LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE

MSc PROJECT REPORT
2017-2018

The timing of carbohydrate intake in UK adults, using the National Dietary and Nutrition Survey (NDNS) 2008-2014 programme

Supervisor:
Professor Luigi PALLA

Submitted in part fulfillment of the requirements for the degree of MSc in Medical Statistics

Candidate number: 110765
Page count: XX from Introduction to Conclusions

September 2018

Declaration of Authorship

I, Chaochen WANG, declare that this thesis titled, "The timing of carbohydrate intake in UK adults, using the National Dietary and Nutrition Survey (NDNS) 2008-2014 programme" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a MSc degree on Medical Statistics at this University.
- No part of this thesis has previously been submitted for a degree or any other qualification at this University 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:

Date:

Acknowledgements

I would like to thank my tutor and supervisor Professor Luigi PALLA, 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 Dietary and Nutrition Survey (NDNS) 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.

I also 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

Abstract

The National Dietary and Nutrition Survey (NDNS) database of detailed four-day food diaries was used to …

Contents

Declaration of Authorship **iii**

Acknowledgements **v**

Abstract **ix**

1 Introduction **1**
Background 1
The National Dietary and Nutrition Survey (NDNS) 2
Aims and objectives 2

2 Methods **3**
Dietary diary collected in the NDNS RP 3
Definition of carbohydrate intake 3
Survey Data 4
Survey Selection Method 4
Response rates 5
Strata and weightings 5
Latent Class Analysis (LCA) 6
Multilevel Latent Class Analysis (MLCA) 7
Parametric approach 7
Non-Parametric approach 9
Strategy of conducting MLCA in the current analysis 10
Association between day-level and individual-level latent classes and other variables 11

3 Results **12**
Model selection, comparison, and interpretation 12
Subsection 2 15
Main Section 2 15

4 Discussion and Conclusion **16**
Main Section 1 16

Subsection 1 16
Subsection 2 16
Main Section 2 16

Bibliography **17**

A R code for importing and manipulating the data **21**

B Mplus code for Multilevel LCA models **26**

C Example of a food diary for one day **27**

List of Figures

3.1 Day Level Latent Classes. 14
3.2 Individual Level Latent Classes. 15

C.1 One day food diary example 6 am to 2 pm 27
C.2 One day food diary example 2 pm to 6 am 28
C.3 Food diary example of home made food recipes 28

List of Tables

3.1 Fit Criteria for Each Model Specification 12
3.2 Day Level Latent Class Solution for Three-Class Model (No Individual level Model) 13

List of Abbreviations

AIC	**A**kaike **I**information **C**riterion
aBIC	**a**djusted **B**ayesian **I**nformation **C**riterion
BMI	**B**ody **M**ass **I**ndex
BIC	**B**ayesian **I**information **C**riterion
DM	**D**iabetes **M**ellitus
EM	**E**xpectation **M**aximazation
FSA	**F**ood **S**tandards **A**gency
HbA1c	**H**emoglobin **A1c**: Glycated hemoglobin
LCA	**L**atent **C**lass **A**nalysis
LCGA	**L**atent **C**lass **G**rowth **A**nalysis
MAFF	**M**inistry of **A**griculture, **F**isheries and **F**ood
MAR	**M**issing **A**t **R**andom
MCAR	**M**issing **C**ompletely **A**t **R**andom
MLCA	**M**ultilevel **L**atent **C**lass **A**nalysis
MNAR	**M**issing **N**ot **A**t **R**andom
ML	**M**aximum **L**ikelihood
NDNS	the **N**ational **D**ietary and **N**utrition **S**urvey
OR	**O**dds **R**atio
PHE	**P**ublic **H**ealth **E**ngland
PSUs	**P**rimary **S**ampling **U**nits

Chapter 1

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 evidence. More importantly, the actual temporal patterns of eating in the population, proportions of people who actually manage/fail to follow this so-called doctrine have not been described thoroughly.

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 (Johnston, 2014; Asher and Sassone-Corsi, 2015).

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. 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 (Uemura et al., 2015). Shift workers have a higher risk of developing metabolic syndrome (De Bacquer et al., 2009) and type 2 diabetes (Pan et al., 2011). Evening intake of energy is positively associated with overweight/obesity (Almoosawi et al., 2016).

More recently, discernible temporal eating patterns that differed by sociodemographic and eating profiles were revealed by latent class analysis using nutrition survey data (Leech et al., 2017; Mansukhani and Palla, 2018). 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 one or two days dietary recall, and therefore could not capture the day-to-day variation in temporal eating patterns.

(some review of articles about carbohydrate eating)

The National Dietary and Nutrition Survey (NDNS)

The National Diet and Nutrition Survey (NDNS) programme (NatCen Social Research, 2018) 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 eleventh 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 (2008/09-2015/16) 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 eating time patterns with diabetes and obesity.

Chapter 2

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 addition to details of what and how much was eaten, participants recorded for each eating occasion; when was it, where they were, who they were eating with. An example, used as guidance for participants, of a food diary for one day is shown in **Appendix C**.

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 (Smithers, 1993), 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

(Department of Health, 2018). Specifically, the main variables that we adopted in the current analysis were defined as:

- Total Energy intake = (protein(gramme) $\times$ 17) + (fat(gramme) $\times$ 37) + (carbohydrate(gramme) $\times$ 16) + (alcohol(gramme) $\times$ 29) kJ;
- Carbohydrate intake = total sugars (gramme) + starch (gramme);

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. 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 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 food 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 over were kept in our analysis.

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 (Office for National Statistics, 2018).

Another set of weights were generated to correct for differential non-response (either due to refusal or inability) to giving a blood sample. Response to the 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 blood sample weight, a logistic regression was used by the NDNS RP study team to model the relationship between response to 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 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 blood sample and therefore receive a low predicted probability. The inverse of these predicted probablities 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 blood sample weights were re-scaled so that the sum of the weights equalled the number of participants who had a nurse visit. The final blood weights should therefore make the blood 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) (Bates et al., 2014; Roberts et al., 2018).

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 (Collins and Lanza, 2010). 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:

$$P(U_{i1} = s_1, U_{i2} = s_2, \cdots, U_{ik} = s_K) = \sum_{t=t}^{T} P(C_i = t) \prod_{k=1}^{K} P(U_{ik} = s_k | C_i = t) \quad (2.1)$$

Where,

- $P(U_{i1} = s_1, U_{i2} = s_2, \cdots, U_{ik} = s_k)$ is the probability of observing a particular vector of responses;

- $P(C_i = t)$ is the probability that a randomly selected ith observation will be in class t;
- $P(U_{ik} = s_k | C_i = t)$ is the probability of a particular observed response pattern $U_{ik} = s_k$ conditional on membership in latent class t.

Equation 2.1 indicates that responses for an observation to the measuring variables are independent of one another given its membership in latent class t. 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 (Vermunt, 2003; Vermunt, 2008) and Asparouhov and Muthén (Muthén and Asparouhov, 2009), 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 model 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 (Snijders and Bosker, 2011).

Thus, for a binary outcome $C_{ij} = 0, 1$ (low $= 0$ or high $= 1$ carbohydrate eating days), where i denotes the observation days $(i = 1, 2, 3, 4)$, and j denotes the individual $(j = 1, 2, \cdots, 6155)$. The 2-level random intercept logistic regression model can be expressed as:

$$\begin{aligned} \text{logit}[P(C_{ij} = 1)] &= \beta_{0j} + \beta_1 x_{ij} \qquad \textbf{(day level)} \\ \beta_{0j} &= \gamma_0 + \gamma_1 w_j + u_{0j} \; \textbf{(individual level)} \\ \Rightarrow P(C_{ij} = 1) &= \frac{\exp\left(\gamma_0 + \beta_1 x_{ij} + \gamma_1 w_j + u_{0j}\right)}{1 + \exp\left(\gamma_0 + \beta_1 x_{ij} + \gamma_1 w_j + u_{0j}\right)} \end{aligned} \tag{2.2}$$

Where we define:

- $P(C_{ij} = 1)$ as the probability that the randomly selected ith observation day of jth individual is a high carbohydrate eating day;
- β_{0j} as the random intercept, for outcome $C_{ij} = 1$;
- the random deviation of the individuals u_{0j} are assumed be normally distributed (i.e. $u_{0j} \sim N(0, \sigma^2_{u_0})$), the magnitude of the u_{0j} variance ($\sigma^2_{u_0}$) indicates the influence of the individuals (level 2);
- x_{ij}, w_j 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 C_{ij} is either low or high carbohydrate eating days as if we already know, it is now replaced by a latent variable G_{ij} which indicates the typologies of carbohydrate eating patterns. Then we can use the day-level data to assess the log-odds of belonging to kth 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 kth type of carbohydrate eating pattern during the survey can be high, but for the other persons, the log-odds of following the kth type of carbohydrate eating pattern can be low.

If the day-level LCA model (carbohydrate eating pattern typologies) is best defined by $T(T \geqslant 2)$ latent classes, then $T - 1$ 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 (U_{ijk}), the MLCA model can be written as follows:

$$P(U_{ij1} = s_1, U_{ij2} = s_2, \cdots, U_{ijk} = s_K) = \sum_{t=1}^{T} P(G_{ij} = t) \prod_{k=1}^{K} P(U_{ijk} = s_k | G_{ij} = t) \tag{2.3}$$

Where,

- U_{ijk} represents the response of eating carbohydrate (one of the following: not eating any food, $< 50\%$ of the energy, or $\geqslant 50\%$ of the energy) on ith day of the survey ($i \in (1, 2, 3, 4)$) in jth individual at the kth time slot of the day ($k \in (1, 2, 3, \cdots, 7)$);
- G_{ij} denotes the latent class membership for jth individuals on ith day of the survey, the total number of day-level latent class is T;
- $P(U_{ijk} = s_k | C_{ij} = t)$ is the probability of a specific response pattern, conditional on membership in latent class t.

The $P(G_{ij} = t)$ in equation 2.3 is what we have already defined in equation 2.2:

$$P(G_{ij} = t) = \frac{\exp\left(\gamma_0 + \beta_1 x_{ij} + \gamma_1 w_j + u_{0j}\right)}{1 + \exp\left(\gamma_0 + \beta_1 x_{ij} + \gamma_1 w_j + u_{0j}\right)} \tag{2.4}$$

Non-Parametric approach

Since the parametric approach discussed above can be extremely computationally demanding (Van Horn et al., 2008; Vermunt, 2008), an alternative approach is using a non-parametric MLCA (Davidian et al., 2008). 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 $T - 1$ random intercepts, where T 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:

$$P(C_{ij} = t | CB_j = m) = \frac{\exp(\gamma_{tm})}{\sum_{r=1}^{T} \exp(\gamma_{tm})} \tag{2.5}$$

Where,

- CB_j is individual-level latent class membership for jth individual;
- γ_{tm} is day-level and individual-level indicators.

According to Finch and French's simulation study (Finch and French, 2014), 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. Specifically, we are interested in exploring both meaningful individual (level 2) latent classes and the daily 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 observation-level latent classes within individuals.

Strategy of conducting MLCA in the current analysis

To identify the best-fitting model, we used the following sequential modelling strategy (Henry and Muthén, 2010):

- 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-level;
- Next, a series of MLCA models were fitted to account for the multi-level structrure of the data. In these models, the number of observational-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, observational-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 Akaike information criterion (AIC), Bayesian information criterion (BIC), adjusted Bayesian information criterion (aBIC) where smaller values indicate better, 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) (Lo, Mendell, and Rubin, 2001; Nylund, Asparouhov, and Muthén, 2007) which compare q vs. $q-1$ class models, where q is the number of latent classes and 3) pattern interpretability. In the step of performing multilevel LCA, where LMR-LRT is 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 (Muthén and Muthén, 2017), the Mplus codes are shown in the Appendix xxx.

Association between day-level and individual-level latent classes and other variables

Day-level latent classes identified by MLCA steps were tabulated according to whether the diary was recorded during weekdays or not. A contingency table giving the frequency of responses across the 7 time slots of the survey days was produced.

Chapter 3

Results

Model selection, comparison, and interpretation

TABLE 3.1: Fit Criteria for Each Model Specification

	Number of day-level classes				
Model	**1 class**	**2 classes**	**3 classes**	**4 classes**	**5 classes**
Fixed effects model					
No. of free parameters	14	29	44	59	74
Log-likelihood	-173793.306	-172669.771	-172039.204	-171633.941	-171377.292
BIC	347728.092	345632.608	344523.06	343864.121	343502.409
Lo-Mendell-Rubun LRT	–	< 0.0001	1e-04	< 0.0001	< 0.0001
Entropy	1	0.31	0.392	0.51	0.481
Random effects model					
2 individual-level classes					
No. of free parameters		59	89	119	
Log-likelihood		-169331.132	-168700.96	-168366.193	
BIC		339258.502	338301.338	337934.968	
Entropy		0.581	0.569	0.555	
3 individual-level classes					
No. of free parameters		89	134	179	
Log-likelihood		-166936.279	-166348.815	-166062.761	
BIC		334771.968	334051.799	333934.448	
Entropy		0.677	0.63	0.644	
4 individual-level classes					
No. of free parameters		119	179		
Log-likelihood		-165441.731	-164845.696		
BIC		332086.045	331500.318		
Entropy		0.729	0.659		

Note:
Abbreviation: No, number; BIC, Bayesian information criterion; Entropy, a pseudo-r-squared index; Lo-Mendel-Rubin LRT, likelihood ratio test comparing q classes models with q-1 classes models.

–

TABLE 3.2: Day Level Latent Class Solution for Three-Class Model (No Individual level Model)

Time slots of the day	**Responses of carbohydrate intake**	n	(%)	**Class 1 (39.5%) High carbo-hydrate day**	**Class 2 (20.4%) Low carbo-hydrate day**	**Class 3 (40.1%) Regular meals day**
6 am – 9 am						
	Not eating	7655	31.2	0.129	0.450	0.320
	Carbohydrate < 50%*	4500	18.4	0.130	0.267	0.128
	Carbohydrate ⩾ 50%†	12328	50.4	0.741	0.283	0.552
9 am – 12 am						
	Not eating	5447	22.2	0.237	0.079	0.401
	Carbohydrate < 50%	7227	29.5	0.158	0.492	0.173
	Carbohydrate ⩾ 50%	11809	48.2	0.605	0.429	0.426
12 noon – 2 pm						
	Not eating	4783	19.5	0.156	0.356	0.019
	Carbohydrate < 50%	11112	45.4	0.405	0.413	0.560
	Carbohydrate ⩾ 50%	8588	35.1	0.439	0.231	0.421
2 pm – 5 pm						
	Not eating	6926	28.3	0.130	0.123	0.659
	Carbohydrate < 50%	8277	33.8	0.249	0.602	0.076
	Carbohydrate ⩾ 50%	9280	37.9	0.621	0.276	0.266
5 pm – 8 pm						
	Not eating	3043	12.4	0.114	0.199	0.034
	Carbohydrate < 50%	14240	58.2	0.516	0.590	0.639
	Carbohydrate ⩾ 50%	7200	29.4	0.370	0.211	0.328
8 pm – 10 pm						
	Not eating	8722	35.6	0.322	0.291	0.480
	Carbohydrate < 50%	8898	36.3	0.266	0.551	0.212
	Carbohydrate ⩾ 50%	6863	28.0	0.412	0.158	0.308
10 pm – 6 am						
	Not eating	16295	66.6	0.680	0.590	0.751
	Carbohydrate < 50%	4144	16.9	0.074	0.294	0.101
	Carbohydrate ⩾ 50%	4044	16.5	0.246	0.115	0.148

Note:
* Carbohydrate < 50% indicates that within the time slot, carbohydrate contributed less than 50% total energy intake.
† Carbohydrate ⩾ 50% indicates that within the time slot, carbohydrate contributed higher or equal to 50% total energy intake.

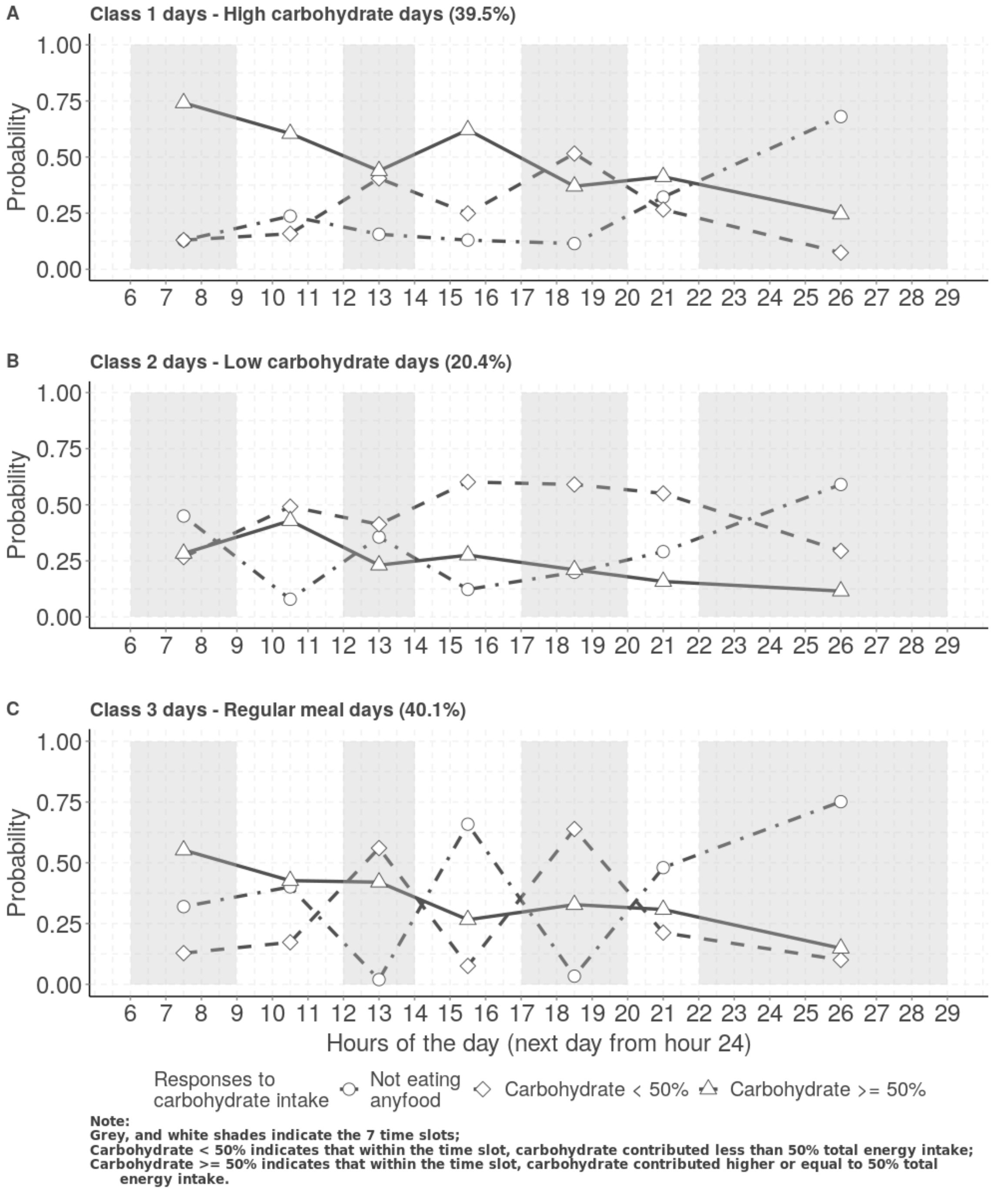

FIGURE 3.1: Day Level Latent Classes.

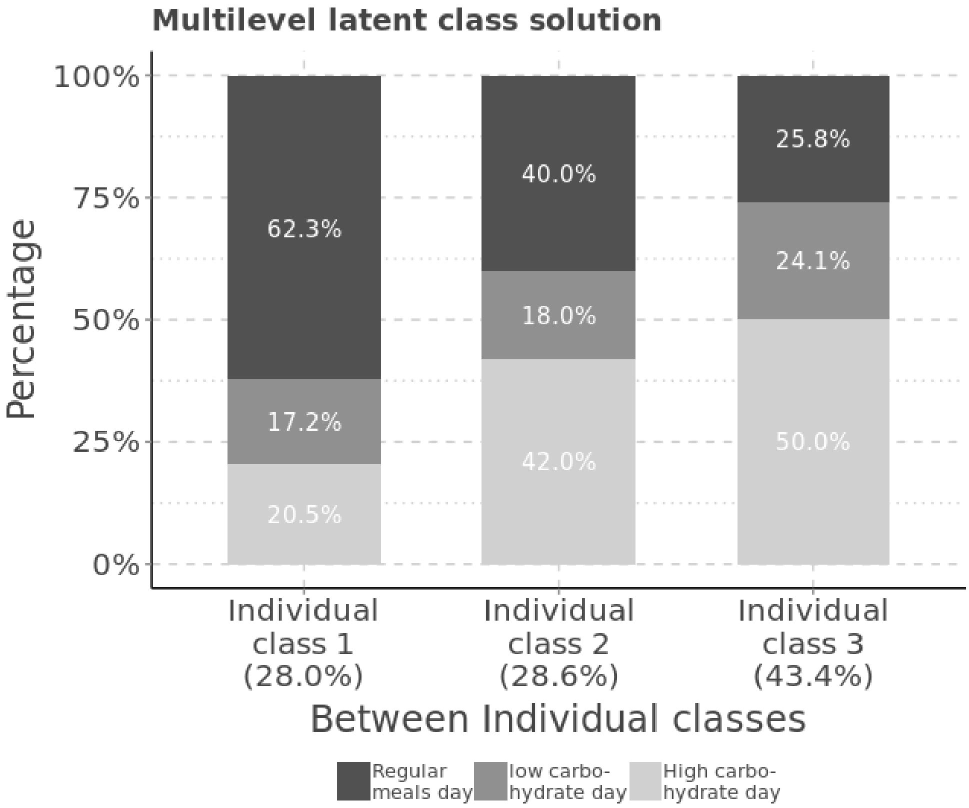

FIGURE 3.2: individual Level Latent Classes.

Subsection 2

Main Section 2

Chapter 4

Discussion and Conclusion

Main Section 1

Subsection 1

Subsection 2

Main Section 2

Bibliography

Almoosawi, S et al. (2016). "Chrono-nutrition: a review of current evidence from observational studies on global trends in time-of-day of energy intake and its association with obesity". In: *Proceedings of the Nutrition Society* 75.4, pp. 487–500.

Asher, Gad and Paolo Sassone-Corsi (2015). "Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock". In: *Cell* 161.1, pp. 84–92.

Bates, Beverley et al. (2014). *National Diet and Nutrition Survey: Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/2009-2011/2012): A survey carried out on behalf of Public Health England and the Food Standards Agency*. Public Health England.

Collins, L.M. and S.T. Lanza (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Wiley Series in Probability and Statistics. Wiley.

Davidian, Marie et al. (2008). "Growth mixture modeling: Analysis with non-Gaussian random effects". In: *Longitudinal Data Analysis*. Chapman and Hall/CRC, pp. 157–180.

De Bacquer, Dirk et al. (2009). "Rotating shift work and the metabolic syndrome: a prospective study". In: *International Journal of Epidemiology* 38.3, pp. 848–854.

Department of Health (2018). *National Diet and Nutrition Survey Rolling Programme*. https://www.gov.uk/government/collections/national-diet-and-nutrition-survey.

Finch, W Holmes and Brian F French (2014). "Multilevel latent class analysis: Parametric and nonparametric models". In: *The Journal of Experimental Education* 82.3, pp. 307–333.

Henry, Kimberly L and Bengt Muthén (2010). "Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors". In: *Structural Equation Modeling* 17.2, pp. 193–215.

Johnston, Jonathan D (2014). "Physiological responses to food intake throughout the day". In: *Nutrition Research Reviews* 27.1, pp. 107–118.

Leech, Rebecca M et al. (2017). “Temporal eating patterns: a latent class analysis approach”. In: *International Journal of Behavioral Nutrition and Physical Activity* 14.1, p. 3.

Lo, Yungtai, Nancy R Mendell, and Donald B Rubin (2001). “Testing the number of components in a normal mixture”. In: *Biometrika* 88.3, pp. 767–778.

Mansukhani, Raoul and Luigi Palla (Jan. 2018). “Investigating eating time patterns in UK adults from The 2008–2012 National Diet and Nutrition Survey”. In: 77.

Muthén, Bengt and Tihomir Asparouhov (2009). “Multilevel regression mixture analysis”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172.3, pp. 639–657.

Muthén, Linda K and Bengt O Muthén (2017). *Mplus: Statistical analysis with latent variables: User's guide*. Muthén & Muthén Los Angeles.

NatCen Social Research (2018). *National Diet and Nutrition Survey Years 1-8, 2008/09-2015/16*. http://doi.org/10.5255/UKDA-SN-6533-8.

Nylund, Karen L, Tihomir Asparouhov, and Bengt O Muthén (2007). “Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study”. In: *Structural Equation Modeling* 14.4, pp. 535–569.

Office for National Statistics (2018). *Mid 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, and 2016 Population Estimates*. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates.

Pan, An et al. (2011). “Rotating night shift work and risk of type 2 diabetes: two prospective cohort studies in women”. In: *PLoS Medicine* 8.12, e1001141.

Roberts, Caireen et al. (2018). “National Diet and Nutrition Survey: results from years 7 and 8 (combined) of the Rolling Programme (2014/2015–2015/2016)”. In:

Smithers, Gillian (1993). “MAFF's nutrient databank”. In: *Nutrition & Food Science* 93.2, pp. 16–19.

Snijders, T.A.B. and R.J. Bosker (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE Publications.

Uemura, Mayu et al. (2015). “Breakfast skipping is positively associated with incidence of type 2 diabetes mellitus: evidence from the Aichi Workers' Cohort Study”. In: *Journal of Epidemiology* 25.5, pp. 351–358.

Van Horn, M Lee et al. (2008). “Using multilevel mixtures to evaluate intervention effects in group randomized trials”. In: *Multivariate Behavioral Research* 43.2, pp. 289–326.

Vermunt, Jeroen K. (2003). *Multilevel Latent Class Models*. Vol. 33. 1, pp. 213–239. DOI: 10.1111/j.0081-1750.2003.t01-1-00131.x.

Vermunt, Jeroen K (2008). “Latent class and finite mixture models for multilevel data sets”. In: *Statistical Methods in Medical Research* 17.1, pp. 33–51.

Appendix A

R code for importing and manipulating the data

```
# NDNS analysis, data management ------------------------------------------

# Change the data path accordingly ----------------------------------------
setwd("/home/wangcc-me/Downloads/UKDA-6533-stata11_se/stata11_se/") # in Ubuntu
library(epiDisplay)
library(plyr)
library(tidyverse)

# Read the data into memory -----------------------------------------------
library(haven)
data <- read_dta("ndns_rp_yr1-4a_foodleveldietarydata_uk.dta")
data56 <- read_dta("ndns_rp_yr5-6a_foodleveldietarydata.dta")
data78 <- read_dta("ndns_rp_yr7-8a_foodleveldietarydata.dta")
names(data)
names(data56)
names(data78)
names(data)[names(data) == "seriali"] <- "id"
names(data56)[names(data56) == "seriali"] <- "id"
names(data78)[names(data78) == "seriali"] <- "id"

# Extract the data we needed ----------------------------------------------
df14d <- data[, c(113, 1, 2, 3, 5, 6, 7, 8, 9, 21, 24, 55, 57, 58,
    59, 60, 61, 62, 63, 64)]
var <- names(df14d)
df56d <- data56 %>% select(var)
```

```
df78d <- data78 %>% select(var)
dfs1 <- rbind(df14d, df56d, df78d)
dfs2 <- dfs1[dfs1$Age >= 19, ]
rm(data, data56, data78)
dfs2

# Calculate the time (minute and hour) when they eat ----------------------

dfs2$MealTime_chr <- as.character(dfs2$MealTime)
dfs2$MealTime_hm <- unlist(strsplit(dfs2$MealTime_chr, " "))[c(FALSE,
    TRUE)]
dfs2$MealHourN <- as.numeric(unlist(strsplit(dfs2$MealTime_hm, ":"))[c(TRUE,
    FALSE, FALSE)])
dfs2$MealMinN <- as.numeric(unlist(strsplit(dfs2$MealTime_hm, ":"))[c(FALSE,
    TRUE, FALSE)])
dfs2$MealMinN0 <- (60 * dfs2$MealHourN) + dfs2$MealMinN
dfs3 <- dfs2[order(dfs2$id, dfs2$DayNo, dfs2$MealMinN0), ]
length(unique(dfs3$id))  ## number of participants = 6155

# Create a subset data with only the first observation of each
# participant --------
NDNS <- dfs3[!duplicated(dfs3$id), ]
with(NDNS, tab1(SurveyYear, graph = FALSE, decimal = 2))

# #SurveyYear :
            # Frequency Percent Cum. percent
# NDNS Year 1       801   13.01        13.01
# NDNS Year 2       812   13.19        26.21
# NDNS Year 3       782   12.71        38.91
# NDNS Year 4      1055   17.14        56.05
# NDNS Year 5       625   10.15        66.21
# NDNS Year 6       663   10.77        76.98
# NDNS Year 7       703   11.42        88.40
# NDNS Year 8       714   11.60       100.00
  # Total          6155  100.00       100.00


# create a variable combine id and day No ---------------------------------
```

```
dfs3 <- dfs3 %>%
mutate(id_dy = paste(id, DayNo, sep = "D"))

# For each subject, the total energy/carbohydrate intake for each eating
# time can be calculated --------
old <- Sys.time()
Energy <- ddply(dfs3, .(id_dy, id, SurveyYear, DayNo, Age, Sex,
                        DiaryDaysCompleted, MealHourN, DayofWeek),
                summarise,
                Tot_Energ = sum(EnergykJ),
                Tot_Carb = sum(Carbohydrateg),
                Tot_Sugar = sum(Totalsugarsg),
                Tot_Starch = sum(Starchg))
new <- Sys.time() - old
print(new)
# Time difference of 3.876385 mins

rm(df14d, df56d, df78d, dfs2)

# Calculate the energy from total carbohydrates ---------------------------
Energy <- Energy %>%
  mutate(KJcarbo = Tot_Carb * 16) %>%
  mutate(CarKJpercentage = KJcarbo/Tot_Energ) %>%
  mutate(Carbo = cut(CarKJpercentage, breaks = c(0, 0.26, 0.75, 2),
        right = FALSE)) %>% mutate(Carbo2 = cut(CarKJpercentage, breaks = c(0,
    0.26, 2), right = FALSE))
Energy0 <- Energy[!(Energy$Tot_Energ == 0), ]
          # some food consumption does not contain any carbohydrates
Energy0$Carbo <- factor(Energy0$Carbo, labels = c("Low_carb", "Med_carb",
    "High_carb"))
Energy0$Carbo2 <- factor(Energy0$Carbo2, labels = c("Low_carb", "Med_or_high_carb")

# Generate data sets for each day ----------------------------------------------
dta_day1 <- Energy0 %>%
  filter(DayNo == 1) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
```

```
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))

dta_day2 <- Energy0 %>%
  filter(DayNo == 2) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))

dta_day3 <- Energy0 %>%
  filter(DayNo == 3) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))

dta_day4 <- Energy0 %>%
  filter(DayNo == 4) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))
vecid1 <- unique(dta_day1$id) # n = 6153
vecid2 <- unique(dta_day2$id) # n = 6153
vecid3 <- unique(dta_day3$id) # n = 6151
vecid4 <- unique(dta_day4$id) # n = 6026

Noday1 <- setdiff(vecid, vecid1)  # two subjects did not have day 1 data
Noday2 <- setdiff(vecid, vecid2)  # two subjects did not have day 2 data
Noday3 <- setdiff(vecid, vecid3)  # four subjects did not have day 3 data
Noday4 <- setdiff(vecid, vecid4)  # 129 subjects did not have day 4 data

# Transform the data shape from long to wide ---------------------------------
dta_d1_wide <- dta_day1[, -7] %>%
```

```
  spread(key = MealHourN, value = Carbo)
names(dta_d1_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

dta_d2_wide <- dta_day2[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d2_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

dta_d3_wide <- dta_day3[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d3_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

dta_d4_wide <- dta_day4[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d4_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")
# recode NA to not eating -------------------------------------------------
for (i in 5:ncol(dta_d1_wide))
  if (is.factor(dta_d1_wide[, i])) levels(dta_d1_wide[,
    i]) <- c(levels(dta_d1_wide[, i]), "Not_eating")

dta_d1_wide[is.na(dta_d1_wide)] <- "Not_eating"

for (i in 5:ncol(dta_d2_wide))
  if (is.factor(dta_d2_wide[, i])) levels(dta_d2_wide[,
    i]) <- c(levels(dta_d2_wide[, i]), "Not_eating")

dta_d2_wide[is.na(dta_d2_wide)] <- "Not_eating"

for (i in 5:ncol(dta_d3_wide))
  if (is.factor(dta_d3_wide[, i])) levels(dta_d3_wide[,
    i]) <- c(levels(dta_d3_wide[, i]), "Not_eating")

dta_d3_wide[is.na(dta_d3_wide)] <- "Not_eating"

for (i in 5:ncol(dta_d4_wide))
  if (is.factor(dta_d4_wide[, i])) levels(dta_d4_wide[,
    i]) <- c(levels(dta_d4_wide[, i]), "Not_eating")

dta_d4_wide[is.na(dta_d4_wide)] <- "Not_eating"
```

Appendix B

Mplus code for Multilevel LCA models

Appendix C

Example of a food diary for one day

Day EXAMPLE			**Day:** Thursday	**Date:** March 31st
Time	**Where? With whom? TV on? Table?**	**What**	**Brand Name**	**Amount eaten**
How to describe what you had and how much you had can be found on pages 20-25				
6am to 9am				
7.30am	Kitchen	Orange juice, unsweetened, UHT	Tesco	Large glass
	Family	Tea	Tesco	Mug
	No TV	Milk, fresh semi skimmed	Tesco	A little
	At table	Sugar white	Silverspoon	2 level teaspoons
		Weetabix		2
		Milk as above		Drowned
		Sugar as above		2 heaped teaspoons
		Toast wholemeal, large loaf	Hovis	2 thin slices
		Butter unsalted	Anchor	thick spread on both
		Strawberry Jam	Co-op	1 teaspoon on one slice
9am to 12 noon				
11am	School playground	Coca cola diet	Coca Cola	330ml can
	With friends	Potato crisps, Salt and Vinegar	Walkers	25g packet from a multipack
12noon	School corridor	Water from water cooler		small plastic cup
	Alone	Mars Bar		1 kingsize
12 noon to 2pm				
12.45pm	School canteen	Sandwich, from home		
	With friends	White bread, large loaf	Kingsmill	2 med slices
	At table	Spread	Flora Light	thin spread on both slices
		Ham unsmoked	Tescos	1 slice
		Cheddar cheese		2 medium slices
		Branston Pickle		1 teaspoon
		Apple with skin from home		1 (left core)
		Ribena Light, Ready to Drink, Blackcurrant, from canteen		220ml carton
		Kitkat from home		2 fingers
1.50pm	School corridor			
	Alone	Chewing gum	Orbit Sugar Free	1 piece

FIGURE C.1: NATIONAL DIET AND NUTRITION SURVEY – Food and Drink Diary Example, from 6 am to 2 pm.

Day EXAMPLE			Day: Thursday	Date: March 31st
Time	Where? With whom? TV on? Table?	What	Brand Name	Amount eaten
		2pm to 5pm		
3.45pm	Bus Alone	Wine gums	Maynards	140g packet
4.30pm	Home, sitting room, With family TV on Not at table	Tea (as above) Chocolate Hob Nobs	Mcvitites	mug 3
		5pm to 8pm		
6.30pm	Friend's kitchen With friends No TV At table	Chicken in tomato sauce made by friend's mum Tomato fresh Sweetcorn tinned Peach yoghurt low fat Lemon squash No Added Sugar	See recipe Mullerlight Sainsbury's	3 tablespoons 3 slices 1 dessertspoon 200g pot medium glass
		8pm to 10pm		
8pm	Home, sitting room Alone TV on, Not at table	Satsuma Cream Crackers (no spread)	 Jacob's	1 4
9.30pm	Kitchen Alone No TV, At table	Thick cut, frozen chips fried in vegetable oil Brown sauce	McCains HP	small portion 1 dessertspoon
		10pm to 6am		
10.30pm	Bedroom Alone TV on Not at table	Hot chocolate drink made with water	Cadbury's	Mug (made with 4 tsp powder)
2am	Bedroom (in bed) Alone No TV	Water tap		½ small glass

FIGURE C.2: NATIONAL DIET AND NUTRITION SURVEY – Food and Drink Diary Example, from 2 pm to 6 am.

Write in recipes or ingredients of made up dishes or take-away dishes			
NAME OF DISH: *Chicken in tomato Sauce*		***Serves:*** *4 people*	
Ingredients	***Amount***	***Ingredients***	***Amount***
Pieces of chicken	*3 pieces*	*Olive oil*	*2 tbsp*
Sauce made with:			
Tinned tomatoes	*1 tin*		
Green pepper	*1 medium*		
Onion	*1 small*		
Brief description of cooking method Chicken pieces fried in olive oil, then mixed in with tomato and vegetable sauce.			

FIGURE C.3: NATIONAL DIET AND NUTRITION SURVEY – Food and Drink Diary Example, home made food recipes.