Author’s response to reviews

Title: The added value of Food Frequency Questionnaire (FFQ) information to estimate the usual food intake based on repeated 24-hour recalls

Authors:
C Ost (cloe.ost@wiv-isp.be)
Karin De Ridder (karin.deridder@wiv-isp.be)
Jean Tafforeau (jean.tafforeau@wiv-isp.be)
Herman Van Oyen (h.vanoyen@wiv-isp.be)

Version: 1 Date: 21 Mar 2017

Author’s response to reviews:

Answer the reviewers comments:

Reviewer 1

Reviewer #1: 24-hour recalls to assess dietary intake are increasingly being used, and respondents are usually asked to provide only two days intake to reduce the burden on them. This method, however, produces wide within-person day-to-day variations and has limitations for providing good estimates of usual intake, particularly of foods where intake is episodic in nature. Increasingly the dietary data gathered in this way is modelled further using programs such as SPADE to produce intake estimates that are closer to usual intakes. The current study evaluates whether an integrated approach using information about the frequency of intake collected by Food Frequency Questionnaire (FFQ) in addition to the 24hr data can produce intake estimates that are closer to usual intakes than if the FFQ data is not used. In particular, information about never-consumers of foods from the FFQ is used.

This is an important area to study, however the manuscript of the current study is confusing and difficult to follow for a variety of reasons, and needs to be re-written more clearly for those who are not very familiar with modelling usual food intake, as well as for those who are familiar with it.

Firstly, it is difficult to follow the different modelling scenarios: I was unsure how many simulation models there were and sometimes I was confused about which of the models the authors were referring to, especially since the names of the different models were similar and I believe the authors did not always refer to the same model every time with the same name.
Near the beginning of the methods section, I suggest adding a flow chart/diagram of the different models/simulations (and the original BNFC2004 data) giving them a name which should be used throughout the manuscript and state whether they include consumers only, whole population, and are simulated with or without FFQ information on never-consumers, etc. Additionally a brief description of the different models could be added near that.

Response:

We added four flowcharts to the methods section:

- One showing the basic steps of the SPADE model with and the SPADE model without the inclusion of never-consumers.
- One showing more detail on how SPADE exactly models the usual intake distribution for consumers.
- One showing how the simulated BNFC2004 with two 24-hour recalls per individual was simulated.
- One showing how the "true" usual intake distribution was simulated.

For the names:

- Original BNFC2004 consumers only versus Original BNFC2004
- Original BNFC2004 consumers only: contains two 24-hour recalls for the individuals that do not indicate to be a never-consumer.
- Original BNFC2004: contains two 24-hour recalls for all respondents (thus including the respondents that indicated to be a never-consumer).
- Simulated BNFC2004 consumers only versus Simulated BNFC2004
- Simulated BNFC2004 consumers only: contains two simulated 24-hour recalls for the individuals that are not never-consumers.
- Simulated BNFC2004: contains two simulated 24-hour recalls for all respondents (thus including the respondents that are simulated to be never-consumers).
- Simulated "true" usual intake distribution consumers only versus simulated "true" usual intake distribution.
• Simulated "true" usual intake distribution consumers only: the usual intake distribution obtained after the simulation the "true" usual intake for the (non-) consumers.

• Simulated "true" usual intake: the usual intake distribution obtained after the simulation of the "true" usual intake for the (non-) consumers and the never-consumers.

* Lines 205-206: At the bottom of the first paragraph of the section 'Simulation of consumers only' is the sentence "The usual intake distributions obtained from the original BNFCS2004 data should be called 'the original BNFCS2004'. However, it is not clear whether any modelling /simulation was done to produce 'the original BNFCS2004'; if not, then to reduce confusion I suggest that sentence be included earlier in methods. Additionally, if no modelling was involved to produce 'the original BNFCS2004' then the intake distributions are unlikely to reflect 'usual' intake. If not I suggest omitting the word 'usual' when talking about the distribution of 'the original BNFCS2004'. Alternatively, if usual intake has been estimated accounting for within-person variations, please make this clear(er).

Response:

We rewrote whole this paragraph, because it was indeed confusing. The "original BNFCS2004 consumers only" is the result of running the SPADE model on the consumers only of the original BNFCS2004 data. It has been rewritten in the article as follows:

“First all never-consumers were excluded from the original BNFCS2004 using FFQ data, resulting in a subdatabase with consumers only. Then the SPADE 2-part model without information on the never-consumers was used to obtain the usual intake distribution for consumers only from the original BNFCS2004. During the modeling some parameters were extracted: the mean usual intake for every age (μage), the within-individual standard deviation (σw), the between-individual standard deviation (σb) and the Box-Cox transformation parameter (λbc)."

Secondly, some parts of the manuscript are difficult to follow because of the sometimes awkward and unclear use of English. Ideally the manuscript needs proof reading by a first language English speaker who understands the methods.

Also

* Generally avoid the use of 'one' followed by a verb in articles and reports (e.g. lines 378, 390, 392, 438, 484, 488, 521, 529 and conclusion of abstract).

Response: All these constructions were changed.
* Please avoid the use of 'in case' (e.g. line 392), instead it may be more appropriate to use 'when'

Response: All these constructions were changed.

* Line 102: Instead of using the term 'blow-up' a more appropriate word could be used. e.g. widen

Response: This was changed.

* Line 372: Do you mean 'somewhat' instead of 'somehow'? Or perhaps it may be better to delete 'somehow'.

Response: This was changed.

* Line 71: delete 'and this'

Response: This was changed.

Lines 111-112 and 119-120: I assume some of the never-consumers may have consumed the foods at some point in their lives, and they are not 'never-ever-consumers'. Therefore, in the definition of never-consumers please mention the time frame of the FFQ that provides the never-consumer information, and the specific time frame of never-consumers in the BNFCS. For instance did the FFQ used in the BNFCS ask whether the respondents have never consumed a food within the last 12 months?

Response: We did not add the time frame into the introduction, because here the goal was more to introduce the concept of the difference between a consumer, a non-consumer and a never-consumer.

We did add the time frame of the FFQ (12 months) in the methods section:

Firstly in the part where the BNFCS 2004 was described:

"A twice repeated non-consecutive face-to-face 24HR and a self-administered FFQ (covering a 12 month period) were used to gather information on food intake."

Secondly in the data simulation part:

"The simulation took place in two stages: (a) simulate the consumers only and (b) simulate the never-consumers only (never-consumers are individuals who indicated in the FFQ that they never consumed the food item during the last 12 months)."
Lines 180-184 and Table 1: A criteria for selecting the food items for modelling was they needed to have different proportions of never-consumers. Table 1 shows that the percentage of never-consumers for 2 of the 3 foods selected (Water and Cheese) are in fact very similar and very low (2% and 5% weighted). I think it would have been useful and interesting to use the FFQ never-consumers data to model the usual distributions of other foods that had similar or higher percentages of never-consumers than that of fat spread.

Response: It is true that the weighted proportion of never-consumers for water and cheese was similar. However while first selecting the food items the "unweighted" percentages (which differ more) were studied, it was only during the simulation phase that the effect of the weighting became obvious.

We would have liked to include a food item where the proportion of never-consumers was higher as compared to the fat spread. However the FFQ only contained 60 question, so the food items questioned were in general broad and not specific (e.g. milk, fruit, nuts,...). As a consequence the proportion of never-consumers was very small for most questions. On the other hand more specific FFQ questions (e.g. giblets) resulted in convergence problems. Namely, the proportion of individuals that consumed the food on both recall days must be sufficiently large. The latter was generally not the case for food items with a large proportion of never-consumers.

Line 184: It may be useful to briefly explain 'convergence' problems.

Response: We added an explanation:

"Convergence problems occur when the available amount of data is insufficient to obtain an adequate model fit.”

Line 296: Why does this sentence state that the simulated "true" usual intake distribution does not require any modelling?

Response: For the simulation of the "true" usual intake, we simulated for each individual 1000 intake days, which would be similar to following these individuals during 2 years and 8 months. Then the median intake over these 1000 days can be considered as the "true" usual intake of an individual. So actually the within-individual variance will have disappeared by following the individuals during 1000 days and as a consequence modeling is no longer needed to integrate this component out.

To make it more clear we added to line 265:
“Thus no additional modeling was needed, since by following the individuals during one thousand days, almost no within-individual variance will be left.”

Line 419: I don't think Goedhart et al. [6] were the first to perform a large simulation study as the authors have stated, I think they were the first to use a number of methods and programs to simulate and compare intake of episodically consumed foods, and to include never-consumer information from an FFQ. Please correct.

Response: We did not want to say that they were the first to perform a simulations study, but that they were the first to perform a simulation study of that size with many different scenarios and programs. But we understand that this can be wrongly interpreted. We changed the sentence as follows:

“Goedhart et al. [6] performed a large simulation study.”

Lines 337-338: The tables 2 and 3 only include one of the 4 simulations undertaken for each of the 3 foods. Would it be more appropriate to include the mean and/or the range of these 4 simulations?

Response: We think that the focus should be on the comparison of the results simulation by simulation, therefore providing the range would be a better option than providing the mean. However, if in tables 2 and 3 the range would be provided in each cell, it would lead to a lot more information in the tables. In addition this could be confusing to then compare percentiles, because then it could be for table 2 that the ranges of P0.25 and P0.50 are overlapping. Therefore we chose to include an additional file where the graphs of figure 1 will be shown simulation by simulation to allow a within-simulation comparison of the effect of including FFQ information on the never-consumers.

Tables 2 and 3: In the tables please state the units used.

Response: The units (g/day) were added in the table titles of table 2 and 3.

Figures: More information is needed in the figures so that they are understandable without having to read the rest of the manuscript. For instance a title is need for figures A, B C to distinguish which is Water, Cheese and Fat spread. Also a title is needed on each Y axis.
Response: We added a label on the y-axis (relative difference), we changed A-C into the names of the food items. In an additional file the plots will be shown simulation by simulation, resulting in four plots per food item, to allow the within simulation comparison of with/without FFQ.

Discussion, limitation section: Modelling usual intake appears to be very involved and time consuming, requiring more expertise than is needed for normal analyses of datasets; potentially this limits the amount and type of analyses that can be produced from datasets that have only 2 days of food intake.

Response: This study was a simulation study that focused on assessing the potential benefits of including FFQ information on the never-consumers during the modeling of the usual intake. It was especially the simulation process which was complicated. However, if the goal is to model the usual intake there is no need for this simulation exercise. When a ready-to-use package (e.g. SPADE) is selected to model the usual intake data, the process should not be too complicated. Convergence issues could still occur, but then it means that the amount of available data is insufficient to have an adequate model fit.

The amount and type of analyses produced from datasets with two days of food intake is indeed limited. However performing three or even four 24HRs will not solve the problem for episodically consumed foods. Therefore I did not add this to the limitations section of the study.

Reviewer 2

Reviewer #2: This is an interesting paper with clear objectives and conclusions. However it is currently a difficult manuscript to get through as the methods are far too detailed, and the results and discussion are not written concisely. This probably reflects the fact that the work is based on a masters project, so it needs to be re-written appropriately for a scientific journal.

Major comments:

1) The methods need to be rewritten to make them accessible to a non-statistician. I appreciate that the modelling is complex, so would suggest a clear overview is given within the text, then enough detail for the work to be replicated is given as supplementary material.

Response: We added four flowcharts to the methods section, as was proposed by the other reviewer, to make the methods section more understandable:

- One showing the basic steps of the SPADE model with and the SPADE model without the inclusion of never-consumers.
• One showing more detail on how SPADE exactly models the usual intake distribution for consumers.

• One showing how the new database with two 24-hour recalls per individual was simulated.

• One showing how the "true" usual intake distribution was simulated.

2) The results and discussion sections need to be made more concise, e.g. lines 345-53 (re Table 2) and lines 365-93 (re Table 3) in the results, and the limitation of simulation by age strata which is mentioned twice in the discussion

Response: Table 2 versus lines 345-353: L345-353 were rewritten to have less repetition.

Table 3 versus lines 365-393: L365-393 were rewritten to have less repetition.

For the discussion: the simulation by age strata has two drawbacks:

• L451-65: the simulation of the “true” usual intake distribution was performed per age stratum and did not need any additional modeling. The simulated BNFCS2004 was simulated the same way, but still needed modeling by SPADE. SPADE will model the intake amount in function of the age, but will only fit one model over all age strata together. As a consequence both models take into account the age, but the “true” usual intake distribution will take the age stronger into account as compared to the usual intake distribution obtained after modeling the simulated BNFCS2004 with SPADE.

• L500-13: the simulation was performed in the three age strata separately, as a consequence the number of individuals consuming the food on both recall days had to be sufficiently large in each of these age strata to avoid convergence problems.

The first problem was discussed in the section "evaluation of the simulation" in this section the goal was to explain the differences that were related to the simulation.

And the second problem was discussed in the "strength and limitations of the study", as it was a choice we made, because usual intakes can vary substantially depending on the age of the individuals. We could also have made the choice to perform the simulations for all age strata at once, but then we would not have taken into account the abovementioned problem.
3) All decisions made need to be justified, e.g. why was the SPADE modelling option chosen in preference to others? / line 198 "for the simulation purpose it was chosen to only to model the intake amounts in function of age"?

Response: The reason why SPADE was chosen was mentioned in the methods, however we reformulated the sentence a bit so it is more clear why SPADE was selected:

“The Statistical Program for the Assessment of Dietary Exposure (SPADE), an R package developed at the Dutch National Institute of Public Health was selected to estimate the usual intake distribution [21-22], because both R and the SPADE package are freeware. In addition SPADE allows including information on the never-consumers without a large increase in the analysis time [21-22].”

The intake frequency could also be simulated in function of age. However, SPADE is sensitive to convergence problems, when SPADE has to model both the intake amount and the intake frequency in function of age convergence problems occur more often. In addition it has been reported that the intake amount can differ substantially depending on the age. The effect of the age on the intake frequency has been less reported. We changed this in the text as follows:

“To avoid convergence problems it was chosen to only model the intake amounts in function of age during the simulation.”

4) Figure 1 needs more thought as it is confusing to have all four simulations on the same plot e.g. within simulation comparison of with/without FFQ is not possible. Also more labelling is needed i.e. the y axes and what A-C refer to

Response: We added a label on the y-axis (relative difference), we changed A-C into the names of the food items.

As an additional file we included the figures simulation by simulation, resulting in four plots per food item, to allow the within simulation comparison of with/without FFQ.

5) When reference 6 is mentioned in the discussion it is a bit of a surprise that work that is so similar has been undertaken before (even though it is a report rather than a published paper) - the similarities and differences between that work and your own should be explained in the background.

Response: We added the information on the study of Goedhart et al. [6] and the difference with our study at the end of the background section:
“Also Goedhart et al. [6] performed a simulation study, where they amongst others assessed the effect of the use of FFQ information to identify the never consumers. However Goedhart et al. [6] used artificial data to assess the effect, while this study will base the simulation on real food items whose intake was assessed in the Belgian population during the BNFCS2004.”

The benefit of our simulation as compared to Goedhart et al. [6] was explained in the discussion:

- L 421-423: the simulation is in this study performed on the basis of data of the BNFCS2004, while Goedhart et al. simulated data artificially.

-L 515-525: where the benefits of L421-423 are explained in more detail.

Minor comments:

1) The N value should be given in the abstract
Response: This was added in the methods part of the abstract:

“Three food items with a different proportion of never-consumers were selected from the database of the Belgian food consumption survey of 2004 (N=3200).”

2) In table 1 it would be helpful to have an explanation of the weights in a footnote.
Response: We added a legend to table 1 as follows:

Legend of table 1: The BNFCS2004 made use of stratified sampling. To make the sample representative for the whole Belgian population (between 15 and 74 years) a weighting factor was calculated to compensate for the unequal sampling probability. The weighting factor can be used to convert the percentage of never-consumers towards the weighted percentage of never-consumers.