MSc Medical Statistics: Project Assessment 2017-18

Candidate's number: 110765	**Grade:** 4

Title of Report: The time and quantity patterns of carbohydrate intake in the UK adults – the National Diet and Nutrition Survey 2008/09-15/16 programme

First Marker Comments:

Overview: This project used data from the National Diet and Nutrition Survey 2008/09-15/16 programme to investigate relationships between eating times and carbohydrate intake, looking at both the amount and the timing of that consumption. It then extended the analysis to look for associations between these eating patterns and hypertension and obesity (for the latter using BMI and waist circumference).

Overall this was a very strong project. The analysis methods used techniques outside the MSc syllabus, specifically the handling of survey data, multilevel latent class analysis and Mplus software. The report was well presented, almost always carefully guiding the reader through the extensive results tables and figures, and very interesting to read. Several of the results were unexpected and the student put forward some suggested explanations for this. However, there were aspects of the report writing (mentioned below) that could have been improved, most notably the need to reference material taken from other sources even when presenting routine background information.

The comments below are presented in the context that this was a very good project so are suggestions for possible improvements.

Abstract:
The Abstract was clear but perhaps a bit long and could have been tightened in places. As a completely new reader it was also difficult immediately to understand the difference between 'eating day patterns' and 'types of carb eaters'. In the conclusions it would have been preferable to avoid phraseology such as "seemed to be" and "probably followed" as it was not clear whether this was author speculation or a loose presentation of statistical test results.

Introduction:
This was clear and informative, although I would have been interested to read in more detail about some of the existing nutritional research in the field, especially given the subsequent unexpected results. The overall report length was under 50 pages to there would have been space to do this.

The main issue was the lack of proper references in the section on NDNS. Different styles of writing made it easy to spot when material had been taken from other sources, for example: "NDNS provides essential evidence on the diet and nutrition of the UK population to enable PHE to identify and address nutritional issues in the population and monitor progress towards public health nutrition objectives." This appears verbatim at https://www.gov.uk/government/collections/national-diet-and-nutrition-survey. The student should review the rules for quoting verbatim and should have provided more careful references for the whole of this section.

Structure of the early part of the report:

It might have been better to group all the background information on NDNS under a standalone section heading, rather than split it between the Introduction and the Methods section. Generally I would expect a Methods section to focus on your own methods, i.e. the material that was presented under Statistical Methods. But the first 5 pages of Methods was primarily background information on the NDNS.

You could then have started the Methods section with a short specific sub-section explaining what NDNS data you chose to use and why e.g. 6,155 adults aged 19+ years, and information specifically about what you yourself did in terms of data management, such as the dichotomisation of the responses etc. That way there would have been a clear distinction between presenting other people's description of their survey data and the information/decisions that were specific to your report – some of the time it was difficult to tell what the student had done themselves.

As previously, some material needed referencing. For example, under 'Survey selection method' (p. 5) the first paragraph was clearly taken from this or a similar document (http://doc.ukdataservice.ac.uk/doc/6533/mrdoc/pdf/6533_ndns_yrs7-8_uk_user_guide.pdf) with only minor editing changes and should have been referenced as such, even though it was background information on the NDNS.

Statistical methods:

I would have liked to read a fuller description and general explanation of what latent class analysis does, why it is useful, and why you chose to use it for this project.

The analysis models were well presented, although as a reader I would have appreciated an earlier paragraph explaining that there would be a decision to be made about whether parametric or non-parametric approaches would be preferable. It came as a bit of a surprise to read in detail about the parametric models only then to be told that they would be too computationally demanding.

That said, you explained these complex models clearly and your strategies for conducting MLCA, and for building the logistic and linear regression models, were well expressed.

On the exploration of associations with hypertension and obesity measures, it would have been good to justify why you looked at men and women separately. If this decision was based on existing research, such information should have been mentioned. I assume you built separate models for men and women, rather than interaction terms, but this was not explicitly stated.

Results:

Again this section was very well laid out, providing explanatory and guiding text for each of the Tables and Figures. The writer made it easy to link the relevant paragraphs to the relevant Table/Figure, and highlighted what to look for in the figures/tables.

However, in the 2nd paragraph it was just stated that the 3x3 model was being adopted rather than the 2x4 model (or any other model). Since this was a key decision for the analysis I would have liked more justification. What was meant by "the most substantively interpretable model"? Appendix C provided some results from other choices, but without text. There was some material on Figures 3.2 vs 3.3 later on but it would have been good to have included in the main text some more information

about the choice the student had about the combination of class numbers and why the lowest entropy 2x4 approach was not followed.

I found the characterisation of the 3 latent groups at the day level to be a bit awkward because it seemed to mix the contrasts of high/low percent carbs with regular/irregular mealtimes. This made interpretability harder and it was more difficult to present trends. After all, someone who keeps regular mealtimes may have high or low percentage carb intake. LCA does throw up classes that sometimes do not lend themselves to easy characterisation and this difficulty might have been worth discussing.

There was a lot of information for the reader to digest in Figure 3.4 and Tables 3.4 to 3.11. The text summaries were clear and accurate but it would have been good to pick out what were considered the key points and perhaps highlight them in bullet points, accompanied by the supporting quantitative results. The reader needed a bit more help here to work out the key messages.

Looking at the definitions and values for a couple of variables raised a couple of thoughts about their role in the modelling. Physical activity (moderate and vigorous), e.g. in Table 3.6 at around 1 hr/day, seemed much higher than is normally reported these days in the UK.

The Living with partner definition used by NDNS is narrow according to the definition given below Tables 3.5, 3.8, 3.9 ("Live with partner was defined as either living with a married husband/wife or a legally recognised civil partnership"), ignoring the large proportion of adults who cohabit long-term without marriage/civil partnership. Since the latter was a variable that featured in the detailed results and the Discussion, this might have been pointed out – it seems curious that it is the formalisation of a relationship rather than the actual living together that would be relevant, but it's possible.

Discussion and conclusion:

This section was interesting to read and the writer had clearly thought about the results carefully. But it would have been good, wherever possible, if the student had also set this section in the context of some other relevant research, in order for the reader to see if there was consistency with whatever other results do exist.

I also found that some statement were made without supporting material. The discussion was set up by the writer stating two lines of thought:

- "… one might anticipate that individuals followed much of class 3) days, the regular meals day, might be eating a healthier diet because of regular eating habits.
- We might also speculate that those who followed either high or low carbohydrate percentage days would consume higher total energy than those who followed mostly regular meals days."

In the absence of evidence from the existing literature both these statement seem open to challenge: regular eating is not necessarily going to be healthier eating; Table 3.3 shows that low carb days have lower total energy than regular meal days.

The writer grappled well with the difficult task of distilling out key messages about carb eating patterns but got it right in saying the study had "highlighted the complexity of eating behaviours". The low carb/high carb/regular categories for the day latent classes made drawing out these key messages difficult, as mentioned earlier.

For instance, when stating that "contrary to what was anticipated at the beginning, people who followed high percentage carbohydrate days for most of their time were potentially eating a healthier diet compared with the other two eating patterns" the writer seemed to be defining healthier diet just in terms of not eating much after 8pm.

It would also have been interesting for the writer to consider whether, in the context of other existing nutritional research, the unexpected results possibly signified a problem with the approach or the actual models.

In the discussion of the associations with health outcomes, the material on the interaction for women with 'living with partner' needed to take on board the narrow definition, as mentioned earlier.

The Limitations and Strengths section was excellent, with very good use of references and setting the points made in a wider context.

Overall this project demonstrated novel techniques that were able to identify and describe some of the complexities in eating patterns in the UK population. The success of the project suggests that further development of these methods would be worthwhile and could hopefully lead to a published paper.

Second Marker Comments:

This project report was very good. It was well-structured and well-written.

The project entailed familiarising with methods not covered in the syllabus of the MSc in Medical Statistics, namely the analysis of survey data and latent class analysis, in a version including hierarchical data/repeated measures. You showed ability to acquire and investigate such methods, particularly latent class analysis which you also describe in the methods and show to thoroughly master in your data-analysis with M-Plus (again a package not taught in the Master) that was particularly impressive. The analysis of survey data is also correctly applied although perhaps the methodological background for it is not as *formally* explained in the methods which, on the other hand, dwell very much on describing the design of the nutrition survey. The latter was extracted from the appendices and report of NDNS, some of which may have not been necessary in such detail and could have just been referenced to save space to report more of what was of your own making and needed for the analysis.

In terms of results reported the figures and tables are very well presented and clear throughout the report.

Perhaps more explanation could have been provided as to the reasons why the alternative MLCA models were not amenable to interpretation, making an example of an alternative one, rather than just reporting them in the appendix. This may have helped also to highlight as additional limitation the difficulty in interpreting an involved model, with different typology of day based on the day time spectrum (expressing the probability of 3 mutually exclusive events at different time slots) which give rise to typologies of people based on the mix of probabilities that they follow each day. What would be the consequence on the interpretations of the results of different choices made?

Another aspect of results which could have been improved concerns how the association (between latent classes and outcomes) analyses were conducted: these appear to have been stratified by sex throughout (so separately by men and women) rather than interaction being calculated first , also with sex like you did with other variables, to be consistent with the fact that you built the latent class model using the whole sample rather than by sex strata (which could have been a viable option).

The discussion is excellent – clear and comprehensive. Again, you could have just reflected/discussed also some alternative models (especially with regard to the difficulties of interpretation) that you did not present (or only mentioned in the appendix) and speculated on possible different methods or choices of modelling that could help proceeding further in unravelling the nutrition issues that motivated this work.

All these are just minor points to help improve further a report that was really very good, well done!