Reviewer’s report

Title: Food purchase patterns: empirical identification and analysis of their association with diet quality, socio-economic factors, and attitudes

Version: 0 Date: 13 Mar 2017

Reviewer: Michelle Morris

Reviewer's report:

Thank you for the opportunity to review this paper. It is really interesting and has the potential to be a great paper.

Background

Line 54 - Please make a stronger link from diet quality in the first sentence to the increasing overweight in Germany in the second sentence.

Line 56 - Please consider introducing 'diet-related diseases' rather than assuming understanding. This could appear after the first sentence to make the background flow better.

Line 58 - Lifestyle - all one word.

Line 62 - 'these disease' - do you mean diet-related diseases?

Line 69 - consider re-phrasing, this does not read well.

Line 83 - I completely agree that food purchase data is potentially more objective. However, you need to say acknowledge that food purchase data is not the same as the food consumption records used in the previous studies. I would recommend expanding on this in the discussion.

Line 84 - I agree that a combination of both types of diet measure could offer deeper insight. As you do not do this in this paper, perhaps you should save this for the discussion.

Line 94 - re-phrase - doesn't read well.

Line 97 - re-phrase - doesn't read well.

Line 100 - It would be beneficial to introduce the concept of individual versus household patterns and behaviours.

Line 97 - onwards could set out the aims of the paper more clearly. I think this paragraph undersells the paper.
Methods

Line 105 - data is from 2011. Please reflect on the timeliness of this data in the discussion.

Line 113 - 'obliged to participate' - is this in the German micro-census or the consumer panel. Please clarify.

Line 116 - manual scanning products for 10 months is extremely labour intensive. You should reflect on the reliability of this data and quality deterioration with time in the discussion.

Line 132 - How many products in total are in the BLS?

Line 137 - Please reflect on the impact of reducing the number of linked products in the discussion.

Line 149 - Can you include an image of the German Food Pyramid in the paper? If you have space for an extra figure this would add value.

Line 150 - take care here. How do the 18 food groups have specific health values? It may be less ambiguous just to state the 18 food groups.

Line 163 - 'a factor called pattern' - Are these your derived dietary patterns?

Line 184 - did you consider using a DAG or other causal diagram to check for confounder and mediators?

Line 200 - 'bought at retail brands'? Do you mean supermarkets?

Table 1 - The table legend could be expanded to make it clearer and reduce ambiguity.

Line 206 - indicates that the chronbach's alpha presented in table 1 is unacceptable. Is this your intended message? If you are implying it is close to acceptable, please clarify.

Line 207 - Please explain why it is relevant to present three goodness of fit measures. For other statistical models this is not appropriate so it would be worthwhile justifying why it is this instance.

Results

Table 2 - household characteristics. Many of these distributions are skewed, as indicated by the mean and std. dev. - therefore is the mean value the most appropriate to report?

Level of prices - how are these results helpful - mean 0 std dev. 0.999?
Can you include (n) for the number of people responding to each question?

Attitudes of the person….these values may be more intuitive presented as proportion of households reporting 'high price awareness', 'paying attention to fat' etc

Table 3 - all decimals should be presented with decimal point rather than a comma. This is confusing and inconsistent.

Line 249 - 'pattern_had' - please remove underscore

Table 5 - What is the journal style for presenting p values? I would prefer to see P=0.000 reported as p<0.001.

Table 4 and 5 - how/why did you select these micronutrients to present?

Table 4 and 5 - be consistent with the number of decimal places you present. Some results have three and others four. Consider carefully how many decimal places is meaningful.

Table 5 - P/S quotient foot note symbol should be 'f'

Line 273, 274,275 - please rephrase - this does not read well.

Line 276 - consider saying 'this processes pattern score'.

Lines 293 - 295 - Is this effect determined by the presence of adults in the household? To be discussed in the discussion.

Line 300 - how does the non-linear relationship alter your modelling?

Line 304 - Where is figure 2 (I cannot find it in the submission). There is no reference to figure 1 in the text - should this be figure 1. No figures included in the submission. Please clarify and update accordingly.

Line 306 - is the processed pattern the cheapest? Please discuss in the discussion.

Table 6 - again - please be consistent with your decimal points versus commas and presentation of p values.

Table 6 legend - consider including that the beta value is a standardised coefficient

Discussion

The discussion in general should include more discussion of the challenges in the study, rather than a detailed description of results. Please see comments above where I have highlighted areas to be expanded in the discussion. In addition, please consider the effect of waste when using
purchase data, the impact of household composition on the findings and implications of using purchase data versus consumption data. The discussion may benefit from some additional subheadings or structured discussion to cover: 'strengths and limitations'. 'Future direction' & 'policy implications'

Line 344 - Why does (-products) appear in brackets here? Same on line 346.

Line 352 - I would like to see this paragraph linked better to the diet quality (in the next paragraph) having potential policy application.

Line 371 and line 391 both begin with 'Furthermore' - please consider another word.

Line 395 - does this really indicate that dietary habits are formed at an early age? Or does it reflect that adult dietary patterns were dominant as they consume more of the household purchases?

Line 400 - care with English - 'particularly' doesn't read right in this sentence.

Line 414 - can you add a reference for this assumption?

Line 414 - another 'Furthermore' - please consider another word.

Line 431 - why only fat consumption?

Line 439 - another 'furthermore' - please consider another word.

Conclusion

This appears to be an extension of the discussion. Please consider re-focussing on what you think are the main conclusions of the study.

It is not clear to me the value of the identified patterns in the discussion about diet quality and how these translate to policy. Please clarify.

Ethics - please expand why ethical approval was not applicable to this study. I would expect to see some ethical approval for secondary data analysis, or at least reference to why it is not required.

Level of interest

Please indicate how interesting you found the manuscript:

An article of importance in its field
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 am a shareholder in a company, Dietary Assessment Ltd. This company provides an online dietary assessment tool www.myfood24.org.

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.