Dear Professor Clare Collins and Dr. Sharon Kirkpatrick,

RE: Rebuttal Manuscript Number: NUTJ-D-18-00233 R1

Please find enclosed our revised manuscript entitled “FFQ versus repeated 24-hour recalls for estimating diet-related environmental impact” (NUTJ-D-18-00233R1). We thank you to reconsider this revised manuscript as an original contribution for publication in your Nutrition Journal.

The reviewer’s comments have all been addressed in the forthcoming pages, and included in the paper.

We hope that you find our revised manuscript suitable for publication and look forward to your response.

Yours sincerely, on behalf of all authors,
Elly Mertens,
PhD candidate at Wageningen University

Title: FFQ versus repeated 24-hour recalls for estimating diet-related environmental impact

1st author: Elly Mertens
Date: 9th November 2018
Article type: Original contribution

In the prepublication stage, correspondence and questions regarding the submitted manuscript should be directed to: Elly Mertens, Division of Human Nutrition, PO BOX 17, 6700 AA Wageningen, the Netherlands, Phone: +31 (0)317 482589, Fax: +31 (0)317 483342, E-mail: elly.mertens@wur.nl

Other authors: Anneleen Kuijsten, Johanna M Geleijnse, Hendriek Boshuizen, Edith JM Feskens, Pieter van ’t Veer

Reply to comments from reviewers:

In the revised manuscript substantive changes have been corrected and highlighted (using yellow highlighted text).

REVIEWER #1 COMMENTS TO AUTHOR:

The authors present an analysis of the relationship between dietary assessment methods, diet quality, and environmental impact. The authors should be commended on conducting a much-needed study, very relevant to current nutrition-environment debates, using a rigorous study design. The notable strengths of this study are:

- Analysis of the relationships between diet quality and environmental impact, using multiple measures of dietary assessment, multiple measures of diet quality, and multiple measures of environmental impact;

- Use of a composite environmental impact score instead of just one or two indicators of environmental impact; and

- Rigorous statistical assessment of variation and very good interpretation of observed variation.
We thank the reviewer for carefully reading our manuscript and for his positive feedback on our analyses.

Please find more specific comments below.

Introduction

Very well-written. Comprehensive yet also concise.

Methods

1. Lines 90-91: "Multiple 24hrs were administered throughout the two-year study period" - this is rather vague; please provide even an approximate frequency, even though it is specified in greater detail later (e.g., every 1-5 months).

The authors agreed that the wording ‘multiple’ is rather vague and could be specified. The frequency of sampling was not identical for each subject. Recall days were randomly selected and scheduled across the year, with at least 40 days in between. As suggested we specified the frequency in the manuscript (line 89-93).

2. It is confusing whether the data were collected longitudinally (line 88) or cross-sectionally (line 91). From Figure 1 it appears that the data were originally collected from 2,048 people (I assume from a longitudinal study), then the final sample was 1,169 (and were analyzed cross-sectionally). But the text is confusing because it doesn't specify that it would be helpful if the authors would be more explicit in the text about how the final sample size was reached, and whether the data were originally collected from a longitudinal sample and why they are being analyzed cross-sectionally.

Agreed and amended accordingly (line 93-97). A total of 2,048 subjects were included in the NQplus study, of which 1,653 subjects completed one FFQ at a baseline and 1,430 subjects completed two replicates of a web-based 24-hour recall spaced over one-to-five month period. We excluded 185 subjects with misreporting for the FFQ, and 37 subjects with misreporting for the 24hR. A total of 1,169 subjects, who completed both an FFQ and two replicates of the 24hR, remained for analysis.

We believed that our analyses were cross-sectional, as our dietary intake data for the FFQ and for the two 24hRs were collected within the same period of time. That is why the association between environmental impact of the diet and dietary quality we studied was studied cross-sectionally, independently of the method of dietary assessment. Both measures of diet-related environmental impact and measures of dietary quality, i.e. a food-based and a nutrient-based diet score, rely on dietary intake data, either obtained from the FFQ of the 24hR in our case.
3. If the participants completed two 24hRs and one FFQ, how can this be a cross-sectional analysis?

Dietary intake data as obtained from the FFQ and the two 24hR reflected the same period of time (see reply on comment #2).

4. Can the authors please clarify the temporal spacing between 24hRs and FFQs for each participant?

The temporal spacing between the first 24hR and the FFQ was on average 147 days, and between the second 24hR and FFQ was on average 182 days, however the sequence of filling out the 24hRs and the FFQ was not the same in all subjects. The FFQ was filled out before both the 24hRs in 739 subjects (63%), with the administration of the FFQ on average 127 days before the first 24hR and on average 212 days before the second 24hR. The FFQ was filled out between the two 24hRs in 169 subjects (15%), with administration of the FFQ on average 39 days after the first 24hR and on average 70 days before the second 24hR. The FFQ was filled out after the two 24hRs in 261 subjects (22%), with administration of the FFQ on average 274 days after the first 24hR and on average 70 days before the second 24hR.

5. Please provide an explanation of the pReCiPe formula in standard mathematical notation. For example, underneath the formula please include "where GHGE represents greenhouse gas emissions in kg of CO2-equivalents", etc. Also, please specify what the values in the formula represent. I understand that the ReCiPe method is provided in citation 24, but this manuscript should provide enough detail on the methods for it to stand alone to some extent.

Agreed and amended accordingly (line 126-128). ReCiPe can be regarded as a recipe (like a recipe from a cookbook) to calculate life cycle impact category indicators based on life cycle impact assessments. A life cycle impact assessment translates emissions of hazardous substances and extraction of natural resources into impact category indicators at the midpoint level (18 in total, including climate change/greenhouse gas emissions, fossil energy use and land use), or at the endpoint level (3 in total, including damage to human health, to the ecosystem and to resource availability). In the present study, we considered the midpoint-oriented method of which greenhouse gas emissions, fossil energy use and land use are the most important ones representing 90% of the total score. The authors specified that values for the formula based on these three midpoint indicators were weighted using the panel approach, and characterised and normalised using the year 2000 as reference year, and information was gathered for the European situation. Short explanation of the values in the formula is added to the revised manuscript (line 127-130).

6. Line 157 and throughout: "Data" should be pluralized (i.e., "data were", rather than "data was").

Agreed and amended accordingly (line 165).
7. Lines 160-162: Please specify how physical activity was categorized from a continuous variable to a categorical variable. Equal tertiles? Prespecified cutoff points?

Physical activity was categorised from a continuous variable to a categorical variable using pre-specified cut-off points, i.e. low physically active when Metabolic Equivalent of Task (MET, in minutes per week) < 500; moderate physically active when MET between 500 and 1000; and high physically active when MET ≥ 1000. The pre-specified cut-off points are added to the revised manuscript (line 170-171).

8. Line 174: Do the authors mean that T represents the known intake? Also, if there was only one FFQ administered, how can there be a within-person error in the FFQ model?

The authors take for granted that the intake values obtained from the 24hR are less biased, and therefore are the best standards to approximate true intake, as mentioned in line 177-178 of the manuscript. Based on this assumption, T represents the true intake, however it is unknown whether what we measure with the 24hR is the true intake.

The author agreed that for the present study with only one FFQ the within-person error in the FFQ model cannot be quantified, however there will be within-person error in the FFQ which can be measured with at least two replicates.

9. Line 195: Usually, "diet-health" associations refer to the associations between a dietary exposure and a health outcome. Since the authors are not assessing health outcomes, I don't feel that this term is appropriate. Actually, I'm not clear on what the intended association is, so it would be appreciated if the authors would please clarify that (do they mean the association between dietary intake and diet quality?).

The authors agreed that the wording ‘diet-health association’ might be confusing as we are not assessing any health outcome, and have therefore changed this into diet associations as we meant the association between diet-related environmental impact and dietary quality (line 204).

10. In the field of nutritional epidemiology, energy-adjustment is most often conducted using the residual method or the density method (Walter Willet has a fundamental paper on this from years ago). Can the authors please explain why they chose to energy-adjust by including energy as a model covariate?

The authors are familiar with the paper of Walter Willet on energy adjusted in which the different models for energy-adjustment are described. Next to residual method and the density method, a standard statistical approach to control for confounding by total energy intake is the inclusion of total energy intake in a multivariate model along with the dietary factor of interest. However, this standard multivariate model of energy adjustment by Willet will only produce the correct inferences for the effect of the dietary factor of interest on the outcome of interest when
the variation in the dietary factor is dependent on total energy intake, as mentioned in Walter Willet’s paper. We addressed this in Table 4.

Results

11. Lines 217-226: If the authors are comparing men vs. women (or any other groups, for that matter), they should be basing their statements on the results of statistical tests. Please indicate the type of test used and the results of those tests for each of the comparisons.

Type of test used were added to the revised Table 1 in footnote a, and results of the statistical tests were significant for all the characteristics presented in this table using p < 0.05, as also mentioned in the footnote.

12. Can the authors please confirm that the values presented in Table 1 have been energy-adjusted? In other words, do men's diets have a greater environmental impact than women's diets simply because men consume more food (i.e. kcal)?

Values presented in Table 1 are crude values, not energy-adjusted, as mentioned in line 232 of the manuscript. Men’s diet do indeed have a higher environmental impact than women’s diet, because of their higher energy intake. This interpretation can be confirmed by the comparison of the energy-adjusted values of environmental impact for men with those for women that shows a much smaller difference, as presented in the bottom-half of online resource 1 and online resource 2, data for men and for women, respectively.

13. Table 1: Are these values from the 24HR or the FFQ? It seems that, instead of comparing men vs. women, the more relevant comparison is 24HR vs. FFQ.

These values are based on the two replicates of the 24hR as stated in the manuscript (line 231) and in footnote e (below Table 1). We agreed that comparing FFQ with the 24hRs would be more relevant, however we also believe that providing the general characteristics of the study population is important to show who was included in the present analyses, with regard to their socio-demographic and lifestyle factors. The authors agreed that the more relevant comparison is between 24hR and FFQ, and this comparison is addressed in Table 2 and 3.

14. Why does table 3 indicate that there are two replicates of the 24HR, while Table 2 does not? Was only one 24HR represented in Table 2? My understanding from the Methods section was that a mixed model was used to estimate usual intake using the 24HRs.

Table 2 is also based on two replicates of the 24hR, and this has therefore been added to the title of this table
15. All tables: please present results of statistical tests of difference, and indicate the tests used.

Statistical tests of difference between men and women for Table 1 was added to the footnote a of this table (see comment #11). For Table 2 on the comparison between FFQ and 24hR in terms of contributions of the different food groups, no statistical testing was performed since this table was added for illustrative purpose. For Table 3, all the relevant statistical information can be found in the footnotes of this table, same holds for Online Resource 1 and 2. Results for Table 4 on the association between dietary quality and environmental impact were obtained from multivariate linear regression, reporting the regression coefficients and its 95% confidence interval. Interpretation of the regression coefficient can be found below the table in footnote a.

Discussion

16. Line 295: Here the authors state that no gold standard exists, but earlier in the manuscript they stated that the 24HR was the gold standard. Please clarify.

There is indeed no gold standard for measuring environmental impact of the diet, however the authors believe that the 24hR is best available standard to approximate true diet-related environmental impact, as stated in line 177-178 of the manuscript. The same holds for measuring dietary quality of the diet, for which we used the 24hR as the alleged gold standard reference, although there is none. By using the word ‘alleged’, we meant to signify that this is an assumption on which we based our results on.

Very well-written.