Author’s response to reviews

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

Authors:
Silke Thiele (sthiele@food-econ.uni-kiel.de)
Jonas Peltner (jpeltne@food-econ.uni-kiel.de)
Almut Richter (RichterA@rki.de)
Gert B.M. Mensink (MensinkG@rki.de)

Version: 1 Date: 15 Jul 2017

Author’s response to reviews:

We want to thank both reviewers for the very valuable comments which helped a lot to improve the manuscript. In the following we explain how we incorporated the comments in our paper.

Reviewer #1: Thank you for truly interesting reading. I believe your study provides highly valuable contribution to the knowledge of food purchase patterns and health. It is thoroughly conducted and I am impressed with the quality of the data - well done!

I only have some minor comments and questions about your manuscript, which, if considered, I believe will improve the manuscript even further.

- I believe you need to explain more about the food grouping. I cannot quite grasp the construction of the groups, not even by looking in your ref 20. For example - how come juice is grouped with vegetables/salad? What is the connection between brown bread and potatoes? Eggs together with sausages and cream? And the subgroups in Fats and oils - why section vegetable oils in several subgroups, and what is the difference between these (all vegetable) oils and 'vegetable fat' in the Lard, vegetable fat group? Oilseed is - to the best of my knowledge - an umbrella term for several different crops used to produce oils.

The groups are given by the used German Food Guide Pyramid. We explained the construction of the groups more clearly in the section “classifying food groups”, added a graphical illustration of the German Food Guide Pyramid (Figure 1) and changed some names of the food groups in table 3.
You haven't stated a significance level in the methods section. In the results, there are mentionings of significance level at some points - but this feels like you've selected level for each test. And also, it leaves me wondering what level is set where this isn't stated?

We chose a p-value of < 0.05 as the significance level and added this information in the section “statistical methods”

Have you considered calculating effect size on your comparative tests? You show really low p-values (the majority is stated as 0.000 - please change to p<0.001) when differences are very small - not surprising as you have so many participants. Effect size would be valuable to interpret the impact of significant differences.

The main problem with effect sizes in our study is interpretability. For an interpretation normally distributed variables are required but this is not the case for the nutrient density variables in our study. Therefore, the effect sizes of our study could not be compared with the ones from other dietary pattern analyses. That’s why we did not report effect sizes in our study. However, looking at the mean differences between the quintiles of the factor groups, also gives a good impression of the strengths of the effects. If the reviewer wishes, we could provide confidence intervals for these differences. This is usually done when effect sizes are less meaningful due to non-normality.

I cannot find drinks other than milk/dairy products and juices in your tables and text. How about other drinks? Sodas etc. Were they included? And if so, which food group were they allocated to? Diet sodas?

Beverages were not included in our study, because they were not fully available in our data set. However, we considered vegetable juices, fruit juices and milk, because they are parts of the food group sides of the German Food Pyramid and were available in our dataset.

Line 149: This is the first reference to a table - shouldn't they be numbered consequently starting from 1? Thus, this would reference to table 1

It was a reference to the food groups of the German Food Pyramid. As we included a figure of the pyramid in the revised version (Figure 1) we refer to Figure 1 now.

Line 168-9: I would like an explanation to how these standardized quantities were calculated.
We included the calculation of the standardized quantities.

- Line 219: Table 2. To me it makes no sense to present data as mean and SD when data are dichotomous or "number of persons…". It would be so much easier to understand if values were given as number/percentages.

- How should we interpret SD when -1SD is below zero? It signals that distribution is really skewed when in fact no measured values can be below zero. If it was up to me, I would consider using median and percentiles instead (given that data is quantitative).

We picked up the idea and used the values “percent”, “median” and “interquartile range”

- Line 233: I find it a bit confusing to state "the first pattern…” and finishing with the given name of that pattern. Consider 'flipping' the phrasing so that the name is given in the beginning of each description. I found myself leaping to the table to find what "the first pattern" was.

We changed that.

- Line 246: Table 3. I would appreciate a footnote or similar explaining why some figures are bold (of course it can be figured out, but better if you state it J) The food group "Meat (higher in fat)" - could you find another wording than higher? Maybe 'moderate fat content'? If I understand correctly, it is an in-between meat group in regards of fat content.

We included a footnote and changed the wording.

- Line 294-5: It is unclear to me what a 'given household size' is.

We mean “holding the household size constant” and changed that in the text.

- Line 304: I cannot find a 'Figure 2' in the manuscript? Neither a 'Figure 1'…

We included both figures in the revised version.
- Line 312-3: You state that the natural pattern uses supplements - but this result is not significant if your significance level is 0.01. However, it is unclear what your stated level of significance is, as commented above.

We used the 0.05 significance level and mentioned this in the section “statistical methods”.

- Line 319: Table 6. I would like an explanation (footnote maybe) to the "Age 2/100".

For all metric independent variables we checked whether they were linear or non-linear associated with the dependent variable. For age it turned out that the association with the processed and traditional pattern was quadratic. The division by 100 was done in order to indicate the effect of the independent variable without showing more decimal places. As suggested, we added a footnote to table 6.

- Line 344 and 368: This may be due to differences between countries, but in my country margarine is not necessarily lower in fat content than butter.

In Germany, many fat reduced and (omega 3) fortified types of margarine are on the market. We added this information in the text.

- Line 398: "Households with a higher adherence to the natural and thus healthier…” The 'thus' implies that natural = healthy which is a generalization that cannot be made. Please rephrase

We removed “thus healthier”

- Line 404-6: I think that one explanation to this controversy lies within the 'natural' thinking. Sugar is 'natural' whilst artificial sweeteners aren't - that's the general view in society, I believe. So - if you want to have the sweet taste but want to eat natural - you need to go with sugar. This might be extra pronounced in household with kids, not so easy to refuse the sweet taste for the children.

This is an interesting point. We added this explanation in the text.
Reviewer #2: 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.

We made a stronger link between the two sentences.

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.

We gave some examples for diet-related-diseases.

Line 58 - Lifestyle - all one word.

We corrected that.

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

We replaced “these diseases” by “diet-related-diseases”.

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

We re-phrased that sentence (“So far, only some studies have been carried out for Germany dealing with the identification of dietary patterns”).

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.

As suggested, we expanded on the differences between the data in the discussion (see line 438 ff)
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.

We deleted the sentence in this paragraph and mentioned this point in the discussion.

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

We re-phrased that sentence (“Hence, it is important to identify both the different diet quality issues in the patterns and characteristics associated with the patterns”).

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

We re-phrased that sentence (“To shed more light on these issues this study aims at identifying food patterns which is conducted on the basis of purchase data for the first time”).

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

We introduced an economic concept in the methods section (“It can be assumed that the attitudes of the person responsible for the household’s food purchases approximately reflect the attitudes of the household as a whole. This assumption can be derived from economic household theory where it is assumed that households are either a unified decision-making unit or they are individual members of multi-person households keeping their own preferences and constraints. Then they have, for example, a social welfare function reflecting a household consensus”).

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

We rephrased that paragraph (“This study aimed at identifying food purchase patterns and specific dietary issues associated with them. The results can be used for creating group specific dietary recommendations. In order to identify the household groups with their specific needs of improvement of diet quality, associations with socio-economic characteristics and attitudes were investigated. The attitudes reflect personal opinions and refer to dietary guidelines, supplements, and fortified foods”).
Methods

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

We reflected on that in the discussion (see line 438 ff).

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

It is the German micro-census. We clarified it in the text.

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.

We reflected on that in the discussion (see line 438 ff).

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

The BLS gives information on nutritional values for 14,814 foods available in the German market. We added this information to the text.

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

We reflected on that in the discussion (see line 438 ff).

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.

As suggested, we included an image of the pyramid to improve the clarity (Figure 1).

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.
We deleted the term “health values”.

Line 163 - 'a factor called pattern' - Are these your derived dietary patterns?
Yes, this is right. We clarified it in the text.

Line 184 - did you consider using a DAG or other causal diagram to check for confounder and mediators?
In principle, it is a good idea to include such a diagram. However, it cannot be included without commenting. As this article is already long we decided not to include a causal diagram. When explicitly required, we could insert it after consultation with the editor.

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.

Packaged foods are usually offered as retail brands (private labels) or manufacturing brands. Normally, retail brands are cheaper than manufacturing brands. We included a legend in Table 1.

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.

As a Cronbach’s Alpha value of 0.7 is an approximate reference value a value of 0.69 can be considered as acceptable. To avoid confusion we rounded the value of 0.69 to 0.7.

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.

We used all three measures mentioned in the literature because there is no consensus of which is the favored one. We mentioned this in the section “statistical methods”.

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?

In the revised version we used the values “percent”, “median” and “interquartile range”.

Can you include (n) for the number of people responding to each question?

We included n for 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.

As suggested, we presented the values as proportion of households and we used a decimal point.

Line 249 - 'pattern_had' - please remove underscore

We removed the 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.

We used the suggested form (journal style).

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

The selected nutrients were those for which an insufficient supply is noted in Germany. We mentioned this in the methods section.

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.
We chose four decimal places for all results.

Table 5 - P/S quotient foot note symbol should be 'f'
We changed that.

Line 273, 274, 275 - please rephrase - this does not read well.
We rephrased the sentence (“The percentage of calories from sugar was clearly positively associated with the score”).

Line 276 - consider saying 'this processes pattern score'.
We changed that.

Lines 293 - 295 - Is this effect determined by the presence of adults in the household? To be discussed in the discussion.
Yes, the reviewer is right. We included this implication in the discussion. (“Whereas household size was positively associated with the adherence to a food pattern in all three pattern groups, the presence of children showed a negative association. This implies that the presence of adults determined a stronger adherence to a pattern. Taking the factors age and the presence of adults and children into account, it can be assumed, that purchase patterns generally seemed to be more pronounced when the household consisted of older people”).

Line 300 - how does the non-linear relationship alter your modelling?
We included a quadratic term in our regression equation. Including the quadratic term led to a better fit of the model which was indicated by the r-square. We described this procedure in the methods section.

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.
We apologize for this mistake. In the revised version we included two figures.

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

Whereas a higher adherence to the natural and traditional pattern was positively associated with the level of prices the processed pattern showed a negative association. This is mentioned in the discussion.

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

We changed that.

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

In table 6 the unstandardized regression coefficients are shown. We replaced the $\beta$ by regression coefficient to avoid misunderstanding.

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 sub-headings or structured discussion to cover: 'strengths and limitations'. 'Future direction' & 'policy implications'

We tried to include all suggested points and inserted sub-headings in the discussion.

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

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

We linked the two paragraphs in the discussion.

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

We replaced “furthermore” by “moreover” or “in addition”

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?

It is one possible explanation. In the revised version, we expressed it more carefully (“This could support a finding of a previous study detecting that dietary habits are mainly formed at a younger age and then often remain stable. However, further analyses are necessary to confirm our assumption”).

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

We removed “particularly”.

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

Unfortunately we cannot give a meaningful reference. In the literature it is often mentioned that purchase data are probably less biased because households are not directly asked about their dietary habits, but there is no evidence.

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

We replaced some of the “furthermore” by “moreover” or “in addition”

Line 431 - why only fat consumption?
Because precise information on “fat contents” is necessary for this kind of analysis but this information is often not available in intake data. However, we removed this note to avoid confusion.

Line 439 - another ‘furthermore’ - please consider another word.

We replaced some of the “furthermore” by “moreover” or “in addition”

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.

As suggested by the reviewer, we removed detailed descriptions but focused more on general results and on policy implications.

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.

The German ‘Gesellschaft für Konsumforschung’ (GfK) have provided us with anonymised data so that no conclusion on certain households is possible. We included this explanation.