Reviewer’s report

Title: Importance of Details in Food Descriptions in Estimating Population Nutrient Intake Distributions

Version: 0 Date: 01 Nov 2018

Reviewer: Thea Zimmerman

Reviewer’s report:

This is interesting work, and definitely has potential for streamlining data collection methodology. I cannot speak to the validity/reasonableness of the random forest method, but developing a scientific approach to identifying extraneous details in dietary interviews has definite value for the scientific community. It is also very important to assess the impact of any changes to the interview, which the authors have done. However, the conclusion that elimination of these facets will have no significant effect on nutrient distributions is not really demonstrated by this work. The results show that 727 of the foods reported in the Dutch national study were changed to a new food code following elimination of the facets identified as least predictive. This represents just 2.7% of all the foods reported in the study (26,679). The lack of effect on nutrient distributions is as likely to be due to the small number of changed food codes as it is to changes in the nutrients resulting from the facet elimination. It may be that in a population where these changes were more prevalent, the effect on nutrients may be greater. This may be one of the points - that the elimination of these facets may simplify the data collection and have little effect on nutrients precisely because it affects so few food items, but that is not the point the authors make. The manuscript would benefit from a discussion of the impact of changing so few food codes.

It is not clear that saving 3 to 4 minutes of a 45 minute interview time is truly an impact - it is not a significant reduction in respondent burden on an individual level.

The authors claim that the reduction in the number of unique food combinations would result in 200 fewer coding hours. The coding hours are more accurately determined by the number of food items reported by respondents, and as the work evaluated in this manuscript is data processing, it doesn't support a reduction in the number of food items reported. Additionally, when coding 7,638 24-hour recalls, a savings of 200 hours is unlikely to be a meaningful reduction.

Comments on lines from the manuscript:

149: Stratified by food group, the importance of a facet is calculated as the percentage increase

150 in prediction error, denoted by %IncMSE, when data for that facet are permuted in the dataset (implying no
predictive power for that facet in predicting the nutrient content), while keeping data for the other facets unchanged.

The sentence is confusing, as it seems to be saying that an increase in the prediction error implies no predictive power for the facet in predicting the nutrient content. Doesn't an INCREASE in prediction error when data for that facet is permuted mean that the facet actually DOES have predictive power for the nutrient content?

In the food group 'Miscellaneous' most facets were unimportant, followed by 'Meat (products)' and 'Dairy (products)'.

Sentence is unclear. What does it mean that meat and dairy "follow miscellaneous"?

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.
I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.