Chaochen Wang

The National Diet and Nutrition Survey (NDNS) database of detailed four-day food diaries was used in exploration of both temporal and quantity eating patterns of carbohydrates in the general adults living in the UK. 6155 participants aged 19 or older were kept in the analysis.  
Multilevel latent class analysis (MLCA) models were applied in finding the latent classes of carbohydrate consumption both at the observation day level, and individual level.

image

MSc Project Report  
2017-2018

*Supervisor:*

**Submitted in part fulfillment of the requirements**  
**for the degree of**

September 2018

I, , declare that this thesis titled, “” and the work presented in it are my own. I confirm that:

* This work was done wholly while in candidature for a MSc degree in Medical Statistics at London School of Hygiene and Tropical Medicine.
* No part of this thesis has previously been submitted for a degree or any other qualification at London School of Hygiene and Tropical Medicine or any other institution.
* Where I have consulted the published work of others, this is always clearly attributed.
* Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
* I have acknowledged all main sources of help.
* Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: image

Date: 2018-09-05

I would like to thank my tutor and supervisor , for his guidance, patience, and help while working on this project and also to Dr. Suzana Almoosawi for her invaluable nutritional academic insight, and suggestions.  
Thanks to Raoul Mansukhani, for sharing his previous work and thoughts on the analyses, methodology in helpful discussions.  
I would like to thank the team of the National Diet and Nutrition Survey Rolling Programme (NDNS RP) who have made their data available to the public for academic study.  
I would like to thank all of the teachers, lecturers, staffs, and fellow course mates in the Department of Medical Statistics for providing their wonderful teaching techniques, sharing their excellent ideas that made this year such fruitful and enjoyable.  
Last but not the least, I want to express my gratitude to my family for their unconditional support, understanding, and encouragement throughout this year.

“*All models are wrong, but some are useful.*”

George E. P. Box

openrightfalse

ll **AIC** & **A**kaike **I**information **C**riterion  
**A1c** & Haemoglobin **A1c**: Glycated hemoglobin  
**BMI** & **B**ody **M**ass **I**ndex  
**BIC** & **B**ayesian **I**information **C**riterion  
**CI** & **C**onfidence **I**nterval  
**DM** & **D**iabetes **M**ellitus  
**FSA** & **F**ood **S**tandards **A**gency  
**HDL** & **H**igh **D**ensity **L**ipoprotein cholesterol  
**LDL** & **L**ow **D**ensity **L**ipoprotein cholesterol  
**LCA** & **L**atent **C**lass **A**nalysis  
**LCGA** & **L**atent **C**lass **G**rowth **A**nalysis  
**LTA** & **L**atent **T**ransition **A**nalysis  
**MAFF** & **M**inistry of **A**griculture, **F**isheries and **F**ood  
**MLCA** & **M**ultilevel **L**atent **C**lass **A**nalysis  
**ML** & **M**aximum **L**ikelihood  
**NDNS** & the **N**ational **D**iet and **N**utrition **S**urvey  
**OR** & **O**dds **R**atio  
**PHE** & **P**ublic **H**ealth **E**ngland  
**PSUs** & **P**rimary **S**ampling **U**nits  
**RMLCA** & **R**epeated **Measures** **LCA**  
**TC** & **T**otal **C**holesterol  
**TG** & **T**ri**g**lycerides  
**WC** & **W**aist **C**ircumference

# Introduction

## Background

The widely accepted standard these days seems to be that we eat three times a day. However, whether this is really an ideal temporal eating pattern for everyone has never been answered with scientific evidence. More importantly, how many temporal patterns of eating are there in the population, proportions of people who actually manage/fail to follow this doctrine, are people consistently following one specific temporal eating pattern or do they switch, have not been studied and described thoroughly either.

Although nutritional studies have extensively examined the influence of the quantity and quality of dietary and nutrients intake and their alteration on morbidity and mortality, investigations on temporal eating patterns and their effects are still scarce. The importance of the circadian rhythm in regulating physiological responses has been recognised for long, while the impact of which on nutrition and metabolism is still largely unknown . Some recent evidence have found that meal timing is associated with a wide variety of health outcomes. Skipping breakfast is associated with higher risk of type 2 diabetes . Shift workers have a higher risk of developing metabolic syndrome and type 2 diabetes . Evening intake of energy is positively associated with overweight/obesity .

More recently, discernible temporal eating patterns that differed by sociodemographic and eating profiles were revealed by latent class analysis using nutrition survey data . Based on total energy consumption, the presence of 3 groups of eaters: grazers, early eaters, and late eaters were identified. So far, the temporal eating patterns were only based on averaging the total energy intake calculated from dietary recall, and therefore could not capture the day-to-day variation in temporal eating patterns. Thus, the question of how much variability of a person follows one or several specific temporal eating patterns in his/her everyday life remains unanswered. Many factors, such as day of the week or season, or culture may contribute to daily variation in dietary intake, however, most of the variation in an individual’s diet may seems to be without an obvious pattern. Intakes of macro-nutrients (carbohydrate, fat, and protein), due to the reason of their large contribution to the total energy intake, may have somewhat moderate degrees of day-to-day variation . Novel analytic methods that can account for this within person day-to-day variation is needed.

In the current report, we focused on temporal eating patterns for carbohydrate consumption in the UK adults. Eating more carbohydrate in the morning has been found to be negatively associated with metabolic syndrome . On the other hand, high total consumption of carbohydrate has been linked with higher risk of type 2 diabetes . Whether the amount or the timing (or both) of carbohydrate consumption during the day actually matters, would be the question concerned by the public.

## The National Diet and Nutrition Survey (NDNS)

The National Diet and Nutrition Survey (NDNS) programme was initially established in 1992 and started off as a joint initiative between the Ministry of Agriculture, Fisheries and Food (MAFF) and the Department of Health. In 2008, a new continuous cross-sectional survey was started, the NDNS Rolling Programme (RP). The NDNS RP is funded by Public Health England (PHE), an executive agency of the Department of Health, and the UK Food Standards Agency (FSA). The survey covers a representative sample of around 1000 people per year. Fieldwork began in 2008 and is now beginning its 11th year. 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.

The NDNS RP has now completed and analysed its eighth year. The sample was randomly drawn from a list of all the addresses, clustered into postcode sectors from across the UK. Overall, for years 1-8 combined, a sample of 39,300 addresses was selected from 799 (year 1-4), 323 (year 5-6), and 316 (year 7-8) postcode sectors. At each address, one household was selected at random (in cases where there were two or more households). For each household, either an adult and a child, or a child only, was selected to participate.

These individuals were asked to keep a four-day diary on their food and drink consumption on consecutive days. An interview and a nurse visit were also conducted to collect information regarding height and weight, smoking and drinking habits, physical activity, blood pressure, prescribed medicines, dietary supplements, fasting blood sample, and 24-hour urine sample.

## Aims and objectives

Our goal is to explore and make use of the NDNS RP (2008/09-15/16) database to describe and identify the potential relationship between the timing of eating within the day and specific nutrient–carbohydrate intake. We aimed at finding time patterns of consumption of carbohydrate and defining latent groups in the UK adults. Subsequently, an additional potential aim, is to investigate the association between carbohydrate eating patterns with hypertension and obesity.

# Methods

## Dietary diary collected in the NDNS RP

Participants were asked to keep a record of everything eaten or drunk over four consecutive days. Interviewers undertook three visits with each participant. At the first visit, the interviewer explained the method followed a protocol, taking participants through the sections in the diary including how to describe details of food and drink and portion size and an example day. The second was a brief visit to check for compliance, answer questions or deal with problems and review the diary to identify and edit possible omissions and missing detail. The third visit was to collect the diary and again review and edit possible omissions.

In the diary, participants were asked to record portion sizes in household measures (e.g. one tablespoon of beans, one Kit Kat finger-size), or for packaged foods to note the weight indicated on the packet. For homemade dishes, participants were asked to record on a separate page in the diary the individual ingredients and quantities for the whole dish along with a brief description of the cooking method and how much of dish they had consumed. In each eating occasion, in addition to the details of what and how much was eaten, participants were also asked to record: when was it, where they were, and who they were eating with. An example, used as guidance for participants, of a food diary for one day is shown in **Appendix [AppendixE]**.

### Definition of carbohydrate intake

Detailed dairy checking was performed to code and convert the food consumption into energy and nutrients intake. Intakes of nutrients were calculated from the food consumption records using a specially adapted Nutrient Databank , which was originally developed by the Ministry of Agriculture, Fisheries and Food (MAFF) for the Dietary and Nutritional Survey of British Adults. Further details of data coding and editing are outlined in Appendix A of the NDNS official reports . Specifically, the main variables that we adopted in the current analysis were defined as:

* Total energy intake = (protein(gram) 17) + (fat(gram) 37) + (carbohydrate(gram) 16) + (alcohol(gram) 29) kJ;
* Carbohydrate intake = total sugars (gram) + starch (gram);

Time across a typical survey day was divided into 7 time slots in the dietary diary of NDNS RP: 6 am to 9 am, 9 am to 12 noon, 12 noon to 2 pm, 2 pm to 5 pm, 5 pm to 8 pm, 8 pm to 10 pm, and 10 pm to 6 am next morning. To produce a sequence of discrete responses regarding the carbohydrate intake we are interested, the energy consumption within each time slot over the four days of survey for each participant were calculated. The percentages of energy that contributed by carbohydrate within each time slot were then estimated. Since we planed to apply latent class analysis (LCA) in the current study, in which the observed indicators for latent classes must be categorical, we then dichotomised the responses according to the carbohydrate contribution to the energy intake at cut-off value of 50%, i.e. if within a time slot where there is any energy intake occured, carbohydrate consumption was categorised into whether it’s energy contribution was lower or higher/equal to 50% of total energy intake within that time slot. Consequently, for each day of the recording, there were 7 data points generated by the diary, each data point included one of the following responses:

* Not eating any food (energy intake = 0 kJ);
* Eating, and carbohydrate contributed less than 50% of the total energy intake;
* Eating, and carbohydrate contributed higher or equal to 50% of the total energy intake.

## Survey Data

### Survey selection method

The NDNS RP participants were drawn from the UK Postcode Address File, a list of all the addresses in the UK. The addresses were clustered into Primary Sampling Units (PSUs), small geographical areas, based on postcode sectors, randomly selected from across the UK. A list of 27 or 28 addresses was then randomly selected from each PSU.

Overall, for years 1 to 8 combined, a sample of 39,300 addresses was selected from 1,438 PSUs. The sampling selection process was:

* Randomly select PSUs from the Postcode Address File;
* Randomly select 27 or 28 addresses in that postcode area;
* Randomly select one household at that address;
* Selected addresses were randomly allocated to one of two groups to determine whether an adult (aged 19 years or older) and a child (aged 1.5 to 18 years), or a child only, were selected for interview.

### Response rates

The response rates for completion of the dietary diary (three or four days) were 56%, 53%, 53%, for years 1 to 4, 5 to 6, and 7 to 8, respectively. A total of 6,155 adults aged 19 years and older were kept in the current study.

### Strata and weightings

It is necessary to apply weighting factors to the data collected in the NDNS RP for two reasons: to remove any bias in the observed results which may be due to differences in the probability of households and individuals being selected to take part; and to attempt to reduce differential non-response bias by age, sex and geographical region.

The strata that used to calibrate proportions in the sample include: age-group (1.5-3, 4-6, 7-10, 11-15, 16-18, 19-24, 25-29, 30-39, 40-49, 50-59, 60-64, 65-69, and over 70 years); sex (men or women); and regions (Northern Ireland, Scotland, Wales, and the nine regions of England).

Two steps of weighting system are designed in the NDNS RP to assure that the combined sample will be representative of the UK population:

1. An overall selection weight, which is the product of the address, dwelling unit, catering (household) unit, and individual selection weights, was generated to correct for the unequal selection probabilities. These weights are the inverse of the selection probabilities at each level of the random sampling process and they can be used to compensate for differences in the chance of selection of an individual.
2. An iterative procedure was used to adjust the selection weights until the distribution of the weighted sample matched that of the population for age-group, sex and geographical region. Population distributions were taken from the mid-year population estimates .

Another two sets of weights were generated to correct for differential non-response (either due to refusal or inability) to 1) nurse visit, and 2) giving a blood sample. Response rates to the nurse visit among those completed a dietary diary was approximately 75%, to blood sample in adults were 51%, 57%, and 50% for years 1 to 4, 5 to 6, and 7 to 8, respectively. In creating the nurse/blood sample weight, a logistic regression model was used by the NDNS RP study team to model the relationship between response to nurse visit/giving blood sample and a set of predictor variables (socio-demographic, participant and catering/household unit characteristics). The model generated a predicted probability for each participant, which is the probability would agree to a nurse visit/provide blood sample, given the characteristics of the individual and the household unit. Participants with characteristics associated with non-response were under-represented in the sample and therefore receive a low predicted probability. The inverse of these predicted probabilities were used as a set of non-response weights so that participants with a low predicted probability got a larger weight, increasing their representation in the sample. Then the nurse/blood sample weights were re-scaled so that the sum of the weights equalled the number of participants who had a nurse visit/who provided a blood sample. The final nurse/blood weights should therefore make the sample participants representative of all eligible persons in the population.

Further details of the weighting system developed by the NDNS RP are described in the Appendix B of the reports published by Public Health England (PHE) .

### Socio-demographic status, lifestyle, physical activity, anthropometric measurements and biochemical analyses

Computer assisted personal interviews were conducted for the selected individuals by trained interviewers to collect background information on smoking habits (current, ex-smokers, and never), ethnicity (white, non-white), education level (lower than degree/degree or above level), living with a partner or not and other socio-demographic related variables. Participants also had their height, weight, blood pressure, waist circumferences (WC) measured by the nurses.

Specifically, blood pressure was measured in a sitting position using an automated, validated machine, the Omron HEM907, after a five minute rest. The means of second and third readings, taken at one minute intervals, were used in the current report. Hypertension was defined as with systolic blood pressure of 140 mmHg or above, and/or diastolic blood pressure of 90 mmHg or above, and/or taking any medication specifically to reduce blood pressure.

A self-completion questionnaire - the Recent Physical Activity Questionnaire (RPAQ, developed by the MRC Epidemiology Unit Cambridge) was used to estimate physical activity from year 2 of the survey. The RPAQ was designed to assess usual physical activity in the last month in four domains: home, work, commuting to work, and leisure activities. Detailed descriptions of the assessment of adult physical activity in the NDNS RP and the processing of data are available in Appendices G and V of the published reports .

Blood samples were stored at 4 C, and sent directly by post to the Department of Haematology and Department of Clinical Biochemistry and Immunology, Addenbrooke’s Hospital, Cambridge within two hours of their collection. Serum samples were obtained by centrifugation of the coagulated blood sample. Serum total, High Density Lipoprotein (HDL) and Low Density Lipoprotein (LDL) cholesterol, triglycerides (TG), fasting blood glucose, haemoglobin A1C were measured. A1C value of 6.5% was used as the cut off point for diagnosing diabetes.

Body mass index (BMI) was calculated as weight in kilograms divided by height in square meters. BMI was then categorised into less than 25 kg/m2 (normal weight), 25 to 30 kg/m2 (overweight), and higher or equal to 30 kg/m2 (obese).

### Ethical approval

Ethical approval for the survey was obtained form the Oxfordshire A Research Ethics Committee. The letters of approval for the original submission and subsequent substantial amendments, together with approved documents, were sent to all Local Research Ethics Committees covering areas where fieldwork was being conducted. Research governance approval was sought for all participating NHS laboratories and obtained where required by the Research and Development Committee for each laboratory. Ethical approval for the current project was obtained from the MSc Research Ethics Committee of London School of Hygiene & Tropical Medicine (LSHTM MSc Ethics Ref: 15624).

## Statistical methods

### Latent Class Analysis (LCA)

Latent class analysis is a statistical technique that identifies categorical latent (unobserved) class variables on the basis of observed categorical variables . It belongs to the family of latent variable models, and is directly analogous to the factor analysis model. The major difference is that the latent variable in LCA is categorical, not continuous as in factor analysis. The basic assumptions in LCA are independent observations, and local independence, as shown in the fundamental expression of a typical LCA model:

Where,

* is the probability of observing a particular vector of responses;
* is the probability that a randomly selected th observation will be in class ;
* is the probability of a particular observed response pattern conditional on membership in latent class .

**Equation [LCA]** indicates that responses for an observation to the measuring variables are independent of one another given its membership in latent class . However, in the NDNS RP data set, the assumption of independent observations is violated. Each individual completed their dietary diary for four consecutive days, their diary recordings were later converted into four sequences of categorical responses reflecting the type of carbohydrate consumption at each time slot of the day. The four observed sequences (observed days) are nested in the participants and therefore are not independent. This nested data structure requires multilevel techniques.

### Multilevel Latent Class Analysis (MLCA)

Multilevel latent class analysis accounts for the nested structure of the data by allowing latent class intercepts to vary across level 2 units and thereby examining if and how level 2 units influence the level 1 latent classes. These random intercepts allow the probability of membership in a particular level 1 (observation days) latent class to vary across level 2 units (e.g., here in the current context are the individuals). Essentially this allows the probability that an observation day will belong to a particular day level latent class to vary across individual levels.

#### Parametric approach

Proposed by Vermunt and Asparouhov and Muthén , a traditional, parametric approach can be applied using a logistic regression model. For example, let’s assume that there are two types of observation days in the dietary survey–high and low carbohydrate eating days. In an unconditional logistic regression model, the probability of the outcome (i.e. an observed high carbohydrate eating day vs. a low carbohydrate eating day) is constant within individual level, which means for each person throughout his/her survey there is some probability of following a high carbohydrate eating day. A random effect considers the individuals (level 2) to be drawn from the adult population in the UK, and the probability of the outcome (i.e. high carbohydrate eating days) across individuals is considered to be a random variable .

Thus, for a binary outcome (low or high carbohydrate eating days), where denotes the observation days , and denotes the individual . The 2-level random intercept logistic regression model can be expressed as:

Where we define:

* ;
* as the probability that the randomly selected th observation day of th individual is a high carbohydrate eating day;
* as the random intercept, for outcome ;
* the random deviation of the individuals are assumed be normally distributed (i.e. ), the magnitude of the variance () indicates the influence of the individuals (level 2);
* is the predictors for day level (weekdays or weekends) and individual level, such as age, and sex.

Same framework can be used to consider random effects in an LCA model, but instead of saying that is either low or high carbohydrate eating days as if we already know, it is now replaced by a latent variable which indicates the typologies of carbohydrate eating patterns. Then we can use the day level data to assess the log-odds of belonging to th type of carbohydrate eating pattern on a specific day of survey, and we allow the log-odds to vary across individuals. Therefore, for some persons the log-odds of having a th type of carbohydrate eating pattern during the survey can be high, but for the other persons, the log-odds of following the th type of carbohydrate eating pattern can be low.

If the day level LCA model (carbohydrate eating temporal pattern typologies) is best defined by latent classes, then random intercept will be specified by a two-level multinomial logistic regression model. Similar to the typical LCA models, the latent class variable in a MLCA is defined by multiple observed indicators (here is defined by the responses of eating carbohydrate within each time slots, throughout 4 consecutive days of survey period). Considering the latent class indicators are indicator variables (), the MLCA model can be written as follows:

Where,

* represents the response of eating carbohydrate (one of the following: not eating any food, of the energy, or of the energy) on th day of the survey () in th individual at the th time slot of the day ();
* denotes the latent class membership for th individuals on th day of the survey, the total number of day level latent class is ;
* is the probability of a specific response pattern, conditional on membership in latent class .

The in equation [MLCA] is what we have already defined in equation [randomLCA]:

#### Non-Parametric approach

Since the parametric approach discussed above can be extremely computationally demanding , an alternative approach is using a non-parametric MLCA . In this approach, separate latent class models are specified for level 1 (observation days) and level 2 (individuals). Similar with the parametric MLCA approach, there are random intercepts, where is the number of level 1 latent classes. However, rather than assuming the random intercepts following a normal distribution, the non-parametric MLCA assumes a multinomial (discrete) distribution of the level 2 latent classes. This approach is less computationally demanding compared with the parametric approach. These level 2 (individual) latent classes reflect differences in the probability of belonging to a specific day level latent class, so that individuals that contain observation days with similar probabilities for the level 1 latent classes will be grouped together. The non-parametric MLCA model can be defined as follows:

Where,

* is individual level latent class membership for th individual;
* is day level and individual level indicators.

According to Finch and French’s simulation study , non-parametric approach generally resulted in more accurate recovery of the underlying latent structure of the data at both levels and provided better latent class model compared with parametric approach. In the current project, we are interested in exploring both meaningful individual (level 2) latent classes and the daily (level 1) carbohydrate consumption classification. Therefore, non-parametric MLCA was employed 1) to identify latent classes of observation days (level 1) based on the subjects’ responses to the 4-day food and drink diary and 2) to form distinct latent classes of individuals (level 2) based on the distribution of day-level carbohydrate eating temporal latent classes within individuals.

### Strategy of conducting MLCA

To identify the best-fitting model, we used the following sequential modelling strategy :

* Firstly, we ignored the multilevel structure of the data and estimated a series of traditional LC models to determine the number of classes at the observational day-level;
* Next, a series of MLCA models were fitted to account for the multilevel structrure of the data. In these models, the number of day-level classes was based on the best fitting LCA model from the first step, and the LCA model at the individual level was estimated to identify the number of individual level latent classes;
* Thirdly, when number of individual level latent classes is defined based on the previous stage, day-level classes was modified (one class lower and one class higher than in the second step) to see the effect of changing level 1 classes and confirm the best fitting model.

The number of classes in either level 1 were determined by 1) the evaluation of model fit indices, including the Bayesian information criterion (BIC), and entropy which is a statistic that summarizes latent class probabilities where values near 1 indicate better latent class separation; 2) the Lo-Mendell-Rubin Likelihood Ratio Test (LMR-LRT) which compares vs. classes models, where is the number of latent classes and most importantly, 3) pattern interpretability. In the steps of performing multilevel LCA, where LMR-LRT is not available, same rules of model fit indices and pattern interpretability were used to determine the optimal combination of latent classes in observation day level and individual level. MLCA models were fitted in Mplus 7.4 , the Mplus syntax and outputs are shown in **Appendix [AppendixB]**.

### Characteristics of day level latent classes and individual level latent classes

Day level latent classes identified by the first step of MLCA were tabulated according to day of week and also whether the diary was recorded during weekends or not. A contingency table giving the frequency of responses across the 7 time slots of the survey days was produced. Descriptive statistics for the dietary day level recordings according to the latent class memberships were presented. Pearson test was used to compare the distribution of categorical variables. One-way Analysis of Variance (ANOVA) was used to compare the means across the multiple groups for continuous variables.

Person level point estimates and 95% confidence intervals (CIs) were determined by applying individual, nurse visiting, and blood sample weights accordingly which account for the probability of participant selection and the clustered survey design. Descriptive statistics for sample characteristics are presented as weighted means (95% CI) or weighted percentages (95%CI). After examining the distribution of the data, the following variables were log-transformed to improve normality: fasting blood glucose, A1C, TC, LDL, HDL, TG, and average physical activity duration per day. Weighted geometric means (95% CI) were used for all log-transformed variables.

To see whether there is any temporal pattern for food intake eating could also be defined at individual level, weighted estimates of nutrients consumption across the 7 time slots of the day were calculated for each individual level latent class. Contributions (%) of the average energy intake within time slots were evaluated by determining the percentages of energy coming from carbohydrate, fat, protein, and alcohol intake.

For continuous variables, the *F* test was used to determine differences between latent classes with Bonferroni correction to account for multiple testing across 2 classes when applicable. For categorical variables, differences between latent classes were assessed using the adjusted Pearson test for survey data.

### Association between individual level latent classes and the prevalence of hypertension, and measurements of obesity

Associations between individual level carbohydrate eating classes and hypertension (yes/no), body mass index (BMI, kg/m2), and waist circumference (WC, cm) were explored in men and women separately. Point estimates of weighted means and proportions and 95%CI of the characteristics were determined by applying either nurse visiting weights (for outcomes of hypertension, BMI, and WC) or blood sample weights (for diagnosis of DM) accordingly. Similarly, *F* tests (for continuous variables) and adjusted Pearson tests (for categorical variables) were used to determine sex-specific differences by hypertension status, and BMI categories.

Survey-designed logistic regression models (for hypertension), and linear regression models (for WC, BMI), were used to test for associations between latent classes of carbohydrate eating patterns and hypertension, BMI, and WC, in the NDNS RP sample, separately. Since diabetic participants might or might not modify their carbohydrate eating habits, we also fitted all the above mentioned regression models restricted to those without diabetes.

For the multiple regression models, model fitting strategies are as follows:

1. The crude association between the carbohydrate eating groups and the outcomes was first examined.
2. Potential confounders of the association between carbohydrate eating groups (exposure) and the outcomes were selected depending on the descriptive statisitical analyses conducted above, i.e. those are associated with both the exposure and the outcome and also not on the causal pathway were selected as potential confounders. Those are strongly related with the outcomes but may not associated with carbohydrate eating groups may reduce the standard errors and so improve the precisions are also considered in the linear regression models.
3. Confounding and/or interaction effect from each of the potential factors were checked one by one. Interaction effect were tested using the adjusted Wald test testing whether the regression coefficients of the interaction terms are simultaneously equal to zero.
4. A preliminary model that includes all of the variables suggested to be confounders in the previous step was established.
5. The remaining variables were added to the preliminary model one by one to see if any of them may be a confounder in condition of the presence of the other covariates.
6. For logistic regression models (hypertension) under the survey data, goodness-of-fit was assessed using the adapted svylogitgof command in Stata . Other diagnostics for regular logistic regression models, such as estimating the pseudo-R2, AIC or BIC, checking the standardized Pearson residuals, or covariate pattern residuals are currently not available for weighted survey data.
7. For linear regression models (WC, BMI), assumption of independent observations is violated as soon as we weighted the sample. General checking such as QQ plots of the residuals (normality), plotting the residuals against fitted values (constant variance) are not available as well. Outliers, leverage, and Cook’s distance cannot be check either, however, participants with extreme weightings (if exist) were checked by removing them and refit the models as a sensitivity analysis.
8. Since under survey design data, the sampling-weighted least squares are not maximum likelihood, it would not be possible to compare models using likelihood ratio test. Instead, adjusted Wald tests with were used as criteria for variable inclusion in the final model. Another Stata command linktest was also used to decide whether quadratic and cubic terms of continuous variables were necessary in improving the fitting of model .

Data manipulation and preparation **(Appendix [AppendixA])** were done in R version 3.5.1 . All statistical analyses, except for MLCA models, were performed with svyset command as implemented in Stata software version 15.1 . The process of model fitting, covariates selection, interaction effect testings for the association between carbohydrate eating patterns and hypertension is shown as an example in **Appendix [AppendixD]**. All *p* values were two-sided.

# Results

### Model selection, and interpretation

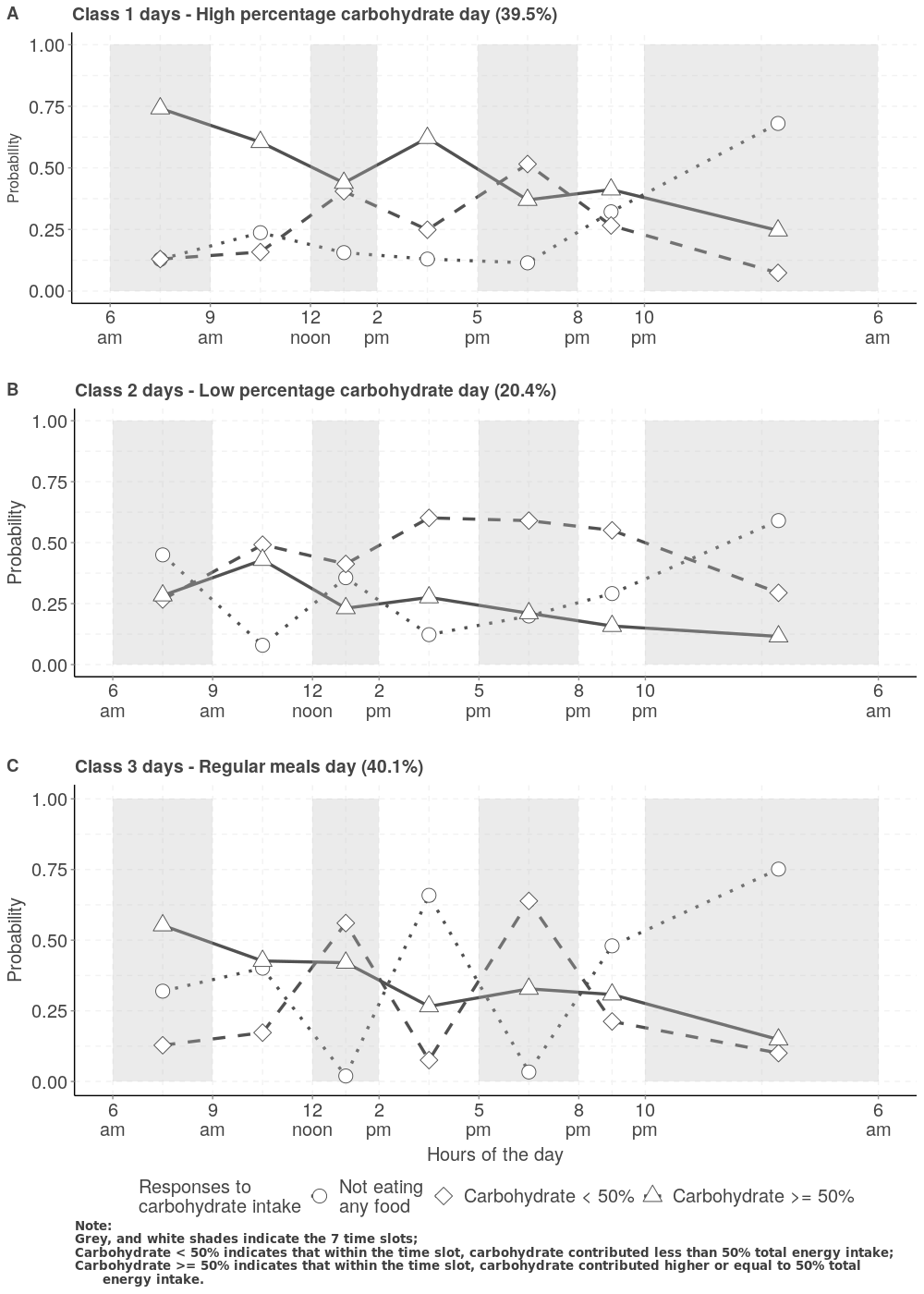
A series of traditional LCA of the responses to carbohydrate intake within 7 time slots of day was first examined. These initial analyses ignored the clustering of observation days within participants of the survey. **Table [tab:mixmodels]** shows the latent class solutions for one to five classes (see rows under the Fixed effects model section). The BIC declines with the number of day level classes increases. However, the improvement of BIC dropped to less than 1000 from 3 classes to 4 classes solutions (658.9) and from 4 classes to 5 classses solutions (361.7). Entropy index indicates that the 4 classes model could explain about 51% percent of the data, while *p* values of Lo-Mendell-Rubun LRT suggest that the more classes we fit, the better model we will have until up to 6 classes (*p* = 0.06 and is not shown in the table). From the parsimony point of view, we extended the model with random effects building on 2 classes, 3 classes and 4 classes solutions.

The results of the random effect included models are presented in **Table [tab:mixmodels]** under the Random effects model section. It is obvious that the BIC improves with the addition of the random effects which account for the nested structure of the data. Entropy indicates that 4 classes in individual level and 2 classes in the day level may be the best solution mathematically. However, after these solutions were checked in more details, the potentially most substantively interpretable model was found to be the 33 random effect model, which is the model with 3 latent classes in the day level, and 3 latent classes in the individual level. We must emphasize that different researchers may have made decision slightly different from ours, we provided the descriptions and figures for other solutions in the **Appendix [AppendixC]** for reference.

In the 33 random effect model solution we have chosen, there were 39.5%, 20.4%, and 40.1% observations classified into 3 latent groups in the day level. The overall counts and percentages for each responses within every time slot and the distributions of the solution are presented in **Table [tab:daylevel]**. The trajectories illustrating the change of the probabilities of each response to carbohydrate eating during the hours of the day are shown separately by three types of days in **Figure [fig:level1]**.

Class 1 days **(Figure [fig:level1]-A)** were given the name of “high percentage carbohydrate day” since in these days of survey, the probabilities of carbohydrate contributed higher or equal to 50% of the energy consumed were always higher than that in the other two types of days. Specifically, high percentage carbohydrate days were characterised with probabilities of over 0.6 in time slots between 6 am to 9 am, 9 am to 12 noon, and also 2 pm to 5 pm, during which the time slots may be interpreted as breakfast, morning snack, and afternoon snack time periods for many participants. Moreover, even during late night time period, such as 8 pm to 10 pm, and 10 pm to 6 am time slots, the probabilities of having higher carbohydrate contained food were still as high as 0.412, and 0.246, respectively.

Class 2 days **(Figure [fig:level1]-B)** were named as “low percentage carbohydrate day” because first of all, in these days the probability of participants skipping breakfast was 0.45. And after 9 am, within these days, the probability of having low carbohydrate contained food (carbohydrate contributed 50% of total energy intake), was always higher than having high carbohydrate contained food (carbohydrate contributed 50% of total energy intake). In class 2 days, participants also turned to have morning snacks (with only 0.079 possibility of **not** eating any food and similar probabilities of having either high or low carbohydrate contained food). This phenomenon may also be interpreted as having a long and late breakfast (brunch) in these mornings. The probability of **not** eating any food was the lowest for low carbohydrate days during the midnight time slot (10 pm to 6 am), with probability of 0.590 compared with 0.680 and 0.751 in class 1 and class 3 days, respectively.



Day Level Latent Classes Solution.

Class 3 days **(Figure [fig:level1]-C)** were called “regular meals day” due to the following reasons: 1) participants’ dietary recordings showed that in these days there was almost 0 possibility of not eating any food at lunch (0.019 between 12 noon and 2 pm) and dinner (0.034 between 5 pm and 8 pm); 2) the probabilities of not eating during morning snack time (9 am to 12 noon) and afternoon snack time (2 pm to 5 pm) were also the highest among the three types of days (0.401 and 0.659). 3) during these days, participants may have some high carbohydrate contained food between 8 pm and 10 pm (probability = 0.308), but the probability of not eating any food during 10 pm to 6 am next morning was 0.751, the highest among the three types of days.

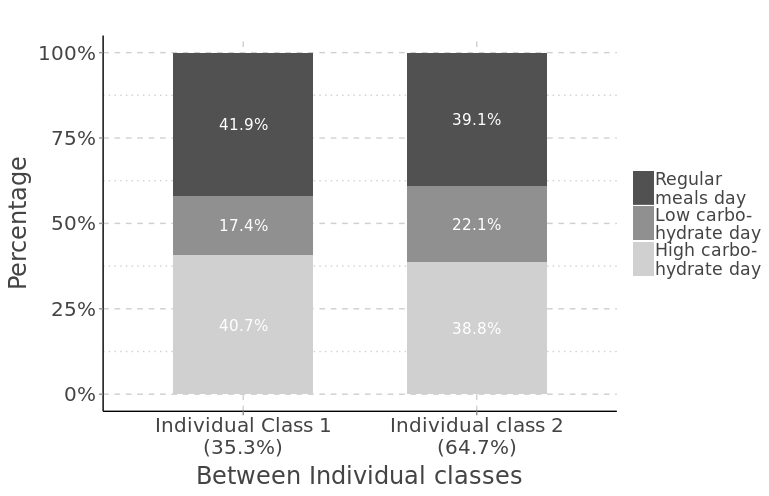
### Features of the three types of carbohydrate eating temporal patterns

The details of the characteristics of the three types of carbohydrate eating time pattern were listed in **Table [tab:day-level-features]**. Specifically, regular meals day turned to be recorded slightly more often in Northern Ireland, and Scotland. In terms of day of week distribution in the three types of days, there is strong evidence (*p* < 0.001) that high carbohydrate days appeared more frequently in weekends (32.5%) compared with low carbohydrate day (26.5%) and regular meals day (30.7%).

As expected, consumption of total energy (7539.98 kJ), total carbohydrate (222.79 g), total sugar (98.63 g), starch (124.07 g), and non-milk extrinsic sugar (59.45 g) were the highest among high percentage carbohydrate days (all *p* < 0.001). On the other hand, the consumption of protein (73.29 g), total fat (67.24 g), and alcohol (13.80 g) were the highest in the regular meals days. Moreover, in high percentage carbohydrate days, participants turned to consume the highest amount of fruit (107.40 g). There was no evidence of any difference for the consumption of yellow, red, or green vegetables across the three types of days (*p* = 0.681).

### Individual level LCA solution

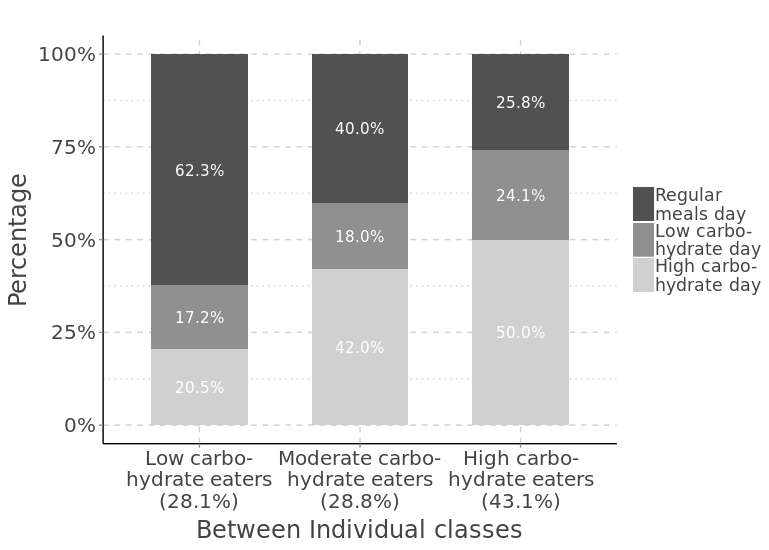
In the random effect models we utilized the non-parametric approach, in which we added a level 2 (individual level) latent classes based on the random means from the level 1 (day level) latent class solution. The results of the individual level LCA solution for 2 and 3 classes are presented in **Figure [fig:CB2level2]**, and **[fig:level2]**.



Multilevel Latent Class Solution, 3 classes in day level, 2 classes in individual level.

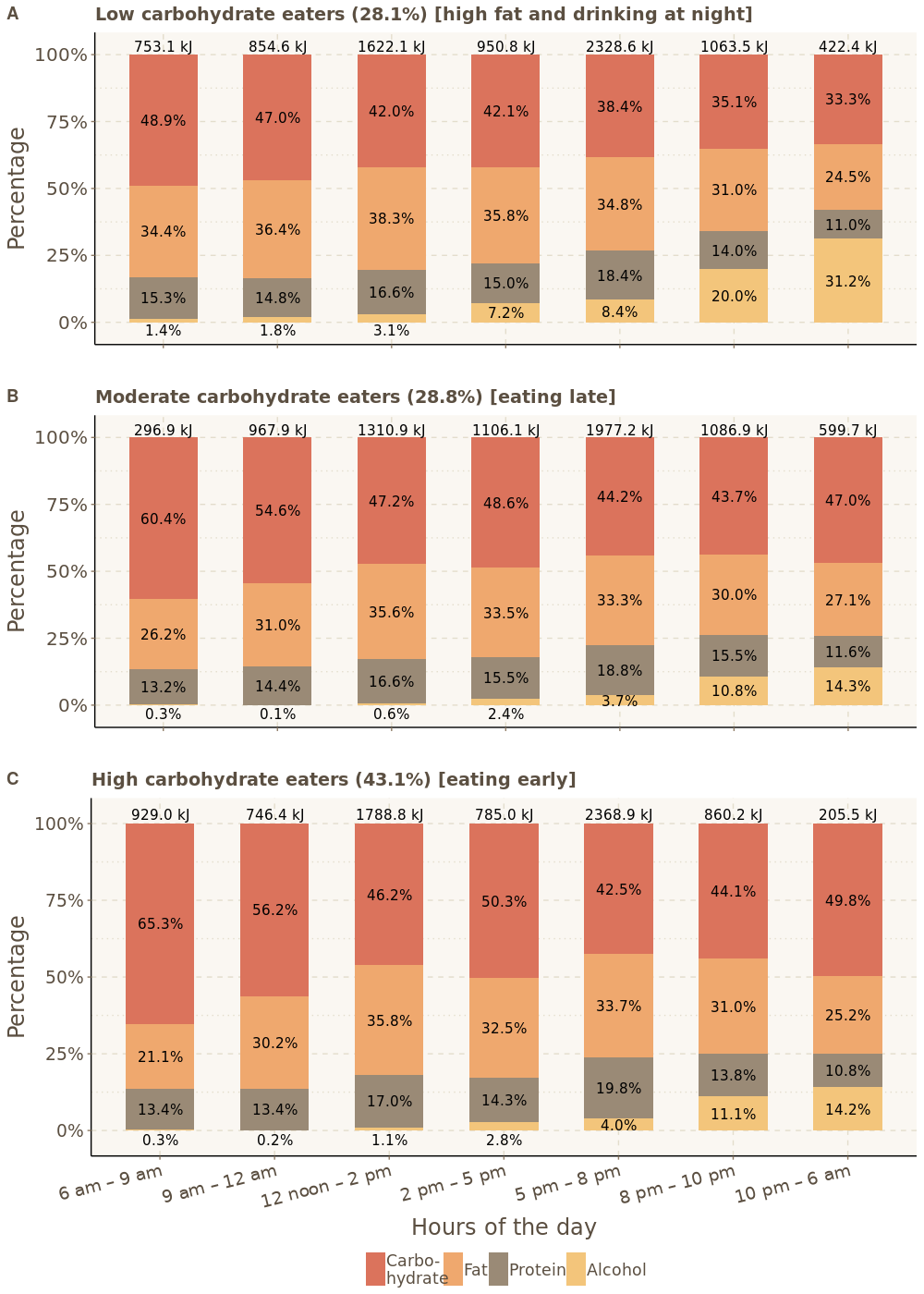
With two individual level latent classes **(Figure [fig:CB2level2])**, one individual class is comprised of individuals with a relatively slightly higher proportion of having “low carbohydrate day” (22.1%) compared to the other (17.4%). This class represents nearly 65% of the individuals. However, we believe these individual classes are not very distinguishable to each other.

With three individual level latent classes **(Figure [fig:level2])**, a low-carbohydrate eaters class, a moderate-carbohydrate eaters class, and a high-carbohydrate eaters class emerges. 43.1% participants were identified as high-carbohydrate eaters, in these individuals, about 50% of the days (2 out of 4 days) of their dietary diary could be classified as having high carbohydrate days. Nearly 1 out of 4 days of their dietary diary were either “regular meals day” or “low carbohydrate day”. 28.1% participants fell into the low carbohydrate eaters class in the left hand side of **Figure [fig:level2]**, their recordings of food intake showed that in more than 60% of their days, they were having “regular meals” which was characterised as with highest amount of fat and alcohol consumptions as already described in **Table [tab:day-level-features]**. Moderate carbohydrate eaters have comparable proportions (42.0% vs. 40.0%) of having high carbohydrate days and regular meals day, 18.0% of their dietary diary were found to be low carbohydrate days.



Multilevel Latent Class Solution, 3 classes in day level, 3 classes in individual level.

After recognising that there were three potential latent groups of carbohydrate eaters in the UK adults, whose food consumption pattern were also probably switching from one to another during the survey, their average carbohydrate contribution to total energy intake (as well as the subtypes of carbohydrate actually consumed) within the 7 pre-defined time slots of the day were still of interest. Survey-design-weighted mean energy intake within each time slot of the day and their composition of contribution are illustrated in **Figure [fig:energysourcesCB]**, weighted mean nutrients intakes are listed in **Table [tab:tab-nutri-indi]**.



Sources of Energy Contribution at Each Time Slot by Individual Carbohydrate Eating Groups.

Among the three types of carbohydrate eaters, the mean of total energy intake over the 4 days of dietary survey was the highest (7985.8 kJ, 95%CI: 7283.3, 8146.3) in the low carbohydrate eaters group, and the lowest (7341.8 kJ, 95%CI: 7172.5, 7511.2) in the moderate eaters group **(Table [tab:tab-nutri-indi])**. Sources of energy for each type of carbohydrate eaters by the 7 time slots were also different. Low carbohydrate eaters **(Figure [fig:energysourcesCB]-A)** never had carbohydrate contributed more than 50% of their total energy throughout the day. Energy from fat were the highest for low carbohydrate eaters most of the time during the day (except for time between 10 pm to 6 am next morning). Most impressively, energy from alcohol were always the highest in low carbohydrate eaters, percentages for energy from alcohol for the 7 times slots were 1.4% (6-9 am), 1.8% (9-12 noon), 3.1% (12-2 pm), 7.2% (2-5 pm), 8.4% (5-8 pm), 20.0% (8-10 pm), and 31.2% (10 - 6 am), respectively. Contribution from different energy sources are quite similar for moderate and high carbohydrate eaters, but their absolute amount of energy consumption at each time slot were largely different. Moderate carbohydrate eaters **(Figure [fig:energysourcesCB]-B)** were characterised as consuming the lowest energy (296.9 kJ) before 9 am, but having higher energy consumption (967.9 kJ) between 9 am and 12 noon time compared with low and high carbohydrate eaters. Moderate carbohydrate eaters may turn to have later breakfast, later lunch, and probably later dinner as well. They had the highest total energy consumption (599.7 kJ) at night (10 pm - 6 am) across three types of eaters. High carbohydrate eaters **(Figure [fig:energysourcesCB]-C)** consumed the highest total energy (929.0 kJ) during 6 am to 9 am in the morning and the lowest total energy between 10 pm to 6 am (205.5 kJ). Specifically, carbohydrate contribution to total energy intake were 65.3% (6-9 am), 56.2% (9-12 noon), 46.2% (12-2 pm), 50.3% (2-5 pm), 42.5% (5-8 pm), 44.1% (8-10 pm), and 49.9% (10-6am). We also noticed that high carbohydrate eaters consumed their energy mainly from three time slots: 6-9 am, 12-2 pm, and 5-8 pm.

As expected, in total, the mean of carbohydrate intake was 203.8 g, 218.3 g, and 233.4 g for low, moderate, and high carbohydrate eaters, respectively **(Table [tab:tab-nutri-indi])**. Energy contribution from carbohydrate was close to 50% in the high carbohydrate eaters, but was only 40.6% in the low carbohydrate eaters. In terms of the subtypes (components) of the carbohydrate consumed at each time slot, high carbohydrate eaters consumed as much as more than 2 times (compared to low carbohydrate eaters) and nearly 4 times (against moderate carbohydrate eaterss) the amount of sugar (37.9g 95% CI: 36.8, 39.2) and non-milk extrinsic sugar (i.e. free sugar) (11.1g 95%CI: 10.7, 11.6) between 6-9 am. Moderate carbohydrate eaters had their carbohydrate intake more spread out. They consumed more sugar and starch during 9-12 noon, 2-5 pm, 8-10 pm, and 10-6 am. Low carbohydrate eaters turned to have similar temporal pattern of consuming carbohydrates but the absolute amount of fibre, sugar, free sugar, and starch were usually lower than that in the high carbohydrate eaters except for time slots of 2-5 pm, and 10-6 am. Strong evidence (*p* < 0.001) suggested that the mean of total fibre consumption for low, moderate, and high carbohydrate eaters were different: 13.7g (13.4, 14.0), 12.5g (12.1, 12.9), and 14.7g (14.4, 14.9) with all 95% CI being exclusive to each other. It is also noteworthy that low carbohydrate eaters consumed the highest average amount of protein (79.9 g, 17.2% of total energy), fat (74.7g, 35.4% of total energy), and alcohol (20.8 g, 6.8% of total energy) as we have described for **Figure [fig:energysourcesCB]**.

The social-demographic characteristics of the UK adults according to their individual level latent class membership are shown in **Table [tab:Level2tab1]**. Moderate carbohydrate eaters were relatively younger (*p* < 0.001), and slightly less from England (*p* = 0.007). Gender distribution across the three types of carbohydrate eaters was fairly even (*p* = 0.119). The distribution of the carbohydrate eater types turned out to be changing with the year of survey. Low carbohydrate eaters represented 32.5% of the population in the first year of survey, but later dropped to lower than 30% (lowest in the third year, 22.6%) until the most recent year. Proportion of high carbohydrate eaters increased from 41.2% to the highest (50.6%) in the second year, but then started to decline to 38.4% in the 8th year of survey (*p* = 0.015). There was no evidence of difference in employment status across three types of carbohydrate eaters. However, strong evidence suggested that high carbohydrate eaters had the highest proportion (61.3%) of living with partner (*p* < 0.001); moderate carbohydrate eaters had the lowest average income (27180.8 /year), highest proportion of non-white population (20.5%), and lower education level (23.3% with degree of higher education) compared with either low or high carbohydrate eaters.

Weighted means, percentages of anthropometric measurements, average of main nutrients intake, as well as biochemical characteristic profiles according to the latent carbohydrate eater groups are given in **Table [tab:tab2]**. Low carbohydrate eaters had higher mean BMI (27.8 kg/m2) and larger mean WC (98.9/89.9 cm in men/women) compared with 27.2, 27.3 kg/m2, and 95.9/88.7 (men/women), 98.1/87.2 (men/women) cm in moderate and high carbohydrate eaters. Moderate carbohydrate eaters had the highest prevalence of being a current smoker (27.8%), shortest time of daily physical activity (geometric mean: 0.87 hours/day), and the lowest prevalence of hypertension (20.2%).

From the results of blood tests, 6.9% of low carbohydrate eaters were found to be diabetic (diagnosed by A1C > 6.5%), while the percentages of diabetes in the moderate and high carbohydrate eaters were 3.5%, and 4.1% (*p* < 0.011), respectively. Although there was some evidence (*p* = 0.027) that fasting blood glucose level may be slightly higher in non-diabetic low carbohydrate eaters, the geometric mean for A1C was probably lower in moderate carbohydrate eaters (4.72, 95%CI: 5.39, 5.47). Total cholesterol, HDL, and LDL were all lower in the moderate carbohydrate eaters, while no evidence of any difference of TG was found across three types of carbohydrate eaters.

### Association between individual level latent classes and hypertension, and obesity.

#### Hypertension

**Table [tab:tab1hypetension]** presents the characteristics of men and women participants in the NDNS RP 2008/09-15/16 by hypertension status. The weighted prevalences of hypertension were 30.4% in men and 27.5% in women. Among both sexes, there were strong evidence of differences by hypertension status for age, education level, living with a partner or not, smoking status, BMI, abdominal obesity (WC), and prevalence of diabetes (*p* < 0.01). No difference was found among either men or women for ethnicity. Strong evidence of difference was suggested in women for average household income (32741.5 /year in non-hypertensive vs. 27862.0 /year in hypertensive, *p* < 0.001), and physical activity level (geometric mean: 0.81 hours/day in non-hypertensive compared with 0.53 hours/day in hypertensive, *p* < 0.001) but not in men. Interestingly, in both sexes, hypertensive participants had higher proportion of being classified as low or high carbohydrate eaters; total energy and carbohydrate intake was higher in people with hypertension (*p* < 0.001).

The sex-specific associations of carbohydrate eating patterns with hypertension (both in total and in participants without diabetes) are shown in **Table [tab:tab2HYT]**. In the crude models, moderate carbohydrate eaters had statistically significant lower odds of having hypertension than low carbohydrate eaters in both men and women irrespective to diabetes status. Among men, after adjustment for selected confounders, which includes: age, live with partner or not, education level, BMI, smoking status, and total energy intake, the odds ratio (OR) comparing moderate with low carbohydrate eaters was 0.68 (95% CI: 0.43, 1.07) and remained borderline significant (*p* = 0.093). 95% CI of the adjusted OR became narrower (OR: 0.64, 95% CI: 0.41, 1.01, *p* = 0.054) when BMI was replaced with WC in model 2. When diabetic men were excluded in the models, the ORs (95%CI) for moderate and high carbohydrate eaters compared with low carbohydrate eaters were 0.65 (0.41, 1.03) and 0.73 (0.51, 1.06), respectively. The negative associations between moderate carbohydrate eating pattern and hypertension were also observed in women, however, without any statistically significant evidence in the fully adjusted models. High carbohydrate eaters also had lower adjusted odds compared with low carbohydrate eaters, while the 95% CIs for the adjusted ORs were all wide and included the null value suggesting no evidence of any association in either men or women for high carbohydrate eating pattern and hypertension.

#### Obesity (BMI and WC)

**Table [tab:tab1BMI]** shows the characteristics for participants according to their obesity status stratified by sex. The survey design-weighted prevalence for being overweight and obese in the UK adults were estimated to be 43.4% and 25.7% in men, and 30.9% and 27.4% in women. Obviously, abdominal obesity (WC) increased significantly with the elevated BMI level in both men and women. Overweight or obese participants were older, having lower total energy intake and lower carbohydrate intake compared with normal weight men and women (*p* < 0.001). Moreover, education level (*p* = 0.022 for men, < 0.001 for women), average household income (*p* = 0.011 for men, < 0.001 for women) were decreasing with increasing BMI. Living with partner or not was strongly positively associated with obesity in men but not in women. Men with obesity were also found to have the lowest proportion of never being a smoker (47.9 %), and the highest proportion of being ex-smoker (32.9%). Association between smoking status and obesity in women was only with very weak evidence (*p* = 0.042) but similar pattern as in men was also observed (higher proportion of ex- and current smokers in overweight or obese women). No difference was found for length of physical activity across obesity levels in men, while in women, somewhat weak inverse association (*p*=0.038) was confirmed. Interestingly, predefined carbohydrate eating patterns were seemingly not associated with BMI in men (*p* = 0.072) or in women (*p* = 0.253).

Results of the multivariable linear regression analyses showed inverse associations between latent classes of carbohydrate eating patterns and BMI among men **(Table [tab:tab2BMI])**. However, the 95%CI of the regression coefficients were all wide and included the null value 0, indicating no statistically supported evidence for the inverse association. However, evidence of interaction effect was found in whether live with a partner or not on the association between carbohydrate eating patterns and BMI (*p* for interaction = 0.014 and 0.036 for women in total and without diabetes). For women who were living with her partner, latent classes of carbohydrate eating patterns were negatively associated with BMI. Compared with women eating low carbohydrate food pattern, women having a high carbohydrate eating pattern were averagely associated with 1.76 kg/m2 lower BMI after adjustment of age, average household income, education level, smoking status, total energy intake and alcohol consumption. 95%CI for the adjusted BMI difference was 0.73 to 2.78 kg/m2, *p* = 0.001. After excluding diabetic women, BMI was still 1.62 kg/m2 (95%CI: 0.58, 2.65, *p* = 0.002) smaller in high carbohydrate eaters versus low carbohydrate eaters on average. On the contrary, latent classes of carbohydrate eating patterns were positively associated with BMI in women who were living by themselves, although the regression coefficients were not statistically different from 0.

Similarly, when looking at the association between carbohydrate eating pattern and abdominal obesity (WC), men who were classified as moderate carbohydrate eaters were found to had about 3 cm (95%CI: 0.52, 5.49 cm, *p* = 0.018) smaller WC compared with low carbohydrate eaters in the crude model **(Table [tab:tab2WC])**. However, after adjustment of age, living with partner or not, average household income, education level, hypertension, smoking status, total energy intake, and alcohol consumption, the association attenuated to no difference. The interaction effect of whether living with partner or not on the association between carbohydrate eating patterns and WC was again found in women in total (*p* for interaction = 0.009) and without diabetes (*p* for interaction = 0.012). Among women who were living with their partners, high carbohydrate eaters had 4.71 cm (95%CI: 2.43, 7.00, *p* < 0.001) smaller WC on average compared with low carbohydrate eaters. The association remained when restricted the sample to non-diabetic women (-3.74 cm, 95% CI: -5.97, -1.51, *p* = 0.001). However, for women who were living alone, moderate carbohydrate eaters had 3.17 cm (95%CI: 0.05, 6.30, *p* = 0.047) larger WC on average compared with low carbohydrate eaters. The evidence for the positive association between moderate carbohydrate eaters and WC in women became weaker but with borderline significant when excluding diabetic women (3.08 cm, 95%CI: -0.09, 6.25, *p* = 0.057).

# Discussion and Conclusion

## Carbohydrate eating patterns

Using multilevel LCA as a novel technique, and the NDNS RP dietary database, this project examined carbohydrate eating temporal patterns firstly in the day level, based on which, individual level carbohydrate eating patterns were also defined subsequently.

Among the dietary diaries collected, there were three distinct latent classes specifically for carbohydrate intake: 1) high probabilities of having high carbohydrate contained food across the hours of day (high percentage carbohydrate day); 2) low carbohydrate food dominant through out the hours of day (low percentage carbohydrate day); and 3) always having lunch and dinner day (regular meals day). And from these day level classifications and their features, one might anticipate that individuals who managed to follow the regular meals day might be eating a relatively healthier diet because it seemed to be a regular temporal eating habit; at this time point, we also believed that those who followed either high or low carbohydrate percentage days would probably consume higher total energy than those who followed mostly regular meals days.

However, when the MLCA extended the model to individual level, three types of persons were further defined depending on their 4-day-diary: 1) low carbohydrate eaters, who mostly followed “regular meals day”; 2) moderate carbohydrate eaters, who had similar probabilities of following either “regular meals day” or “high percentage carbohydrate day”; 3) high carbohydrate eaters, who followed “high percentage carbohydrate day” for half of their survey. For the first time, as far as we know, the day-to-day food intake pattern variation within individuals was successfully captured by MLCA models. Results from the MLCA models showed that from the perspective of carbohydrate consumption, people were indeed changing their diet from day to day even within a short term period of survey. The MLCA models allowed the probability of following a certain type of carbohydrate eating day to vary across individuals. This properly accounted for the fact that for some people, their probability of following a type of food eating pattern during the survey could be higher/lower than that in the others. This finding also suggested that assuming a person will always follow a certain type of food intake pattern is not appropriate.

Surprisingly, low carbohydrate eaters whose dietary recordings suggested that they were mostly following a regular temporal meals pattern turned out to be consuming the highest amount of total energy among the three types of carbohydrate eaters. Detailed profiling of energy composition according to the time slots revealed that low carbohydrate eaters actually had higher proportion of energy contributed by both alcohol and fat. High percentage of fat consumption was shown in all 7 time slots, energy coming from alcohol exceeded more than one fifth of the total energy after 8 pm. These findings explained why they were actually consuming the highest energy among the three types of carbohydrate eaters. However, we also found that participants consuming low carbohydrate food had higher prevalence of diabetes, hypertension, and obesity. These health issues might possibly lead them (or advices were given from their physicians) to replace carbohydrates in their diet to other energy sources such as fat, protein, or even alcohol. Therefore, there is a possibility that they chose to follow low carbohydrate diets out of health purpose, but many of them were replacing carbohydrates with higher energy condensed food or even alcohol at night which might indeed be a public health concern.

Next, when looking into the details of the timing and composition of the energy intake among the moderate carbohydrate eaters, we realised that although these individuals did not consume as much alcohol as low carbohydrate eaters did at night, they consumed the highest amount of energy, especially during time period as late as after 10 pm. People fell into moderate carbohydrate eaters group seemed to have the tendency of having their food or meals later than high carbohydrate eaters. They consumed the highest amount of carbohydrates and also total energy among three types of carbohydrate eaters during the following time slots: 9-12 noon, 2-5 pm, 8-10 pm, and 10pm-6am. These individuals in the NDNS RP were younger, largely single, with lower average income, and lower education level. They might possibly correspond to the “late eaters” defined by previous studies .

Finally, the high carbohydrate eaters identified by our MLCA models had the highest absolute total amount of carbohydrate intake. Most (nearly 70%) of their total energy intake occurred during 6-9 am, 12-2 pm, and 5-8 pm time slots although their total energy consumption was not the highest (between low and moderate carbohydrate eaters). High carbohydrate eaters were also found to be the ones that consumed the least amount of energy after 8 pm. Therefore, contrary to what was anticipated at the beginning, people who followed high percentage carbohydrate days for most of their time, were actually eating a healthier diet compared with the other two eating patterns.

Our analyses looking for different temporal carbohydrate also highlighted the complexity of eating pattern behaviours in the population and the utility of exploratory, data-driven methods to objectively identify eating patterns that reflect both timing and quantities of food intake, which may have not been detected so far in the literature.

## Association between carbohydrate eating patterns and health outcomes

Among men, who were classified as moderate carbohydrate eaters probably had lower odds of having hypertension, after adjustment of age, live with partner or not, educational level, BMI, smoking status and total energy intake. As discussed above, moderate carbohydrate eaters turned to have meals (or energy consumption) later in time compared with high carbohydrate eaters, but low carbohydrate eaters also consumed large amount of energy (from both fat and alcohol) at night. Therefore, considering that moderate carbohydrate eaters were younger than low carbohydrate eaters (although age was adjusted in the full models), there is probably reverse causality exists here (also due to the nature of cross-sectional study). That is, they were potentially both late eaters, however, with increased age (and so as increased health-related problems/concern) some of them modified their habits, such as quit smoking, replace carbohydrate food with other energy sources which lead to the phenomenon of lower odds of hypertension in moderate carbohydrate eaters. Although these hypotheses cannot be determined by the cross-sectional data from NDNS RP, if they were true, the energy sources they used to replace carbohydrate in their diet were apparently not very wisely chosen.

Among women, whether living with a partner became an interaction factor for the associations between carbohydrate eating patterns and BMI and abdominal obesity (WC). Directions of the associations were opposite to each other depending on whether women were living with a partner or not. This interaction effect was more obvious when looking at abdominal obesity measurement. High carbohydrate eaters had lower BMI and WC in those lived with their partner, while moderate carbohydrate eaters had higher WC in those who lived alone after adjustment of age, education level, smoking status, total energy intake, and alcohol consumption. High carbohydrate eaters who were characterised with high and early in the day energy consumption and low fat and alcohol intake may reflect a healthier diet and lifestyle, but this might be different between women who lived alone and those who lived with their partners. It was often assumed that live alone may associated with lower diversity of food intake, and a higher likelihood of having an unhealthy dietary pattern . Therefore, there may be differences in the actual contents consumed in high carbohydrates eating women, or there may be other social, psychological or lifestyle related factors related with living alone which were not measured or did not included in the models, so that the inverse association between high carbohydrate eating pattern and BMI or abdominal obesity were only observed among women who lived with partners, further analysis is needed. Whereas the reason why moderate carbohydrate eaters’ WC was larger than low carbohydrate eaters only among women who lived alone is unknown, given that the evidence of this association was weak and borderline significant, whether it was just a false positive finding should be explored in other studies.

## Limitations and strengths

There are several limitations in the current project that merits consideration.

Firstly, we ignored the order of observation days in the MLCA models. The food consumption diaries were accomplished by participants for at least 3 out of 4 consecutive days. In the multilevel analyses, these 3 or 4 days’ observations were treated as if they were exchangeable in the models. Since one’s diet might change according to season, day of the week (weekdays or weekends), or sometimes depends on what they had consumed the day before, exchangeability of their diaries is a strong assumption and cannot be overcome by the MLCA models adopted here. Other statistical techniques, which could take the order or the longitudinal nature of the data into account, such as repeated measures latent class analysis (RMLCA), latent transition analysis (LTA) , or latent class growth analysis (LCGA) are not applicable for the NDNS RP dataset. For the purpose of maintaining the response rate, flexibility was allowed for the NDNS RP participants in their choice of which day to begin their diary. But the above mentioned alternative models require that the 3 or 4 repeated measurements of food diaries to be recorded at the same time points longitudinally. Specifically, LCGA models (also called growth mixture modelling) will also need to model the change of the odds of the probability over time as a function (quadratic, or cubic) which is apparently not the objective in the current project.

Secondly, the classification of individuals to latent carbohydrate eating classes were defined by maximum posterior probability assignment rule. This approach assigns individuals to the class for which they have the highest posterior probability of membership . The other approach is multiple pseudo-class draws , which was proposed in attempt to account for uncertainty in class assignment. However, the maximum probability rule is still believed to be able to minimize the number of incorrect assignments . A Monte Carlo simulation study also demonstrated that maximum probability assignment is less biased than multiple pseudo-class draws. Moreover, this simulation study also found that an inclusive LCA (i.e. LCA with covariates) would perform better than non-inclusive LCA, and has the potential to reduce bias of class assignment. However, whether this advantage can be extended to MLCA in the current project is unknown, and also due to the reason that the associations between the latent classes and distal outcomes (hypertension, and BMI) in our analyses are mostly still under exploration, therefore, the model for LCA with covariates can never be well-known before our analyses. Thus, as a first step, given the complexity of the NDNS RP dataset itself, we chose to conduct the MLCA modellings without any covariate in either day level or individual level models. Future studies may be advised to consider incorporating other predictors in the MLCA models to see whether the classifications in both levels can be improved or not.

Thirdly, information regarding the detailed occupations of the participants was not available and because of which we could not exclude those who had a shift work involving night work. But considering that completing a 4-consecutive-day food diary would be a large burden for the participants, those had a shift work involving night work might possibly performed relatively lower compliance and could not complete their diary for at least 3 days. But this hypothesis cannot be verified in the NDNS RP dataset. If people working with shift work were included in the NDNS RP sample, the classification in both levels as well as their associations with either hypertension or BMI would be biased. Furthermore, under-reporting due to the burden of writing down everything they eat at each occasion is also a concern in the food diaries collected. To evaluate the influence of under-reporting in the NDNS RP sample, study team of NDNS RP conducted doubly labelled water (DLW) sub-study within the survey. The details of this sub-study can be accessed from the Appendix X of the Official Reports provided by Public Health England . Briefly, in healthy adults, if, for a given period of time, energy consumed in food matches total energy expended, they are in energy balance. In NDNS RP sub-study, estimates of energy intake (EI) from the four-day diary were compared with measurements of total energy expenditure (TEE) using the DLW technique in a sub-sample of survey participants. Results of the sub-study showed reasonable agreement between EI and TEE (overall ratio: 0.73). Reasons such as misreporting of actual consumption; under-reported or modified usual intake due to the burden of the survey/DLW sub-study are potential factors may contributed to under-reporting of their EI. We cannot extrapolate the estimated under-reporting to the whole sample since other individuals’ diet might be affected by under-reporting differentially.

Lastly, it should be noted that findings from data-driven, exploratory methods may not be generalizable to populations in other countries. Since the carbohydrate eating patterns in both day level and individual level probably only represent the socio-cultural and life-style characteristics of the adult population in the UK. Further researches are warranted to explore and better understand the eating patterns in other populations. Besides, the cross-sectional study design also refrained us from deducing any causal effect between the carbohydrate eating patterns and hypertension, BMI or abdominal obesity.

Strengths of this study include the large, nationally representative sample of men and women in the UK. We applied a novel, objective approach using MLCA to examine the carbohydrate eating patterns while applying standardised criteria to determine the numbers of latent classes. The process of finding the classifications through model-based, data-driven procedure minimises reliance on researchers’ preconcieved notions of eating patterns. Eating patterns were determined from 4 consecutive days of well-designed, fully-examined food diaries. MLCA correctly accounted for the multilevel structure of the data which means that the 4-day diet diaries were nested within the participants. Our findings also captured the day-to-day variation of following different carbohydrate eating patterns within individuals. Moreover, the examination of associations between the individual level carbohydrate eating patterns and hypertension, BMI, abdominal obesity were conducted correctly considering the probability of participant selection and the clustered survey design. Health outcomes such as weight, height, waist circumferences, and blood pressure, diabetes status were measured by trained nurses and blood tests which could also minimise bias caused by misclassification and under-reporting of these measurements.

## Conclusions

We have successfully defined carbohydrate eating patterns in the general population in the UK adults using the NDNS RP database in both observation day level and individual level. Low carbohydrate eaters turned to have more energy that contributed by both fat and alcohol. Moderate carbohydrate eaters consumed the lowest total energy, while they had the tendency of having meals later in time-of-day. High carbohydrate eaters consumed most of their carbohydrate as well as energy earlier in time-of-day. These dietary patterns specifically for carbohydrate intake were found to be differed by timing, quantity and contributions to energy consumption. Compared with low carbohydrate eaters, men had moderate carbohydrate eating pattern may associated with a lower prevalence of hypertension, women in this latent class who lived alone may associated with a larger waist circumferences. Among women who lived with partners, high carbohydrate eating pattern was associated with both lower BMI and smaller waist circumferences. Longitudinal designed studies are needed in finding whether the eating patterns themselves are changing over time and in investigating how such circadian eating patterns may relate to the change of blood pressure, obesity, and even mortality.