Author’s response to reviews

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

Authors:

Liangzi Zhang (liangzi.zhang@wur.nl)
Anouk Geelen (anouk.geelen@wur.nl)
Hendriek Boshuizen (hendriek.boshuizen@wur.nl)
José Ferreira (jose.ferreira@rivm.nl)
Marga Ocké (liangzi.zhang@wur.nl)

Version: 2 Date: 22 Jan 2019

Author’s response to reviews:

NUTJ-D-18-00263

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

Dear Prof. Kirkpatrick and the reviewers,

Thank you very much for your quick response for my revised version of our manuscript “Importance of Details in Food Descriptions in Estimating Population Nutrient Intake Distributions”. We believe that our responses to the valuable and helpful comments of the reviewers have resulted in an improved manuscript. Comments from the reviewers and our responds are detailed below. Changes are highlighted directly in the text.

Sincerely,

Liangzi Zhang

Reviewer #1:

Question 1: I appreciate the authors' responses to my comments, and most have been adequately addressed by the revision. However, thorough editing for language, grammar, voice, and tense is
needed. There are numerous grammatical errors and sentences that are quite difficult to understand at first read.

Answer 1: We thank the reviewer for raising this point, we have revised and highlighted several parts of the manuscript that were not easy to understand before.

Question 2: Permutation means to change the order of something, reverse, rearrange. Is this truly the meaning intended in this sentence (lines 151-153): "Stratified by food group, the importance of a facet (denoted by %IncMSE), was calculated as the percentage increase in prediction error, when data for that facet were permuted in the dataset, while keeping data for the other facets unchanged." My understanding is that the facets are removed, not rearranged.

Answer 2: The values of the facets are permuted randomly. This means that they are shuffled (reordered) between the observations. This means for each observation the facet gets a new (shuffled) value, which now is a random value, i.e. this value is not associated any more with any other variable of the observation (including the value that will be predicted). Because of this, the facet value no longer has any predictive value. This is the standard procedure used in random forest. According to one sentence in the reference Breiman, L. Random Forests. Machine Learning 2001, 45, 19, “Suppose there are M input variables. After each tree is constructed, the values of the mth variable in the out-of-bag examples are randomly permuted and the out-of-bag data is run down the corresponding tree.” We think “permutation” is the right way to describe the procedure.

Question 3: I appreciate the addition of lines 274-276, discussing the fact that many foods remain linked to the same code in the Dutch National Food Composition Database. This is an important point that actually runs counter to the idea of reducing the detail questions. The authors mention the importance of retaining more details in the interview to allow national surveillance (lines 340 and on), but make no mention of times when researchers not conducting a national surveillance study may still want to collect details even though the Food Composition Database does not contain food codes that reflect that information, such as intervention studies modifying certain aspects of the diet. In the absence of nutrient composition data reflecting those modifications, only the questionnaire data will tell researchers whether the study participants have modified their diets. Although questionnaire simplification is a worthy goal, limitations should be acknowledged.

Answer 3: We agree with the reviewer that other study aims than estimating nutrient intake can be reasons to keep detail questions. We indeed tried to explain this in the discussion in lines 306-309 and added some further explanation to stress that the simplification is study aim specific. “Similarly, in DNFCS the facet preparation method should be kept for some foods like meats since it is an important food characteristic to identify microbiological risks. The facets and descriptors of the GloboDiet software can be tailored for any new study [17]. Researchers use this software should thus consider carefully which food characteristics are important for their study aims before the start of a study.”.
In addition, in the method section lines 116-118, we briefly mentioned the study specific adaptation of GloboDiet software. “The actual selection of facets and descriptors could be adjusted according to country and study-specific situations. For the DNFCS 2007-2010 version of the software, a total of 16 facets with varying numbers of descriptors was selected by experienced dieticians based on the knowledge of the food market and the insight into the purposes of the data collected (Table 1). “

Question 4: Lines 328-335: I questioned the true effect of deleting facets on the coding effort, as the number of foods reported has more of an effect on coding time than the details reported. The authors’ response indicated that most foods were auto-coded. However, the manuscript reports that the average time needed to link a combination to the food composition table was 5-10 minutes, and that 442 hours would be saved by facet removal. This is describing a manual coding effort, and although the authors added an explanation that only new food-descriptor combinations would require linking, my concern still stands: the number of food items that need to be manually coded more correctly determines the coding time, not the number of details collected about the food. Perhaps the authors intended to provide an idea of how the unimportant facets could lead to new food-descriptor combinations, but that is not the argument they have made.

Answer 4: We apologize for any misleading sentences we might had before, because the food items that need to be manually coded is the same thing with the unique food-descriptor combinations. As indicated in the methods section:” The total collected consumption data from all participants for the two 24HRs has 219,006 food records....This results in a number of 26,679 unique combinations of foods with descriptors.” This means only 26,679 combinations need to be linked manually.

I revised the discussion from lines 316-321 into the following:” Moreover, less extensive food description during data collection would lead to fewer unique food-descriptor combinations reported in a survey. In the data handling phase, each unique food-descriptor combination needs to be linked manually to the food composition database, which would cost 5 to 10 minutes approximately. Hence, a reduced number of 3534 unique combinations (from 26679 to 23145) after deleting less important facets would save around 442 hours.