

**An investigation of dietary patterns in UK adults as a method for developing a brief diet quality assessment tool**

**Katharine Elissa Roberts**

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Medicine, Dentistry and Health

School of Health and Related Research

April 2017

**Funding**

This research has been funded through a MRC scholarship.

**Acknowledgements**

The first person on my list to thank needs to be my husband, Andy, who, despite some initial reservations resulting from his witnessing of my levels of stress in doing a MSc while working full-time, has supported me with patience, encouragement and love. There have been many weekends where spending ‘quality time’ together has had to be sacrificed for me to work towards a goal that has sometimes felt unreachable. Thank you so much, I couldn’t have done it without you.

I would like to thank Professor Paul Bissell for offering me the amazing opportunity to do this PhD through the MRC scholarship which made it possible for me to pursue my studies on a part-time basis.

I would like to acknowledge Ben Kearns, friend and Research Associate at SCHARR, for helping me to understand and undertake some pilot Principal Component Analysis early in the development of my research questions and thesis.

I would like to give a huge thank you to my lovely and brilliant supervisors, Michelle, Jeremy and Janet. Thank you all for your patience and unwavering support and for your enthusiasm with gentle pragmatism for my changing and meandering plans. I have greatly appreciated the individual and different contributions that you have made to my PhD experience and to my personal and professional development.

Michelle, as my principal supervisor, your professional and personal support has been incredible, particularly during the ups and downs of life that can happen over a 6 year period – thank you.

Jeremy, thank you for your patience with me and your amazing ability to explain statistics so clearly and in such a way as to make them enjoyable!

Janet, thank you for your enthusiasm, positivity, encouragement and brilliant insights that have shaped the overall ‘story’ and purpose of my thesis.

**List of publications & communications**

*Oral communications*

Roberts, K., Holdsworth, M., Dawson, J and Cade, J. ‘Dietary patterns in the UK and their association with socio-demographic characteristics, lifestyle and diet quality’. Conference of the International Society of Behavioural Nutrition and Physical Activity, Cape Town, June 2016.

*Posters at conference*

Roberts, K., Holdsworth, M., Dawson, J and Cade, J. ‘Exploring dietary patterns in the UK adult population towards the development of a brief diet quality assessment tool.’ International Conference on Diet and Activity Methods, Brisbane, September 2015.

*Journal articles in preparation*

1. Dietary patterns in the UK population and their association with socio-economic status, lifestyle factors, nutrient intake and nutrient biomarkers.

*Planned to submit to: British Journal of Nutrition*

1. Development of a UK Nutrient-based Diet Quality Score

*Planned to submit to: Public Health Nutrition*

1. The use of dietary patterns analyses methods to develop the Brief Diet Quality Assessment Tool (BDQAT) for public health research.

*Planned to submit to: European Journal of Clinical Nutrition*

**List of contents**

Funding and Acknowledgements………………………………………………………………...2

List of publications and communications……………………………………………….............3

List of contents……………………………………………………………………………………..4

List of tables……………………………………………………………………………………….11

List of figures………………………………………………………………………………………14

List of abbreviations and acronyms…………………………………………………………….15

Abstract……………………………………………………………………………………………16

Chapter 1. Introduction and background……………………………………………………….18

* 1. Aims and objectives………………………………………………………………...18
  2. Background: diet as a determinant of health in the UK…………………………18
     1. Obesity: prevalence, causes and consequences for health………..19
     2. Type 2 diabetes: prevalence, causes and consequences for health.20
     3. Cardiovascular diseases: causes and consequences for health……22
     4. Diet-related cancers: causes and consequences for health…………23
  3. Monitoring diet at population level to inform public health policy and evidence based practice…………………………………………………………………………24
     1. Population level dietary assessment methods…………………………26
        1. Weighed or estimated food diaries…………………………27
        2. Dietary recall………………………………………………….28
        3. Food Frequency Questionnaires (FFQs)…………………..28
        4. The use of new technologies in dietary assessment……..29
  4. Population level dietary surveys in the UK………………………………………….30
     1. The National Diet and Nutrition Survey (NDNS)……………………….30
     2. The Health Survey for England………………………………………….31
     3. The Living Costs and Food Survey……………………………………..32
     4. The Active People Survey………………………………………………..32
  5. Diet as a complex behaviour: the wider determinants of diet and their relevance in population monitoring of diet…………………………………………..………….32
     1. Economic influences on diet…………………………………………….34
     2. Social and cultural influences on diet…………………………………..34
     3. Physical environmental influences on diet……………………………..35
     4. Psycho-social influences on diet………………………………………..35
     5. Conceptual models for the wider determinants of diet……………….36
  6. Population level diet quality in the UK………………………………………………37
     1. Dietary recommendations in the UK……………………………………37
     2. Inequalities, lifestyle and patterns in the diets of UK adults…………38
        1. Measures of socio-economic status (SES)……………….38
        2. Smoking status………………………………………………38
        3. Age……………………………………………………………38
  7. Dietary patterns……………………………………………………………………….40
     1. Dietary patterns and their application in nutritional epidemiology…..40
     2. Dietary patterns analyses: methods……………………………………41
        1. Empirically or data-driven methods *(a posteriori)…….....*42
        2. Theoretically driven methods *(a priori)*……………………44
     3. Empirically-derived dietary patterns studies in the UK: an overview..46
     4. Theoretically-derived dietary patterns relating to diet quality: an overview……………………………………………………………………48
        1. Diet quality scores and indices in the U.S. and Europe….48
        2. Diet quality scores and indices in the UK………………….50
  8. Dietary assessment tools in the UK: an overview………………………………….52
  9. Research needs to facilitate improved availability of UK dietary data to support evidence based practice and policy in public health……………………………….54
     1. National dietary intake data that is robust at local level……………….54
     2. A UK Diet Quality Score………………………………………………….55
     3. Diet quality data by socio-economic measures…………………………55
     4. Brief, validated diet quality assessment tools for population level dietary surveillance and evaluation……………………………………..55
     5. Objectives of this thesis that will support population level dietary…..56 surveillance and evaluation

Chapter 2. An exploration and examination of dietary patterns in UK adults using empirical methods (Study 1)…………………………………………………………………………………58

2.1. Introduction, aims and objectives………………………………………………….58

2.2. Methods…………………..…………………………………………………………..59

2.2.1. Dataset…………………………………………………………………….59

2.2.2. Statistical analysis………………………………………………………..62

2.2.2.1. Empirically-derived dietary patterns analyses………………62

2.2.2.2. Key foods characterising the dietary patterns………………64

2.2.2.3. Associations with wider determinants of health…………….64

2.2.2.4. Internal consistency and stability of the dietary patterns…..64

2.2.2.5. Associations with nutrient intake and biomarkers………….65 2.2.3. Identification of items for inclusion in the YHS-DQ…………………..66

2.3. Results………………….……………………………………………………………67

2.3.1. Sample characteristics ……………………………………………………67

2.3.2. PCA results (NDNS 2008-2011)………………………………………….68

2.3.3. Characterising of dietary patterns ………………………………………..68

2.3.4. Main effects regression between dietary patterns and sample characteristics, socio-economic measures and lifestyle factors….……..……72

2.3.5. Evaluating internal consistency and stability of the dietary patterns….76

2.3.5.1. PCA in NDNS 2008-2012 dataset……………………………..76

2.3.5.2 .Food and food groups that characterised dietary patterns in the 2008-2012 dataset compared with 2008-2011 dataset…………..79

2.3.5.3. Association of dietary patterns with sample characteristics, socio-economic measures and lifestyle factors (NDNS 2008-2012)………………………………………………………………………..84

2.3.5.4. Internal stability and reproducibility (PCA with random 50% sample of NDNS 2008-2012)…………………………………………….85

2.3.6. Regression analysis with nutrient intake and nutrient biomarkers…….88

2.3.7. Identifying items for the YHS-DQ based on the dietary patterns analyses…………………………………………………………………………….90

2.3.8. Internal consistency and stability of tools selected for the YHS-DQ….91

2.4. Discussion…………………………………………………………………………..92

2.4.1. Summary of findings…………………………………………………….…92

2.4.2. Deriving the data-driven dietary patterns…………………………….….94

2.4.3. The consistency and stability of the dietary patterns……………….….94

2.4.4. The limitations of using empirically-derived dietary patterns to inform the YHS-DQ………………………………………………………………………..95

Chapter 3. The development and validation of a theoretically-derived nutrient-based UK Diet Quality Score (Study 2)…………………………………………………………………………….98

* 1. Introduction, aims and objectives………………………………………………..98
  2. Objective 1. Development of a theory driven, Nutrient-based Diet Quality Score (NDQS)………………………………………………………………………99

3.2.1. Methods……………………………………………………………………99

3.2.1.1. Dataset…………………………………………………………....99

3.2.1.2. Scope and nature of the score………………………………....99

3.2.1.3. Weighting and scoring rationale………………………………101

3.2.1.4. Macronutrients in excess at population level………………..102

i) Percentage of total energy from fat…………………………102

ii) Percentage of total energy from saturated fat…………….102

iii) Percentage of total energy from cis-PUFA…………….....103

iv) Percentage of total energy from NMES…………………...103

v) Alcohol………………………………………………………...104

3.2.1.5. Macronutrients deficient at population level…………………104

i) NSP …………………………………………………………....104

3.2.1.6. Macronutrients adequate at population level………………..105

i) Protein………………………………………………………….105

3.2.1.7. Micronutrients in excess at population level………………...105

i) Sodium…………………………………………………………105

3.2.1.8. Micronutrients deficient at population level………………….106

i) Iron……………………………………………………………..106

3.2.1.9. Micronutrients adequate at population level………………..106

i) Calcium………………………………………………………..106

ii) Vitamin C and Folate………………………………………..107

3.2.1.10. Other considerations in the development of the NDQS….107

i) Adequacy versus optimal intake…………………………….107

ii) Total fat………………………………………………………..108

iii) Total energy: as independent determinant of diet quality.108

iv) Total energy: estimating systematic under-reporting of energy intake and testing the sensitivity of the NDQS……...109

3.2.2. Results……………………………………………………………………112

3.2.2.1. Nutrient-based Diet Quality Score (NDQS)……………….…112

3.2.2.2. Distribution of the NDQS……………………………………...115

3.2.2.3. NDQS quartiles by demographics, socio-economic measures and lifestyles factors……………………………………………………..115

3.2.2.4. Distribution in sub-sample of estimated non under- reporters…………………………………………………………………..119

3.2.3. Discussion………………………………………………………………..120

3.3. Objective 2. Evaluation of the content validity of the NDQS against nutrient biomarkers……………………………………………………………………………….. 123

3.3.1. Methods………………………………………………………………….123

3.3.1.1. Dataset………………………………………………………..123

3.3.1.2. Statistical analysis……………………………………………123

3.3.1.3. Biomarker selection and interpretation……………………..123

3.2.2. Results…………………………………………………………………….127

3.2.3. Discussion……………………………………………..…………………128

3.4. Objective 3. Exploration of the associations between the NDQS and sociodemographic and lifestyle variables……………………………………..130

3.4.1. Methods…………………………………………………………………130

3.4.1.1. Dataset……………………………………………………….130

3.4.1.2. Statistical analyses………………………………………….130

3.4.2. Results……………………………………………………………………130

3.4.3. Discussion………………………………………………………………..133

Chapter 4. An evaluation of indicator variables for diet quality derived using empirically driven and theory driven methods (Study 3)……………………………………………………134

4.1. Introduction, aims and objectives……………………………………………….134

4.2. Objective 1. Exploration of associations between empirically-derived dietary patterns and diet quality (NDQS)………………………………………………………..136

4.2.1. Methods…………………………………………………………………..136

4.2.1.1. Dataset……………………………………………………….136

4.2.1.2. Data analysis………………………………………………..136

4.2.2. Results……………………………………………………………………136

4.2.3. Discussion………………………………………………………………..139

4.3. Objective 2. Identification of indicator variables for diet quality using empirically driven and theory driven methods………………………………………….141

4.3.1. Methods……………………………………………………………………141

4.3.1.1. Dataset……………………………………………………....141

4.3.1.2. Statistical analyses…………………………………………142

i) Correlation analysis ………………………………………….142

ii) Backwards elimination regression………………………….143

iii) Confirmatory analysis……………………………………….143

4.3.2. Results……………………………………………………………………144

4.3.2.1. Correlation analysis……………..…………………………….144

4.3.2.2. Backwards elimination regression……………………………146

4.3.2.3. Confirmatory analysis………………………………………….148

4.3.3. Discussion………………………………………………………………..151

4.4. Objective 3. Identification of the most parsimonious model predictive of diet quality to inform the development of a Brief Diet Quality Assessment Tool………………………………………………………………………………...154

4.4.1. Methods………………………………………………………………......154

4.4.1.1. Dataset………………………………………………………..154

4.4.1.2. Statistical analyses………………………………………….154

4.4.2. Results……………………………………………………………………155

4.4.3. Discussion………………………………………………………………..158

Chapter 5. Discussion……………………………………………………………………………..160

5.1. Summary of findings and key messages………………………………………160

5.1.1. Dietary patterns in the UK………………………………………………160

5.1.2. Development of a UK nutrient-based diet quality score (NDQS)…...161

5.1.3. Identification of indicator variables for diet quality to inform the development of a brief diet quality assessment tool………………………….162

5.2. Importance of research areas in the context of the existing literature and relevance to public health policy………………………………………………..162

5.2.1. Theory driven definition of diet quality…………………………………163

5.2.2. Empirically-derived (data-driven) dietary patterns analyses………...164

5.2.3. Identification of indicator variables to inform the development of a brief diet quality assessment tool…………………………………………………….165

5.3. Strengths and limitations………………………………………………………..166

5.3.1. Data source………………………………………………………………166

5.3.2. Theory driven Nutrient-based Diet Quality Score (NDQS)………….168

5.3.3. Dietary patterns analyses……………………………………………….170

5.3.4. Development of Brief Diet Quality Assessment Tools (BDQAT)……173

5.4. Agenda for further research……………………………………………………..175

5.4.1. Exploration of different variables included in the models……………176

5.4.2. Further development and validation of the BDQATs to facilitate efficient and standardised use…………………………………………………..176

5.4.3. Further refining of the NDQS to account for energy intake………….177

5.4.4. Development of tools for use in children……………………………...178

5.5. Conclusion………………………………………………………………………...178

Appendix A. Questionnaires…………………………. ……………...…………………………..179

Appendix B. Residual plots……………………………………………………………………….180

Appendix C. Quadratic plots………………………………………………………………………194

**List of Tables**

* 1. SACN population level dietary recommendations and estimated intake from national diet surveys in the UK…………………………………………………..38
  2. Intake of key nutrients, fruit and vegetables by age group (2013 Living Costs and Foods Survey)………………………………………………………………...40

2.1. List of NDNS food variables included in the PCA …………………………...…61

2.2. Sample characteristics (NDNS 2008-2011 and 2008-2012)…………………..68

2.3a. Rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2011)…………………………………………………………..70

2.3b. Un-rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2011)…………………………………………………………..71

2.4. Foods characterising each dietary pattern (NDNS 2008-2011)……………….72

2.5. Main effects regression model (95% confidence intervals) of each dietary pattern for sociodemographic characteristics and lifestyle factors (NDNS 2008-2011)………………………………………………………………………………………..74

2.6a. Rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2012)………………………………………………………….77

2.6b. Un-rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2012)………………………………………………………….78

2.7. Food variables characterising each dietary pattern (NDNS 2008-2012)…….79

2.8. Similarities and differences between the foods with moderate/strong associations with each dietary pattern in 2008/11 and 2008/12 datasets……………81

2.9. Main effects regression model (95% confidence intervals) of dietary patterns for sociodemographic characteristics and lifestyle factors (NDNS 2008-2012)……..82

2.10. Food variables and factor loadings for each Principal Component (random 50% sample NDNS 2008-2012)…………………………………………………………..86

2.11. Foods characterising each dietary pattern (random 50% sample NDNS 2008-2012)…………………………………………………………………………………………87

2.12. Main effects regression model (95% confidence intervals) of dietary patterns for nutrient intakes and nutrient biomarkers……………………………………………..89

2.13. Items selected for the YHS-DQ and the direction of their association with each dietary patterns in the 2008/11 and 2008/12 NDNS datasets…………………………91

3.1. Estimated BMR, mean EIRep and mean EER (SD) at PALs of 1.4 and 1.54..........................................................................................................................111

3.2. Mean ratios of EIrep to EER for at BMR. PAL of 1.4 and 1.54………………111

3.3. Mean ratios of EIRep to EER by BMI category………………………………..111

3.4. Mean PAL (SD) and ranges for adults aged 19 and over in the NDNS (2008/09) doubly labelled water sub sample study………………………………………………..111

3.5. Nutrient-based Diet Quality Score (NDQS): Criteria, scoring and national comparators ………………………………………………………………………………113

3.6. Proportion in each NDQS quartile and mean (SD) scores by sample characteristics…..…………………………………………………………………………118

3.7. Biomarkers selected for inclusion in the validation study of the NDQS, their function and hypothesised direction of correlation with the NDQS…………………125

3.8. Correlations between NDQS and nutrient biomarkers……………………....127

3.9. Main effects regression model (95% confidence intervals) NDQS for gender, ethnicity, socio-economic status, income, total energy, BMI and smoking status (NDQS 2008-2012)………………………………………………………………………131

4.1. Food variables with moderate/high factor loadings for each dietary pattern……………………………………………………………………………………...136

4.2. Main effects regression (95% confidence intervals) of NDQS for dietary patterns…………………………………………………………………………………….137

4.3. Main effects regression model (95% confidence intervals) of NDQS for dietary patterns, sociodemographic characteristics and lifestyle factors……………………139

4.4. Summary of years of NDNS data used for analyses in the thesis and associated rationale………………………………………………………………………141

4.5. Spearman’s rho correlation coefficients between food variables and NDQS (NDNS Years 2 and 4)……………………………………………………………………145

4.6. Food only indicator variable models the most strongly predictive of the NDQS (NDNS Year 2 and Year 4)………………………………………………………………147

4.7. Food and wider characteristics indicator variable models the most strongly predictive of the NDQS (NDNS Year 2 and Year 4)…………………………………148

4.8. Comparison of model fit of empirically driven method model and theory driven method model with NDQS (food variables only and food variables plus age and smoking status)………………………………………………………………………….149

4.9. Main effects of theory driven method indicator variables (food only) for the NDQS……………………………………………………………………………………..149

4.10. Main effects of theory driven method indicator variables (food plus age and smoking) for the NDQS………………………………………………………………….150

4.11. Main effects of empirical (data-driven) method indicator variables (food only) for the NDQS……………………………………………………………………………..150

4.12. Main effects of empirical (data-driven) method indicator variables (food plus age and smoking) for the NDQS………………………………………………………..151

4.13. Exploratory models and the overall effect sizes with the NDQS…………….156

4.14. Main effect regression (95% confidence intervals) of Brief Diet Quality Assessment Tool (food only) with the NDQS………………………………………….157

4.15. Main effects regression (95% confidence intervals) of Brief Diet Quality Assessment Tool (food plus alcohol) with the NDQS…………………………………158

4.16. Main effects regression (95% confidence intervals) of Brief Diet Quality Assessment Tool (food plus alcohol and smoking) with the NDQS…………………158

**List of Figures**

* 1. The ‘stream of causation’ model in relation to obesity…………………………37
  2. Flow chart of the aims and objectives of the thesis…………………………….57

2.1. Scree plot for PCA of 60 food variables (NDNS 2008-2011)………………….69

2.2. Scree plot for PCA of 60 food variables (NDNS 2008-2012)……………….…76

3.1. Distribution of NDQS scores in UK population………………………………...115

3.2. Distribution of NDQS scores in non-eUR sub-sample………………………..119

3.3. Quadratic plot of BMI and NDQS regression…………………………………..132

3.4. Quadratic plot of total energy (kcal) and NDQS regression……………...…..132

A.1. Yorkshire Health Study – Diet Questionnaire………………………………….179

A.2. Derbyshire Heart Study Questionnaire…………………………………………179

B.1.1. - B.1.12. Residual plots for dietary patterns with age, BMI and household income (NDNS 2008-2011)……………………………………………………………..180

B.2.1.- B.2.12. Residual plots for dietary patterns with age, BMI and household income (NDNS 2008-2012)……………………………………………………………..186

B.3.1.- B.3.4. Residual plots for NDQS and data-driven dietary patterns (2008-12)………………………………………………………………………………………….192

C.1.1. Quadratic plot of household income and FVOF dietary pattern (NDNS 2008-2011)……………………………………………………………………………………….194

C.1.2. Quadratic plot of household income and SFFFD dietary pattern (NDNS 2008-2012)……………………………………………………………………………………….194

C.2.1. Quadratic plot of NDQS and BMI (NDNS 2008-2012)………………………..195

C.2.2. Quadratic plot of NDQS and household income (NDNS 2008-2012)……….195

C.2.3. Quadratic plot of NDQS and total energy (NDNS 2008-2012)………………196

C.3.1.- C.3.10. Quadratic plots of NDQS and food variables (NDNS 2008)….196

C.4.1.- C.4.9. Quadratic plots of NDQS and food variables (NDNS 2012)….201

References………………………………………………………………………………206

**List of abbreviations and acronyms**

BMI Body Mass Index

CHD Coronary heart disease

CVD Cardiovascular disease

DRV Dietary Reference Values

DQI Diet Quality Index

DQS Diet Quality Score

EAR Estimated Average Requirement

ECI Eating Choices Index

FFQ Food Frequency Questionnaire

HDI Healthy Diet Index

HEI Healthy Eating Index

HSE Health Survey for England

LRNI Lower Reference Nutrient Intake

MDQ Mediterranean Diet Score

MUFA Monounsaturated fatty acids

NDNS National Diet and Nutrition Survey

NMES Non-milk extrinsic sugars

NSSEC National Statistics Socio-economic Classification

NSP Non starch polysaccharides

PCA Principal Component Analysis

PHE Public Health England

PHOF Public Health Outcomes Framework

PUFA Polyunsaturated fatty acids

RNI Reference Nutrient Intake

SACN Scientific Advisory Committee on Nutrition

SES Socio-economic status

**Abstract**

**Background**

Dietary patterns analyses reduce detailed dietary intake data to underlying patterns, derived empirically or based on *a priori* knowledge. Brief tools to assess diet quality are needed for dietary surveillance and research in the UK.  This study explores dietary patterns in UK adults as a method for developing such a tool.

**Methods**

Data analyses were conducted in National Diet and Nutrition Survey (NDNS) data (n=2083). Dietary patterns were derived through Principal Component Analysis. Associations with socio-demographic characteristics, lifestyle factors, nutrient intake and biomarkers were explored.

A nutrient-based diet quality score (NDQS) was developed based on *a priori* knowledge and validated against nutrient biomarkers. NDNS respondents were scored against the NDQS. Backwards elimination regression identified variables independently predictive of the NDQS.

Confirmatory analysis compared the indicator variables generated from empirically and theoretically driven methods as predictors of the NDQS. Regressions identified the most parsimonious model predictive of the NDQS.

**Results**

Four dietary patterns, explaining13.4% of the sample variance, were labelled according to the foods characterising them: ‘fruit, vegetables and oily fish’ (FVOF), ‘snacks, fast food and fizzy drinks’ (SFFFD), ‘sugary foods and dairy (SFD)’ and ‘meat, potatoes and beer (MPB)’. FVOF was positively associated with being female, non-white, older, a non-smoker and a higher NDQS score. SFFFD was the inverse.

The NDQS was positively associated with biomarkers of vitamins C, D, B6, total carotenoids and age and negatively with urinary sodium, being white and smoking. Both methods generated models predictive of diet quality (adjusted R2 = 0.29-0.31, 0.33 respectively). A tool with 5 items, fruit, vegetables, wholemeal bread, sugary drinks and coated chicken/turkey, was predictive of diet quality (adjusted R2 = 0.26).

**Conclusion**

Diet quality is an important measure for informing public health policy. This study found dietary patterns analyses to be useful methods for developing a brief, diet quality assessment tool.

**An investigation of dietary patterns in UK adults as a method for developing a brief diet quality assessment tool.**

**Chapter 1. Introduction and background**

This chapter describes the aims and objectives of this thesis and provides the background and rationale for the research in the context of public health nutrition policy and research in the UK and the peer reviewed literature. The aims and objectives of the research are described and are related to the relevant sections of the thesis and in a flow diagram presented at the end of this chapter (Figure 1.2.)

* 1. **Aims and objectives**

The overarching aim of this thesis was to examine dietary patterns in UK adults using theoretical and empirical methods to identify a brief, valid set of indicators that can be used to assess dietary quality at population level. This thesis is comprised of three studies that each contributes to the overarching aim:

Study 1. To explore and examine dietary patterns in UK adults derived using empirically driven methods and to describe the foods that characterise them, evaluate their internal validity and analyse their association with demographic characteristics, socio-economic measures, lifestyle factors, nutrient intake and nutrient biomarkers.

Study 2. To define and quantify a nutrient-level Diet Quality Score based on UK dietary recommendations and to explore its association with demographic characteristics, socio-economic measures, wider determinants of health and biomarkers of nutritional status in a UK dataset.

Study 3. An evaluation of indicator variables for diet quality derived using empirically driven and theory driven methods to identify the most parsimonious model that is predictive of diet quality.

* 1. **Background: Diet quality as a determinant of health in the UK**

Poor diet is associated with a number of chronic diseases, such as obesity ([1](#_ENREF_1)), type 2 diabetes ([2](#_ENREF_2)), cardio-vascular diseases and hypertension ([3](#_ENREF_3)) and some cancers ([4](#_ENREF_4)). The ‘Global burden of disease study 2010’ reported that of the top 20 risk factors for disease globally, 9 of them were dietary and these were in addition to diet-related risk factors such as high BMI and high blood pressure. Globally, a diet low in fruit ranked fourth after tobacco smoking, high blood pressure and high BMI in attributable risk ([5](#_ENREF_5)). The same study identified that in the UK, dietary risks are the number one risk factor based on death, disability-adjusted life years and years of life lost in all ages and both genders. The contribution of diet-related risk factors to the burden of illness and disease, including high Body Mass Index (BMI), is second only to tobacco use ([6](#_ENREF_6)). An earlier study estimated that the burden of diet-related ill health in terms of mortality and morbidity in the UK was similar to that of smoking ([7](#_ENREF_7)). The cost to the UK National Health Service was more than twice that of smoking, with cardiovascular disease, diabetes and cancer contributing the most. Improving diet is therefore a key priority for national and local public health agendas.

* + 1. **Obesity prevalence, causes and consequences for health**

In the UK, the proportion of the population who were categorised as obese (BMI ≥30kg/m2) increased between 1993 and 2012 from 13.2% to 24.4% among men and from 16.4 per cent to 25.1 per cent among women. The proportion of adults with a healthy BMI decreased between 1993 and 2012 from 41.0 per cent to 32.1 per cent among men and from 49.5 per cent to 40.6 per cent among women ([8](#_ENREF_8)). In children, the National Child Measurement Programme for the 2014/2015 school year reported that 21.7% of boys and 17.9% of girls in school Year 6 were classified as obese, according to the British 1990 population definition of obesity. The figure for boys and girls combined in Year 6 who were classified as obese rose from 19.1% in 2014/15 school year to 19.8% in 2015/16 ([9](#_ENREF_9)).

The increasing prevalence of obesity amongst adults and children is a major public health challenge. Being overweight or obese can increase the risk of developing a range of other health problems, such as coronary heart disease (CHD), type 2 diabetes, some cancers or stroke, therefore reducing life expectancy. Childhood obesity is strongly associated with the presence of cardiovascular risk factors in childhood ([10](#_ENREF_10)). Obesity in childhood tends to persist into adulthood, as do the cardiovascular effects which are predictors of premature mortality ([11](#_ENREF_11)).

The prevalence of obesity is not equally distributed throughout the UK population. Diet-related health outcomes such as obesity and diabetes are strongly associated with socio-economic indicators. Obesity prevalence in England is known to be associated with many indicators of socio-economic status, with higher levels of obesity found among more deprived groups ([12](#_ENREF_12)). The association is stronger for women than for men ([13](#_ENREF_13)), a pattern that has been observed in many other developed countries ([14](#_ENREF_14)). National Child Measurement Programme data from 2006/7 to 2011/12 indicate that prevalence of obesity in both Reception year and Year 6 children is approximately double in the most deprived 10% compared with the least deprived 10%.

As an important determinant of energy balance, dietary intake is a major contributor to overweight and obesity prevalence ([15](#_ENREF_15)). Diets that are ‘energy dense’, i.e. high in saturated fat and sugar are associated with obesity ([16](#_ENREF_16), [17](#_ENREF_17)), as well as type 2 diabetes ([18](#_ENREF_18), [19](#_ENREF_19)) and cardiovascular diseases including coronary heart disease (CHD) ([20](#_ENREF_20), [21](#_ENREF_21)).

Despite a number of public health campaigns and environmental interventions in the UK by the Department of Health, such as ‘Five a Day’ and ‘Change 4 Life’ ([22](#_ENREF_22)), the prevalence of adult obesity continues to increase . By 2050, it is predicted that over 60% of adult men, 50% of women and 25% of children will be obese and will cost the NHS an estimated £9.7 billion, with wider costs to society of an estimated £49.9 billion annually ([23](#_ENREF_23)).

* + 1. **Type 2 diabetes prevalence, causes and consequences for health**

Public Health England has estimated that 9% of the adult population aged 16 and over in England have diabetes with 90% of these cases being type 2 diabetes ([24](#_ENREF_24)). Type 2 diabetes is associated with lifestyle factors including being overweight. In higher income countries such as the UK, lower socio-economic status groups tend to be disproportionately affected ([25](#_ENREF_25), [26](#_ENREF_26)). This is partially due to the inverse association between some socio-economic measures such as deprivation and obesity ([27](#_ENREF_27)).

Resistance to insulin is the underlying abnormality in most people who develop type 2 diabetes ([28](#_ENREF_28)). This may be preceded by states of impaired glucose metabolism, Impaired Fasting Glucose (IFG) or Impaired Glucose Tolerance (IGT), where glucose levels are consistently elevated above normal, or high glucose levels remain two hours after consuming glucose respectively. Type 2 diabetes is associated with a number of other conditions such as hypertension and cardiovascular disease, renal disease and retinal problems ([29](#_ENREF_29)). Age-adjusted mortality rates in people with diabetes are 1.5-2.5 times higher than the general population ([30](#_ENREF_30)).

Whilst there are inherent and genetic risk factors for type 2 diabetes, such as ethnicity, familial aggregation, age and gender ([30](#_ENREF_30)), there are several modifiable risk factors. Obesity is a strong predictor of type 2 diabetes and a rising prevalence of obesity is usually accompanied by increasing prevalence of type 2 diabetes ([30](#_ENREF_30)). A systematic review concluded that exercise combined with dietary changes are effective in preventing the onset of diabetes type 2 in high risk individuals (those with impaired glucose tolerance or metabolic syndrome), but that more research is required to understand the effects of exercise alone ([31](#_ENREF_31)).

Dietary intake influences insulin response and resistance, therefore diet is an important factor in the development of the disease. There are only a few diet-only randomised controlled trials to prevent or treat type 2 diabetes ([32](#_ENREF_32)) and even fewer that examine food consumption at a nutrient level, therefore much of the evidence associating particular foods, nutrients and dietary patterns with type 2 diabetes comes from observational studies. The amount of fat in the diet may modify glucose tolerance and insulin sensitivity through a number of mechanisms. For example, high levels of dietary fat may impair glucose transport ([30](#_ENREF_30)). The type of fat consumed is also important. Higher intake of polyunsaturated fat and long chain n-3 fats may be beneficial, whereas higher intakes of saturated fat and trans fats could adversely affect glucose metabolism ([2](#_ENREF_2)). Epidemiological evidence and controlled studies suggest that replacing saturated fats and trans fats with unsaturated (polyunsaturated and monounsaturated) fats has a beneficial impact on insulin sensitivity, therefore reducing risk of type 2 diabetes ([33](#_ENREF_33))

There is evidence that the consumption of carbohydrates is important in the development of type 2 diabetes. The type of carbohydrate rather than the amount may be more important, as different types of carbohydrate result in different types of physiological response. This range of responses is defined by the ‘Glycaemic Index’ (GI) which ranks carbohydrates and foods containing carbohydrate according to their post-prandial glucose responses ([34](#_ENREF_34)). Diets comprised of low GI foods have been shown in many studies to be beneficial in improving glycaemic control ([35](#_ENREF_35)). Dietary fibre positively influences post-prandial glucose response and consumption of dietary fibre has been shown in cohort studies to have a protective effect in terms of preventing type 2 diabetes ([36](#_ENREF_36)). High fibre intake also increases satiety and reduces hunger, therefore contributing to reduced energy intake and weight management ([37](#_ENREF_37), [38](#_ENREF_38)).

Several dietary patterns have been associated with preventing type 2 diabetes, including the ‘Mediterranean’ style diet ([38](#_ENREF_38)) the ‘Prudent’ diet and the ‘low-Glycaemic Index’ diet ([39](#_ENREF_39), [40](#_ENREF_40)).These dietary patterns are heterogeneous and which components of the diet are most important in preventing or treating type 2 diabetes remains controversial. However, dietary patterns that are protective from type 2 diabetes seem to be generally characterised by the following: high consumption of soluble and insoluble fibre in the form of fruit and vegetables and whole-grains; consumption of a wide variety of fruit and vegetables; consumption of polyunsaturated and monounsaturated fatty acids (PUFA and MUFA), for example, through olive oil, nuts, seeds and oily fish; consumption of lean meat and fish; low intakes of fat overall, but in particular saturated fat and trans fats through low intake of processed meats, red meat and full fat dairy foods; and low intakes of refined, simple carbohydrates and ‘high energy density’ processed foods ([40](#_ENREF_40)).

* + 1. **Cardiovascular diseases and hypertension – prevalence, causes and consequences for health**

Cardiovascular disease (CVD) is the leading cause of mortality in the UK and in Europe ([41](#_ENREF_41)). CVD includes diseases of the heart and circulation such as CHD and stroke. In 2011, 30% of all deaths in the UK were as a result of CVD, and the most significant cause of disability was stroke ([42](#_ENREF_42)). Hypertension is a risk factor for CVD and the most important risk factor in stroke. Overweight and obese individuals have increased risk of hypertension ([43](#_ENREF_43)) and those with diabetes are twice as likely to have hypertension compared with those without diabetes ([44](#_ENREF_44)).

Numerous experimental and observational studies have indicated that a major risk factor for hypertension is consumption of a high salt (sodium) diet ([45](#_ENREF_45)). In addition, evidence from prospective studies and clinical trials have indicated that several nutrients and foods are linked with risk of CVD and hypertension. The ‘Global burden of disease study 2010’ reported that Ischaemic Heart Disease (IHD) is associated with low intake of fruit, vegetables, whole grains, nuts and seeds, fibre, omega 3 fatty acids, polyunsaturated fatty acids and high intake of red meat, processed meat, trans fatty acids and sodium. Diets low in fruit and vegetables and high in sodium are also associated with stroke ([5](#_ENREF_5)).

Prospective studies and randomised controlled trials have shown a protective effect of marine n-3 PUFA (eicosapentaenoic (EPA) and docosahexaenoic (DHA)) against CHD risk factors and outcomes. Findings of the U.S. Nurse’s Health Study reported that consumption of oily fish and derived consumption of n-3 PUFA was inversely associated with risk of fatal and non-fatal CHD as well as all-cause mortality ([46](#_ENREF_46)). Randomised controlled trials have also shown n-3 PUFA (EPA and DHA) to be effective in the secondary prevention of CHD outcomes in post-myocardial infarction patients ([47](#_ENREF_47), [48](#_ENREF_48)). Dietary patterns that comply with the ‘DASH’ diet, ‘prudent’ diet or the ‘Mediterranean’ pattern are associated with lower risk of cardio-vascular diseases such as CHD and hypertension.

* + 1. **Diet-related cancers - prevalence, causes and consequences for health**

In the UK, 331,396 cases of cancer were diagnosed in 2011 and it is estimated that about one-third of the most common cancers could be prevented through changes to diet, physical activity and weight ([49](#_ENREF_49)). There is evidence for an inverse correlation between socio-economic status and cancer incidence and mortality, which may be due to inequalities in lifestyle factors associated with cancer such as smoking, alcohol consumption, diet, physical activity levels and obesity status ([50](#_ENREF_50)).

There is some evidence that dietary components and foods are associated with cancers of various sites in the body. The synthesis of evidence undertaken by the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research in 2007 ([4](#_ENREF_4)) ranked the evidence for causal associations between food, nutrition and decreased risk of cancer of various sites in the body as ‘convincing’, ‘probable’, ‘limited-suggestive decreased risk’, ‘limited-suggestive increased risk’, ‘probable increased risk’, ‘convincing increased risk’ and ‘substantial effect on risk unlikely’ ([4](#_ENREF_4)). There were several sites where the evidence of decreased cancer risk was judged to be ‘probable’: non-starchy vegetables (including green leafy vegetables such as spinach, cruciferous such as cabbage) and fruits and cancers of the mouth, larynx, pharynx, oesophagus and stomach; fruits and lung cancer; allium vegetables and stomach cancer; garlic and colorectal cancer; foods containing folate and pancreatic cancer; foods containing carotenoids and mouth, larynx pharynx and lung cancer; foods containing beta-carotene and vitamin and oesophagus cancer; foods containing lycopene and selenium and prostate cancer; milk and calcium and colorectal cancer.

In terms of meat, fish and egg consumption, the review concluded that there was ‘probable’ evidence that red and processed meat increase the risk of stomach cancer; and ‘limited’ evidence that red meat increases the risk of oesophagus, lung, pancreas and endometrium cancer; and processed meat increases the risk of oesophagus, lung, stomach and prostate cancer. There was also some ‘limited’ evidence that smoked, grilled or barbequed animal foods increase the risk of stomach cancer ([4](#_ENREF_4)). The WCRF are expected to publish an updated evidence review on cancer prevention in 2017.

* 1. **Monitoring diet at population level to inform public health policy and evidence based practice**

This section will consider the relevance of monitoring diet at population level for public health policy, the methods available for conducting such research, their limitations and the associated challenges for public health policy makers and practitioners.

Nutritional epidemiology underpins public health nutrition: it is the study of how the occurrence of disease in a population is related to diet and nutrition ([51](#_ENREF_51)). Information about the dietary intakes of populations is the cornerstone of public health nutrition policy at national and local level. These data can be used to plan policies and interventions to treat and prevent diet-related disease ([52](#_ENREF_52)).

In order for public health resources to be allocated and appropriate services that address local need commissioned, policy makers and practitioners need robust intelligence relating to the diet and other health related behaviours of their populations at national and local level and across relevant sub-groups. Data that is robust, nationally consistent and comparable at lower geographical levels and at sub-group level, for example, by age, ethnicity or deprivation level is vital for local authorities to be able to assess their local populations both broadly in the context of other local areas and the national picture, but also to identify communities where there are particular public health problems. Robust data and intelligence at local level are also required for statutory plans and official statistics. For example, upper tier local authorities are required to produce Joint Strategic Needs Assessments which use local data and intelligence to provide evidence-based justifications for where public health resources are needed ([53](#_ENREF_53)). In addition, local level data are needed for statistics published in the Public Health Outcomes Framework ([54](#_ENREF_54)).

Some types of public health relevant data are routinely collected through registrations, GP surgeries, hospital admissions, schools and other points of contact with health professionals or official bodies. However, data about health-related behaviour, such as for smoking, physical activity, diet, alcohol consumption and drug taking are, to a large extent, collected through cross-sectional surveys. For diet in particular, cross-sectional surveys are employed to identify sub-groups of a population that is ‘at risk’ of particular nutritional deficiencies or particular diet-related disease ([55](#_ENREF_55)). They can also be used where appropriate to evaluate large scale, population level nutrition or diet-related interventions through the collection of dietary intake or behaviour data at baseline, then at the end of an intervention and at one or more follow-up data points ([55](#_ENREF_55)). An essential element of collecting dietary intake data through cross-sectional surveys is the nature of the questions asked and their validity, appropriateness and reliability in relation to the population surveyed and the methodology employed.

A wide range of methods can be utilised for dietary assessment, including anthropometric measurements using a variety of methods from skinfold callipers to full body composition scanning equipment, to evaluating biochemical markers from blood and urine samples, or employing self-reported methods such as dietary records, histories or diaries, 24-hour recall questionnaires or interviews and Food Frequency Questionnaires (FFQ) ([55](#_ENREF_55)). Each of these methods has advantages and disadvantages for the researcher and the method employed will depend on the objectives of the research and the skills, resources and time available. More invasive methods such as those involving nurse visits to take blood and urine samples are expensive and resource-intensive and are a heavy burden on participants and tend not to be practical for larger scale research, particularly research undertaken by public health practitioners and policy makers rather than academic institutions. The methods that are more frequently feasible for gathering data at a population level are considered in Section 1.3.1.

* + 1. **Population level dietary assessment methods**

This section will consider the different dietary assessment methods available for population level monitoring and surveillance and their strengths and limitations. These methods will be considered in the context of the practical resource constraints experienced by public health practitioners and policy makers at national and local level.

Dietary intake is a complex behaviour which can be measured at multiple levels from biomarkers of nutritional status collected from urine and blood plasma, to consumption of foods and food groups, to broader assessment of dietary patterns and habitual dietary related behaviour. It can be measured to assess individual level risk for a particular health outcome that is related to particular foods or nutrients, for example, measuring levels of cholesterol in the diet as an indicator for risk of CHD or to provide an overall assessment of dietary adequacy or quality to identify where intervention may be needed. Due to the inherent challenges associated with collecting accurate, unbiased dietary intake data from free living individuals, diet has been referred to as the most difficult exposure to assess ([56](#_ENREF_56)). There are many different methods available to assess dietary intake, ranging from objective methods, such as those derived from biomarkers in the blood or urine, to self-reported methods such as diaries, food frequency questionnaires and 24-hour recall ([55](#_ENREF_55)).

As data to assess population level dietary intake need to be collected from free-living individuals, self-reported methods are frequently the most feasible. Self-reported dietary intake data can be prospective (in the case of diaries) and retrospective (in the case of food frequency questionnaires, 24-hour recalls or diet histories). Each of these methods has strengths and weaknesses relating to ease of administration, participant burden, the level and type of dietary intake detail generated (for example the number of items in the construct, food or food group level of detail, estimated portion sizes, food group level or nutrient level analyses), the timeframe, sensitivity, validity, reliability and cost ([55](#_ENREF_55)). In addition, self-reported methods are known to be subject to systematic and recall error and ‘social desirability bias’ ([57](#_ENREF_57)). These factors need to be taken into account when deciding which method of dietary assessment is most appropriate for the proposed research ([58](#_ENREF_58)). Local public health teams need data relating to the diet and other health related behaviours of their local populations for health needs assessments, evaluating interventions, identifying ‘at risk’ sub-groups and monitoring local level targets, such as those enshrined in the Public Health Outcomes Framework ([54](#_ENREF_54)). Resource and time constraints mean some research methods are unfeasible and questionnaires using brief, self-report methods and proxy indicators of more complex behaviours are employed ([59](#_ENREF_59), [60](#_ENREF_60)).

* + - 1. *Weighed or estimated food diaries*

A weighed food diary is often considered to be the most accurate self-reported measurement available for assessing diet at an individual level ([55](#_ENREF_55)). When using a weighed food diary, participants in the study are instructed to weigh all foods and beverages consumed during a period of time, usually 4 to 7 days. This includes recording in detail the amounts of individual ingredients and raw foods used in composite dishes, cooking methods used as well as the total weight of the final dish. This method therefore requires a large commitment from participants. It can be expensive for a number of reasons. Standardised equipment (for example weighing scales) or financial incentives may need to be issued to participants, and specialised software is often needed to be employed in order to analyse the resulting data. Diaries, as with other self-reported methods of dietary assessment, may be subject to systematic under-reporting ([61](#_ENREF_61), [62](#_ENREF_62)). Additionally, individuals may change their usual eating patterns either to make the process of recording their intake simpler, or as a result of social desirability bias ([63](#_ENREF_63)). As a result of the high level or particular burden in using this method, there is evidence that recording can become increasingly inaccurate over time ([64](#_ENREF_64)).

In order to reduce the burden on the participant, estimated food diaries can be used. Using this methodology, participants provide detailed descriptions of all foods and beverages consumed, as with a weighed diary, except that portion size is estimated by the participant (or parent or guardian in children) perhaps using standard household measures such as cups, spoons or handfuls, sometimes with the help of supporting pictures or guidance ([55](#_ENREF_55)). This method is clearly more likely to produce recording errors and inconsistencies of measurement between study participants than weighed intake.

* + - 1. *Dietary recall*

Dietary recall methods are typically administered by an interviewer and can be carried out face-to-face or by telephone. The participant is usually asked to recall everything they have had to eat and drink in the previous 24-hour period ([55](#_ENREF_55)). Interviewers can ask probing questions in order to aid memory and improve the accuracy of the data collected. In more recent studies memory aids such as digital photographs and web-based tools have been successfully employed to improve accuracy of recall ([65](#_ENREF_65), [66](#_ENREF_66)). One advantage of the 24-hour recall method is that it reduces the likelihood of an individual altering their eating behaviours due to the measurement process itself, as with prospective methods.

This method has been employed for population level dietary surveillance and monitoring. For example, for the Low-Income Diet and Nutrition Survey (LIDNS), several 24-hour recalls were conducted over a period of time, ensuring that weekend days and week days are included ([67](#_ENREF_67)) to provide a measure of within-subject variability to help establish how accurately participants have reported their habitual diet. Some have suggested that compared with food frequency questionnaires and other self-report methods, this is an appropriate method for evaluating food-based interventions at a local level ([58](#_ENREF_58)). However, this method can be resource-intensive and requires significant time and expertise for analysing data.

Some studies have demonstrated that when compared with the ‘doubly labelled water’ method, an objective measure that estimates energy expenditure and thus energy intake in weight stable adults through the use of a trace isotope in water ([68](#_ENREF_68)), energy intake estimates from 24-hour recall methods are susceptible to significant under-reporting ([69](#_ENREF_69)). This effect may occur regardless of whether the recall method was administered by phone or face-to-face interview ([70](#_ENREF_70)). However, this level of under-reporting may be similar to that observed in other self-report methods ([71](#_ENREF_71)).

* + - 1. *Food Frequency Questionnaires (FFQs)*

FFQs aim to assess the frequency of the habitual consumption of particular foods or nutrients over a period of time ([55](#_ENREF_55)). FFQs have become one of the most widely used research tools in nutritional epidemiology ([72](#_ENREF_72)). The aim was to develop a rapid, less costly method for identifying the intake of certain foods, types and combinations of foods and other variables in the diet, to enable valid and reliable classification of particular dietary practices in a population ([73](#_ENREF_73), [74](#_ENREF_74)). FFQs consist of a checklist of one or more foods with options or categories for frequency of occasions consumed within a particular time frame. They vary in length, method of administration and level of detail depending on the nature of the study and the dietary outcomes being investigated.

A semi-systematic review of their design, validation and utilisation concluded that FFQs were most commonly used in cross-sectional surveys and the majority were designed to be disease specific ([72](#_ENREF_72)). The review also identified a number of aspects of design which improved accuracy including having pre-defined portion sizes; administering the FFQ by interview (either face-to-face or by telephone); and having a higher number of items, although this depended on the specific measurement objectives of the list. However, in the majority of studies reviewed, the FFQs were validated against other self-reported dietary assessment methods such as diaries, as opposed to more objective measures such as biomarkers ([72](#_ENREF_72)) which may provide a source of reporting bias. As with any self-reported dietary assessment method, FFQs are also subject to social desirability and recall bias ([55](#_ENREF_55)).

FFQs are commonly used in large scale, population level studies. They work on the assumption that habitual intake of a food or group of foods over a period of time is more important as an exposure than detailed accurate dietary intake data on specific days ([74](#_ENREF_74)). However, the choice of foods or food groups, units of time frequency, method of administration and length are key elements of the design of an FFQ, which need to be carefully considered when designing a new FFQ and may influence the accuracy of data they yield ([74](#_ENREF_74)).

*1.3.1.4. The use of new technologies in dietary assessment*

The use of new technologies such as the internet, mobile and smart phones, sensor technology and voice recognition and digital imaging software have become popular in the field of dietary assessment ([75](#_ENREF_75), [76](#_ENREF_76)). These technologies have been shown to improve accuracy of data collected in relation to several aspects of dietary assessment such as estimation of portion size ([77](#_ENREF_77)), analysis ([78](#_ENREF_78)), estimation of energy values for foods consumed ([79](#_ENREF_79)) and relative validity ([80](#_ENREF_80)). In particular, internet-based 24-hour recall methods have negated the need for resource-intensive, lengthy interviews and allowed for larger cohort studies ([81](#_ENREF_81)), and may ultimately prove to improve accuracy of recall. These developments may mean that internet-based 24-hour recall methods are appropriate and feasible for population level surveillance and monitoring.

* 1. **Population level dietary surveys in the UK**

This section will consider the existing sources of national level dietary surveillance in England that are government funded and available in the public domain, describing the methods employed for collection, the data that these surveys yield and their respective advantages and limitations.

* + 1. **The National Diet and Nutrition Survey**

The most detailed dataset regarding the diet of the UK population is the National Diet and Nutrition Survey (NDNS). It was originally a series of cross-sectional surveys of diet and nutritional status of the population ([82](#_ENREF_82)). The survey covered four age groups during different years: pre-school children in 1992-1993; older adults in 1994-1995; school-age children in 1997; and adults in 2000-2001. In April 2008, the NDNS changed to a rolling programme with data collected annually from a representative sample of approximately 500 adults (aged >19) and 500 children (>18 months old) ([83](#_ENREF_83)).

The NDNS has a number of elements that provide an estimate of food consumption, nutrient intake and derived nutritional status. Fasting blood samples are taken from a sub-sample (in years 2008-2011 this was 50% of adults aged 19-64 years and 38% of children aged 11-18 years) and analysed for a range of biomarkers or indicators of nutritional status ([84](#_ENREF_84)). Height and weight measurements and waist and hip circumferences are also taken from all adults, as well as urine samples.

Data on an individual’s habitual dietary behaviour are collected through face-to-face interviews and includes whether the individual has access to amenities for storage and preparation of food, is vegetarian, takes dietary supplements, is currently trying to lose weight or uses artificial sweeteners. Dietary intake is recorded using a diary of estimated (un-weighed) food over four consecutive days. Headline results are published annually by Public Health England (PHE). This includes data on food and nutrient intake such as: the amount of food consumed; energy and macronutrient intake (protein, total carbohydrate, NMES, total fat and saturated fat); vitamins and minerals (vitamin A, thiamin, riboflavin, niacin, vitamin B6, vitamin B12, folate, vitamin C, vitamin D, iron, calcium, magnesium, potassium, zinc, iodine and copper), alcohol; dietary supplement use and blood analyses (haematology, ferritin, C-reactive protein, water soluble vitamins and plasma total homocysteine, fat soluble vitamins and carotenoids, blood lipids, selenium and zinc)([84](#_ENREF_84)).

The NDNS provides robust, representative, detailed dietary intake data from representative samples of adults and children across the UK. This enables policy-makers to assess the overall diet of the population and identify where there may be areas for national-level intervention. As the sample size of the NDNS each year is relatively small, annual estimates at regional, local or sub-group level are not robust. However year-on-year data can be pooled to provide sufficient power for these types of analyses. In addition, despite the rigour of the data collection method, under-reporting in the survey is high ([85-87](#_ENREF_85)). Table 1.1 summarises the headline dietary intake data for the survey years 2012/13 to 2013/14.

* + 1. **The Health Survey for England**

The annual Health Survey for England (HSE) has been carried out since 1991 with the aim of collecting data to measure general health and health-related behaviours in private households in England. The survey collects data on several areas of health, some of which are collected consistently each year, and others which are additional ad hoc elements. The methodology combines self-reported data with interviewer led questions. A sub-sample receive a visit from a nurse to enable anthropometric and other measurements to be taken such as blood pressure and saliva tests ([88](#_ENREF_88), [89](#_ENREF_89)). The only dietary intake data that are collected biennially relate to fruit and vegetable intake. These data are gathered through a number of 24-hour recall questions. The sample size varies but is usually around 16,000 adults, including approximately 2,000 children.

The HSE provides a useful source of trend data for fruit and vegetable consumption as it has collected these data using a consistent methodology since 2001. The most recent estimates which are from 2014 ([90](#_ENREF_90)) are compared with other national level estimates of fruit and vegetable intake in Table 1.1. The dataset is also suitable for analysis of fruit and vegetable consumption by measures of socio-economic status (SES) such as Index of Multiple Deprivation (IMD) and household income ([91](#_ENREF_91)). This sample size is insufficiently powered to provide robust local estimates. Modelling methods have been used to generate ‘synthetic’ estimates of adult consumption of fruit and vegetables at local area level ([92](#_ENREF_92)).

* + 1. **The Living Costs and Food Survey (formerly Expenditure and Food Survey)**

The Living Costs and Food Survey (LCFS) collects self-reported data from a sample of private households in the UK of all food and drink purchases. This includes food eaten outside of the home, over a two week period, partially from receipts. Estimates of consumption are then calculated, which include a proportion for wastage. ‘Family Food’ annual reports are produced each year by the Department for Food and Rural Affairs (DEFRA), the most recent of which, ‘Family Food 2014’ contains data from 6,000 UK families ([93](#_ENREF_93)).

* + 1. **The Active People Survey and Public Health Outcomes Framework diet indicator**

The Active People Survey has been commissioned by Sport England since 2005 relating primarily to evaluate population levels of all types of physical activity. Self-reported data are collected annually from over 165000 people aged ≥14 years using a telephone survey. The survey allows for robust estimates to be made at local authority level ([94](#_ENREF_94)), and since 2014 data have been collected on the number of portions of fruit and vegetables consumed on a ‘usual’ day. These data are used by local authorities to annually report on the Public Health Outcomes Framework (PHOF) indicator for the ‘proportion of the (local authority) adult population meeting the recommended '5-a-day’ on a usual day’ ([95](#_ENREF_95)).

* 1. **Diet as a complex behaviour: the wider determinants of diet and their relevance in population level monitoring of diet.**

This section will summarise the evidence for the wider determinants of dietary intake, describe how these determinants affect eating habits and consider their relevance for population level dietary assessment and monitoring.

At individual level, dietary intake is not an unconstrained choice. It is a set of complex behaviours related to an individual’s biology, physiology and psychology (for example their metabolism, insulin sensitivity and fitness levels, emotion-driven relationships with food, disordered eating, preferences, attitudes) ([74](#_ENREF_74)) which occur within and are influenced by a social, economic and environmental context ([96](#_ENREF_96)). Social epidemiologists have stressed the importance of understanding health-related behaviours within the multi-dimensional environment within which they occur, the systematic interrelations and feedback loops between the layers of this physical and social environment and biological and physiological mechanisms, and how these affect risk factors and health outcomes ([97](#_ENREF_97)). Sociologists have also drawn on theories of ‘habitus’ (the ingrained habits and dispositions that we have as a result of our life experiences, which are not always conscious) to explain or describe the complex web of psycho-social structures influencing an individual’s choices relating to their health including their dietary intake. This social context brings different exposures at different stages of the life course ([98](#_ENREF_98)), the timing of these exposures may be relevant to their impact ([99](#_ENREF_99)) and their effects may be cumulative ([100](#_ENREF_100)). This ecological approach to explaining health-related behaviour can be usefully applied to dietary intake and is relevant to understanding dietary behaviour at a population level.

There is evidence that dietary intake, including dietary patterns associated with varying levels of diet quality, is influenced by a number of wider determinants of health including income ([101](#_ENREF_101)), education and occupation ([102](#_ENREF_102)), social mobility ([103](#_ENREF_103)), preference ([104](#_ENREF_104)), self-efficacy and perceived benefit ([105](#_ENREF_105)), knowledge ([106](#_ENREF_106)), self-control ([107](#_ENREF_107)), perceptions ([108](#_ENREF_108)), the wider social culture ([109](#_ENREF_109)), lifestyle behaviours such as smoking ([110](#_ENREF_110)), and cost and availability of healthy food ([108](#_ENREF_108)).

It is useful to consider whether measures relating to wider determinants of health may be valuable to include in a brief diet quality assessment tool. Inclusion of such variables in a construct intended as a proxy for diet quality may increase the sensitivity and validity of the tool; there may be some determinants that are stronger predictors of diet quality than foods or food groups. This may be helpful where the aim is to produce a tool with very few items and where the data may be easier to collect from participants and potentially less prone to bias than diet-related data, for example, occupation or education level.

* + 1. **Economic influences on diet.**

Income level and the cost of food are important influences on dietary intake. The term ‘food poverty’ can be used to describe the *‘inability to afford, or to have access to food to make up a healthy diet’* ([111](#_ENREF_111)), with low-income or unemployed families with dependent children being among the most at risk ([112](#_ENREF_112)). Data from sources of dietary intake data in the UK show that people on lower incomes spend less on their diets than those on higher incomes ([113](#_ENREF_113)), and that there is a positive correlation between levels of income and expenditure on fruit and vegetables ([114](#_ENREF_114)). Section 1.6.2 describes data relating to inequalities in dietary intake in the UK in more detail. Evidence from a longitudinal UK dataset suggests that ‘unhealthier’ foods, based on the Food Standards Agency UK Nutrient Profiling model for TV advertising to children ([115](#_ENREF_115)), have been consistently cheaper since 2002 than ‘healthy’ foods and that the gap between the cost is growing ([116](#_ENREF_116)). Data from the UK in 2007 showed that income and the cost of food are important determinants of dietary intake and that ‘healthy foods’ were perceived as expensive ([117](#_ENREF_117)). It has also been suggested that in the face of budgetary pressures other household expenditure such as fuel and rent, may be considered less flexible than the household food budget, as items such as fresh fruit and vegetables are replaced with cheaper alternatives ([118](#_ENREF_118)).

* + 1. **Social and cultural influences on diet.**

Qualitative studies have shown that the social environment, particularly partners, peers and family members have some influence on food choice ([119](#_ENREF_119), [120](#_ENREF_120)). In a survey of a sample of materially deprived older people, social isolation was an influencing factor, with those who ate alone less likely to eat a nutritionally adequate diet ([121](#_ENREF_121)). It is likely that a complex and mutually reinforcing relationship exists between an individual and their physical and social environments which influences their diet ([122](#_ENREF_122)).

There are variations in dietary patterns across ethnic groups. A systematic review of the nutritional composition of children’s diets found that compared with White Europeans, children from South Asian ethnic groups and Bangladeshi children reported higher mean total energy intakes. Black African and Black Caribbean children had lower fat intakes ([123](#_ENREF_123)). A qualitative study examining the eating practices of British Pakistanis and Indians with type 2 diabetes found that many respondents found it difficult to balance the risk of eating some less healthy South Asian foods against feelings of alienating themselves from their culture and community ([124](#_ENREF_124)).

* + 1. **Physical environmental influences on diet.**

Research looking at environmental influences on diet has been primarily concerned with physical access to, and availability of food and drink. This relates to the physical location of food outlets, including their relative accessibility by public transport or on foot. It also relates to whether outlets that are physically accessible provide a range of healthy and affordable foods. Studies from the US, Australia and New Zealand have shown that in many poorer neighbourhoods healthy food is either not available, unacceptable, inaccessible or unaffordable. This has not been conclusively shown in the UK ([125](#_ENREF_125), [126](#_ENREF_126)). However, the evidence in the UK is not conclusive. In relation to accessibility, two ‘natural experiments’ evaluating the impact of new large food retail stores in low-income communities have had differing results. One study found positive changes in fruit and vegetable consumption and the other did not ([127](#_ENREF_127), [128](#_ENREF_128)). Two further studies conducted in major urban centres in the UK found no independent association between the food environment (in terms of availability) and individual diet and fruit and vegetable consumption ([129](#_ENREF_129), [130](#_ENREF_130)).

The evidence supporting a relationship between area deprivation and diet in the UK is also mixed. For example, one large-scale, cross-sectional, population-based study has shown that residential area deprivation predicts fruit and vegetable consumption independently of individual educational level and occupational social class ([131](#_ENREF_131)). However, other evidence indicates a more complex picture. A recent qualitative study suggests that the relationship between area deprivation, food environment and dietary intake at individual level is inconsistent, and is influenced by multiple interrelated social, economic and cultural factors ([132](#_ENREF_132)).

* + 1. **Psycho-social influences on diet.**

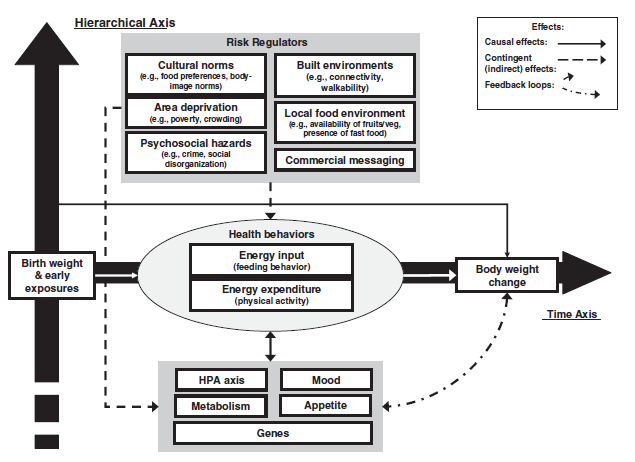
Dietary intake is influenced by a number of psycho-social factors such as preferences ([133](#_ENREF_133)), knowledge, perceived benefit and self-efficacy ([105](#_ENREF_105)), self-control ([107](#_ENREF_107)) and social norms ([134](#_ENREF_134)). In one qualitative study, whilst women of higher educational attainment were more likely to eat higher quality diets than those of lower educational attainment, perceived control, to some extent, mediated for these differences in observed dietary patterns ([135](#_ENREF_135)). However, it may be that educational attainment is a proxy for other psycho-social influences such as self-confidence or self-esteem. Another study showed that compared to women of higher educational attainment, women of lower educational attainment had less perceived control over their family's food choices, less support for attempts to eat healthily, fewer opportunities to observe and learn good food-related practices, a greater number of perceived environmental constraints and more ambiguous beliefs about the consequences of eating a nutritious diet for particular foods ([136](#_ENREF_136)). Cross-sectional studies in the US and UK have concluded that greater knowledge of nutrition and dietary guidelines increases the likelihood of consuming a diet which meets dietary guidelines, and that nutrition knowledge at least partially mediates for the association between socio-economic position and dietary intake ([120](#_ENREF_120), [137](#_ENREF_137), [138](#_ENREF_138)).

* + 1. **Conceptual models for the wider determinants of diet.**

Several authors have attempted to conceptualise the multiple determinants of food choice and dietary intake through models ([139-141](#_ENREF_139)). Figure 1.1 illustrates one such model which uses a social ecological approach to obesity and a visual explanation of what has been referred to as the ‘obesogenic’ environment ([23](#_ENREF_23)). In this model, the types of factors influencing dietary intake referred to above are described as mediating ‘risk regulators’. The authors define these as a category of variables that capture aspects of a social structure that are below the macro, large-scale, organisational levels of influence but that are nevertheless relatively stable features of a particular combination of social and physical environments. They are not risks in themselves in the sense of a single exposure that can be regulated or controlled for, but they nevertheless pose both day-to-day opportunities and constraints for the individual ([96](#_ENREF_96)). In the model, the effects of the ‘risk regulators’ are illustrated with dotted arrows and the ‘causal’ effects of biological and behavioural variables are illustrated with solid arrows.

In attempting to identify appropriate indicators of diet quality at population level, it is feasible that indicators relating to one or more of these ‘risk regulators’ are predictive of levels of diet quality.

**Figure 1.1. The ‘stream of causation’ model in relation to obesity (**[**96**](#_ENREF_96)**).**



* 1. **Population level dietary quality in the UK**

This section describes the current dietary recommendations in the UK for the prevention of diet–related disease and the promotion of health and current levels of adherence to these recommendations in adults in the UK.

* + 1. **Dietary recommendations in the UK**

The UK Scientific Advisory Committee on Nutrition (SACN) review evidence and advise the government regarding recommendations for the levels of foods and nutrients to be consumed for health. The ‘Nutritional Wellbeing of the British Population’ report ([142](#_ENREF_142)) presents recommendations for population level dietary health, some of which were updated in the recommendations for carbohydrates published in 2015 ([143](#_ENREF_143)). These recommendations are summarised in Table 1.1 and dietary intake data from NDNS 2012/13-2013/14, Family Food 2014 and HSE 2013 for comparison.

**Table 1.1. UK population level dietary recommendations and estimated intake from national diet surveys**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Food or macro-nutrient | Recommendation for adults (population average[[1]](#footnote-1),[[2]](#footnote-2)) | NDNS (adults aged 19-64 years[[3]](#footnote-3)) | Family Food[[4]](#footnote-4) | HSE[[5]](#footnote-5) (adults ≥16 years) |
| Fruit and vegetables (portions per day) | At least 5 (400 grams) | 4 portions | 3.9 portions | 3.5 portions |
| Oily fish | 1 portion per week (140 grams) | 54-87 grams | N/A | N/A |
| Total fat | ≤ 35% food energy | 34.2% | 38.4% | N/A |
| Saturated fatty acids | ≤ 11% food energy | 12.7% | 14.5% | N/A |
| Trans fatty acids | ≤ 2% food energy | 0.5% | N/A | N/A |
| NMES | ≤ 11% food energy | 12.3% | 13.1% | N/A |
| ‘Free’ sugars | ≤ 5% food energy | No data reported |  | N/A |
| Fibre (Non-starch polysaccharides  Englyst method) | ≥18 grams per day | 14 grams | 14.2 grams | N/A |
| Fibre (AOAC method) | ≥ 30 grams per day | No data reported | No data reported | N/A |

* + 1. **Inequalities, lifestyle and patterns in the diet of the UK population.**

As described in Section 1.5, patterns of dietary intake in the UK are not equally distributed and are patterned by wider determinants such as measures of socio-economic status and lifestyle factors.

* + - 1. *Measures of socio-economic status (SES)*

Data from the HSE 2013 show that consumption of portions of fruit and vegetables is patterned by income, with the lowest proportion of adults meeting the recommended levels of fruit and vegetable consumption coming from the lowest income group, and the highest proportion consuming the recommended levels coming from the highest income group ([146](#_ENREF_146)). This pattern is also observed in data from the NDNS 2008-2012 ([147](#_ENREF_147)) and the Family Food report 2014 ([93](#_ENREF_93)). Data from the NDNS 2008-2012 show that in adults aged 16-64 years there is little variance in the distribution of fat or saturated fat intake across the income groups. However, NMES and non-starch polysaccharide (NSP) consumption varies with income, with consumption decreasing as household income increases in NMES and the inverse in NSP ([146](#_ENREF_146)).

In 2005, a survey was undertaken specifically in a sample of the 15% most deprived households in the UK in order to examine the association between diet and income more closely ([117](#_ENREF_117)). The sample was based on an index of material deprivation in relation to food access and affordability based on factors such as receipt of benefits, household composition, car ownership and employment status. While these data are now over a decade old, they include important findings indicating the extent to which dietary quality is influenced by socio-economic measures. Compared with a sample from the general population, lower income groups were more likely to consume fat spreads and oils, processed meats, non-diet soft drinks, red meats, pizza, whole milk and table sugar.

*1.6.2.2. Smoking status*

Data from the HSE (2013) suggests that fruit and vegetable consumption is patterned by smoking status. The data show that 31.4% of those who reported never to have smoked consume the recommended five or more portions of fruit and vegetables per day compared with 28.3% of ex regular smokers and 15.5% of current smokers. Similarly, 14.4% of current smokers report consuming no portions of fruit and vegetables per day compared with 4.4% of ex regular smokers and 4.6% of those who reported to have never smoked ([148](#_ENREF_148)).

*1.6.2.3. Age*

Data from the 2013 Living Costs and Food Survey suggest that age is an important determinant of some aspects of dietary quality. Table 1.2 shows that incremental increases in age group, until the oldest groups of 70 years and over, bring increases in consumption of fruit and vegetables. The same incremental increases in age group also bring increases in consumption of sodium, percentage of energy from saturated fat (until the 70 years and over age group) and NMES.

**Table 1.2. Intake of key nutrients, fruit and vegetables by age group (2013 Living Costs and Food Survey)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age band (yrs) | Sodium per person per day | % Energy sat fat | % Energy from NMES | Fruit (grams) per person per day | Veg (grams) per person per day |
| 16-30 | 2.5 | 14.3 | 13.7% | 137 | 164 |
| 30-39 | 2.8 | 14.4 | 12.5% | 159 | 197 |
| 40-49 | 2.9 | 14.6 | 13.0% | 166 | 209 |
| 50-59 | 3.1 | 14.8 | 12.7% | 187 | 230 |
| 60-69 | 3.1 | 15.4 | 12.7% | 212 | 246 |
| 70-79 | 2.9 | 15.3 | 13.0% | 214 | 217 |
| ≥80 | 2.5 | 16.0 | 13.7% | 206 | 204 |

Which foods are the best proxies for dietary quality still needs to be elucidated, but clearly it is important to give due consideration to the influence of wider determinants of diet both as potential predictors of dietary intake and as population sub-groups for whom the diet quality assessment tool must be relevant.

* 1. **Dietary patterns**

This section describes how dietary patterns are analysed, their use as an exposure or outcome in nutritional epidemiology and why they are helpful compared with studies examining single nutrients or foods. The section provides an overview of the literature relating to dietary patterns analyses.

**1.7.1. Dietary patterns and their application in nutritional epidemiology**

Dietary intake varies from day to day, therefore it is the underlying consistent patterns that need to be identified for the purposes of epidemiology ([149](#_ENREF_149)). In recent years, nutritional epidemiology has begun to focus on ‘dietary patterns’ rather than consumption of macronutrients, micronutrients or foods. This method examines diet as a multidimensional exposure, examining relationships with the whole diet and health rather than the more ‘reductionist’ approach examining simply foods, food groups or nutrients ([150](#_ENREF_150)).

Whilst looking at single dietary components remains an important focus for nutritional epidemiological studies as particular nutrient and foods are correlated with disease risk and health, this has a number of limitations for certain types of investigation. Firstly, individuals do not purchase or consume dietary components, nutrients and, frequently, foods as single items in isolation. Individuals usually consume multiple nutrients in one food item and often consume a number of food items as part of a meal. It is likely that diseases such as cancer are influenced by multiple biochemical and physiological interactions between nutrients and foods. Secondly, other substances within food with no nutritive value, such as phytochemicals, may also be influential in disease development or protection ([151](#_ENREF_151), [152](#_ENREF_152)). Thirdly, biochemical and metabolic interactions in the body between micronutrients, dietary components and foods may complicate or confound attempts to identify associations between foods, nutrients and disease ([153](#_ENREF_153), [154](#_ENREF_154)). Indeed, conclusions about the relationship between a single nutrient, dietary component or food with a particular health outcome may be misleading ([155](#_ENREF_155)).

Investigation of dietary patterns can also be very useful for public health nutrition policy and research as they demonstrate the importance of looking at total diet in interventions as opposed to single food groups (e.g. fruit and vegetables), dietary components (e.g. fat) or micronutrients (e.g. exposure to particular vitamins). Studies have identified several types of ‘dietary pattern’ that are associated with positive health outcomes including prevention and control of diabetes type 2 ([156](#_ENREF_156)) and the prevention of cardiovascular diseases ([157](#_ENREF_157)), such as the ‘Mediterranean Diet’. Other patterns, such as high energy density diets, consisting of high sugar and high fat foods have been associated with obesity and diabetes ([16](#_ENREF_16)). Dietary patterns analyses have reiterated findings from studies that have shown associations between SES and diet with ‘healthier’ dietary patterns being associated with higher SES. ([158](#_ENREF_158), [159](#_ENREF_159)).

Analysis of dietary patterns can be undertaken in three ways: data-driven *(a posteriori),* theoretically driven *(a priori)* or a combination of both ([160](#_ENREF_160)).

* + 1. **Dietary pattern analyses: methods**

This section describes the methods that can be employed to analyse dietary patterns. They are described here to provide background to the overview of the dietary patterns analyses literature below. This section also provides a more detailed description of some of the methods used in the analyses in this thesis in Chapters 2-4.

* + - 1. *Empirically or data-driven methods (a posteriori)*

Data-driven dietary patterns are derived by reducing existing dietary intake data, collected from a sample, using statistical methods such as factor analysis (and Principal Component Analysis (PCA), which is a type of factor analysis) and cluster analysis. These are methods of representing or describing a number of initial variables by a smaller number of hypothetical variables ([161](#_ENREF_161)). Factor analysis and PCA reduce the data variables into patterns based on the inter-correlations between dietary items or variables. Cluster analysis reduces the data based on maximal differences in dietary intake patterns between groups of individuals ([160](#_ENREF_160)). Cluster analysis differs from factor analysis because instead of grouping ‘factors’ the method clusters individuals according to consistencies in their food intake. The aim is therefore to categorise individuals to a particular ‘cluster’ in which habitual food consumption is relatively homogenous. Dietary pattern differences within the ‘cluster’ should be relatively small. Whether factor analyses or cluster analyses are employed as the method for exploring dietary patterns will depend on specific objectives of the study ([162](#_ENREF_162)).

These types of statistical analyses have been used to investigate associations between dietary patterns and health outcomes such as hypertension ([163](#_ENREF_163)), CHD ([164](#_ENREF_164)), BMI and obesity ([165](#_ENREF_165)) and waist circumference ([166](#_ENREF_166)). However, such methods have been criticised by some authors as the dietary patterns generated by these types of analyses may not be stable over time, reproducible or generalisable to other populations ([153](#_ENREF_153), [167](#_ENREF_167), [168](#_ENREF_168)). These types of method for deriving dietary patterns have also yielded inconsistent results in terms of associations with health outcomes ([169](#_ENREF_169)) and to an extent, disregard *a priori* knowledge about diet and health outcomes. Some studies, however, have shown that data-derived patterns do reflect what might be expected *a priori* ([153](#_ENREF_153)). In addition, one systematic review concluded, from four validation studies, that dietary patterns derived from factor analyses showed moderate to good validity in relation to assessing adequacy of intake of a wide range of nutrients (vitamin B12 and vitamin E were the exceptions regardless of whether patterns were derived *a priori* or *a posteriori*) ([170](#_ENREF_170)).

Factor analyses methods are also sensitive to important decisions that have to be made regarding the definition of the variables, number of variables that are included in the analyses, how many factors are extracted, the method of rotation and the subjective labelling of patterns once they are identified ([153](#_ENREF_153)). In PCA studies of dietary intake the nature of the variables in terms of whether food variables are defined by the absolute weight consumed, energy adjusted weight or binary intake (consumed or not consumed) can have an influence on the nature of the dietary patterns generated and will have advantages and disadvantages depending on the objectives of the study ([171](#_ENREF_171)). Generally, the larger the number of variables, the smaller the proportion of the overall variation explained and the fewer the variables, the more variation is explained ([172](#_ENREF_172)). This may have implications for the reliability of the resulting ‘dietary patterns’ and the sensitivity of these patterns to variance in dietary intake in the population. The relative importance of the granularity of dietary intake data in the resulting patterns and the proportion of variance explained will be influenced by the outcome of interest and *a priori* knowledge. In a study where it is important to identify the impact of particular nutrients within an overall dietary pattern on an outcome, reducing the granularity of the diet variables included in the analyses to broad groups with a range of nutrient content such as ‘yoghurts and dairy desserts’ or ‘savoury snacks’ may reduce the sensitivity of the resulting patterns and potentially render them unhelpful in the context of the objective.

The definition of the variables will also depend on the dietary assessment method used and the variables available in the resulting dataset. Thus, a key area for future research is to test the impact of these decisions on the sensitivity, reproducibility and validity of the resulting patterns. Confirmatory factor analysis, for example, can be used to evaluate how robust and useful the derived patterns are in the context of the outcomes under examination. Also, because the patterns are usually derived from empirical data, collected from free living individuals, it is likely that patterns will vary depending on culture, socio-economic measures, ethnicity and other wider determinants. It is important that this is taken into consideration when results are interpreted and relationships between dietary patterns and health outcomes are reported.

* + - 1. *Theoretically driven methods (a priori)*

Theoretically driven dietary patterns involve developing a standard method of measuring adherence to a particular whole diet or dietary pattern that has been defined based on *a priori* nutritional knowledge and evidence, as opposed to measuring individual aspects of diet in a population such as consumption of one type of food or nutrient considered to be ‘healthy’ or ‘unhealthy’. The resulting composite measure is usually referred to as an ‘index’, ‘score’ or ‘indicator’. This method can be useful in ascertaining a population’s adherence to an *a priori* defined dietary pattern, such as ‘diet quality’ or ‘healthy eating’ as defined by dietary recommendations such as the U.S. Healthy Eating Index ([173](#_ENREF_173)) or a particular type of dietary pattern where epidemiological evidence shows an association with lower risk of a disease, for example, the Mediterranean dietary pattern ([174](#_ENREF_174)) or the Dietary Approaches to Stop Hypertension (DASH) diet. It can also be used to validate hypothesised associations between particular dietary patterns and health outcomes. Indices such as these have advantages in that they are based on existing knowledge of nutrition-health relationships and provide a clear benchmark ([175](#_ENREF_175)).

Whilst indices and scores such as these are useful tools in simplifying complex dietary intake patterns, a systematic review of predefined quality scores undertaken in 2007 concluded that none of those available at the time of review predicted morbidity or mortality significantly better than individual dietary factors ([176](#_ENREF_176)). Since this systematic review, a number of existing scores and indices have been refined and adapted and new indices have been developed that have been robustly validated and shown to be at least moderately inversely related to health outcomes such as CVD and cancer mortality ([177](#_ENREF_177)). As well as undergoing validation for health outcomes, several indices have also been validated, with positive results, using biomarkers to reflect levels of key nutrients or nutritional status. This provides evidence that there is a place for the use of these tools in epidemiological studies and population level dietary monitoring, but it is important to be clear about their validity and generalisability before adapting them for use in other studies ([177](#_ENREF_177)).

There are some methodological challenges in developing indices and scores such as these. The definition of dietary quality differs depending on the health outcomes and population under investigation. For example, there is some evidence, particularly from low and middle income countries, that have shown that ‘dietary diversity’ is associated with positive health outcomes and therefore provides an adequate proxy for ‘diet quality’ ([178](#_ENREF_178)); whereas in studies in some higher income countries this association is less clear ([179](#_ENREF_179)). The ratios, attributes or weightings of particular dietary components, foods or nutrients within an index differ depending on the particular focus and the evidence available at the time of development. There are decisions relating to the cut-off values and weighting of each element of the score for which the rationale may be somewhat subjective. In addition, indices developed from analyses of existing data depend on the context and characteristics of that sample, therefore results may not be applicable to other population groups ([155](#_ENREF_155)).

These challenges are reflected in the wider range of existing indices, the methods employed to develop them and variability in the quality and extent of their validity. Indices vary in the focus of investigation from assessing nutrient adequacy or adherence to broad healthy eating recommendations to preventing or screening for specific health outcomes. They are constructed based on intake of nutrients, foods, food groups or a combination ([155](#_ENREF_155)) and are frequently constructed to reflect existing dietary recommendations in the population of study ([176](#_ENREF_176)).

* + - 1. *Combined methods of dietary pattern analyses*

In response to the limitations of the solely empirically-derived or theoretically-derived methods of deriving dietary patterns, alternative data analyses methods have been proposed that combine an empirical approach with a priori knowledge. One such method is Reduced Rank Regression (RRR), which is a combined data reduction and prediction method that has been adopted and employed more recently in nutrition epidemiology ([17](#_ENREF_17)). The method includes both predictor and response variables ([161](#_ENREF_161)) and does not ignore *a priori* knowledge as the other data-driven methods do ([169](#_ENREF_169)). RRR enables the inclusion of both a hypothesis orientated and an exploratory approach and provides a more robust method for identifying patterns associated with disease by giving linear combinations that explain the maximum variation in the response variables ([169](#_ENREF_169)). RRR has been used to identify dietary patterns associated with obesity in children ([17](#_ENREF_17)) and associated with fat and bone mass in young children ([180](#_ENREF_180)). RRR is most useful in prospective cohort or longitudinal data. Not all studies, however, have found RRR to be the most effective methods for identifying dietary patterns associated with health outcomes ([181](#_ENREF_181)). A disadvantage of RRR is that it can only be carried out using specific statistical analysis software.

* + 1. **Empirically-derived dietary patterns studies in the UK: an overview**

The most recent systematic review of empirically-derived dietary patterns studies was published in 2004 ([160](#_ENREF_160)). Therefore a basic key word search of the literature to identify key papers relevant to the thesis published from 2004 onwards was undertaken of Medline in January 2016. The following key words were used in the search: ‘eating patterns’, ‘dietary patterns’, ‘food patterns’, ‘factor analysis and diet’, ‘cluster analysis and diet’. Abstracts and full papers were only extracted for potential inclusion in this summary if they were conducted in data from adult populations and were conducted in the UK and were available in English. This was not a systematic review and the papers described below are not a comprehensive representation of the existing literature. The review is intended to provide context for the methods employed in this thesis and an overview of the most recent literature with a focus on studies conducted in the UK. Where considered relevant or appropriate for context, papers from pre 2004 or in child or adolescent populations are described.

A wide range of dietary patterns analyses studies have been undertaken in the UK with many of the studies identified being in children and adolescents ([17](#_ENREF_17), [158](#_ENREF_158), [159](#_ENREF_159), [182-184](#_ENREF_182)). In the 2004 systematic review described above, six studies in the UK, two of which were in Northern Ireland, were identified ([160](#_ENREF_160)) which explored associations between data-driven dietary patterns and a range of outcomes including all-cause mortality ([185](#_ENREF_185)), nutrient biomarkers, ([186](#_ENREF_186)), BMI ([165](#_ENREF_165), [186-188](#_ENREF_186)) and metabolic syndrome ([189](#_ENREF_189)).

Studies in the UK from 2004 onwards identified through the rapid review had a similarly wide range of areas of investigation from simply defining the dietary patterns of a population, to examining associations with risk factors, markers of disease or long-term health outcomes, to examining the stability of diet over time and its relationship with wider determinants. Data from the Southampton Women’s Survey study group have been utilised for several dietary analyses studies. Analyses of dietary patterns comparing cluster analysis and PCA identified two similar patterns from each method, a ‘prudent’ and a ‘high-energy’ pattern ([190](#_ENREF_190)). The patterns identified were employed in a separate study to explore the associations between maternal diet and children’s dietary patterns ([191](#_ENREF_191)). Another study explored the stability in the dietary patterns of young women over a two year period and the wider factors influencing any changes in diet ([192](#_ENREF_192)).

Dietary patterns analyses of the NDNS 2000/2001 cohort data identified distinct dietary patterns that were associated with consumption of recommended levels of sodium, fat and sugars ([193](#_ENREF_193)). An exploration of data-driven dietary patterns and cardiovascular risk markers in the UK LIDNS identified four dietary patterns with a ‘fast food’ pattern characterised by high saturated fat foods explaining the majority of the variance and the ‘health aware’ pattern being inversely associated with markers of cardiovascular health ([194](#_ENREF_194)).

Many dietary patterns studies have been undertaken using data from the Avon Longitudinal Study of Parents and Children (ALSPAC). Findings reported from empirically driven dietary patterns analyses of this dataset have included: energy dense dietary patterns were associated with childhood obesity ([17](#_ENREF_17), [183](#_ENREF_183)); dietary patterns in children were associated with the child’s mother’s social, demographic and lifestyle characteristics ([158](#_ENREF_158)); dietary patterns of expectant mothers were socially patterned ([159](#_ENREF_159)) and were associated with nutrient intakes and adequately characterise dietary intake ([195](#_ENREF_195)); dietary patterns tend to be stable during childhood ([182](#_ENREF_182)); dietary patterns of men are associated with socio-economic measures and nutrients ([196](#_ENREF_196)); poor diet in early childhood may be associated with small reductions in IQ in later childhood ([197](#_ENREF_197)); dietary patterns are associated with educational attainment in children ([184](#_ENREF_184)); and dietary patterns of pregnant women with lifetime eating disorders were similar to those of healthy controls ([198](#_ENREF_198)). This dataset was also used to explore whether dietary patterns derived from PCA were stable in women at pregnancy and then four years post-partum ([199](#_ENREF_199)).

Longitudinal changes in dietary patterns during adult life from 1989-1999 were explored using factor analyses, which identified three dietary patterns for women and two for men. Scores for each pattern were found to change significantly over three time points in men and women ([200](#_ENREF_200)). Data from the British Prospective Cohort were analysed to examine the association between dietary patterns from childhood through to adulthood and breast density ([201](#_ENREF_201)). A study in a sample of adults aged ≥65 years explored data-driven dietary patterns to examine their association with mortality ([202](#_ENREF_202)). Dietary patterns in older adults were also explored in data from the Hertfordshire Cohort Study where PCA identified two dietary patterns, a ‘prudent’ and ‘traditional’ diet ([203](#_ENREF_203)). A study examining the association between dietary patterns and adult lung function and COPD concluded that a ‘prudent’ diet may protect against impaired lung function and COPD ([204](#_ENREF_204)).

Further dietary patterns analyses studies in UK cohorts could add to this body of evidence and provide further insight into UK dietary patterns and how these patterns are distributed in the population. In particular, there is a need for studies examining the association between data-driven dietary patterns and long term health outcomes (in prospective or longitudinal datasets) and biomarkers, nutrient intakes or theoretically defined diets (in cross sectional studies).

* + 1. **Theoretically-derived dietary patterns relating to diet quality: an overview**

This section provides an overview of literature relating to theoretically-derived dietary patterns relating to ‘diet quality’ or ‘healthy eating’. This method of dietary patterns analysis is used to develop standard composite measures, scores or indices, based on existing nutrition knowledge, against which the diets of individuals and populations can be assessed. There are only two such tools that have been developed in the UK for use in adults, therefore the majority of the literature in this area comes from the U.S. and Europe. A number of studies in the UK have adapted scores and indices developed in the U.S. and Europe to investigate associations between diet quality and health outcomes in the UK or to validate dietary assessment tools ([205-207](#_ENREF_205)).

* + - 1. *Diet quality scores and indices in the U.S and Europe*

A systematic review conducted in 2009 ([177](#_ENREF_177)) identified 25 indices of ‘diet quality’ that had been developed up to 2007 and included the results of a systematic review published in 1994 ([155](#_ENREF_155)). Studies were included where the theoretically defined dietary patterns or measure of diet quality were developed *a priori* and based on current nutritional evidence ([177](#_ENREF_177)), and excluded those indices developed *a posteriori* from food consumption data generated using statistical methods such as factor or cluster analysis. The review included ‘dietary variety’ in its definition of dietary quality. Tools specifically and exclusively intended to measure dietary diversity or variety were not included in the review of tools in this thesis, because the evidence for dietary diversity as an indicator of dietary quality or positive dietary related health outcomes is inconclusive, particularly in Organisation for Economic Cooperation and Development (OECD) countries ([179](#_ENREF_179), [208](#_ENREF_208)).

Some of the most frequently cited, revised and adapted indices of this nature include: the Healthy Eating Index (HEI) ([173](#_ENREF_173), [209](#_ENREF_209)) and the Diet Quality Index (DQI) ([210](#_ENREF_210), [211](#_ENREF_211)) both developed from data from the U.S; the Healthy Diet Indicator (HDI) ([212](#_ENREF_212)) developed from data from Italy, Finland and the Netherlands; and the Mediterranean Diet Score (MDS) developed from data from an elderly sample ([174](#_ENREF_174), [213](#_ENREF_213), [214](#_ENREF_214)). All four of these indices are made up of a combination of nutrient and food based parameters based on the appropriate dietary recommendations for that geographical area. In the case of the HDI, the indicator was based on the World Health Organisation’s (WHO) guidelines for the prevention of chronic diseases ([215](#_ENREF_215)) and includes between 8-10 items (adapted versions of the indices have added items).

The DQI was based on the U.S. dietary recommendations for reducing chronic disease risk published in 1989 ([216](#_ENREF_216)) and scores against a mixture of nutrients (total fat, saturated fat, cholesterol, complex carbohydrates, protein, sodium, calcium) and food (portions of fruit and/or vegetables) ([210](#_ENREF_210)). The tool has been shown to be predictive of short-term all-cause, all-circulatory disease and all-cancer mortality in a large prospective cohort study in US adults ([211](#_ENREF_211)), and be associated with smaller weight gain in US adults over time ([217](#_ENREF_217)). A revised version of the DQI (DQI-Revised) included fruit and vegetable consumption separately and included two items measuring ‘dietary diversity’ and ‘dietary moderation’. The latter was a composite measure with combined scoring criteria for ranges of intake of added sugar, discretionary fat, sodium and alcohol ([218-220](#_ENREF_218)).

The HDI includes a combination of nutrients (saturated fatty acids, polyunsaturated fatty acids, protein, complex carbohydrates, dietary fibre, monosaccharides and disaccharides, cholesterol) and foods (fruit and vegetables, pulses, nuts and seeds). The HDI has been shown to be inversely associated with mortality and lower cognitive impairment in the elderly ([212](#_ENREF_212), [221](#_ENREF_221), [222](#_ENREF_222)).

The HEI was developed to measure and monitor diet quality using a method that incorporated nutrient requirements and dietary guidelines into one measure ([173](#_ENREF_173)). Five of the scored components (grains, vegetables, fruits, milk, meat) are based on the 1992 US ‘Food Guide Pyramid’, and five (total fat, saturated fat, cholesterol, sodium, dietary variety defined by 16 kinds of food items over a three day period) are based on the 1990 Dietary Guidelines for Americans ([223](#_ENREF_223)). Studies have found the HEI to be positively associated with a wider list of micronutrients ([173](#_ENREF_173)), greater dietary variety, higher intakes of fruit, lower intakes of fat and saturated fat and higher plasma concentrations of carotene, lutein and vitamin C ([224](#_ENREF_224)) and be positively associated with biomarkers of fruit and vegetables intake ([225](#_ENREF_225)). The HEI has been revised and adapted, the latest iteration being the HEI-2010 which includes 9 adequacy components (total fruit, whole fruit, total vegetables, greens and beans, wholegrains, dairy, total protein foods, seafood and plant protein, fatty acids: a ratio of polyunsaturated and monounsaturated to saturated fatty acids) and 3 moderation components (refined grains, sodium, empty calories) ([226](#_ENREF_226)).

The MDS was originally developed as an 8 point index scoring key components of a traditional Mediterranean diet (vegetables, legumes, fruit and nuts, dairy products, cereals, meat and meat products, monounsaturated fats to polyunsaturated fats ratio, ethanol) adjusted for energy intake ([213](#_ENREF_213)). It was hypothesised *a priori* that those adhering more to the MDS would have beneficial health effects and indeed adherence to the overall Mediterranean dietary pattern was associated with reduced risk of mortality. An adapted MDS that included an item for fish was inversely associated with overall mortality and with death due to CHD and cancer ([174](#_ENREF_174), [214](#_ENREF_214)). Various adaptations and revisions of the MDS with items added and omitted have been considered ([222](#_ENREF_222)). A modified MDS (mMDS) adapted for a North American population including scores for items not associated with a Mediterranean style diet such as fast foods and desserts, showed that greater adherence to a Mediterranean-style dietary pattern was inversely associated with metabolic syndrome, LDL-cholesterol and reported weight gain, and was positively associated with higher HDL-cholesterol ([227](#_ENREF_227)).

* + - 1. *Diet quality scores and indices in the UK*

There are currently two tools that have been developed in the UK to assess diet quality and healthy eating in adults. Both tools were published during the development and writing of this thesis. A food and nutrient-based diet quality score (DQS) ([205](#_ENREF_205)) was derived from a 21 item Short Form Frequency Questionnaire (SFFFQ) and includes five dietary components reflecting the WHO dietary recommendations for prevention of chronic disease ([1](#_ENREF_1)). The SFFFQ is described in more detail in Section 1.8. The five components of the score were: fruit, vegetables, oily fish, fat and NMES. The score was used to evaluate the SFFFQ as a tool that could be used to assess diet quality at population level ([205](#_ENREF_205)). Whilst the DQS has been developed from an existing and widely used tool (the HDI), it has not been validated in a UK cohort to assess its validity as a measure of diet quality against a more robust or objective measure such as nutrient biomarkers, and does not reflect the broader aspects of the UK’s dietary recommendations ([205](#_ENREF_205), [228](#_ENREF_228)).

The Eating Choices Index (ECI) was developed in 2014 to discriminate between individuals who make healthy choices from those who do not for use in analysis of large surveys and was kept deliberately brief in number of items for ease of data collection ([229](#_ENREF_229)). The index was originally developed to represent diet quality in a study on diabetes and required further testing against more detailed dietary intake data in order to assess its validity as a representation of diet quality. The ECI includes four components: consumption of breakfast, consumption of two portions of fruit per day, type of milk consumed and type of bread consumed. The score is described as having been developed based on exploration of dietary intake data in the Medical Research Council National Survey of Health and Development (NSHD 1946 British birth cohort), experience, previous findings in the literature and UK dietary recommendations ([229](#_ENREF_229)). No further detail regarding the exploratory or developmental methods for this index is described.

Correlation analyses with intake of nutrient indicators defined as being associated with a healthy diet (fat, protein and carbohydrate as a proportion of energy intake, NSP, vitamin C, iron, calcium and folate) found that the ECI was significantly positively with all these nutrients (r=0.2-0.5; p<0.001) and negatively associated with fat (r=-0.2; p<0.001) which was as hypothesised. The ECI was not associated with total energy intake. Individuals who had a lower score were significantly more likely to be men, overweight or obese, have lower SES, smoke more and less physically active. The strength of this tool is that it is suitable for employment in large scale or practical settings. With only four items, the data are easy to collect and the nature of the items means that they do not require energy adjustment. However, the tool has not yet been validated or tested for its reliability as a questionnaire or in relation to health outcomes. The ECI has also not been validated against an objective measure such as nutrient biomarkers and was intended to discriminate between those who make broadly healthy and unhealthy choices as opposed to being a proxy measure reflective of the detail of UK DRVs and broader dietary recommendations ([229](#_ENREF_229)).

A nutritional scoring system for assessing diet quality in children aged ≤10 years in the UK was developed in 2012 ([230](#_ENREF_230)). Each child was allocated a score against a number of nutrients that were selected from the NDNS feedback letter to children. The nutrients included in the score were NMES, fibre, vitamin C, folate, calcium and iron. The scoring method was based on the UK Reference Nutrient Intakes (RNI) for children and therefore some nutrients varied with age. The maximum scores for all nutrients were weighted the same but a higher maximum score was given to NMES to align with the UK government’s current guidelines for reducing free sugar consumption in children ([230](#_ENREF_230)).

* 1. **Dietary assessment tools in the UK: an overview**

There are a number of dietary assessment tools available that have been developed and validated in the UK. Few of these focus on assessing diet quality. A 2015 systematic review of brief (<35 items) dietary questionnaires for clinical use in the prevention and management of obesity, CVD and type 2 diabetes identified 35 tools, 10 of which assessed healthy eating or healthy dietary patterns, two assessed adherence to the Mediterranean diet, 18 assessed dietary fat intake and 5 assessed fruit and vegetable intake ([231](#_ENREF_231)). The majority of the tools (n=20) were developed in North America and there was wide variability in study design, settings, validity testing and populations. Only two tools were developed in the UK, neither of which were intended to measure overall dietary quality or healthy eating and both of which are now over 20 years old ([232](#_ENREF_232), [233](#_ENREF_233)). The Heart Disease Prevention Project Screener (HDPPS) is a ten question tool, developed in 1981 with a mixture of FFQ and behaviour related questions and was developed and validated for the purpose of assessing levels of dietary fat intake alone, in men ([232](#_ENREF_232)). The tool was validated at group level only against data collected using a 3 day food diary (total score with grams of saturated fat r= -0.30, p<0.05, n=100, male only) and for reliability using a test re-test method (re-test time 3 to 4 months). The Dietary Instrument for Nutrition Education (DINE) tool is a 29 question FFQ developed in 1994 and validated against data collected from a 4 day weighed food diary (DINE fat score r=0.28 (% energy from total fat) to r= 0.57 (grams of saturated fat), DINE fibre score r=0.46 (grams of fibre), n=206), with the purpose of screening for levels of ‘high’ or ‘low’ fat and fibre intake ([233](#_ENREF_233)).

The authors of the review concluded that if tools are to be used in countries or for purposes that differ from those of their original development and evaluation, they need to be adapted or evaluated locally to ensure that they are reliable and valid ([231](#_ENREF_231)). As the tools described above were developed for the purposes of screening diets in clinical settings for high levels of fat and fibre, they are inappropriate for use in population level surveys or for measuring dietary quality or healthy eating patterns. One of the tools is 29 items which, despite being classified as a ‘short-form FFQ’ in academic terms, may be considered too long by public health researchers or practitioners to be practically feasible for inclusion in population level dietary surveillance or evaluation surveys. It is also likely that the questions in the tools are no longer culturally appropriate for the current dietary intake patterns of the UK population, given that they were developed in 1981 and 1994.

As referred to in Section 1.7.4.2, since the above systematic review was published, a 21 item short-form FFQ (SFFFQ) has been developed, based on a number of existing tools including the DINE tool ([233](#_ENREF_233)), tools developed outside the UK ([205](#_ENREF_205), [234](#_ENREF_234)) and expert knowledge which was used to tailor the SFFFQ for the UK population ([222](#_ENREF_222)). The SFFFQ was compared against a previously developed and validated 217-item FFQ ([235](#_ENREF_235)) and a 24 hour dietary recall ([205](#_ENREF_205)), and all 3 methods were compared with a DQS (described in Section 1.7.4.2.) reflecting the WHO dietary recommendations for prevention of chronic disease ([1](#_ENREF_1)). The results showed that a number of single items in the SFFFQ were predictive of a higher score against the DQS. The odds ratio (and 95% confidence intervals) of having a healthy diet were found to be 1.27 (1.09, 1.49) for an increase in fruit of 1 portion per day and were 0.33 (0.21, 0.53) for an increase in crisps of 1 portion per day ([205](#_ENREF_205)). The overall DQS showed fair agreement (k=0.38, p<0.001) between the SFFFQ and the FFQ but poor agreement between the SFFFQ and the diet recall (k=0.02-0.07) except for fruit intake where agreement was fair. The individual components of the DQS were all significantly correlated when comparing the SFFFQ with the FFQ but comparison of the SFFFQ with the dietary recall only showed significant correlations for fruit (p=0.02) and vegetable intake (p=0.02). Whilst the aim of the SFFFQ was to produce a tool that could be used in large population surveys to assess diet quality, it has a greater number of items than, for the purposes of this thesis, considered pragmatic for a brief tool for use by public health teams. In addition, the DQS was developed based on the HDI (as opposed to UK dietary recommendations) and was not validated against a more robust or objective measure of diet quality such as nutrient biomarkers.

As described in Section 1.4.4, the UK Active People Survey routinely collects data on fruit and vegetable consumption in adults for the purposes of reporting for the PHOF ([54](#_ENREF_54)). This indicator is used broadly by local authorities and public health bodies as an assessment of diet quality. Fruit and vegetables intake is frequently used as an indicator of diet quality in part due to the evidence of its association with positive health outcomes ([236](#_ENREF_236), [237](#_ENREF_237)). However, the two question tool used to collect these data, whilst adapted from a validated tool ([238](#_ENREF_238)), has not been validated in itself either as a measure of fruit and vegetable consumption or of diet quality ([54](#_ENREF_54)).

* 1. **Research needs to facilitate improved availability of dietary data in the UK to support evidence based practice and policy in public health**

This section will consider current research needs in the UK regarding the availability of dietary data and intelligence to support evidence based public health nutrition policy and practice at local and national level. It will also consider how the research questions and objectives in this thesis will contribute to these areas of research need.

*1.9.1. National dietary intake data that is robust at local level*

Section 1.4 described the availability and nature of dietary intake data at population level in England. There are a number of gaps in the availability of robust local level dietary intake data to support the work of local public health teams. As described previously, data on adult fruit and vegetables consumption are collected nationally to produce local level estimates for the PHOF through the Active People Survey. This is the only diet-related data that is collected nationally using a consistent data collection method and is of sufficient sample size to be robust at local area level. However, the data is limited and may not be a good indicator of diet quality or broader patterns of a healthy or unhealthy diet.

*1.9.2. A UK Diet Quality Score*

As described in Section 1.7.4 there is a paucity of research done in the UK to define ‘diet quality’ or to develop a standardised, food or nutrient-based score that reflects UK dietary recommendations, the equivalent of those available in other countries such as the HEI-2010 in the U.S ([226](#_ENREF_226)). Indeed, there has been no research conducted to develop and validate a score that has the aim of reflecting the full breadth of the UK government and SACN recommendations for a high quality diet ([228](#_ENREF_228), [239](#_ENREF_239)). This work is needed so that a standardised assessment can be made of levels of diet quality in the UK. A standardised measure enables behaviour change over time, between local geographies and across key population sub-groups to be monitored, compared and evaluated. A diet quality score would also enable researchers to validate dietary assessment tools against a valid, standardised composite measure of diet quality.

*1.9.3. Diet quality data by socio-economic measures*

The current availability of surveillance data in relation to socio-economic inequalities in diet is not comprehensive. The HSE reports annually on fruit and vegetable consumption and is sufficiently powered to allow for robust analysis by income group. However, evidence about what can be inferred from fruit and vegetable intake data in relation to diet quality is currently inconclusive. Data from the EFS can be used to provide estimates of intake of certain food groups by income level. However, data are not available at macro-nutrient or micro-nutrient level and are based on expenditure on food as opposed to consumption. The socio-economic measure reported for both surveys is income level. Income level is not the only socio-economic determinant of dietary intake and analysis of dietary patterns by other measures would contribute significantly to this area of research, potentially providing support for targeted interventions. In addition, neither of these data sources are large enough to allow for socio-economic inequalities within different regional or local geographies to be analysed.

*1.9.4. Brief, validated diet quality assessment tools for population level dietary surveillance and evaluation*

Increasing prevalence of obesity and other diet-related illnesses such as diabetes in adults and children has led to an increase in national and local level public health interventions to improve diet-related health in the UK. In order to evaluate whether these interventions are effective, there is a need for standardised outcome data collected using validated tools ([59](#_ENREF_59), [240](#_ENREF_240)). This is the case whether the aim is to monitor population level dietary change in relation to a national level intervention such as a sugar levy, or large scale campaign, or for a local, short-term intervention to improve the diets of low income families in a community. In clinical or primary care settings, specific aspects of diet may be of greater or lesser importance in relation to particular health outcomes. For use in public health, a standardised diet quality assessment tool, based on the key UK dietary recommendations, would enable changes in dietary patterns overall to be measured and monitored. Any such tools developed would require validity testing appropriate to their intended use. For example, tools for evaluating a short-term intervention would need to be tested for sensitivity to change over time as well as being practical to administer and low participant burden.

* + 1. *Objectives of this thesis that will support the research needs identified above.*

The following objectives will support the research needs described in the section above. A flow diagram showing the aims and objectives of the thesis is shown below in Figure 1.2.

1. Development and validation of a theoretically-derived UK Diet Quality Score.
2. Analysis of dietary patterns and their association with population characteristics, socio-economic measures, lifestyle factors and diet quality.
3. Identification of indicator variables and the development of brief tools for assessing diet quality for population level dietary surveillance and evaluation.

**Figure 1.2. Flow chart of the aims and objectives of this thesis**



**Chapter 2.**

**An exploration and examination of dietary patterns in UK adults using empirical methods (Study 1).**

**2.1. Introduction, aims and objectives**

The overarching aim of this study is to explore empirically-derived dietary patterns in a sample of UK adults; to evaluate their reliability; examine their association with wider determinants of health and nutrient biomarkers and assess their potential suitability as a method for generating indicator variables of diet quality to inform the development of a brief diet quality assessment tool.

During the development of the methodology for this first study within the thesis overall, public health researchers in the School of Health and Related Research at the University of Sheffield were seeking recommendations regarding the identification of a brief, validated tool, questionnaire or set of variables suitable for inclusion in the self-report survey for the second wave of the Yorkshire Health Study (YHS) to assess diet quality. The YHS survey had been developed by the University of Sheffield as part of a long-term health study in Yorkshire investigating lifestyle, use of health related services and long term health conditions ([241](#_ENREF_241)). Questions on diet had not been included in the first wave of the research. It was suggested that interim results from this thesis could inform the diet questionnaire in the YHS survey (YHS-DQ) and that this may also provide an opportunity for testing the external validity of the questions against an alternative dietary assessment method in a sub-sample of the YHS cohort in the future.

Whilst the interim results were agreed to be used for this purpose, it was acknowledged that due to the timescale for the second wave of the survey, the dietary questions would have to be included before some of the analyses and validity testing work would be completed. Therefore some of the analyses described in this chapter was undertaken after the YHS-DQ was finalised. In order to describe the stages of the analyses logically, the process by which the variables were selected and included in the YHS-DQ is described retrospectively as opposed to chronologically (see objective 5).

The research questions and objectives for this study are described below.

**Research questions**

1. Do distinct, empirically driven dietary patterns exist in UK adults and how are they characterised?
2. How reliable are these patterns and what are their associations with wider determinants of health (such as demographic, socio-economic and lifestyle variables) and nutrient biomarkers?
3. What are the strengths and limitations of using empirically-derived dietary patterns to inform a population level dietary intake questionnaire?

**Objectives**

1. To explore empirically-derived dietary patterns in a representative sample of UK adults.
2. To identify the key foods characterising each dietary pattern.
3. To explore the association of these dietary patterns with demographic and socio-economic characteristics, lifestyle factors and nutrient biomarkers.
4. To assess the internal consistency and stability of the dietary patterns.
5. To use the findings to inform the diet quality questions in the Yorkshire Health Study questionnaire.

This chapter will describe the methods employed in this study to answer the research questions and achieve the objectives as described above. The findings of these analyses will also be described. The results, strengths and limitations of the analyses and findings will be discussed in the context of the research questions.

**2.2. Methods**

*2.2.1. Dataset (Objectives 1-5)*

The dataset used for the analyses undertaken in Objectives 1 to 5 were from the UK NDNS Rolling Programme. As described in Section 1.4.1, the NDNS is an annual cross-sectional survey undertaken in the UK since 2008. Dietary intake data are gathered from a self-reported 3 or 4 day food diary. Blood plasma and urine are also collected from a consenting sub-sample from which biomarkers of nutritional status is derived. These blood and urine samples were taken approximately 8 weeks after the self-reported diary data were collected, however, as with the self-reported data, the biomarkers are, for the purposes of the analyses in this thesis, assumed to be representative of habitual dietary intake and nutrient status.

The NDNS contains data on more than 7000 foods which are aggregated into 59 ‘main’ food groups and then disaggregated from these ‘main’ categories into ‘subsidiary’ food groups. The ‘main’ food groups categorise all dietary intake in the dataset. Each food code has a value for 54 nutrients including energy, sugars, vitamins and minerals ([242](#_ENREF_242)). The food codes, food group names and nutrient content of foods is sourced from the UK Nutrient Databank ([243](#_ENREF_243)). The UK Nutrient Databank was originally developed in 1990 and is updated yearly according to the food supply. The nutrient data assigned to foods originate from a Department of Health programme of nutrient analysis, food manufacturers and information from food labels and recipe calculations for homemade dishes and for some manufactured products ([243](#_ENREF_243)).

All ‘main’ food groups from the NDNS dataset were included in the analyses excluding ‘Commercial Toddlers Foods and Drinks’ (which was irrelevant for the adult sub-sample). The data for the ‘main’ food group ‘Dietary sweeteners’ were not provided in the dataset downloaded from the UK Data Archive and were therefore also excluded from analyses. ‘Subsidiary’ food group variables for different types of spreads had already been disaggregated in the dataset and were also included in the analyses to reflect their varying composition and contribution to levels of consumption of different types of fatty acids. In addition, the ‘subsidiary’ food group variables from the ‘main’ food group ‘miscellaneous’ were included. This was due to the fact that the food group ‘miscellaneous’ covered foods and drinks that were varied in nature rendering the category as a whole meaningless in terms of overall nutritional or diet quality value. The ‘subsidiary’ variables in this category are: ‘dry weight beverages’ (drinks such as drinking chocolate, Ovaltine, milk shake powders), ‘soup homemade or manufactured’ and ‘savoury sauces, pickles, gravies and condiments’ ([242](#_ENREF_242)). Table 2.1 below lists the 60 pre-defined variables included in the analyses. Foods and food groups were included in all analyses in the thesis in average grams consumed daily by each individual as derived from the food diaries and available in the dataset.

Initially, data from years 1 to 3 (2008-2011) were obtained from the UK Data Service containing data from a representative sample of 1491 adults (aged ≥19 years, mean age 49 years). This dataset was used for the initial dietary patterns analyses. During the period of time during which this study was undertaken, an additional year of data from the NDNS was made available (2011-2012). Therefore a larger sample of 2083 adults in total from years 2008-2012 (aged ≥19 years, mean age 49 years) ([244](#_ENREF_244)) was available for analyses. The additional year of data was incorporated into the study sample so that it was as up-to-date and large a sample as possible. The methods and results are presented chronologically.

Consideration was given to the exclusion of ‘outlier’ responses or potential under-reporters of intake of nutrients and energy from the sample for analyses. In line with other studies in the NDNS, the full sample was included in the analyses ([83](#_ENREF_83)). This was firstly as it was not considered possible to separate under-reporters from under-consumers, those who were unwell, for example. Secondly, previous studies in this dataset have estimated levels of under-reporting in the NDNS to be significant ([87](#_ENREF_87), [147](#_ENREF_147)) therefore exclusion based on a population level estimate of physical activity level (as individual level data were not available) would be likely to lead to a significantly reduced sample size available for analysis. Under-reporting and nutrient intake for energy intake are given consideration at later stages of analyses.

**Table 2.1. List of NDNS pre-defined food variables included in the PCA**

|  |  |
| --- | --- |
| Food Group Variable (grams) |  |
| 1. 1% milk | 31. High fibre breakfast cereals |
| 1. Beef, veal and dishes | 32. Lamb and dishes |
| 1. Butter | 33. Liver products and dishes |
| 1. Other margarine, fats and oils | 34. Meat pies and pastries |
| 1. Other milk and cream | 35. Nuts and seeds |
| 1. PUFA margarine oils | 36. Oily fish |
| 1. Semi skimmed milk | 37. Other bread |
| 1. Whole milk | 38. Other meat and meat products |
| 1. Wholemeal bread | 39. Pasta rice and other cereals |
| 1. Ice cream | 40. Pork and dishes |
| 1. White fish coated or fried | 41. Salad and other raw vegetables |
| 1. Beer, lager, cider and perry | 42. Sausages |
| 1. Crisps and savoury snacks | 43. White bread |
| 1. Fruit juice including smoothies | 44. Yogurt, fromage frais and dairy desserts |
| 1. Soft drinks (low calorie) | 45. Other breakfast cereals |
| 1. Soft drinks (not low calorie) | 46. Other potatoes, potato salads and dishes |
| 1. Spirits and liqueurs | 47. Other white fish, shellfish, fish dishes |
| 1. Sugar confectionery | 48. Vegetables not raw |
| 1. Tea coffee and water | 49. Chicken and turkey dishes |
| 1. Wine | 50. Puddings |
| 1. Bacon and ham | 51. Brown granary and wheat germ bread |
| 1. Biscuits | 52. Skimmed milk |
| 1. Buns, cakes, pastries and fruit pies | 53. Sugars, preserves and sweet spreads |
| 1. Burgers and kebabs | 54. Dry weight beverages |
| 1. Cheese | 55. Low fat spread not polyunsaturated |
| 1. Chips (fried), roast potatoes and potato products | 56. Low fat spread polyunsaturated |
| 1. Chocolate confectionery | 57. Reduced fat spread not polyunsaturated |
| 1. Coated chicken and turkey | 58. Reduced fat spread polyunsaturated |
| 1. Eggs and egg dishes | 59. Savoury sauces, pickles, gravies condiments |
| 1. Fruit | 60. Soup homemade and retail |

*2.2.2. Statistical analysis (Objectives 1-5)*

*2.2.2.1. Empirically-derived dietary patterns analyses (Objective 1)*

All analyses were conducted in SPSS Version 22 statistical software package (with the exception of a pilot of the methodology on this dataset, which was conducted in STATA with the assistance of a Sheffield University Researcher in the early stages of the thesis development.)

As previously described in Section 1.7, a number of methods have been used widely in nutritional epidemiology to reveal the underlying patterns in the diet of a sample or population. These methods can be either theory or empirically driven or a mixture of both. Exploratory factor analysis and in particular, PCA which is a variation on exploratory factor analysis, have been used in many studies to reduce large sets of dietary variables into smaller sets of variables and to identify underlying ‘dietary patterns’ ([160](#_ENREF_160), [245](#_ENREF_245)). In this study, PCA is used to reduce detailed dietary intake data from the UK population, into dietary patterns to inform the development of a brief assessment tool for monitoring diet at population level. PCA was selected as the appropriate method for this thesis as this method uses the degree to which foods are correlated with each other to derive a new, smaller set of composite variables which are uncorrelated with each other and therefore can be considered as representative of discrete dietary patterns. The extracted sets of composite variables are ‘components’ and the variables within them are ‘factors’.

The data-driven dietary patterns generated are characterised by high and low consumption of particular foods and drinks. Each individual in the sample is scored to assess how similar their dietary intake is to each pattern that emerges, therefore it is also possible to explore which sample characteristics and other variables in a dataset are associated positively and negatively with each pattern. This can be used to broadly assess the generalisability and content validity of the patterns in terms of whether observed associations with sample characteristics, lifestyle factors and socio-economic variables are as might by hypothesised based on the existing literature regarding patterns of diet in the UK.

SPSS standardises the food amounts in the analysis to account for any exaggerated influence of foods with larger variances due to their nature or the subjectively, predefined foods and food groups in the variable list (for example white bread compared with 1% milk).

Prior to conducting the PCA, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett’s test of sphericity[[6]](#footnote-6) were undertaken to ensure that the data were suitable for factor analysis. (KMO=0.5 which reached the acceptable limit of 0.5 ([246](#_ENREF_246)) and Bartlett’s test of sphericity was significant indicating that correlations between items were sufficiently large for PCA).

The 60 pre-defined food variables (average grams consumed per day) shown in Table 2.1 above were entered into the PCA. Orthogonal rotational (varimax) was applied. The purpose of applying the rotation is that it redistributes the explained variance for each component and thus achieves a simpler structure ([159](#_ENREF_159)). Results for both the rotated and un-rotated solutions are shown in Tables 2.3a and 2.3b.

PCA will produce as many components as there are variables, but the first few will explain the largest proportion of the total variance. In this analysis, the number of components was subjectively chosen based on both a visual assessment of the scree plot (Figure 2.1) and those with eigenvalues above the values of approximately 1.5 in order to identify the fewest number of patterns that explained the largest proportions of variance. The scree plot displays the order of the components against their eigenvalues (eigenvalues are calculated as the sum of a component's squared factor loadings for all the variables and measures the variance in all the variables that is accounted for by that factor) ([246](#_ENREF_246)). If there is a noticeable ‘elbow’ in the plot, then this may be used to select the number of PCs for further analysis (the number selected is usually one less than the point at which the elbow occurs) ([245](#_ENREF_245)).

*2.2.2.2.**Key foods characterising the dietary patterns generated (Objective 2)*

Each dietary pattern was characterised by positive and negative associations with particular foods and food groups. All foods and foods groups have an associated factor loading with each component. For the purposes of this analyses only those factors (or food variables) with an association of ≥0.2 or ≤-0.2 were retained and reported in the results as they were the variables that most characterised each component. Foods and food groups with associations of ≥0.25 or ≤-0.25 were considered ‘moderate’ and ≥0.3 or ≤-0.3 were considered ‘strong’ in line with previous studies ([158](#_ENREF_158), [159](#_ENREF_159), [247](#_ENREF_247)).

*2.2.2.3. Associations with wider determinants of health (Objective 3)*

SPSS generated a score for each individual in the dataset which represents the extent to which their individual food choices deviate from each component (or ‘dietary pattern’). Regression analysis was conducted to identify the associations between the dietary patterns and several wider determinants of health: key demographics, socio-economic variables and lifestyle factors to see whether the variables were associated with each pattern in the direction that would be expected. The following continuous variables were included in the model: age; household income (standardised to adjust for the number of individuals in the household) and BMI (kg/m2) and the following categorical variables were included in the model: sex (male/female), ethnicity (white/non-white), National Statistics Socio-economic Classifications (NSSEC)*[[7]](#footnote-7)* and smoking status (smoker/non-smoker).

Plots of residuals were visually assessed for evidence of homoscedasticity, constancy of variance and outliers (Figures B.1.1-B.1.12, Appendix B). Where variables were not normally distributed, regressions including the square of the independent variables were carried out to test for evidence of curvilinearity. Where the coefficient of the squared term was significant (thus indicating curvilinearity), the coefficients for both the value and the squared values were plotted to visually assess for potential curvilinear (quadratic) effects (Figure C.1.1, Appendix C).

*2.2.2.4. Internal consistency and stability of the dietary patterns (Objective 4)*

As an updated NDNS dataset including another year of dietary intake data were made available during the study period, it was possible to test the reproducibility of these patterns in a larger and more up-to-date version of the same dataset (2008-2012). The consistency and stability of the dietary patterns generated was tested through repeating the PCA in the 2008-2012 dataset and in a random 50% split of this dataset. The dietary patterns generated from the PCA of each of these datasets and the foods that characterised them were compared for consistency. Associations between the dietary patterns generated from the 2008-2012 PCA with wider determinants of health were also analysed for an assessment of consistency with the 2008-2011 patterns using methods as described in 2.2.2.3 (Figures B.2.1-B.2.12, Appendix B for relevant residual plots and Figure C.1.2, Appendix C for quadratic plot).

*2.2.2.5. Associations with nutrient intakes nutrient biomarkers (Objective 3)*

Regression analyses were used to examine the associations between each dietary pattern generated by the PCA of the 2008-2012 dataset and nutrient intake, derived from self-reported dietary diaries, and nutrient status, derived from biomarkers from blood plasma and urine samples. As described in 2.2.1 the biomarkers are assumed to be representative of the usual nutrient status of an individual.

*2.2.3. Identification of items for inclusion in the YHS Diet Questionnaire (YHS-DQ) (Objective 5)*

The aim of this objective is to explore whether the food variables that characterised the dietary patterns are useful indicators of diet quality for inclusion in the YHS questionnaire. As described in the introduction to this chapter, researchers at the University of Sheffield had requested that the interim results from this section of the thesis might be used to inform diet quality questions in the survey used for the second wave of the Yorkshire Health Study. The method for selecting items for inclusion in the YHS-DQ was influenced by a number of pragmatic constraints. For example, in addition to a tight timescale which did not allow for the completion of all the validation analyses described in this chapter, there was also limited space in the survey for the questions therefore allowing for a tool with a maximum of approximately 12-13 items only.

The initial method for developing the YHS-DQ was to include all foods with strong positive or negative (≥0.3 or ≤-0.3) factor loadings for each of the four dietary patterns generated by the 2008-2011 PCA with the rationale being that they would be representative of the overall dietary pattern. This method was then modified to take account of the timescales involved, the requirement for a brief tool with as few items as possible, participant cognition and food groups that were clear (for the benefit of the participants) and easy to define in terms of diet quality (for the purposes of analysis.)

The following rationale was applied to item selection:

1. At least one food or food group that is strongly associated (positively or negatively) with each dietary pattern was included. The reason for this was to select foods which ‘typified’ certain dietary patterns and discriminated between these, rather than selecting foods eaten in large quantities by everyone.
2. Foods that are already highlighted by UK dietary recommendations as those that should be eaten more or less frequently at population level were prioritised (e.g. oily fish, fruit, vegetables, foods clearly high in sugar, saturated fat or salt).
3. Where a food or food group was strongly associated with more than one pattern and the association was in the same direction with each pattern the food was excluded (for example, chocolate confectionary).
4. Some of the foods and food group variables (as defined in the NDNS dataset) were not useful from an epidemiological perspective as they contained very broad ranges of foods or drinks that could not easily be categorised or assessed in terms of diet quality, nutrient density or associated health impacts. On this basis, the following categories were excluded: ‘tea, coffee and water’; ‘soft drinks (low calorie’, ‘yoghurt, fromage frais and dairy desserts’, ‘other potatoes, potato salads and dishes’, ‘nuts and seeds’ and ‘savoury sauces, pickles, gravies and condiments’.
5. Foods that were strongly positively associated with one pattern and strongly negatively associated with another pattern were prioritised.
6. For the purposes of clarity and ensuring participant cognition, ‘reduced fat spread, not polyunsaturated’ was excluded, ‘chips (fried), roast potatoes and potato products’ was re-defined as simply ‘chips’ and ‘soft drinks (not low calorie)’ was defined as ‘sugary drinks (fizzy pop, squash)’.

**2.3. Results**

*2.3.1. Sample characteristics*

Sample characteristics for both datasets are described in Table 2.2. In both NDNS datasets (2008-2011 and 2008-2012) there were significantly more females than males and white than non-white participants (p<0.05). The spread of proportions of the sample in each NSSEC grouping is representative of the UK as reported in the General Household Survey. ([147](#_ENREF_147))

**Table 2.2. Sample characteristics NDNS 2008-2011 and 2008-2012**

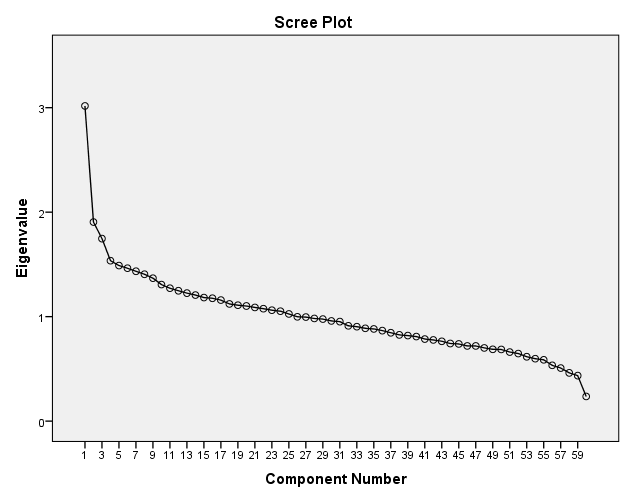
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2008 - 2011 | | | | 2008 - 2012 | | | |
| Characteristics | n | % | | Mean (SD) | n | | % | Mean  (SD) |
|  | 1491 |  | |  | 2083 | |  |  |
| Sex |  |  | |  |  | |  |  |
| Males | 650 | 43.6 | |  | 901 | | 43.3 |  |
| Females | 841 | 56.4 | |  | 1182 | | 56.7 |  |
| Age (years) | 1491 |  | | 49  (17) | 2083 | |  | 49  (17) |
| Ethnicity |  |  | |  |  | |  |  |
| White | 1364 | 91.5 | |  | 1877 | | 91.5 |  |
| Non-white | 127 | 8.5 | |  | 168 | | 8.5 |  |
| NSSEC Group |  |  | |  |  | |  |  |
| Higher managerial and professional | 219 | 14.7 | |  | 309 | | 14.9 |  |
| Lower managerial and professional | 403 | | 27 |  | | 553 | 26.6 |  |
| Intermediate occupations | 135 | | 9.1 |  | | 191 | 9.2 |  |
| Small employers | 161 | | 10.8 |  | | 222 | 10.7 |  |
| Low supervisory and technical  Occupations | 155 | | 10.4 |  | | 203 | 9.8 |  |
| Semi-routine occupations | 195 | 13.1 | |  | 289 | | 13.9 |  |
| Routine occupations | 164 | 11 | |  | 235 | | 11.3 |  |
| Never worked | 33 | 2.2 | |  | 43 | | 2.1 |  |
| Smoking status |  |  | |  |  | |  |  |
| Smoker | 371 | 24.9 | |  | 488 | | 24 |  |
| Non-smoker | 1120 | 75.1 | |  | 1557 | | 76 |  |
| Equivalised household income (£) | 1275 |  | | 32288  (23963) | 1777 | |  | 32035  (23570) |
| BMI (kg/M2) | 1365 |  | | 27.5  (5.4) | 1902 | |  | 27.7  (5.4) |
| Total energy intake diet only (kcal) | 1491 |  | | 1804  (580) | 2083 | |  | 1803  (575) |
|  |  |  | |  |  | |  |  |

*2.3.2. PCA results 2008-2011 NDNS data (Objective 1)*

The scree plot (Figure 2.1) indicated inflexions provided a rationale for retaining 3 or 4 components as those that explained the largest proportions of variance in the dataset. As the inflexions in the scree plot after the fourth component were all very close together the first four were retained for reporting which explained 13.7% of the total variance (4%, 3.9%, 3% and 2.7% respectively).

*2.3.3. Characterisation of dietary patterns (Objective 2)*

Tables 2.3a shows the rotated solution of the PCA with the food variables and factor loadings for each principal component and Table 2.3b shows the un-rotated solution. Factor loadings of ≥0.3/≤-0.3 were classified as ‘strong’ and factor loadings between ±0.25-0.3 were ‘moderate’. Table 2.4 below shows the foods that most strongly characterised each dietary pattern and the subjective label given to each pattern.

**Figure 2.1. Scree plot for PCA of 60 food variables (NDNS 2008-2011)**

The first component was labelled ‘Fruit, vegetables and oily fish’ as the food variables with the highest factors loadings were ‘fruit’, ‘salad and other raw vegetables’, ‘yoghurt, fromage frais and dairy desserts’ and ‘oily fish’. The food variables with the lowest factor loadings were ‘white bread’ and ‘chips’. The second component was labelled ‘snacks, fast food and fizzy drinks’ as the foods with the highest factor loadings were ‘crisps and savoury snacks’, ‘coated chicken and turkey’ and ‘soft drinks (not low calorie)’. The foods with the lowest factor loadings were ‘tea, coffee and water’ and ‘high fibre breakfast cereals’. The third component was labelled ‘sugary foods and dairy’ as the food variables with the highest factor loadings were ‘sugar, preserves and sweet spreads’, ‘buns, cakes, pastries and fruit pies’, ‘biscuits’ and ‘semi skimmed milk’. The fourth component was labelled ‘meat, potatoes and beer’ as the food variables with the highest factor loadings were ‘savoury sauces and gravies’, ‘other potatoes and dishes’, ‘beer, lager and cider’, ‘pork and dishes’ and ‘white bread’.

**Table 2.3a. Rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2011).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Food/food group | Principal Component (Variance explained) | | | |
| 1 (4%) | 2 (3.9%) | 3 (3%) | 4 (2.7%) |
| Fruit (g) | 0.60 | -0.25 | 0.17 | 0.003 |
| Salad and other raw vegetables (g) | 0.54 | -0.11 | 0.04 | 0.07 |
| Yogurt, fromage frais and dairy desserts (g) | 0.42 | -0.08 | 0.17 | -0.08 |
| Oily fish (g) | 0.40 | -0.16 | -0.04 | 0.05 |
| White bread (g) | -0.37 | 0.28 | 0.28 | 0.33 |
| Nuts and seeds | 0.31 | 0.05 | 0.13 | 0.06 |
| Vegetables not raw (g) | 0.28 | -0.25 | 0.03 | 0.24 |
| Whole milk (g) | -0.27 | -0.14 | 0.11 | 0.27 |
| Wholemeal bread (g) | 0.26 | -0.19 | 0.12 | -0.02 |
| White fish coated or fried (g) | -0.26 | -0.09 | 0.02 | 0.004 |
| Fruit juice including smoothies | 0.24 | 0.08 | 0.02 | 0.12 |
| Skimmed milk (g) | 0.22 | -0.04 | -0.01 | 0.01 |
| Other white fish, shellfish and fish dishes (g) | 0.21 | -0.08 | -0.08 | -0.04 |
| Meat pies and pastries (g) | -0.20 | 0.10 | 0.08 | 0.10 |
| Other bread (g) | 0.20 | 0.04 | 0.02 | 0.02 |
| Other milk and cream (g) | 0.16 | 0.002 | -0.02 | 0.05 |
| Brown, granary and wheat germ bread | 0.13 | 0.01 | 0.06 | -0.04 |
| Other margarine, fats and oils (g) | 0.13 | -0.05 | 0.10 | 0.12 |
| Low fat spread not polyunsaturated (g) | 0.06 | -0.02 | -0.06 | 0.05 |
| Soft drinks not low calorie (g) | -0.02 | 0.53 | -0.03 | 0.13 |
| Crisps and savoury snacks (g) | -0.003 | 0.52 | 0.21 | -0.001 |
| Coated chicken and turkey (g) | -0.002 | 0.40 | -0.11 | -0.02 |
| Soft drinks low calorie (g) | 0.06 | 0.40 | -0.05 | -0.05 |
| Burgers and kebabs (g) | -0.09 | 0.39 | -0.04 | 0.00 |
| Chips, fried and roast potatoes and potato products (g) | -0.32 | 0.33 | 0.06 | 0.18 |
| High fibre breakfast cereals (g) | 0.13 | -0.30 | 0.13 | -0.09 |
| Sugar confectionery (g) | 0.15 | 0.28 | 0.06 | -0.05 |
| Pasta, rice and other cereals (g) | 0.22 | 0.23 | -0.10 | -0.14 |
| Soup homemade and retail (g) | 0.05 | -0.19 | 0.05 | -0.07 |
| Chicken and turkey dishes (g) | 0.15 | 0.19 | -0.07 | -0.10 |
| Beef, veal and dishes (g) | -0.04 | 0.12 | 0.00 | -0.12 |
| Lamb and dishes (g) | 0.01 | -0.08 | -0.05 | 0.05 |
| PUFA margarine and oils (g) | 0.07 | -0.07 | 0.05 | 0.04 |
| One percent milk (g) | 0.01 | -0.02 | 0.00 | 0.00 |
| Sugar, preserves and sweet spreads (g) | -0.29 | -0.10 | 0.45 | 0.18 |
| Buns, cakes, pastries and fruit pies (g) | 0.10 | -0.11 | 0.42 | 0.07 |
| Biscuits (g) | 0.11 | 0.11 | 0.40 | -0.24 |
| Chocolate confectionery (g) | 0.10 | 0.37 | 0.38 | -0.07 |
| Semi skimmed milk (g) | 0.01 | -0.05 | 0.37 | -0.20 |
| Reduced fat spread not polyunsaturated (g) | -0.17 | 0.19 | 0.35 | 0.13 |
| Tea, coffee and water (g) | 0.28 | -0.31 | 0.34 | 0.01 |
| Cheese (g) | 0.10 | 0.03 | 0.29 | 0.10 |
| Puddings (g) | -0.02 | -0.10 | 0.29 | 0.03 |
| Ice cream (g) | 0.09 | 0.02 | 0.17 | 0.03 |
| Reduced fat spread polyunsaturated (g) | -0.11 | -0.13 | 0.15 | -0.08 |
| Other breakfast cereals (g) | -0.06 | -0.03 | 0.12 | -0.05 |
| Low fat spread polyunsaturated (g) | 0.03 | -0.05 | 0.06 | -0.01 |
| Sauces, pickles and gravies (g)' | 0.12 | 0.06 | 0.01 | 0.40 |
| Other potatoes, potato salads and dishes (g) | 0.07 | -0.28 | 0.11 | 0.38 |
| Beer, lager, cider and perry (g) | -0.10 | 0.22 | -0.05 | 0.37 |
| Pork and dishes (g) | -0.01 | -0.01 | -0.04 | 0.32 |
| Bacon and ham (g) | -0.04 | 0.16 | 0.23 | 0.30 |
| Spirits and liqueurs (g) | 0.03 | 0.18 | -0.18 | 0.27 |
| Wine (g) | 0.26 | 0.04 | -0.23 | 0.27 |
| Butter (g) | -0.07 | -0.09 | 0.06 | 0.27 |
| Sausages (g) | -0.13 | 0.05 | 0.10 | 0.23 |
| Other meat and meat products (g) | -0.003 | 0.01 | 0.02 | 0.22 |
| Eggs and egg dishes (g) | 0.09 | -0.08 | -0.05 | 0.21 |
| Liver and dishes (g) | 0.002 | -0.09 | -0.03 | 0.19 |
| Dry weight beverages (g) | -0.04 | -0.05 | 0.05 | -0.15 |

**Table 2.3b. Un-rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2011).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Food/food group | Principal Component (Variance explained) | | | |
| 1 (4%) | 2 (3.9%) | 3 (3%) | 4 (2.7%) |
| Fruit (g) | 0.61 | 0.06 | 0.27 | 0.05 |
| White bread (g) | -0.48 | 0.39 | 0.09 | 0.07 |
| Chips, fried and roast potatoes and potato products (g) | -0.48 | 0.12 | 0.07 | 0.06 |
| Salad and other raw vegetables (g) | 0.46 | -0.03 | 0.27 | 0.16 |
| Tea, coffee and water (g) | 0.43 | 0.29 | 0.09 | -0.10 |
| Soft drinks not low calorie (g) | -0.39 | -0.11 | 0.34 | 0.13 |
| Oily fish (g) | 0.38 | -0.05 | 0.12 | 0.14 |
| Yogurt, fromage frais and dairy desserts (g) | 0.38 | 0.01 | 0.26 | -0.06 |
| Vegetables not raw (g) | 0.33 | 0.15 | 0.03 | 0.25 |
| Burgers and kebabs (g) | -0.33 | -0.12 | 0.19 | 0.002 |
| Wholemeal bread (g) | 0.33 | 0.07 | 0.08 | -0.02 |
| High fibre breakfast cereals (g) | 0.32 | 0.11 | -0.07 | -0.12 |
| Coated chicken and turkey (g) | -0.28 | -0.21 | 0.23 | 0.04 |
| Meat pies and pastries (g) | -0.22 | 0.14 | -0.01 | 0.004 |
| Other white fish, shellfish and fish dishes (g) | 0.21 | -0.12 | 0.04 | 0.05 |
| Skimmed milk (g) | 0.18 | -0.05 | 0.10 | 0.06 |
| Soup homemade and retail (g) | 0.18 | 0.04 | -0.08 | -0.07 |
| Other milk and cream (g) | 0.10 | -0.03 | 0.09 | 0.09 |
| PUFA margarine and oils (g) | 0.10 | 0.06 | 0.02 | 0.02 |
| Low fat spread polyunsaturated (g) | 0.06 | 0.05 | 0.01 | -0.03 |
| One percent milk (g) | 0.02 | 0.003 | -0.01 | 0.00 |
| Sugar, preserves and sweet spreads (g) | -0.14 | 0.54 | -0.05 | -0.13 |
| Buns, cakes, pastries and fruit pies (g) | 0.16 | 0.37 | 0.16 | -0.13 |
| Other potatoes, potato salads and dishes (g) | 0.19 | 0.34 | -0.08 | 0.28 |
| Whole milk (g) | -0.13 | 0.33 | -0.20 | 0.11 |
| Reduced fat spread not polyunsaturated (g) | -0.24 | 0.32 | 0.18 | -0.10 |
| Bacon and ham (g) | -0.17 | 0.30 | 0.19 | 0.14 |
| Puddings (g) | 0.07 | 0.27 | 0.04 | -0.12 |
| Pasta, rice and other cereals (g) | 0.02 | -0.27 | 0.25 | -0.02 |
| Cheese (g) | 0.06 | 0.25 | 0.20 | -0.03 |
| Butter (g) | -0.03 | 0.22 | -0.06 | 0.18 |
| Sausages (g) | -0.15 | 0.21 | 0.01 | 0.12 |
| Chicken and turkey dishes (g) | -0.01 | -0.19 | 0.18 | -0.02 |
| Other margarine, fats and oils (g) | 0.12 | 0.12 | 0.09 | 0.07 |
| Chocolate confectionery (g) | -0.14 | 0.13 | 0.46 | -0.22 |
| Crisps and savoury snacks (g) | -0.34 | 0.01 | 0.44 | -0.09 |
| Sugar confectionery (g) | -0.07 | -0.09 | 0.30 | -0.04 |
| Soft drinks low calorie (g) | -0.22 | -0.19 | 0.28 | 0.00 |
| Nuts and seeds | 0.20 | 0.04 | 0.28 | 0.05 |
| Fruit juice including smoothies | 0.10 | -0.01 | 0.21 | 0.15 |
| White fish coated or fried (g) | -0.12 | 0.10 | -0.21 | -0.06 |
| Other bread (g) | 0.11 | -0.04 | 0.16 | 0.05 |
| Ice cream (g) | 0.06 | 0.12 | 0.14 | -0.03 |
| Brown, granary and wheat germ bread | 0.10 | -0.01 | 0.10 | -0.04 |
| Wine (g) | 0.11 | -0.11 | 0.11 | 0.39 |
| Biscuits (g) | 0.07 | 0.13 | 0.29 | -0.37 |
| Sauces, pickles and gravies (g)' | -0.01 | 0.17 | 0.15 | 0.36 |
| Semi skimmed milk (g) | 0.09 | 0.20 | 0.11 | -0.35 |
| Spirits and liqueurs (g) | -0.15 | -0.06 | 0.09 | 0.33 |
| Beer, lager, cider and perry (g) | -0.27 | 0.11 | 0.09 | 0.32 |
| Pork and dishes (g) | -0.04 | 0.14 | -0.01 | 0.29 |
| Eggs and egg dishes (g) | 0.08 | 0.07 | -0.004 | 0.22 |
| Liver and dishes (g) | 0.03 | 0.10 | -0.06 | 0.18 |
| Other meat and meat products (g) | -0.04 | 0.12 | 0.03 | 0.18 |
| Reduced fat spread polyunsaturated (g) | 0.03 | 0.14 | -0.10 | -0.17 |
| Dry weight beverages (g) | 0.03 | -0.02 | -0.05 | -0.16 |
| Other breakfast cereals (g) | -0.01 | 0.09 | -0.02 | -0.11 |
| Beef, veal and dishes (g) | -0.09 | -0.09 | 0.05 | -0.10 |
| Low fat spread not polyunsaturated (g) | 0.05 | -0.03 | 0.004 | 0.09 |
| Lamb and dishes (g) | 0.05 | 0.01 | -0.06 | 0.06 |

**Table 2.4. Foods characterising each dietary pattern (NDNS 2008-2011).**

|  |  |  |
| --- | --- | --- |
| ‘Dietary Pattern’ | Foods with moderate/strong positive factor loadings\* (≥0.25) | Foods with moderate/strong negative factor loadings (≤-0.25)\* |
| ‘Fruit, vegetables and oily fish (FVOF)’ | Fruit  Salad and other raw vegetables  Yoghurt, fromage frais and dairy deserts  Oily fish  Nuts and seeds  Wine  Vegetables not raw  Tea, coffee and water  Wholemeal bread | Chips (fried), roast potatoes and potato products  White bread  Sugar, preserves and sweet spreads  Whole milk  White fish coated or fried |
| ‘Snacks, fast food and fizzy drinks (SFFFD)’ | Crisps and savoury snacks  Coated chicken and turkey  Soft drinks (not low calorie)  Soft drinks (low cal)  Burgers and kebabs  Chocolate confectionery  Chips (fried), roast potatoes and potato products  Sugar confectionery  White bread | Tea, coffee and water  High fibre breakfast cereals  Other potatoes, potato salads and dishes  Fruit  Vegetables not raw |
| ‘Sugary food and dairy (SFD)’ | Sugars, preserves and sweet spreads  Buns, cakes, pastries and fruit pies  Biscuits  Chocolate confectionery  Semi skimmed milk  Reduced fat spread not PUFA  Tea, coffee and water  Cheese  Puddings  White bread | Wine |
| ‘Meat, potatoes and beer (MPB)’ | Savoury sauces, pickles, gravies and condiments  Other potatoes, potato salads and dishes  Beer, lager, cider and perry  White bread  Pork and dishes  Bacon and ham  Whole milk  Spirits and liqueurs  Butter |  |

\*Foods are listed in descending order of factor loading value.

*2.3.4. Main effects between dietary patterns and sample characteristics, socio-economic measures and lifestyle factors (objective 3)*

Main effects regression models were used to explore adjusted associations between each dietary pattern and sample characteristics (age, sex, ethnicity), socio-economic variables (household income and NSSEC) and lifestyle factors (smoking status and BMI). A visual assessment of residual plots concluded that the conditions for the model were satisfied (see Figures B.1.1-B.1.12, Appendix B). As BMI and household income were not normally distributed, regressions with the values and the squared term were undertaken with each dietary pattern. The results showed evidence of potential non-linearity for household income with the ‘FVOF’ dietary pattern. A quadratic plot demonstrated this was not the case and that the relationship was linear (see Figure C.1.1).

Table 2.5 below shows the results of these analyses. The results reflected patterns as might be hypothesised based on estimates from other national level data and literature as described in Section 1.6.2. For all analyses the level of significance was set at p<0.05. People scoring higher on (and were therefore more likely to consume) the FVOF dietary pattern were significantly more likely to be older (0.002) and less likely to be male (-0.15), white (-0.28), smokers (-0.34) and be in routine occupations (-0.53). People scoring higher on the SFFFD dietary pattern were significantly more likely to be white (0.21), smokers (0.14) and male (0.22). This pattern was inversely associated with age (-0.03) and positively associated with BMI (0.10). Both the SFD and the MPB dietary patterns were significantly positively associated with age (0.04, 0.07 respectively), being white (0.37, 0.53 respectively) and being male (0.57, 0.36 respectively). The SFD pattern was significantly positively associated with being a smoker (0.21) whereas the MPB patterns was significantly, negatively associated with this lifestyle factor (-0.17). The MPB pattern was also negatively associated with BMI (-0.02).

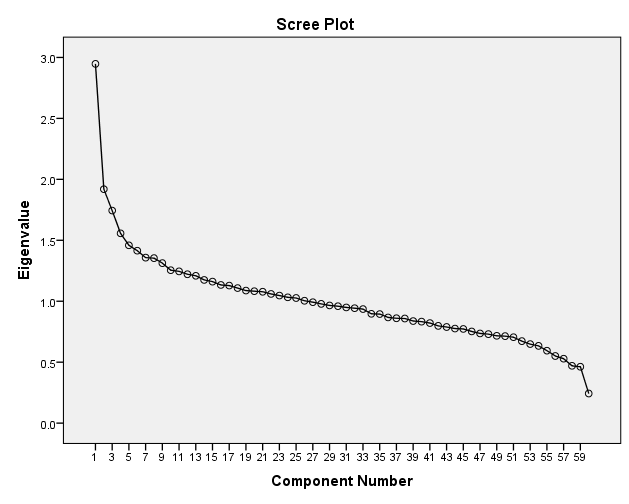
**Table 2.5. Main effects regression model (95% confidence intervals) of each dietary pattern for sociodemographic characteristics and lifestyle variables (NDNS 2008-2011)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample characteristic (n) | Fruit, veg & oily fish | | Snacks, fast food and fizzy drinks | | Sugary foods and dairy | | Meat, potatoes and beer | |
|  | Coefficient (95% CI) | P-Val | Coefficient (95% CI) | P-Val | Coefficient (95% CI) | P-Val | Coefficient (95% CI) | P-Val |
| *Sex* |  | | | | | | | |
| Male (523) | -0.15  (-0.26, -0.05) | <0.01 | 0.22 | <0.001 | 0.57 (0.46, 0.68) | <0.001 | 0.36 (0.25, 0.47) | <0.001 |
| Female (665) | *Reference* | | | | | | | |
| *Age (1491)* | 0.002  (-0.001, 0.005) | <0.001 | -0.28 (-0.031,  -0.025) | <0.001 | 0.004 (0.001, 0.008) | 0.12 | 0.007 (0.004, 0.01) | <0.001 |
| *Ethnicity* |  | | | | | | | |
| White (1091) | -0.28 (-0.48,  -0.08) | <0.01 | 0.21 (0.13, 0.32) | 0.02 | 0.37 (0.16, 0.57) | <0.001 | 0.53 (0.33. 0.73) | <0.001 |
| Non-white (97) | *Reference* | | | | | | | |
| *NSSEC Group* | | |  |  |  |  |  |  |
| Higher managerial and professional (188) | 0.18 (-0.32, 0.67) | 0.49 | 0.11  (-0.34, 0.56) | 0.64 | -0.10  (-0.62, 0.42) | 0.70 | 0.28 (-0.23, 0.79) | 0.29 |
| Lower managerial and professional (325) | 0.10 (-0.38,  0.59) | 0.68 | 0.13 (-0.31, 0.57) | 0.57 | -0.80 (-0.58, 0.42) | 0.76 | 0.10 (-0.40, 0.60) | 0.71 |
| Intermediate occupations (107) | -0.15 (-0.66, 0.35) | 0.56 | 0.12 (-0.34, 0.58) | 0.62 | -0.14 (-0.66, 0.38) | 0.60 | 0.09 (-0.43, 0.61) | 0.74 |
| Small employers (121) | -0.12 (-0.62, 0.35) | 0.65 | -0.05 (-0.50, 0.41) | 0.83 | 0.14 (-0.38, 0.66) | 0.59 | 0.11(-0.41, 0.63) | 0.68 |
| Low supervisory and technical occupations (126) | -0.21 (-0.71, 0.29) | 0.41 | 0.21 (-0.43, 0.48) | 0.93 | -0.08 (-0.59, 0.44) | 0.77 | 0.20 (-0.32, 0.71) | 0.45 |
| Semi-routine occupations (156) | -0.23 (-0.72, 0.26) | 0.36 | 0.10 (-0.35, 0.54) | 0.68 | 0.001 (-0.51, 0.51) | 1.00 | 0.11 (-0.40, 0.62) | 0.67 |
| Routine occupations (123) | -0.53 (-1.03,  -0.03) | 0.04 | 0.15 (-0.30, 0.60) | 0.52 | 0.16 (-0.36, 0.68) | 0.54 | 0.12 (-0.50, 0.53) | 0.95 |
| Never worked (27) | *Reference* | | | | | | | |
| *Smoking status* |  | | | | | | | |
| Smoker (287) | -0.34 (-0.47, -0.21) | <0.001\* | 0.14 (0.03-0.39) | 0.02 | 0.21 (0.07, 0.34) | 0.003 | -0.17 (-0.30,  -0.04) | 0.01 |
| Non-smoker (901) | *Reference* | | | | | | | |
| *Household income (1275)* | 7.44E-6 (4.94E-6, 9.936E-6) | <0.001 | 1.04E-7 (-2.17E-6, 2.375E-6) | 0.93 | -2.07E-7  (-2.38E-6, 2.79E-6) | 0.02 | -2.96E-6  (-5.53E-6, -3.84E-7) | 0.88 |
| *BMI (1365)* | | |  |  |  |  |  |  |
|  | 0.03 (-0.006, 0.13) | 0.5 | 0.10  (0.002,0.02) | 0.02 | 0.002  (-0.008, 0.012) | 0.69 | -0.02  (-0.03, -0.01) | <0.001 |

*2.3.5. Evaluating internal consistency and stability of the dietary patterns (Objective 4)*

* + - 1. *PCA in NDNS 2008-2012 dataset*

Prior to conducting PCA, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett’s test of sphericity was undertaken to verify the sampling adequacy for the analysis. KMO=0.5 which reaches the acceptable limit of 0.5 and Bartlett’s test of sphericity was significant indicating that correlations between items were sufficiently large for PCA ([246](#_ENREF_246)). PCA was conducted on the NDNS 2008-2012 dataset to compare the empirically-derived dietary patterns with those identified from 2008 to 2011. Four components with eigenvalues of above 1.5 were identified (Figure 2.2 below).

**Figure 2.2. Scree plot for PCA of 60 food variables (NDNS 2008-2012)**

Similarly to the analysis of years 1 to 3 of data, four components explained 13.6% of the total variance (3.9%, 3.7%, 3.1% and 2.8% respectively). Tables 2.6a and 2.6b below show the rotated and un-rotated solutions respectively for the PCA. Factor loadings of ≥0.3/≤-0.3 were classified as ‘strong’ and factor loadings between ±0.25-0.3 were ‘moderate’. Each component was again subjectively labelled according to the food variables most strongly characterising it (Table 2.7 below).

**Table 2.6a. Rotated PCA solution: food variables (g) and factor loadings for each Principal Component (NDNS 2008-2012).**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Food variables (average g/day) | Principal Component (variance explained) | | | | |
|  | PC 1 (3.9%) | PC 2 (3.7%) | PC 3 (3.1%) | PC 4 (2.8%) | |
| Soft drinks not low calorie | 0.51 | 0.03 | 0.10 | 0.03 |
| Crisps and savoury snacks | 0.48 | -0.01 | 0.00 | 0.24 |
| Soft drinks low calorie | 0.42 | 0.11 | -0.07 | 0.07 |
| Coated chicken and turkey | 0.40 | -0.02 | -0.07 | 0.06 |
| Tea, coffee and water | -0.38 | 0.20 | 0.05 | 0.25 |
| Burgers and kebabs | 0.37 | -0.08 | 0.004 | -0.01 |
| High fibre breakfast cereals | -0.35 | 0.09 | -0.05 | 0.14 |
| Chips, fried and roast potatoes and potato products | 0.32 | -0.28 | 0.25 | 0.07 |
| Wholemeal bread | -0.28 | 0.16 | -0.08 | 0.16 |
| Sugar confectionery | 0.27 | 0.18 | -0.05 | 0.24 |
| Pasta, rice and other cereals | 0.26 | 0.19 | -0.20 | -0.05 |
| Soup homemade and retail | -0.20 | 0.06 | -0.04 | 0.02 |
| Chicken and turkey dishes | 0.17 | 0.14 | -0.12 | -0.09 |
| Low fat spread polyunsaturated | -0.06 | 0.03 | 0.03 | 0.03 |
| Fruit | -0.30 | 0.56 | -0.12 | 0.19 |
| Salad and other raw vegetables | -0.14 | 0.51 | -0.05 | 0.02 |
| Yogurt, fromage frais and dairy desserts | -0.13 | 0.39 | -0.13 | 0.22 |
| Oily fish | -0.19 | 0.38 | -0.01 | -0.06 |
| Vegetables not raw | -0.23 | 0.34 | 0.22 | 0.00 |
| Fruit juice including smoothies | 0.07 | 0.27 | 0.11 | 0.00 |
| Nuts and seeds | 0.02 | 0.27 | -0.06 | 0.13 |
| Skimmed milk | -0.03 | 0.25 | -0.06 | 0.02 |
| White fish coated or fried | -0.07 | -0.25 | 0.04 | 0.06 |
| Other margarine, fats and oils | -0.04 | 0.21 | 0.09 | 0.03 |
| Other bread | 0.02 | 0.21 | 0.05 | -0.05 |
| Other white fish, shellfish and fish dishes | -0.06 | 0.20 | -0.12 | -0.09 |
| Other milk and cream | 0.03 | 0.19 | 0.04 | -0.04 |
| Brown, granary and wheat germ bread | 0.01 | 0.15 | -0.03 | 0.00 |
| PUFA margarine and oils | -0.05 | 0.09 | 0.04 | 0.02 |
| Other potatoes, potato salads and dishes | -0.23 | 0.14 | 0.47 | 0.03 |
| Sauces, pickles and gravies | 0.06 | 0.18 | 0.47 | -0.05 |
| White bread | 0.30 | -0.27 | 0.44 | 0.13 |
| Butter | -0.09 | -0.02 | 0.35 | 0.04 |
| Sugar, preserves and sweet spreads | -0.10 | -0.26 | 0.34 | 0.29 |
| Bacon and ham | 0.18 | 0.05 | 0.34 | 0.09 |
| Beer, lager, cider and perry | 0.22 | -0.07 | 0.29 | -0.16 |
| Meat pies and pastries | 0.09 | -0.14 | 0.28 | 0.04 |
| Sausages | 0.06 | -0.11 | 0.27 | 0.03 |
| Whole milk | -0.07 | -0.18 | 0.25 | -0.03 |
| Pork and dishes | -0.01 | 0.03 | 0.21 | -0.05 |
| Eggs and egg dishes | -0.07 | 0.16 | 0.20 | -0.15 |
| Other meat and meat products | 0.02 | 0.05 | 0.18 | -0.02 |
| Beef, veal and dishes | 0.06 | -0.10 | -0.15 | -0.004 |
| Liver and dishes | -0.08 | -0.002 | 0.10 | -0.04 |
| Low fat spread not polyunsaturated | -0.01 | 0.03 | -0.03 | -0.03 |
| Biscuits | 0.04 | 0.04 | -0.13 | 0.47 |
| Semi skimmed milk | -0.17 | -0.12 | -0.03 | 0.43 |
| Chocolate confectionery | 0.35 | 0.13 | 0.02 | 0.40 |
| Buns, cakes, pastries and fruit pies | -0.12 | 0.11 | 0.19 | 0.38 |
| Wine | 0.03 | 0.28 | 0.13 | -0.31 |
| Puddings | -0.13 | 0.01 | 0.12 | 0.28 |
| Cheese | 0.01 | 0.15 | 0.12 | 0.28 |
| Ice cream | 0.03 | 0.07 | -0.002 | 0.24 |
| Reduced fat spread not polyunsaturated | 0.12 | -0.14 | 0.19 | 0.21 |
| Spirits and liqueurs | 0.17 | 0.08 | 0.15 | -0.20 |
| Reduced fat spread polyunsaturated | -0.08 | -0.13 | 0.01 | 0.16 |
| Other breakfast cereals | 0.001 | -0.04 | 0.001 | 0.15 |
| Dry weight beverages | -0.07 | -0.04 | -0.06 | 0.09 |
| Lamb and dishes | -0.02 | -0.02 | 0.03 | -0.08 |

**Table 2.6b. Un-rotated PCA solution: food variables and factor loadings for each Principal Component (NDNS 2008-2012).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Food variables (average g/day) | Principal Component (variance explained) | | | |
|  | PC 1 (3.9%) | PC 2 (3.7%) | PC 3 (3.1%) | PC 4 (2.8%) |
| Fruit | 0.62 | 0.05 | 0.26 | 0.02 |
| White bread | -0.50 | 0.34 | 0.10 | 0.07 |
| Chips, fried and roast potatoes and potato products | -0.47 | 0.15 | 0.06 | 0.01 |
| Salad and other raw vegetables | 0.45 | -0.03 | 0.23 | 0.17 |
| Tea, coffee and water | 0.40 | 0.28 | 0.03 | -0.11 |
| Yogurt, fromage frais and dairy desserts | 0.39 | 0.01 | 0.27 | -0.06 |
| Oily fish | 0.38 | -0.002 | 0.08 | 0.18 |
| Soft drinks not low calorie | -0.37 | -0.08 | 0.34 | 0.13 |
| Wholemeal bread | 0.34 | 0.09 | 0.01 | -0.12 |
| Vegetables not raw | 0.33 | 0.22 | 0.07 | 0.24 |
| High fibre breakfast cereals | 0.33 | 0.14 | -0.08 | -0.14 |
| Burgers and kebabs | -0.31 | -0.12 | 0.16 | 0.03 |
| Beer, lager, cider and perry | -0.29 | 0.08 | 0.004 | 0.26 |
| Coated chicken and turkey | -0.27 | -0.17 | 0.26 | -0.02 |
| Meat pies and pastries | -0.23 | 0.22 | -0.01 | 0.07 |
| Skimmed milk | 0.20 | -0.05 | 0.14 | 0.06 |
| Other white fish, shellfish and fish dishes | 0.20 | -0.13 | 0.04 | 0.08 |
| Soup homemade and retail | 0.19 | 0.04 | -0.07 | -0.04 |
| Other margarine, fats and oils | 0.14 | 0.08 | 0.12 | 0.12 |
| Brown, granary and wheat germ bread | 0.10 | -0.04 | 0.10 | 0.05 |
| PUFA margarine and oils | 0.09 | 0.05 | 0.03 | 0.04 |
| Sugar, preserves and sweet spreads | -0.17 | 0.48 | -0.06 | -0.15 |
| Other potatoes, potato salads and dishes | 0.13 | 0.46 | -0.03 | 0.26 |
| Buns, cakes, pastries and fruit pies | 0.12 | 0.37 | 0.19 | -0.13 |
| Butter | -0.04 | 0.33 | -0.04 | 0.14 |
| Pasta, rice and other cereals | -0.01 | -0.30 | 0.24 | 0.06 |
| Puddings | 0.08 | 0.28 | 0.08 | -0.15 |
| Bacon and ham | -0.18 | 0.25 | 0.18 | 0.17 |
| Reduced fat spread not polyunsaturated (g) | -0.21 | 0.23 | 0.10 | -0.09 |
| Whole milk (g) | -0.14 | 0.22 | -0.16 | 0.06 |
| Sausages (g) | -0.18 | 0.22 | -0.02 | 0.09 |
| Chicken and turkey dishes (g) | 0.00 | -0.21 | 0.14 | 0.09 |
| Beef, veal and dishes (g) | -0.07 | -0.13 | -0.03 | -0.11 |
| Liver and dishes (g) | 0.02 | 0.09 | -0.07 | 0.07 |
| Low fat spread polyunsaturated (g) | 0.05 | 0.06 | 0.003 | 0.004 |
| Low fat spread not polyunsaturated (g) | 0.03 | -0.04 | -0.004 | 0.01 |
| One percent milk (g) | 0.02 | -0.02 | -0.02 | 0.02 |
| Chocolate confectionery (g) | -0.14 | 0.08 | 0.50 | -0.16 |
| Crisps and savoury snacks (g) | -0.33 | -0.05 | 0.41 | -0.10 |
| Sugar confectionery (g) | -0.04 | -0.04 | 0.40 | -0.07 |
| Soft drinks low calorie (g) | -0.20 | -0.18 | 0.35 | 0.03 |
| Cheese (g) | 0.08 | 0.21 | 0.24 | -0.06 |
| Nuts and seeds | 0.19 | -0.02 | 0.24 | 0.01 |
| Fruit juice including smoothies | 0.10 | 0.04 | 0.20 | 0.19 |
| Ice cream (g) | 0.04 | 0.10 | 0.19 | -0.13 |
| White fish coated or fried (g) | -0.12 | 0.11 | -0.16 | -0.15 |
| Lamb and dishes (g) | -0.01 | -0.01 | -0.07 | 0.06 |
| Wine (g) | 0.11 | -0.08 | 0.02 | 0.42 |
| Semi skimmed milk (g) | 0.07 | 0.26 | 0.06 | -0.39 |
| Sauces, pickles and gravies (g)' | -0.05 | 0.32 | 0.12 | 0.38 |
| Biscuits (g) | 0.07 | 0.11 | 0.30 | -0.37 |
| Spirits and liqueurs (g) | -0.12 | -0.04 | 0.05 | 0.29 |
| Eggs and egg dishes (g) | 0.09 | 0.10 | -0.03 | 0.27 |
| Reduced fat spread polyunsaturated (g) | -0.02 | 0.12 | -0.04 | -0.18 |
| Pork and dishes (g) | -0.03 | 0.14 | -0.02 | 0.16 |
| Other bread (g) | 0.10 | -0.01 | 0.11 | 0.16 |
| Other milk and cream (g) | 0.09 | -0.01 | 0.11 | 0.14 |
| Other meat and meat products (g) | -0.03 | 0.13 | 0.03 | 0.13 |
| Other breakfast cereals (g) | -0.02 | 0.08 | 0.06 | -0.12 |

**Table 2.7. Foods variables characterising each dietary pattern (NDNS 2008-2012).**

|  |  |  |
| --- | --- | --- |
| Dietary Pattern  label | Foods with moderate/strong positive factor loadings (≥0.25) | Foods with moderate/strong negative factor loadings (≤-0.25) |
| ‘Fruit, vegetables and oily fish (FVOF)’ | Fruit  Salad and other raw vegetables  Yoghurt, fromage frais and dairy deserts  Oily fish  Vegetables not raw  *Fruit juice including smoothies\*\**  Wine  Nuts and seeds | Chips (fried), roast potatoes and potato products  White bread  Sugar, preserves and sweet spreads |
| ‘Snacks, fast food and fizzy drinks (SFFFD)’ | Soft drinks (not low calorie)  Crisps and savoury snacks  Coated chicken and turkey  Burgers and kebabs  Chocolate confectionery  Chips (fried), roast potatoes and potato products  White bread  Sugar confectionery  *Pasta, rice and other cereals* | Tea, coffee and water  High fibre breakfast cereals  Fruit  *Wholemeal bread*  Vegetables, not raw |
| ‘Sugary food and dairy (SFD)’ | Biscuits  Semi skimmed milk  Chocolate confectionery  Buns, cakes, pastries and fruit pies  Sugars, preserves and sweet spreads  Puddings  Cheese  Sugar confectionery | Wine |
| ‘Meat, potatoes and beer (MPB)’ | Other potatoes, potato salads and dishes  Savoury sauces, pickles, gravies and condiments  White bread  *Butter*  *Sugar, preserves and sweet spreads*  Bacon and ham  Beer, lager, cider and perry  *Meat pies and pastries*  *Sausages*  Pork and dishes  Whole milk |  |

\*Foods are listed in descending order of the factor loading value.

\*\**Foods in italics did not have moderate/strong factor loadings in the 2008-2011 PCA*

*2.3.5.2. Foods and food groups that characterised dietary patterns in the 2008-2012 dataset compared with 2008-2011*

The four patterns that were identified in the PCA of the 2008-2012 data were, as expected, very similar to those identified from the 2008-2011 data as obviously there was a large overlap in the samples. There were some small differences such as the ‘FVOF’ dietary pattern explained the most variance in the 2008-2011 dataset (4%), whereas the ‘SFFFD’ component explained the most variance in the 2008-2012 dataset (3.9%). There were also some small differences in terms of the foods that characterised each pattern. In the 2008-2012 dataset, ‘wholemeal bread’ was moderately negatively associated with the ‘SFFFD’ pattern but was not moderately, positively associated with the ‘FVOF’ pattern as it was in the 2008-2011 data. There were several other small differences but none that altered the overall dietary pattern or the label allocated to it. Table 2.8 shows the similarities and differences in the foods characterising each dietary pattern identified in each year of analysis.

**Table 2.8. Similarities and differences between the foods with moderate/strong associations with each dietary pattern in 2008/2011 and 2008/2012\* datasets**

|  |  |  |  |
| --- | --- | --- | --- |
| **Dietary Pattern** | **Only 2008/2011** | **Both years** | **Only 2008/2012** |
| Fruit, vegetables and oily fish | Tea, coffee and water  Wholemeal bread | Fruit  Salad and other raw vegetables  Vegetables not raw  Oily fish  Wine  Yoghurt, fromage frais and dairy deserts  Nuts and seeds | Fruit juice including smoothies |
| *Whole milk*  *White fish coated or fried* | *Chips (fried), roast potatoes and potato products*  *White bread*  *Sugar, preserves and sweet spreads* |  |
|  |  |  |  |
| Snacks, fast foods and fizzy drinks | Soft drinks (low cal) | Soft drinks (not low calorie)  Crisps and savoury snacks  Coated chicken and turkey  Burgers and kebabs  Chocolate confectionery  Chips (fried), roast potatoes and potato products  White bread  Sugar confectionery | Pasta, rice and other cereals |
| *Other potatoes, potato salads and dishes* | *Tea, coffee and water*  *High fibre breakfast cereals*  *Fruit*  *Vegetables not raw* | *Wholemeal bread* |
|  |  |  |  |
| Sugary foods (and dairy) | Reduced fat spread not PUFA  Tea, coffee and water  White bread | Sugars, preserves and sweet spreads  Buns, cakes, pastries and fruit pies  Biscuits  Chocolate confectionery  Semi skimmed milk |  |
|  |  | Puddings  Cheese  Sugar confectionery  *Wine* |  |
|  |  |  |  |
| Meat, potatoes and beer | Spirits and liqueurs | Other potatoes, potato salads and dishes  Bacon and ham  Pork and dishes  Savoury sauces, pickles, gravies and condiments  White bread  Beer, lager, cider and perry  Whole milk  Butter | Sugar, preserves and sweet spreads  Meat pies and pastries  Sausages |
|  |  |  |

*\*Foods in italics showed negative associations with the patterns.*

**Table 2.9. Main effects regression model (95% confidence interval) of each dietary pattern for socio-demographic characteristics and lifestyle factors (NDNS 2008-2012, n=2083).**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample characteristics (n) | Snacks, fast food & fizzy drinks | | Fruit, veg & oily fish | | Meat, potatoes & beer | | Sugary foods & dairy | |
|  | Coefficient (95% CI) | P-Val | Coefficient (95% CI) | P-Val | Coefficient (95% CI) | P-Val | Coefficient  (95% CI) | P-Val |
| *Sex* |  | | | | | | | |
| Male (711) | 0.24 (0.16, 0.33) | <0.001 | -0.08  (-0.08, -0.08) | <0.001 | 0.63  (0.54, 0.72) | <0.001 | 0.19 (0.10, 0.29) | <0.001 |
| Female (914) | *Reference* | | | | | | | |
| *Age (2083)* | -0.03  (-0.03, -0.03) | <0.001 | 0.002 (0.002, 0.002) | <0.001 | 0.01 (0.004, 0.01) | <0.001 | 0.003 (<0.001. 0.01) | <0.001 |
| *Ethnicity* |  | | | | | | | |
| White (1495) | 0.13  (-0.02, 0.23) | 0.09 | -0.20  (-0.20, -0.19) | <0.001 | 0.56  (0.38, 0.73) | <0.001 | 0.46 (0.29, 0.64) | <0.001 |
| Non-white (130) | *Reference* | | | | | | | |
| *NSSEC Group* | | |  |  |  |  |  |  |
| Higher managerial  and professional (260) | 0.12  (-0.20, 0.3) | 0.46 | 0.25 (0.24, 0.27) | <0.001 | 0.03  (-0.33, 0.37) | 0.89 | 0.45 (0.09, 0.81) | 0.02 |
| Lower managerial  and professional (453) | 0.21  (-0.10, 0.51) | 0.18 | 0.14 (0.13, 0.15) | <0.001 | -0.04  (-0.37, 0.30) | 0.83 | 0.43 (0.08, 0.78) | 0.02 |
| Intermediate  Occupations (149) | 0.21  (-0.11, 0.53) | 0.20 | -0.02  (-0.03,  -0.01) | 0.002 | 0.04  (-0.31, 0.40) | 0.81 | 0.38 (0.01, 0.75) | 0.04 |
| Small employers (165) | 0.05  (-0.27, 0.34) | 0.76 | -0.03  (-0.04,  -0.02) | <0.001 | 0.25  (-0.10, 0.61) | 0.16 | 0.37 (0.007, 0.74) | 0.05 |
| Low supervisory  and technical  occupations (160) | 0.16  (-0.16, 0.47) | 0.33 | -0.12  (-0.14,  -0.11) | <0.001 | 0.09  (-0.27, 0.44) | 0.63 | 0.53 (0.16, 0.89) | 0.005 |
| Semi-routine  Occupations (228) | 0.27  (-0.4, 0.56) | 0.09 | -0.11  (-0.12,  -0.10) | <0.001 | 0.21  (-0.13, 0.56) | 0.23 | 0.52 (0.16, 0.87) | 0.004 |
| Routine occupations (177) | 0.20  (-0.11, 0.52) | 0.21 | -0.36  (-0.37,  -0.35) | <0.001 | 0.20  (-0.16, 0.55) | 0.27 | 0.25 (-0.11, 0.62) | 0.17 |
| Never worked (33) | *Reference* | | | | | | | |
| *Smoking status* |  | | | | | | | |
| Smoker (377) | 0.16 (0.06, 0.26) | 0.002 | -0.37  (-0.38, -0.37) | <0.001 | 0.21  (0.10, 0.33) | <0.001 | -0.21  (-0.33, -0.10) | <0.001 |
| Non-smoker (1248) | Reference | | | | | | | |
| *Household income (1777)* | 7.627E-07  (-1.207E, 2.733E-06) | 0.45 | 7.112E-06 (7.044E-06, 7.181E-06) | <0.001 | -9.357E-07  (-3.146E-06, 1.274E-06) | 0.41 | -2.192E-06  (-4.471E-06, 8.576E-08) | 0.06 |
| *BMI (1902)* | | |  |  |  |  |  |  |
|  | 0.13 (0.005, 0.02) | 0.01 | -0.002  (-0.002,  -0.001) | <0.001 | -0.003  (-0.12, 0.005) | 0.42 | -0.02  (-0.03, -0.01) | <0.001 |

*2.3.5.3. Association of dietary patterns with sample characteristics, socio-economic measures and lifestyle factors (NDNS 2008-2012)*

As for the 2008-2011 dataset, main effects regression models were used to explore adjusted associations between each dietary pattern and sample characteristics (age, sex, and ethnicity), socio-economic variables (household income and NSSEC) and lifestyle factors (smoking status and BMI). Plots of residuals were visually assessed for evidence of homoscedasticity, constancy of variance and outliers and these conditions of the model were considered to be satisfied (see Figures B.2.1-B.2.12, Appendix B).

As they were not normally distributed, regression analyses for BMI and household income including the value and the squared term with each dietary pattern were carried out. There was no evidence of non-linearity apart from between household income and the ‘SFFFD’ dietary pattern. A quadratic plot indicated that any curvilinear effects were negligible and not sufficient to affect the model (see Figure C.2.2, Appendix C)

Table 2.9 summarises the results of these analyses. The majority of the associations between the dietary patterns and the wider variables were as might be hypothesised from population level data (see Section 1.6.2) and reflected the associations observed in the dietary patterns identified in years 2008-2011 data. The SFFFD pattern was significantly (p≤0.05) positively associated with being male (0.24), being a smoker (0.16) and a higher BMI (0.13) and was negatively associated with age (-0.03). The FVOF pattern was significantly negatively associated with being male (-0.08), being white (-0.20), being a smoker (-0.37) and BMI (-0.002). The pattern was positively associated with age (0.002).

There was a clear gradient for the FVOF pattern with NSSEC which was significant for all categories of SES, with a lower SES being negatively associated with this pattern and a higher SES being positively associated this pattern. The SFD category was significantly positively associated with being male (0.19), being white (0.46) with age (0.003) and with all categories of occupation other than routine occupations. The pattern was most strongly associated with lower supervisory and technical and semi routine occupations (0.53 and 0.52 respectively, p<0.01). It was significantly inversely associated with being a smoker (-0.21) and with BMI (-0.02) which is contrary to the relationship that might be hypothesised for this dietary pattern. The MPB pattern was significantly positively associated with being male (0.63), being white (0.56), age (0.01) and being a smoker (0.21).

*2.3.5.4. Internal stability and reproducibility (PCA with 50% random sample of NDNS 2008-2012)*

Prior to conducting PCA, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett’s test of sphericity was undertaken to verify the sampling adequacy for the analysis (KMO=0.49 which is just below the limit deemed acceptable of 0.5 and Bartlett’s test of sphericity was significant indicating that correlations between items were sufficiently large for PCA) ([246](#_ENREF_246)).

PCA was conducted in a 50% random sample of the NDNS 2008-2012 dataset generated by SPSS. When extraction was based on eigenvalues >1.5, six components were initially extracted. On examination of the resulting dietary patterns, three patterns mirrored those identified in the previous analysis while three further patterns were less obviously similar. Observation of the scree plot suggested a clearer cut-off inflection after the fourth principal component than the sixth. On this basis, the PCA was then re-run, specifying the extraction of four components ([246](#_ENREF_246)). Food variables with factor loadings of ≥0.2 or ≤-0.2 with a component were extracted and retained for reporting (Table 2.10). Foods that were moderately or strongly correlated with a component (≥0.25 or ≤-0.25) were used to characterise and label the dietary patterns. The four components or ‘dietary patterns’ have similar characteristics to those in the previous two analyses demonstrating some internal validity and stability.

**Table 2.10. Food variables and factor loadings ≥0.25 or ≤-0.25** **for each Principal Component (random 50% sample NDNS 2008-2012).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Food/food group | Principal Component | | | |
| PC 1 | PC2 | PC3 | PC4 |
| Butter |  | 0.33 |  |  |
| Semi skimmed milk |  |  |  | 0.51 |
| Whole milk |  |  | -0.23 |  |
| Wholemeal bread | 0.36 |  |  |  |
| Beer, lager, cider and perry | -0.21 | 0.37 |  |  |
| Crisps and savoury snacks | -0.24 |  | 0.43 | 0.22 |
| Soft drinks (low calorie) |  |  | 0.49 |  |
| Soft drinks (not low calorie) | -0.32 |  | 0.33 |  |
| Sugar confectionery |  |  | 0.36 |  |
| Tea coffee and water | 0.43 |  |  | 0.21 |
| Wine | 0.24 | 0.28 |  | -0.27 |
| Bacon and ham |  | 0.41 | 0.23 |  |
| Biscuits |  |  |  | 0.36 |
| Buns, cakes, pastries and fruit pies |  |  |  | 0.46 |
| Burgers and kebabs | -0.26 |  | 0.21 |  |
| Chips (fried), roast potatoes and potato products | -0.33 | 0.40 |  |  |
| Chocolate confectionery |  |  | 0.46 | 0.32 |
| Coated chicken and turkey |  |  | 0.29 |  |
| Eggs and egg dishes | 0.22 | 0.21 |  |  |
| Fruit | 0.60 | -0.21 |  |  |
| High fibre breakfast cereals | 0.23 |  | -0.33 |  |
| Meat pies and pastries |  | 0.26 |  |  |
| Nuts and seeds |  |  | 0.21 |  |
| Oily fish | 0.46 |  |  |  |
| Other meat and meat products |  | 0.20 |  |  |
| Pasta rice and other cereals |  | -0.27 | 0.24 |  |
| Pork and dishes |  | 0.23 |  |  |
| Salad and other raw vegetables | 0.49 |  |  |  |
| Sausages |  | 0.39 |  |  |
| White bread | -0.31 | 0.40 |  | 0.22 |
| Yogurt, fromage frais and dairy desserts | 0.38 |  |  |  |
| Other potatoes, potato salads and dishes | 0.32 | 0.39 |  |  |
| Vegetables not raw | 0.47 |  |  |  |
| Puddings |  |  |  | 0.33 |
| Skimmed milk | 0.21 |  |  |  |
| Sugars, preserves and sweet spreads |  |  | -0.21 | 0.45 |
| Reduced fat spread not polyunsaturated |  |  |  | 0.24 |
| Savoury sauces, pickles, gravies condiments |  | 0.39 |  |  |

Foods with the highest factor loadings characterising each component are the same as for those identified in the PCA of the 2008-2011 dataset (Table 2.11) therefore each component or ‘dietary pattern’ has been subjectively labelled in the same way.

**Table 2.11. Foods characterising each dietary pattern (random 50% sample NDNS 2008-2012)**

|  |  |  |
| --- | --- | --- |
| ‘Dietary Pattern’  Label | Foods with moderate/high positive factor loadings (≥0.25 or ≤-0.25)\* | Foods with moderate/high negative factor loadings (≥0.25 or ≤-0.25)\* |
| ‘Fruit, vegetables and oily fish’ | **Fruit\*\*\***  **Salad and other raw vegetables**  **Oily fish**  **Vegetables not raw**  **Tea, coffee and water**  **Yoghurt, fromage frais and dairy deserts**  **Wholemeal bread**  *Other potatoes, potato salads and dishes\*\** | **Chips (fried), roast potatoes and potato products**  *Soft drinks (not low calorie)*  **White bread**  *Burgers and kebabs* |
| ‘Snacks, fast food and fizzy drinks’ | Soft drinks (low calorie)  **Crisps and savoury snacks**  **Chocolate confectionery**  **Sugar confectionery**  **Soft drinks (not low calorie)**  **Coated chicken and turkey** | **High fibre breakfast cereals** |
| ‘Sugary food and dairy’ | **Semi skimmed milk**  **Buns, cakes, pastries and fruit pies**  **Sugars, preserves and sweet spreads**  **Biscuits**  **Chocolate confectionery**  *Puddings* | **Wine** |
| ‘Meat, potatoes & beer’ | **Bacon and ham**  **White bread**  **Chips, fried and roast potatoes and potato products**  *Sausages*  **Other potatoes, potato salads and dishes**  **Savoury sauces, pickles, gravies and condiments**  **Beer, lager, cider and perry**  *Butter*  *Wine*  *Meat pies and pastries* | *Pasta rice and other cereals* |

\*Foods are listed in descending order of the value of their factor loading;

*\*\*Foods in italics did not have moderate/strong factor loadings in the same dietary pattern in the 2008-2011 analysis;*

**\*\*\***Foods that have moderate/strong factor loadings in the same dietary pattern in the 2008-2011 analysis are in bold.

*2.3.6. Regressions with nutrient intake and nutrient biomarkers (objective 3).*

Table 2.12 shows the results of the regression analyses undertaken with each dietary pattern and nutrient intakes derived from self-reported dietary intake diaries and nutrient levels derived from urine and blood plasma samples. The FVOF pattern was significantly positively associated with intake per 1000 calories of vitamins C, D, E, B12, B6, iron, folate and magnesium; proportion of food energy from n-3 and n-6 PUFAs; NSP and biomarkers of vitamin C, D (25-hydroxy vitamin D) and A (retinol), iron (ferritin) and total carotenoids. It was significantly negatively associated with the proportion of food energy from NMEs, saturated fat, total fat, starch and biomarkers of triglycerides and urinary sodium. The SFFFD pattern was negatively associated with intake per 1000 calories of vitamins C, D, E, B6, B12, folate, magnesium, proportion of food energy from n-3, grams of NSP and biomarkers of vitamins C, D, A, total cholesterol and total carotenoids. This pattern was positively associated with proportion of food energy from NMEs, total fat, n-6 PUFA, starch and with urinary sodium. The MPB dietary pattern was negatively associated with intake of vitamins C and E, iron and magnesium and with biomarkers for total carotenoids. It was positively associated with proportion of food energy from NMEs, saturated fat, total fat and NSP and biomarkers of triglycerides, urinary sodium and ferritin. The SFD pattern had very similar associations to that of the SFFFD pattern with the exception of being positively associated with the proportion of food energy from saturated fat and absolute intake of NSP and negatively associated with n-6 PUFA, starch and biomarkers of retinol, ferritin and urinary sodium.

**Table 2.12. Main effects regression model (95% confidence intervals) of dietary patterns for nutrient intakes and biomarkers (NDNS 2008-12)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Snacks, fast foods fizzy drinks | Fruit, vegetables and oily fish | Meat, potatoes and beer | Sugary foods and dairy |
|  | Coefficients (95% Confidence intervals), \*p<0.05, \*\*p<0.01, \*\*\*p<0.001 | | | |
| ***Nutrient intake*** |  | | | |
| Vitamin C per 1000 kcal | -0.01 (-0.006, -0.004)\*\*\* | 0.01 (0.011, 0.014)\*\*\* | -0.003 (-0.004, -0.002)\*\*\* | -0.003 (-0.004, -0.002)\*\*\* |
| Vitamin D per 1000 kcal | -0.20 (-0.23, -0.16)\*\*\* | 0.17 (0.133, 0.20)\*\*\* | 0.01 (-0.03, 0.41) | -0.10 (-0.13, -0.6)\*\*\* |
| Vitamin E per 1000 kcal | -0.04 (-0.05, -0.02)\*\*\* | 0.08 (0.07, 0.10)\*\*\* | -0.06 (-0.7, -0.4)\*\*\* | -0.02 (-0.4, -0.01)\*\* |
| Vitamin B6 per 1000 kcal | -0.19 (-0.27, -0.11)\*\* | 0.09 (0.01, 0.18)\* | -0.03 (-0.16, 0.05) | -0.40 (-0.48, -0.32)\*\*\* |
| Vitamin B12 per 1000 kcal | -0.10 (-0.12, -0.08)\*\*\* | 0.04 (0.02, 0.05)\*\*\* | -0.01 (-0.03, 0.01) | -0.05 (-0.07, -0.03)\*\*\* |
| Iron per 1000 kcal | -0.12 (-0.14, -0.10)\*\*\* | 0.11 (0.09, 0.14)\*\*\* | -0.08 (-0.10, -0.06)\*\*\* | -0.14 (-0.16, -0.12)\*\*\* |
| Folate per 1000 kcal | -0.01 (-0.006, -0.005)\*\*\* | 0.003 (0.003, 0.004)\*\*\* | -0.001 (-0.001, 0.001) | -0.003 (-0.004, -0.002)\*\*\* |
| Magnesium per 1000 kcal | -0.01 (-0.01, -0.008)\*\*\* | 0.01 (0.009, 0.011)\*\*\* | -0.01 (-0.006, -0.004)\*\*\* | -0.01 (-0.006, -0.004)\*\*\* |
| % Food Energy NMES | 0.05 (0.04, 0.05)\*\*\* | -0.01 (0.21, 0.007)\*\*\* | 0.02 (0.02, 0.03)\*\*\* | 0.03 (0.02, 0.03)\*\*\* |
| % Food Energy Sat Fat | -0.12 (-0.024, 0.001) | -0.04 (-0.05, -0.02)\*\*\* | 0.08 (0.07, 0.09)\*\*\* | 0.06 (0.05, 0.07)\*\*\* |
| % Food Energy Total Fat | 0.01 (0.01, 0.02)\*\*\* | -0.01 (-0.02, -0.002)\*\* | 0.04 (0.03, 0.05)\*\*\* | 0.01 (0.01, -0.02)\*\*\* |
| % Food Energy n-3 PUFA | -0.21 (-0.30, -0.12)\*\*\* | 0.45 (0.36, 0.54)\*\*\* | 0.08 (-0.01, 0.17) | -0.31 (-0.40, -0.22)\*\*\* |
| % Food Energy n-6 PUFA | 0.06 (0.03, 0.08)\*\*\* | 0.03 (0.002, 0.06)\* | -0.004 (-0.03, 0.02) | -0.04 (-0.07, -0.01)\*\* |
| % Food Energy Starch | 0.01 (0.001, 0.02)\* | -0.04 (-0.05, -0.03)\*\*\* | -0.02 (-0.03, -0.01)\*\*\* | -0.02 (-0.03, -0.01)\*\*\* |
| Non-starch polysaccharides (fibre) | -0.03 (-0.04, -0.02)\*\*\* | 0.11 (0.10, 0.12)\*\*\* | 0.03 (0.02, 0.04)\*\*\* | 0.07 (0.06, 0.08)\*\*\* |
| ***Nutrient level from biomarkers*** |  | | | |
| Vitamin C | -0.004 (-0.007, -0.001)\* | 0.02 (0.01, 0.02)\*\*\* | -0.004 (-0.01, -0.001)\* | 0.001 (-0.002, 0.004) |
| 25-Hydroxy Vitamin D | -0.002 (-0.004, 0.001) | 0.01 (0.004, 0.009)\*\*\* | -0.001 (-0.003, 0.002) | 0.0004 (-0.002, 0.003) |
| Retinol (Vitamin A) | -0.10 (-0.20, 0.004) | 0.13 (0.02, 0.23)\* | 0.12 (-0.001, 0.21) | -0.18 (-0.29, -0.08)\*\*\* |
| Ferritin (iron) | 4.908E-05 (-0.001, -0.001) | 0.001 (5.153E-5, 0.001)\* | 0.001 (0.001, 0.002)\*\*\* | -0.001 (-0.001, -3.55)\* |
| Triglycerides | 0.01 (-0.05, 0.08) | -0.06 (-0.13, 0.002) | 0.13 (0.06, 0.19)\*\*\* | -0.02 (-0.09, 0.04) |
| Total cholesterol | -0.07 (-0.13, -0.02)\*\* | 0.04 (-0.01, 0.10) | -0.03 (-0.08, 0.03) | -0.07 (-0.13, -0.02) |
| Urinary sodium | 0.01 (0.005, 0.008)\*\*\* | -0.01 (-0.006, -0.003)\*\*\* | 0.002 (0.0004, 0.003)\* | -0.002 (-0.003, 0.0001)\* |
| Total carotenoids | -0.08 (-0.15, -0.02)\* | 0.32 (0.26, 0.38)\*\*\* | -0.16 (-0.23, -0.09)\*\*\* | -0.03 (-0.09, 0.04) |

* + 1. *Selecting items for the YHS-DQ based on the dietary patterns analyses results (Objective 5)*

Each dietary pattern generated in the 2008-2011 PCA had 6 items that had high factor loadings apart from the SFFFD component which had 8 items. As ‘biscuits’ and ‘buns, cakes, pastries and fruit pies’ were both strongly positively associated with the ‘sugary foods’ dietary pattern only and are both categories of foods that are similar in nutrient composition, they were combined to create a broader, derived category of ‘biscuits, cakes and pastries’. In addition to ‘salad and other raw vegetables’, consumption of portions of fruit and vegetables was asked using two separate validated questions that reflected those in the national Active People Survey, in order to be able to compare results with the national indicator. Whilst ‘wholemeal bread’ scored 0.26 which was just below the factor loading cut-off for consideration of 0.3, it was included in the YHS-DQ, as it was the only indicator reflective of the government dietary recommendation to consume wholemeal versions of starchy carbohydrates rather than non-wholegrain varieties ([143](#_ENREF_143)).

On the basis of the nature of the results and the method described in 2.2.2.6, the following items remained for consideration for inclusion in the YHS-DQ: ‘semi-skimmed milk’, ‘beer, lager, cider and perry’, ‘crisps and savoury snacks’, ‘soft drinks (not low calorie)’, ‘bacon and ham’, ‘biscuits, cakes and pastries’, ‘burgers and kebabs’, ‘chips’, ‘coated chicken and turkey’, ‘oily fish’, ‘salad and other raw vegetables’, ‘white bread’, ‘wholemeal bread’, ‘fruit’ and ‘vegetables’.

Due to the limited space on the survey and thus the number of items that could be included in the tool, it was necessary to reduce this list by two foods. The rationale for removal of items was based on a process of elimination according to their contribution to diet quality overall (contribution to intake of key nutrients) and their importance as a key characterising food of a particular dietary pattern (e.g. fruit in the FVOF pattern). On this basis, ‘burgers and kebabs’ and ‘semi-skimmed milk’ were excluded. Appendix A shows the resulting questions included in the YHS questionnaire. Table 2.13 below shows the balance of items reflecting each dietary pattern and includes the direction of the association with each selected food in both the 2008-2011 and the 2008-2012 dietary patterns analysis.

* + 1. *Internal consistency and stability of foods selected for the YHS-DQ*

As described in Section 2.3.5 above, the dietary patterns identified through PCA of the 2008-2011 NDNS dataset were demonstrated to be consistent and stable through application of the same method in the 2008-12 dataset and a random 50% split of the 2008-2012 dataset. The dietary patterns identified in each dataset were labelled in an identical way as the foods that characterised each pattern were very similar. Table 2.13 below shows the items that were selected from the results of the initial 2008-2011 PCA for inclusion in the YHS-DQ and the direction of their association with each dietary pattern. The direction of the association in the 2008-2012 analysis (carried out after the YHS-DQ was finalised) is included in brackets in order to provide some further indication of internal validity and consistency.

**Table 2.13. Items selected for the YHS-DQ and the direction of their association with each dietary pattern generated from the 2008-2011 and 2008-2012 datasets\***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Food/drink group | Fruit veg and oily fish | Snacks, fast foods and fizzy drinks | Sugary foods and dairy | Meat, potatoes and beer |
| Beer, lager, cider and perry |  |  |  | Positive  *(Positive)* |
| Crisps and savoury snacks |  | Positive  *(Positive)* |  |  |
| Sugary drinks (fizzy pop, squash) |  | Positive  *(Positive)* |  |  |
| Bacon and ham |  |  |  | Positive  *(Positive)* |
| Biscuits, cakes and pastries |  |  | Positive  (Positive) |  |
| Chips | Negative  *(Negative)* | Positive  *(Positive)* |  |  |
| Chicken (coated, fried) |  | Positive  *(Positive)* |  |  |
| Oily fish | Positive  *(Positive)* |  |  |  |
| Salad and other raw vegetables | Positive  *(Positive)* |  |  |  |
| White bread | Negative  *(Negative)* | *(Positive)* |  | Positive  *(Positive)* |
| Wholemeal bread | Positive | (Negative) |  |  |

\*2008-2012 dataset in italics

* 1. **Discussion**

*2.4.1. Summary of findings*

* In a representative sample of UK adults, 13.7% of the variance in dietary intake can be described by and reduced to ‘dietary patterns’ from an initial 60 variables included in the analyses.
* The patterns were demonstrated to be internally stable in a 50% random sample of the data and when including an additional year of more recent data in the sample.
* The dietary patterns were characterised by consumption of higher or lower quantities of particular foods and food groups.
* The four patterns explaining the most variance were labelled as ‘fruit, vegetables and oily fish’, ‘snacks, fast foods and fizzy drinks’, ‘meat, potatoes and beer’ and ‘sugar foods and dairy’ based on the foods that featured most dominantly in each pattern.
* The patterns were significantly associated with population characteristics, wider determinants of health, nutrient intake and biomarkers of nutritional status with the first two patterns showing most relationships that were as hypothesised.
* The FVOF pattern was associated with biomarkers of nutrient status that were considered indicative of a better quality diet. The reverse was true of the SFFFD dietary pattern.
* Those scoring higher on the FVOF pattern were more likely to be older, have a lower BMI, have a higher income and be of a higher NSSEC. They were less likely to be male, white, smokers and be in routine occupations. Those scoring higher on the SFFFD pattern were more likely to be male, smokers, younger and have a higher BMI.

*2.4.2. Deriving the data-driven dietary patterns*

PCA was used as a method to reduce detailed dietary intake data from a representative sample of the UK population into dietary patterns. Four broad dietary patterns were identified explaining only 13.7% of the variance in the data. This is in line with, if somewhat lower than, similar studies in the UK which have explained between 6% and 40% of variance in the dietary data ([193](#_ENREF_193), [195](#_ENREF_195), [247-250](#_ENREF_247)). The fact that the four dietary patterns generated explained only 13.7% of the variability in the dataset could be interpreted as providing evidence that the diet of the UK adult population (as defined by the 60 variables included in the PCA) is too complex to be reduced to ‘dietary patterns’ in this way. If only 13.7% of the variability is explained, then 86.3% of the variability remained unexplained.

Studies with a lower number of dietary intake variables (thus less detailed dietary intake information) included in the PCA have been shown to result in a higher proportion of the variability explained ([193](#_ENREF_193)). Where food group categories are broad, foods that are weakly associated with a pattern may be classified in the same category as foods more strongly associated thus increasing the amount of information that a specific pattern may capture ([172](#_ENREF_172)). However, this in turn has been shown to have an impact on the sensitivity of the components in terms of their associations with risk factors for disease. This may indicate that when investigating an exposure such as dietary intake, the greater detail may be more important in extracting patterns than the amount of variance explained ([172](#_ENREF_172)) and may be why many authors do not report the proportion of variance explained by the factors ([160](#_ENREF_160)).

This methodology for examining dietary patterns in the population is also limited by the subjective decisions that are made regarding the level of detail that should be included in the PCA, including the initial categorisation and definition of foods and food groups. Dietary intake data can be collected and analysed at multiple levels of detail and how these data are classified can have an impact of the sensitivity of the resulting patterns and their interpretation. In the case of the NDNS dataset, nutrient-level data are available but food and food group categories have been pre-defined in the dataset for analysis ([147](#_ENREF_147)). Whilst the 60 variables included in the PCA were more detailed than those included in some dietary patterns studies, some of the food group categories were broad and thus included a range of foods with different nutritional characteristics, which may have fallen into differing ‘dietary patterns’ had they been differentiated. Therefore some categories may have been improved through aggregation or disaggregation of existing food and food group categories for the purposes of the research questions in this thesis. For example, the category ‘pasta, rice and other miscellaneous cereals’ contains a range of foods that differ widely in terms of nutrient profile and impact on health - from pizza and manufactured pasta ready meals to porridge or homemade risotto ([147](#_ENREF_147)). Ideally, the categories would be framed to ensure they are most appropriate for the focus of the research.

*2.4.3. The consistency and stability of the dietary patterns*

The dietary patterns identified reflected those observed in other similar studies investigating empirically-derived dietary patterns in the UK adult population using PCA. A study in a sample of young Northern Irish adults aged 20 to 25 years identified four dietary patterns: ‘drinker/social’ characterised by white bread, alcohol, fats and meat dishes; ‘healthy’ characterised by fruits, vegetables and brown bread; ‘western’ characterised by soft drinks, crisps and chips and ‘sweet tooth’ which was characterised by puddings, chocolate and confectionery ([249](#_ENREF_249)). A study of older men and women aged 59 to 70 in Hertfordshire, only two patterns were identified: a ‘prudent’ pattern characterised by fruit, vegetables, oily fish and wholemeal cereals and a ‘traditional’ pattern characterised by processed and red meat, vegetables and puddings ([203](#_ENREF_203)). A study in a sample of 12, 053 pregnant women identified five dietary patterns: ‘health conscious’ characterised by salad, fruit, pulses, non-white bread, fish and oat and bran based breakfast cereals; ‘traditional’ characterised by vegetables, meat and poultry; ‘processed’ characterised by meat pies, sausages and burgers, fried foods, pizza, chips and crisps; ‘confectionery’ characterised by chocolate, sweets, biscuits, cakes and puddings and ‘vegetarian’ characterised by meat substitutes, nuts, pulses and herbal teas ([159](#_ENREF_159)). Despite the fact that all of these studies were conducted in different samples of adults in the UK, with differing dietary assessment and data collection methods, there are some similarities in the patterns that are revealed. This provides support to the external validity and generalisability of the patterns identified in this thesis.

Further support for their validity comes from their associations with sample characteristics and wider determinants of health which reflect the patterns of variance observed in other national UK datasets and in previous analyses of intake of single nutrients and foods in the NDNS dataset ([147](#_ENREF_147)). The ‘fruit, vegetables and oily fish’ dietary pattern in particular reflects the socio-economically stratified patterns of fruit and vegetable intake described in Section 1.6.2. This pattern was also inversely associated with BMI and smoking whereas the more energy dense, low nutrient dietary pattern ‘snacks, fast food and fizzy drinks’ was positively associated with BMI and smoking. This also reflects the patterns described in Section 1.6.2 from analysis of the UK HSE data.

The internal validity and stability of the dietary patterns were tested through undertaking the same PCA in the updated NDNS dataset, the same sample with the addition of the 2011/12 data and also in a 50% randomly split sample of the four year (2008-2012) dataset. The dietary patterns results from the three sets of analyses were characterised, in the main, by the same foods and were labelled to reflect this. This provides support for the assumption that the patterns are internally valid and stable.

*2.4.4. The limitations of using empirically-derived dietary patterns to inform the Yorkshire Health Study Diet Questionnaire (YHS-DQ)*

In the initial PCA from the 2008-2011 dataset, four distinct dietary patterns were identified explaining 13.7% of the variance in the data. The development of the YHS-DQ was constrained by two practical parameters. Firstly, the timescale for the second wave of the YHS survey to be completed was prescribed and did not allow for the results of all the analyses to be conducted before the questions to be included were finalised. Secondly that there was limited space on the paper questionnaire, therefore limiting the number of questions that could be asked to approximately 12. These are typical constraints that are faced by public health practitioners in local settings. This process resulted in items being included in the YHS-DQ on the basis of a clear rationale but without robust testing of their content or external validity.

As stated above, some degree of validity of the items in the YHS-DQ came from exploration of the associations between the 2008-2011 data generated dietary patterns and sample characteristics such as gender, age, smoking status, income and BMI status. For example, those scoring higher on the ‘fruit, vegetables and oily fish’ pattern were less likely to be smokers or in routine occupations and were more likely to have a higher household income. Those scoring highly on the ‘snacks, fast food and fizzy drinks’ pattern were more likely to be smokers and be younger. These patterns reflect the associations that would be expected based on other data and evidence in the literature, given the nature of the dietary patterns revealed. This provided some confidence in the patterns and in the foods characterising them as being typical of patterns on dietary intake in the UK population.

The question of whether the PCA methodology provided valid and useful indicators of diet quality for inclusion in the YHS-DQ will be further explored in Chapters 3 and 4. In terms of responding pragmatically and in a timely manner to the request for food based questions for the study, there are two positive points. Firstly, the opportunity to include questions generated through the first study of this thesis allowed for a set of indicators generated by one of the exploratory methods to be tested in a distinct sample for analysis for associations with other variables included in that dataset such as socio-economic variables, lifestyle factors and long term health outcomes such as type 2 diabetes, depression and hypertension. Should the opportunity arise, these self-reported questions could be robustly validated in a sub-sample of the YHS through use of a more robust and detailed dietary assessment method such as a weighed diary or (resources permitting) through the collection of biomarkers through urine samples and blood plasma.

Secondly, whilst the methods used to generate the questions in the YHS-DQ may not be comprehensive, they have been generated using a robust method of reducing detailed, national, representative dietary intake data to dietary patterns. Thus the ‘indicator foods’ provided for the tool are, to some degree, evidence based and representative of broad patterns of diet. The fact that there is some support for the internal stability and external validity of the patterns support their use within a diet monitoring questionnaire.

As described in Section 2.3.7, some compromises were made in terms of the food items included in the questionnaire due to space and ease of understanding for the questionnaire recipient. Some foods were prioritised over others due to their relevance in the context of the current UK government dietary guidelines and others were excluded on the basis of the food group category being too broad to be meaningfully interpreted. Therefore, some decisions were made regarding the items included in the final YHS-DQ that were, to an extent, subjective albeit with a rationale. Studies 2 and 3 will assess whether the use of these empirically-derived dietary patterns to generate ‘indicators’ for inclusion in such a tool is valid and useful without a more theoretically driven justification for the content of the questions.

**Chapter 3.**

**The development and validation of a theoretically-derived UK Nutrient-level Diet Quality Score (Study 2)**

**3.1 Introduction, aims and objectives**

As described in Section 1.7.4.2, at the time of investigation there was no existing standard Diet Quality Score or Healthy Diet Index for assessing adults’ adherence to current population level UK Dietary Reference Values (DRVS) and wider government guidelines relating to diet ([251-253](#_ENREF_251)). For the purposes of this thesis, a theory driven, nutrient-based composite measure of diet quality based on the most recent SACN nutrient-level dietary guidelines ([143](#_ENREF_143), [239](#_ENREF_239), [253](#_ENREF_253)) and key public health nutrition issues in the UK was needed. The purpose of the score was to provide a standard definition and measure of diet quality against which proxy indicators for diet quality and the brief assessment tool could be tested.

The research questions and objectives for Study 2 are set out below. This study will contribute to the over-arching objective by providing a standard, composite measure of diet quality, reflecting UK DRVs and dietary guidelines against which a brief diet quality assessment tool can be developed, evaluated and validated.

**Research questions:**

1. How should diet quality be defined, quantified and measured in the UK adult population?

2. Which foods, socio-demographic and lifestyle variables are associated with a theory driven definition of dietary quality?

**Objectives:**

1. To develop a theory driven, nutrient-based Diet Quality Score (NDQS) for use in the UK adult population based on UK dietary recommendations.

2. To evaluate the content validity of the NDQS through exploring associations with nutrient biomarkers.

3. To assess the generalisability of the NDQS through exploring associations with socio-demographic and lifestyle variables.

**3.2. Objective 1. Development of a theory driven, nutrient-based Diet Quality Score**

The aim of this part of the study is to develop a theory driven, nutrient-based Diet Quality Score (NDQS) for use in the UK adult population.

**3.2.1. Methods**

*3.2.1.1. Dataset*

The dataset used was the UK NDNS. Data were available for four years 2008/09, 2009/10, 2010/11 and 2011/12. The NDNS dataset contains individual level data for intakes (average grams per day) of 54 nutrients (macronutrients and micronutrients) from a sample of adults aged ≥19 years (n=2083, 43% males, mean age = 49, 92% white, full sample characteristics are shown in Table 2.2.) derived from the disaggregation of food intake data collected using a three or four day food diary. The data are in the public domain and were downloaded from the UK Data Service ([244](#_ENREF_244)).

*3.2.1.2. Scope and nature of the score*

As described in Section 1.7.4, diet quality or healthy eating scores and indices have been developed based on consumption of foods or food groups ([254](#_ENREF_254)), nutrient only ([255](#_ENREF_255)) or, more frequently, a mixture of both foods and nutrients ([211](#_ENREF_211), [256](#_ENREF_256)). Several indices also include measures of dietary moderation or overall dietary diversity ([176](#_ENREF_176)). The former often includes a composite measure scoring levels of consumption of key aspects of dietary intake such as total fat, saturated fat, added (or ‘free’ or ‘non milk extrinsic’) sugars and alcohol. Dietary diversity measures the range of different types of foods or variety of food groups consumed within the diet habitually. This has been shown in some studies to be an indicator of diet quality and in other to be associated with positive health outcomes such as reduced diabetes risk ([257](#_ENREF_257)). The methods by which dietary diversity and dietary quality are defined and measured differ significantly by the research objectives and the population being studied and thus it may not be a generalisable indicator ([179](#_ENREF_179), [208](#_ENREF_208)).

For the purposes of this thesis, it was considered that the score needed to be at nutrient level. The rationale for this was predominantly theoretical in that a nutrient level definition was considered to be a more robust measure of dietary quality for the purposes of validation. It would avoid large scale misclassifications due to the potential breadth (and thus wide range of potential nutrient content) within some of the different foods and food group variables such as ‘pasta, rice and other cereals’ or ‘yoghurts, fromage frais and dairy desserts’.

Secondly the use of a nutrient-based scale avoids any tautology when analysing the associations between food based indicator variables and diet quality. As the variables included in the dietary patterns analyses (see Table 2.1) were foods and food group variables defined in the NDNS dataset, the NDQS was developed at nutrient level to ensure that associations analysed between the dietary patterns and a theoretically-derived measure of diet quality were not tautological (i.e. the predictors and the outcomes in the model are equivalent). This would also be avoided by ensuring that different years of the NDNS dataset were used in assessing content validity.

Thirdly, a nutrient-based measure provided scope for more sensitive ranges of scores to be set according to the specific and detailed SACN guidelines for DRVs (RNIs and EARs) for the UK population ([239](#_ENREF_239)). Indices such as the WHO HDI have prioritised simplicity of data collection, scoring and analysis by having a dichotomised score or 0 or 1 for each item ([212](#_ENREF_212)) whilst others such as the Diet Quality Index Revised (DQI-R) provide an opportunity for increased sensitivity to a range of dietary quality by including a range of scores for each item ([220](#_ENREF_220), [230](#_ENREF_230)). In order to take account of the diverse range of variations and combinations of dietary intake that could be broadly categorised as being ‘high’ or ‘low’ quality, it was considered that there should be a minimum of 10 items included and that the scoring range should be developed to be as sensitive as possible within the context of the SACN DRV guidelines. Each item should be weighted and scored according to its priority for public health nutrition and health in the UK, the mean (and standard deviation of) consumption in the UK ([147](#_ENREF_147)) compared with dietary guidelines and the UK DRV ([258](#_ENREF_258)).

The following existing diet quality/healthy eating indices and their validation studies were referred to in developing the NDQS construct and providing a rationale for the scoring methodology where one was not clearly indicated by the UK DRVs (they are described in more detail in Section 1.7.4):

* World Health Organisation Healthy Diet Indicator ([1](#_ENREF_1));
* U.S. Healthy Eating Index 2005 ([256](#_ENREF_256)),
* Diet Quality Index Revised ([220](#_ENREF_220)),
* Alternative Healthy Eating Index ([259](#_ENREF_259)),
* Australian Dietary Guideline Index ([175](#_ENREF_175))
* The nutritional scoring system for assessing diet quality for children aged 10 years and under in the UK ([230](#_ENREF_230)).

Table 3.5 summarises the NDQS, including which nutrients were included, their relative scoring range and unit and the mean consumption in the UK as estimated by the NDNS 2008-2012 ([147](#_ENREF_147)). The rationale for inclusion of the items is described below in Section 3.2.1. Items within the score were classified as ‘deficient’, ‘excess’ or ‘adequate’ to indicate where population level consumption was either too little or too much in relation to current dietary guidelines and therefore where scored weighting may have been considered. A diet considered to be of the highest quality according to this scale scored 100 and 0 was the minimum score possible indicating a very poor quality diet.

The macronutrients included in the score were saturated fat (percentage contribution to total energy), trans fat (percentage contribution to total energy), cis n-3 and cis n-6 polyunsaturated fats (percentage contribution to total energy), NMES (percentage contribution to total energy), protein (grams per day) and NSP (grams per day). These items provide a reflection of intake across all macronutrient groups. The micronutrients included in the score were sodium, calcium, iron, vitamin C and folate (all grams per day). These micronutrients were selected for inclusion because they were key contributors (both negative and positive) to health outcomes and diet-related disease in the UK or were nutrients that the UK population were deficient in. For these micronutrients there was also some evidence in the literature of their use as indicators of diet quality in previous studies ([175](#_ENREF_175), [230](#_ENREF_230)).

*3.2.1.3. Weighting and scoring rationale*

Indices were weighted either a maximum of 5 or 10 (double) points on the NDQS. Key macronutrients that the UK general population or a significant number of sub-groups within it were not meeting the current recommendations (saturated fat, non-milk extrinsic sugars (NMEs), non-starch polysaccharide (NSP), sodium, iron and alcohol) were prioritised within the construct and were weighted double the other items and given a maximum score of 10. Saturated fat and NMEs were also doubly weighted as diets high in fat and sugar tend to be energy dense ([260](#_ENREF_260)). Energy dense diets are associated with obesity and markers of chronic diseases such as diabetes type 2 ([261](#_ENREF_261)) which are high priority public health issues in the UK ([262](#_ENREF_262)). Vitamin C was also weighted 10 as there was evidence that it was a key nutrient-based indicator of diet quality and consumption of a fruit and vegetables intake ([263](#_ENREF_263), [264](#_ENREF_264)). Similar approaches to weighting some components of a score have been taken in the development of other scores and indices ([265](#_ENREF_265)).

Minimum and maximum scores were set based on the UK DRVs, taking into account whether there was any further SACN guidance provided for individual maximum and minimum requirements ([239](#_ENREF_239)). The mean intakes for nutrients for the UK population as observed in the NDNS 2008-2012 was also taken into account when considering the scoring and cut-offs. Where necessary, other indices including nutrient-based measures such as the WHO Healthy Diet Indicator, the U.S. Healthy Eating Index 2005 and the Diet Quality Index Revised were referred to for supporting evidence of scoring ranges and cut-offs applied in other indices (e.g. saturated fat, sodium).

*3.2.1.4. Macronutrients in excess at population level*

*i) Percentage of total energy from fat*

Where UK DRVs existed for a type of fat as a proportion of total energy, and the data were available in the NDNS dataset for analysis, these were included in the NDQS. No separate measure for total fat was included, despite the existence of a DRV for this, in order not to duplicate scoring at individual level.

*ii) Percentage of total energy from saturated fat*

The average population DRV for the percentage of total energy from saturated fat is < 10% ([239](#_ENREF_239)). This nutrient is a high priority for public health nutrition in the UK due to its association with CVD ([266](#_ENREF_266)) and energy dense diets which can lead to obesity ([1](#_ENREF_1)). The criteria for the maximum score for this item was set at <10% of total energy contribution from saturated fat in line with the UK DRV, WHO recommendations and the U.S. indices. As the UK mean was 12.4% ([147](#_ENREF_147)), a pragmatic range of scores was set from the maximum (< 10%) to the minimum score (> 15%). The criteria for a minimum score total energy from saturated fat >15%. This cut-off was adapted from the U.S. Healthy Eating Index-2010 ([226](#_ENREF_226)) with the minimum score allocated as ≥15%, as no equivalent UK guideline exists**.**

*iii) Percentage of total energy from cis-PUFA (n-3 and cis n-6)*

There were no population average DRVs set for cis n-3 and n-6 PUFA separately however there were individual minimum levels set at 0.2% and 1.0% of total energy respectively and the population average UK DRV for percentage of total energy contribution of total PUFAs is 6% ([239](#_ENREF_239)). The WHO goals for population level intake were that 5-8% of total energy contribution should be from n-6 PUFA and 1-2% from n-3 PUFA ([1](#_ENREF_1)). The mean average contribution in the UK population for n-3 PUFA was 1% of total energy for n-6 PUFA was 4.8% total energy ([147](#_ENREF_147)). The maximum and minimum score ranges and cut-offs for the NDQS for n-3 and n-6 PUFA were set taking these data into consideration with the range for achieving the maximum score of 5 set at 2 to 8% for n-6 PUFA and 0.2 to 2% for n-3 PUFA. The minimum score of 0 was given for any percentage greater than or less than these ranges (see Table 3.5).

*iv) Percentage of total energy from NMES*

A key type of carbohydrate that is currently over-consumed in the UK population compared with the DRV is NMES. Therefore this nutrient was included in the index. The most recent UK guidelines regarding carbohydrate consumption have replaced the definition of NMES with ‘free sugars’. This is a slightly less stringent definition compared with NMES as it does not include any dried or processed fruit within in a dish/food and has reduced the DRV for adults to a maximum of 5% ([252](#_ENREF_252)). As the published NDNS dataset did not include data for ‘free sugars’ according to these updated definitions, the definition of ‘NMES’ and the respective DRV was used for developing the NDQS. In order to take account of the more stringent DRV relating to the consumption of ‘free sugars’, the maximum score was given for individuals whose consumption of NMEs was ≤5% of their total energy. To compensate for the fact that NMEs includes more sugars than ‘free sugars’, therefore making this cut-off more stringent than the current UK guidelines and with consideration of the UK mean consumption of NMEs (11.3% of total energy), the scoring was set to be realistic with individuals scoring 8 out of a possible 10 if they consumed 5-8% of their total energy from NMEs, 6 if they consumed 8-11% of their total energy from NMEs, 4 for 11-13%, 2 for 13-15% and 0 for >15%.

*v) Alcohol*

Whilst alcohol is not a nutrient, it is included here as a key element of dietary quality that is currently excessively consumed at population level. It is positively associated with many negative health outcomes including a higher risk of some cancers, liver disease and heart conditions. Alcohol also contributes to overall energy intake and thus, potentially weight gain. Chronic over consumption is associated with nutritional deficiencies ([267](#_ENREF_267)). Alcohol is included in several existing indices either as a single item or within composite scores for ‘moderation’ ([175](#_ENREF_175), [226](#_ENREF_226), [256](#_ENREF_256)).

The alcohol score was set based on the UK public health guidelines for alcohol consumption which was set at 14 units a week for men and women (an average of 2 units or 8 grams per day) ([268](#_ENREF_268)). The item was weighted 10 as the current population mean shows that people in the UK are consuming, on average, more than is recommended. In the first iteration of the score, the maximum score was given to a mean consumption of 0-4 g of alcohol with scores awarded at intervals of 2 points between this maximum score and anything over 16g which received the minimum score of 0. However, in the final iteration of the score, this was considered too stringent as the mean consumption in the NDNS 2008-2012 was 24g (3 units). Therefore in the final version of the UKDQS, the maximum score was allocated for a mean consumption of less than 16g per day (2 units), the minimum score was allocated for a mean consumption of more than 24g (3 units) and a score of 5 was given for between 16 and 24g.

*3.2.1.5. Macronutrients deficient at population level*

*i) NSP*

A key carbohydrate group that is currently under consumed by the UK population compared with the DRVs non-starch polysaccharides (NSP) or ‘fibre’. Therefore this nutrient was included in the index. The recent SACN guidelines regarding carbohydrate consumption provided updated guidance for ‘fibre’ consumption based on an alternative definition of ‘fibre’ from NSP, the Englyst method of calculation. The Englyst method includes lignin and allows for comparisons with other European countries who use this definition (the Association of Official Analytical Chemists or AOAC method of calculation) ([143](#_ENREF_143)).

As the published NDNS dataset did not include figures for Englyst calculation or ‘AOAC fibre’ intake according to these updated definitions, the previous ‘NSP’ definition and DRV has been used in the development of the score (18g per day). The existing population average DRV for NSP was used as the basis for the maximum score, however the scoring range was developed to reflect minimum and maximum DRVs that are set for individuals ([258](#_ENREF_258)).

*3.2.1.6. Macronutrients adequate at population level*

*i) Protein*

Protein was included in the measure due to it being an essential component of dietary quality and a key element of the UK government’s dietary recommendations ([228](#_ENREF_228)). The score for levels of protein intake was adjusted according to age and sex as per the DRV. The mean intake in the UK is currently adequate in all age groups ([147](#_ENREF_147)). The DRV for protein is provided as an absolute figure, not as a proportion of total energy therefore it was included as such in the NDQS with the maximum score being 5.

*3.2.1.7. Micronutrients in excess at population level*

*i) Sodium*

Reduction of consumption of sodium (or salt) is a high priority for public health nutrition policy in the UK due to the association between lower levels of sodium consumption with lower risk for hypertension, coronary heart disease and stroke ([269](#_ENREF_269)). Sodium was weighted with a maximum score of 10 due to its priority in terms of the UK population dietary health. The Lower Reference Nutrient Intake (LRNI) and a Reference Nutrient Intake (RNI) for adults is 575mg and 1600 mg respectively ([258](#_ENREF_258)). Subsequently, a maximum level of 2400 mg sodium was set for the UK adult population ([147](#_ENREF_147), [251](#_ENREF_251)). The UK population mean consumption is currently 2213 mg (SD 830), which is higher than the current RNI but lower than the maximum target.

Several different approaches have been taken to scoring sodium consumption on existing indices in countries where the target is the same as the UK. The lowest score on U.S. Healthy Eating Index ([173](#_ENREF_173)) was given for intake of sodium that was >4800mg per day (the optimal dietary score was given for ≤), however the updated HEI-2005 adjusted this for total energy by allocating the minimum score for greater than 2000 mg/per 1000kcal per day. The US Diet Quality Score allocated the score associated with the poorest diet to those consuming over 3,400 mg per day ([210](#_ENREF_210), [211](#_ENREF_211)). To take account of this and increase sensitivity to the wide range of levels of consumption of sodium in the UK diet, the minimum score was allocated for consumption >3000 mg per day with a range of scores set around these cut-offs (2700mg-3000mg=2, 2400-2700=4, 2100-2400=6, 1600-2100=8, 575-1600=10). The aim of this scoring is to reflect the range of dietary recommendations and also to be sensitive to realistic levels of intake by allocating scores for those who just fall short of meeting the target.

*3.2.1.8. Micronutrients deficient at population level.*

*Iron*

Iron is a component of haemoglobin, myoglobin and many enzymes and contributes to red blood cell, hormone and neurotransmitter synthesis and in the functioning of the immune system ([267](#_ENREF_267)). Certain groups of the UK population are more vulnerable to iron deficiency due to having higher physiological requirements to meet tissue growth (e.g. in infants, adolescents, pregnant women), high losses (e.g. menstruating women) or poor absorption (e.g. the elderly) ([258](#_ENREF_258)). In the UK, 23% of adult women (aged 11-64 years) some of whom are particularly vulnerable to deficiency due to menstruation, are not meeting the RNI for iron ([147](#_ENREF_147)). On this basis, combined with its essential contribution to physiological functioning, iron was included with a weighted maximum score of 10. It has also been included as an item in other indices and scores such as the U.S. Diet Quality Index ([220](#_ENREF_220)) and has been included as an validating indicator or determinant of diet quality in several studies ([254](#_ENREF_254), [255](#_ENREF_255)). The cut-offs for maximum and minimum score allocation are based on the EAR and RNI for males and females aged 19-50 years and >50 years. A maximum score was given for those consuming above the RNI and minimum score for those consuming less than the EAR.

*3.2.1.9. Micronutrients adequate at population level*

*i) Calcium*

The maximum score for calcium was 5 as the population mean is currently higher than the DRV. Calcium was included in the score as there is evidence that calcium intake is a useful indicator independently of dietary quality ([270](#_ENREF_270)) and is also a marker of dairy food consumption. For adults aged ≥19 years the UK LRNI is 400mg/day, the Estimated Average Requirement (EAR) 525mg/day and the RNI 700mg/day. The EAR and RNI were used as the cut-offs for achieving the minimum and maximum scores respectively and a mid-range of score was set between these figures reflecting the mean and range of intakes in the UK population.

*ii) Vitamin C and Folate*

Vitamin C and folate have been shown in previous studies to be indicative of diet quality, diet diversity and consumption of fruits and vegetables, therefore they were considered key items for inclusion in the diet quality score ([264](#_ENREF_264), [271](#_ENREF_271)). Vitamin C is also an important antioxidant, as it assists in the absorption of non-haem iron, prevents scurvy and aids in the healing of wounds ([258](#_ENREF_258)). Vitamin C was considered a key micronutrient and indicator for health and diet quality independently and therefore was given a weighted maximum score of 10. Folate was given the unweighted maximum score of 5. The basis of the scoring system for both nutrients was the EAR and RNI; with the maximum score being allocated when the individual consumed above the RNI and the minimum score being allocated when the individual consumed below the EAR. A mid-range score of 5 for Vitamin C and 2 for folate was given for consumption levels between the EAR and the RNI.

*3.2.1.10. Other considerations in the development of the NDQS*

*(i) Adequacy versus optimal intake*

In defining a high or low level of diet quality for the UK population, the question of whether to define scores relating to nutrient adequacy or optimal intake was considered. DRVs (RNIs and EARs) are not intended to provide optimal levels of nutrient intake (for health or disease preventions). They are either the average requirement of a group of people (EAR) or an amount of the nutrient that is enough to ensure that approximately 97% of the population are not deficient. For some nutrients a ‘safe’ level of intake was defined as there is not sufficient evidence to set an EAR or and RNI ([258](#_ENREF_258)). Evidence supporting what the most beneficial or optimal intake levels of macro and micro nutrients for either optimal health or disease prevention is the subject of ongoing research. The quality and extent of the evidence varies for different nutrients in part driven by the extent to which a population is deficient and also depending on the extent to which a particular disease or condition is conclusively associated with intake of a particular nutrient. There were no consistently agreed recommendations for optimal levels of intake of nutrients for the UK population therefore, the definition of dietary quality was driven by existing, widely accepted SACN recommendations for nutrient ‘adequacy’.

*(ii) Total fat*

Some indices allocate a score for total fat intake either as a single item or within a composite item for ‘moderation’ ([226](#_ENREF_226)). Whilst there is a population average DRV for the maximum proportion of total energy that should come from total fat and the current mean intake in UK adults is in excess of this, this item was not included in the score. This was on the basis that several of the sub-categories of fatty acids (saturated fats, cis n-3 and cis n-6 PUFA and trans fatty acids) are covered as separate items in the score and therefore to include total fat would be duplicating a measurement. Including the different categories of fat as opposed to just one measure for ‘total fat’ allowed the score to more sensitively classify diets containing high and low levels of different types of fat that have different contributions to diet-related health.

*(iii) Total energy: An independent determinant of diet quality*

Total energy was a key factor for consideration when developing the score. Inevitably, the greater the number of kilocalories consumed the greater the likelihood of reaching the recommended intakes of absolute grams of nutrients (albeit a mixture of those that require moderation and those that require increased consumption) ([176](#_ENREF_176), [272](#_ENREF_272)) thus potentially misclassifying individuals in terms of their overall relative diet quality and nutrient density. Some indices such as the Mediterranean Diet Score adjust for energy intake by standardising grams of nutrients consumed per 1,000 kilocalories ([176](#_ENREF_176)) while other indices do not account for total energy variation.

Consideration was given as to whether it was possible to include ‘cut-offs’ or scores for total energy as a separate item within the score as total energy can be considered an independent contributor to diet quality as it is possible to consume a diet that is highly nutrient dense but is still excessive and therefore can lead to weight gain and obesity. Allocating a generalised maximum level ‘cut-off’ (for example >2000 kilocalories for women and >2500 kilocalories for men) is problematic as the energy requirements of an individual depend on a number of variables such as gender, age, height, weight and activity levels ([273](#_ENREF_273)). In order to accurately calculate whether the individual is reporting consuming more kilocalories than they require, these data fields are therefore required to be in the dataset.

It is possible to test whether energy intake is more than is required because under conditions of weight stability (such as during the short period of time during which the NDNS data were collected), energy intake (EI) equates to total energy expenditure (TEE). Therefore the reported energy intake of a group can be examined through estimation of TEE at either individual or group level ([55](#_ENREF_55)). TEE can be defined as Basal Metabolic Rate (BMR) plus the energy expended through the level of physical activity (PAL) which can be expressed as a multiple of BMR. In estimating TEE, the assumption has to be made that PAL can also be expressed as the ratio of Reported Energy Intake to Basal Metabolic Rate (EIrep:BMR) ([274](#_ENREF_274)).

However, the NDNS dataset only has an objective measure of physical activity measure for a sub-sample ([147](#_ENREF_147)). Without these physical activity data, any ‘cut-off’ could lead to individuals who are very active, for example, being incorrectly scored. Therefore a level above which kilocalorie consumption is judged as too high would be to a large extent, arbitrary. In order to account for total energy intake to some degree in the NDQS itself, the scoring of the key macronutrients which contribute to excess energy intake (i.e. fats and sugars) were standardised by including them as a proportion of total energy as opposed to an absolute value. This is in line with the UK dietary recommendations for these macronutrients ([239](#_ENREF_239)).

Total energy was also included as a covariate in regression models examining whether the empirically-derived dietary patterns identified are predictive of the NDQS later in the thesis. This method of adjusting for energy intake in regression models at a later stage of analyses is consistent with other studies and recommendations ([275](#_ENREF_275)).

*(iv) Total energy: estimating for systematic under-reporting of energy intake and testing the sensitivity of the NDQS*

It is likely that there will be systematic bias in the scoring of individuals due to under-reporting of kilocalorie intake. Self-reported dietary assessment methods are widely acknowledged to result in under-reporting of energy intake, in particular in groups such as those who are obese and who are the most athletic ([274](#_ENREF_274), [276](#_ENREF_276)) and studies have demonstrated under-reporting in the NDNS ([87](#_ENREF_87)). Nutrient intake is strongly positively associated with energy intake, therefore it is likely that if energy intake is under reported or under estimated, then nutrient intake will be too ([274](#_ENREF_274)).

As described above, the inclusion of total energy as a separate item in the NDQS score was not possible as it would have required an accurate calculation of energy requirement versus energy intake. Therefore, to evaluate the sensitivity of the NDQS to under-reporting more generally in terms of whether the distribution of NDQS score in the general sample reflected that of a sample where the estimated under-reporters (eUR) had been removed, levels of energy requirement (eER) were estimated through multiplying individual level BMR and by a population level PAL.

Estimated energy requirements (EERs) were calculated for four years of data (2008/09 to 2011/12). As sex, age, and weight were available for the majority of the sample (n=1927), it was possible to calculate BMR using Schofield equations ([276](#_ENREF_276)). EERs were then calculated using the Goldberg cut-off method ([277](#_ENREF_277)) PAL of 1.54, which assumes 4 days of dietary intake data and a sample size of <2000 ([55](#_ENREF_55)) and 1.4, the standard physical activity multiplier specified by COMA for both sexes engaging in ‘light’ occupational and non-occupational physical activity ([258](#_ENREF_258)). This latter PAL of 1.4 is also considered by the UK SACN to reflect the inactive lifestyle of much of the population and is used as the basis for Estimated Average Requirements (EARs) for energy for groups of ages and weights for men and women ([239](#_ENREF_239)). These estimates were used to calculate ratios which were compared to those calculated from a doubly labelled water study which was conducted in a sub-set of the NDNS 2008/2009 sample ([147](#_ENREF_147)). Doubly labelled water is the most accurate and objective method for assessing energy expenditure and thus energy intake and allows the identification of ‘implausible reporters’ ([74](#_ENREF_74)).

Table 3.1 shows the results of the mean (and Standard Deviations) estimated Basal Metabolic Rate (BMR Est), the mean total energy intake reported (EIRep) and the mean estimated Estimated Energy Requirement (EER) using the different Physical Activity Levels (PAL) for males and females aged 19 and over. Table 3.2 shows the ratios of reported energy intake (EIrep) to EER for each of the PALs for males and females. Table 3.3 shows the mean ratios of reported energy intake to EER by BMI category.

**Table 3.1. Estimated BMR, mean EIRep and mean EER (SD) at PALs of 1.4 and 1.54**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | EstBMR | EIRep | EER PAL 1.4 | EER Goldberg 1.54 |
| Male | 1861 (281) | 2100 (609) | 2605 (393) | 2884 (436) |
| Female | 1420 (169) | 1576 (426) | 1988 (237) | 2201 (262) |

**Table 3.2. Mean ratios of EIrep to EER for at BMR. PAL of 1.4 and 1.54**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | EIRep | EIRep:EER (BMR) | EIrep:EER  (PAL 1.4) | EIrep:EER  (Goldberg 1.54) |
| Male | 2100 (609) | 1.13 | 0.81 | 0.73 |
| Female | 1576 (426) | 1.11 | 0.80 | 0.72 |

**Table 3.3. Mean ratios of EIRep to EER by BMI category**

|  |  |  |
| --- | --- | --- |
| Valid BMI | n | EIRep:EER |
| <18.5 | 27 | 0.94 |
| 18.5-24.99 | 635 | 0.83 |
| 25-29.99 | 697 | 0.73 |
| 30-40 | 491 | 0.64 |
| >40 | 52 | 0.55 |

These findings reflect those from a study in which individual level EE calculations and Physical Activity Levels (PALs) were calculated from doubly labelled water (DLW) in a representative sub-set of the NDNS 2008/2009 sample (n=197 adults aged >19) ([147](#_ENREF_147)). Table 3.4 shows the results from the findings from the DLW study. The mean EI:TEE for all adults aged 19 and over in the sub-set was 0.73 (SD 0.26).

**Table 3.4. Mean PAL (SD) and ranges for adults aged 19 and over in the NDNS (2008/09) doubly labelled water sub sample study**

|  |  |  |
| --- | --- | --- |
|  | Mean PAL(SD) | Min-Max (Range) |
| Males >19 (n=99) | 1.78 (0.28) | 1.19-2.73 (1.54) |
| Females >19 (n=98) | 1.68 (0.22) | 1.18-2.11 (0.93) |
| Total (n=197) | 1.73 (0.26) | 1.18-2.73 (1.55) |

The ratio of reported EI:EER in the findings in this study that reflected the ratio in the doubly labelled water sub-study of 0.73 is the ratio resulting from the Goldberg cut-off of 1.54. Therefore this was the multiplier adopted from which to calculate the mean level of under-reporting for each sex. Using the Goldberg cut-off, the mean level of under-reporting was 784kcal for men and 625 kcal for women or 27.2% and 28.4% of estimated energy requirements respectively. These were more conservative estimates than identified in the DLW sub sample study where adults aged ≥16 years were estimated to be under reporting by 907 kcal for men and 785 kcal for women ([147](#_ENREF_147)), but are consistent with a previous study evaluating under reporting of energy NDNS 2000 ([87](#_ENREF_87)). This study compared the Goldberg cut-off method with an ‘Estimated Energy Requirements’ method and found that using the latter method, the median under-reporting in men was found to be 837 kcal in men and 675 kcal in women and that there were no significant differences between the estimates from each method ([87](#_ENREF_87)). As observed in other studies ([87](#_ENREF_87), [274](#_ENREF_274)) the mean ratio of reported energy intake to EER (BMR x 1.54) was inversely correlated with BMI. With each increasing grouped BMI category, reported energy intake became a smaller and smaller proportion of EER (Table 3.3).

Whilst some studies have shown that the Goldberg cut-off method can lead to the misclassification of some individuals, particularly those with higher energy expenditures ([87](#_ENREF_87), [278](#_ENREF_278)) and can therefore introduce a systematic bias, for the purposes of evaluating the sensitivity of the NDQS (as opposed to for example, excluding participant’s data from the analyses), it is considered appropriate. Section 3.2.2 below show the distribution of NDQS scores in the NDNS 2008/09-2011/12 sample and the distribution of scores of the sub-sample of those estimated to be under-reporters using the methods described above.

**3.2.2. Results (Objective 1)**

**3.2.2.1. Nutrient-based Diet Quality Score (NDQS)**

Table 3.5 shows the Nutrient-based Diet Quality Score (NDQS) that was developed, including the scoring range and the NDNS 2008-2012 population mean for each nutrient (and alcohol).

**Table 3.5 Nutrient-based Diet Quality Score (NDQS): Criteria, scoring and national comparators (DRVs and population means\*\*)**

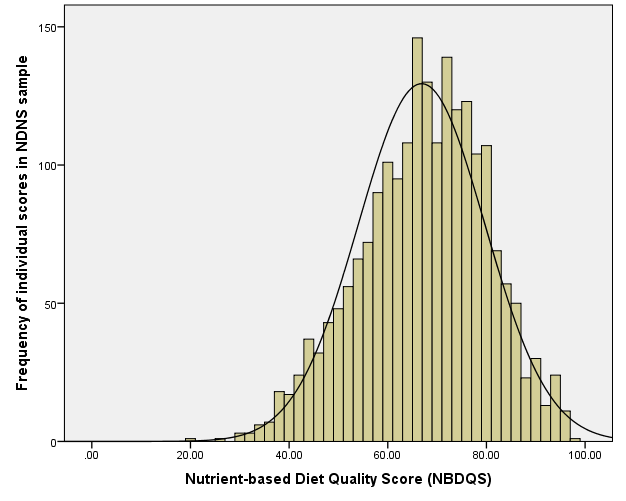
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nutrient | Mean intake (SD) | Max score | DRV\* | Score Range | Maximum score | Minimum score |
| Saturated fat % Total Energy (TE) | 12.4 (3.4) | 10 | 10 | 10.1-11 8  11.1-12 6  12.1-13 4  13.1- 15 2 | <10 | >15 |
| Trans fat % TE | 0.6 (0.3) | 5 | 2 | 0/5 | ≤2 | ≥2 |
| Cis n3 PUFA % TE | 1 (0.5) | 5 | Indiv min 0.2 | 0/5 | 0.2-2 | <0.2 >2 |
| Cis n6 PUFA % TE | 4.8 (1.5) | 5 | Indiv min 1 | 0/5 | 2-8 | <2 >8 |
| NMES % TE | 11.3 (6) | 10 | 10 | >5- 8 8  >8-11 6  >11-13 4  >13-15 2 | ≤5 | >15 |
| Protein (g/day) | 72.9 (25.2) | 5 | *M<50* 55.5  *M>50* 53.3  *F<50* 45  *F>50* 46.5 | 0/5 | *M<50* 55.5-111  *M>50* 53.3-106.6  *F<50* 45-90  *F>50* 46.5-93 | *M<50* <55.5>111  *M>50* <53.3 >106.6  *F<50* <45 >90  *F>50* <46.5 >93 |
| Non starch polysaccharide (g/day) | 13.8 (5.1) | 10 | 18  Indiv min 12  Indiv max 24 | 12-13.9 5  14-15.9 8  16-19.9 10  20-21.9 8  22-24 5 | 16-20 | <12  >24 |
|  |  |  |  |  |  |  |
| Sodium (mg) | 2213.2 (830.5) | 10 | LNRI 575  RNI 1600  Dietary target max 2400 | 1600-2100 8  2100-2400 6  2400-2700 4  2700-3000 2 | 575-1600 | >3000  <575 |
| Calcium (mg) | 815.3 (310.6) | 5 | LNRI 400  EAR 525  RNI 700 | 525-699.9 2 | ≥700 | <525 |
| Iron (mg/day) | 10.5 (3.6) | 10 | *M 19-50*  LNRI 4.7  EAR 6.7  RNI 8.7  *F 19-50*  LNRI 8.0  EAR 11.4  RNI 14.8  *>50*  LNRI 4.7 | 6.7-8.69 5  11.4-14.79 5  6.7-8.69 5 | *M 19-50* ≥8.7  *F 19-50* ≥14.8  50+ ≥8.7 | <6.7  <11.4  <6.7 |
|  |  |  | EAR 6.7 |  |  |  |
| Alcohol (g/day) (8g=1 unit) | 24 (26.3) | 10 | RNI 8.7  2 units (14 units a week) | 16.01-24 5 | ≤16 | >24 |
| Vitamin C (mg/day) | 83.1 (61.9) | 10 | LNRI 10  EAR 25  RNI 50 | 25-40 5 | >40 | <25 |
| Folate (µg/day) | 257.2 (104.9) | 5 | LNRI 100  EAR 150  RNI 200 | 150-200 2 | >200 | <150 |

\*Population average unless otherwise stated \*\* Population means from NDNS 2008/12

**3.2.2.2. Distribution of the NDQS**

Figure 3.1 shows the distribution of the score in the NDNS sample (2008-2012, n=2083). The distribution was normal with a mean score of 66.9 (SD 12.84). The range was 20 to 98, with only one individual scoring in the lowest quintile of the score.

**Figure 3.1. Distribution of NDQS scores in UK population (NDNS 2008-2012, n=2083, aged 19+).**



**3.2.2.3. NDQS quartiles by demographics, socio-economic measures and lifestyle factors**

Table 3.6 below shows the breakdown of the proportions of the sample in each NDQS quartile and mean scores by age, income group, gender, ethnicity, smoking status, NSSEC and WHO BMI classification ([279](#_ENREF_279)). The lowest quartile indicates the lower levels of dietary quality. T-tests, chi squared linear-by-linear test (for ordinal variables such as NDQS quartiles) and one-way ANOVA with Bonferroni post-hoc tests were used to investigate the differences in mean scores and proportions by each characteristic across the NDQS quartiles.

*Age*

The mean age increased with each incremental increase in NDQS quartile suggesting that older people tend to have a better quality diet than younger people. The mean ages for each quartile were significantly different from each other apart from between the first and the second quartile (p<0.05).

*Income*

The mean income increased with each incremental increase in NDQS quartile suggesting that people with higher incomes tend to have better quality diets. The mean income for the fourth quartile was significantly higher than for the first and second quartiles (p<0.001).

*Gender*

Whilst the difference in mean scores for men and women was seemingly small, with men scoring 67.7 and females scoring 66.2, the difference was statistically significant (p<0.001). The highest proportion of females was in the bottom quartile of the NDQS (28.2%) whereas the highest proportion of males (28.4%) was in the second quartile. Chi squared test, linear-by-linear result showed that the differences between the proportions are significantly different across the quartiles (8.46, p<0.01).

*Ethnicity*

A higher proportion of individuals of white ethnicity were in the bottom quartile of the NDQS than non-white and there was a higher proportion of non-white individuals in the highest quartile than white. Chi squared test linear-by-linear result indicated that the differences in proportions across the quartiles were significant (3.76, p<0.05). The difference between the mean scores for white and non-white individuals was statistically significant (p<0.05).

*Smoking status*

The proportion of smokers in the bottom quartile of the NDQS was significantly greater than the proportion of non-smokers (45.4% and 18.8% respectively). The inverse was also true, with only 12% of smokers achieving scores in the top quartile and 27.1% of non-smokers. Chi squared test linear-by-linear result indicated that the differences in proportions across the NDQS quartiles were significant (141.91, p<0.001). There was a significant difference in mean score between smokers and non-smokers (p<0.001).

*NSSEC (Socio-economic Status)*

The NDQS mean scores for the higher and lower managerial and professional NSSEC groups were significantly higher than semi routine and routine occupations (p<0.001). Intermediate occupations, small employers and lower supervisory and technical occupational groups were significantly higher than routine occupations (p<0.001). No other differences between mean scores were significant.

The highest socio-economic group also had the highest proportion in the top quartile of the NDQS (27.5%). The socio-economic group with the highest proportion in the bottom quartile of the NDQS was ‘routine occupations.’ There was a clear gradient in the top quartile of the NDQS from the lowest proportion achieving this level of diet quality coming from the lowest socio-economic group and the highest proportion achieving this level of diet coming from the highest socio-economic group. The patterns were less clear for the lower quartiles. For quartiles 1 and 2 the patterns were less clear. Chi squared linear-by-linear test result indicated that the differences were significant across the quartiles (45.91, p<0.001).

*BMI WHO Classification Group*

The mean score of 58.1 for the ‘underweight’ classification was significantly lower than all the other classification groups apart from the ‘morbidly obese’ group (p<0.01). There were no other significant differences between mean scores across the groups. The chi squared linear-by-linear reflected this indicating that the differences in proportions across the NDQS quartiles were not significant (0.009, p=0.92).

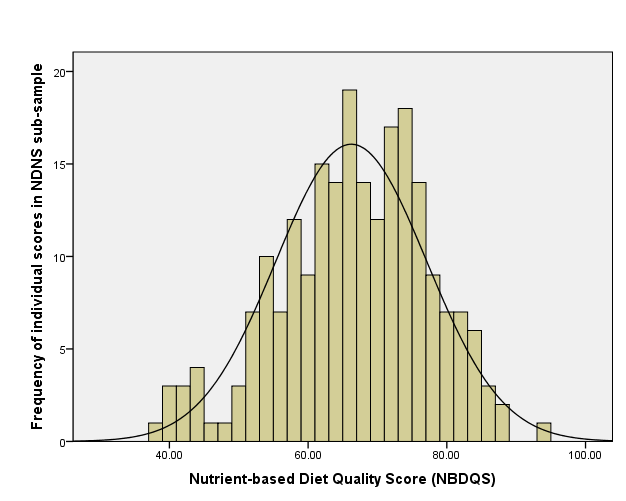
As described in Section 1.6.2. these patterns reflect those from other national data sources ([114](#_ENREF_114)). The BMI group with the highest proportion of individuals in the top quartile is the overweight group. The BMI group with the highest proportion of individuals in the bottom quartile and the lowest proportion of individuals in the top quartile is the underweight group. The highest proportion of individuals in the very obese group were in the bottom quartile. The highest proportion of healthy weight individuals were in the second quartile.

**Table 3.6. Proportion in each NDQS quartile and mean (SD) scores by sample characteristics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Characteristic | Quartile | | | |  |
|  | 1 | 2 | 3 | 4 |  |
| *Mean in each NDQS quartile* |  |  |  |  |  |
| Age (years, n=2083) | 44.6 | 47.1 | 50.3 | 55.2 |  |
| Income (£s, n=1777) | 27638 | 33542 | 33171 | 33733 |  |
|  |  |  |  |  |  |
| *Proportions in each NDQS quartile* |  |  |  |  | *Mean (SD)* |
| Male (n=901) | 21.2 | 28.4 | 26.2 | 24.2 | 67.7 (12.3) |
| Female (n=1182) | 28.2 | 27.4 | 21.5 | 22.9 | 66.2 (13.2) |
| White (n=1905) | 25.6 | 28.0 | 23.3 | 23.1 | 66.7 (12.9) |
| Non-white (n=178) | 20.2 | 26.4 | 25.8 | 27.5 | 68.7 (12.4) |
| Smoker (n=1583) | 45.4 | 27.2 | 15.4 | 12.0 | 60.4 (13.1) |
| Higher managerial and professional occupations (n=309) | 16.2 | 28.5 | 27.8 | 27.5 | 69.8 (11.4) |
| Lower managerial and professional occupations (n=553) | 21.9 | 27.3 | 23.9 | 26.9 | 68.2 (12.5) |
| Intermediate occupations (n=191) | 22.0 | 30.9 | 22.5 | 24.6 | 67.5 (11.7) |
| Small employers and own account workers (n=222) | 24.8 | 23.9 | 27.9 | 23.4 | 67.2 (12.7) |
| Lower supervisory and technical professions (n=203) | 24.6 | 26.1 | 26.1 | 23.2 | 67.3 (12.8) |
| Semi routine occupations (n=289) | 32.2 | 28.4 | 20.1 | 19.4 | 64.5 (13.6) |
| Routine occupations (n=235) | 39.1 | 28.1 | 16.6 | 16.2 | 62.1 (13.5) |
| Never worked (n=43) | 30.2 | 34.9 | 18.6 | 16.3 | 64.1 (15.2) |
| Underweight (n=27) | 48.1 | 29.6 | 14.8 | 7.4 | 58.1 (12.9) |
| Healthy Weight (n=635) | 25.0 | 26.8 | 24.4 | 23.8 | 67.1 (12.7) |
| Overweight (n=697) | 24.7 | 26.3 | 23.7 | 25.4 | 67.2 (12.9) |
| Obese (n=491) | 23.6 | 30.8 | 23.4 | 22.2 | 66.8 (12.5) |
| Very Obese (n=52) | 32.7 | 25 | 28.8 | 13.5 | 64.5 (11.7) |

**3.2.2.4. Distribution in sub-sample of estimated non under-reporters (eUR)**

Section 3.2.1.10 (iv) describes the method used to identify the estimated under reporters using the Goldberg cut-off method and assuming significant systematic under-reporting in the dataset as evidenced by the literature ([277](#_ENREF_277)). In order to test the sensitivity of the NDQS to under-reporting, the distribution of the NDQS scores was examined in the individuals in the data analysed in Section 3.2.1.10.iv who were estimated as not being under-reporters. Figure 3.2 below shows the distribution of NDQS scores in this sub-set of the sample (n=219). The distribution mirrors that shown above in the general sample and is normal. The mean score is 66.2 (10.9) and is not significantly different from the mean score in the general sample.

**Figure 3.2. Distribution of NDQS scores in non-eUR sub-sample (NDNS 2008-2012, n=219, aged 19+)**

**3.2.3. Discussion (Objective 1)**

The first objective of this study in the thesis was to develop a theory-driven, nutrient-based Diet Quality Score based on UK DRVs and government dietary recommendations ([228](#_ENREF_228), [239](#_ENREF_239)). In order to do this, the methods used for the development of a number of similar indices for other countries were reviewed and adapted for this score. The rationale for inclusion of nutrients in the score was based on *a priori* knowledge of key nutrients for health and prevention of disease, with a focus on those diet-related health issues that are most prevalent in the UK. The rationale for weighting and scoring of nutrient items in the score was defined by UK DRVs and also by population level mean intake levels. In the few cases where UK recommendations did not exist, it was guided by scoring used for existing indices. The NDQS was developed in the NDNS dataset and analyses was undertaken to examine the distribution of the score by population sub-groups and excluding estimated under-reporters.

As discussed in Section 1.7.3, there are limitations and challenges in developing diet quality or healthy eating indices. Many of the decisions relating to weighting, scoring and inclusion of items, whilst carefully considered and justified in the context of the thesis, are subjectively made. Different decisions relating to the inclusion or exclusion of items and how they are scored impacts on the distribution and variance of scores in a sample or population. Weighting items within an index that relates to the health value of a dietary intake is problematic. It is difficult to evaluate and make a relative or absolute health related judgement or comparison between items within the scale. For example, assessing the overall health outcome related value of under-consumption of Vitamin C compared with an over-consumption of saturated fat is likely to be, to an extent subjective, particularly when making a population level ‘judgement’ ([255](#_ENREF_255)). Similarly, the creation of an aggregate index has the potential to allocate the same scores for intakes of very different typologies of diet in terms of food and food group intake. This means that inevitably there will be some loss of sensitivity in terms of understanding the nature of the ‘true’ underlying dietary pattern represented by the score.

The fact that the NDQS construct does not include an item that allows for an assessment of an individual’s energy intake levels is a limitation. It is a particularly salient limitation in a score that is intended to takes into account the most prevalent public health nutrition issues in the UK. As discussed in Section 3.2.1.10, energy intake could be considered to be a vital component of diet quality as under consumption or over consumption of calories, from any source, can lead to diet-related disease. Over consumption of calories, regardless of the nutrient density of the calories consumed, leads to weight gain, increasing the risk of obesity which is associated with a multitude of diseases and is a public health crisis in the UK and globally. The score does take account of nutrient density in terms of grams of nutrient per 1000 calories consumed, for some of the key macronutrients such as saturated fat and non-milk extrinsic sugars. However, an individual level score relating to total energy intake versus energy requirement could not be included as the NDNS dataset does not include a valid measure of physical activity at individual level therefore any estimation energy requirement would be subject to potential error.

Another limitation is that the scores generated for individuals in the NDNS dataset, were based on nutrient intake levels derived from dietary intake data using food composition tables ([147](#_ENREF_147)). Estimates of nutrient levels in foods in food composition tables are subject to several types of potential inaccuracy such as random variability in food composition (in some foods there is significant within-food variation due to production methods; storage, processing or cooking practices; or factors such as the soil and fertilisers applied), bias from incorrectly identification of food items and differences in bioavailability of individual nutrients due to individual level physiology and the presence of absorption enhancers or inhibitors due to other foods consumed ([55](#_ENREF_55)).

Despite these limitations, the methods used to develop the NDQS included actions to mitigate against some of the challenges described above, including reviewing existing, well-validated and widely cited indices and referring extensively to existing UK dietary reference values and population level intakes to inform the items included and the details of the scoring and weighting system.

The validity of the measure is supported by the results showing differences in NDQS quartile and mean scores between groups of different demographics, socio-economic status and lifestyle measures (Table 3.6). Many of the differences reflect the patterns and inequalities in diet quality described in Section 1.6.2 and are supported by evidence in the wider literature. When grouped by NSSEC, there was a clear inverse relationship between SES and NDQS. This reflects other studies in the UK showing that higher SES groups were more likely to consume higher quality diets than lower SES groups ([280](#_ENREF_280)). Also, while there were no significant differences in mean scores between ‘healthy weight’, ‘overweight’ and ‘obese’ individuals, those who were ‘underweight’ and ‘very obese’ had significantly lower scores than these groups indicating that the score is likely to be sensitive to extremes of energy intake and is somewhat indicative of unhealthy weight ranges that are very high and very low BMI. Individuals who smoked also scored significantly lower than both the general population mean and those who were non-smokers. This also reflects evidence in the literature suggesting that smokers have lower diet quality than non-smokers ([110](#_ENREF_110)). Analysis of the correlations between the NDQS and objective measures of nutrient intake derived from blood plasma and urine will provide further support for the score being a valid indicator of dietary quality.

**3.3. Objective 2. Evaluation of the content validity of the NDQS against nutrient biomarkers.**

The aim of this part of the study is to evaluate the content validity of the NDQS through exploring correlations with objectively measured (from blood or urine samples) nutrient biomarkers that can be used as positive and negative proxy indicators of dietary quality.

**3.3.1. Methods**

*3.3.1.1. Dataset*

As for Objective 1 described in 3.2 above, the dataset used was the NDNS (2008/09-2011/12). In addition to the food based variables used in the previous analyses, the dataset contains data for a number of biomarkers (more detail regarding biomarkers is provided in 3.3.2) derived from blood plasma and urine obtained from a sub-sample (n=1769). Whilst the blood and urine samples were taken at a different time from the collection of the dietary data, the food diary is assumed to be representative of habitual intake ([147](#_ENREF_147)) and therefore the biomarkers to be representative of ‘usual’ levels of the biomarker. Where there was evidence in the literature to support it, either because the biomarker was a direct measure of the intake levels of a particular nutrient or because the biomarker had been used in previous studies as an indicator of dietary quality, biomarkers were used to evaluate the content validity of the NDQS.

*3.3.1.2. Statistical analysis*

Correlation analysis was undertaken to calculate coefficients between the NDQS and a number of biomarkers available in the NDNS 2008-2012 dataset. Correlation was the simplest and most appropriate method as the aim was to simply show the amount of variance shared by two variables as opposed to using the findings for any prediction. The Spearman’s rho correlation analysis method was used because some nutrients were not normally distributed.

*3.3.1.3. Biomarker selection and interpretation*

Biomarkers are indicators of nutritional status with respect to intake or metabolism of nutrients ([281](#_ENREF_281)). For some micronutrients, status can be directly measured from the concentration of the nutrient in the blood whereas for others it is derived from a functional measure, for example by the degree of activation of vitamin dependent enzymes. Other biomarkers are indicators of health outcomes or disease risk for example circulating concentrations of serum total cholesterol and Low Density Lipoprotein cholesterol are predictive of Coronary Heart Disease ([147](#_ENREF_147)).

The biomarkers in the NDNS dataset were derived from analysis of blood plasma and urine samples that were taken from a sub-sample of consenting participants two to eight weeks after the initial nurse visit. Whilst the specimens were not taken during the same period as the dietary intake diary was taken, the biomarkers were taken on the assumption of habitual intake of dietary patterns ([147](#_ENREF_147)) and were therefore assumed to be representative of habitual intake for the purposes of this thesis and therefore appropriate for validating the NDQS.

Whilst for some nutrients there was an absolute optimal blood plasma level or threshold ([147](#_ENREF_147)) a broad approach that assessed the direction and significance of the correlation between the NDQS and the blood analyte in relation to the nature of the analyte itself and its implications for diet quality and health evidenced in the academic literature, was considered appropriate for the purposes of this study.

Table 3.7 shows the biomarkers selected from those available in the dataset and the rationale for their inclusion in the validation study. Some of the analytes have been directly measured from blood plasma concentration levels (Vitamin C, Vitamin B12, zinc) and others are derived through the measurement of another substance (for example Vitamin B6 from concentrations of Pyridoxal-5-Phosphate, ferritin as an indicator of iron stores). Some of these micronutrients are also useful indicators of intake of particular types of food associated or dietary quality for example studies have shown plasma carotenoids to be indicative of fruit and vegetable intake and high quality dietary patterns more generally ([282](#_ENREF_282), [283](#_ENREF_283)).

**Table 3.7. Biomarkers selected for inclusion in validation study of NDQS, their function and the hypothesised direction of correlation with NDQS.**

|  |  |  |
| --- | --- | --- |
| Biomarker | Hypothesised direction of correlation with NDQS | Function of biomarker/Rationale for inclusion as an indicator of diet quality |
| Zinc (µmol/L) | Positive | Zinc has an important role in the regulation of some enzymes and essential to the functioning of the metabolism ([284](#_ENREF_284)). |
| Selenium (µmol/L) | Positive | Selenium is essential to the metabolism including antioxidant systems and the functioning of the thyroid ([147](#_ENREF_147)). |
| Beta-Carotene (µmol/L) | Positive | Studies have shown that carotenoids are beneficial  to human health including reducing the risk of certain cancers and eye diseases ([285](#_ENREF_285)). Studies have shown that plasma concentrations of α-carotene, β-carotene, lycopene and lutein can be used to assess levels of intake of fruit and vegetables ([264](#_ENREF_264), [286](#_ENREF_286)). |
| Alpha-Carotene (µmol/L) | Positive |
| Lycopene (µmol/L) | Positive |
| Beta-Cryptoxanthin (µmol/L) | Positive |
| Alpha-Cryptoxanthin (µmol/L) | Positive |
| Lutein (µmol/L) | Positive |
| Total carotenoids (µmol/L) | Positive |
| Alpha-Tocopherol (µmol/L) | Positive | α-tocopherol plasma concentration can be used as a measure of vitamin E status ([147](#_ENREF_147)). Vitamin E is an antioxidant and is important in limiting radical damage resulting from oxidation of polyunsaturated fatty acids ([284](#_ENREF_284)). |
| Gamma-Tocopherol (µmol/L) | Positive | β-tocopherol plasma concentrations can be used as a measure of vitamin E status ([287](#_ENREF_287)). |
| Retinol (µmol/L) | Positive | Plasma retinol is an indicator of dietary intake of vitamin A ([147](#_ENREF_147)). |
| Vitamin B6 Pyridoxal-5-Phosphate (PLP) (nmol/L) | Positive | Vitamin B6 is needed for protein synthesis. PLP is the biologically active form. |
| Vitamin B1 Status (ETKAC) | Negative | Vitamin B1 (or thiamine) is important in maintaining the nervous system and metabolising carbohydrates. Thiamine status is measured by Erythrocyte Transketolase Activation Coefficient (ETKAC) where the higher the ETKAC the more likely there is to be a deficiency. |
| Vitamin B2 Status (EGRAC) | Negative (the higher the EGRAC the more likely there is to be deficiency) | Vitamin B2 (or riboflavin) has a central role in yielding energy from food. Riboflavin status is measured by Erythrocyte Glutathione Reductase Activation Coefficient (EGRAC) where the higher the EGRAC the more likely there is to be a deficiency ([147](#_ENREF_147), [284](#_ENREF_284)). |
| Vitamin C (µmol/L) | Positive | Vitamin C supports the maintenance of healthy connective tissues and can act as an antioxidant ([147](#_ENREF_147)) and is correlated with increased consumption of fruit and vegetables |
| Homocysteine (µmol/L) | Negative | Increased levels of homocysteine are a risk factor for cardiovascular diseases ([288](#_ENREF_288)). |
| 25-Hydroxy vitamin D (nmol/L) | Positive | Plasma 25-hydroxy D is a measure of vitamin D status. It indicates the availability of vitamin D from both dietary sources and sunlight ([289](#_ENREF_289)). |
| Ferritin (g/L) | Positive | The concentration of plasma ferritin is an indicator of the level of total iron stores in the body (in the absence of liver disease and inflammation) ([290](#_ENREF_290)). |
| Vitamin B12 (pmol/L) | Positive | Vitamin B12 has a key role in the functioning of the nervous system and brain. It is also important in blood cell formation ([147](#_ENREF_147)). |
| Low Density Lipoproteins (mmol/L) from fasted blood sample | Negative | High concentrations of circulating serum total cholesterol and LDL cholesterol are predictive of coronary heart disease and other vascular diseases ([147](#_ENREF_147), [291](#_ENREF_291)). |
|  |  |
| Total Cholesterol (mmol/L)Triglycerides (mmol/L) from fasted blood sample | Negative |
| C-Reactive Protein (mg/L) | Negative | C-Reactive Protein is a biomarker and mediator of inflammation and cardiovascular diseases ([292](#_ENREF_292)). |
| Urinary sodium (mmol/L) | Negative | Urinary sodium can be used to measure intake of salt. Increased intake of sodium is associated with risk of hypertension, stroke and coronary heart disease ([269](#_ENREF_269)). |

**3.3.2. Results**

Table 3.8 shows the results of the correlation analysis between the NDQS and biomarkers from serum plasma and urine samples. Many of the biomarkers were correlated significantly with the NDQS in the hypothesised direction to provide some validity for an increasing score at individual level on the NDQS being associated with improving dietary quality. The NDQS was significantly (p<0.05) positively associated with blood plasma levels of selenium, beta-carotene, alpha-carotene, lycopene, beta-cryptoxanthin, alpha-cryptoxanthin, lutein, alpha-tocopherol, total carotenoids, vitamins B6, C and D. The NDQS was significantly negatively associated with blood plasma levels of ETKAC and EGRAC (indicative of a lower risk of vitamins B1 and B2 deficiency respectively), homocysteine, c-reative protein and urinary sodium. There was no significant correlation between the score and either of the measures of cholesterol, vitamin B12, ferritin, retinol, gamma-tocopherol and zinc.

**Table 3.8. Correlations between NDQS and nutrient biomarkers**

|  |  |  |  |
| --- | --- | --- | --- |
| Nutrient | Correlation coefficient | P | n |
| Zinc (µmol/L) | 0.04 | 0.25 | 953 |
| Selenium (µmol/L) | 0.26\*\*\* | <0.001 | 954 |
| Beta-Carotene (µmol/L) | 0.21\*\*\* | <0.001 | 984 |
| Alpha-Carotene (µmol/L) | 0.20\*\*\* | <0.001 | 984 |
| Lycopene (µmol/L) | 0.06\* | 0.05 | 984 |
| Beta-Cryptoxanthin (µmol/L) | 0.23\*\*\* | <0.001 | 984 |
| Alpha-Cryptoxanthin (µmol/L) | 0.11\*\* | 0.001 | 984 |
| Lutein (µmol/L) | 0.19\*\* | <0.001 | 984 |
| Gamma-Tocopherol (µmol/L) | -0.06 | 0.08 | 986 |
| Alpha-Tocopherol (µmol/L) | 0.11\*\* | 0.001 | 986 |
| Total carotenoids (µmol/L) | 0.22\*\*\* | <0.001 | 985 |
| Retinol (µmol/L) | 0.03 | 0.34 | 973 |
| Vitamin B6 Pyridoxal-5-Phosphate (PLP) (nmol/L) | 0.19\*\*\* | <0.001 | 964 |
| Vitamin B1 Status (ETKAC) | -0.12\*\*\* | <0.001 | 927 |
| Vitamin B2 Status (EGRAC) | -0.20\*\*\* | <0.001 | 930 |
| Vitamin C (µmol/L) | 0.22\*\*\* | <0.001 | 924 |
| Homocysteine (µmol/L) | -0.09\*\* | 0.01 | 989 |
| 25-Hydroxy Vitamin D (nmol/L) | 0.12\*\* | <0.001 | 989 |
| Ferritin (g/L) | 0.03 | 0.36 | 997 |
| Vitamin B12 (pmol/L) | 0.06 | 0.08 | 990 |
| Low Density Lipoproteins (mmol/L) | 0.05 | 0.13 | 994 |
| Total Cholesterol (mmol/L) | 0.03 | 0.32 | 1016 |
| C-Reactive Protein (mg/L) | -0.11\*\* | <0.001 | 1017 |
| Urinary sodium (mmol/L) | -0.09\*\* | 0.001 | 1269 |

**3.3.3. Discussion**

In order to evaluate the content validity of the NDQS, correlation analyses was undertaken with a number of nutrient biomarkers available at individual level in the NDNS dataset. Correlation analysis was conducted as the aim was simply to investigate the amount of variance shared by two variables as opposed to using the findings for any prediction.

There are limitations to this correlation analysis as a content validity study for the NDQS. For several of the biomarkers, there are factors wider than dietary intake alone that are relevant and influence the nutritional status of an individual that were not adjusted for in this study. For example, vitamin D (and 25-Hyrdroxy vitamin D) is also synthesised through sunlight as well as consumed in the diet, therefore the correlation with the NDQS, suggesting that the diet is sufficient in vitamin D rich foods may be misleading. Serum carotenoid levels are influenced by confounders such as age and lifestyle factors such as BMI and smoking status ([282](#_ENREF_282)).

There are more biomarkers included in the validity study that are positive for health than negative, which potentially presents a skewed picture of the validity of the NDQS, particularly as the two cholesterol based biomarkers did not show a significant correlation. In addition there was no weighting applied to the specific biomarkers to account for between-individual or within-individual variability, or lack thereof in the case of micronutrients that are homeostatically regulated or to account for more or less importance in terms of contribution to dietary quality or health outcomes. By using a correlation method, the evaluation was crude and did not take account of specific thresholds where they existed for individual biomarkers. Regression analyses adjusting for important confounders, particularly those that effect the absorption and plasma levels of some nutrients, such as smoking, would have resulted in more robust findings and potentially added validity to the final NDQS.

The biomarker samples were also taken approximately eight weeks after dietary intake data were collected. However, as stated above, for the purposes of this thesis and other population level estimates published from analyses of this dataset both the dietary and nutrient intake data derived from the self-reported diaries and the biomarker indicators of nutritional status are assumed to be representative of habitual intake and nutritional status ([147](#_ENREF_147), [193](#_ENREF_193), [293](#_ENREF_293)).

Total energy was not included as an independent item in the NDQS and was only taken account of in the items that were relative measures to be moderated, i.e. proportion of total calorie intake for saturated fat, trans fats, polyunsaturated fats and NMES. Whilst this will mitigate against the confounding effects of increased total energy to an extent, there may still be some effect from total energy increasing the likelihood of having optimal levels of nutrients ([274](#_ENREF_274)). The analysis might have been improved by undertaking regression analysis with the directly measured biomarkers rather than correlation and including total energy in the model.

However, this relatively crude method of using biomarkers from blood and urine specimens from the same study sample has been widely used in other studies to validate both data-driven dietary patterns ([282](#_ENREF_282), [294](#_ENREF_294)) and theoretically driven healthy eating and dietary quality indices ([218](#_ENREF_218), [254](#_ENREF_254)). It can be concluded from the analysis that the NDQS is a robust measure by which to assess dietary quality in the UK population.

**3.4. Objective 3. Exploration of the associations between the NDQS and socio-demographic and lifestyle variables.**

The aim of this part of the study is to assess the generalisability of the NDQS through exploring associations with socio-demographic and lifestyle variables.

**3.4.1. Methods**

*3.4.1.1. Dataset*

NDNS 2008/09-2011/12 dataset was used as per previous analyses undertaken in this chapter. As previously described, a NDQS score as per the criteria described in Section 3.2.1 had been calculated and assigned for each individual in the dataset.

*3.4.1.2. Statistical analyses*

Main effects regression models were used to examine the associations between the NDQS and the following variables: age, sex, ethnicity (white/non-white), smoking status, total energy (kcal) and household income (equivalised for numbers in household). Quadratic plots (Figures 3.3. and 3.4) showed evidence of curvilinear effects for BMI and total energy therefore the squared values of these variables were included in the model. All analyses were conducted in SPSS version 22.

**3.4.2. Results**

The regression model (Table 3.8) shows that the NDQS was significantly (p<0.001) positively associated with age and negatively associated with smoking and being white. For each additional year older someone is, the expected increase in NDQS, all else being equal, is 0.16. On average, when all else remains equal, the score is 3.48 lower for white individuals compared with non-white and 5.85 lower for smokers compared with non-smokers. While the direction of the coefficient appears to change as might be expected for the high to low categories of socio-economic status (as measured by NSSEC), none of the associations were significant, therefore the direction of these associations may be due to chance. There were curvilinear relationships between the NDQS and BMI (not significant) and total energy intake (see Figures 3.3 and 3.4). The plots for these regressions shows a similar relationship between each of these variables and the NDQS in that there is a positive relationship between the score and both BMI and total energy to an optimal point after which the association becomes negative.

**Table 3.9. Main effects regression model (95% confidence intervals) of NDQS for gender, ethnicity, socio-economic status, income, total energy, BMI and smoking status (NDNS 2008/09-2011/12)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable (n) | Coefficient | P-value | 95% Confidence Interval | |
| Lower | Upper |
| *Gender* |  |  |  |  |
| Males (711) | 0.80 | 0.241 | -0.54 | 2.13 |
| Females (914) | Reference | | | |
| *Ethnicity* |  |  |  |  |
| White (1495) | -3.48 | 0.001 | -5.63 | -1.34 |
| Non-white (130) | Reference | | | |
| *Smoking status* |  |  |  |  |
| Smoker (377) | -5.85 | <0.001 | -7.27 | -4.44 |
| Non-smoker (130) | Reference | | | |
| *Socio-economic status* |  |  |  |  |
| Higher managerial and professional occupations (260) | 1.75 | 0.43 | -2.59 | 6.08 |
| Lower managerial and professional occupations (453) | 1.20 | 0.57 | -3.00 | 5.38 |
| Intermediate occupations (149) | 0.57 | 0.80 | -3.86 | 5.00 |
| Small employers and own account workers (165) | 0.60 | 0.79 | -3.79 | 5.00 |
| Lower supervisory and technical occupations (160) | 1.15 | 0.61 | -3.26 | 5.55 |
| Semi-routine occupations (228) | 0.17 | 0.93 | -4.10 | 4.46 |
| Routine occupations (177) | -2.36 | 0.29 | -6.73 | 2.00 |
| Never worked (33) | Reference | | | |
| Total Energy (kcal) (2083) | 0.02 | <0.001 | 0.01 | 0.02 |
| Total Energy2 (kcal) | -3.621E-06 | <0.001 | -4.640E-06 | -2.603E-06 |
| BMI (1902) | 0.35 | 0.43 | -0.52 | 1.21 |
| BMI2 | -0.01 | 0.36 | -0.02 | 0.01 |
| Equivalised household income (1777) | 2.625E-05 | 0.06 | -1.149E-06 | 5.364E-05 |
| Age (2083) | 0.16 | <0.001 | 0.12 | 0.20 |

**Figure 3.3. Quadratic plot of BMI and NDQS regression**

**Figure 3.4. Quadratic plot of total energy (kcal) and NDQS regression**

**3.4.3. Discussion (objective 3)**

Main effects regression analyses of the NDQS with demographic, socio-economic and lifestyle variables support the validity of the score as they indicate that the score is associated as might expected from previous evidence in the literature (as described in Section 1.6.2) in particular in relation to smoking and diet ([187](#_ENREF_187)). Individuals that were smokers were significantly likely to have a score that was, on average, 5.85 less than non-smokers. The findings also showed that ethnicity and age were significantly associated with the score, with those of white ethnicity being likely to have a score 3.48 lower than those of non-white ethnicity and each year of age meaning a likely increase in NDQS of 0.16.

Associations with socio-economic status and household income were not significant. This latter result does not reflect evidence in the literature which suggests a positive association between diet quality and socio-economic status or social class ([260](#_ENREF_260)). However, evidence from the UK tends to be limited to analyses of single foods and nutrients ([116](#_ENREF_116)) as opposed to a composite measure of dietary quality and as noted previously in Section 1.6.2. the relationship between different socio-economic measures and dietary intake varies with the measure and by gender ([114](#_ENREF_114)).

The curvilinear relationship with total energy and BMI, is also as might have been expected as total energy was not included as an independent item in the NDQS. The plots suggest that for both BMI and total energy an increase in the latter variables is associated with a higher diet quality score to an optimal point (the optimal BMI for NDQS is 29.8) beyond which the diet quality score reduces as the BMI and total energy continues to increase. The breakdown of mean score by BMI category in Section 3.2.2.3 also reflects this. This relationship may also be due to under-reporting in the NDNS dataset.

**Chapter 4.**

**An evaluation of indicator variables for diet quality derived using empirically driven and theory driven methods (Study 3)**

**4.1. Introduction, aims and objectives**

The aim of this chapter is to assess whether the indicator variables generated from the empirically driven dietary patterns analyses in Chapter 2, were predictive of a nutrient-based Diet Quality Score (NDQS) based on *a priori* nutritional knowledge and to compare their ability to predict diet quality with indicators derived from a theory-driven method. The ultimate aim was to identify the most parsimonious model for predicting diet quality to inform the development of a brief diet quality assessment tool. The findings from these analyses will be summarised and the conclusions and limitations will be discussed in the context of the research questions.

As described in Chapters 1 and 2, a number of methods can be used to reveal the underlying patterns in the diet of a sample or population. These methods can be theory driven, empirically driven or a mixture of both. In Study 1, presented in Chapter 2, PCA was used to identify data-driven ‘dietary patterns’ that were underlying in the dietary intake of the UK population. Food and food groups with moderate to strong positive or negative factor loadings on each pattern were described as ‘characterising’ each pattern or typology of diet and the patterns were subjectively labelled according to these foods as representing diets of higher and lower quality. Regression analyses was used to explore associations with demographic, socio-economic and lifestyle variables to identify additional indicator variables that may be useful predictors of diet quality. This analyses was also used to evaluate whether the patterns reflected variance that would be expected in the UK populations’ diet according to socio-economic measures and lifestyle/behavioural factors. Some of the foods that characterised the data-driven patterns were included in the YHS-DQ.

As described in Chapter 1, at the time of investigation, there was no standard, validated measure of nutrient-based diet quality in the UK. Study 2 in Chapter 3 described the theory based methods used to define diet quality in the UK, drawing on *a priori* nutritional knowledge summarised in the UK DRVs ([239](#_ENREF_239)). The content validity of the NDQS was evaluated through correlation analyses with a range of nutrient biomarkers in the NDNS dataset. The distribution of the score in the NDNS including associations with demographic, socio-economic and lifestyle variables, was also investigated.

Study 3 in Chapter 4 describes the next stage in the thesis, addressing the research questions and objectives below.

**Study 3**

**Research questions**

1. Are the empirically-derived dietary patterns identified in Study 1 associated with diet quality (as defined by the NDQS)?
2. Which indicator variables (foods and wider characteristics) are most predictive of diet quality?
3. How do empirical and theory driven methods compare in terms of their ability to generate models that are predictive of diet quality?
4. Which is the most parsimonious model that is predictive of diet quality (NDQS)?

**Objectives**

1. To explore whether the empirically-derived dietary patterns are associated with a theory driven definition of diet quality (NDQS).

2. To identify indicator variables predictive of diet quality using a theory driven method (NDQS) and compare these with the indicator variables identified using empirically driven methods.

3. To identify the most parsimonious model that is predictive of diet quality on which to base the brief diet quality assessment tool(s).

**4.2. Objective 1. Exploration of associations between empirically-derived dietary patterns and diet quality (NDQS).**

The aim of this part of the study is to explore whether the empirically-derived dietary patterns identified in Study 1 are predictive of diet quality (NDQS) both unadjusted and when adjusted for age, gender, ethnicity, NSSEC, household income, smoking status, BMI and total energy intake.

**4.2.1 Methods**

*4.2.1.1. Dataset*

As for the previous studies, the data used for these analyses were Years 1-4 (2008/09-2011/12) from the UK NDNS. Table 2.1 in Study 1 describes the 60 foods and food group categories that were included in the PCA used to derive the dietary patterns. The food variables that were identified as being moderately or strongly associated with the four dietary patterns generated from the 2008-2011 and the 2008-2012 dataset are shown in Table 4.1. Foods shown in italics were negatively associated with the dietary pattern.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4.1. Food variables with moderate/high factor loadings for each dietary pattern\*** | | | |
| Fruit, vegetables and oily fish | Snacks, fast food and fizzy drinks | Sugary foods and dairy | Meat, potatoes and beer |
| **Fruit**  **Salad and other raw vegetables**  **Oily fish**  Wine  Yoghurt, fromage frais and dairy deserts  Nuts and seeds | **Soft drinks (not low calorie)**  **Crisps and savoury snacks**  **Coated chicken and turkey**  Burgers and kebabs  Chocolate confectionery  **Chips (fried), roast potatoes and potato products**  **White bread**  Sugar confectionery | Sugars, preserves and sweet spreads  **Buns, cakes, pastries and fruit pies**  **Biscuits**  Chocolate confectionery  Semi skimmed milk | Other potatoes, potato salads and dishes  **Bacon and ham**  Pork and dishes  Savoury sauces, pickles, gravies and condiments  **White bread**  **Beer, lager, cider and perry**  Whole milk  Butter |
| ***Chips (fried), roast potatoes and potato products***  ***White bread***  *Sugar, preserves and sweet spreads* | *Tea, coffee and water*  *High fibre breakfast cereals*  ***Fruit***  ***Vegetables not raw*** | *Wine* |  |

\**Food variables in italics had negative factor loadings;* **Foods shown in bold were included in the YHS-DQ**

* + - 1. *Data analysis*

Associations between the dietary patterns and NDQS score were investigated using regression analyses. The analysis was carried out in years 1-4 (2008/09-2011/12), combined, of the data. Regression models were used to calculate the crude and adjusted coefficients (main effects model adjusted for total energy, age, gender, ethnicity, smoking status, BMI, socio-economic status and household income). Plots of residuals were visually assessed for evidence of homoscedasticity, constancy of variance and outliers and these conditions of the model were considered to be satisfied (Figures B.3.1-B.3.4, Appendix B). Where covariates were not normally distributed (BMI, total energy and household income), linear regressions including the variable and its square with the NDQS were carried out to test for evidence of curvilinear effects. Where the coefficient of the squared term was significant, the coefficients for both the value and the squared values were plotted to visually assess for potential curvilinear (quadratic) effects (Figures C.2.1-C.2.3, Appendix C).

* + 1. **Results**

Table 4.2 summarises the results of the unadjusted regression for any other variables. All patterns were significantly predictive of the NDQS. The SFFFD and MPB patterns were negatively associated with the NDQS. The SFFFD pattern was more strongly inversely predictive of the NDQS than the MPB pattern with, when all things remained equal, an expected decrease in NDQS score of 3.7 with every incremental increase in SFFFD factor score, compared with an expected decrease in NDQS score of 1.2 with every increase in MPB factor score. The ‘FVOF’ and the SFD patterns were positively associated with the NDQS with the FVOF pattern being the more strongly positively predictive of the NDQS than the SFD. For every incremental increase in FVOF score, the expected NDQS increase would be 4.5 and for SFD it would be 2.4.

**Table 4.2. Main effects regression model (95% confidence intervals), NDQS and dietary patterns**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dietary Pattern | Coefficient | P value | 95% Confidence Interval | |
| Lower | Upper |
| Snacks, Fast Food and Fizzy Drinks (SFFFD) | -3.727 | <0.001 | -4.205 | -3.249 |
| Fruit, Vegetables and Oily Fish (FVOF) | 4.514 | <0.001 | 4.037 | 4.992 |
| Meat, Potatoes and Beer (MPB) | -1.183 | <0.001 | -1.660 | -0.705 |
| Sugary Foods and Dairy (SFD) | 2.425 | <0.001 | 1.948 | 2.903 |

Table 4.3 shows the results of the regression model adjusted for age, gender, SES, household income, ethnicity, smoking status, BMI and total energy intake. For BMI and total energy intake, the terms and their squares were included in the model due to the curvilinear relationships between these variables and the NDQS (Figures C.2.1 and C.2.3). The adjusted model did not result in any significant changes in the directions or strength of the relationships between the data-driven dietary patterns and the NDQS.

When adjusted for gender, smoking status, ethnicity, SES, household income, BMI and total energy the SFFFD and MPB remained negatively predictive of the NDQS. The effect was attenuated slightly in the SFFFD pattern compared with the unadjusted model. For each incremental increase in factor score for the pattern, all else being equal, the NDQS was expected to decrease by 3.3 compared with 3.7 in the unadjusted model. The effect of the MPB pattern on the NDQS was strengthened in the adjusted model, with an expected 1.8 decrease on the NDQS compared with 1.2 in the unadjusted model.

The FVOF and the SFD patterns remained positively predictive of the NDQS with the effects of both patterns being slightly attenuated in the adjusted model (with expected increases in NDQS of 3.8 and 1.3 respectively compared with 4.5 and 2.4 in the unadjusted model).

**Table 4.3. Main effects regression model (95% confidence intervals), NDQS and dietary patterns for demographic characteristics, socio-economic measures and lifestyle factors.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Independent variable | Coefficient | P value | 95% Confidence Interval | |
| Lower | Upper |
| Snacks, Fast Food and Fizzy Drinks | -3.31 | <0.001 | -3.97 | -2.64 |
| Fruit, Vegetables and Oily Fish | 3.79 | <0.001 | 3.19 | 4.40 |
| Meat, Potatoes and Beer | -1.76 | <0.001 | -2.43 | -1.09 |
| Sugary Foods and Dairy | 1.31 | <0.001 | 0.69 | 1.93 |
| Male | 2.71 | <0.001 | 1.46 | 3.95 |
| Female | Reference | | | |
| Age | 0.06 | 0.001 | .026 | 0.10 |
| Smoker | -3.30 | <0.001 | -4.63 | -1.97 |
| Non-smoker | Reference | | | |
| White | -1.92 | 0.06 | -3.92 | 0.08 |
| Non-white | Reference | | | |
| Higher managerial and professional occupations | 0.68 | 0.74 | -3.28 | 4.64 |
| Lower managerial and professional occupations | 0.76 | 0.70 | -3.05 | 4.58 |
| Intermediate occupations | 0.98 | 0.64 | -3.07 | 5.02 |
| Small employers and own account workers | 0.88 | 0.67 | -3.13 | 4.90 |
| Lower supervisory and technical occupations | 1.65 | 0.42 | -2.38 | 5.68 |
| Semi-routine occupations | 1.20 | 0.55 | -2.72 | 5.11 |
| Routine occupations | -0.33 | 0.87 | -4.32 | 3.67 |
| Never worked | Reference | | | |
| Household income | 3.524E-06 | 0.79 | -2.184E-05 | 2.889E-05 |
| BMI | 0.44 | 0.28 | -0.35 | 1.23 |
| BMI Squared | -0.01 | 0.28 | -0.02 | 0.01 |
| Total Energy (kcal) Squared | -3.137E-06 | <0.001 | -4.088E-06 | -2.185E-06 |
| Total Energy (kcal) | 0.01 | <0.001 | 0.01 | 0.02 |

* + 1. **Discussion (Objective 1)**

Regression analyses were undertaken to explore the relationship between the empirically-derived dietary patterns and the NDQS. Analyses were conducted both unadjusted and adjusted for demographic, socio-economic and lifestyle variables. The findings showed that all of the dietary patterns were significantly predictive of the NDQS. The SFFFD and MPB were inversely associated with the score and the FVOF and SFD patterns were positively associated. With all else being equal, the effect was attenuated for all patterns in the adjusted model apart from the MPB pattern, where the effect was strengthened.

The relationship between the patterns and diet quality was in the direction as would have been expected for the SFFFD, FVOF and MPB patterns but was the inverse of what might have been expected for the SFD pattern. This latter pattern was positively associated with the NDQS in both the unadjusted and the adjusted model with the adjusted model attenuating the effect from 2.4 to 1.3. This result is surprising as the foods characterising this dietary pattern were predominantly high in sugar and saturated fat which would contribute towards a lower score on the NDQS. In addition, as described in Section 2.3.6, regression analyses with nutrient intakes (derived from self-reported food diary data) and nutrient biomarkers showed this pattern to be negatively associated with intake per 1000 calories of vitamins C, folate, proportion of food energy from n-3 and n-6 PUFA and positively associated with proportion of food energy from saturated fat, NMEs, total fat and NSP. These nutrients were included in the NDQS and aside from NSP the relationships described here would result in a lower NDQS score. The pattern was also negatively associated with biomarkers of vitamins C, D, A, retinol, ferritin and total carotenoids. One explanation might be that as this pattern explained the least amount of variance in the PCA (2.8%) that it is not representative of a larger population and is what has been described as a ‘rogue’ factor ([295](#_ENREF_295)). It is also possible that other elements of the NDQS such as low levels of sodium and alcohol and high levels of NSP which are elements of this pattern have outweighed some of the other elements that might have led to a lower score.

* 1. **Objective 2. Identification of indicator variables for diet quality using empirically driven and theory driven methods.**

The aim of this part of the study was to identify the indicator variables that were the most predictive of diet quality using a theory-driven method and to compare this with the indicator variables generated from the empirically-derived dietary patterns identified in Chapter 2 and used in the YHS-DQ.

**4.3.1. Method**

*4.3.1.1. Dataset*

The dataset used was the NDNS as described previously. Separate and combined years of the data that differ from those used in the analyses in Studies 1 and 2 and in Study 3, Objective 1, were used in the following analyses for the purposes of testing repeatability and validity of the findings.

Table 4.4 summarises which years of data were used for which analysis in the thesis and the associated rationale. All analyses were undertaken in SPSS version 22.

**Table 4.4. Summary of years of NDNS data used for analyses in the thesis and associated rationale.**

|  |  |  |  |
| --- | --- | --- | --- |
| Study &  Objective | Type of analysis | Data year(s) | Rationale |
| Study 1  Objective 1 & 3 | PCA  Regressions with dietary patterns and sample characteristics | Years 1-3 (2008/09 -2010/11) | All years of data available at time of investigation. Dataset with largest sample size was selected to increase stability of patterns and reduce chance of sampling error, misclassification of items and ‘splintered’ factors ([295](#_ENREF_295)). |
| Study 1  Objective 4 & 5 | PCA  Regressions with dietary patterns and sample characteristics, nutrient intake and nutrient biomarkers | Years 1-4  (2008/09- 2011/12) | Another year of data became available during investigation. As above and larger sample size to increase power of regression analysis. |
| Study 1  Objective 4 | PCA | 50% random sample from years 1-4 | A random split of the data was used to test the stability of the dietary patterns identified. |
| Study 3  Objective 1 | Regression with dietary patterns and NDQS | Years 1-4 combined (2008/09-2011/12) | The largest sample size was selected in order to optimise the power of the analysis. |
| Study 3  Objective 2 | Correlation analysis (food variables and NDQS) | Years 2 & 4 combined  (2009/10 & 2011/12) | The data were split into two halves for this study to enable confirmatory analysis, comparing the predictive ability of the models generated from both data-driven and theory-driven methods to be tested in a different cohort from that from which the theory driven models were generated. |
| Study 3  Objective 2 | Backwards elimination regression (to identify variables independently most predictive of NDQS) | Years 2 & 4 separately | Analysis was conducted in two separate years of data to reduce the sample size to identify the fewest and most predictive models predictive of the NDQS. Larger sample sizes using this method case can increase the number of retained models ([296](#_ENREF_296)). |
| Study 3  Objective 2 | Confirmatory analysis (to compare model generated from data-driven methods with models generated from theory driven methods). | Years 1 & 3 (2008/09 & 2010/2011) combined | A separate cohort was used for the confirmatory analysis as demonstrating the repeatability of the analysis (in relation to the predictive capacity of the models) strengthens the external validity of the findings. |

* + - 1. *Statistical Analysis*

1. *Correlation analysis with food variables and NDQS*

To explore the relationships between the food variables used in the initial PCA in Study 1, presented in Chapter 2 and the NDQS, correlation analyses was conducted. The analyses was conducted with each of the 60 food variables and the NDQS in years 2 and 4 combined of the NDNS dataset. The aim was also to investigate whether the particular foods that characterised the empirically-derived dietary patterns also correlate with higher and lower scores on the NDQS. For the majority of the food variables the distribution was not normal, therefore Spearman’s rho for non-parametric data was used.

1. *Backwards elimination regression*

To explore which of the food variables independently predict the most variance in the NDQS, backwards elimination regression was undertaken, again with the 60 food group variables from the NDNS dataset that were included in the PCA presented in Chapter 2. Backwards elimination regression was undertaken firstly with the food variables only and then with the food variables plus the wider demographic and lifestyle variables. The wider variables that were included were only those that were either continuous or binary variables and were therefore suitable for inclusion in the regression model and those that were considered feasible for data collection using a brief self-reported tool (sex, age, BMI and smoking status).

Backwards elimination regression was conducted on two separate years of data (year 2 and year 4) in order to reduce the sample size and identify the fewest number of independent predictor variables in each year ([296](#_ENREF_296)). The variables common to both years were then taken forwards in the next stage of analysis. The entry p value was set at 0.001 and the removal p value at 0.01 in order to reduce the resulting list of coefficients to the strongest independent predictors of the NDQS.

Evidence of curvilinear effects was tested by calculating the square of each food variable and conducting regression analyses with both the food variable and its squared term as predictors of the NDQS. Evidence of curvilinear effects was assumed when the effect with the squared term was significant to a level of <0.01 or the squared term is significant and the raw correlation with the original food variable is not. Where this was the case, visual assessment of quadratic plots confirmed that there was no evidence of curvilinear effects for any of the food variables that were considered significant enough to effect the model (Figures C.3 and C.4, Appendix C).

1. *Confirmatory analysis (indicator variables of diet quality derived from the data-driven dietary patterns as included in the YHS-DQ).*

Confirmatory analysis of the indicator variables of diet quality derived from empirically driven dietary patterns was undertaken by comparing the ‘model fit’ or how predictive of diet quality (as defined by the NDQS) it was with the final models generated from the backwards elimination regression in (ii) above in a different cohort of the NDNS (as summarised in Table 4.4). Regression analyses were conducted exclusive and inclusive of the wider demographic and lifestyle related variables as in (iii) above. All analyses were undertaken in SPSS version 22.

As described in Section 2.3.7, in some cases the exact definition of the food variables included in the YHS-DQ was slightly amended from the wording of the food groups defined by the NDNS dataset. This was in order to improve the sense for the respondent and due to lack of physical space on the questionnaire, e.g. ‘chips, fried potatoes and potato products became ‘chips’; ‘biscuits’ and ‘buns, cakes, pastries and fruit pies’ became ‘biscuits, cakes and pastries’; ‘soft drinks not low calorie’ became ‘sugary drinks (‘fizzy pop, squash’). In addition a separate question was asked regarding portions of fruit and vegetables consumed on a typical day separately from the FFQ in order to enable comparison with data collected nationally through the Active People Survey. Therefore for the purposes of the confirmatory analysis, the food categories from the NDNS dataset included in the regression were those based on the original PCA results described in Tables 2.6-2.10. The additional two fruit and vegetables questions were considered to be reflected in the main food group categories relating to fruit and vegetables listed below. The variables included were: white bread; bacon or ham; buns, cakes, pastries and fruit pies; vegetables not raw; coated chicken and turkey; wholemeal bread; beer, lager, cider and perry; oily fish; chips, fried potatoes and potato based products; biscuits; soft drinks not low calorie; fruit; crisps and savoury snacks and salad and other raw vegetables.

**4.3.2. Results**

*4.3.2.1. Correlation analysis with 60 food variables and NDQS (Year 2 and 4 NDNS combined)*

Table 4.4 shows that the NDQS was statistically significantly (p<0.05) correlated in the direction that might be hypothesised with several of the foods that characterised the data-driven dietary patterns. Using Cohen’s effect size recommendations for interpreting the size of the coefficient ([297](#_ENREF_297)) the strength of the associations were most frequently ‘negligible’ (<0.30) or ‘low’ (0.30 to 0.50). There were a few associations categorised as ‘moderate’ and no associations classed as ‘high’ or ‘very high’ in strength. For the purposes of this thesis, correlation coefficients above 0.10 with p<0.05 are considered relevant.

Food group variables that were significantly positively associated with the NDQS with a coefficient >0.10 were: ‘semi skimmed milk’, ‘wholemeal bread’, ‘tea, coffee and water’, ‘fruit’, ‘high fibre breakfast cereals’, ‘nuts and seeds’, ‘oily fish’, ‘salad and other raw vegetables’, ‘potatoes, potato salad and potato dishes’, There were two food groups where the direction of the relationship was not as might have been expected. ‘Buns, cakes, pastries and fruit pies’ and ‘puddings’ were significantly, positively associated with the NDQS. Both these categories contain foods that are high in added sugar and/or high in saturated fat which would score lower on the NDQS for these items.

**Table 4.5. Spearman’s rho correlation coefficients between food variables and the NDQS (NDNS 2008 and 2011, n=1075)**

|  |  |  |
| --- | --- | --- |
| Food variable | Correlation coefficient (Spearman’s Rho) | P value (2 tailed) |
| One percent milk | -0.03 | 0.327 |
| Beef, veal and dishes | 0.10 | 0.001\*\* |
| Butter | -0.05 | 0.88 |
| Other margarine, fats and oils | 0.06 | 0.05\* |
| Other milk and cream | 0.05 | 0.09 |
| PUFA margarine and oils | 0.06 | 0.05\* |
| Semi skimmed milk | 0.12 | <0.001\*\*\* |
| *Whole milk* | -0.16 | <0.001\*\*\* |
| Wholemeal bread | 0.26 | <0.001\*\*\* |
| Ice cream | 0.03 | 0.39 |
| White fish coated or fried | 0.06 | 0.04\* |
| *Beer, lager, cider and perry* | -0.17 | <0.001\*\*\* |
| *Crisps and savoury snacks* | -0.14 | <0.001\*\*\* |
| Fruit juice including smoothies | 0.12 | <0.001\*\*\* |
| Soft drinks low calorie | -0.05 | 0.10 |
| *Soft drinks not low calorie* | -0.17 | <0.001\*\*\* |
| *Spirits and liqueurs* | -0.12 | <0.001\*\* |
| Sugar confectionery | -0.03 | 0.06 |
| Tea, coffee and water | 0.24 | <0.001\*\*\* |
| Wine | 0.01 | 0.78 |
| Bacon and ham | -0.09 | 0.004\*\* |
| Biscuits | 0.09 | 0.06 |
| Buns, cakes, pastries and fruit pies | 0.10 | 0.002\*\* |
| *Burgers and kebabs* | -0.12 | <0.001\*\*\* |
| Cheese | 0.03 | 0.40 |
| Chips, fried and roast potatoes, potato products | -0.05 | 0.09 |
| Chocolate confectionery | -0.04 | 0.18 |
| *Coated chicken and turkey* | *-0.09* | *0.004\*\** |
| Eggs and egg dishes | 0.02 | 0.60 |
| Fruit | 0.51 | <0.001\*\*\* |
| High fibre breakfast cereals | 0.38 | <0.001\*\*\* |
| Lamb and lamb dishes | -0.02 | 0.58 |
| Liver and dishes | -0.01 | 0.73 |
| *Meat pies and pastries* | *-0.14* | *<0.001\*\** |
| Nuts and seeds | 0.14 | <0.001\*\* |
| Oily fish | 0.17 | <0.001\*\*\* |
| Other bread | 0.03 | 0.26 |
| Other meat and meat products | 0.003 | 0.91 |
| Pasta, rice and other cereals | 0.08 | 0.01 |
| Pork and dishes | -0.002 | 0.95 |
| Salad and other raw vegetables | 0.26 | <0.001\*\*\* |
| *Sausages* | *-0.07* | *0.02\** |
| *White bread* | *-0.25* | *<0.001\*\*\** |
| Yoghurt, fromage frais and dairy deserts | 0.27 | <0.001\*\*\* |
| Other breakfast cereals | 0.06 | 0.04\* |
| Other potatoes, potato salads and dishes | 0.19 | <0.001\*\*\* |
| Other white fish, shellfish and fish dishes | 0.04 | 0.21 |
| Vegetables not raw | 0.31 | <0.001\*\*\* |
| Chicken and turkey dishes | 0.06 | 0.05 |
| Puddings | 0.12 | <0.001\*\*\* |
| Brown, granary and wheat germ bread | 0.13 | <0.001\*\*\* |
| Skimmed milk | 0.15 | <0.001\*\*\* |
| *Sugar preserves and sweet spreads* | *-0.13* | *<0.001\*\*\** |
| Dry weight beverages | 0.10 | 0.001\*\* |
| Low fat spread not polyunsaturated | 0.05 | 0.12 |
| Low fat spread polyunsaturated | 0.11 | <0.001\*\*\* |
| *Reduced fat spread not polyunsaturated* | *-0.07* | *0.02\** |
| Reduced fat spread polyunsaturated | 0.005 | 0.86 |
| Sauces, pickles and gravies | 0.06 | 0.045\* |
| Soup homemade and retail | 0.04 | 0.18 |

*Food variables with negative correlations with the NDQS are highlighted in Italics*

*4.3.2.2. Backwards elimination regression with NDQS*

*Food variables only (NDNS years 2 and 4 separately)*

Table 4.6 shows the results of the backwards elimination regression of all food variables with the NDQS in year 2 and year 4 NDNS datasets. There were 19 variables in the final model of the backwards elimination regression in year 2 and 21 variables in the final model of year 4. Foods highlighted in bold were observed in the final regression model in both datasets. A final list of 12 food variables were observed in the final regression model in both years of data: beef, veal and dishes; wholemeal bread; beer, lager, cider and perry; spirits and liqueurs; wine; chips, fried and roast potatoes and potato products; fruit; pasta, rice and other cereals; salad and other raw vegetables; vegetables (not raw); brown, granary and wheat germ bread and sugar, preserves and sweet spreads.

**Table 4.6. Food only indicator variables for the NDQS (Years 2 and 4 NDNS)\***

|  |  |
| --- | --- |
| Year 2 | Year 4 |
| **Beef, veal and dishes** | **Beef, veal and dishes** |
| **Wholemeal bread** | Butter |
| **Beer, lager, cider and perry** | Semi skimmed milk |
| Soft drinks not low calorie | **Wholemeal bread** |
| Soft drinks low calorie | **Beer, lager, cider and perry** |
| **Spirits and liqueurs** | **Spirits and liqueurs** |
| **Wine** | Tea, coffee and water |
| **Chips, fried and roast potatoes and potato products** | **Wine** |
| Eggs and egg dishes | **Chips, fried and roast potatoes and potato products** |
| **Fruit** | Chocolate confectionery |
| High fibre breakfast cereals | **Fruit** |
| **Pasta, rice and other cereals** | Other bread |
| **Salad and other raw vegetables** | **Pasta, rice and other cereals** |
| Other breakfast cereals | **Salad and other raw vegetables** |
| Other potatoes, potato salads and dishes | White bread |
| **Vegetables not raw** | Other potatoes, potato salads and dishes |
| **Brown, granary and wheat germ bread** | **Vegetables not raw** |
| **Sugar, preserves and sweet spreads** | **Brown, granary and wheat germ bread** |
|  | Skimmed milk |
|  | **Sugar, preserves and sweet spreads** |

\*Foods highlighted in bold were in the final model for both years of data

*Food variables plus sex, age, BMI and smoking status*

Table 4.7 shows the results of the final models in the backwards elimination regression analyses including the food variables, demographic and lifestyle variables in the model in years 2 and 4 data separately. There were 16 variables in the final model for year 2 and 20 for year 4. When the additional variables sex, age, BMI and smoking status were included in the model, ‘salad and other raw vegetables’, ‘wholemeal bread’, ‘wine’ and ‘sugar preserves and sweet spreads’ were no longer independent predictors in the final model for both years of data. ‘Age’ and ‘smoking status’ were in the final model for both years of data. The final list of 11 common variables was: age; smoking status; beef, veal and dishes; beer, lager, cider and perry; spirits and liqueurs; chips, fried and roast potatoes and potato products; fruit; pasta, rice and other cereals; vegetables not raw; chicken and turkey dishes and brown, granary and wheat germ bread.

**Table 4.7. Food and wider characteristics variables models that were the most predictive of diet quality (Years 2 and 4 NDNS).**

|  |  |
| --- | --- |
| Year 2 | Year 4 |
| **Age** | **Age** |
|  | Sex |
| **Smoking status** | **Smoking status** |
| **Beef, veal and dishes** | **Beef, veal and dishes** |
|  | Butter |
|  | Semi skimmed milk |
|  | Wholemeal bread |
| **Beer, lager, cider and perry** | **Beer, lager, cider and perry** |
| **Spirits and liqueurs** | **Spirits and liqueurs** |
| Wine |  |
| **Chips, fried and roast potatoes and potato products** | **Chips, fried and roast potatoes and potato products** |
|  | Chocolate confectionery |
| **Fruit** | **Fruit** |
| High fibre breakfast cereals | Meat pies and pastries |
|  | Other bread |
| **Pasta, rice and other cereals** | **Pasta, rice and other cereals** |
| Salad and other raw vegetables | Other potatoes, potato salads and dishes |
| Other breakfast cereals |  |
| **Vegetables not raw** | **Vegetables not raw** |
| **Chicken and turkey dishes** | **Chicken and turkey dishes** |
| **Brown, granary and wheat germ bread** | **Brown, granary and wheat germ bread** |
| Sugar, preserves and sweet spreads | Skimmed milk |

\*Foods highlighted in bold were in the final model for both years of data

*4.3.2.3. Confirmatory analysis (empirically-derived indicator variables of diet quality as included in the YHS-DQ).*

Table 4.8 shows the findings of the analyses comparing the model fit for the NDNS for the 4 models. The adjusted R2 for all four models demonstrates a moderate model fit for diet quality varying from 0.29-0.33. The model fit with the NDQS for the theory-driven method models is higher than for the data-driven method models with an R2 of 0.33 for both models compared with 0.29 and 0.31 for the food only model and food plus wider characteristics model respectively. This means that the theory driven method models are stronger predictors of the NDQS than data-driven method models.

**Table 4.8. Comparison of model fit of empirically-driven method model and theory-driven method model with NDQS (food variables only and food variables plus age and smoking status)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | R | R squared | Adjusted R squared |
| 1. Theory driven predictors of NDQS(12 food variables) | 0.58 | 0.34 | 0.33 |
| 2. Theory driven predictors of NDQS (9 food variables + age and smoking) | 0.58 | 0.34 | 0.33 |
| 3. Empirically-derived predictors of NDQS as included in YHS-DQ (14 food variables) | 0.55 | 0.30 | 0.29 |
| 4. Empirically-derived predictors of NDQS plus theory driven wider determinant variables YHS-DQ (14 food variables + age and smoking) | 0.57 | 0.32 | 0.31 |

Tables 4.9 to 4.12 show the main effects for the indicator variables in each model for diet quality (NDQS). ‘Beer, lager, cider and perry’ was significantly negatively associated with the NDQS in all 4 models with a main effect ranging from -0.15- -0.17. This means that, with all other things equal, with each incremental increase of ‘beer, lager, cider and perry’ intake we would expect a decrease in NDQS score of between -0.15 and -0.17. ‘Fruit’ and ‘vegetables not raw’ were significantly positively associated with the NDQS across all 4 models. For every incremental increase in ‘fruit’ intake, we would expect to see an increase in NDQS score of 0.33-0.35 and for ‘vegetables not raw’ 0.20-0.23. ‘Wholemeal bread’ was significantly positively associated with the NDQS in 3 models and was associated with an increase in NDQS score of 0.10-0.12.

**Table 4.9. Main effects of theory driven method indicator variables (food only) for diet quality (NDNS)**

|  |  |  |
| --- | --- | --- |
| Model | Main effect (Beta) | P value |
| Beer, lager, cider and perry | -0.17 | <0.001 |
| Chips, fried and roast potatoes and potato products | 0.10 | <0.001 |
| Fruit | 0.35 | <0.001 |
| Vegetables not raw | 0.23 | <0.001 |
| Wholemeal bread | 0.12 | <0.001 |
| Salad and other raw vegetables | 0.08 | 0.01 |
| Spirits and liqueurs | -0.04 | 0.10 |
| Wine | -.10 | <0.001 |
| Beef, veal and dishes | 0.10 | 0.003 |
| Pasta, rice and other cereals | 0.09 | <0.001 |
| Brown, granary and wheat germ bread | 0.14 | <0.001 |
| Sugar, preserves and sweet spreads | -0.11 | <0.001 |

**Table 4.10. Main effects of theory driven method indicator variables (food and wider characteristics) for diet quality (NDQS)**

|  |  |  |
| --- | --- | --- |
| Model | Main effect (Beta) | P value |
| Age | 0.17 | <0.001 |
| Smoking | 0.10 | <0.001 |
| Beef, veal and dishes | 0.08 | 0.002 |
| Beer, lager, cider and perry | -0.16 | <0.001 |
| Spirits and liqueurs | -0.03 | 0.22 |
| Chips, fried and roast potatoes and potato products | 0.10 | <0.001 |
| Fruit | 0.34 | <0.001 |
| Pasta, rice and other cereals | 0.14 | <0.001 |
| Vegetables not raw | 0.22 | <0.001 |
| Chicken and turkey dishes | 0.06 | 0.02 |
| Brown, granary and wheat germ bread | 0.10 | <0.001 |

**Table 4.11. Main effects of empirical (data-driven) method indicator variables (food only) for diet quality (NDQS)**

|  |  |  |
| --- | --- | --- |
| Model | Main effect (Beta) | P value |
| Beer, lager, cider and perry | -0.16 | <0.001 |
| Chips, fried and roast potatoes and potato products | 0.09 | 0.002 |
| Wholemeal bread | 0.10 | <0.001 |
| Salad and other raw vegetables | 0.08 | 0.01 |
| Vegetables not raw | 0.21 | <0.001 |
| Fruit | 0.35 | <0.001 |
| Soft drinks not low calorie | -0.08 | 0.001 |
| Crisps and savoury snacks | 0.02 | 0.45 |
| Bacon and ham | -0.02 | 0.40 |
| Biscuits | 0.04 | 0.13 |
| Buns, cakes, pastries and fruit pies | 0.03 | 0.21 |
| Coated chicken and turkey | -0.07 | 0.01 |
| Oily fish | -0.01 | 0.78 |
| White bread | 0.04 | 0.19 |

**Table 4.12. Main effects of empirical (data-driven) method indicator variables (food and wider determinants) for diet quality (NDQS)**

|  |  |  |
| --- | --- | --- |
| Model | Main effect (Beta) | P value |
| Age | 0.11 | <0.001 |
| Smoking | 0.11 | <0.001 |
| Beer, lager, cider and perry | -0.15 | <0.001 |
| Chips, fried and roast potatoes and potato products | 0.10 | <0.001 |
| Fruit | 0.33 | <0.001 |
| Vegetables not raw | 0.20 | <0.001 |
| Coated chicken and turkey | -0.06 | 0.03 |
| Biscuits | 0.03 | 0.19 |
| Buns, cakes, pastries and fruit pies | 0.01 | 0.75 |
| Oily fish | -0.02 | 0.53 |
| White bread | 0.05 | 0.08 |
| Wholemeal bread | 0.10 | <0.001 |
| Crisps and savoury snacks | 0.05 | 0.06 |
| Soft drinks not low calorie | -0.05 | 0.05 |

**4.3.3. Discussion (Objective 2)**

The aim of the analyses in objective 2 was to identify the most predictive indicator variables for diet quality (as defined by the NDQS) and to compare the findings of an empirically driven and theory driven method in independent datasets to demonstrate repeatability and validity. The findings showed that both methods are useful in identifying indicator variables that are predictive of diet quality.

The correlation analyses confirmed, to a large extent, that the individual food categories were correlated with the NDQS in the direction that would be expected. The exception to this was the ‘buns, cakes, pastries and fruit pies’ and ‘puddings’ categories which were positively correlated with the NDQS. It is possible that this was due to the breadth of these categories that include a wide range of different types of food and product that are consumed by individuals with a wide range of different types of diet. Disaggregation of some of these broader food groups before including them in the initial analyses may have led to less ambiguity.

Many of the food categories with statistically significant correlations with the NDQS were foods that had high factor loadings in the dietary patterns derived from the PCA in Study 1. This too is a positive result and reiterates that the foods that were interpreted as ‘characterising’ a dietary pattern were also associated with an *a priori* defined measure of diet quality, as was also indicated by the analysis in objective 1 of this study showing the associations between the data-driven dietary patterns and the NDSQ (Table 4.2).

The backward elimination analyses, largely confirmed previous analyses in identifying key independent predictors of diet quality. There were a number of indicator variables that were in the final backwards elimination models for NDQS in both years of the backwards elimination regression, that also featured in the list of indicator variables using the data-driven method such as ‘fruit’, ‘vegetables not raw’, beer, lager, cider and perry’ and ‘chips, fried and roast potatoes and potato products’. Only 3 of these indicator variables 3 of these indicators also had main effects of >0.10 across all 4 models: ‘beer, lager, cider and perry’ was significantly negatively associated with the NDQS (-0.15 - -0.17) and ‘fruit’ and ‘vegetables not raw’ were positively with the NDQS (0.33-0.35 and 0.20-0.22 respectively).

There were some food categories that were highlighted as being significantly predictive of the NDQS that had not been highlighted using the empirically driven methods such as ‘beef, veal and dishes’ and ‘pasta, rice and other cereals’. As a result there were some differences in the food categories and number of variables included in the final diet quality predictor models, even though there was only a small amount of difference in their overall predictive capability (adjusted R2 = 0.33 theory driven method, 0.29-0.31 data-driven method).

As wider variables such as smoking status had not been included in the PCA for methodological reasons, they could only be included in the backwards elimination analyses from which the theory driven models were derived. ‘Age’ and ‘sex’ were shown to be significant predictors. When ‘age’, ‘sex’, ‘BMI’ and ‘smoking status’ were included in the backwards elimination regression, ‘age’ and ‘sex’ were included in the final model for both years of data and ‘salad and other raw vegetables’, wholemeal bread’, ‘wine’ and ‘sugar preserves and sweet spreads’ were not included.

For almost all the food categories, the direction of the coefficient was as expected. However, the category ‘chips, fried and roast potatoes and potato products’ was positively associated with the NDQS in all three of the models that it occurred in. The correlation analyses confirmed that this variable was significantly negatively correlated with the NDQS. Therefore this result in the coefficient may be due to a ‘negative suppression effect’ which occurs when ≥2 independent variables are correlated with each other therefore confounding the relationship in a regression ([298](#_ENREF_298)).

* 1. **Objective 3. Identification of the most parsimonious model predictive of diet quality to inform the development of a Brief Diet Quality Assessment Tool.**

The aim of this part of the study is to explore the ‘trade-offs’ between predictive capability and reduction of number of variables in the models to identify the most parsimonious model for inclusion in a brief, diet quality assessment tool.

**4.4.1. Methods**

*4.4.1.1. Dataset*

The analyses were conducted in years 1 and 3 of the NDNS dataset as per the backwards elimination analysis above so that the fit for each of the models for the NDQS were confirmed in a different dataset from which the indicator variables were generated.

*4.4.1.2. Statistical analysis*

In order to identify the most parsimonious model predictive of diet quality from the findings described in Section 4.3.2, the following exploratory regression analyses with the NDQS as the dependent variable were undertaken and the model fits were compared:

1. The variables that had a statistically significant (p<0.001) main effect of ≥0.10 or ≤-0.10 and occurred in all four final models in Table 4.8 were included in regression analyses with the NDQS. Regressions were conducted with the food variables only. Models were explored with different combinations of variables. The independent variables included in the food-only regressions were: ‘fruit’, ‘vegetables (not raw)’ and ‘beer, lager, cider and perry’.
2. Food variables that had a statistically significant (p<0.001) coefficient of ≥0.10 or ≤-0.10 and occurred in three of the four models in Table 4.8 were then included in exploratory regression models. The variables included were ‘wholemeal bread’ and ‘chips, fried and roast potatoes and potato products.’ Models were explored with different combinations of variables. ‘Fruit’ and ‘vegetable (not raw)’ were included as items in all iterations due to the strength of their association with the NDQS and the data-driven dietary patterns and their key contribution to dietary recommendations and diet quality. ‘Beer, lager, cider and perry’ was excluded from some of the iterations in order to attempt to identify a reduced model that did not require measurement of alcohol consumption that would exclude teetotallers.
3. The variables ‘soft drinks not low calorie’, ‘salad and other raw vegetables’ and ‘coated chicken and turkey’ were included in the exploratory analyses. These variables had a statistically significant coefficient (p≤0.05) in two of the four models in Table 4.8 and featured in at least one of the data-driven models so as to reflect actual consumption (for example, ‘beef, veal and dishes’ was not included as it featured in the theory-driven method model but not the data-driven method model and therefore was not considered to reflect empirical patterns of consumption in the population). ‘Soft drinks not low calorie’ and ‘salad and other raw vegetables’ were considered particularly relevant in terms of their contribution to population level adherence to a dietary recommendations pertaining to consumption of fruit and vegetables and consumption of free sugars ([143](#_ENREF_143), [147](#_ENREF_147)).
4. The initial regressions from a) were repeated including the ‘smoking status’ variable. This was shown to be a significant predictor of the NDQS and was also practical to collect in a brief self-report tool.

**4.4.2. Results**

Table 4.13 summarises the results of the exploratory analyses using the criteria defined above to identify the most parsimonious model(s) predictive of diet quality (NDQS) to inform the Brief Diet Quality Assessment Tool. As would be expected, the models with the higher number of variables tended to have the greater model fit with the NDQS. There were particular variables, however, such as ‘beer, lager, cider and perry’ and ‘smoking status’ that appeared to be particularly key to improving the fit of the model with the NDQS suggesting that other lifestyle variables such as alcohol consumption and smoking are as important in predicting diet quality as individual food variables.

**Table 4.13. Exploratory models and the overall effect sizes with the NDQS**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Exploratory models of indicator variables | | Number of Items | | R | | R2 | | Adjusted R2 | |
| a) | |  | |  | |  | |  | |
| 1. ‘Fruit’, vegetables not raw’ | | 2 | | 0.49 | | 0.24 | | 0.24 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’ | | 3 | | 0.51 | | 0.26 | | 0.26 | |
| b) | |  | |  | |  | |  | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’, wholemeal bread’ | | 4 | | 0.53 | | 0.28 | | 0.27 | |
| 1. ‘Fruit, ‘vegetables not raw’, ‘beer, lager, cider and perry’, ‘chips etc.’ | | 4 | | 0.52 | | 0.27 | | 0.26 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’ | | 3 | | 0.50 | | 0.25 | | 0.25 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘chips etc.’ | | 3 | | 0.49 | | 0.24 | | 0.24 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘chips etc.’, ‘wholemeal bread’ | | 4 | | 0.50 | | 0.25 | | 0.25 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’, ‘wholemeal bread’, ‘chips etc.’   c) | | 5 | | 0.53 | | 0.28 | | 0.28 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, ‘soft drinks not low calorie’ | | 4 | | 0.51 | | 0.26 | | 0.25 | |
| 1. ‘Fruit’, ‘vegetables not raw’, wholemeal bread’, ‘soft drinks not low calories’, ‘chips etc.’ | | 5 | | 0.51 | | 0.26 | | 0.25 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, ‘soft drinks not low calorie’, ‘chips etc.’, ‘coated chicken and turkey’ | | 6 | | 0.51 | | 0.26 | | 0.26 | |
| 1. **‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, ‘soft drinks not low calorie’, ‘coated chicken and turkey’** | | **5** | | **0.51** | | **0.26** | | **0.26** | |
| 1. Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, soft drinks not low cal’, ‘coated chicken and turkey’, ‘salad and other raw vegetables’ | | 6 | | 0.51 | | 0.26 | | 0.26 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’, ‘wholemeal bread’, ‘chips, fried and roast potatoes and potato products’, ‘salad and other raw vegetables’. | | 6 | | 0.53 | | 0.28 | | 0.28 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’, ‘wholemeal bread’, ‘chips, fried and roast potatoes and potato products’, ‘soft drinks not low calorie’ | | 6 | | 0.54 | | 0.29 | | 0.28 | |
| 1. **‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’, ‘wholemeal bread’, ‘soft drinks not low calorie’** | | **5** | | **0.53** | | **0.28** | | **0.28** | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, soft drinks not low calorie’, ‘chips, fried and roast potatoes and potato products’, ‘coated chicken and turkey’, salad and other raw vegetables’ | | 7 | | 0.52 | | 0.27 | | 0.26 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, soft drinks not low calorie’, ‘chips, fried and roast potatoes and potato products’, ‘coated chicken and turkey’, salad and other raw vegetables’, ‘beer, lager, cider and perry’ | | 8 | | 0.54 | | 0.30 | | 0.29 | |
| d) | |  | |  | |  | |  | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘beer, lager, cider and perry’ + ‘smoking’ | | 4 | | 0.53 | | 0.28 | | 0.28 | |
| 1. ‘Fruit’, ‘vegetables not raw’ + ‘smoking’ | | 3 | | 0.51 | | 0.26 | | 0.26 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’ + ‘smoking’ | | 4 | | 0.52 | | 0.27 | | 0.27 | |
| 1. ‘Fruit’, ‘vegetables not raw’, ‘wholemeal bread’, ‘beer, lager, cider and perry’ + ‘smoking’ | | 5 | | 0.54 | | 0.29 | | 0.29 | |

Tables 4.14 – 4.16 show the coefficients for the 3 models identified as being the most suitable to fulfil the over-arching objective of this thesis, in terms of predicting the greatest amount of diet quality related variance with the least number of items to develop a Brief Diet Quality Assessment Tool (BDQAT). As the inclusion of the ‘beer, lager, cider and perry’ and ‘smoking status’ variables were key predictors but were also non-food variables, brief tools have been described below that include ‘food and non-alcoholic drinks variables only’, ‘food and alcohol variables’, ‘food and alcohol variables and smoking’. The variable names have been simplified for the purposes of use in a data collection tool. In the final tools, ‘fruit’, ‘vegetables’ and ‘wholemeal bread’ were positively associated with diet quality, ‘coated chicken and turkey’, ‘soft drinks (not low cal), ‘beer, lager, cider and perry’ and ‘smoking’ were negatively associated with diet quality.

**Table 4.14. Main effects regression (95% confidence intervals) of Brief Diet Quality Assessment Tool (food only) for diet quality (NDQS)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Main effect (B) | Upper CI | Lower CI | P value |
| Fruit | 0.043 | 0.036 | 0.049 | <0.001 |
| Vegetables | 0.033 | 0.024 | 0.041 | <0.001 |
| Wholemeal bread | 0.039 | 0.018 | 0.060 | <0.001 |
| Coated chicken and turkey | -0.049 | -0.094 | -0.005 | 0.030 |
| Soft drinks (not low cal, diet, sugar free) | -0.004 | -0.007 | -0.001 | 0.012 |

**Table 4.15. Main effects regression (95% confidence intervals) of Brief Diet Quality Assessment Tool (food and alcohol variables) for diet quality (NDQS)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Coefficient (B) | Upper CI | Lower CI | P value |
| Fruit | 0.042 | 0.035 | 0.048 | <0.001 |
| Vegetables | 0.033 | 0.025 | 0.041 | <0.001 |
| Beer, lager and cider | -0.004 | -0.005 | -0.003 | <0.001 |
| Wholemeal bread | 0.041 | 0.020 | 0.061 | <0.001 |
| Soft drinks (not low cal, diet, sugar free) | -0.004 | -0.008 | -0.001 | 0.006 |

**Table 4.16. Main effects regression (95% confidence intervals) of Brief Diet Quality Assessment Tool (food, alcohol and smoking) for diet quality (NDQS)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Coefficient (B) | Upper CI | Lower CI | P value |
| Fruit | 0.041 | 0.048 | 0.035 | <0.001 |
| Vegetables | 0.033 | 0.041 | 0.025 | <0.001 |
| Beer, lager and cider | -0.004 | -0.002 | -0.005 | <0.001 |
| Smoking | -3.893 | -2.271 | -5.516 | <0.001 |

**4.4.3. Discussion (Objective 3)**

The aim of this objective was to use the indicator variables most predictive of diet quality identified through the confirmatory analyses in objective 2 and explore which combinations of indicators were most useful for informing a brief population level dietary quality assessment tool. The findings show that:

* By combining the most predictive indicator variables from both methods to reflect actual intake as well as theory defined intake, brief models with <12 items were shown to be predictive of diet quality (adjusted r2=0.25-0.29).
* Three models (food variable only, food and alcohol variables and food, alcohol and smoking variables) were identified that were ≤5 items and remained predictive of diet quality (r2=0.26-0.28).
* These models can be used as Brief Diet Quality Assessment Tools (BDQAT) in population level research.

An important strength of this study is the test for repeatability and validity shown from undertaken the analyses in different cohorts of the NDNS. The backwards elimination regression in year 2 and 4 combined of the NDNS dataset allowed for the key (food and wider determinant) variables that were independently predictive of diet quality, as measured by the NDQS, to be distilled from the list of food variables included in the original data-driven dietary patterns analysis in Study 1. These were compared with the list of variables that were generated by the empirically driven dietary patterns analyses conducted in years 1-3 and 1-4 of the NDNS and included in the Yorkshire Health Study Diet Questionnaire. This confirmatory analysis, comparing these sets of was conducted in years 1 and 3 combined of the NDNS thus ensuring that associations between the key predictor variables were robust. The nutrient-based diet quality score against which the models have been assessed was developed with a rigorous methodology and was validated both through assessing its relationship with wider determinants and through correlations with biomarkers of nutrients.

There are several limitations with this study. The model that was the most strongly associated with the NDQS only had a moderate association of 0.33. Therefore, around two-thirds of the variation in the dataset in terms of diet quality is unaccounted for even in the model with the strongest association with diet quality. This could lead to substantial misclassification or inaccuracy in the prediction. This limitation will be discussed in more detail in Chapter 5.

As the wider determinant variables could not be included in the PCA in Study 1 (as the method would not be appropriate for mixed data types), the inclusion of the ‘age’ and ‘smoking status’ variables in the fourth ‘empirically-derived predictors’ model (Table 4.8) was a pragmatic as opposed to methodologically robust next step in the analyses as ‘age’ and ‘smoking status’ had been identified as key predictor variables through the backwards elimination analyses.

**Chapter 5. Discussion**

This section will discuss the findings of this thesis in the context of its aims and objectives. The strengths and limitations of the work will be discussed in the context of the existing literature. Implications for and relevance to public health policy makers will also be discussed as well as an agenda for future work and research that is generated by this thesis.

**5.1. Summary of findings and key messages**

The over-arching aim of this thesis was to explore dietary patterns in UK adults using theoretical and empirical methods to identify a brief, valid set of indicator variables that could be used to develop a brief dietary quality assessment tool to support dietary assessment in public health research. The three studies described in the thesis contributed towards this aim through the following objectives:

* To explore and examine empirically-derived (data-driven) dietary patterns in UK adults, the foods that characterise them, their internal validity and their association with demographic characteristics, socio-economic measures, wider determinants of health and nutrient biomarkers.
* To define and quantify a nutrient-level Diet Quality Score (NDQS) based on UK dietary recommendations and to explore its association with demographic characteristics, socio-economic measures, wider determinants of health and biomarkers of nutritional status.
* To evaluate the indicator variables derived using empirical and theory-driven methods as predictors of diet quality and to identify the most parsimonious model that is predictive of diet quality (as defined by the NDQS).

This section will present a summary of the findings and key messages from this thesis. These findings will be discussed in terms of their strengths and limitations and in the context of the existing literature.

**5.1.1. Dietary patterns in the UK**

In a representative sample of UK adults, 13.6% of the variance in dietary intake can be described by and reduced to ‘dietary patterns’ from an initial 60 variables included in the analyses. These patterns of dietary intake were characterised by consumption of higher or lower quantities of particular foods and food groups. The four patterns explaining the most variance were labelled as ‘fruit, vegetables and oily fish (FVOF)’; ‘snacks, fast foods and fizzy drinks (SFFFD)’; ‘meat, potatoes and beer (MPB)’; and ‘sugar foods and dairy (SFD)’; based on the foods that featured most dominantly in each pattern. The patterns were demonstrated to be consistent in a 50% random sample of the data and in an additional year of more recent data in the sample. The patterns identified were significantly associated with population characteristics, wider determinants of health, nutrient intake, nutritional status and diet quality with the first two patterns showing most relationships that were as hypothesised.

The FVOF pattern was associated with nutrient intake and biomarkers of nutrient status that were also considered indicative of a better quality diet such as intake per 1000 calories of vitamins C, D, E, B12, B6, iron, folate and magnesium; proportion of food energy from n-3 and n-6 PUFAs and NSP and biomarkers of vitamin C, D (25-hydroxy vitamin D) and A (retinol), iron (ferritin) and total carotenoids. Those scoring higher on the FVOF pattern were more likely to be older, have a lower BMI, have a higher income and be of a higher NS-SEC. They were less likely to be male, white, smokers and be in routine occupations.

The SFFFD pattern was associated with nutrient intake and biomarkers of nutrient status that were indicative of a poorer quality diet such as being negatively associated with intake per 1000 calories of vitamins C, D, E, B6, B12, folate, magnesium, proportion of food energy from n-3, grams of NSP and biomarkers of vitamins C, D, A and total carotenoids. This pattern was positively associated with proportion of food energy from NMEs, total fat, n-6 PUFA, starch and with urinary sodium. Those scoring higher on the SFFFD pattern were more likely to be male, smokers, younger and have a higher BMI.

The next stage of the thesis was to develop a theory-driven, valid, composite measure, which could be used to evaluate the data-driven dietary patterns and other models as predictors of diet quality.

**5.1.2.** **Development of a UK nutrient-based diet quality score (NDQS)**

A 13 item composite, nutrient-based measure of diet quality based on and scored according to UK dietary recommendations was correlated with biomarkers from blood and urine samples to a large extent in the hypothesised direction. A higher diet quality score was significantly positively associated with biomarker status for most vitamins, minerals and total carotenoids and negatively with c-reative protein, homocysteine and sodium. The NDQS was inversely associated with being white and being a smoker and positively associated with being older. There was no significant difference between the mean NDQS score in males and females but those of higher SES scored more highly than those of lower SES, smokers scored lower than non-smokers and the mean score for those who were obese and very obese was lower than for those of a healthy weight. In a sample of estimated under-reporters, the mean score was not significantly different from that of the general sample. Energy intake for energy requirement was not a specific measure included within the diet quality construct. The relationship between the NDQS and BMI and total energy intake was curvilinear. The food variables that were significantly and most strongly correlated with the NDQS were fruit, vegetables (not raw), high fibre breakfast cereals and wholemeal bread (positive correlation) and white bread, soft drinks (not low calorie) and beer, lager, cider and perry (negative correlations).

**5.1.3. Identification of indicator variables for diet quality to inform the development of brief diet quality assessment tools**

Data-driven and theory-driven dietary patterns analyses methods can be utilised to generate proxy indicators that are associated with diet quality. Varying the combination of the most predictive variables and number of items in a model varies the strength of prediction of diet quality. Overall, the strongest independent food variable predictors of diet quality in both methods of exploratory analysis were fruit and vegetables. The strongest independent wider determinants of health variable predictors of diet quality in the models examined were age and smoking status. Three tools of ≤ five items were predictive of diet quality and were used to inform the development of three tools (food variables only, food and alcohol variables, food, alcohol and wider determinants variables) and can be used to develop Brief Diet Quality Assessment Tools for population level research.

**5.2. Importance of this research area in the context of the existing literature and relevance to public health research and policy**

As described in Chapter 1, there were a number of gaps in public health nutrition research related to dietary surveillance in the UK that would be guided by the findings this thesis. In summary these gaps were: nationally consistent data on diet quality robust at local and population sub group level; a UK score of index of diet quality; and brief, validated diet quality assessment tools for population level surveys and evaluation of interventions. Obviously, these areas of research all interlink, overlap and support each other to facilitate the collection and analysis of robust, population level dietary data and nutritional epidemiology research in the UK. The sections below describe how this thesis contributes to supporting these areas of research need.

**5.2.1. Theory driven definition of diet quality**

Defining, measuring and monitoring population-level diet quality in the UK is under researched ([205](#_ENREF_205), [229](#_ENREF_229)). Unlike a number of other countries such as the U.S., Australia and countries in Europe, the UK does not have a theory driven composite measure of diet quality or healthy eating specifically developed in and defined for the UK population, and based on UK dietary recommendations. Perhaps as a result of this there is a paucity of research in the UK investigating the associations between diet quality and disease or other health outcomes in nutritional epidemiology.

Where UK studies do exist, the measures and tools utilised to assess levels of diet quality or consumption of a healthy diet vary widely in their content, validity and the methods used to develop them leading to a lack of consistency, repeatability and comparability. Studies with ‘diet quality’ as an outcome in the UK (children and adults) have defined and measured it using simple food based tools measuring fruit and vegetable consumption; tools generated from dietary patterns analyses ([299](#_ENREF_299), [300](#_ENREF_300)); adapted versions of indices and scores developed for other populations ([301](#_ENREF_301)); or lists of nutrients identified as key for disease prevention in particular sub-groups of the population ([302](#_ENREF_302)).

The development of the NDQS as described in study 2, provides a standard score that can be used in nutritional epidemiological studies to measure diet quality in the UK. Particularly useful might be analyses using the NDQS in future NDNS datasets which is the key data source for the government’s national level dietary monitoring, and informs key decisions relating to changes in dietary guidance and targeting of resources. The score provided the objective measure against which the proposed indicators for diet quality were validated in this study and could be used similarly in other studies.

A limitation of the use of the NDQS is that because it is based on nutrient intake, it will only be relevant for epidemiological studies where dietary intake data can be analysed at the nutrient level. For example, studies collecting data based on short-form FFQs may not allow for nutrient level disaggregation of the data and therefore would not be able to use the score. This limitation is mitigated somewhat by the cross-validation of the score with other types of variables in the NDNS, demonstrating that other forms of data can be used as a proxy for diet quality.

**5.2.2. Empirically-derived (data-driven) dietary patterns analyses**

The empirically-derived dietary patterns analyses undertaken in the NDNS provides significant support for this method as a means of exploring dietary intake and dietary habits in the UK population. The findings supported the hypothesis that there is a proportion of the population who consume particular types of foods and food groups habitually and that these overall dietary patterns are associated not only with demographic characteristics but also with wider determinants of health such as socio-economic measures and lifestyle factors. The findings also support the use of data-driven dietary patterns as proxies for diets of higher or lower diet quality and as a method for generating lists of foods for inclusion in FFQs with such a purpose. Whilst there is some evidence of using dietary patterns analyses for these purposes in the literature ([299](#_ENREF_299), [300](#_ENREF_300)) it is not a widely used method for informing FFQ or dietary assessment tool development.

The empirically-derived dietary patterns analyses undertaken also support the use of ‘whole diet’ as opposed to single nutrient or single food approaches in interventions as well as in epidemiological research. The findings support the concept that free-living individuals tend to consume types of foods habitually and are likely to consume particular types of foods together as meals habitually also. This means that education based on healthy eating and weight loss interventions need to focus on behavioural change interventions that support individuals not just in understanding which foods are high and low quality but also how to change their habitual behaviours, potentially including wider aspects of dietary intake such as shopping and cooking. There are a number of challenges with this approach. Dietary behaviour is complex and is influenced by a range of physiological, psycho-social, economic and environmental factors, and education alone is not likely to bring about behaviour change ([23](#_ENREF_23), [120](#_ENREF_120), [280](#_ENREF_280)). A ‘whole systems’ approach is needed that intervenes at multiple levels of influence and takes account of the adaptive nature of complex systems such as diet and diet-related public health issues such as obesity and diabetes type 2 ([23](#_ENREF_23), [303](#_ENREF_303), [304](#_ENREF_304)).

**5.2.3. Identification of indicator variables to inform the development of brief diet quality assessment tools**

There is a paucity of tools for assessing diet quality that have been developed and validated in and for the UK population. It is likely that this lack of validated tools has reinforced an existing gap in nutritional epidemiological studies relating to diet quality in the UK. In addition, the lack of tools that can be used and analysed easily is a problem for many public health nutrition policy makers.

The development of the validated BDQATs (food variables only, food variables plus alcohol or food variables plus alcohol and smoking) and the testing for the diet quality predictive power of the longer tools provided in the findings, including the tool developed for inclusion in the YHS survey, provides a pragmatic solution to the above research gaps for public health researchers, nutritional epidemiologists and public health policy makers. The appropriate tool or set of variables can be selected depending on the specific purpose of the research, the target group and the requirement for brevity versus predictive power. The brevity of the final BDQATs means that they are pragmatic for inclusion in a large scale self-reported survey or for collection through face-to-face interview during a session of a community intervention.

The necessity for these tools in the public health ‘field’ is demonstrated by the fact that the food-only, 5 item, BDQAT is already being used to evaluate weight management and dietary interventions commissioned in Derbyshire and in Bedfordshire. In addition, the wider set of variables developed from the empirical dietary patterns analyses and tested in the final models was developed into a 13 item dietary quality food frequency-style questionnaire for use in the Yorkshire Health Study survey. Data are currently being collected using this tool in the YHS survey and have been used in a study to analyse the association between diet quality, obesity and proximity of fast food outlet (Hobbs et al, in submission).

A strength of this study is that the wider findings present other combinations, in addition to the BDQATs, of indicator variables that are predictive of diet quality that may be more appropriate for the nature of the particular study. For example, if it is feasible to include a longer list of variables in a questionnaire or data collection, the findings show which combinations of foods are most predictive of diet quality and which single foods and wider variables are negatively and positively associated with diet quality. These findings are also highly relevant to other researchers who may wish to investigate diet quality focusing on intake of specific foods and drinks or who wish to investigate dietary patterns further for example in relation to specific meals, habits or in particular population sub-groups.

The finding in this study also add to the literature in this area due to the novel methodological approach taken. There were no studies identified that directly compared theoretically-derived and empirically-derived dietary patterns analyses methods as a means to developing dietary assessment tools. In addition, this comparison was undertaken through confirmatory analysis in a distinct dataset (different years of NDNS data) to ensure that the findings were robust and repeatable.

**5.3. Strengths and limitations of the study**

**5.3.1. Data source**

The NDNS was used for all the data analyses in this thesis. The NDNS is a robust dataset containing detailed dietary intake data from a representative sample of the UK population (more detail in provided in Section 1.4.1). The methods for data collection, recruitment of participants, processing and analysis of the data (including sourcing and updating food composition data) have been developed with close scrutiny and oversight from the commissioning departments in government ([147](#_ENREF_147)), and some of data reported from the NDNS are National Statistics which means that legally they must be collected and reported according to a Code of Compliance for Official Statistics ([305](#_ENREF_305)). As four years of data were available and each year was independently representative of the UK population, separate datasets could be used for different parts of the analyses in the thesis in order to test internal stability and repeatability.

The NDNS dataset contains individual level food intake data collected from three or four day food diaries and in its most granular form there are over 7,000 food variables which are aggregated into 59 ‘main’ food groups and ‘subsidiary’ food groups within each of these. These pre-defined ‘main’ food categories were used to dictate the variables included in the empirically driven and the theory driven analyses methods. Using these pre-defined variables meant that all food groups were covered in the analyses and that these aggregated food group variables had been processed using stringent, quality assured methods.

A limitation of using these food group variables is that, from the perspective of dietary quality and nutrient content, some of the combinations of foods included in one single category were not helpful in terms of interpretation. For example, one of the ‘main’ food group categories including in the analyses is ‘yoghurt, fromage frais and other dairy desserts’. This category combines a wide variety of dairy products that can have significantly differing levels of key nutrients such as saturated fat, protein and NMEs. It was therefore difficult to interpret what this category represented in terms of diet quality and nutrient intake. A high consumption of foods in this category could represent a diet high in plain, low fat Greek yoghurt which may be high in calcium and protein and low in saturated fat, and NMEs or a chocolate mousse which may be the inverse. Other categories such as ‘tea, coffee and water’, ‘chips, fried and roast potatoes and potato products’ were similarly broad in terms of the nutrient content of the products and foods that they covered. It may have been useful to have disaggregated some of the ‘main’ food groups and further aggregated some other categories such as those relating to red and processed meat to ensure that there was a clearer interpretation of broad nutritional value for each variable. However, this may have proved difficult given the wide variety of nutrient content of processed products and decisions may have been made subjectively.

The use of the pre-defined categories in this way also dictated the number of items included in the PCA and therefore influenced the findings overall. The number of items included in the PCA also had an influence on the amount of variance explained in the sample. If the categories had been differently defined and the number of items including in the PCA increased or decreased, this may have led to some different results. Disaggregating the food categories and thus increasing the number of items included would have increased the sensitivity to the wide variety of dietary intake in the UK population, but is likely to have reduced the proportion of the overall variance explained by the dietary patterns identified and the reverse is also likely.

Another limitation of the NDNS dataset is that it is known to contain a significant proportion of under-reporters which could have led to the misclassification of some individuals by the NDQS, leading to spurious associations with other variables in the analyses. This under-reporting was demonstrated by a Doubly Labelled Water (DLW) sub-study in years 1 and 3 of the dataset ([147](#_ENREF_147)), and was supported by analyses estimating energy expenditure from BMRs and application of a single PAL to the sample undertaken in this thesis (Section 3.2.1.10). As the DLW sub-study was undertaken in a sub-study and the analyses undertaken as part of this thesis relied on estimates calculated using broad population level PALs, no adjustment for under-reporting was made in the analyses. Without an accurate individual level measure of physical activity, it was not possible to make an adjustment that was not potentially biased by misclassification due to the use of a single cut-off as utilised in Section 3.2.1.10. The NDQS was developed to score several macronutrient groups as proportions, as opposed to absolute values, in order to account for misreporting and reporting biases to some extent. There is some evidence that under-reporting is more significant in over-weight and obese groups ([306](#_ENREF_306)). Therefore analyses in this thesis examining relationships between dietary quality and BMI status may be biased by misreporting.

**5.3.2. Theory driven nutrient-based diet quality score (NDQS)**

Study 2 described in chapter 3 of this thesis resulted in the development of a nutrient-based composite measure of diet quality for UK adults based on the current DRVs, RNIs, EARs and taking into account broader dietary guidelines, population level deficiency and excess in particular nutrients and key areas for public health concern in the UK. The NDQS was demonstrated to be positively associated with a range of biomarkers of key vitamins and minerals such as vitamins C, D, B1, B2, B6 and total carotenoids and was negatively associated with urinary sodium. The NDQS was also shown to be sensitive to gender, age and NSSEC group.

It is a limitation of any such score or index that any rationale for decisions made relating to the inclusion and exclusion of particular items, weightings, scoring and ‘cut-offs’ will be, to an extent, subjective and may have advantages and disadvantages ([155](#_ENREF_155)). The NDQS scores individuals on their consumption of 13 items, 12 nutrients and alcohol. As described in Section 3.2.1, the selection of the specific nutrients included in the measure and their relative weighting and scoring was undertaken with careful consideration reflecting current UK dietary guidelines and DRVs, priority areas for UK public health nutrition and with reference to other similar scores developed for populations with broadly comparable diets, such as in Australia and the U.S ([175](#_ENREF_175), [226](#_ENREF_226)).

The items included in the NDQS were in line with those included in several other scores and indices developed outside the UK, although other scores were not based solely on nutrient intake as they included food items. For example, the DQI included a measure of the proportion of saturated fat and amount of protein, sodium and calcium consumed compared with the U.S. Recommended Daily Allowances (RDA) ([211](#_ENREF_211)), the A-HEI includes a measure of trans fats as a proportion of total energy intake as well as alcohol intake ([259](#_ENREF_259)), the DQI-R includes measures of saturated fat as a proportion of energy intake, calcium and iron intake and a measure of dietary moderation which includes alcohol intake ([218](#_ENREF_218)), and the Aussie-DQI includes percentage energy from saturated fat, percentage energy from sugar, alcohol and sodium ([265](#_ENREF_265))

Whilst the NDQS was developed with a systematic and justifiable rationale, arguments could be made for an alternative set of items in the construct which would potentially have resulted in different findings. The inclusion of alcohol, for example, as discussed in Section 3.2.1.4, may be considered to be peripheral to a construct measuring ‘diet quality’. For the purposes of this thesis, it was considered that it was a key aspect of dietary intake due to its excessive consumption at population level, its contribution to overall energy intake and its association with a variety of negative health outcomes. The evaluation of levels of consumption of one nutrient against another in terms of relative value of contribution to overall health is also problematic. For example, quantifying the value of under-consumption of vitamin C against the over-consumption of saturated fat is difficult, particularly at population level. Each nutrient has different qualities in relation to health and their relative impact will vary depending on the individual and their diet overall (as well as numerous other variables).

A limitation of the NDQS score is its lack of ability to include an accurate assessment for appropriate levels of energy intake. As described above, this is due to the lack of a valid measure of physical activity in the NDNS dataset, meaning that individual level energy requirements could only be estimated from BMR from sex, weight and age and a population estimate of physical activity (PAL) ([307](#_ENREF_307)). This was considered too broad an assumption to be used for calculating individual level NDQS scores but the method was used to identify estimated under-reporters of calories in the dataset for the purpose of testing the sensitivity of the NDQS (Section 3.2.1).

The inability of the NDQS to take account of calorie intake is problematic firstly as consumption of excess calories, leading to weight gain and high prevalence of obesity is a priority issue for public health in the UK. Secondly, inevitably, there is a positive association between energy intake and nutrient intake ([272](#_ENREF_272)). This means that individuals who are consuming a diet that, while high in nutrients is also high in calories (to the extent that it could lead to obesity), may still be consuming adequate absolute levels of nutrients as defined by UK DRVs ([239](#_ENREF_239)). This generates the question of whether a diet can be considered to be of good or high quality if it is one in which an individual consumes too many calories, potentially leading to weight gain and negative health outcomes, regardless of its absolute levels of nutrient intake ([228](#_ENREF_228)).

However, the NDQS score did adjust for energy intake for some of the key macronutrients in that individuals were scored based on the proportion of their average daily total energy intake for saturated fat, n-3 and n-6 PUFA, trans fats and NMEs, as opposed to an absolute value. This could not be calculated for some macronutrients, such as NSP or for micronutrients, as their respective DRVs are based on absolute levels of intake ([239](#_ENREF_239)).

**5.3.3. Dietary Patterns Analyses**

PCA has been used widely in nutritional epidemiological studies to reduce the detailed complexity of dietary intake into reduced sets of variables that represent a proportion of the overall variance in the data and ‘patterns’ of consumption that can be labelled for easy interpretation ([160](#_ENREF_160)). The method differs from cluster analysis in that it focuses on the grouping of input variables by the inter-correlations between them as opposed to grouping of individuals according to consistencies in their food intake ([160](#_ENREF_160)). The PCA method was therefore appropriate for the research questions posed at the beginning of this thesis as one of the key objectives was to identify the variables (food based or wider determinants of health) that were representative of, and therefore could potentially be used as proxy measures for, broader groups of inter-correlated variables. Methods such as PCA have been criticised for their lack of repeatability between studies ([46](#_ENREF_46), [168](#_ENREF_168)). This study has demonstrated some evidence of repeatability by conducting PCA and comparing dietary patterns derived from different years of the dataset and a random 50% sample of the data.

As discussed above, the number of variables included in the PCA was partially pre-defined through the ‘main’ and ‘subsidiary’ food group categories available in the dataset ([147](#_ENREF_147)). There were 60 variables included in the PCA in this thesis which generated four Principal Components explaining 13.6% of the variance in the data. The number of variables included was greater than other studies undertaking similar analyses in the UK ([158](#_ENREF_158), [159](#_ENREF_159), [195](#_ENREF_195), [249](#_ENREF_249)) where 31-51 variables were included in the PCA, and greater proportions of variance explained in the sample ranging from 23.5%-32.7% have been reported. The UK study with the nearest amount of variance explained, of 16.5% compared with the findings in this thesis was conducted in a sample of a comparable size (n=2931). This included 51 food variables in the PCA and used comparable food categories to the PCA in this thesis ([194](#_ENREF_194)). There was evidence from one study that the more variables that are included in the PCA, the less variance tends to be explained and that the inverse is true ([172](#_ENREF_172)). The latter was also detrimental to generating precise estimates of disease risk. This suggests that reducing the number of variables to increase the amount of variance explained is not always desirable, as it potentially attenuates sensitivity and the identification of relevant associations with disease and other variables.

There are some studies in the UK that have identified dietary patterns similar to those identified in this thesis which supports their generalisability and external validity. One study undertaking PCA on data from the Low Income Diet and Nutrition Survey (LIDNS) (n=2931, mean age=49.4 years) ([117](#_ENREF_117)) used the same data collection and processing methodology to the NDNS, including categorising food groups similarly but in a sample of low income households ([194](#_ENREF_194)). In the LINDS, 51 variables were entered into the PCA from which four dietary patterns were identified that explained 16.5% of the total variance. The dietary patterns were labelled as ‘fast food’ (characterised by pasta, rice and pizza, coated chicken and turkey, chips, crisps and confectionery), ‘health aware’ (characterised by wholemeal bread, yoghurt and dairy desserts, oily fish, vegetables and fruit), ‘traditional’ (characterised by white bread, bacon and ham, sausages and potatoes) and ‘sweet’ (characterised by biscuits, buns, cakes, pastries and puddings) ([194](#_ENREF_194)). These patterns mirrored the four patterns identified in this thesis which were labelled as ‘snacks’, fast food and fizzy drinks’ (characterised by soft drinks, crisps, coated chicken and turkey, chips and confectionery), ‘fruit, vegetables and oily fish’ (characterised by fruit, vegetables, oily fish and yoghurt), ‘meat, potatoes and beer’ (characterised by potato dishes, bacon and ham, white bread and beer) and ‘sugary foods and dairy’ (characterised by sweet spreads, buns, cakes, pastries and puddings).

A study analysing dietary patterns in pregnant women in the ALSPAC dataset identified five dietary patterns explaining 32.7% of variance which were also similar to those identified in this thesis. The patterns were described as: ‘health conscious’ characterised by fruit, pulses, oat and bran cereals and non-white bread; ‘traditional’ characterised by meat and vegetables; ‘processed’ characterised by high fat and processed foods such as pizza, crisps and chips; ‘confectionery’ characterised by chocolate and other sweet foods such as biscuits, cakes and puddings, and ‘vegetarian’ which was characterised by meat substitutes and pulses ([159](#_ENREF_159)). The PCA in men in the same dataset identified four dietary patterns which were labelled similarly as: ‘health conscious’, ‘traditional’, ‘confectionery/processed’ and ‘semi-vegetarian’ ([196](#_ENREF_196)). Both of these studies identified distinct patterns that were labelled as being ‘health conscious’ which were similar to the FVOF pattern in this thesis and ‘traditional’, which was similar to the MPB pattern. The ‘processed’ pattern was similar to the ‘SFFFD’ pattern and the ‘confectionery’ pattern was similar to the ‘SFD’ pattern. No ‘vegetarian’ pattern was identified in this thesis. However, some differences in the resulting patterns would be expected when the data collection methods, data processing and input variables were different in these studies compared with this thesis.

A limitation of the PCA in this thesis is that there was no adjustment of input variables to take account of energy intake. Studies investigating this have found that it is not necessary to adjust for energy intake before entry into the PCA ([275](#_ENREF_275)), and one study showed that there was little difference in the dietary patterns generated from PCA regardless of whether variables were included as absolute values or adjusting for weight for energy intake ([171](#_ENREF_171)).

**5.3.4. Development of Yorkshire Health Study Diet Questionnaire and the Brief Diet Quality Assessment Tools (BDQATs)**

The final sets of analyses in the thesis led to the development of three brief tools that were demonstrated to be predictive of diet quality: a food only tool, a food and alcohol tool and a food, alcohol and smoking tool. The importance of developing tools such as these and their necessity in public health research ‘in the field’ is clearly demonstrated by the fact that the food-based tool has already been adopted by two adult weight management interventions, one in Derbyshire (‘Heart of Derbyshire’ see Appendix A, Figure A.2) and one in Bedfordshire (‘Beezeebodies’). The food based tool is being used in these interventions for assessing the levels of diet quality of the participants and evaluating the effectiveness of one of the intervention objectives. Further discussions have also take place with a national level intervention regarding use of the tool.

In addition, the interim results from the empirically driven dietary patterns analyses were used to generate variables that were developed into a food frequency-style questionnaire for inclusion in the Yorkshire Health Study survey. These variables were validated against the NDQS later in the thesis and were found to be moderately associated with diet quality. This tool has been used to collect dietary data from a sample of nearly 3,000 adults from across Yorkshire along with other detailed demographic, socio-economic and health related information. Dietary data from this study has been used to analyse the association between diet quality, obesity and proximity of residence to fast food outlets (309). Further studies examining associations between diet quality, obesity, long term health conditions, use of health services, socio-economic variables and physical activity are possible from this dataset in the future.

Two studies had used brief measures of diet quality for different purposes and developed tools using different methods in the UK ([205](#_ENREF_205), [229](#_ENREF_229)). The Eating Choices Index (ECI) is a four item questionnaire intended to measure diet quality in large population level surveys and has been demonstrated to be associated with eight nutrient level indicators of diet quality ([229](#_ENREF_229)). The DQS is a five item measure, based on the list of dietary components in the WHO prevention of chronic disease report and used to validate a 20 item FFQ as a tool for measuring dietary quality ([205](#_ENREF_205)).

In this thesis, the indicator variables of diet quality that have been identified are in line with these two studies ([205](#_ENREF_205), [229](#_ENREF_229)). The four items in the ECI are: consumption of breakfast; consumption of two portions of fruit per day; type of bread consumed and type of milk consumed. The five item DQS used to validate the SFFFQ included consumption of fruit, vegetables, oily fish, fat and NMES with cut-offs based on UK recommendations ([205](#_ENREF_205)).

During the validation of the SFFFQ to assess dietary quality, it was reported that a greater number of portions of fruits and vegetables and wholemeal bread consumed was associated with a higher DQS score (OR 1.27, 1.27 and 1.25 respectively), and consumption of sugary fizzy drinks and processed chicken based foods were associated with a lower DQS score (OR 0.20 and 0.05 respectively) ([205](#_ENREF_205)). This reflects the food items included in the food-only BDQAT, which includes fruit, vegetables and wholemeal bread as positive indicators of diet quality and sugary drinks and coated chicken/turkey as negative indicators of diet quality.

Similarly, the identification in this thesis of consumption of fruit and vegetables as the foods that are the most strongly predictive of diet quality, compared with the other food variables included in the analyses, reflects the findings of the validity study of the ECI score ([229](#_ENREF_229)). The validation study of the ECI showed that more frequent consumption of at least two portions of fruit daily was positively associated with nutrient indicators of a healthy diet ([229](#_ENREF_229)). The ECI also identified that consuming wholemeal bread more regularly was associated with a healthy diet, in line with the results of this thesis.

In terms of the methods used, a strength of the BDQATs compared with these other UK tools is that they have been developed and validated using robust analytical methods that incorporate indicators derived from empirical patterns of dietary intake associated with diet quality, as well as indicators of a theory-driven, nutrient-based ‘ideal’ diet score (NDQS) based on UK DRVs and dietary recommendations. The development and validation methods used in this thesis include testing against objective measures of nutrient biomarkers from blood and urine samples and confirmatory analysis for repeatability undertaken in different cohorts of the NDNS. Whilst the ECI was validated against intake of nutrients that were indicative of healthy eating, this was derived from self-reported intake as opposed to an objective measure ([229](#_ENREF_229)). Similarly, whilst the DQS was developed from an existing and widely used tool (the HDI) it has not been validated in a UK cohort to assess its validity as a measure of diet quality against a more robust or objective measure such as nutrient biomarkers, and does not reflect the broader aspects of the UK’s dietary recommendations ([205](#_ENREF_205), [228](#_ENREF_228)).

The findings in this thesis are also in line with food-based diet quality measurement tools developed outside the UK. The Healthy Diet Indicator based on WHO guidelines, the Danish Diet Quality Score ([254](#_ENREF_254)) and the U.S. HEI-2010 both include fruit and vegetables ([212](#_ENREF_212), [226](#_ENREF_226)) and the latter also includes whole grains. Again, this reflects the findings in this thesis that fruit and vegetables are the strongest food based positive predictors of diet quality and that wholemeal bread is also indicative of a higher quality diet. The Mediterranean Diet Score and all adaptations of it also include fruit and vegetables ([174](#_ENREF_174), [213](#_ENREF_213)) with some more recent variations including wholegrains ([219](#_ENREF_219)).

A limitation of the findings and the BDQATs is that the models that were the most predictive of diet quality were only moderately predictive (maximum adjusted R2=0.33). Therefore two thirds of the variance in diet quality is unaccounted for, which could lead to misclassification and inaccuracy of prediction. This is, however, comparable with other studies where the levels of effect tend to be fair to moderate although different measures of diet quality and statistical tests have been used from this thesis ([205](#_ENREF_205), [229](#_ENREF_229), [300](#_ENREF_300)).

A further limitation of the tools is that whilst they are pragmatic for public health research and population level surveillance in the brevity of the number of items in the constructs and in their simplicity, they do not provide clear scoring and cut-offs for discriminating between different levels of dietary quality. A standardised method for collecting and interpreting data collected using these tools is a key area for further research to enable the tools to be utilised effectively, consistently and importantly to allow for robust comparison between studies where the tools have been used.

**5.4. Agenda for further research**

As described above, the findings in this thesis provide some answers to gaps in the provision of tools suitable for population level monitoring of diet quality in the UK and also the assessment of nutrient level dietary quality in nutritional epidemiological studies. The wide range of findings in the thesis are likely to be of relevance to a range of different types of research relating to the use of dietary patterns analyses in epidemiology, the development of dietary assessment methods, methods of evaluations of complex interventions and the definition and relevance of diet quality as a predictor and outcome variable in epidemiological studies.

With a more focused approach, there are opportunities to build on the methods and findings presented here to improve the validity of the BDQATs, to develop them to further enhance their fitness for purpose for use in pragmatic public health practice, and to explore methodological options for mitigating against some of the limitations described in Section 5.3. The following sections cover specific areas of further research which are proposed as being generated by the findings in this thesis.

**5.4.1. Exploration of differently defined food variables included in the models**

As described above, the variables included in both the PCA and the backwards elimination regressions to develop data-driven and theory-driven dietary patterns were pre-defined in the NDNS dataset. There were some broadly defined variables that were difficult to categorise in terms of the outcome of interest (diet quality) for example, ‘yoghurt, fromage frais and dairy desserts’. The nature and number of the variables included in the analyses significantly influences the findings in this type of research and further consideration of whether disaggregation or aggregation of some of the food variables may provide clearer patterns. This should be considered in the context of the number of variables included which, as described in Section 5.3.1 may also have an impact on the amount of variance in the data explained.

**5.4.2. Further development and validation of BDQATs to facilitate efficient and standardised use**

The final tools developed provide sets of indicators that are associated with diet quality and evidence as to the strength and direction of their association. Further work is required to develop these existing resources into tools that can be easily administered, analysed and the outcome data interpreted in a standardised and consistent way by public health practitioners and researchers. This might be facilitated through the development of a FFQ or other type of dietary questionnaire as per similar studies in the literature ([300](#_ENREF_300)).

In order for the tools to be most useful to public health practitioners and researchers, a clear system for scoring and interpretation that discriminates between individuals and groups consuming diets of different quality levels is required. This would enable the tools to be used to generate comparisons between different samples. An example would be a FFQ that scored individuals against each item and provided standard cut-offs for ‘high’, ‘medium’ and ‘low’ quality diets. Work to define what constitutes a ‘good’ or ‘poor’ quality diet in terms of the sensitivity and specificity of the NDQS score, and to develop a justifiable rationale for scoring ‘cut-offs’ is therefore needed.

In order to ensure that the resulting FFQ has been comprehensively validated, it should be tested and re-tested at an appropriate length interval against a more robust dietary assessment method such as an existing validated FFQ or 24-hour recall method ([72](#_ENREF_72)). Cognitive testing in the target population should also be considered. If the tool is intended for the purpose of evaluating interventions, testing should also be carried out to ensure that the tool is sensitive to behaviour change. The validity of the tools should ideally be tested in different groups such as different ethnicities, ages and socio-economic groups.

As the food-based tool has already been adopted by two interventions (as described in Section 5.3.4), there are potential opportunities to analyse existing data collected to examine patterns by demographic, socio-economic and other wider lifestyle variables being routinely collected. There may also be an opportunity to undertake further validity and cognitive testing of the tools in sub-samples of the participants taking part in these interventions with additional funding.

* + 1. **Further refining of the NDQS to account for energy intake**

The key limitation of the NDQS is the lack of inclusion of an item scoring individuals according to their individual energy intake for energy requirement, which is a key predictor of diet quality and health. The regression analyses in Section 3.4.2 showed that both total energy intake and BMI had a curvilinear relationship with the NDQS where they were positively associated with the score until an optimal point when the relationship became negatively associated. Also, the findings in Section 3.2.2 showed no significant difference in the mean NDQS score for healthy weight and obese individuals. This supports the need for the score to be amended to take account of energy intake, at least in epidemiological studies where an individual level measure of physical activity is included in the dataset, therefore permitting such a score to be calculated on the basis of estimated Basal Metabolic Rate and Physical Activity Level.

* + 1. **Development of tools for use in children**

A key area for further development of this work would be to carry out similar analyses to develop equivalent tools for use in children. The diet quality of children is a high priority for public health in the UK particularly due to high levels of sugar consumption compared to recommended levels in children under 11 years ([147](#_ENREF_147)) which can lead to obesity and dental caries ([252](#_ENREF_252)). While there has been some work defining diet quality at nutrient level for children’s diets ([230](#_ENREF_230), [308](#_ENREF_308)), the methods described in this thesis would add to this existing research and provide a robust approach to developing a nutrient-based score for diet quality for use in children in the UK as well as brief tools for assessing diet quality in population level surveillance.

**5.5. Conclusions**

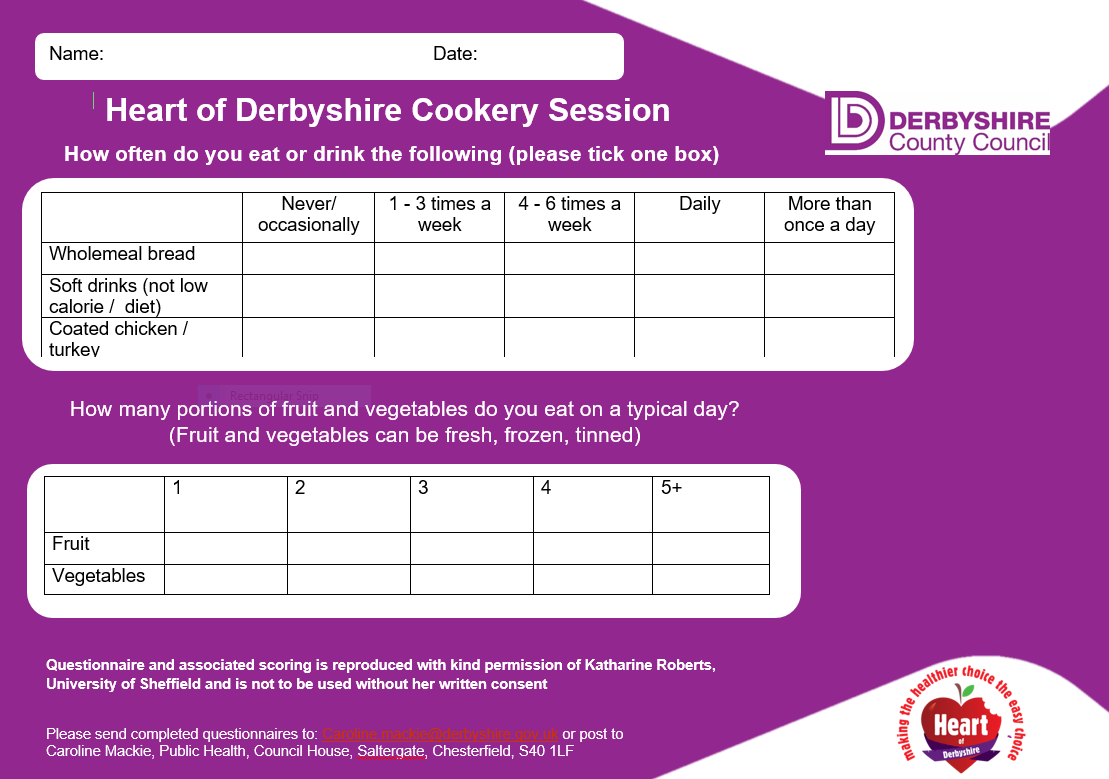
Diet quality is an important measure for nutritional epidemiological studies and to inform public health policy and decision making. In this thesis, a novel approach based on empirically and theoretically-derived dietary patterns analyses methods has been taken to develop brief, diet quality assessment tools that can be utilised to facilitate the collection of robust dietary data to support different types of public health nutrition research. Whilst the findings in this thesis have their limitations and further work is required to ensure that the brief, diet quality assessment tools can be used effectively and consistently in a variety of settings, the aims and objectives of the thesis have been successfully addressed. The tools produced in this thesis will facilitate further research into diet quality and improved dietary surveillance to inform public health policy in the UK. The tools have already been adopted for use in a variety of public health research projects, clearly demonstrating their immediate usefulness and relevance.

**Appendix A.**

**Figure A.1. Yorkshire Health Study Questionnaire – diet questions**



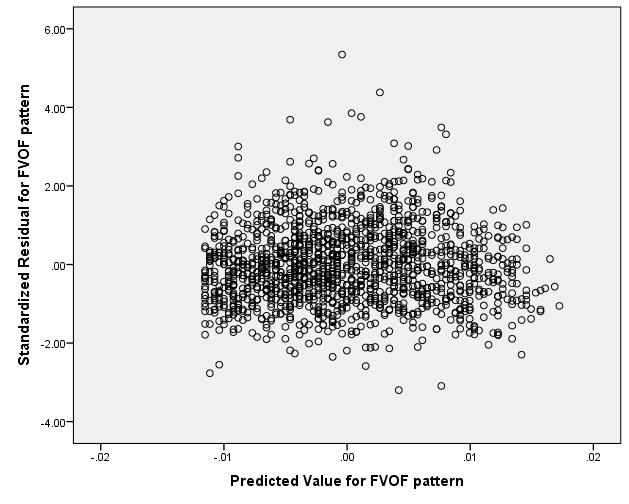
**Figure A.2. Derbyshire Heart Study Questionnaire (food based BDQAT)**

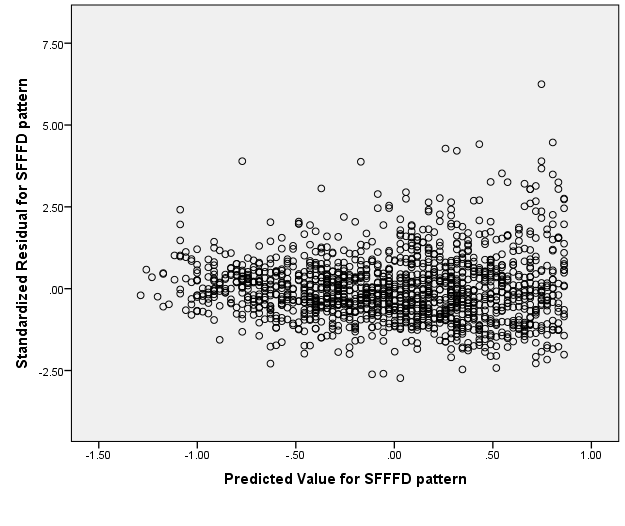


**Appendix B. Residual plots**

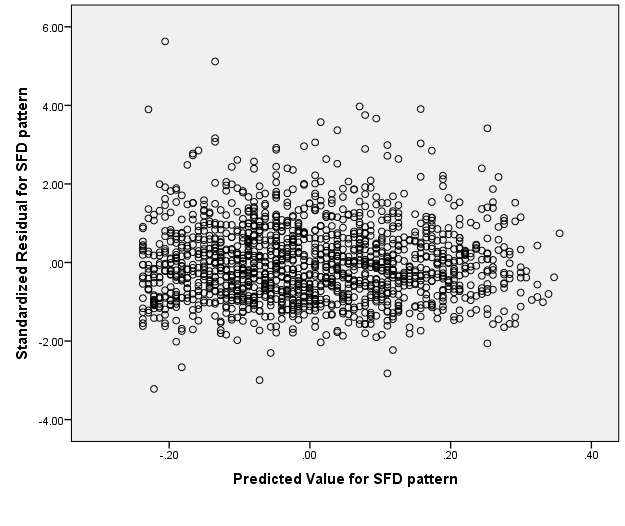
**B.1. Residual plots for dietary patterns with age, BMI and household income (NDNS 2008-2011)**

**Figure B.1.1. Residual plot FVOF dietary pattern with age**

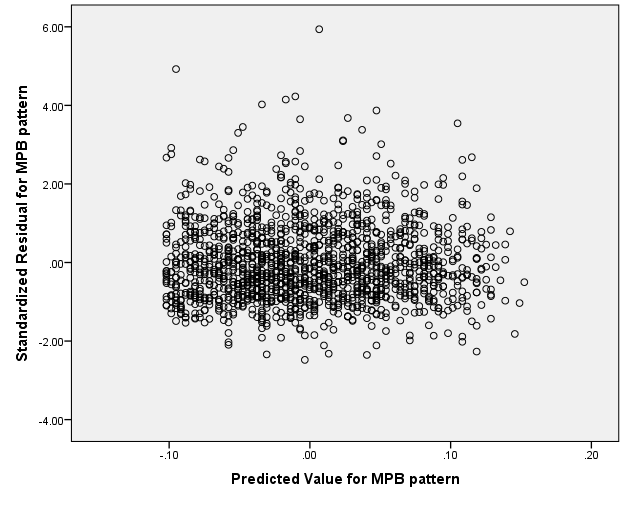
****

**Figure B.1.2 Residual plot SFFFD dietary pattern with age **

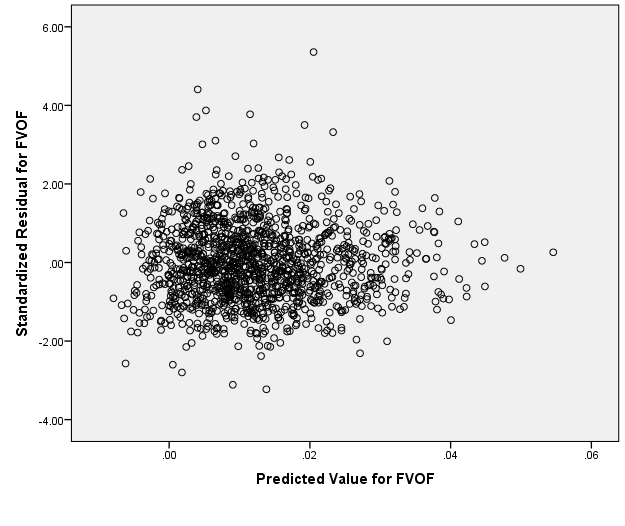
**Figure B.1.3. Residual plot SFD dietary pattern with age**



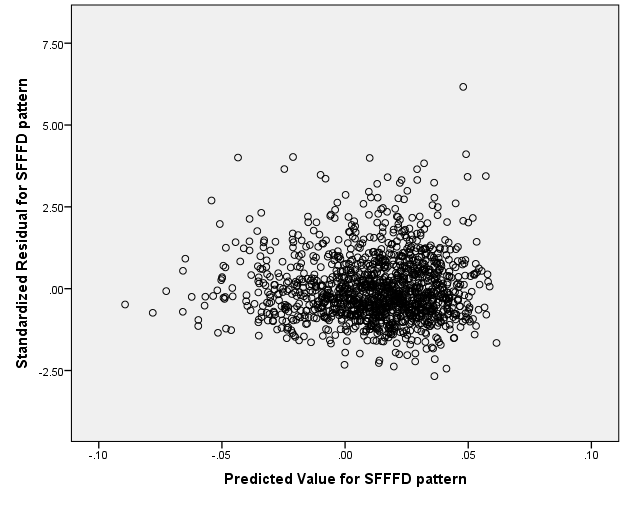
**Figure B.1.4. Residual plot for MPB dietary pattern with age**



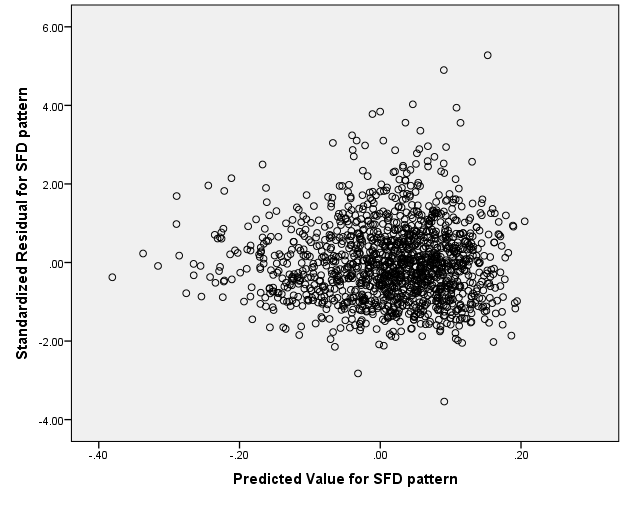
**Figure B.1.5. Residual plot for FVOF dietary pattern with BMI**



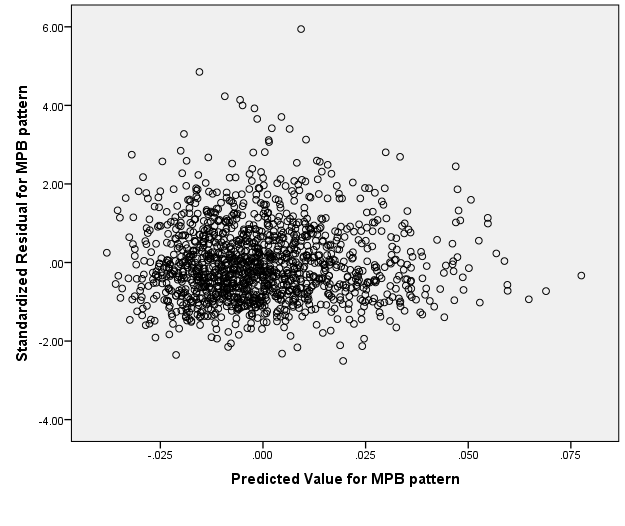
**Figure B.1.6. Residual plot for SFFFD dietary pattern with BMI**



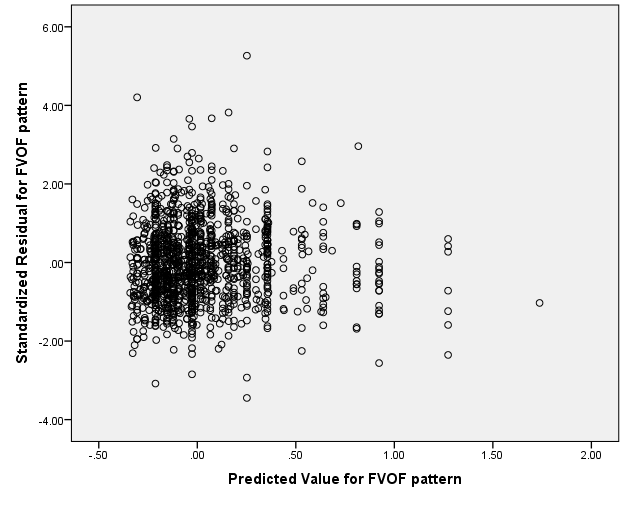
**Figure B.1.7. Residual plot for SFD dietary pattern with BMI**

****

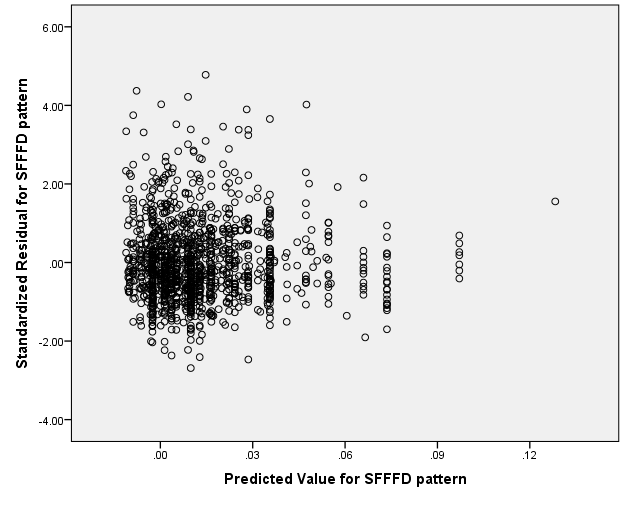
**Figure B.1.8. Residual plot for MPB dietary pattern with BMI**

****

**Figure B.1.9. Residual plot for FVOF pattern with household income**

****

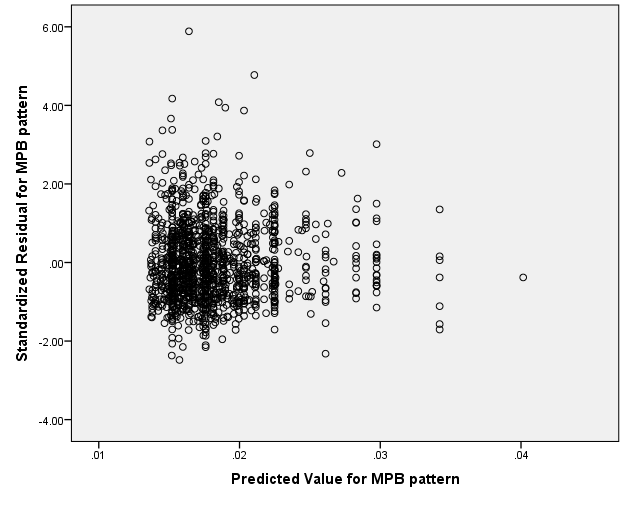
**B.1.10. Residual plot for SFFFD pattern with household income**

****

**B.1.11. Residual plot for SFD pattern with household income**

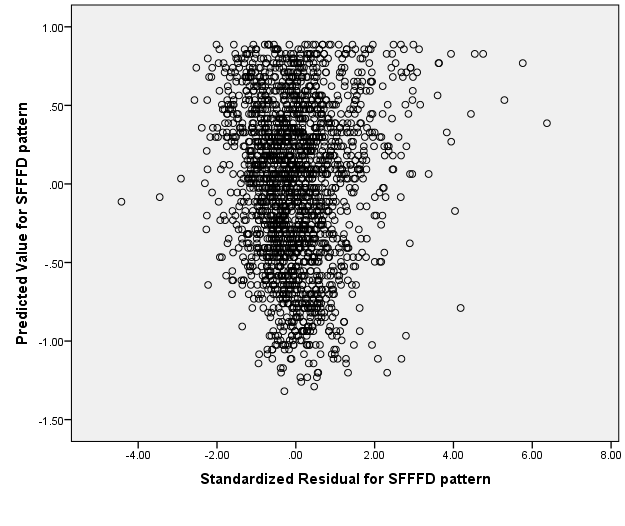
****

**B.1.12. Residual plot for MPB pattern with household income**

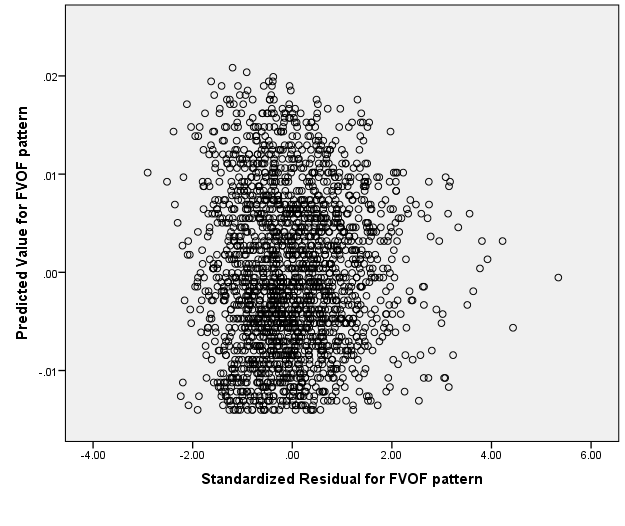
****

**B.2. Residual plots for dietary patterns and age, BMI and household income (NDNS 2008-2012)**

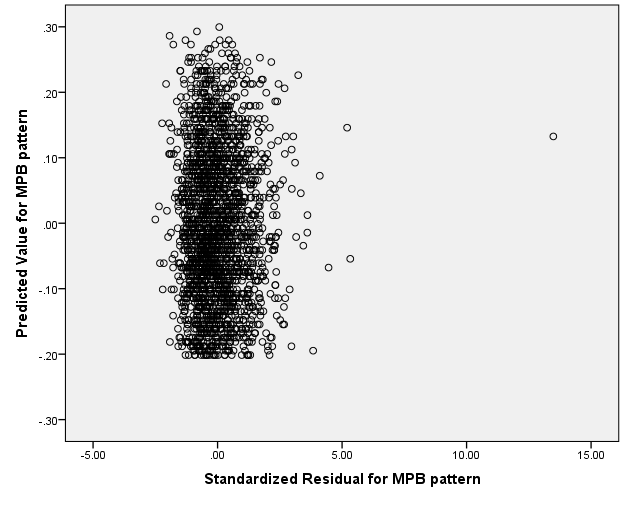
**B.2.1. Residual plot for SFFFD pattern and age**

****

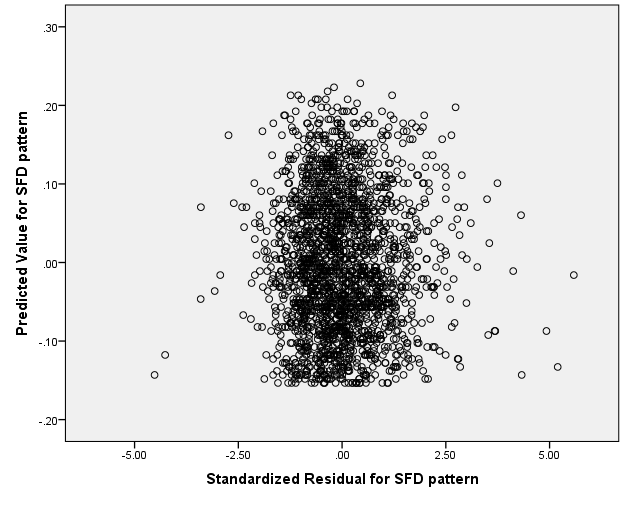
**B.2.2. Residual plot for FVOF pattern and age**

****

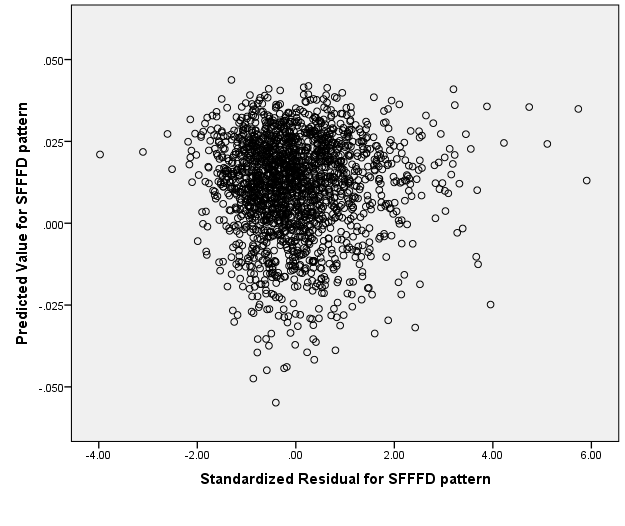
**B.2.3. Residual plot for MPB pattern and age**

****

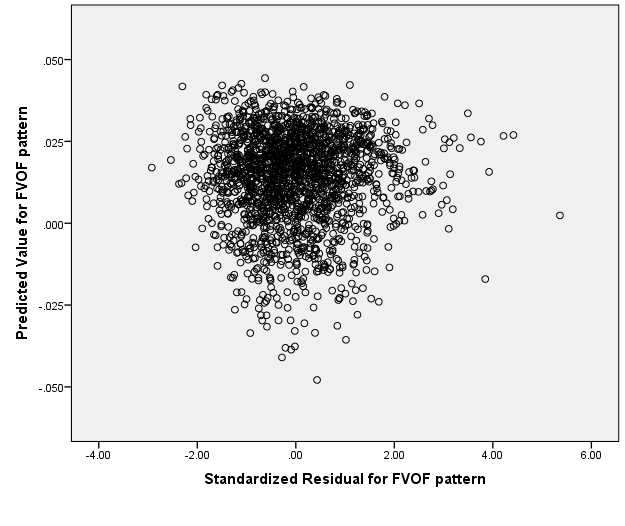
**B.2.4. Residual plot for SFD pattern and age**

****

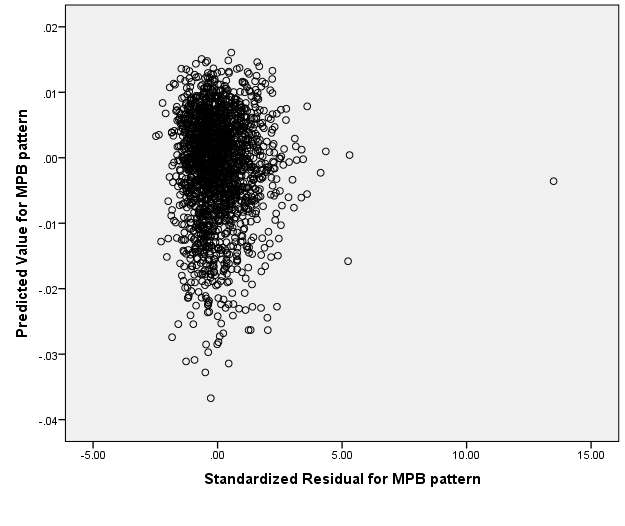
**B.2.5. Residual plot for SFFFD pattern and BMI**

****

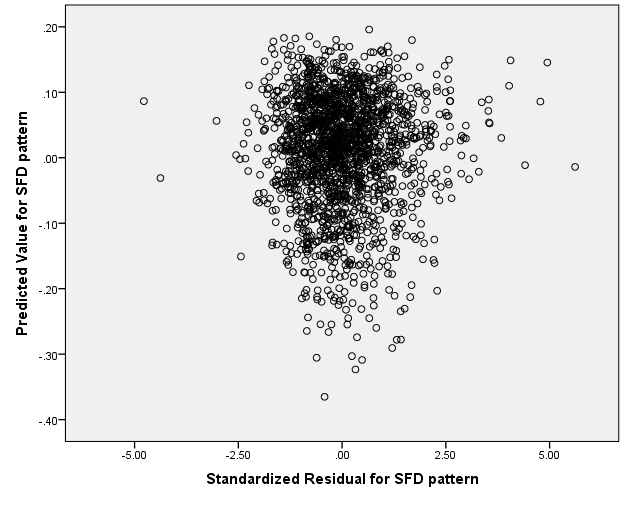
**B.2.6. Residual plot for FVOF pattern and BMI**

****

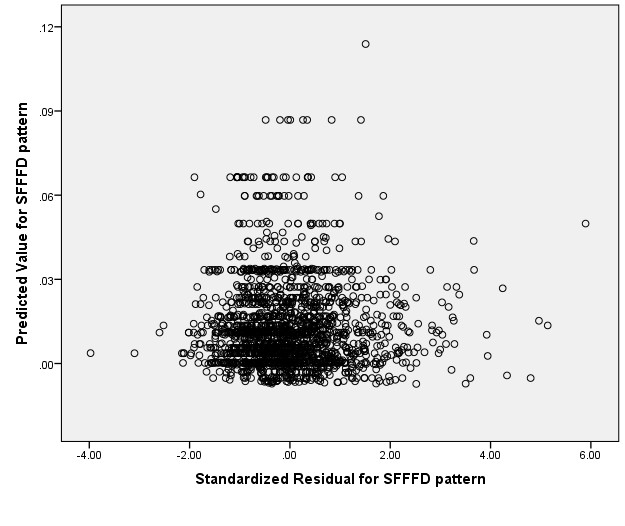
**B.2.7. Residual plot for MPB pattern and BMI**

****

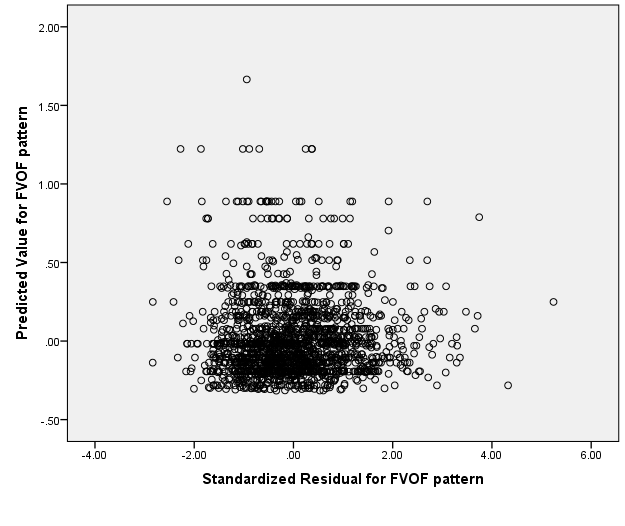
**B.2.8. Residual plot for SFD pattern and BMI**

****

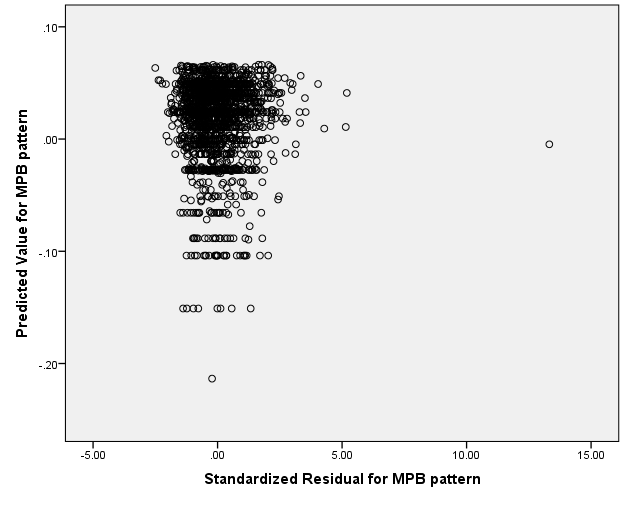
**B.2.9. Residual plot for SFFFD pattern and household income**

****

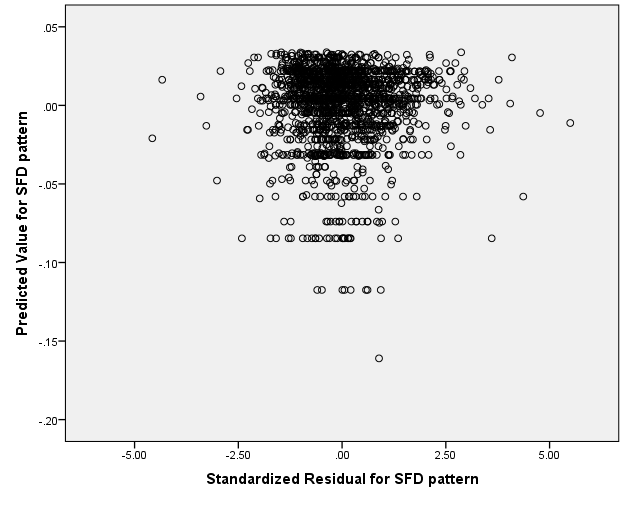
**B.2.10 Residual plot for FVOF pattern and household income**

****

**B.2.11. Residual plot for MPB pattern and household income**

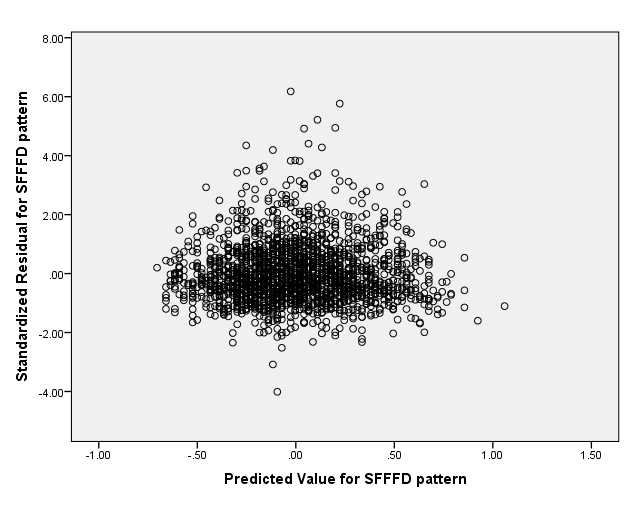
****

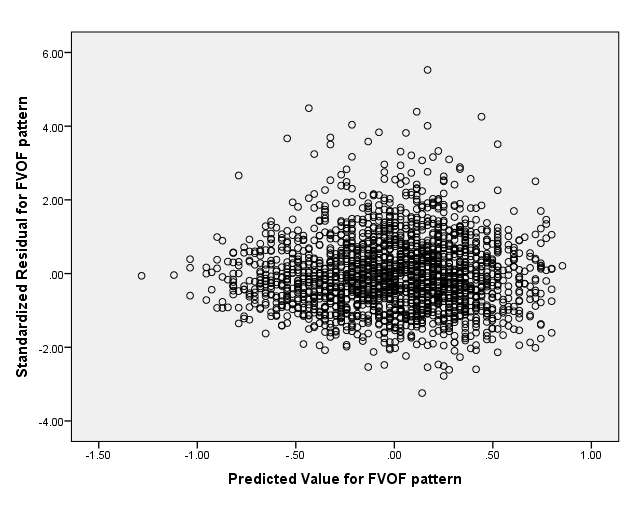
**B.2.12. Residual plot for SFD pattern and household income**

****

**B.3. Residual plots for NDQS and data-driven dietary patterns (2008-2012 NDNS)**

**Figure B.3.1. Residual plot of NDQS and ‘Snacks, fast food and fizzy drinks’ pattern**

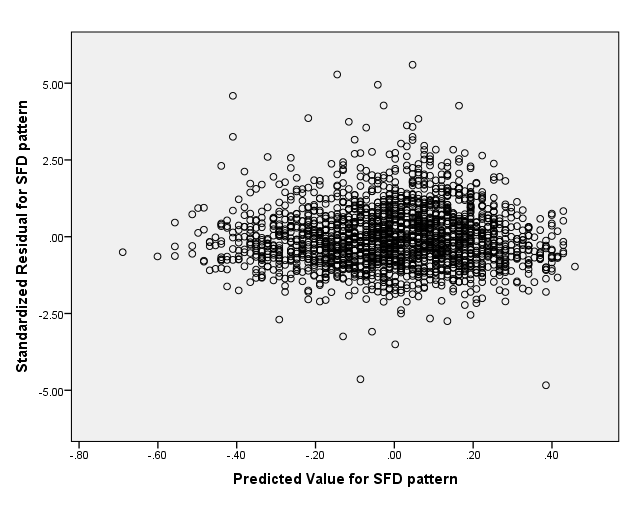


**Figure B.3.2. Residual plot of NDQS and ‘Fruit, vegetables and oily fish’ pattern**

**Figure B.3.3. Residual plot of NDQS and ‘Meat, potatoes and beer’ pattern**

****

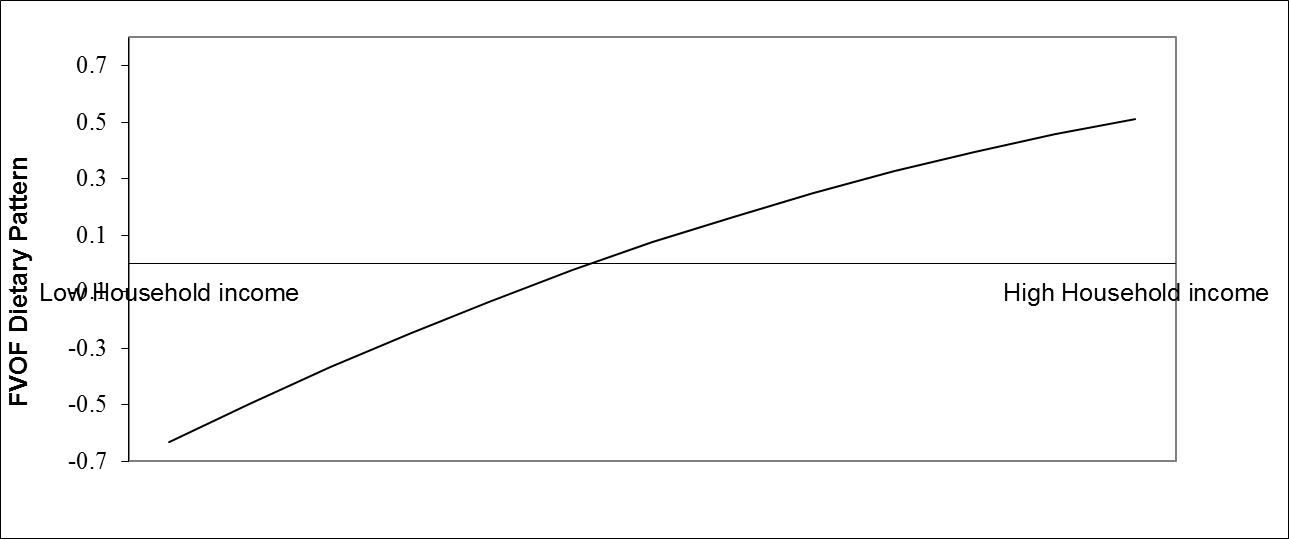
**Figure B.3.4. Residual plot of NDQS and ‘Sugary foods and dairy’ pattern**

****

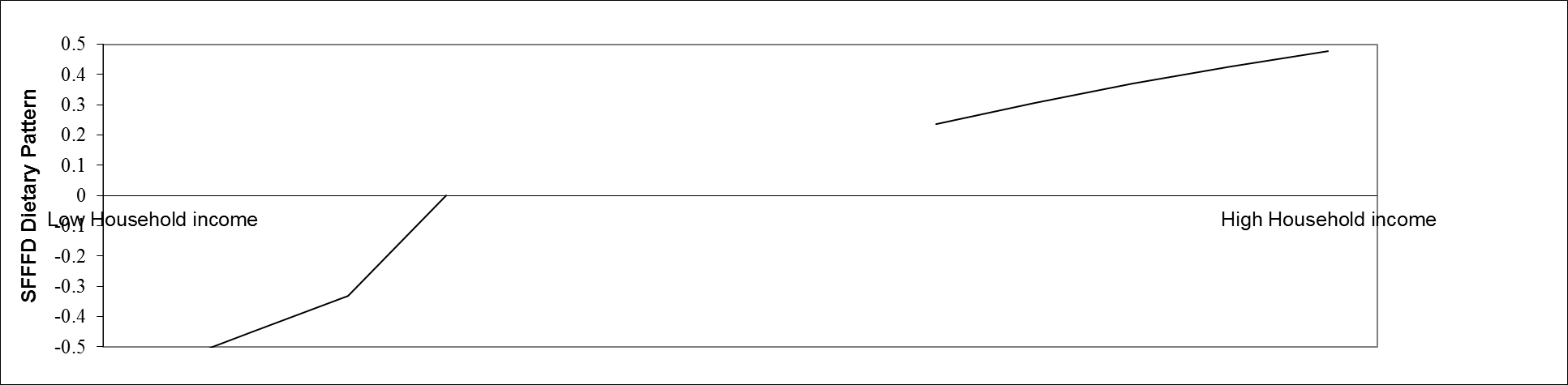
**Appendix C. Quadratic plots**

**C.1. Quadratic plots: data-driven dietary patterns and sample characteristics**

**Figure C.1.1. Quadratic plot: Household income and FVOF dietary pattern (NDNS 2008-2011)**



**Figure C.1.2. Quadratic plot: Household income and SFFFD dietary pattern (NDNS 2008-2012)**



**C.2. Quadratic plots: NDQS and sample characteristics (NDNS Years 1-4)**

**Figure C.2.1. Quadratic plot: NDQS and BMI**

**Figure C.2.2. Quadratic plot: NDQS and household income.**

**Figure C.2.3. Quadratic plot: NDQS and total energy.**

**C.3. Quadratic plots: NDQS and Food Variables (Year 2 NDNS only)**

**Figure C.3.1. Quadratic plot: NDQS and Skimmed Milk**

**Figure C.3.2. Quadratic plot: NDQS and yoghurt, fromage frais, dairy dessert**

**Figure C.3.3. Quadratic plot: NDQS and white bread**

**Figure C.3.4. Quadratic plot: NDQS and salad and other raw vegetables**

**Figure C.3.5. Quadratic plot: NDQS and meat pies and pasties**

**Figure C.3.6. Quadratic plot: NDQS and high fibre breakfast cereals**

**Figure C.3.7. Quadratic plot: NDQS and tea, coffee and water**

**Figure C.3.8. Quadratic plot: NDQS and spirits and liqueurs**

**Figure C. 3.9. Quadratic plot: NDQS and wholemeal bread**

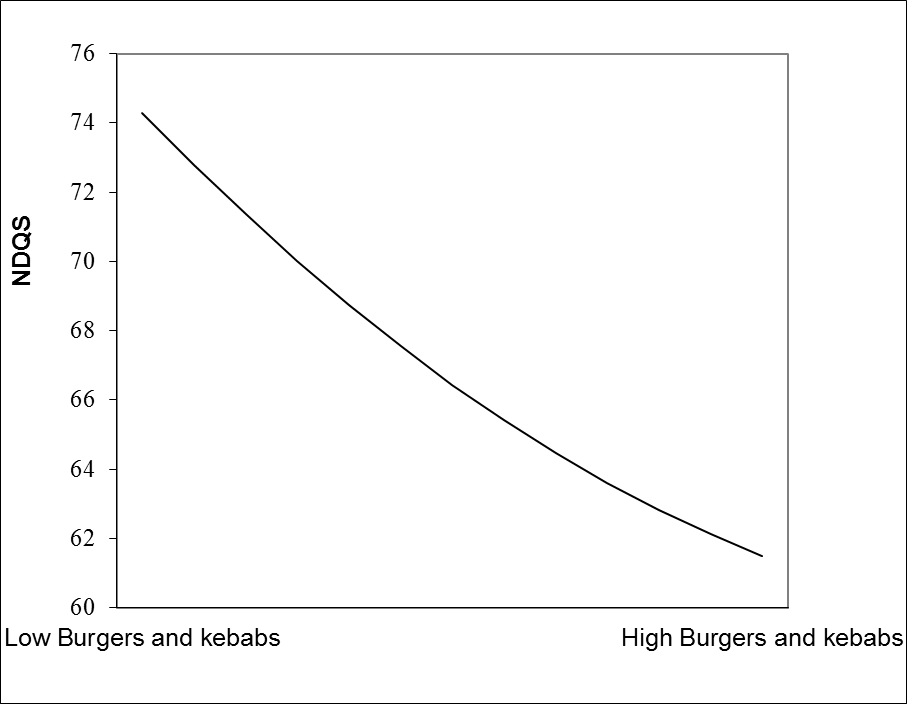
**Figure C.3.10. NDQS and fruit**

**C.4. Quadratic plots: NDQS and Food Variables (Year 4 NDNS only)**

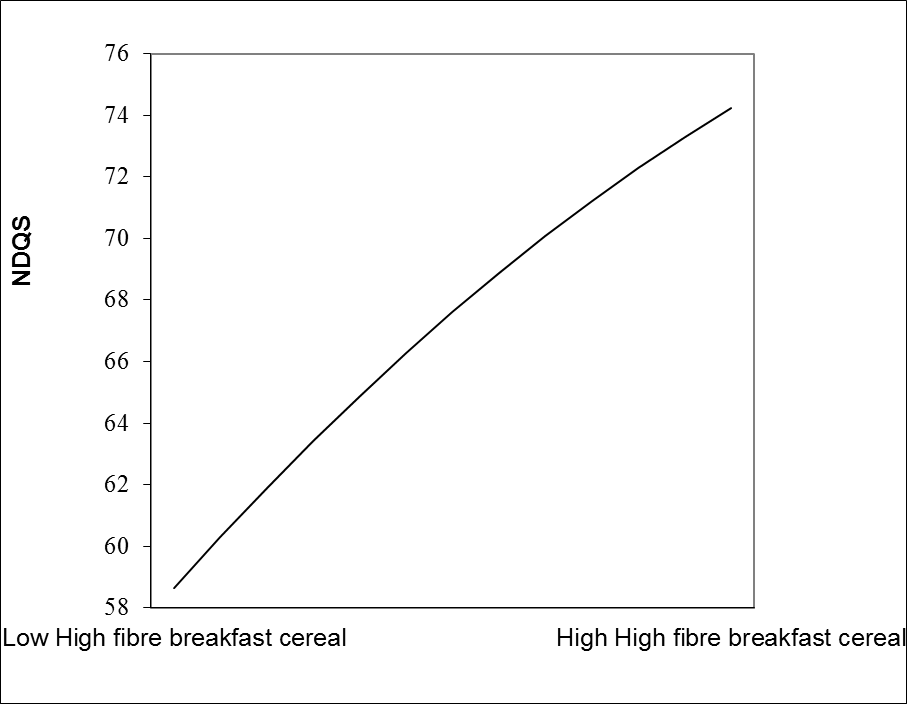
**Figure C.4.1. Quadratic plot: NDQS and wholemeal bread**

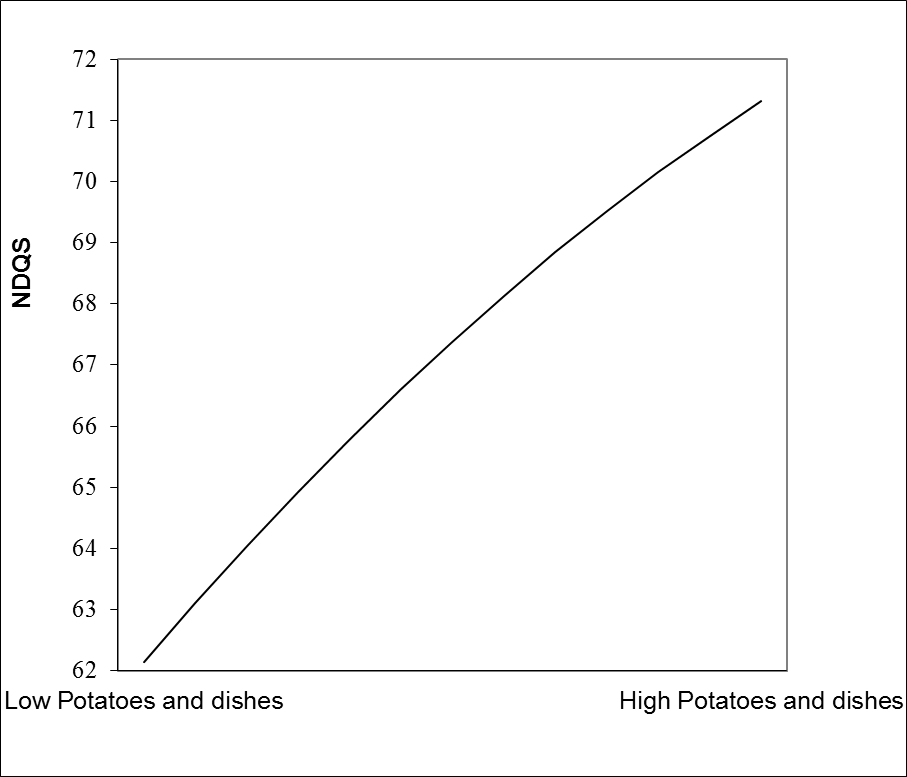
**Figure C.4.2. Quadratic plot: NDQS and salad and other raw vegetables**

**Figure C.4.3. Quadratic plot: NDQS and fruit**

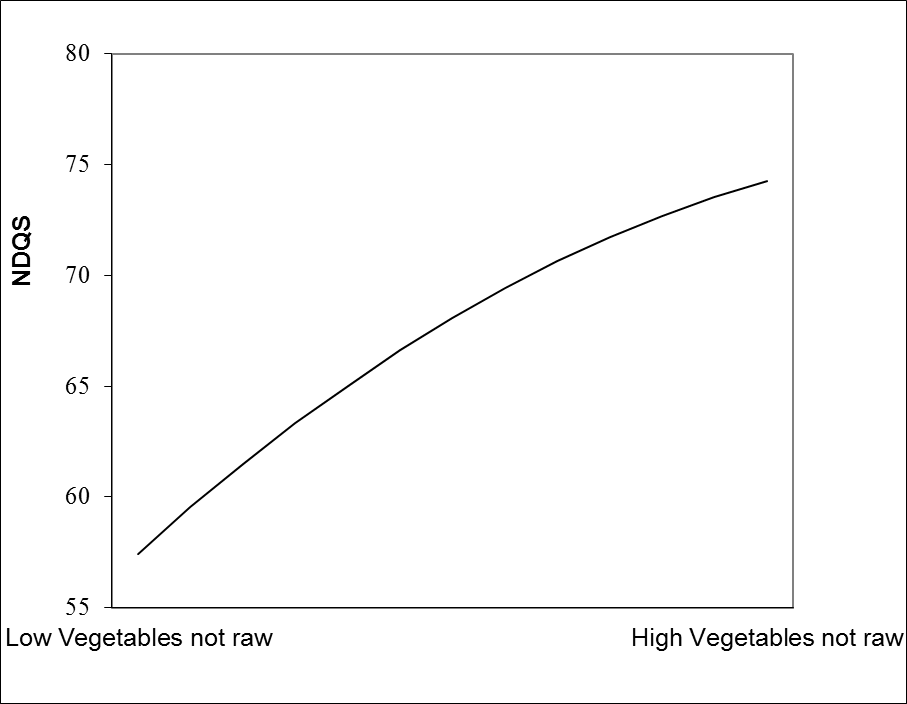
**Figure C.4.4. Quadratic plot: NDQS and burgers and kebabs**

**Figure C.4.5. Quadratic plot: NDQS and high fibre breakfast cereal**

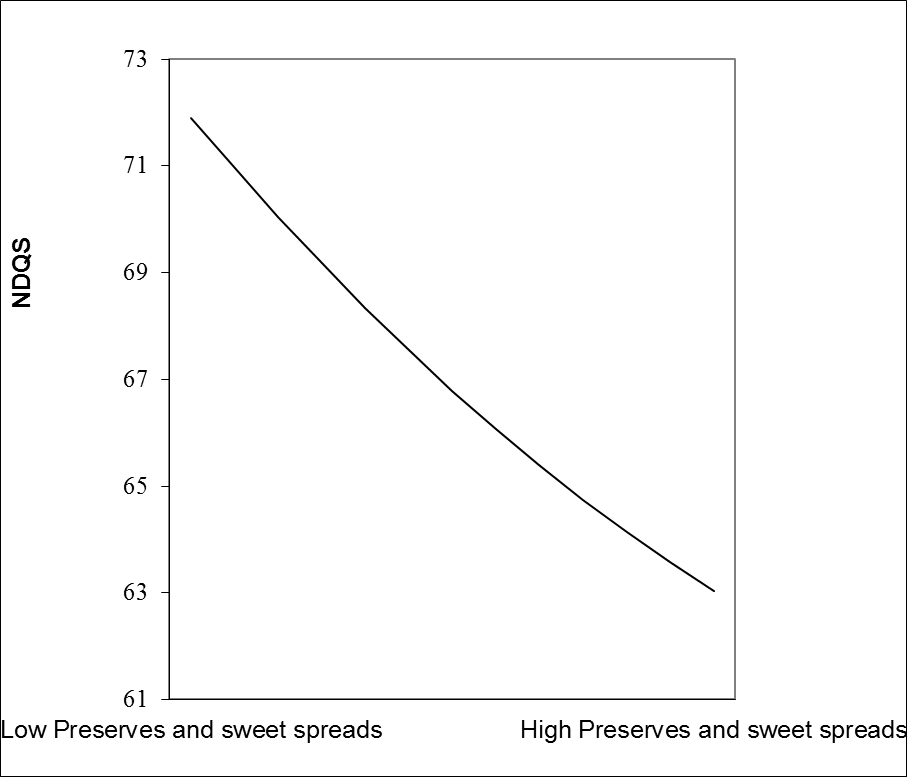
****

**Figure C.4.6. Quadratic plot: NDQS and other potatoes and dishes**

**Figure C.4.7. Quadratic plot: NDQS and vegetables (not raw)**

****

**Figure C.4.8. Quadratic plot: NDQS and preserves and sweet spreads**

****

**Figure C.4.9. Quadratic plot: NDQS and dry weight beverages**

**References**

1. World Health Organisation. *Diet, Nutrition and the Prevention of Chronic Diseases.* WHO Technical Report Series. 2003; 916.

2. Hu FB, van Dam RM, Liu S. Diet and risk of Type II diabetes: the role of types of fat and carbohydrate. *Diabetologia.* 2001;44(7):805-17.

3. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. *The New England Journal of Medicine.* 2000;343(1):16-22.

4. World Cancer Research Fund/American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective.* Washington DC: AICR. 2007.

5. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012;380(9859):2224-60.

6. Jessop E. Health in the UK:could do even better? *The Lancet.* 2013;381(9871):970-2.

7. Rayner M, Scarborough P. The burden of food related ill health in the UK. *Journal of Epidemiology and Community Health.* 2005;59(12):1054-7.

8. NHS Digital. *Statistics on Obesity, Physical Activity and Diet*. Available from: <http://content.digital.nhs.uk/catalogue/PUB13648>. [Accessed January 2017].

9. NHS Digital. *National Child Measurement Programme 2015/16 School Year.* Available from: <http://content.digital.nhs.uk/catalogue/PUB22269/nati-chil-meas-prog-eng-2015-2016-rep.pdf>. [Accessed April 2017].

10. Reilly JJ, Methven E, McDowell ZC, Hacking B, Alexander D, Stewart L, et al. Health consequences of obesity. *Archives of Disease in Childhood.* 2003;88(9):748-52.

11. Berenson GS, Bogalusa Heart Study Group. Health consequences of obesity. *Pediatric Blood & Cancer.* 2012;58(1):117-21.

12. The Marmot Review Team. *Fair Society, Healthy Lives: Strategic Review of the Health Inequalities in England post-2010: The Marmot Review.* London: The Marmot Review; 2010.

13. The Information Centre. *Statistics on Obesity, Physical Activity and Di*e*t: England.* Leeds: The Information Centre; 2010.

14. McLaren L. Socio-economic status and obesity. *Epidemiologic Reviews.* 2007;29:29-48.

15. Tohill, BC. Dietary intake of fruit and vegetables and management of body weight. *World Health Organisation (electronic resource).* Available from: http://www.who.int/dietphysicalactivity/publications/f&v\_weight\_management.pdf. [Accessed January 2017].

16. Mendoza JA, Drewnowski A, Christakis DA. Dietary energy density is associated with obesity and the metabolic syndrome in U.S. adults. *Diabetes Care.* 2007;30(4):974-9.

17. Ambrosini GL, Emmett PM, Northstone K, Howe LD, Tilling K, Jebb SA. Identification of a dietary pattern prospectively associated with increased adiposity during childhood and adolescence. *International Journal of Obesity.* 2012;36(10):1299-305.

18. Wang J, Luben R, Khaw KT, Bingham S, Wareham NJ, Forouhi NG. Dietary energy density predicts the risk of incident type 2 diabetes: the European Prospective Investigation of Cancer (EPIC)-Norfolk Study. *Diabetes Care.* 2008;31(11):2120-5.

19. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *The New England Journal of Medicine.* 2001;345(11):790-7.

20. Howard BV, Wylie-Rosett J. Sugar and cardiovascular disease: A statement for healthcare professionals from the Committee on Nutrition of the Council on Nutrition, Physical Activity, and Metabolism of the American Heart Association. *Circulation.* 2002;106(4):523-7.

21. Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D, et al. Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. *American Journal of Epidemiology*. 1999;149(6):531-40.

22. Blackshaw JR. Public health nutrition in the civil service (England): approaches to tackling obesity. *The Proceedings of the Nutrition Society.* 2016;75(3):356-66.

23. Kopelman P, Jebb SA, Butland B. Executive summary: Foresight 'Tackling Obesities: Future Choices' project. *Obesity Reviews.* 2007;8 Suppl 1:vi-ix.

24. Public Health England. *3.8 million people in England now have diabetes*. Available from: <https://www.gov.uk/government/news/38-million-people-in-england-now-have-diabetes>. [Accessed January 2017].

25. Stringhini S, Tabak AG, Akbaraly TN, Sabia S, Shipley MJ, Marmot MG, et al. Contribution of modifiable risk factors to social inequalities in type 2 diabetes: prospective Whitehall II cohort study. *British Medical Journal*. 2012; 345:e5452.

26. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *International Journal of Epidemiology.* 2011;40(3):804-18.

27. National Obesity Observatory. *Inequalities in obesity prevalence*. Oxford: Public Health England; 2013. Available from: <http://www.noo.org.uk/NOO_about_obesity/inequalities>. [Accessed January 2016].

*28. Saad MF, Knowler WC, Pettitt DJ, Nelson RG, Charles MA, Bennett PH. A two-step model for development of non-insulin-dependent diabetes. The American Journal of Medicine. 1991;90(2):229-*35.

29. Ramachandran AS. Diabetes Mellitus, In: Gibney M (ed.). *Public Health Nutrition.* Tunbridge Wells: Blackwell; 2004. p. 330-340.

30. Steyn NP, Mann J, Bennett PH, Temple N, Zimmet P, Tuomilehto J, et al. Diet, nutrition and the prevention of type 2 diabetes. *Public Health Nutrition.* 2004;7(1A):147-65.

31. Orozco LJ, Buchleitner AM, Gimenez-Perez G, Roque IFM, Richter B, Mauricio D. Exercise or exercise and diet for preventing type 2 diabetes mellitus. *The Cochrane Database of Systematic Reviews*. 2008(3):CD003054.

32. Nield L, Summerbell CD, Hooper L, Whittaker V, Moore H. Dietary advice for the prevention of type 2 diabetes mellitus in adults. *The Cochrane Database of Systematic Reviews*. 2008(3):CD005102.

33. Riserus U, Willett WC, Hu FB. Dietary fats and prevention of type 2 diabetes. *Progress in Lipid Research.* 2009;48(1):44-51.

34. Brand JC, Colagiuri S, Crossman S, Allen A, Roberts DC, Truswell AS. Low-glycemic index foods improve long-term glycemic control in NIDDM. *Diabetes Care.* 1991;14(2):95-101.

35. Hermansen ML, Eriksen NM, Mortensen LS, Holm L, Hermansen K. Can the Glycemic Index (GI) be used as a tool in the prevention and management of Type 2 diabetes? *The Review of Diabetic Studies.* 2006;3(2):61-71.

36. Priebe MG, van Binsbergen JJ, de Vos R, Vonk RJ. Whole grain foods for the prevention of type 2 diabetes mellitus. *The Cochrane Database of Systematic Reviews.* 2008(1):CD006061.

37. Slavin JL. Dietary fiber and body weight. *Nutrition.* 2005;21(3):411-8.

38. Schroder H. Protective mechanisms of the Mediterranean diet in obesity and type 2 diabetes. *The Journal of Nutritional Biochemistry.* 2007;18(3):149-60.

39. Buyken AE, Mitchell P, Ceriello A, Brand-Miller J. Optimal dietary approaches for prevention of type 2 diabetes: a life-course perspective. *Diabetologia*. 2010;53(3):406-18.

40. Salas-Salvado J, Martinez-Gonzalez MA, Bullo M, Ros E. The role of diet in the prevention of type 2 diabetes. *Nutrition, Metabolism, and Cardiovascular Diseases.* 2011;21 Suppl 2:B32-48.

41. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe: epidemiological update. *European Heart Journal.* 2013;34(39):3028-34.

42. NHS/National End of Life Care Intelligence Network. *Deaths from cardiovascular diseases: implications for end of life care in England.* Available from: file:///C:/Users/user/Downloads/Deaths\_from\_CVD\_FINAL%20(1).pdf. [Accessed January 2016].

43. Kotsis V, Stabouli S, Papakatsika S, Rizos Z, Parati G. Mechanisms of obesity-induced hypertension. *Hypertension Research : official journal of the Japanese Society of Hypertension.* 2010;33(5):386-93.

44. Sowers JR, Epstein M, Frohlich ED. Diabetes, hypertension, and cardiovascular disease: an update. *Hypertension*. 2001;37(4):1053-9.

45. Drenjancevic-Peric I, Jelakovic B, Lombard JH, Kunert MP, Kibel A, Gros M. High-salt diet and hypertension: focus on the renin-angiotensin system. *Kidney & Blood Pressure Research.* 2011;34(1):1-11.

46. Hu FB, Bronner L, Willett WC, Stampfer MJ, Rexrode KM, Albert CM, et al. Fish and omega-3 fatty acid intake and risk of coronary heart disease in women. *The Journal of the American Medical Association*. 2002;287(14):1815-21.

47. Marchioli R, Schweiger C, Tavazzi L, Valagussa F. Efficacy of n-3 polyunsaturated fatty acids after myocardial infarction: results of GISSI-Prevenzione trial. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. *Lipids*. 2001;36 Suppl:S119-26.

48. Burr ML. Secondary prevention of CHD in UK men: the Diet and Reinfarction Trial and its sequel. *The Proceedings of the Nutrition Society.* 2007;66(1):9-15.

49. World Cancer Research Fund. *Cancer Statistics.* Available from: https://www.wcrf-uk.org/uk/preventing-cancer/cancer-preventability-statistics. [Accessed January 2016].

50. Gordon-Dseagu V. *Cancer and Health Inequalities: an introduction to the current evidence*. Cancer Research UK. Available from: http://www.cancerresearchuk.org/prod\_consump/groups/cr\_common/@nre/@pol/documents/generalcontent/crukmig\_1000ast-3344.pdf. [Accessed May 2014].

51. Neuhouser ML and Patterson R. *Overview of Nutritional Epidemiology*, In: Coulston A, M & Bouchey, C, J. (eds.). Nutrition in the Prevention and Treatment of Disease. 2nd ed. London: Elsevier; 2008. p. 97-125.

52. Margetts BM and Nelson N. *Overview of the principles of nutritional epidemiology,* In: Margetts BM and Nelson N. (eds). Design Concepts in Nutritional Epidemiology. Oxford: Oxford University Press; 1997. p. 3-63.

53. Department of Health. *Statutory Guidance on Joint Strategic Health Assessments and Joint Health and Wellbeing Strategies.* London: Department of Health; 2011.

54. Department of Health. *Improving Outcomes and Supporting Transparency: Part 2.* London: Department of Health; 2012.

55. Gibson R. *Principles of Nutritional Assessment.* New York: Oxford University Press; 2005.

56. Michels KB. Nutritional epidemiology--past, present, future. *International Journal of Epidemiology.* 2003;32(4):486-8.

57. Hebert JR, Clemow L, Pbert L, Ockene IS, Ockene JK. Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International Journal of Epidemiology.* 1995;24(2):389-98.

58. Coates J, Colaiezzi B, Fiedler JL, Wirth J, Lividini K, Rogers B. A program needs-driven approach to selecting dietary assessment methods for decision-making in food fortification programs. *Food and Nutrition Bulletin*. 2012;33(3 Suppl):S146-56.

59. Roberts K, & Cavill, N. *Standard Evaluation Framework for Dietary Interventions.* Oxford: National Obesity Observatory; 2012.

60. Cavill N, Roberts, K and Ells, L. *Evaluation of weight management, physical activity and dietary interventions.* London: Public Health England; 2015.

61. Livingstone MB, Prentice AM, Strain JJ, Coward WA, Black AE, Barker ME, et al. Accuracy of weighed dietary records in studies of diet and health. *British Medical Journal*. 1990;300(6726):708-12.

62. Hill RJ and Davies PS. The validity of self-reported energy intake as determined using the doubly labelled water technique. *The British Journal of Nutrition.* 2001;85(4):415-30.

63. Patterson REP and Pietinen P. *Assessment of nutritional status in individuals and populations,* In: Gibney MJ, Margetts BM, Kearney JM & Arab L. (eds). Public Health Nutrition. Oxford: Blackwell Science Ltd; 2004. p. 66-82.

64. Johnson RK. Dietary intake--how do we measure what people are really eating? *Obesity Research.* 2002;10 Suppl 1:63S-8S.

65. Lazarte CE, Encinas ME, Alegre C, Granfeldt Y. Validation of digital photographs, as a tool in 24-h recall, for the improvement of dietary assessment among rural populations in developing countries. *Nutrition Journal*. 2012;11:61.

66. Baranowski T, Islam N, Baranowski J, Martin S, Beltran A, Dadabhoy H, et al. Comparison of a Web-based versus traditional diet recall among children. *Journal of the Academy of Nutrition and Dietetics*. 2012;112(4):527-32.

67. Nelson M, Erens B, Bates B, Church S & Bosher T. *Low income diet and nutrition survey - summary of key findings.* Norwich: TSO; 2007.

68. Trabulsi J, Troiano RP, Subar AF, Sharbaugh C, Kipnis V, Schatzkin A, et al. Precision of the doubly labeled water method in a large-scale application: evaluation of a streamlined-dosing protocol in the Observing Protein and Energy Nutrition (OPEN) study. *European Journal of Clinical Nutrition.* 2003;57(11):1370-7.

69. Scagliusi FB, Ferriolli E, Pfrimer K, Laureano C, Cunha CS, Gualano B, et al. Underreporting of energy intake in Brazilian women varies according to dietary assessment: a cross-sectional study using doubly labeled water. *Journal of the American Dietetic Association.* 2008;108(12):2031-40.

70. Tran KM, Johnson RK, Soultanakis RP, Matthews DE. In-person vs telephone-administered multiple-pass 24-hour recalls in women: validation with doubly labeled water. *Journal of the American Dietetic Association*. 2000;100(7):777-83.

71. Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't Veer P. Misreporting of energy and micronutrient intake estimated by food records and 24-hour recalls, control and adjustment methods in practice. *The British Journal of Nutrition*. 2009;101 Suppl 2:S73-85.

72. Cade JE, Burley VJ, Warm DL, Thompson RL, Margetts BM. Food-frequency questionnaires: a review of their design, validation and utilisation. *Nutrition Research Reviews.* 2004;17(1):5-22.

73. Stefanik PA, Trulson MF. Determining the frequency intakes of foods in large group studies. *The American Journal of Clinical Nutrition.* 1962;11:335-43.

74. Willett WC. *Nutritional Epidemiology.* New York: Oxford University Press; 2013.

75. Stumbo PJ. New technology in dietary assessment: a review of digital methods in improving food record accuracy. *The Proceedings of the Nutrition Society.* 2013;72(1):70-6.

76. Carter MC, Hancock N, Albar SA, Brown H, Greenwood DC, Hardie LJ, et al. Development of a New Branded UK Food Composition Database for an Online Dietary Assessment Tool. *Nutrients.* 2016;8(8).

77. Khanna N, Boushey CJ, Kerr D, Okos M, Ebert DS, Delp EJ. An Overview of The Technology Assisted Dietary Assessment Project at Purdue University. *Proceedings of the International Symposium on Multimedia.* 2010:290-5.

78. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C et al. The Automated Self-Administered 24-Hour Dietary Recall (ASA24): A Resource for Researchers, Clinicians and Educators from the National Cancer Institute. *Journal of the Academy of Nutrition and Dietetics.* 2012 Aug; 112(8): 1134–1137.

79. Lee CD, Chae J, Schap TE, Kerr DA, Delp EJ, Ebert DS, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. *Journal of Diabetes Science and Technology.* 2012;6(2):428-34.

80. Albar SA, Alwan NA, Evans CE, Greenwood DC, Cade JE. Agreement between an online dietary assessment tool (myfood24) and an interviewer-administered 24-h dietary recall in British adolescents aged 11-18 years. *The British Journal of Nutrition.* 2016;115(9):1678-86.

81. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *American Journal of Epidemiology.* 2012;175(4):340-7.

82. Swan G. Findings from the latest National Diet and Nutrition Survey. *The Proceedings of the Nutrition Society*. 2004;63(4):505-12.

83. Whitton C, Nicholson SK, Roberts C, Prynne CJ, Pot GK, Olson A, et al. National Diet and Nutrition Survey: UK food consumption and nutrient intakes from the first year of the rolling programme and comparisons with previous surveys. *The British Journal of Nutrition.* 2011;106(12):1899-914.

84. Bates B, Lennox A, Prentice A, Bates C and Swan G. *National Diet and Nutrition Survey: Headline results from Years 1, 2 and 3 (combined) of the Rolling Programme (2008/2009 - 2010/2011).* London: Department of Health; 2012.

85. Rennie KL, Jebb SA, Wright A, Coward WA. Secular trends in under-reporting in young people. *The British Journal of Nutrition*. 2005;93(2):241-7.

86. Public Health England. *National Diet and Nutrition Survey. Results from Years 1-4 (combined) of the Rolling Programme (2008/2009 – 2011/12): Appendix X Misreporting in the National Diet and Nutrition Survey Rolling Programme (NDNS RP): summary of results and their interpretation.* London: Public Health England; 2014.

87. Rennie KL, Coward A, Jebb SA. Estimating under-reporting of energy intake in dietary surveys using an individualised method. *The British Journal of Nutrition.* 2007;97(6):1169-76.

88. Roberts K. *Dietary Surveillance and Nutritional Assessment in England: what is measured and where are the gaps?* Oxford: National Obesity Observatory; 2010.

89. Health and Social Care Information Centre. *Health Survey for England 2014 Volume 2: Methods and Documentation.* Available from: http://content.digital.nhs.uk/catalogue/PUB19295/HSE2014-Methods-and-docs.pdf. [Accessed January 2016].

90. Office of National Statistics. *Health Survey for England 2015 - trend tables.* Available from: https://www.gov.uk/government/statistics/health-survey-for-england-trend-tables-health-survey-for-england-trend-tables-2015. [Accessed February 2017].

91. The Information Centre. Health Survey for England 2009 Volume 1 Health and Lifestyles. Available from: http://content.digital.nhs.uk/catalogue/PUB00414/heal-surv-heal-life-eng-2009-rep-v2.pdf. [Accessed April 2017].

92. Association of Public Health Observatories. *Sources of data on lifestyle risk factors in local populations.*  Available from: http://content.digital.nhs.uk/catalogue/PUB00414/heal-surv-heal-life-eng-2009-rep-v2.pdf. [Accessed February 2017].

93. Department for the Environment, Food and Rural Affairs. *Family Food 2014*. York: DEFRA; 2015.

94. Sport England/TNS BMRB. *Active People Survey 5-7 Technical Report*. London: BMRB; 2013.

95. Public Health England. *Public Health Outcomes Framework: Indicator definitions and supporting information*. Available from: <http://www.phoutcomes.info/public-health-outcomes-framework#page/6/gid/1000042/pat/6/par/E12000004/ati/102/are/E06000015/iid/91477/age/164/sex/4>. [Accessed February 2017].

96. Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. *Social Science & Medicine.* 2006;62(7):1650-71.

97. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology*. 2001;30(4):668-77.

98. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology.* 2002;31(2):285-93.

99. Singh-Manoux A, Ferrie JE, Chandola T, Marmot M. Socio-economic trajectories across the life course and health outcomes in midlife: evidence for the accumulation hypothesis? *International Journal of Epidemiology.* 2004;33(5):1072-9.

100. Singh-Manoux A, Marmot M. Role of socialization in explaining social inequalities in health. *Social Science & Medicine*. 2005;60(9):2129-33.

101. Cade J, Upmeier H, Calvert C, Greenwood D. Costs of a healthy diet: analysis from the UK Women's Cohort Study. *Public Health Nutrition*. 1999;2(4):505-12.

102. El-Sayed AM, Scarborough P, Galea S. Unevenly distributed: a systematic review of the health literature about socio-economic inequalities in adult obesity in the United Kingdom. *BMC Public Health.* 2012;12:18.

103. McLaren L, Kuh D. Women's body dissatisfaction, social class, and social mobility. *Social Science & Medicine.* 2004;58(9):1575-84.

104. Drewnowski A, Hann C. Food preferences and reported frequencies of food consumption as predictors of current diet in young women. *The American Journal of Clinical Nutrition.* 1999;70(1):28-36.

105. Steptoe A, Perkins-Porras L, Rink E, Hilton S, Cappuccio FP. Psychological and social predictors of changes in fruit and vegetable consumption over 12 months following behavioral and nutrition education counseling. *Health Psychology.* 2004;23(6):574-81.

106. Parmenter K, Wardle J. Development of a general nutrition knowledge questionnaire for adults. *European Journal of Clinical Nutrition.* 1999;53(4):298-308.

107. Gerrits JH, O'Hara RE, Piko BF, Gibbons FX, de Ridder DT, Keresztes N, et al. Self-control, diet concerns and eater prototypes influence fatty foods consumption of adolescents in three countries. *Health Education Research.* 2010;25(6):1031-41.

108. Withall J, Jago R, Cross J. Families' and health professionals' perceptions of influences on diet, activity and obesity in a low-income community. *Health & Place.* 2009;15(4):1078-85.

109. Marshall D and Pettinger C. Revisiting British Meals In: Meiselman HL (ed.). *Meals in Science and Practice.* Interdisciplinary research and Business applications: CRC Press; 2009.p. 638-659.

110. Alkerwi A, Baydarlioglu B, Sauvageot N, Stranges S, Lemmens P, Shivappa N, et al. Smoking status is inversely associated with overall diet quality: Findings from the ORISCAV-LUX study. *Clinical Nutrition.* 2016.

111. Department of Health. *Choosing a better diet: a food and health action plan.* London: Deoartment of Health; 2005.

112. Faculty of Public Health. *Food Poverty and Health: a briefing paper.* Available from: http://www.fph.org.uk/uploads/bs\_food\_poverty.pdf. [Accessed February 2017].

113. Timmins KA, Hulme C, Cade JE. The monetary value of diets consumed by British adults: an exploration into sociodemographic differences in individual-level diet costs. *Public Health Nutrition.* 2013:1-9.

114. Roberts K, Cavill, N, Hancock, D & Rutter, H. *Inequalities in diet and physical activity.* Oxford: Public Health England; 2013.

115. Rayner M, Scarborough P, Lobstein. T. *The UK Ofcom Nutrient Profiling Model. Defining 'healthy' and 'unhealthy' foods and drinks for TV advertising to children.* Available from: https://www.ndph.ox.ac.uk/cpnp/about/publications-and-reports/group-reports/uk-ofcom-nutrient-profile-model.pdf. [Accessed March 2017].

116. Jones NR, Conklin AI, Suhrcke M, Monsivais P. The growing price gap between more and less healthy foods: analysis of a novel longitudinal UK dataset. *PloS One*. 2014;9(10):e109343.

117. Food Standards Agency. *Low income diet and nutrition survey: summary findings.* London: TSO; 2007.

118. Dowler E, Caraher M and Lincoln, P. Tackling inequalities in food and nutrition. In: Dowler ES, N (ed.). *Health Inequalities: from Acheson to 'Choosing Health'*. Bristol: Policy Press; 2007. p. 127-56.

119. Ball K, Crawford D, Mishra G. Socio-economic inequalities in women's fruit and vegetable intakes: a multilevel study of individual, social and environmental mediators. *Public Health Nutrition.* 2006;9(5):623-30.

120. Wardle J, Parmenter K, Waller J. Nutrition knowledge and food intake. *Appetite.* 2000;34(3):269-75.

121. Holmes BA, Roberts CL. Diet quality and the influence of social and physical factors on food consumption and nutrient intake in materially deprived older people. *European Journal of Clinical Nutrition*. 2011;65(4):538-45.

122. Cummins S. Neighbourhood food environment and diet: time for improved conceptual models? *Preventive Medicine.* 2007;44(3):196-7.

123. Rees R, Oliver K, Woodman, J and Thomas, J. *Children's views about obesity, body size, shape and weight: a systematic review.* London: EPPI Centre; 2009.

124. Lawton J, Ahmad N, Hanna L, Douglas M, Bains H, Hallowell N. 'We should change ourselves, but we can't': accounts of food and eating practices amongst British Pakistanis and Indians with type 2 diabetes. *Ethnicity & Health*. 2008;13(4):305-19.

125. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: a systematic review. *Health & Place.* 2012;18(5):1172-87.

126. Cummins S, Macintyre S. "Food deserts"--evidence and assumption in health policy making. *British Medical Journa*l. 2002;325(7361):436-8.

127. Wrigley N. Deprivation, diet and food retail access: findings from the Leeds 'food deserts' study. *Environment and Planning.* 2003;3:151-88.

128. Cummins S, Petticrew M, Higgins C, Findlay A, Sparks L. Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *Journal of Epidemiology and Community Health.* 2005;59(12):1035-40.

129. White M, Bunting J, Williams E, Raybould S, Adamson A, Mathers JC. Do 'food deserts' exist? *A multi-level geographical analysis of the relationship between retail food access, socio-economic position and dietary intake.* Newcastle: Food Standards Agency; 2004.

130. Pearson T, Russell J, Campbell MJ, Barker ME. Do 'food deserts' influence fruit and vegetable consumption?--A cross-sectional study. *Appetite*. 2005;45(2):195-7.

131. Shohaimi S, Welch A, Bingham S, Luben R, Day N, Wareham N, et al. Residential area deprivation predicts fruit and vegetable consumption independently of individual educational level and occupational social class: a cross sectional population study in the Norfolk cohort of the European Prospective Investigation into Cancer (EPIC-Norfolk). *Journal of Epidemiology and Community Health*. 2004;58(8):686-91.

132. Thompson C, Cummins S, Brown T, Kyle R. Understanding interactions with the food environment: an exploration of supermarket food shopping routines in deprived neighbourhoods. *Health & Place*. 2013;19:116-23.

133. Drewnowski A, Henderson SA, Levine A, Hann C. Taste and food preferences as predictors of dietary practices in young women. *Public Health Nutrition.* 1999;2(4):513-9.

134. Croker H, Whitaker KL, Cooke L, Wardle J. Do social norms affect intended food choice? *Preventive Medicine*. 2009;49(2-3):190-3.

135. Barker M, Lawrence W, Crozier S, Robinson S, Baird J, Margetts B, et al. Educational attainment, perceived control and the quality of women's diets. *Appetite.* 2009;52(3):631-6.

136. Lawrence W, Skinner C, Haslam C, Robinson S, Inskip H, Barker D, et al. Why women of lower educational attainment struggle to make healthier food choices: the importance of psychological and social factors. *Psychology & Health.* 2009;24(9):1003-20.

137. Kolodinsky J, Harvey-Berino JR, Berlin L, Johnson RK, Reynolds TW. Knowledge of current dietary guidelines and food choice by college students: better eaters have higher knowledge of dietary guidance. *Journal of the American Dietetic Association.* 2007;107(8):1409-13.

138. Beydoun MA, Wang Y. Do nutrition knowledge and beliefs modify the association of socio-economic factors and diet quality among US adults? *Preventive Medicine.* 2008;46(2):145-53.

139. Krebs-Smith SM, Kantor LS. Choose a variety of fruits and vegetables daily: understanding the complexities. *The Journal of Nutrition.* 2001;131(2S-1):487S-501S.

140. Rose D, Bodor JN, Hutchinson PL, Swalm CM. The importance of a multi-dimensional approach for studying the links between food access and consumption. T*he Journal of Nutrition.* 2010;140(6):1170-4.

141. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. *Annual Review of Public Health*. 2008;29:253-72.

142. Scientific Advisory Committee on Nutrition. *The Nutritional Wellbeing of the British Population*. TSO: London; 2008.

143. Scientific Advisory Committee on Nutrition. *Carbohydrates and Health*. Available from: https://www.gov.uk/government/publications/sacn-carbohydrates-and-health-report. [Accessed October 2015].

144. Public Health England. *National Diet and Nutrition Survey Results from Years 5 and 6 (combined) of the Rolling Programme (2012/2013 –2013/2014.* London: Public Health England; 2016.

145. Office of National Statistics. Health Survey for England 2014: Trend Tables 2015. Available from: <http://content.digital.nhs.uk/catalogue/PUB19297>. [Accessed January 2017].

146. Public Health England. *Adult Diet Data Factsheet*. Available from: http://www.noo.org.uk. [Accessed October 2016].

147. Public Health England. *National Diet and Nutrition Survey Results from Years 1, 2, 3 and 4 (combined) of the Rolling Programme (2008/09-2011/12)*. Public Health England: London; 2014.

148. Health and Social Care Information Centre. *Health Survey for England 2013*. Available from: http://content.digital.nhs.uk/catalogue/PUB16076. [Accessed December 2016].

149. Willett W. *Nutritional Epidemiology.* New York: Oxford University Press; 1998.

150. Kant AK. Dietary patterns and health outcomes. *Journal of the American Dietetic Association.* 2004;104(4):615-35.

151. Craig WJ. Phytochemicals: guardians of our health. *Journal of the American Dietetic Association*. 1997;97(10 Suppl 2):S199-204.

152. Liu RH. Potential synergy of phytochemicals in cancer prevention: mechanism of action. *The Journal of Nutrition.* 2004;134(12 Suppl):3479S-85S.

153. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Current Opinion in Lipidology. 2002;13(1):3-9.

154. Mertz W. Foods and nutrients. *Journal of the American Dietetic Association.* 1984;84(7):769-70.

155. Kant AK. Indexes of overall diet quality: a review. *Journal of the American Dietetic Association.* 1996;96(8):785-91.

156. Esposito K, Maiorino MI, Ceriello A, Giugliano D. Prevention and control of type 2 diabetes by Mediterranean diet: a systematic review. *Diabetes Research and Clinical Practice.* 2010;89(2):97-102.

157. de Lorgeril M, Salen P. The Mediterranean-style diet for the prevention of cardiovascular diseases. *Public Health Nutrition.* 2006;9(1A):118-23.

158. North K, Emmett P. Multivariate analysis of diet among three-year-old children and associations with socio-demographic characteristics. The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) Study Team. *European Journal of Clinical Nutrition*. 2000;54(1):73-80.

159. Northstone K, Emmett P, Rogers I. Dietary patterns in pregnancy and associations with socio-demographic and lifestyle factors. *European Journal of Clinical Nutrition.* 2008;62(4):471-9.

160. Newby PK, Tucker KL. Empirically-derived eating patterns using factor or cluster analysis: a review. *Nutrition Reviews.* 2004;62(5):177-203.

161. Cunha DB, Almeida RM, Pereira RA. A comparison of three statistical methods applied in the identification of eating patterns. *Escola Nacional de Saude Publica.* 2010;26(11):2138-48.

162. Reedy J, Wirfalt E, Flood A, Mitrou PN, Krebs-Smith SM, Kipnis V, et al. Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study. *American Journal of Epidemiology.* 2010;171(4):479-87.

163. Schulze MB, Hoffmann K, Kroke A, Boeing H. Risk of hypertension among women in the EPIC-Potsdam Study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. *American Journal of Epidemiology*. 2003;158(4):365-73.

164. Osler M, Helms Andreasen A, Heitmann B, Hoidrup S, Gerdes U, Morch Jorgensen L, et al. Food intake patterns and risk of coronary heart disease: a prospective cohort study examining the use of traditional scoring techniques. *European Journal of Clinical Nutrition.* 2002;56(7):568-74.

165. Greenwood DC, Cade JE, Draper A, Barrett JH, Calvert C, Greenhalgh A. Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *European Journal of Clinical Nutrition*. 2000;54(4):314-20.

166. Lin H, Bermudez OI, Tucker KL. Dietary patterns of Hispanic elders are associated with acculturation and obesity. *The Journal of Nutrition.* 2003;133(11):3651-7.

167. Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus. *The American Journal of Clinical Nutrition.* 2009;90(4):1075-83.

168. Jacques PF, Tucker KL. Are dietary patterns useful for understanding the role of diet in chronic disease? *The American Journal of Clinical Nutrition.* 2001;73(1):1-2.

169. Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *American Journal of Epidemiology.* 2004;159(10):935-44.

170. Roman-Vinas B, Ribas Barba L, Ngo J, Martinez-Gonzalez MA, Wijnhoven TM, Serra-Majem L. Validity of dietary patterns to assess nutrient intake adequacy. *The British Journal of Nutrition.* 2009;101 Suppl 2:S12-20.

171. Smith AD, Emmett PM, Newby PK, Northstone K. Dietary patterns obtained through principal components analysis: the effect of input variable quantification. *The British Journal of Nutrition*. 2013;109(10):1881-91.

172. McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. *Public Health Nutrition.* 2001;4(5):989-97.

173. Kennedy ET, Ohls J, Carlson S, Fleming K. The Healthy Eating Index: design and applications. *Journal of the American Dietetic Association.* 1995;95(10):1103-8.

174. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a Mediterranean diet and survival in a Greek population. *The New England Journal of Medicine.* 2003;348(26):2599-608.

175. McNaughton SA, Ball K, Crawford D, Mishra GD. An index of diet and eating patterns is a valid measure of diet quality in an Australian population. *The Journal of Nutrition.* 2008;138(1):86-93.

176. Waijers PM, Feskens EJ, Ocke MC. A critical review of predefined diet quality scores. *The British Journal of Nutrition*. 2007;97(2):219-31.

177. Wirt A, Collins CE. Diet quality--what is it and does it matter? *Public Health Nutrition.* 2009;12(12):2473-92.

178. Foote JA, Murphy SP, Wilkens LR, Basiotis PP, Carlson A. Dietary variety increases the probability of nutrient adequacy among adults. *The Journal of Nutrition.* 2004;134(7):1779-85.

179. Drewnowski A, Henderson SA, Shore AB, Fischler C, Preziosi P, Hercberg S. Diet quality and dietary diversity in France: implications for the French paradox. *Journal of the American Dietetic Association.* 1996;96(7):663-9.

180. Wosje KS, Khoury PR, Claytor RP, Copeland KA, Hornung RW, Daniels SR, et al. Dietary patterns associated with fat and bone mass in young children. *The American Journal of Clinical Nutrition.* 2010;92(2):294-303.

181. DiBello JR, Kraft P, McGarvey ST, Goldberg R, Campos H, Baylin A. Comparison of 3 methods for identifying dietary patterns associated with risk of disease. *American Journal of Epidemiology*. 2008;168(12):1433-43.

182. Northstone K, Emmett PM. Are dietary patterns stable throughout early and mid-childhood? A birth cohort study. *The British Journal of Nutrition*. 2008;100(5):1069-76.

183. Johnson L, Mander AP, Jones LR, Emmett PM, Jebb SA. Energy-dense, low-fiber, high-fat dietary pattern is associated with increased fatness in childhood. *The American Journal of Clinical Nutrition*. 2008;87(4):846-54.

184. Feinstein L, Sabates R, Sorhaindo A, Rogers I, Herrick D, Northstone K, et al. Dietary patterns related to attainment in school: the importance of early eating patterns. *Journal of Epidemiology and Community Health*. 2008;62(8):734-9.

185. Whichelow MJ, Prevost AT. Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. *The British Journal of Nutrition*. 1996;76(1):17-30.

186. Pryer JA, Nichols R, Elliott P, Thakrar B, Brunner E, Marmot M. Dietary patterns among a national random sample of British adults. *Journal of Epidemiology and Community Health*. 2001;55(1):29-37.

187. Martikainen P, Brunner E, Marmot M. Socio-economic differences in dietary patterns among middle-aged men and women. *Social Science & Medicine.* 2003;56(7):1397-410.

188. Barker ME, McClean SI, Thompson KA, Reid NG. Dietary behaviours and sociocultural demographics in Northern Ireland. *The British Journal of Nutrition*. 1990;64(2):319-29.

189. Williams DE, Prevost AT, Whichelow MJ, Cox BD, Day NE, Wareham NJ. A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. *The British Journal of Nutrition*. 2000;83(3):257-66.

190. Crozier SR, Robinson SM, Borland SE, Inskip HM, Group SWSS. Dietary patterns in the Southampton Women's Survey. *European Journal of Clinical Nutrition*. 2006;60(12):1391-9.

191. Fisk CM, Crozier SR, Inskip HM, Godfrey KM, Cooper C, Robinson SM, et al. Influences on the quality of young children's diets: the importance of maternal food choices. *The British Journal of Nutrition*. 2011;105(2):287-96.

192. Borland SE, Robinson SM, Crozier SR, Inskip HM, Group SWSS. Stability of dietary patterns in young women over a 2-year period. *European Journal of Clinical Nutrition.* 2008;62(1):119-26.

193. Gibson S, Ashwell M. Dietary patterns among British adults: compatibility with dietary guidelines for salt/sodium, fat, saturated fat and sugars. *Public Health Nutrition*. 2011;14(8):1323-36.

194. Hamer M, Mishra GD. Dietary patterns and cardiovascular risk markers in the UK Low Income Diet and Nutrition Survey. *Nutrition, Metabolism, and Cardiovascular Diseases.* 2010;20(7):491-7.

195. Northstone K, Emmett PM, Rogers I. Dietary patterns in pregnancy and associations with nutrient intakes. *The British Journal of Nutrition*. 2008;99(2):406-15.

196. Northstone K, Emmett PM. Dietary patterns of men in ALSPAC: associations with socio-demographic and lifestyle characteristics, nutrient intake and comparison with women's dietary patterns. *European Journal of Clinical Nutrition*. 2010;64(9):978-86.

197. Northstone K, Joinson C, Emmett P, Ness A, Paus T. Are dietary patterns in childhood associated with IQ at 8 years of age? A population-based cohort study. *Journal of Epidemiology and Community Health*. 2012;66(7):624-8.

198. Micali N, Northstone K, Emmett P, Naumann U, Treasure JL. Nutritional intake and dietary patterns in pregnancy: a longitudinal study of women with lifetime eating disorders. *The British Journal of Nutrition*. 2012;108(11):2093-9.

199. Northstone K, Emmett PM. A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis. *The British Journal of Nutrition*. 2008;99(5):1099-106.

200. Mishra GD, McNaughton SA, Bramwell GD, Wadsworth ME. Longitudinal changes in dietary patterns during adult life. *The British Journal of Nutrition*. 2006;96(4):735-44.

201. Mishra GD, dos Santos Silva I, McNaughton SA, Stephen A, Kuh D. Energy intake and dietary patterns in childhood and throughout adulthood and mammographic density: results from a British prospective cohort. *Cancer Causes & Control.* 2011;22(2):227-35.

202. Hamer M, McNaughton SA, Bates CJ, Mishra GD. Dietary patterns, assessed from a weighed food record, and survival among elderly participants from the United Kingdom. *European Journal of Clinical Nutrition*. 2010;64(8):853-61.

203. Robinson S, Syddall H, Jameson K, Batelaan S, Martin H, Dennison EM, et al. Current patterns of diet in community-dwelling older men and women: results from the Hertfordshire Cohort Study. *Age and Ageing.* 2009;38(5):594-9.

204. Shaheen SO, Jameson KA, Syddall HE, Aihie Sayer A, Dennison EM, Cooper C, et al. The relationship of dietary patterns with adult lung function and COPD. *The European Respiratory Journal.* 2010;36(2):277-84.

205. Cleghorn CL, Harrison RA, Ransley JK, Wilkinson S, Thomas J, Cade JE. Can a dietary quality score derived from a short-form FFQ assess dietary quality in UK adult population surveys? *Public Health Nutrition*. 2016;19(16):2915-23.

206. Maynard M, Ness AR, Abraham L, Blane D, Bates C, Gunnell DJ. Selecting a healthy diet score: lessons from a study of diet and health in early old age (the Boyd Orr cohort). *Public Health Nutrition*. 2005;8(3):321-6.

207. Jennings A, Welch A, van Sluijs EM, Griffin SJ, Cassidy A. Diet quality is independently associated with weight status in children aged 9-10 years. *The Journal of Nutrition*. 2011;141(3):453-9.

208. Kennedy E. Dietary diversity, diet quality, and body weight regulation. *Nutrition Reviews.* 2004;62(7 Pt 2):S78-81.

209. Miller PE, Mitchell DC, Harala PL, Pettit JM, Smiciklas-Wright H, Hartman TJ. Development and evaluation of a method for calculating the Healthy Eating Index-2005 using the Nutrition Data System for Research. *Public Health Nutrition*. 2011;14(2):306-13.

210. Patterson RE, Haines PS, Popkin BM. Diet quality index: capturing a multidimensional behavior. *Journal of the American Dietetic Association.* 1994;94(1):57-64.

211. Seymour JD, Calle EE, Flagg EW, Coates RJ, Ford ES, Thun MJ, et al. Diet Quality Index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. *American Journal of Epidemiology.* 2003;157(11):980-8.

212. Huijbregts P, Feskens E, Rasanen L, Fidanza F, Nissinen A, Menotti A, et al. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and The Netherlands: longitudinal cohort study. *British Medical Journal.* 1997;315(7099):13-7.

213. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, et al. Diet and overall survival in elderly people. *British Medical Journal*. 1995;311(7018):1457-60.

214. Lagiou P, Trichopoulos D, Sandin S, Lagiou A, Mucci L, Wolk A, et al. Mediterranean dietary pattern and mortality among young women: a cohort study in Sweden. *The British Journal of Nutrition*. 2006;96(2):384-92.

215. World Health Organisation. Diet, nutrition, and the prevention of chronic diseases. Report of a WHO Study Group. *World Health Organization Technical Report Series*. 1990;797:1-204.

216. National Research Council. *Diet and health: implications for reducing chronic disease risk.* Washington DC: National Academy of Sciences; 1989.

217. Quatromoni PA, Pencina M, Cobain MR, Jacques PF, D'Agostino RB. Dietary quality predicts adult weight gain: findings from the Framingham Offspring Study. *Obesity.* 2006;14(8):1383-91.

218. Newby PK, Hu FB, Rimm EB, Smith-Warner SA, Feskanich D, Sampson L, et al. Reproducibility and validity of the Diet Quality Index Revised as assessed by use of a food-frequency questionnaire. *The American Journal of Clinical Nutrition*. 2003;78(5):941-9.

219. Fung TT, McCullough ML, Newby PK, Manson JE, Meigs JB, Rifai N, et al. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. *The American Journal of Clinical Nutrition*. 2005;82(1):163-73.

220. Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index revised: a measurement instrument for populations. *Journal of the American Dietetic Association*. 1999;99(6):697-704.

221. Huijbregts PP, Feskens EJ, Rasanen L, Fidanza F, Alberti-Fidanza A, Nissinen A, et al. Dietary patterns and cognitive function in elderly men in Finland, Italy and The Netherlands. *European Journal of Clinical Nutrition*. 1998;52(11):826-31.

222. Knoops KT, Groot de LC, Fidanza F, Alberti-Fidanza A, Kromhout D, van Staveren WA. Comparison of three different dietary scores in relation to 10-year mortality in elderly European subjects: the HALE project. *European Journal of Clinical Nutrition*. 2006;60(6):746-55.

223. US Department of Agriculture/Health and Human Services. *Nutrition and your health: Dietary Guidelines for Americans.* US Government: Washington DC; 1990.

224. Hann CS, Rock CL, King I, Drewnowski A. Validation of the Healthy Eating Index with use of plasma biomarkers in a clinical sample of women. *The American Journal of Clinical Nutrition*. 2001;74(4):479-86.

225. Weinstein SJ, Vogt TM, Gerrior SA. Healthy Eating Index scores are associated with blood nutrient concentrations in the third National Health And Nutrition Examination Survey. *Journal of the American Dietetic Association.* 2004;104(4):576-84.

226. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HA, Kuczynski KJ, et al. Update of the Healthy Eating Index: HEI-2010. *Journal of the Academy of Nutrition and Dietetics.* 2013;113(4):569-80.

227. Yang J, Farioli A, Korre M, Kales SN. Modified Mediterranean diet score and cardiovascular risk in a North American working population. *PloS One.* 2014;9(2):e87539.

228. England PH. Government Dietary Recommendations. London: Public Health England; 2016.

229. Pot GK, Richards M, Prynne CJ, Stephen AM. Development of the Eating Choices Index (ECI): a four-item index to measure healthiness of diet. *Public Health Nutrition*. 2014;17(12):2660-6.

230. Simon A, O'Connell, R and Stephen, A. Designing a nutritional scoring system for assessing dietary quality for children aged 10 years and under in the UK. *Methodological Innovation Online.* 2012;7(2):27-40.

231. England CY, Andrews RC, Jago R, Thompson JL. A systematic review of brief dietary questionnaires suitable for clinical use in the prevention and management of obesity, cardiovascular disease and type 2 diabetes. *European Journal of Clinical Nutrition*. 2015;69(9):977-1003.

232. Heller RF, Pedoe HD, Rose G. A simple method of assessing the effect of dietary advice to reduce plasma cholesterol. *Preventive Medicine*. 1981;10(3):364-70.

233. Roe L, Strong C, Whiteside C, Neil A, Mant D. Dietary intervention in primary care: validity of the DINE method for diet assessment. *Family Practice.* 1994;11(4):375-81.

234. Rifas-Shiman SL, Willett WC, Lobb R, Kotch J, Dart C, Gillman MW. PrimeScreen, a brief dietary screening tool: reproducibility and comparability with both a longer food frequency questionnaire and biomarkers. *Public Health Nutrition*. 2001;4(2):249-54.

235. Cade JE, Burley VJ, Greenwood DC, Group UKWsCSS. The UK Women's Cohort Study: comparison of vegetarians, fish-eaters and meat-eaters. *Public Health Nutrition*. 2004;7(7):871-8.

236. Aljadani HM, Patterson A, Sibbritt D, Hutchesson MJ, Jensen ME, Collins CE. Diet quality, measured by fruit and vegetable intake, predicts weight change in young women. *Journal of Obesity.* 2013;2013:525161.

237. Wang Q, Chen Y, Wang X, Gong G, Li G, Li C. Consumption of fruit, but not vegetables, may reduce risk of gastric cancer: results from a meta-analysis of cohort studies. *European Journal of Cancer.* 2014;50(8):1498-509.

238. Cappuccio FP, Rink E, Perkins-Porras L, McKay C, Hilton S, Steptoe A. Estimation of fruit and vegetable intake using a two-item dietary questionnaire: a potential tool for primary health care workers. *Nutrition, Metabolism, and Cardiovascular Diseases*. 2003;13(1):12-9.

239. Scientific Advisory Committee on Nutrition. *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom.* London: TSO; 2011.

240. Ells LJ, Cavill N, Roberts K, Rutter H. Development of a standard evaluation framework for weight management interventions. *Public Health.* 2013;127(4):345-7.

241. Relton C, Bissell P, Smith C, Blackburn J, Cooper CL, Nicholl J, et al. South Yorkshire Cohort: a 'cohort trials facility' study of health and weight - protocol for the recruitment phase. *BMC Public Health.* 2011;11:640.

242. Food Standards Agency/Public Health England. *National Diet and Nutrition Survey Rolling Programme User Guide for UK Core Sample Data.* London: PHE; 2014.

243. Food Standards Agency/PHE. *Nutrition Diet and Nutrition Survey Rolling Programme - User Guide for UK Core Sample Data.* London: NatCen Social Research; 2015.

244. UK Data Service. *National Diet and Nutrition Survey 1-4, 2008/09-2011/12*. Available from: https://discover.ukdataservice.ac.uk/catalogue/?sn=6533. Accessed February 2014.

245. Cunha DB, Varnier, R. M, de Almeda, R, Pereira, R.A. A comparison of three statistical methods applied in the identification of eating patterns. *Cadernos de Saude Publica*. 2010;26(11):2138-48.

246. Field A. *Discovering Statistics using SPSS*. London: Sage Publications; 2009.

247. Pot GK, Stephen AM, Dahm CC, Key TJ, Cairns BJ, Burley VJ, et al. Dietary patterns derived with multiple methods from food diaries and breast cancer risk in the UK Dietary Cohort Consortium. *European Journal of Clinical Nutrition*. 2014;68(12):1353-8.

248. Bakolis I, Hooper R, Thompson RL, Shaheen SO. Dietary patterns and adult asthma: population-based case-control study. *Allergy.* 2010;65(5):606-15.

249. McCourt HJ, Draffin CR, Woodside JV, Cardwell CR, Young IS, Hunter SJ, et al. Dietary patterns and cardiovascular risk factors in adolescents and young adults: the Northern Ireland Young Hearts Project. *The British Journal of Nutrition*. 2014;112(10):1685-98.

250. Northstone K, Emmett P. Multivariate analysis of diet in children at four and seven years of age and associations with socio-demographic characteristics. *European Journal of Clinical Nutrition*. 2005;59(6):751-60.

251. Committee on Medical Aspects of Food Policy. *Nutritional Aspects of Cardiovascular Disease,* *Department of Health Report on Health and Social Subjects.* London: HMSO; 1994.

252. Public Health England. *Why 5%? An explanation of SACN's recommendations about sugars and health.* Available from: https://discover.ukdataservice.ac.uk/catalogue/?sn=6533. [Accessed February 2017].

253. Scientific Advisory Committee on Nutrition. *Salt and Health.* London: TSO; 2003.

254. Toft U, Kristoffersen LH, Lau C, Borch-Johnsen K, Jorgensen T. The Dietary Quality Score: validation and association with cardiovascular risk factors: the Inter99 study. *European Journal of Clinical Nutrition*. 2007;61(2):270-8.

255. Thiele S, Mensink GB, Beitz R. Determinants of diet quality. *Public Health Nutrition*. 2004;7(1):29-37.

256. Guenther PM, Reedy J, Krebs-Smith SM. Development of the Healthy Eating Index-2005. *Journal of the American Dietetic Association.* 2008;108(11):1896-901.

257. Conklin AI, Monsivais P, Khaw KT, Wareham NJ, Forouhi NG. Dietary Diversity, Diet Cost, and Incidence of Type 2 Diabetes in the United Kingdom: A Prospective Cohort Study. *PLoS medicine.* 2016;13(7):e1002085.

258. Committee on Medical Aspects of Food Policy. *Dietary reference values for food energy and nutrients for the United Kingdom. Report of the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy.* Report on Health and Social Subjects. 1991;41:1-210.

259. McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, et al. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *The American Journal of Clinical Nutrition*. 2002;76(6):1261-71.

260. Darmon N, Drewnowski A. Does social class predict diet quality? *The American Journal of Clinical Nutrition*. 2008;87(5):1107-17.

261. Vernarelli JA, Mitchell DC, Rolls BJ, Hartman TJ. Dietary energy density is associated with obesity and other biomarkers of chronic disease in US adults. *European Journal of Nutrition.* 2015;54(1):59-65.

262. Public Health England. *Sugar Reduction: The evidence for action*. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/470179/Sugar\_reduction\_The\_evidence\_for\_action.pdf. [Accessed March 2017].

263. Jin Y, Gordon MH, Alimbetov D, Chong MF, George TW, Spencer JP, et al. A novel combined biomarker including plasma carotenoids, vitamin C, and ferric reducing antioxidant power is more strongly associated with fruit and vegetable intake than the individual components. *The Journal of Nutrition*. 2014;144(11):1866-72.

264. Pennant M, Steur M, Moore C, Butterworth A, Johnson L. Comparative validity of vitamin C and carotenoids as indicators of fruit and vegetable intake: a systematic review and meta-analysis of randomised controlled trials. *The British Journal of Nutrition*. 2015;114(9):1331-40.

265. Zarrin R, Ibiebele TI, Marks GC. Development and validity assessment of a diet quality index for Australians. *Asia Pacific Journal of Clinical Nutrition*. 2013;22(2):177-87.

266. Aranceta J, Perez-Rodrigo C. Recommended dietary reference intakes, nutritional goals and dietary guidelines for fat and fatty acids: a systematic review. *The British Journal of Nutrition*. 2012;107 Suppl 2:S8-22.

267. Barasi M. *Human Nutrition*. London: Hodder Arnold; 2003.

268. Department of Health. *UK Chief Medical Officers' Alcohol Guidelines Review*. London: Williams Lea; 2016.

269. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: systematic review and meta-analyses. *British Medical Journal.* 2013;346:f1326.

270. Heaney RP. Dairy intake, dietary adequacy, and lactose intolerance. *Advanced Nutrition.* 2013;4(2):151-6.

271. Brevik A, Vollset SE, Tell GS, Refsum H, Ueland PM, Loeken EB, et al. Plasma concentration of folate as a biomarker for the intake of fruit and vegetables: the Hordaland Homocysteine Study. *The American Journal of Clinical Nutrition*. 2005;81(2):434-9.

272. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *The American Journal of Clinical Nutrition*. 1997;65(4 Suppl):1220S-8S; discussion 9S-31S.

273. Gerrior S, Juan W, Basiotis P. An easy approach to calculating estimated energy requirements. *Preventing Chronic Disease.* 2006;3(4):A129.

274. Livingstone MB, Black AE. Markers of the validity of reported energy intake. *The Journal of Nutrition*. 2003;133 Suppl 3:895S-920S.

275. Northstone K, Ness AR, Emmett PM, Rogers IS. Adjusting for energy intake in dietary pattern investigations using principal components analysis. *European Journal of Clinical Nutrition*. 2008;62(7):931-8.

276. Mendez MA, Popkin BM, Buckland G, Schroder H, Amiano P, Barricarte A, et al. Alternative methods of accounting for underreporting and overreporting when measuring dietary intake-obesity relations. *American Journal of Epidemiology*. 2011;173(4):448-58.

277. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, et al. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *European Journal of Clinical Nutrition*. 1991;45(12):569-81.

278. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *International Journal of Obesity and Related Metabolic Disorders.* 2000;24(9):1119-30.

279. World Health Organisation. Physical Status: the use and interpretation of anthropometry. Technical Report of a WHO Expert Committee Series No: 854. Available from: http://www.who.int/childgrowth/publications/physical\_status/en/. [Accessed March 2017].

280. Maguire ER, Monsivais P. Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data. *The British Journal of Nutrition*. 2015;113(1):181-9.

281. Potischman NaF, J.L. Biomarkers of Nutritional Exposure and Nutritional Status: An overview. *American Society for Nutritional Sciences*. 2003;133:873S-4S.

282. Markussen MS, Veierod MB, Sakhi AK, Ellingjord-Dale M, Blomhoff R, Ursin G, et al. Evaluation of dietary patterns among Norwegian postmenopausal women using plasma carotenoids as biomarkers. *The British Journal of Nutrition*. 2015;113(4):672-82.

283. Neuhouser ML, Patterson RE, King IB, Horner NK, Lampe JW. Selected nutritional biomarkers predict diet quality. *Public Health Nutrition*. 2003;6(7):703-9.

284. Gibney M, Macdonald, I.A and Roche, H.M. Human Nutrition. Oxford: Blackwell Science 2002.

285. Krinsky NI, Johnson EJ. Carotenoid actions and their relation to health and disease. *Molecular Aspects of Medicine*. 2005;26(6):459-516.

286. Brevik A, Andersen LF, Karlsen A, Trygg KU, Blomhoff R, Drevon CA. Six carotenoids in plasma used to assess recommended intake of fruits and vegetables in a controlled feeding study. *European Journal of Clinical Nutrition*. 2004;58(8):1166-73.

287. El-Sohemy A, Baylin A, Ascherio A, Kabagambe E, Spiegelman D, Campos H. Population-based study of alpha- and gamma-tocopherol in plasma and adipose tissue as biomarkers of intake in Costa Rican adults. *The American Journal of Clinical Nutrition*. 2001;74(3):356-63.

288. Ganguly P, Alam SF. Role of homocysteine in the development of cardiovascular disease. *Nutrition Journal.* 2015;14:6.

289. Zerwekh JE. Blood biomarkers of vitamin D status. *The American Journal of Clinical Nutrition*. 2008;87(4):1087S-91S.

290. World Health Organisation. Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations. Geneva: World Health Organisation; 2011. Available from: <http://www.who.int/vmnis/indicators/serum_ferritin.pdf>. {Accessed October 2016].

291. Upadhyay RK. Emerging risk biomarkers in cardiovascular diseases and disorders. *Journal of Lipids.* 2015;2015:971453.

292. Shrivastava A, Singh, H, Raizada, A and Singh, S. . C-reactive protein, inflammation and coronary heart disease. *The Egyptian Heart Journal.* 2015;67(2):89-97.

293. Almoosawi S, Cole D, Nicholson S, Bayes I, Teucher B, Bates B, et al. Biomarkers of diabetes risk in the National Diet and Nutrition Survey rolling programme (2008-2011). *Journal of Epidemiology and Community Health*. 2014;68(1):51-6.

294. Newby PK, Muller D, Tucker KL. Associations of empirically-derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *The American Journal of Clinical Nutrition*. 2004;80(3):759-67.

295. Costello A, B., and Osbourne, J.W. . Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation.* 2005;10(7).

296. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Biological Medicine.* 2008;3:17.

297. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*. 2012;24(3):69-71.

298. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prevention Science.* 2000;1(4):173-81.

299. Golley RK, Smithers LG, Mittinty MN, Emmett P, Northstone K, Lynch JW. Diet quality of U.K. infants is associated with dietary, adiposity, cardiovascular, and cognitive outcomes measured at 7-8 years of age. *The Journal of Nutrition*. 2013;143(10):1611-7.

300. Crozier SR, Inskip HM, Barker ME, Lawrence WT, Cooper C, Robinson SM, et al. Development of a 20-item food frequency questionnaire to assess a 'prudent' dietary pattern among young women in Southampton. *European Journal of Clinical Nutrition*. 2010;64(1):99-104.

301. Akbaraly TN, Ferrie JE, Berr C, Brunner EJ, Head J, Marmot MG, et al. Alternative Healthy Eating Index and mortality over 18 y of follow-up: results from the Whitehall II cohort. *The American Journal of Clinical Nutrition*. 2011;94(1):247-53.

302. Ruxton C. Women's Diet Quality in the UK. *Nutrition Bulletin.* 2010;35(2):126-37.

303. Allender S, Owen B, Kuhlberg J, Lowe J, Nagorcka-Smith P, Whelan J, et al. A Community Based Systems Diagram of Obesity Causes. *PloS One.* 2015;10(7):e0129683.

304. Finegood DT, Merth TD, Rutter H. Implications of the foresight obesity system map for solutions to childhood obesity. *Obesity.* 2010;18 Suppl 1:S13-6.

305. UK Statistics Authority. *Types of Official Statistics/List of national statistics*. Available from: <https://www.statisticsauthority.gov.uk/national-statistician/types-of-official-statistics/list-of-national-statistics/>. [Accessed January 2017].

306. Shaneshin M, Jessri M, Rashidkhani B. Validity of energy intake reports in relation to dietary patterns. *Journal of Health, Population and Nutrition.* 2014;32(1):36-45.

307. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. *Frontiers in Nutrition.* 2014;1:5.

308. Rayner M, Scarborough P, Lobstein. T. *The UK Ofcom Nutrient Profiling Model. Defining 'healthy' and 'unhealthy' foods and drinks for TV advertising to children.* Available from: https://www.ndph.ox.ac.uk/cpnp/about/publications-and-reports/group-reports/uk-ofcom-nutrient-profile-model.pdf. [Accessed March 2017].

309. Hobbs M, Green MA, Roberts K, Griffiths C, McKenna J. Understanding the relationship between fast food outlets and body mass index: a structural equation modelling approach. *Plos One.* May 2017. In submission.

1. 142. Scientific Advisory Committee on Nutrition. The Nutritional Wellbeing of the British Population. 2008. [↑](#footnote-ref-1)
2. 143. Scientific Advisory Committee on Nutrition. Carbohydrates and Health. In: England PH, editor. London: TSO; 2015. [↑](#footnote-ref-2)
3. 144. Public Health England. National Diet and Nutrition Survey Results from Years 5 and 6 (combined) of the Rolling Programme (2012/2013 –2013/2014. In: England PH, editor. London: Public Health England; 2016. [↑](#footnote-ref-3)
4. 93. Department of Environment FaRA. Family Food 2014. In: Department of Environment FaRA, editor. York: DEFRA; 2015. [↑](#footnote-ref-4)
5. 145. Office of National Statistics. Health Survey for England 2014: Trend Tables 2015 [Available from: http://content.digital.nhs.uk/catalogue/PUB19297. [↑](#footnote-ref-5)
6. The KMO statistic varies between 0 and 1 with a value of 0 indicating that the sum of partial correlations is large relative to the sum of correlations, indicating diffusion in the pattern of correlations (i.e. factor analysis is likely to be inappropriate). A value close to 1 indicates that patterns of correlations are relatively compact and so factor analysis should yield distinct and reliable factors 246. Field A. Discovering Statistics using SPSS. London: Sage Publications; 2009. [↑](#footnote-ref-6)
7. National Statistics socio-economic Classifications: 1-Higher managerial, administrative and professional occupations; 2 - Lower managerial, administrative and professional occupations; 3 - Intermediate occupations; 4-Small employers and own account workers; 5- Lower supervisory and technical occupations; 6-Semi-routine occupations; 7-Routine occupations; Never worked and long-term unemployed ([http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html Accessed 21st June 2016](http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html%20Accessed%2021st%20June%202016)). [↑](#footnote-ref-7)