bmj/authors/types-of-article/research>). For every clinical trial - and for any other registered study- the last line of the abstract must list the study registration number and the name of the register.
 - c. Introduction: This should cover no more than three paragraphs, focusing on the research question and your reasons for asking it now.
 - d. Methods: For an intervention study the manuscript should include enough information about the intervention(s) and comparator(s) (even if this was usual care) for reviewers and readers to understand fully what happened in the study. To enable readers to replicate your work or implement the interventions in their own practice please also provide (uploaded as one or more supplemental files, including video and audio files where appropriate) any relevant detailed descriptions and materials. Alternatively, please provide in the manuscript urls to openly accessible websites where these materials can be found.
 - e. Results: Please report statistical aspects of the study in line with the Statistical Analyses and Methods in the Published Literature (SAMPL) guidelines <http://www.equator-network.org/reporting-guidelines/sampl/>. Please include in the results section of your structured abstract (and, of course, in the article's results section) the following terms, as appropriate:
 - i. For a clinical trial: Absolute event rates among experimental and control groups; RRR (relative risk reduction); NNT or NNH (number needed to treat or harm) and its 95% confidence interval (or, if the trial is of a public health intervention, number helped per 1000 or 100,000.)
 - ii. For a cohort study: Absolute event rates over time (eg 10 years) among exposed and non-exposed groups; RRR (relative risk reduction.)
 - iii. For a case control study:OR (odds ratio) for strength of association between exposure and outcome.
 - iv. For a study of a diagnostic test: Sensitivity and specificity; PPV and NPV (positive and negative predictive values.)
 - v. For a systematic review and/or meta-analysis: Point estimates and confidence intervals for the main results; one or more references for the statistical package(s) used to analyse the data, eg RevMan for a systematic review. There is no need to provide a formal reference for a very widely used package that will be

very familiar to general readers eg STATA, but please say in the text which version you used. For articles that include explicit statements of the quality of evidence and strength of recommendations, we prefer reporting using the GRADE system.

f. Discussion: To minimise the risk of careful explanation giving way to polemic, please write the discussion section of your paper in a structured way. Please follow this structure: i) statement of principal findings of the study; ii) strengths and weaknesses of the study; iii) strengths and weaknesses in relation to other studies, discussing important differences in results; iv) what your study adds (whenever possible please discuss your study in the light of relevant systematic reviews and meta-analyses); v) meaning of the study, including possible explanations and implications for clinicians and policymakers and other researchers; vi) how your study could promote better decisions; vi) unanswered questions and future research

g. Footnotes and statements

Online and print publication: All original research in The BMJ is published with open access. Our open access policy is detailed here: <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse>. The full text online version of your article, if accepted after revision, will be the indexed citable version (full details are at <http://resources.bmj.com/bmj/about-bmj/the-bmjs-publishing-model>). The print and iPad BMJ will carry an abridged version of your article. This abridged version of the article is essentially an evidence abstract called BMJ pico, which we would like you to write using the template downloadable at <http://resources.bmj.com/bmj/authors/bmj-pico>. Publication of research on bmj.com is definitive and is not simply interim "epublication ahead of print", so if you do not wish to abridge your article using BMJ pico, you will be able to opt for online only publication. Please let us know if you would prefer this option. If your article is accepted we will invite you to submit a video abstract, lasting no longer than 4 minutes, and based on the information in your paper's BMJ pico evidence abstract. The content and focus of the video must relate directly to the study that has been accepted for publication by The BMJ, and should not stray beyond the data.

+++This editorial checklist has been carefully followed and the manuscript has been edited accordingly. Notably, all required statements have been added at the end of the manuscript pages 23-25. We have modified the structure of the abstract to comply with The BMJ rules of presentation (p3), and we have added a "what is already known/What this paper adds" box (page 2). We thank you again for your interest in our work.

END

Reference List

- (1) Andreeva VA, Salanave B, Castetbon K, Deschamps V, Vernay M, Kesse-Guyot E et al. Comparison of the sociodemographic characteristics of the large NutriNet-Sante e-cohort with French Census data: the issue of volunteer bias revisited. *J Epidemiol Community Health* 2015; 69(9):893-898.
- (2) INCa. The French National Institute of Cancer INCA - Cancer in France, 2014 Edition (In French: Les Cancers en France, Editions 2014); Available at URL: <http://www.unicancer.fr/sites/default/files/Les%20cancers%20en%20France%20-%20Edition%202014%20-%20V5.pdf>. [Accessed on 26-10-2017].
- (3) Monteiro CA, Cannon G, Levy RB, Moubarac JC, Jaime PC, Martins AP et al. NOVA. The star shines bright. *World Nutrition* 2016; 7(1-3):28-38.
- (4) Monteiro CA, Cannon G, Moubarac JC, Levy RB, Louzada ML, Jaime PC. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. *Public Health Nutr* 2017;1-13.
- (5) Moubarac JC, Parra DC, Cannon G, Monteiro CA. Food Classification Systems Based on Food Processing: Significance and Implications for Policies and Actions: A Systematic Literature Review and Assessment. *Curr Obes Rep* 2014; 3(2):256-272.
- (6) ANSES (French Agency for Food, Environmental and Occupational Health & Safety). Individual national study for dietary intakes 3 (INCA 3) [Étude individuelle nationale des consommations alimentaires 3 (INCA 3)] (in French). 17-7-2017.
- (7) Rauber F, Campagnolo PD, Hoffman DJ, Vitolo MR. Consumption of ultra-processed food products and its effects on children's lipid profiles: a longitudinal study. *Nutr Metab Cardiovasc Dis* 2015; 25(1):116-122.
- (8) Mendonca RD, Pimenta AM, Gea A, Fuente-Arrillaga C, Martinez-Gonzalez MA, Lopes AC et al. Ultraprocessed food consumption and risk of overweight and obesity: the University of Navarra Follow-Up (SUN) cohort study. *Am J Clin Nutr* 2016; 104(5):1433-1440.
- (9) Mendonca RD, Lopes AC, Pimenta AM, Gea A, Martinez-Gonzalez MA, Bes-Rastrollo M. Ultra-Processed Food Consumption and the Incidence of Hypertension in a Mediterranean Cohort: The Seguimiento Universidad de Navarra Project. *Am J Hypertens* 2017; 30(4):358-366.
- (10) Lahmann PH, Hoffmann K, Allen N, van Gils CH, Khaw KT, Tehard B et al. Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). *Int J Cancer* 2004; 111(5):762-771.
- (11) Dartois L, Fagherazzi G, Baglietto L, Boutron-Ruault MC, Delalogue S, Mesrine S et al. Proportion of premenopausal and postmenopausal breast cancers attributable to known risk factors: Estimates from the E3N-EPIC cohort. *Int J Cancer* 2016; 138(10):2415-2427.
- (12) Morois S, Mesrine S, Josset M, Clavel-Chapelon F, Boutron-Ruault MC. Anthropometric factors in adulthood and risk of colorectal adenomas: The French E3N-EPIC prospective cohort. *Am J Epidemiol* 2010; 172(10):1166-1180.

- (13) Touvier M, Kesse E, Clavel-Chapelon F, Boutron-Ruault MC. Dual Association of beta-carotene with risk of tobacco-related cancers in a cohort of French women. *J Natl Cancer Inst* 2005; 97(18):1338-1344.
- (14) Touvier M, Kesse-Guyot E, Mejean C, Pollet C, Malon A, Castetbon K et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. *Br J Nutr* 2011; 105(7):1055-1064.
- (15) Lassale C, Castetbon K, Laporte F, Camilleri GM, Deschamps V, Vernay M et al. Validation of a Web-based, self-administered, non-consecutive-day dietary record tool against urinary biomarkers. *Br J Nutr* 2015; 113(6):953-962.
- (16) Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol* 2012; 176(3):190-195.