Author’s response to reviews

Title: Changes in food intake patterns during 2000-2007 and 2008-2016 in the population-based prospective Northern Sweden Diet Database

Authors:

Ena Huseinovic (ena.huseinovic@gu.se)
Agneta Hörnell (agneta.hornell@umu.se)
Ingegerd Johansson (ingegerd.johansson@umu.se)
Anders Esberg (anders.esberg@odont.umu.se)
Bernt Lindahl (bernt.lindahl@umu.se)
Anna Winkvist (anna.winkvist@nutrition.gu.se)

Version: 1 Date: 16 May 2019

Author’s response to reviews:

Dear Sir,

We have now again uploaded the manuscript main file with the figure legends inserted AFTER the references.

Sincerely,

Anna Winkvist

Dear Professors Collins and Kirkpatrick,

Thank you for considering our manuscript entitled “Changes in food intake patterns during 2000-2007 and 2008-2016 in the Northern Sweden Diet Database” and for providing us with the reviewers’ helpful comments. We have modified the manuscript in response to the reviewers’ comments and rewritten sections of the manuscript according to the suggestions (in red). We hope this complies with the reviewers’ remarks. Below, we respond to the comments point counter point. We have also now included our co-worker Anders Esberg as formal co-author. Anders has been involved in the manuscript all along but his efforts were largely increased in the current revisions because of all the new images that had to be created. Thus, we believe that he now has contributed on the level of being a full co-author.
Sincerely,

Anna Winkvist with co-writers

Reviewer 1

Reviewer #1: This paper describes dietary patterns, derived using latent class analysis, in a very large population-based sample of northern Swedish men and women spanning across two 8-year time-periods. While the dietary patterns are based on dietary information collected via a 64-item food frequency questionnaire, the examination of the dietary patterns over time and their associations with sociodemographic and health behaviour characteristics provide important insights into shifts in the patterning of food intake in this population and the identification of population subgroups with poorer dietary habits. Overall, this study is well-written and the analysis thoroughly conducted; I only have a minor comments for the authors.

Thank you for your helpful comments. Addressing these helped improve and clarify the manuscript.

Introduction

Lines 76-78: "Further work is needed to identify pockets in the population…” It is not clear why there is this gap in knowledge, and how the present study plans to address it. Numerous dietary pattern studies exist in the literature - what makes this study different (apart from the repeated measures and large study sample)? I think there is an opportunity here to emphasise the novelty of the approach taken to analyse the dietary patterns. Latent class analysis (LCA) has been rarely applied in nutrition research, and has a number of statistical advantages over the more commonly used cluster analysis. LCA is also a "person-centred" approach, which is in contrast to the "variable-centred" approach for extracting dietary patterns (e.g. PCA - the method most commonly applied in dietary pattern research to date). A person-centred approach identifies groups of people who have dietary behaviours in common which may be more useful for identifying at risk population groups. In summary, I would like to see the introduction better emphasise the novelty of the present study, with reference to the statistical approaches used to capture dietary patterns.

Thank you for this suggestion, we have now added and highlighted the strengths with LCA in the background section (p 4, lines 88-94).

Methods

Line 101-109: "Over the years, residents may have participated more than once". Please clarify how many of the same participants completed the assessments during both 8-year time periods.
We have now clarified that 32 visits were removed because they were repeat visits by the same individual within the same time window (Fig 1 and p 7, line 140).

Line 155-156: Please clarify how the participants were classified into the different physical activity categories. For example, what are the cut-offs for being allocated a physical activity category based on the participants' occupational and leisure time physical activity? Were the cut-offs based on time and days being active or intensity of activity or both?

We have now added details about this classification (p 8, lines 173-179).

Line 177: Continuous indicators were used in the Latent Class Analysis (often referred to as Latent Profile analysis when continuous data are used). Please clarify the nature of the data (e.g. whether it had a skewed distribution) and whether LCA is equipped to handle skewed dietary data

If an indicator is defined using the 'continuous' scale type, the normal distribution is assumed. However, the assumption is that the distribution is normal within each latent class which means that the distribution is characterized by a normal mixture distribution, not a normal distribution. As such, transforming it to normal is not a good idea.

Line 180-182: The authors have used a number of model fit indices to determine the number of classes (or clusters) present in the data. For transparency and reproducibility, I would like to see this information presented in a supplementary table. Also, please clarify if likelihood ratio tests were used (e.g. Lo-Mendell-Rubin LRT or Bootstrap LRT) to compare classes and, if so, please include this information in the table, and, if not, please justify why.

When one or more indicators are continuous, Chi-squared fit statistics (including the L2 statistic) are not available as the L2 statistic as well as the p-value associated with the L2 fit statistic are meaningless and these have therefore been omitted. Similarly, a bootstrap of the Lsq statistic is not available. Instead, the more general LL and the associated BIC have been used to determine the number of classes (the model with the smallest BIC). However, in addition to examining BIC values, the decision on number of classes/clusters was also based on class size and interpretability. Therefore, solely the BIC values are not sufficient to guide the reader on the number of classes selected as the best model fit.

Line 191: It is not fully clear to me why this sensitivity analysis was conducted using ). If the authors wanted to examine the reproducibility of the dietary patterns extracted using LCA, they could have repeated the LCA on a random split sample. If the sensitivity analysis was done to see if the type of patterns extracted were similar across different methods, then this needs to be clearer. Again, this highlights to me the need for more background information on the methods used to extract dietary patterns and the novelty of the approach taken in the study (as suggested in my earlier comment).

Thank you for this valuable point. We agree that the analyses with PCA may not represent a sensitivity analysis but rather a complementary analysis using a variable-centered approach. Our ambition was to evaluate the meaningfulness of the results from LCA by comparing with another
data driven method. This has now been strengthened in the manuscript (p 4, lines 88-94 and p 10, lines 213-215) and by naming it complementary analysis rather than sensitivity analysis.

Figures 2-5: The clarity of the figures could be improved as they are difficult to read.

We have now remade the figures in a new programme to improve the clarity.

Tables 3 and 4: Among women, the 'pulses and tea' cluster and 'soda and sweets' cluster represent <2% of the study population. These clusters could perhaps be considered as "outlier" clusters. Please discuss whether understanding these outlier clusters are useful from a public health perspective. As suggested previously, please also provide supporting data showing the model fit indices for the cluster solutions ranging from 2 to 7.

We agree that these are small clusters; however, as these consumption patterns were very prominent in both LCA and PCA, and in the extremes from a healthy diet aspect, we decided to keep the clusters and examine if specific background characteristics were related to these dietary structures compared to the other dietary patterns. With respect to the question on model fit, please see our previous answer.

Results

Lines 249-251: The "Healthy sandwiches" cluster included consumption of "cold cuts" (meat, sausage and liverwurst). Please clarify if cold cuts includes processed sandwich meats. If so, and given the WCRF recommendations to limit intakes of processed meats, then please reconsider the label and descriptions applied to this cluster.

Thank you for the suggestion, we have now renamed the cluster to “Sandwiches”.

Discussion

Lines 342. Please provide an overview of the Nordic recommendations on food habits somewhere earlier in the manuscript (e.g. introduction). This information will help the reader to interpret this statement about the study's findings and add needed context for understanding the shift in food intakes over time. Please also indicate if the Nordic recommendations have changed at all during the time-period under study.

We have now added a clarification about the Nordic Nutrition Recommendations in the Introduction (p 4, lines 74-81).

General discussion point: Is the shift in dietary intake towards higher consumption of high-fat dairy and high-fat dairy and vegetable oil spreads really of concern? That is, are the types of fats in these products actually harmful to cardiovascular health or is it merely the extra calories in the high fat foods that are of concern, given the current obesity prevalence? Please discuss the
evidence (including any recent evidence that may not be captured in the current Nordic dietary guidelines) in relation to the dietary advice for intakes these foods.

We have now added a paragraph regarding the relation between dairy intake and cardiovascular health (p 17, lines 392-399).

Reviewer 2

Reviewer #2: General comments

This article aims to understand changes in dietary patterns over a 16-y period and the associated sociodemographic characteristics by using a very large population-based study. The aim is very relevant, however it is unclear that the methods used are the most appropriate. The analysis and results are very hard to follow. The authors need to do a much better job synthesizing, interpreting and presenting in a clearer way their results. Moreover, it is not convincing that the main conclusion can be reached by their analysis.

Thank you for your comments, addressing these helped improve and clarify the manuscript. We have revised the manuscript thoroughly to improve clarity in relation to the specific comments made by the reviewers and also in general. This includes several clarifications on the methodology and its strengths and limitations. We think that the fact that we used both an individual centered and variable centered method to define food cluster groups and that these indicated similar clusters, support and substantiate our conclusion that these clusters exist in the target population.

When analyzing changes in dietary patterns, two type of changes could be assessed: 1) changes in the structure of the patterns (which foods are characteristic of each pattern), or 2) changes in the degree of adherence to dietary patterns (proportion of subjects in the pattern, or scores if doing PCA o factor analysis). However, to be able to look at 2) the patterns need to be the same, otherwise it is impossible to distinguish between the two types of changes. The authors compare the proportion of subjects in a similar pattern, but the patterns are not exactly the same, so the comparison is not straightforward. I believe that PCA o factor analysis is a more intuitive method for making comparisons of patterns over time, as factor loadings (for type 1), and subjects scores to applied factors (for type 2) could be compared. Otherwise, with latent class/cluster analysis it might be more suitable to perform the analysis with a dataset that includes both time periods and then compare the proportions of subjects from each period in the same pattern.

We agree that it is a challenge to use individual-centered approaches when comparing different populations or different time windows, because the clusters to be compared are not identical between the different populations/time windows. Still, for the current manuscript we wanted to use this method (LCA) as our core method because of its statistical properties. We also applied PCA to strengthen our argument that we have indeed identified meaningful patterns of intake. Including the entire time window in one analysis would mean that all individuals drive the patterns that emerge. Hence, patterns of importance to only one time window may be
overshadowed by patterns of importance to both time windows. We wanted to optimize our ability to identify clear patterns driven by women vs men and therefore separated them in the analyses, and we did the same for the two time windows. We believe that it is of interest to readers to see what kinds of clusters that emerge for women vs men, and for the two time windows, separately. That by itself is a finding. The comparison between results for one time window with another time window is similar to when comparing results from inductive qualitative research from two areas: different images may have emerged and this is the interesting result. Still, to help readers identify important differences between time windows we have now edited Figs 2-3 (previously Figs 2-5) so that both time windows appear together for each sex.

Moreover, the main conclusion "increase in intake of high-fat spread and high fat dairy and decrease in intake of fruit and vegetables" is not 100% supported by the analysis. Changes in dietary patterns cannot be translated into changes in absolute intake without further consideration. For instance, in men, the conditional mean of high-fat spread between the two periods do not seem that different.

Thank you for this comment. Focusing on the results of the analysis we have now rephrased this into: “..., the increase in clusters with a high conditional mean for high-fat spread and high-fat dairy and decrease in clusters with a high conditional mean for fruit and vegetables,...” (p 2, lines 45-47 and p 20, lines 478-479).

Specific comments:

Introduction:

Line 88, the aim is misleading, the only health outcome analyzed was BMI, and the main focus of the article was on sociodemographic characteristics.

We have now revised the manuscript and instead write “We examined food intake patterns and related sociodemographic and individual characteristics” (p 2, lines 28-29 and p 5, lines 103-104).

Methods:

An intervention was performed on the study population? If so, this should be stated and it should be addressed if the intervention could influence the results. This is mention on the discussion, but only briefly.

A comment has been added regarding the intervention, p 16, lines 374-387.

Line 111: What do the authors mean by "food intake level" and how is it different to energy intake.
We have now clarified this and describe that food intake level was calculated as the estimated total energy intake divided by basal metabolic rate (p 5, lines 125-126).

Line 112: It was described that assessments were made at 40, 50 and 60 years of age, please explain the <29 years exclusion criteria.

Until 1996, 30-year olds were also invited, but for financial reasons today this only persists in some communities. This information was missing in the manuscript and has now been added for clarity (p 5, lines 114-115).

What was the proportion of subjects with measurements in both periods analyzed?, the title refers to a prospective study, but by looking at mean age in table 1, it seems that it was mainly two cross-sections.

Thank you for noting this. We have now removed the word ”prospective” from the title and clarified that the data analyses in the present study were cross-sectional despite the longitudinal design of VIP (title page and p 6, lines 128-129).

Why was the frequency of food used for the patterns instead of total intake of the food group (kcal, %kcal or gr)?

In the literature, frequencies are more commonly used compared to absolute intake (in gram/kcal etc) for dietary pattern analysis. Also, in the VIP study, participants gave their responses as frequencies in the FFQ; thus, we decided to keep the frequencies in the analyses to reduce the risk of introducing further uncertainties and assumptions. Also, in previous research we have noted that results hardly differ at all when based on frequencies or on kcal or grams. This reflects that the original data are in the form of frequencies and the conversion to kcal or grams mostly happens via transformations using the same constant for all (eg the calories or the weight of one apple).

When selecting the best latent class solution, was there any criteria on the minimum proportion of subjects in each class? (e.g. <2% in some clusters might not be suitable).

Yes, less than 2% of participants is a common criteria used when selecting the best model fit and number of clusters. In the analyses, two of the clusters are <2%; however, as these consumption patterns were very prominent in both LCA and PCA, and in the extremes from a healthy diet aspect, we decided to keep the clusters and examine if specific background characteristics were related to these dietary structures compared to the other dietary patterns.”

Results:

The presentation of results is very hard to follow and understand. The authors need to re-think their presentation to make their key findings more easily identified. The figures are illegible and one table per period/gender makes comparison hard. Also, the authors might consider if that many food groups or doing the analysis by gender is necessary.
Thanks for the comments and we agree that we present many results. We have now revised previous figures 2-5 (now figures 2-3) so that both time windows are included for each gender, hence facilitating comparisons over time. Still, we want to separate results for the two genders and presenting them in one table would be difficult. We believe that the number of food groups are optimal for our analyses and would like to keep our presentation of the results as is.

The description of the sensitivity analysis is very long (for being a "sensitivity analysis") yet it is unclear. I was unable to follow it.

We agree that these results do not represent a sensitivity analysis in the true sense, and now describe them as a complementary analysis to strengthen the argument that we have identified meaningful patterns. We hope that makes this results part clearer.

Discussion:

Line 359 "Although the study design made evaluation of causality impossible" Please clarify, to which study the authors are referring? Causality between what and what?

We have now clarified this in the discussion section, p 16, lines 383-385.

In the 2nd, 4th and 6th paragraphs of the discussion the authors make comparisons with other studies that assessed trends in the intake of nutrients or food groups. As described above, the results from this analysis cannot be directly translated into trends in absolute intake. This is most likely why the findings are different (not the differences in dietary instrument nor the population as the authors implied).

We have now tried to clarify this in the discussion (p 18, lines 428-429). For example “Although the results from the present analysis cannot be directly translated into trends in absolute intake, a decreased applicability of the FFQ may also partly explain the finding of a decreased proportion of participants in clusters characterized by intake of fruit and vegetables, which is contradictory to findings from the national studies”