



# 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

<b>Declaration of Authorship</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
Background . . . . .	1
The National Dietary and Nutrition Survey (NDNS) . . . . .	1
Aims and objectives . . . . .	2
<b>2 Methods</b>	<b>3</b>
Dietary diary collected in the NDNS RP . . . . .	3
Definition of carbohydrate intake . . . . .	3
Latent class analysis (LCA) . . . . .	4
Multilevel latent class analysis (MLCA) . . . . .	5
Parametric approach . . . . .	5
Non-Parametric approach . . . . .	6
Strategy . . . . .	7
Survey Data . . . . .	7
Survey Selection Method . . . . .	7
Response rates . . . . .	7
<b>3 Results</b>	<b>9</b>
Main Section 1 . . . . .	9
Subsection 1 . . . . .	9
Subsection 2 . . . . .	9
Main Section 2 . . . . .	9
<b>4 Discussion and Conclusion</b>	<b>11</b>
Main Section 1 . . . . .	11
Subsection 1 . . . . .	11
Subsection 2 . . . . .	11
Main Section 2 . . . . .	11
<b>Bibliography</b>	<b>13</b>
<b>A R code for importing and manipulating the data</b>	<b>15</b>
<b>B SAS code for mixed effect LCA analysis</b>	<b>20</b>

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

**21**

## List of Figures

C.1	One day food diary example 6 am to 2 pm . . . . .	21
C.2	One day food diary example 2 pm to 6 am . . . . .	22
C.3	Food diary example of home made food recipes . . . . .	22



# List of Tables





# List of Abbreviations

<b>AIC</b>	<b>Akaike Information Criterion</b>
<b>aBIC</b>	<b>adjusted Bayesian Information Criterion</b>
<b>cAIC</b>	<b>consistent Akaike Information Criterion</b>
<b>BIC</b>	<b>Bayesian Information Criterion</b>
<b>EM</b>	<b>Expectation Maximization</b>
<b>FSA</b>	<b>Food Standards Agency</b>
<b>LCA</b>	<b>Latent Class Analysis</b>
<b>LTA</b>	<b>Latent Transition Analysis</b>
<b>MAFF</b>	<b>Ministry of Agriculture, Fisheries and Food</b>
<b>MAR</b>	<b>Missing At Random</b>
<b>MCAR</b>	<b>Missing Completely At Random</b>
<b>MLCA</b>	<b>Multilevel Latent Class Analysis</b>
<b>MNAR</b>	<b>Missing Not At Random</b>
<b>ML</b>	<b>Maximum Likelihood</b>
<b>NDNS</b>	<b>the National Dietary and Nutrition Survey</b>
<b>OR</b>	<b>Odds Ratio</b>
<b>PHE</b>	<b>Public Health England</b>



# Chapter 1

## Introduction

### Background

The widely accepted norm these days seems to be that we eat three times a day. However, is this really an ideal temporal eating pattern for everyone? 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).

### 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 rolling programme 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**

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

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

- Total sugars = sum of all individual sugars (gram).

To produce a sequence of discrete responses about the carbohydrate intake we are interested, we calculated the energy consumption per hour for each participant. The percentage of energy that contributed by carbohydrate within each hour were calculated for every every participant, each single day. 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 an hour if there is any energy intake occurred, carbohydrate consumption was categorised into whether it's energy contribution was higher or equal to 50% of total energy. Consequently, for each day of the recording, there were 24 data points generated by the dairy.

## 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:

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

Where,

- $P(U_{i1} = s_1, U_{i2} = s_2, \dots, U_{ik} = s_k)$  is the probability of observing a particular vector of responses;
- $P(C_i = t)$  is the probability that a randomly selected  $i$ th 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 4 sequences of categorical responses reflecting the type of carbohydrate consumption at each hour of the day. The 4 observed sequences (observations) 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 (observations) 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 will belong to a particular level 1 latent class to vary across Level 2 units (individuals).

### 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. In an unconditional logistic regression model, the probability of the outcome (i.e. being in latent class  $k$ ) is constant within the 4-day survey for each individual (level 2). Therefore, say when we are fitting a model with  $k(k = 1, \dots, K)$  latent classes in level 1, then in each individual (level 2), there is a probability of being in latent class  $k$ . A random effect model consider the individual (level 2) to be drawn from a population of adults in the UK, and the probability of the outcome (i.e. being in latent class  $k$ ) across individuals is considered to be a random variable (Snijders and Bosker, 2011). The 2-level random intercept effect regression model can be expressed as:

$$\begin{aligned} \text{logit}[P(C_{ij} = t)] &= \beta_{0j} && \text{(level 1)} \\ \beta_{0j} &= \gamma_0 + \gamma_1 w_j + u_{0j} && \text{(level 2)} \\ \Rightarrow P(C_{ij} = t) &= \frac{\exp(\gamma_0 + \gamma_1 w_j + u_{0j})}{1 + \exp(\gamma_0 + \gamma_1 w_j + u_{0j})} && (2.2) \end{aligned}$$

Where we define:

- $P(C_{ij} = t)$  as the probability that the randomly selected  $i$ th observation of  $j$ th individual is belonging to latent class  $t$ ;
- $u_{0j}$  as the random intercept for  $j$ th individual;
- the random intercept are assumed be normally distributed (i.e.  $u_{0j} \sim N(0, \sigma_{u_0}^2)$ ), the magnitude of the  $u_{0j}$  variance ( $\sigma_{u_0}^2$ ) indicates the influence of the individuals (level 2);
- $w_j$  is the predictor for individual (level 2), such as age, and/or sex.

Same as in 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 hour, over 24 hours and during 4 consecutive days of survey period). Considering that the latent class indicators are indicator variables ( $U_{ijk}$ ), the MLCA model can be written as follows:

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

Where,

- $U_{ijk}$  represents the response of eating **high/low** carbohydrate on  $i$ th day of the survey ( $i \in (1, 2, 3, 4)$ ) in subject  $j$  (level 2) at the  $k$ th hour of the day ( $k \in (1, 2, 3, \dots, 24)$ );
- $C_{ij}$  denotes the latent class membership for subject  $j$  on  $i$ th day of the survey;
- A specific latent class is referred to as  $t$ , and the total number of level 1 latent classes is denoted by  $T$ ;
- $P(U_{ijk} = s_k | C_{ij} = t)$  is the probability of a specific response pattern, conditional on membership in latent class  $t$ .

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

### Non-Parametric approach

Another approach is using a non-parametric MLCA, in which separate latent class models are specified for level 1 and level 2. 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 latent classes reflect differences in the probability of belonging to a specific level 1 latent class, so that clusters (i.e., individuals) that contain observations with similar probabilities for the level 1 latent classes will be grouped together. The non-parametric MLCA model can be defined as follows:

$$P(C_{ij} = t | CB_j = m) = \frac{\exp(\gamma_{tm})}{\sum_{r=1}^T \exp(\gamma_{rm})} \quad (2.4)$$

Where,

- $CB_j$  is level 2 latent class membership for  $j$ th individual;
- $\gamma_{tm}$  is level 1 and level 2 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 meaningful individual (level 2) latent classes rather than their daily consumption classification. Therefore, non-parametric MLCA was employed 1) to identify latent classes of observations (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

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 structure 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.

## Survey Data

### Survey Selection Method

### Response rates



## **Chapter 3**

# **Results**

### **Main Section 1**

#### **Subsection 1**

#### **Subsection 2**

### **Main Section 2**



## **Chapter 4**

# **Discussion and Conclusion**

### **Main Section 1**

#### **Subsection 1**

#### **Subsection 2**

### **Main Section 2**



# Bibliography

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

### **SAS code for mixed effect LCA analysis**

## Appendix C

### Example of a food diary for one day

Day			Day: Thursday	Date: March 31 <sup>st</sup>
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				
<i>6am to 9am</i>				
7.30am	Kitchen Family No TV At table	Orange juice, unsweetened, UHT Tea Milk, fresh semi skimmed Sugar white Weetabix Milk as above Sugar as above Toast wholemeal, large loaf Butter unsalted Strawberry Jam	Tesco Tesco Tesco Silverspoon  Hovis Anchor Co-op	Large glass Mug A little 2 level teaspoons 2 Drowned 2 heaped teaspoons 2 thin slices thick spread on both 1 teaspoon on one slice
<i>9am to 12 noon</i>				
11am	School playground With friends	Coca cola diet Potato crisps, Salt and Vinegar	Coca Cola Walkers	330ml can 25g packet from a multipack
12noon	School corridor Alone	Water from water cooler Mars Bar		small plastic cup 1 kingsize
<i>12 noon to 2pm</i>				
12.45pm	School canteen With friends At table	Sandwich, from home White bread, large loaf Spread Ham unsmoked Cheddar cheese Branston Pickle Apple with skin from home Ribena Light, Ready to Drink, Blackcurrant, from canteen Kitkat from home	Kingsmill Flora Light Tescos	2 med slices thin spread on both slices 1 slice 2 medium slices 1 teaspoon 1 (left core) 220ml carton 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			Day: Thursday	Date: March 31 <sup>st</sup>
EXAMPLE				
Time	where? with whom? TV on? Table?	what	Brand Name	Amount eaten
<i>2pm to 5pm</i>				
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
<i>5pm to 8pm</i>				
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
<i>8pm to 10pm</i>				
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
<i>10pm to 6am</i>				
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		$\frac{1}{2}$ 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: <i>Chicken in tomato Sauce</i>		Serves: <i>4 people</i>	
Ingredients	Amount	Ingredients	Amount
<i>Pieces of chicken</i>	<i>3 pieces</i>	<i>Olive oil</i>	<i>2 tbsp</i>
<i>Sauce made with:</i>			
<i>Tinned tomatoes</i>	<i>1 tin</i>		
<i>Green pepper</i>	<i>1 medium</i>		
<i>Onion</i>	<i>1 small</i>		
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.