# Spatial analysis of dietary cost patterns and implications for health

**Michelle Anne Morris**

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

**University of Leeds**

**School of Food Science and Nutrition**

**School of Geography**

**School of Medicine**

**December 2013**

# Intellectual property and publications

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given in the thesis where reference has been made to the work of others.



*Jointly Authored Publications*

Chapter 3, Section 3.4 of this thesis is based on work completed in equal collaboration with fellow PhD student Kate Timmins.

This work has been published as:

Title: Comparability of methods assigning monetary costs to diets: derivation from household till receipts versus cost database estimation from four-day food diaries.

Reference: *European Journal of Clinical Nutrition,* 67**,** 1072-1076.

Authors: Timmins K, Morris M, Hulme C, Edwards K, Clarke G, Cade J.

# Academic Acknowledgements

This research has been funded through an ESRC/MRC interdisciplinary PhD studentship.

I would like to acknowledge: Kate Timmins with whom I jointly completed the analysis for the cost database evaluation, detailed in chapter 3; Kevin Tarbutt (funded via a Rank Prize Fund vacation studentship) and Edmund Parks who created and updated the DANTE cost database; Claire Oyston and Joan Ransley who created the SNIP dataset and Andrea Smyth for creating single women in the UKWCS validation dataset; Lorna Fraser for advice regarding regression analysis for chapter 5 and Alexandra Hatherley for assistance with postcode cleaning to be used in the spatial analysis chapters.

Thank you to the Centre for Spatial Analysis and Policy at the University of Leeds for funding the majority of the research visit to the University of Queensland, Australia. Gita Mishra helped to make the visit possible with guidance through the data application process and encouragement throughout. During this visit the support and guidance from Gita Mishra, Caroline Jackson and David Fitzgerald made the visit a success and enjoyable.

The research visit to the University of Washington was mostly funded by an ESRC Overseas Institution visit award. This visit was possible through the support of Adam Drewnowski in the application process and also during the visit. Support and encouragement from Adam Drewnowski , Anju Aggarwal, Pablo Monsivais and Phil Hurvitz during the visit contributed to the success and positive experience.

Thanks to my transfer viva assessors Victoria Burley and Darren Dahly for their feedback. Thank you also for your encouragement and advice during my MSc studies which contributed to my choice to pursue a PhD in the first place.

Thank you to the participants of the UKWCS, SNIP, ALSWH and SOS.

Finally, suggestions received from my viva assessors Victoria Burley and Jamie Pearce were greatly appreciated and further enhanced the thesis.

# Acknowledgements

To acknowledge everyone who has made such an impact on my PhD experience is no easy task as so many people have contributed to making the last three years so enjoyable and memorable.

Firstly I'd like to thank my Mum and Dad for making me independent and ambitious enough to pursue a PhD in the first place! I'd also like to quote my Gran who has regularly reminded me "if it were easy everyone would do it" which has often been reassuring over the last three years.

James, thank you for being you, keeping me sane and grounded, making me go out and run when I get too highly strung and keeping the tea and biscuits flowing over the final few weeks of writing.

My PhD supervisors, Janet, Claire, Graham and Kim have been supportive and good company throughout. Thank you for helping to tame my ambitious ideas, but encouraging and supporting me along the way. I appreciate you all for your different contributions to my PhD experience and my development as a person and a researcher. To have been included equally in your research groups, despite my interdisciplinary studentship sharing my focus across the three areas has meant a lot. I feel very lucky to have worked with you all.

Thank you to the Nutritional Epidemiology Group, the Centre for Spatial Analysis and Policy and the Health Economics Group - staff and fellow PhD students. Being surrounded by such brilliant minds and personalities has made this PhD such a great experience - and I have gained some amazing friends.

# Abstract

*Background:* Chronic diseases such as obesity, cardiovascular disease, diabetes and cancer are a large burden on society, for which diet is the leading modifiable risk factor. 'Diet' can encompass a variety of aspects of food consumption. Dietary patterns arguably best reflect food as consumed. The determinants of diet are varied and include cost and availability of foods. The aim of this research is to investigate if dietary patterns and diet cost vary spatially and whether this influences health outcomes, specifically obesity and breast cancer.

*Methods:* Using data from the UK Women's Cohort Study, data driven dietary patterns were previously determined. Cost of diet was assigned using a food cost database. Spatial measures for Government Office Region, North South, Urban/ Rural and the Output Area Classification were assigned matched to postcode of the women. Weight status is calculated from self reported height and weight. Time to event analysis investigated association between dietary pattern, diet cost and breast cancer incidence at each spatial scale.

*Results:* There is some spatial variation in dietary patterns both between and within regions. A healthy in more expensive per day than a less healthy one: £6.63 compared to £3.29. The overweight/ obese pay more for their food. Urban and Northern areas have significantly higher BMI than Rural and Southern areas respectively. Those in areas Constrained by Circumstance have highest prevalence of overweight and obesity. There is some spatial variation in breast cancer incidence and variation by dietary pattern. In postmenopausal women, positive association exists between weight status and risk of breast cancer incidence. Methods used for estimating small area dietary patterns and health outcomes may be applicable for use in other developed populations.

*Conclusion:* Understanding determinants of dietary patterns remains important for public health and making healthy diets accessible to all is important. However, while expensive dietary patterns reflect a healthier diet, they do not appear to be the mechanism for which obesity prevalence and breast cancer incidence occur. Geodemographic classifications, combined with other spatial measures could aid more effective targeting of public health nutrition policy.

# Table of contents

# Table of tables

# Table of figures

# Abbreviations

| | |
|---|---|
| AHEI | Alternate Healthy Eating Index |
| ALWSH | Australian Longitudinal Study for Women's Health |
| AMED | Alternate Mediterranean Diet Score |
| BMI | Body Mass Index |
| CIDER | Centre for Interaction Data Estimation and Research |
| CPI | Consumer Price Index |
| CNPP | Centre for Nutrition Policy and Promotion |
| DANTE | Diet and Nutrition Tool for Evaluation |
| DRV | Dietary Reference Value |
| FFQ | Food Frequency Questionnaire |
| HEI | Healthy Eating Index |
| HDI | Healthy Diet Indicator |
| IMD | Index of Multiple Deprivation |
| INSEE | French National Institute of Statistics |
| MED | Mediterranean Diet Score |
| NCI | National Cancer Institute (USA) |
| NEG | Nutritional Epidemiology Group |
| NHANES | National Health and Nutrition Evaluation Survey |
| NHS | National Health Service |
| NHSIC | National Health Service Information Centre |
| NUTS1 | Nomenclature of Territorial Units for Statistics 1 |
| OAC | Output Area Classification |
| ONS | Office of National Statistics |
| SEIFA | Socio-Economic Indexes for Areas |
| SES | Socioeconomic status |
| SNIP | Supermarket Nutrition Information Project |

| | |
|---|---|
| SOS1 | Seattle Obesity Study 1 |
| UKWCS | United Kingdom Women's Cohort Study |
| US | United States |
| USA | United States of America |
| USDA | United States Department of Agriculture |
| WCRF | World Cancer Research Fund |
| WHO | World Health Organisation |

# Chapter 1: Introduction

## 1.1 Overview

This chapter provides some general background to the relevance of my research and places the research in context. The aims and objectives are described and related to the relevant sections of the thesis. The overall flow of the thesis is also presented.

## 1.2 Background

*Diet as a determinant of health*

Non communicable diseases - such as obesity, cardiovascular disease, diabetes and cancer - are a large burden on society, for which diet is a common modifiable risk factor. Diet is considered to be the leading risk factor for non-communicable diseases, such as obesity, cardiovascular disease, diabetes and cancer in the developed world (Key et al., 2002, Doll and Peto, 1981, US Burden of Disease Collaborators, 2013, Institute for Health Metrics and Evaluation, 2013). Diet is a broad term which can encompass a variety of aspects of food consumption. This study will look at dietary patterns which will allow for analysis of food habits taking account of the interactive effect of food combinations. This is a practical approach given that people eat food combinations rather than macronutrients or micronutrients in isolation.

*Determinants of diet*

After taste, cost is one of the main factors influencing diet (Glanz et al., 1998, Lennernas et al., 1997). Cost of food can be a barrier to consumption of a 'healthy' diet, which has been shown to be more expensive than more energy dense alternatives (Maillot et al., 2007a). This may cause a greater problem in low income groups.

It has been suggested that the environmental context in which we live can influence diet (Swinburn et al., 1999), and that the social composition of where we live also plays a role in diet and health (Macintyre et al., 2002, Pearce et al., 2007). Spatial analysis methods are a powerful means of analysing and visualising patterns within a given population. In the UK there are clearly defined boundaries at different population levels, which facilitate spatial analysis; for example, Nomenclature of Territorial Units for Statistics 1 (NUTS1) areas which incorporate the Government Office Regions of England and Scotland and Wales. The UK is often dichotomised by a North-South divide. Much research considers differences according to whether an individual lives in an urban or rural environment which can influence access to facilities and also behaviour, depending on the nature of the surroundings.

In addition to the physical location in which an individual lives, people can be classified into groups based on their demographic characteristics and the composition of an area in which they live, at a small geographical scale, referred to as geodemographic or market segmentation classifications. They are widely used in marketing, but to a lesser extent in health (Abbas et al., 2009). The Output Area Classification (OAC), is one such classification, created using variables from the 2001 UK census at the geographical unit of Output Area (Vickers et al., 2005). This research will use this classification in a cohort of UK women to investigate spatial dietary cost patterns. Profiling dietary patterns and diet cost by geodemographic group may mean that results are more generalisable to the UK population, and not just a specific geographic area. The classification developed by Vickers et al is particularly useful in research as the methodology used is reported in full. Other classifications are available, such as Cameo (Callcredit Information Group) and Acorn (CACI Ltd). However, these have been developed for commercial use and therefore methods used to develop them are less transparent. Using geodemographic classifications personal characteristics, such as education and determinants of socio-economic status, are already accounted for, so additional adjustment for these characteristics are not required in statistical models, making for a more straightforward tool in analysis.

*Available data*

In the UK national data is collected on diet and nutrition (Department of Health and Food Standards Agency, 2012), family spending on food (Department for Environment Food and Rural Affairs, 2004) and health (Health & Social Care Information Centre, 2011) but sample sizes are often limited for research purposes, with insufficient power to detect meaningful differences. Other countries do not have such good quality national data meaning that researchers and policy makers alike need to maximise research effort using the data which they do have, such as cohort data for various research questions. While there are ethical issues looking at patterns at small area geographies, where it could be possible to identify individuals, it is of utmost importance that we understand diet and health variation at a small area geography, in order to develop policy and provide relevant public health services. Geodemographic classifications could be the key to achieving this without jeopardising anonymity because, whilst a small area unit of geography is used to match an individual to a geodemographic group, once the group is assigned the person is no longer identifiable from that group. Typically a whole population is grouped into between 5 and 60 groups, hence an individual would not be personally identifiable.

*Relevance of the spatial variations on diet and health*

Over recent decades, lifestyle factors, including diet, have changed, resulting in higher rates of non-communicable diseases such as obesity, diabetes, cardiovascular disease and certain types of cancer. Responsibility for the management of lifestyle occurs at a number of levels. Some research focuses on the responsibility of an individual, while other research will look at community or society factors and other environmental surroundings. This can be summarised by the widely published Ecological Model which illustrates how individuals, environment and policy can interact to impact on behaviour (Olsen, 2009, Bronfenbrenner, 1977, Gibney, 2004). Examples of similar compositional environments may exist at different geographic locations.

Figure 1.1 shows the Ecological model as the environment composition along with how the geographic location - environment context - may also contribute to influences on behaviour, independently or via the environment composition.

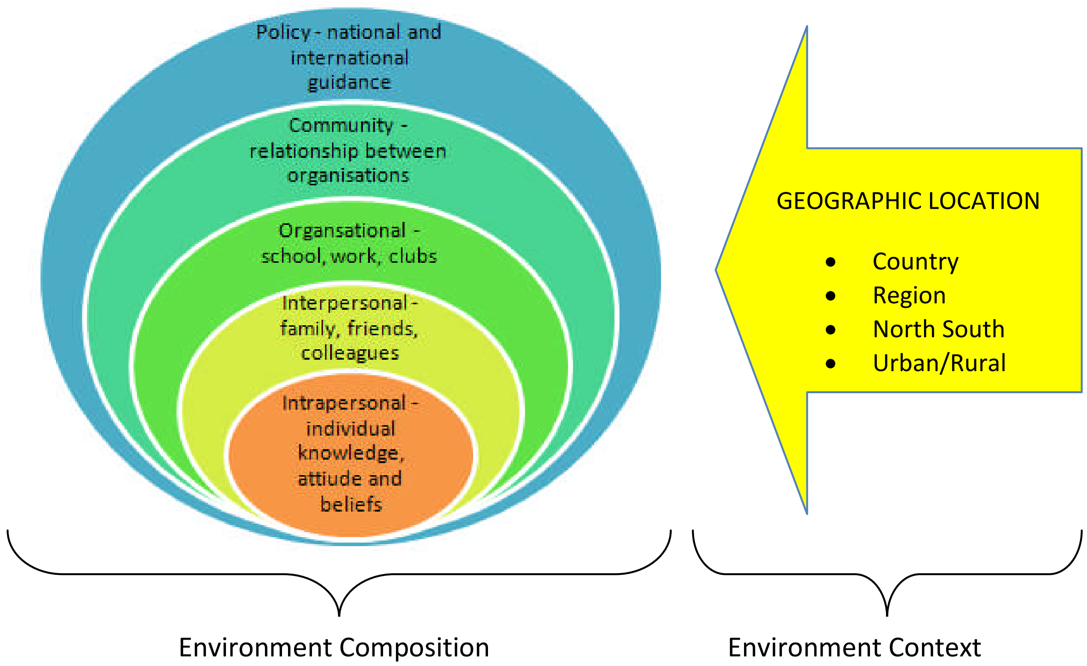



Figure 1.1 - Environment context and composition summarised using the Ecological model and geographic scales. Source: the environment composition section is adapted from Bronfenbrenner's ecological framework (Bronfenbrenner, 1977).

Better understanding of how environmental influences interact with behaviours, such as diet, could go a long way in helping to promote a healthy diet and prevent diet related chronic diseases. The research in this thesis focuses primarily on the environment context but considers environment composition to some extent using a geodemographic classification and through accounting for socio-demographic characteristics.

## 1.3 Aims and objectives

The thesis aims are described in section 1.3.1. Objectives are described in each chapter and summarised in section 1.3.2, Table 1.1.

### 1.3.1 Aims

- To investigate geographies of diet and diet cost patterns
- To investigate the impact of geographies of diet and diet cost patterns on health, specifically obesity and breast cancer
- To evaluate use of different spatial measures in diet and health research
- To investigate the value of geodemographic methods for estimating diet and health patterns in the UK and elsewhere

### 1.3.2 Objectives

| | Objective | Chapter |
|---|---|---|
| 1 | To summarise the existing evidence relating to the cost of a healthy diet. | 2 |
| 2 | To review the literature relating to geographic variation in the cost of diet. | 2 |
| 3 | To summarise the existing evidence relating to geographic variations in diet. | 2 |
| 4 | To explain the data and methods used to meet objectives 5, 7-23. | 3 |
| 5 | To evaluate the DANTE food cost database as an effective tool in estimating population diet costs. | 3 |
| 6 | To categorise the healthiness of dietary patterns in the UKWCS | 4 |
| 7 | To analyse the cost of dietary patterns in the UKWCS and how this cost is related to healthiness of the diet. | 4 |
| 8 | To evaluate the use of geodemographic classifications for public health. | 5,6 & 7 |
| 9 | Explore variations in dietary patterns across the UK according to four spatial scales: at GOR level, by North South, by Urban/Rural classification and using geodemographic Supergroups (which are created at Output Area level). | 5 |
| 10 | Explore variations in diet cost across the UK according to four spatial scales: at GOR level, by North South, by Urban/Rural classification and using geodemographic Supergroups (which are created at Output Area level). | 5 |
| 11 | Use the results generated, by geodemographic classification, to estimate small area dietary patterns and diet cost for a large UK city. | 5 |

| | | |
|---|---|---|
| 12 | Consider the application of results at different spatial scales in a public health context. | 5 |
| 13 | Describe overweight/obesity in the UKWCS using GOR, North South, Urban/Rural and OAC Supergroup. | 6 |
| 14 | Describe cost of diet by weight status and also by spatial measures. | 6 |
| 15 | Investigate associations between overweight/obesity, diet cost and GOR, North South, Urban/Rural and OAC in the UKWCS, using both a continuous measure of BMI and by dichotomising the sample as normal weight or overweight/obese. | 6 |
| 16 | To describe the characteristics of UKWCS women according to whether they have developed breast cancer or not. | 7 |
| 17 | To investigate whether there is a difference in the observed and expected breast cancer incidence in a time to event analysis by predictor variables: GOR, North South, Urban/Rural, dietary pattern, diet cost and weight status. | 7 |
| 18 | To investigate the hazard ratio for breast cancer incidence by predictor variables: GOR, North South, Urban/Rural, dietary pattern, diet cost and weight status. | 7 |
| 19 | Use Cameo geodemographic classifications to describe diet (fruit and vegetable consumption) and weight status in the Australian Longitudinal Study for Women's Health (ALSWH) and the Seattle Obesity Study (SOS1). | 8 |
| 20 | Explore use of Cameo geodemographic classification to estimate small area diet or health patterns in an Australian and USA city. | 8 |
| 21 | To discuss the findings in this thesis and their relevance to public health and policy. | 9 |
| 22 | To identify possible future work following on from the analysis in this thesis. | 9 |
| 23 | To conclude the pertinent findings of this thesis. | 10 |

Table 1.1 - Research objectives

## 1.4 Thesis overview

This thesis comprises an introduction, literature review, methods and data chapter, two main results chapters and three case study results chapters followed by a conclusion, which are summarised in Figure 1.2.

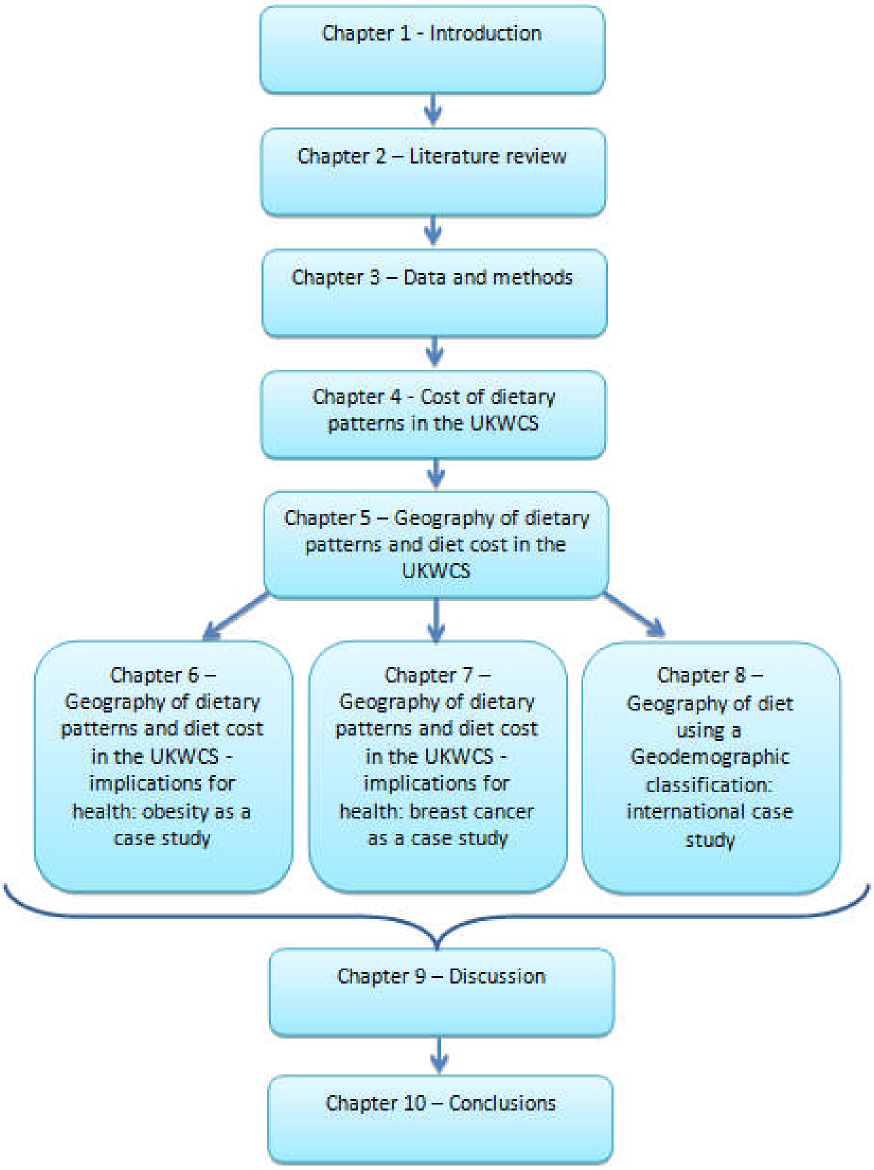



Figure 1.2 - Thesis flow

This research uses data collected for the UK Women's Cohort Study (UKWCS), which recruited approximately 35000 women in the late 1990s and collected in-depth diet and lifestyle information. These women are still being followed up with respect to cancer and cause of death outcomes. Further description of this data is provided in Chapter 3.

The type of food we eat is likely to be influenced by cost and how food is valued by the consumer. There is some notion that a healthy diet is an expensive one (Darmon et al., 2004). In addition, research shows that individuals on a lower income are more likely to consume a poorer quality of diet, comprising energy dense, nutrient poor, foods (Drewnowski et al., 2007). This can lead to inferior health states, thus introducing inequality in health due to differences in an individual's budget for food. Measuring the cost of diet is difficult and it has been noted that linking diet cost databases to food records is an important area for future research (Willett, 2013). Chapter 3 of this thesis includes an evaluation of a food cost database and a discussion regarding its applicability for use in population research. Applying the cost data to dietary records and analysing how the cost of dietary patterns in UK women vary, is discussed in Chapter 4. Chapter 4 also describes how data driven dietary patterns from the UKWCS have been assigned a healthiness score by comparing the contents to the Department of Health's Eatwell Plate, in order that dietary patterns and health outcomes can be discussed in relation to how healthy a diet is.

Recent messages from the UK Chief Medical Officer have highlighted the importance of including spatial analysis in research (Davies, 2011). Four different geographic measures are used and discussed with regards to the relevance to public health. Spatial profiling of dietary patterns and diet cost in a large cohort of UK women is carried out in Chapter 5. Using these profiles it is possible to estimate the diet of women in specific neighbourhoods of the UK using the OAC groups as a unit for estimation. This method is one means of providing neighbourhood estimates of diet without the overhead of collecting data at this level, which could mean that individuals are identifiable. This method of estimation could be useful in other developed countries for which geodemographic classifications exist. Using the Cameo geodemographic classification diet and aspects of health have been profiled for Australia and the USA and are described in Chapter 8.

Understanding determinants of diet is important with respect to long term health. An obvious impact of diet is on obesity - a result of chronic energy imbalance - which is rife in the UK and worldwide. Obesity has a number of associated co-morbidities, such as cardiovascular disease, hypertension, diabetes and cancer. Spatial variations in weight status in the UKWCS are discussed in Chapter 6, along with how spatial variations in dietary patterns and diet cost impact on rates of overweight and obesity. Diet has also been linked to a number of other chronic diseases, aside from obesity, including cancer. In the UK, breast cancer is the most prevalent cancer, despite it predominantly affecting women. There is limited, non conclusive research regarding the role of diet on breast cancer risk, but an overall healthy diet, with limited alcohol, shows decreased risk (Key et al., 2002, WCRF, 2007). Breast cancer rates vary geographically, therefore Chapter 7 explores spatial variation in breast cancer in the UKWCS and investigates whether dietary pattern, diet cost or weight status are the mechanisms for such variation.

## 1.5 Discussion

*Interdisciplinary research*

This research has been carried out during an ESRC/MRC Interdisciplinary PhD Studentship combining research areas: Nutritional Epidemiology and Geography and Health Economics. Interdisciplinary research is important (Foresight, 2007), to try to address current public health concerns in the UK and worldwide. Researchers need to come together and pool their resources and skills and develop researchers with specific cross discipline expertise. The research in this thesis is one example of how this can work to contribute to a better understanding of public health matters relating to nutrition, obesity and inequalities in health.

Nutritional epidemiologists are interested in the effects of food on health over time in a population, psychologists are interested in food choice behaviours and health economists in how the price of food alters behaviour and the cost of diet choices in relation to health outcomes. Geographers investigate the influence of the environment on food and activity choice. The reality is that a combined, interdisciplinary approach is required to better understand the different aspects of determinants of behaviour (in this case dietary behaviour) at a combination of societal levels.

*Potential impact*

The results from this thesis will add to the scant literature on the cost of dietary patterns, how these vary spatially and how these might impact upon health. The results have the potential to have an impact on public health nutrition policy, to add to the literature on inequalities in health and also add to existing methods for estimating small area dietary patterns in the UK and in other developed countries.

Many recent reports such as, the National Institute of Clinical Excellence (NICE) 2012 guidelines (National Institute for Clinical Excellence, 2012), the Chief Medical Officer's report

2011 (Davies, 2011) and the 'Fair Society Healthy Lives' Marmot review 2010 (Marmot, 2010), list priorities aimed at reducing inequalities and inequities and creating happier and healthier communities. The importance of the local community is stressed. Results from this thesis will help inform policy makers to achieve this.

To date there is no research in the UK investigating spatial variations in dietary cost patterns and how they might impact on health, supporting the need for the research in this thesis.

## 1.6 Summary and context

This chapter has set the scene for the thesis. Chapter 2 will provide a review of the existing literature to date.

# Chapter 2: Literature review

## 2.1 Overview

This chapter will set the research in context of the existing literature. Due to the interdisciplinary nature of this research the search areas are diverse, with a large body of literature surrounding them. Literature relating to the case studies in chapters 6, 7 and 8 will be described in those specific chapters. A reminder of how this review fits into the thesis flow can be seen in Figure 2.1.

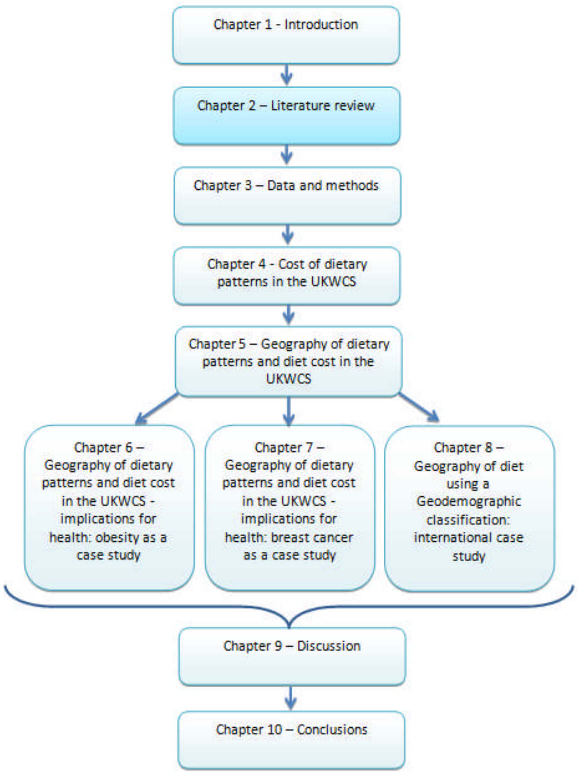



Figure 2.1 - Overview showing how Chapter 2 fits into the overall thesis flow

## 2.2 Introduction

Cost of diet is an important factor in food choice. Food prices are not static and can vary between geographical regions, between retail outlets in the same region and between different brands on the same shelf in a store. The amount which we are willing to spend on food is heavily influenced by the budget we have for food which itself is often determined by income. Other socioeconomic factors such as education and occupation are associated with food choice. Such factors are closely linked with the area in which we live, and in some cases the terms are used almost interchangeably. Where you live can also influence access to foods and food behaviours. There is not a wide body of literature specifically investigating spatial variations in cost of diet, so this chapter will also discuss the cost of diet and geographic variations in diet.

### 2.2.1 Objectives

The objectives of this Chapter, as described in Chapter 1, Table 1.1 are to:

- To summarise the existing evidence relating to the cost of a healthy diet
- To review the literature relating to geographic variation in the cost of diet
- To summarise the existing evidence relating to geographic variations in diet

## 2.3 Methods

The initial literature search was of a semi-systematic nature. Medline, Embase and Food Science and Technology Abstract databases were searched using OvidSP. Search terms are listed in Table 2.1. These were combined using Boolean operators. Wild cards and adjacency searching were applied as appropriate. Inclusion and exclusion criteria are detailed in Table 2.2. Titles and abstracts for search results were reviewed and relevant papers were imported into EndNote for de-duplication across the databases. Only the full text of relevant articles were obtained. Some further studies were then excluded after full text review.

The initial searching identified a number of key papers in the area, and some prolific authors. Using a snowball strategy and citation searching, a number of papers were identified to discuss in order to meet the aims of this review. Notification alerts were set up in order to remain abreast of new papers in the field.

| Search component | Search terms |
|---|---|
| Diet | diet*<br>snack*<br>fruit*<br>vegetable*<br>diet* adj4 (pattern* or provision*)<br>nutrition* adj4 pattern*<br>food* adj4 (pattern* or grocer* or provision* or habit* or consum* or intake)<br>eat* adj4 (pattern* or habit* or provision* or regime* or preference) |
| Cost | Costs and Cost Analysis<br>budget<br>cost<br>cost* adj2 (effective* or utility* or benefit* or minimi* or evaluat* or analy* or study or studies or consequenc* or compar* or efficienc*)<br>price* or pricing*<br>finance$ or financial*<br>fee* or expend* or spend*<br>value adj1 (money or monetary) |
| Spatial | (cluster or urban or rural or output or classification) adj3 (analysis or area*)<br>"hot spot" or "cold spot" or "spatial analysis" or cluster* or spatial* or geodemograph* or geo-demograph* or "ecological correlation"<br>"enumeration district" or topograph* or boundar* or "output area" or "super group" or super?group or geo?references or cartograph*<br>Socio* adj2 (demograph* or geograph* or econom*) |

Table 2.1 - Literature review search terms

| Included | Excluded |
|---|---|
| Original studies | Studies which report only beverage consumption: Alcohol, drinking water, Sugar sweetened beverages |
| Reviews | Studies which report on food which is not for human consumption |
| | Fortification of foods |
| | Supplements |
| | Breast feeding |
| | Animal studies |
| | Studies not written in the English language |
| | Studies only including children, adolescents, pregnant or nursing mothers, elderly |
| | Non-western populations |
| | Studies relating to taxes, subsidies, price elasticity, affordability of food |

Table 2.2 - Inclusion/exclusion criteria for diet cost literature searches

## 2.4 Results and discussion

It is difficult to be able to compare findings of these studies due to the very different measures used to: (1) assess diet, (2) measure cost of diet and (3) apply to a geographical scale. These observations highlight the need to use a framework in order to report results in a consistent manner. The ANGELO framework incorporates environment and cost (Swinburn et al., 1999) at both a macro and micro environment scale, covering areas relating to physical, economic, political and sociocultural environments. With this in mind the results of this literature review will be presented in line with relevant areas of the ANGELO framework, which are summarised in

Table 2.3.

| | Type of setting | | | |
|---|---|---|---|---|
| **Size of setting** | Physical | Economic | Political | Sociocultural |
| Microenvironment | See sections 2.4.1, 2.4.3 and 2.4.6 | See sections 0, 2.4.3 and 2.4.7 | See section 2.4.3.1 | See sections, 2.4.6, 2.4.6 and 2.4.7 |
| Macroenvironment | | | | |

Table 2.3 - Summary of the 2x4 format of the ANGELO framework

### 2.4.1 Physical environment

A snowball strategy has been used to search for relevant literature to summarise the evidence of geographic variations in diet. Key papers are discussed. These have not been tabulated as for the diet cost section of this review, due to the diversity in methods of assessing a food environment, thus making it more intuitive to discuss rather than tabulate.

Understanding food environments is important, not just in relation to food consumption, but to the subsequent effects of diet on health. It is well known that diet affects many non-communicable chronic disease health outcomes. There is a vast body of evidence discussing food environments at a range of geographical scales (both macro and micro environments). Access to food, both geographically and via affordability, along with whether certain types of areas adopt different food choices and what the suggested mechanisms for this are of particular interest in this review.

#### *2.4.1.1 Access to healthy food*

Compositional (social makeup) verses context (location) effects of an area have been debated (Macintyre et al., 2002). Macintyre et al review the literature relating to the 'place effect' exposing mixed results of place on health. Some area differences in health disappear when controlling for socioeconomic variables, yet in other studies differences remain. Some results show that it is economic and social inequality in an area which is associated with biggest differences in health outcomes as it is this inequality which influences behaviour (Wilkinson and Pickett, 2010).

Diet as consumed is not recorded in many of these studies, but instead the availability of food is. Many studies measure the distance between place of residence and food retail outlet: a fast food restaurant, a health food store or a supermarket. The hypothesis for many of these studies is that the closer you live to a retail outlet, the higher your likelihood of making purchases from that outlet and hence influencing your diet and subsequent health outcomes, such as obesity. However, this is not always the case. In New Zealand, Pearce et al (2007) found that those who lived closer to supermarkets selling fresh fruit and vegetables, were in fact those who lived in more deprived neighbourhoods, who are most likely to be overweight or obese than those in less deprived areas, therefore suggesting that it is it not proximity to facilities which necessarily influences dietary behaviour and subsequent weight status (Pearce et al., 2007). In a later study using a national sample in New Zealand, Pearce et al (2008) observed no association with proximity to supermarkets and convenience stores with fruit and vegetable intake (Pearce et al., 2008). In fact, while not convincing, there was some evidence to suggest a negative effect of living in close proximity to convenience stores (a type of retail outlet) on vegetable intake. The authors suggest that convenience stores may be a proxy measure of energy dense convenience foods, although in the absence of dietary consumption data it is not possible to be certain.

An article by Cummins (2007) discusses the ‘local trap’, a reference to the immediate environment surrounding a person’s place of residence and how this does not necessarily reflect their ‘action’ or ‘activity’ space. An activity space is defined as “the space which we live in from day to day”. So an activity space may imply that people may choose to shop somewhere on their commute to work, a school run, or a regular weekly trip to a place of leisure, which is external to their residential neighbourhood but part of their activity space (Cummins, 2007). Activity spaces are more difficult to measure, yet there is interesting emerging research investigating nutrition related behaviour within an activity space (Gustafson et al., 2013). Larger geographical units (macroenvironment) such as region, or in the UK a North-South split, while highly likely to capture activity spaces, may lose detail which a smaller unit (microenvironment) could capture. Also, larger regions have a much more varied demographic characteristic spread.

*Food deserts*

Whether we consider place of residence, or an activity space of an individual, if there is no accessible retail outlet selling ‘healthy’ food, we may call this a ‘food desert’. There is a wealth of literature discussing food deserts, and whether or not they are the reason why the poor have less healthy diets and subsequent inequalities in health, or whether they have just become a ‘factoid’ (i.e. something which is presented as the truth, but is in fact not proven) (Wrigley, 2002, Walker et al., 2010, Cummins and Macintyre, 2002a).

A review by Larson et al (2009) discusses the literature relating to retail food stores and restaurants with respect to influence on dietary intake, obesity and access inequalities in the USA. They conclude from the 54 research studies they evaluated, that residents of low income, minority and rural neighbourhoods are most often affected by poor access to supermarkets and healthy food - a food desert. It is also in these areas where there is sometimes a higher availability of fast food outlets. As alluded to in the first sentence of this section on food deserts, how a food desert is defined can influence research outcomes. A recent USA study by Jiao et al (2012) considers the importance of how a food desert is defined. In their research results differ considerably by modifying the definition subtly to perhaps better reflect access to food outlets. For example, as few as 3% low income populations could walk to a low-cost supermarket. However, almost all could take a ten minute drive or bus ride to a low or medium cost supermarket. So by changing the transport mode and the supermarket budget type access has changed from almost none to almost the entire sample, highlighting that it is essential to assign definitions relevant to the study population.

Some studies in the UK and elsewhere have shown that introducing a new retail outlet into a food desert can increase fruit and vegetable consumption in the surrounding area, by a small but statistically significant amount (Wrigley et al., 2002). However, in a similar natural experiment which controlled for age, sex, education and employment, no effect was observed (Cummins, 2005b). Larsen and Gilliland (2009) found that the introduction of a farmers market (and therefore adding a greater element of competition) into a food desert in Canada reduced fruit and vegetable prices in the surrounding area by 12% over three years, having a positive impact for consumers.

### 2.4.2 Economic environment

The area of cost of diet is of emerging importance and there are a number of researchers keenly following developments in this area. This can be seen by a number of recent published reviews.

In October 2013, Lee et al published a review 'Monitoring the price and affordability of foods and diets globally' (Lee et al., 2013). This review sets in context different areas of food cost influence and how these can relate to a 'healthy diet'. Many areas relating to diet cost are covered at a high level, with key papers identified. This review only skims the surface of the literature, but in doing so sets the scene well for the requirement of a framework in order to ensure that future research is consistent and meets criteria which is useful and comparable in different populations. One message in this paper is that more needs to be learnt from existing data using minimal new data collection. The authors also argue that there is an urgent need to assess the price and affordability of healthy and less healthy foods. This paper is published as part of a series released by the INFORMAS (International Network for Food and Obesity/non-communicable diseases Research, Monitoring, Action and Support) team – a group established to monitor and benchmark food environments on a global scale. The INFORMAS team also suggest a framework for measuring the environment and its effect on dietary consumption through monitoring both public and private sector action (Brinsden et al., 2013, Swinburn et al., 2013). The INFORMAS group support existing policy and frameworks, such as the NOURISHING framework set up by the World Cancer Research Fund (WCRF). The group includes researcher, Boyd Swinburn, author of the ANGELO framework described in
Table 2.3 above.

In 2011, a different Lee et al completed a systematic review investigating the 'Influence of food cost on diet quality and risk factors for chronic disease' (Lee et al., 2011). This review highlights different components of the cost of diet including; cost of healthy and less healthy foods, food prices and food consumption, affordability of food, association between food cost and diet quality and links between food price and chronic disease risk factors. The author concludes that altering cost of foods at a macroenvironment level and monitoring the effect over time is essential to better understand the impact of diet cost on long term health.

A third example is a systematic review and meta-analysis 'Do healthier foods or diet patterns cost more than less healthy options? A systematic review and meta-analysis', which to date only has an abstract published (Rao et al., 2013), with a full paper under review (as of Oct 2013). While this review doesn't take into account any geographical variation in diet, it will offer some promising results with regards to quantification of the cost of dietary patterns, although given the limitations identified in the two Lee reviews related to inconsistencies in how diet cost is recorded, a meta analysis may be challenging.

#### *2.4.2.1 Are healthier diets more expensive?*

With such a broad body of literature relating to diet cost, it has been important for me to focus on the literature specifically relevant to this research: the cost of diet as consumed, which occurs at a microenviromental level, as it is relating to the individuals consumption. In order to answer this, the literature relating to diet cost and diet quality of food consumed will be discussed. Diet quality alludes to the healthiness of diet and is commonly expressed using dietary patterns. Studies which investigate this are limited due to the lack of reliable food cost data. This must either be done through collection of food expenditure data, or by using a food cost database linked to dietary data. It is these studies linking cost of foods to dietary data which will be discussed. Papers relating to this are summarised in Table 2.4.

| Reference | Country | Population | Dietary record method | Diet cost method | Dietary quality measure | Summary of results |
|---|---|---|---|---|---|---|
| (Aggarwal et al., 2012) | USA | 1266 adults (mean age 56 years) comprising 804 women and 462 men in 2008/9 | 152 item FFQ (384 component foods) | Lowest Seattle supermarket retail prices for 384 FFQ component items | Nutrient consumption of vitamins A, C, D, E, B12, beta carotene, folate, iron, calcium, potassium, magnesium and fibre. Also saturated fats, trans fats and added sugars | Higher nutrient intake associated with higher diet cost. Lower nutrient intake associated with lower diet cost |
| (Andrieu et al., 2006) | France | 1474 adults - aged 15 years and over (672 men and 802 women) in 1998 | 7 day food record - with photographic portion sizes | French mean national 1997 retail prices obtained from marketing research for 760 foods. Other prices obtained from supermarket websites and the French National Institute of Statistics | Selected vitamins and mean energy density | Energy dense diets are nutrient poor and less expensive |
| (Bernstein et al., 2010) | USA | 78191 women with data collected in 2002 | 116 item FFQ (467 food components) | 2001/2002 online USDA food-cost database. For 27 components not found, national average prices or online prices were assigned and corrected used consumer price index to February 2002 prices | Alternative Healthy Eating Index (AHEI) | Those in highest AHEI quintile spend 24% more than those in the lowest, but a range of costs occurred in each quintile. Greater spending on nuts, soy and beans and whole grains were associated with higher AHEI score. Greater spending on red and processed meats and high fat and dairy were associated with lower AHEI score |

| Reference | Country | Population | Dietary record method | Diet cost method | Dietary quality measure | Summary of results |
|---|---|---|---|---|---|---|
| (Beydoun et al., 2008) | USA | 7331 adults (aged >20 <65) (3721 men and 3610 women collected between 1994 and 1996 | 2x 24h recalls 3-10 days apart. Food consumed grouped into broader categories | Food price data from American Chamber of Commerce Researchers Association (ACCRA). Prices were matched to the city closest to which the subject lived and to the survey year they were recruited. Average lowest prices were averaged against the 4 seasons | Fruit and veg price index (FVPI) derived from 7 fruit/veg items.<br>Fast food price index (FFPI) derived from 3 fast food items.<br>USDAs healthy eating index (HEI) and the Alternative Mediterranean diet score (aMED) - no price assigned to these | $1 increase in FFPI associated with higher fibre (independent of income) lower saturated fat from total energy intake and increased aMED score.<br>$1 increase in FVPI associated with improved diet quality - 10.8 higher on HEI score and also reduction in sodium and cholesterol intake |
| (Cade et al., 1999) | UK | 15191 Women (mean age 53 years) in 1995 | 217 item FFQ - in past year and a telephone interview | 1995 national food survey and 1997 Tesco home shopping catalogue applied to FFQ items per individual | Healthy diet indicator score 0-8 developed based on WHO recommendations | Healthy diets are most expensive. Largest difference between the least healthy and the second most healthy hdi group |
| (Darmon et al., 2004) | France | 837 adults aged 18-76 years (361 men and 476 women) collected in 1988-89 | Interviewer led diet history questionnaire for habitual diet over the last 6 months | Costs for 57 foods commonly consumed by adults | French National Institute of Statistics (INSEE) provided prices for each of the 57 foods for the year 2000 | Energy dilute diets cost more than energy dense ones |

| Reference | Country | Population | Dietary record method | Diet cost method | Dietary quality measure | Summary of results |
|---|---|---|---|---|---|---|
| (Drewnowski et al., 2007) | France | 1474 adults aged 15-92 years (672 men and 802 women) collected in 1999 | 7 day food record - with photographic portion sizes | French mean national 1997 retail prices obtained from marketing research for 760 foods | Vitamin C as a marker of diet quality | Higher vitamin C was associated with higher diet cost - at each quintile of energy intake.<br>As energy density increases, diet cost decreases in each quintile of energy intake for both men and women |
| (Lopez et al., 2009a) | Spain | 17197 University graduates recruited between 1999 and 2008 (mean age 38.6 years) 60% women | Validated 136 item FFQ | Ministry of Industry, Tourism and Commerce of Spain food prices assigned to each FFQ food item to calculate a cost per day (€) | Likelihood of not meeting 3 or more dietary recommendations | Increased cost was associated with increased micronutrient intake. Decreasing cost was associated with higher likelihood of not meeting 3 or more recommended nutrient intakes.<br>Association strongest in women |
| (Lopez et al., 2009b) | Spain | 17197 University graduates recruited between 1999 and 2008 (mean age 38.6 years) 60% women | Validated 136 item FFQ | Ministry of Industry, Tourism and Commerce of Spain food prices assigned to each FFQ food item to calculate a cost per day (€) | Dietary patterns - Western and Mediterranean - derived using principal component analysis | Highest quintile of western diet spent least money per 1000kcal. The opposite is true for the Mediterranean diet |

| Reference | Country | Population | Dietary record method | Diet cost method | Dietary quality measure | Summary of results |
|---|---|---|---|---|---|---|
| (Maillot et al., 2007a) | France | 1332 adults (596 men and 736 women) aged 15-92 years in 1999 | 7 day food record - with photographic portion sizes | French mean national 1997 retail prices obtained from marketing research for 619 foods | Nutrient density score for 23 qualifying nutrients and 3 disqualifying | Increasing nutrient density is associated with increasing cost. Lower cost also associated with higher disqualifying nutrients. Cost by 7 food groups and 25 food sub groups |
| (Maillot et al., 2007b) | France | 1332 adults (596 men and 736 women) aged 15-92 years in 1999 | 7 day food record - with photographic portion sizes | French mean national 1997 retail prices obtained from marketing research for 620 foods | Mean Adequacy ratio (MAR) based on 23 nutrients | Increasing tertiles of energy density was associated with decreased MAR and diet cost in both men and women. |
| (Monsivais and Drewnowski, 2009) | USA | 164 adults (university staff) aged 25 to 65 years in 2005/2006 | 152 item FFQ (384 component foods) | Supermarket chain food prices in Seattle | Cost per 2000 kcal and energy density | Energy density was inversely associated with diet cost. This association was stronger in the diet of women. |
| (Rehm et al., 2011) | USA | 4744 adults ages 20 years and over from the NHANES 2001-2002 | One 24h recall | Centre for Nutrition Policy and Promotion (CNPP) food price data based on the Neilson Homescan Consumer Panel during the same period in 2001-2002 | Healthy Eating Index(HEI) - 2005 | Highest quintile of diet cost had an associated 13 point rise in HEI score. Associations were strongest in women |

| Reference | Country | Population | Dietary record method | Diet cost method | Dietary quality measure | Summary of results |
|---|---|---|---|---|---|---|
| (Schroder et al., 2006) | Spain | 3179 adults (1547 men and 1615 women) aged 25-74 years in 1999/2000 | 165 item FFQ | Price database from the Secretaria de Estado de Turismo y Comercio de Espania which is updated weekly. Prices from last week of May 2005 used | Mediterranean Diet Score (MED) and HEI | Higher adherence to both the MED and HEI were associated with higher cost |
| (Townsend et al., 2009) | USA | 112 low income women aged 18-45 years in 2006 | 152 item FFQ in past 3 month (384 component foods) | Average local prices from supermarkets and local stores in each of the 4 study counties used by the low income women | 10 selected nutrients and 2 calculations of energy density | Lower energy density is associated with higher energy cost |
| (Waterlander et al., 2010) | Netherlands | Two cohorts<br>1 - 373 young adults. Mean age 36 years in 2000.<br>2 - 200 elderly. Mean age 69 years in 2007 | Computerised face to face interview to ascertain general diet patterns over previous 4 weeks (cohort 1) and 24h recalls (cohort 2) | National food prices based on 2 Dutch leading supermarkets between February and April 2008 | Dietary energy density and fruit and vegetable intake | Higher energy density quartiles were associated with higher total energy intake, decreased fruit and vegetable intake and lower diet costs |

Table 2.4 - Diet cost papers: summary of cost of diets consumed

All the studies summarised in Table 2.4 are of cross sectional design. While some of the studies use baseline data from cohort or observational studies, assigning a cost to a fixed diet history record means that the analysis is cross sectional. All of the studies reported, conclude that a healthy diet does in fact cost more, despite the fact that measures used to define 'healthiness' differ.

Adherence to a healthy dietary pattern is of importance with regards to promoting health and reducing risk of NCDs such as obesity and cardiovascular disease. There are many measures available to assess whether a diet is healthy or not. One common method is the Healthy Eating Index (HEI), which assesses conformance to federal dietary guidelines in the USA. This has been applied to a number of populations, which conclude that better adherence to the pattern is associated with higher monetary cost (Schroder et al., 2006, Rehm et al., 2011). Within quintiles of the Alternative HEI (AHEI) score, there is an increased cost associated with increased adherence (Bernstein et al., 2010). The same is seen with increasing adherence to the Mediterranean Diet (MED) score (Schroder et al., 2006, Lopez et al., 2009b). Conversely a Western dietary pattern (generally deemed unhealthy), typified by consumption of red meats, processed meats, eggs, sauces, pre-cooked food, fast food, calorific soft drinks whole fat dairy and potatoes, has decreasing cost with increasing adherence (Lopez et al., 2009b).

Cade et al (1999) found that the second healthiest diet, measured using a Healthy Diet Indicator was the most expensive, with a clear trend of increasing cost as the healthiness of the diet increases. There is a slight dip in price for the most expensive diet, which could be attributed to low numbers of women consuming this diet in the study population (<1%). The second healthiest group, containing 805 women (5% of subjects) consumed the most expensive diet, which cost 73% more than the least healthy diet. Where studies have assigned costs to food as consumed it has been recorded that more nutrient dense dietary habits are more expensive (Aggarwal et al., 2012, Bernstein et al., 2010, Cade et al., 1999, Darmon et al., 2004, Drewnowski et al., 2007, Maillot et al., 2007a, Maillot et al., 2007b, Monsivais and Drewnowski, 2009, Townsend et al., 2009). This is supported by findings that if individuals were to consume diets which met dietary recommendations, their diet would likely cost more (Monsivais et al., 2011), or conversely, that less expensive diets are more likely not to achieve recommended targets (Lopez et al., 2009a).

Where studies have investigated dietary cost in both men and women, a stronger association between increased diet cost and increased dietary healthiness was observed in women (Monsivais and Drewnowski, 2009, Lopez et al., 2009a, Rehm et al., 2011). This observation was observed in two USA and a Spanish cohort. Individuals with diet costs in the highest quintile were more likely to be women (Lopez et al., 2009b).

Results by Maillot et al (2007a) found that in general food groups with low nutritional quality had the lowest associated cost and conversely those with highest nutrient quality had highest associated cost. However, starch and grain were one group with a high nutritional quality yet provide low cost energy. This is similar to the finding reported by Bernstein et al (2010) who states that nuts, soy, beans and whole grains provide low cost nutrition. Maillot et al (2007a) summarise this by suggesting that preferentially selecting food groups which have the highest nutritional quality to price ratio means that a healthier diet can be achieved at a lower price.

It is interesting to propose that low cost healthy diets can be achieved, because so many studies claim that healthy diets are expensive. Perhaps it is healthy and convenient diets that are more expensive. Many people have time constraints preventing them from achieving such nutritional low cost alternatives. Perhaps it is education related - as suggested by Cade et al (1999). Bernstein et al also show that higher education levels are associated with higher AHEI scores (Bernstein et al., 2010). It doesn't cost more to reduce fat intake, as suggested by Townsend et al (Townsend et al., 2009). However, in low income populations, this may introduce other problems if individuals cannot consume enough calories to maintain a healthy weight. However, given that obesity is increasing in low socioeconomic status groups (those most likely to be susceptible to food insecurity) in the USA, this would be a feasible suggestion. Suggestions by Bernstein et al to reduce consumption of red meat, processed meats and high fat dairy, not only reduces cost through eliminating purchases, but also increases health benefits. Despite these foods not being of lower relative cost (Lopez et al., 2009b) they still must be paid for, so removing such products remove cost and the health benefits increase further cost effectiveness.

In a world where health inequalities are a problem, it is not acceptable to pay 165% more to attain 10% less energy intake (as reported by (Andrieu et al., 2006)) when buying nutrient dense foods. Nutrient dense foods promote health and prevent risk of many NCDs, yet such diets become unobtainable, or difficult to financially justify when money is scarce. This offers an explanation for why in the USA individuals in the highest quintile of the consumption of saturated fats, trans fats and added sugar have significantly lower diet costs than those in the lowest intake quintiles (Aggarwal et al., 2012).

Many studies conclude that energy dense diets are also nutrient poor (Waterlander et al., 2010, Townsend et al., 2009). These are typically low in foods such as fruit and vegetables and other products which have high nutrient content. It can be assumed therefore that these diets are less 'healthy'. This is the case in Western cultures, where food availability is good and most people consume a typically 'Western diet'. In non western countries, differences are observed where energy dense diets are not necessarily nutrient poor, due to the different types of foods which are typically consumed (Murakami et al., 2007). This is an example of why non-western populations were excluded from this review.

#### *2.4.2.2 Methods of recording diets and assigning diet costs*

Recording diet to an individual dietary record is a microeconomic process. These differing methods are discussed and also include the methods used in the geography of diet costs which could fall into physical micro or macroenvirnoment, depending on the geographic scale, but will be discussed here for continuity.

*Diet records*

Methods of dietary assessment used by the different studies vary, as does the assignment of diet cost. While the gold standard of dietary assessment is considered to be food diaries, or multiple 24 hour recalls, many of these studies identified use Food Frequency Questionnaires (FFQs) to derive diets to calculate costs for (Cade et al., 1999, Townsend et al., 2009, Monsivais and Drewnowski, 2009, Schroder et al., 2006, Darmon et al., 2004, Lopez et al., 2009a, Lopez et al., 2009b, Aggarwal et al., 2012, Bernstein et al., 2010). However, some do use food diaries (Maillot et al., 2007a, Andrieu et al., 2006, Drewnowski et al., 2007, Maillot et al., 2007b) and 24 hour recalls (Beydoun et al., 2008, Waterlander et al., 2010, Rehm et al., 2011). It may be expected that using diet diaries or 24 hour recalls would mean that less generalisation occurs, in that prices are assigned to actual products consumed, as with nutrient reporting from the same diet record methods. This is potentially the most reliable method of estimating diet cost. However, studies which do use diet diaries and 24 hour recalls often aggregate foods into more general groups. Beydoun et al (2008) group the food diary records into broad food groups, reducing the number of prices assigned. Waterlander et al (2010) report assigning a cost to every single food item consumed, but do not report the total number of foods which have a cost assigned. Maillot et al (2007a) also use 7 day food records, but aggregate up to 7 major food groups and 25 food subgroups. FFQ methods are well documented though and for the purposes of this type of analysis rank dietary costs in relation to one another well.

All of the studies assign diet costs from various food price databases which is the most intuitive method for reliably estimating cost of food consumed. However, until recently no study has attempted to validate such methods. In 2013, two studies have been published which compare using food price databases linked to diet records with food expenditure and show promising levels of agreement (Timmins et al., 2013, Monsivais et al., 2013) (the first of which I jointly authored and includes work reported in Chapter 3 of this thesis). These papers add to a study by Murakami et al (2008) which examined the comparability of different methods of dietary assessment to estimate cost of diet.

### 2.4.3 Physical and economic environments

Papers identified which address both the cost of diet and the geography of diet are summarised in Table 2.5. Few studies consider the cost of diet as consumed. Instead the cost of a standard food basket is studied. A ‘healthy’, compared to a ‘less healthy’, basket is a common unit for discussion. These studies investigate food availability rather than consumption, but imply that availability is a proxy measure of diet. Considering that collecting dietary data is expensive, time and labour intensive and prone to measurement error, it is reasonable that availability of food in retail outlets is used by so many (Cummins et al., 2010, Larsen and Gilliland, 2009). Studies using these proxy measures of diet and diet as consumed are combined in Table 2.5 in order to obtain an overview of what has been reported to date.

With the exception of Larsen and Gilliland, all studies reported in this section are cross sectional. Larsen and Gilliland used a two stage field survey to examine before and after effects of introducing a farmers market into a food desert (Larsen and Gilliland, 2009).

Evidence differs in relation to the geography of the cost of a healthy diet. While some report that a healthy diet costs more in the poorest areas (Beydoun et al., 2008), others report the opposite (Mooney, 1990). Interestingly some studies attribute difference in the geography of diet cost to the access to different types of retail outlet (Liese et al., 2007). This concurs with evidence reported which relates to the geography of diet. In the study by Cummins et al (2010) results demonstrate how evidence can differ, in that price of fruit and vegetables decrease with increasing store size, fruit is most expensive in poor areas, yet vegetables are least expensive in these areas. This highlights how variable the geography of cost of diet can be and how sensitive it is to the way in which it is measured. Differences relating to the methods used to measure diet and diet cost are discussed in section 2.4.2.2. Relevance of these results in the context of this thesis are discussed in section 2.4.9.

| Reference | Country | Population | Measure of diet | Measure of diet cost | Measure of geography | Summary of findings |
|---|---|---|---|---|---|---|
| (Beydoun et al., 2008) | USA | 7331 adults aged between 20 and 65years (3721 men and 3610 women) collected between 1994 and 1996 | 2x 24h recalls 3-10 days apart. Food consumed grouped into broader categories<br>Fruit and veg price index derived from 7 fruit/veg items.<br>Fast food price index derived from 3 fast food items.<br>USDAs healthy eating index (HEI) and the Alternative Mediterranean diet score (aMED) - no price assigned to these | Food price data from American Chamber of Commerce Researchers Association (ACCRA). Prices were matched to the city closest to which the subject lived and to the survey year they were recruited. Average lowest prices were averaged against the 4 seasons | Poverty income ratio is used as a measure of area of deprivation, dividing the population into three deprivation categories | Poorest deprivation categories were exposed to highest food prices. The middle category were exposed to the cheapest food |
| (Cole et al., 2010) | USA | No person data was collected. Data from different types of retail outlet was used | Cost and quality of 18 different fruits and 21 different vegetables.<br>Quality was recorded on a scale of 1-5, where 1 is poor - i.e. brown or wilted - and 5 is excellent and products were near to perfection | Average price per item | Two racially different (determined at census tract level) Brooklyn community districts (BCDs).<br>Three groups were determined: (1) predominantly black; (2) predominantly white; and (3) mixed race | For all but 2 items, price was highest in predominantly white neighbourhoods. No consistent quality patterns observed.<br>Less supermarkets exist in Black or Mixed race areas than in White areas |

| Reference | Country | Population | Measure of diet | Measure of diet cost | Measure of geography | Summary of findings |
|---|---|---|---|---|---|---|
| (Cummins and Macintyre, 2002b) | UK | No person data collected | 57 foods described as a 'modest but adequate diet' dietary pattern | Cheapest price, branded price and availability were collected from 325 retail outlets of different types | Carstairs-Morris deprivation category (DEPCAT) | Shop type - not location or deprivation level - best predicted price of food |
| (Cummins et al., 2010) | UK | No person data collected. 310 stores in 10 diverse areas were surveyed | Data on 15 fruit and vegetable items which comprised the fruits and vegetable sections of the Healthy Eating Indicator Shopping Basket (HEISB) | Price of 15 fruit and vegetable items | Urban/Rural environments and neighbourhood deprivation | Fruit and vegetable availability increases and price decreases with increasing store size. Fruit is more expensive in the most deprived areas, and vegetables are least expensive in these areas, although this difference is statistically non-significant |
| (Jiao et al., 2012) | USA | Data from the 2000 USA census was obtained for those residing in Seattle's Urban Growth Boundary, which comprises 90% of King County's population | Diet was not recorded. Supermarket access was used as a measure of food | Diet cost was assessed at a supermarket level, categorising supermarkets as either low cost, medium cost or high cost | Block group income levels. (A geographic unit which contains 600-1500 individuals) | Criteria used to define low income populations and to define food deserts greatly influences results. As few as 3% low income populations could walk to a low-costs supermarket, however almost all could take a ten minute drive or bus ride to a low or medium cost supermarket |

| Reference | Country | Population | Measure of diet | Measure of diet cost | Measure of geography | Summary of findings |
|---|---|---|---|---|---|---|
| (Larsen and Gilliland, 2009) | Canada | No person data collected.<br>Data was collected in a Food desert, with a farmers market introduced between 2005 and 2008 | Ontario Nutritious Food Basket (ONFB) | Cost of filling an ONFB basket, using only the lowest priced items from 11 supermarkets across the city | Urban and suburban locations in areas of contrasting incomes | Price of basket increases 9.12% over 3 years, after controlling for inflation.<br>In the Food desert areas however, prices reduced, which has been attributed to the introduction of farmers market |
| (Liese et al., 2007) | USA | No person data collected.<br>Supermarket data collected in 2004 | Limited number of staple foods, such as: fresh vegetables; low-fat/non-fat, reduced fat and whole milk; regular and low-fat options for chicken and beef; canned tuna and salmon in water; eggs; packaged roasted or smoked turkey; and low-fibre and high-fibre bread | Lowest price for available items were recorded. | Rural county (population 91582; 1106 square miles). No comparison was made to other regions | Convenience stores were typically more expensive than supermarkets, with much less availability. Healthier alternatives were typically more expensive than the less healthy versions |

| Reference | Country | Population | Measure of diet | Measure of diet cost | Measure of geography | Summary of findings |
|---|---|---|---|---|---|---|
| (Mooney, 1990) | UK | No person data collected | Shopping basket of 'healthy' foods and one of 'less healthy foods' | Price and availability of the baskets in 9 or the largest supermarkets in Hampstead, London (4 in the affluent areas and 5 in the deprived areas. Smallest and largest pack sizes priced, along with the most common size, named for this study as 'standard' | Hampstead, London, which contains areas of affluence alongside areas of deprivation | The healthy basket cost considerably more throughout the district. Baskets in poorer areas were cheaper than in more affluent areas. In the most deprived area difference in price between baskets was greatest. Buying in small packs cost about 10% more than standard size packs |
| (Sooman et al., 1993) | UK | No person data used in this analysis | Shopping basket of 'healthy' foods and one of 'unhealthy foods' | Price and availability of the baskets in 10 different stores in each of the localities (smallest available pack sizes) | A poor neighbourhood and an affluent neighbourhood | Both baskets costs more in poorer areas, however the price difference was greater in poorer areas. The availability of products was more likely in the more affluent area |

Table 2.5 - Geography and diet cost papers: summary of findings

#### 2.4.3.1 Healthy baskets

Estimating the cost of a healthy diet using a shopping basket approach is common. In this review, this is seen in studies investigating how the cost of diet varies according to environment. There are two ways in which this has been carried out. First, a shopping basket has included foods required to meet dietary recommendations (Larsen and Gilliland, 2009). The second method compares a 'healthy' basket to a 'less healthy' basket, again based around dietary recommendations (Mooney, 1990, Sooman et al., 1993). These methods draw association with what a diet would cost were someone to consume the foods included in the basket. It is a straightforward method of food cost data collection which is important to inform policy makers with regards to how much it costs to meet dietary recommendations. By investigating area differences in diet cost, it is possible to investigate whether price of meeting government dietary recommendations differ depending of the sort of area in which you reside, thus introducing potential systematic inequalities in access to affordable food. In a similar vein, other studies have used the cost and availability of fresh fruit and vegetables as a marker of cost of diet (Cummins et al., 2010, Cole et al., 2010) or staple food stuffs (Liese et al., 2007). Foods identified in a typical 'modest but adequate diet' were sourced and priced in another alternative to pricing food as consumed - but not calling itself a basket analysis (Cummins and Macintyre, 2002b).

#### 2.4.3.2 Type of supermarket

Residential proximity to a type of supermarket has been used to infer cost of diet in certain areas. This type of study is subject to a number of assumptions which are discussed in this review, but offer an alternative metric for investigating spatial diet costs. For example, grouping supermarkets as low cost, medium cost and high cost supermarkets is one way to allude to the cost of diet (Jiao et al., 2012). In the UK, examples of such supermarkets could be Aldi, Tesco and Waitrose, respectively. However, this would not take into account the price of luxury brands compared to budget brands within a supermarket.

Interestingly the review by Lee et al (2013) for the INFORMAS project suggest that new data collection should be undertaken to assess cost of healthy and less healthy diets. When considering that the INFORMAS project proposed to assign retail prices, or even use a commercial database to assign food cost to diet records, it could have been more cost effective to assign these values to dietary records which have been collected for another purpose, for example at baseline or follow up data collection for large cohort studies. Not only would this reduce overheads for new data collection, but could also bring researchers together in a collaborative manner.

### 2.4.4 Political environment

There are factors of the political environment which influence diet, diet cost and health. However, these are outside of the scope of this review. Some key examples include policies relating to the taxing of sugar sweetened beverages and unhealthy foods. They may also include studies investigating the effect of promotions such as ‘buy one get one free’ offers in supermarkets.

### 2.4.5 Sociocultural environment

Environment and socio economic status (SES) or deprivation level are often used interchangeably, as these refer to the context and composition of the environment. At times it is not always possible to differentiate between the effects of each (Macintyre et al., 2002). Physical location of where someone lives (microenvironment) is highly influenced by their SES. SES influences many types of behaviours. A collective factor, relating to the sociocultural and historical behaviours of an area (macroenvironment), can also play a role in the geography of diet and other behaviours and subsequent health.

A systematic review published in 2006, which investigates environmental determinants of fruit and vegetable consumption, found that income and being married were identified as consistently showing a positive association with fruit and vegetable intake (Kamphuis et al., 2006). This review considered environmental influences as ‘all factors external to the individual’, which is somewhat broad. They broke this down into four categories: accessibility and affordability, social conditions, cultural conditions and material conditions, which in turn encompass context and composition of the environment.

Often neighbourhoods will be described according to their level of deprivation. In the UK there are a number of indices which can be assigned to the neighbourhood in which we live, for example IMD (Data.gov.uk, 2010). Australia also have such indices (Australian Bureau of Statistics, 2008) as does New Zealand - NZDep (University of Otago, 2013). However, in the USA such an index is not widely available and less formal measures, such as income and education, are used. If poor residents are residing in a poor area then we may see cases of amplified deprivation (Macintyre, 2007). While evidence is clear that people living in deprived areas in general consume poorer diets, and suffer more from non-communicable diseases such as obesity, the mechanism for these behaviours and health outcomes is not clear. There may be a combined effect of access - both in respect to physical location and to affordability - and societal perceptions, education or socioeconomic standing.

A study in the UK investigated the cost and quality of food for sale in supermarkets across a city, which covers a range of levels of deprivation (Black et al., 2012). While this study did not consider what was actually bought or consumed and by who, it was able to conclude that those in more affluent areas had access to a greater variety of healthy foods, than those in more deprived neighbourhoods.

Shohaimi et al (2004) used the large EPIC Norfolk cohort to investigate whether area level deprivation is associated with fruit and vegetable consumption, and whether this is independent of individual SES markers - education and social class. They found the effect of area level deprivation was observed predominantly in those of manual employment with no education and that also all three of these factors independently predicted low fruit and vegetable intake. This is an example of deprivation amplification in that area deprivation effects are only displayed in those with low SES.

In France Wyndels et al (2011) found that education level correlated with adherence to national guidelines differently according to the region in which individuals lived. However, association with income tax level was consistent for all regions. This suggests in France that financial constraints affect everyone but in some regions people’s nutritional status benefits from better education, promoting the use of educational programmes in some regions.

### 2.4.6 Physical and sociocultural environment

There is body of evidence discussing context (physical) and composition (sociocultural) of a neighbourhood and its influence on health behaviours, but much less which tries to combine such approaches. Geodemographic classifications (sometimes described as market segmentation) offer some kind of hybrid measure. Living in the same place will often mean that individuals have similar access to facilities, so pockets of like types of people with similar behaviours form. A geodemographic classification formally classifies such pockets. Marketing companies exploit these classifications for economic gain, yet they are often ignored in health. There is evidence that forms of social marketing could benefit health, yet they remain underused (Stead et al., 2007, Abbas et al., 2009). While geodemographic classifications incorporate demographic characteristics - which are often determinants of behaviours - they do not necessarily relate to lifestyle behaviours (Slater and Flora, 1991). Slater and Flora (1991), attempt to quantify the extent to which lifestyle and demographic characteristics correlate. In the end, they conclude that lifestyle segmentation combined with social marketing should improve understanding of target audiences. While geodemographic systems may be far from ideal to estimate lifestyle factors, they may be much better than other methods - with a practical application of results. Pitfalls of such classifications may be that they can miss regional variations, by combining like types of people and areas, irrespective of their region of residence (Petersen et al., 2010). This is also one of the benefits of such classifications in that they allow for grouping of characteristics at a small area geography level and escape the generalising effect of grouping across a large geographic scale. Another potential problem described in the literature is ecological fallacy which is often observed in neighbourhood level research when researchers imply individual level effects which were actually observed at larger geographic scale. With a geodemographic classification, the unit of geography is the smallest available, so this is minimised in some respects, but is still evident. With larger scale geographies this must also be considered.

### 2.4.7 Economic and sociocultural environment

#### *2.4.7.1 Is diet cost independent of socioeconomic status?*

The association between diet cost and socioeconomic markers was not discussed in the first part of this review; however, it seems relevant to discuss some points here, to illustrate how closely dietary pattern, diet cost, socioeconomic status and related geographies may be linked.

How the cost of diet influences the type of diet which we consume has been discussed, with the majority of papers concluding a 'healthier' diet is more expensive than a 'less healthy' one. It is difficult to conclude in some cases whether the outcome of a healthy diet is independent of compositional factors such as markers of socio-economic status (SES). With increasing diet cost, an increase in education level - a marker of SES - is commonly observed (Cade et al., 1999, Monsivais and Drewnowski, 2009). In the study by Cade et al (1999), those consuming the healthiest diet are also those consuming the most daily calories. Given lower BMIs in these women, this suggests high levels of physical activity. Therefore, it is reasonable to suggest that diet cost is associated with healthier lifestyle factors (not just diet). Alternatively, it could be an effect of higher education or higher physical activity level being associated with higher diet cost.

Aggarwal et al (2012) show significant association between those in the lowest quintiles of diet cost and being less likely to be of higher SES, where income and education are used as markers of SES. Other results from the USA show that education is a stronger predictor of diet cost than income (Monsivais and Drewnowski, 2009) and that the income-diet cost-diet quality pathway is moderated by education level (Aggarwal et al., 2011). This is in agreement with other findings. For the NHANES cohort, Rehm et al (2011) report that lower income-to-poverty ratio and educational attainment, are associated with lower HEI-2005 scores.

Waterlander et al (2010) found that there is no observed difference in energy density or energy cost between income levels, suggesting in fact that diet cost is not an effect of SES, with no difference in diet cost, energy density, energy consumed or fruit and vegetable intake by income group. However, Townsend et al (2009) found energy dense diets are the least expensive in a sample of low income women. It might have been expected that in a sample where income is low, that everyone is forced to consume an energy dense nutrient poor diet, but this is not the case. So while markers of SES in some cases appear to influence diet cost, it is not that straightforward with other studies finding no effect.

### 2.4.8 Study population - how generalisable are results?

Results presented in Table 2.4 were reported about populations from different counties (macroenvironments); the UK, USA, France, Spain and the Netherlands. They all conclude the same - a healthier diet does cost more. Given the overall consistency in findings, it could be concluded that results for cost of a healthy diet are transferrable to other western populations. Nutrient dense, healthier diets have been shown to be more expensive than less healthy alternatives in the UK, France, the USA, the Netherlands and Spain. Waterlander et al (2010) estimated daily expenditures to be around €5.00 in the Netherlands, which is similar to Darmon et al (€5.00) and Maillot et al (€5.26) in France. In the UK, Cade et al found mean diet costs in women to be £3.43, which (using an exchange rate from 30th June 1999) equates to €5.20. The data for these studies were all collected in 1999/2000 so despite them being different European countries, the diet costs are very similar. The results also show patterns of increasing cost in line with measures of increasing 'healthiness'.

Some studies recruited a nationally representative sample (Andrieu et al., 2006, Maillot et al., 2007a), although this is not necessarily a requirement for studies investigating diet and health. In national surveys, it should be expected. For example the National Diet and Nutrition Survey (NDNS) is geographically representative of the UK, using postcode sampling units to recruit participants (Department of Health and Food Standards Agency, 2012). However, the sample is still of a limited size, so the worth of such sampling should be considered, especially when dietary data is subject to so many pitfalls.

Some studies focused on recruiting a low income population, although when considering cost of diet research, a positive gradient in diet cost in association with nutrient density is still observed (Townsend et al., 2009). A wealth of research has been completed on Scottish populations (Sooman et al., 1993, Cummins and Macintyre, 2002b, Cummins et al., 2010) and also in French populations (Andrieu et al., 2006, Darmon et al., 2004, Drewnowski et al., 2007, Maillot et al., 2007a, Maillot et al., 2007b, Wyndels et al., 2011) and findings are consistent with other Western populations. Results from rural communities, those in cities and those with different ethnic populations, all show a gradient in cost of diet with markers of deprivation (Liese et al., 2007, Cole et al., 2010, Mooney, 1990).

### 2.4.9 Application for this thesis

This review highlights existing evidence for spatial variations in the cost of diet. However, cost of diet calculated from cost assigned to diet records using a validated food cost database has not been carried out, making the results somewhat subjective. The majority of studies have used modest sample sizes, which can be lacking in power to detect differences in diet cost with real confidence. Having rich geographic identifiers, which allow diet at a small area geography to be investigated, combined with cost of diet as consumed have not been reported, suggesting that these are not available or that no one has yet carried out such research. This is where this research will add to the gap in evidence.

This research is well timed as many recent examples highlight the importance of such new evidence. In July 2013 the journal Public Health Nutrition dedicated an issue to reporting on this area. The editorial by Ball and Thornton (2013) summarises the importance of such work, referring to a meeting of the USA National Cancer Institute (NCI) in partnership with others, in 2007, which identified knowledge gaps and future research priorities for measuring food and physical activity environments. In the October 2013 issue of Obesity Reviews, a supplement issue was included, publishing work from the INFORMAS project. These papers discuss the effect of price on diet and also the effect of environment, suggesting frameworks for future monitoring and investigating such relationships.

Methods of collecting price data and applying to food intake is of importance (Willett, 2013, Lee et al., 2013), yet could be impractical if it places too high a financial burden on governments or researchers. Using a cost database which can be applied to dietary intakes reported by diet records in large cohort studies could be a straightforward application of existing cohort dietary record data. Often such cohorts also report their addresses at the time of data collection. These can be used to carry out important spatial analyses.

Using a geodemographic classification is one method of incorporating demographic characteristics to small area geography without the risk of identifying individuals. This would optimise on readily available classifications by adding them to rich cohort data and gaining from the long established tools used by market research companies.

Concerns over price and availability of healthy foods, according to where you live, are not new. In 1993 Sooman et al carried out some research, following on from a Chief Medical Officer white paper in Scotland, to investigate price and availability of recommended healthy foods in two different neighbourhoods in Glasgow, Scotland (Sooman et al., 1993). Yet evidence is still emerging and recent reports from the UK Chief Medical Office in 2010, reiterate the importance of geography when considering determinants of health (Davies, 2011).

## 2.5 Summary and context

Chapter 1 described the aims and objectives and provided an overview of the thesis. This Chapter has reviewed and discussed the literature relating to cost of diet, geography of diet and geography of diet cost. Most studies have found that a healthier diet does in fact cost more. This is particularly evident in areas of low income populations, whose economic and physical access to healthy foods is limited. Studies have rarely combined valid methods of dietary assessment with application of costs from a reliable cost database, with small scale geographic units. The work reported in this thesis will add evidence to help fill this gap.

The data and methods used to achieve the aims and objectives of this thesis are discussed in Chapter 3. Chapter 3 also includes an evaluation of a diet cost database which has been published in the European Journal of Clinical Nutrition (Timmins et al., 2013).

# Chapter 3: Data and methods

## 3.1 Overview

Chapter 1 described the aims and objectives and provided an overview of the thesis. Chapter 2 reviewed the existing literature in this area. Now, Chapter 3 explains the data and methods which have been used to produce the results described in Chapters 4-8 in order to meet the aims and objectives described in Chapter 1. Figure 3.1 serves as a visual reminder of where this appears in the thesis flow.

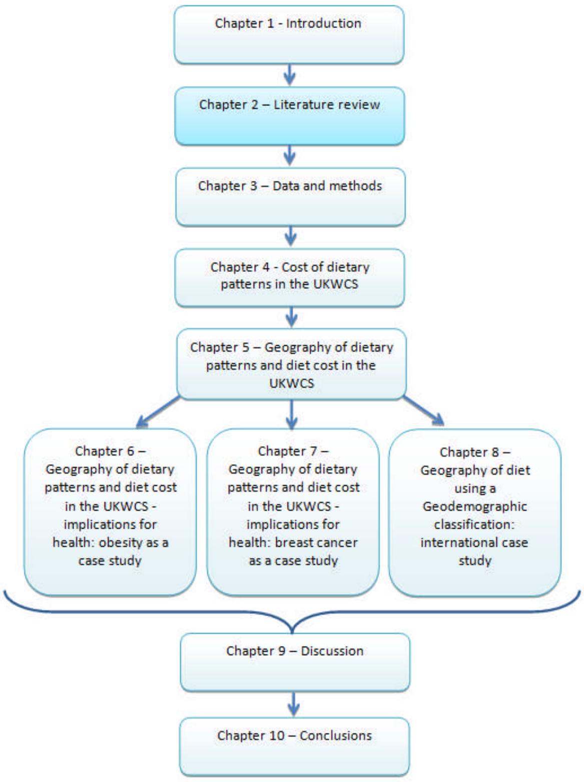



Figure 3.1 - Overview showing how Chapter 3 fits into the overall thesis flow

This thesis uses existing datasets, such as the UK Women's Cohort Study, and classifications, such as the Output Area Classification in order to complete the analysis. This chapter is in two distinct sections. In the first part, the datasets and classifications are described in the Data section (3.2), followed by details of statistical analyses, data linking and spatial analyses carried out in the Methods section (3.3). Methods of specific chapters are put into context in the related chapter. The second part details an evaluation of a food cost database (3.4). This section includes results and discussion of the food cost database evaluation, which has been published in the European Journal of Clinical Nutrition in October 2013 (Timmins et al., 2013).

## 3.2 Data

A number of different data sources were used for the analysis in this thesis. A summary of each of these is presented below.

### 3.2.1 The UK Women's Cohort Study

The UK Women's Cohort Study (UKWCS) was set up in the 1990s to investigate links between diet and health. Women were recruited from a World Cancer Research Fund (WCRF) mailing list. There was some stratified sampling in order to ensure over representation of those consuming a vegetarian diet, to provide sufficient power to detect differences in health due to diet.

At baseline, between 1995 and 1998, 35372 women were recruited. At this time an extensive lifestyle questionnaire was completed, along with a 217 item FFQ (Cade et al., 2004). Full postal address, including postcode, was collected for each woman. Approximately five years later, the women were followed up, where approximately 14000 responded, completing a further lifestyle questionnaire, a four day food diary and a 24 hour physical activity diary.

All the women agreed to have updates on cancer incidence and death provided to the research group automatically from the National Health Service Information Centre (NHSIC) (as it is currently known).

The UKWCS data is used in Chapters 4, 5, 6 and 7 of this thesis.

#### *3.2.1.1 Data cleaning*

The UKWCS data has been widely used for research and has undergone extensive data cleaning in the past. Specific for these analyses, data cleaning of postcodes was carried out. Missing postcodes were found by searching on the rest of the postal address using www.google.co.uk. Common mis-typing errors, such as mixing 0 with O and 1 with I were also corrected.

#### *3.2.1.2 Body Mass Index*

Body Mass Index (BMI) was calculated using self reported height (m) and weight (kg) using the formula:

BMI ($kg/m^2$) = weight (kg)/height $(m)^2$

This produces a continuous measure of BMI. In addition to this a categorical measure of BMI is useful. The World Health Organisation (WHO) publishes cut points for underweight, normal weight, overweight and obese to ensure consistent recording of these weight status (World Health Organisation, 2006). Each of the women was assigned a BMI category. The WHO cut points are summarised in Table 3.1.

| Weight status | BMI range |
|---|---|
| Underweight | <18.5 |
| Normal weight | 18.5-24.9 |
| Overweight | 25-29.9 |
| Obese | >=30 |

Table 3.1 - WHO BMI categories

### *3.2.1.3 Updating Breast Cancer Cases in the UKWCS*

As Chapter 7 of this thesis is a case study of how spatial diet cost patterns impact on incidence of breast cancer, it was important to ensure that the UKWCS cancer data was up to date. As described in section 3.2.1 the NHSIC provides updates cancer incidence data on a quarterly basis. These data are applied to UKWCS Microsoft Access database, which is carried out by the Nutritional Epidemiology Group (NEG) Database Manager. These data were then updated into Stata (StataCorp, 2012).

In order to ensure that data sent to us from NHSIC was added to the Nutritional Epidemiology Group (NEG) databases consistently it was important to create a step by step procedure, which was developed with a colleague. This procedure also includes comprehensive update testing guidelines (see Appendix A).

### *3.2.1.4 Dietary patterns*

Data driven dietary patterns are used in Chapters 4-7 of this thesis. These dietary patterns were previously identified using k means cluster analysis (Greenwood et al., 2000). Seven dietary patterns were identified from FFQ data collected at baseline for the UKWCS. These dietary patterns are summarised in Table 3.2. Greenwood et al have shown stability in these patterns are over time (Greenwood et al., 2003). While it would have been possible to have used other types of dietary patterns in this research, it seemed most suitable to use patterns derived for this cohort as opposed to making the diets of these women fit pre-defined dietary patterns.

| Dietary pattern | High quantities | Moderate quantities | Low quantities |
|---|---|---|---|
| Monotonous Low Quantity Omnivore | White bread, milk, sugar | Potatoes, meat | Most other foods |
| Health Conscious | Bran, potatoes, wholemeal food, yoghurt, low-fat dairy products, pulses, fish, vegetables, salad, fruit | Most other foods | Chips, sugar |
| Traditional Meat, Chips and Pudding Eater | White bread, chips, meat, sugar, high-fat and creamy food, biscuits, cakes | Most other foods | Wholemeal food, soya products, vegetables, salad, fruit |
| Higher Diversity Traditional Omnivore | Chips, white pasta and rice, high-fat and creamy food, eggs, meat, fish, chocolate, biscuits, crisps. More fish and salad and general diversity than the Traditional Meat Chips and Pudding Eater. | Vegetables, fruit and alcohol. | Less cakes and puddings than the Traditional Meat Chips and Pudding Eater. |
| Conservative Omnivore | - | Most food, including potatoes, meat, fish, eggs, fruit, vegetables | Cereals, chips, wholemeal food, nuts, pulses, spreads and dressings, chocolate, crisps, biscuits. Less red meat, less chips and less puddings than the Traditional Meat Chips and Pudding Eater and the Higher Diversity Traditional Omnivore. |
| Low Diversity Vegetarian | Wholemeal bread, soya products, pulses, fruits (not exotic fruit), vegetables. | Cereals | Butter, eggs, meat, fish |
| High Diversity Vegetarian | Wholemeal bread, cereals, wholemeal pasta and rice, soya products, spreads, nuts, pulses, vegetables, fruit, herbal tea (generally higher consumption of these products that the Low Diversity Vegetarian). | - | White bread, meat, fish |

Table 3.2 - Characteristics of food consumption in the seven dietary patterns

### 3.2.2 Australian Longitudinal Study for Women's Health

The Australian Longitudinal Study for Women's Health (ALSWH) data were used in Chapter 8, and is one of two international case studies investigating how effective the use of geodemographic classifications to profile diet and health are in other developed populations. The ALSWH cohort was conceived and developed by groups of interdisciplinary researchers at the Universities of Newcastle and Queensland in Australia and is funded by the Australian Department of Health and Ageing. It was set up in 1995 to investigate various areas of women's health and wellbeing (Lee et al., 2005). Five key areas are explored: health service use; health related behaviours such as diet and exercise; time use (be-it for employment, voluntary work or leisure); life stages and key events including childhood, divorce and widowhood; and violence against women. Dietary information was collected in the form of an 80 item food frequency questionnaire, in addition to specific questionnaire points relating to diet. Three age groups of women were recruited (aged 18-23, 45-50 and 70-75), forming three cohorts. For this research the cohort of middle aged women, ages 45-50 at recruitment was used.

This research uses data from the third survey of middle aged women which took place in 2001.

### 3.2.3 Seattle Obesity Study

The Seattle Obesity Study (SOS) has also been used in Chapter 8, the international case study chapter. Participants for the initial wave of the Seattle Obesity Study (SOS1) were recruited between October 2008 and March 2009 from King County, Washington to participate in a study designed to investigate social disparities, diet and health (Aggarwal et al., 2011). Adult residents from a stratified random sample of 2001 completed a twenty minute telephone questionnaire. Geocoded addresses are available for 1992 of these participants. Of these 1297 completed a FFQ to assess dietary intake, which comprise the sample for this analysis.

### 3.2.4 Spatial measures

In order to carry out spatial analysis the postcode of each UKWCS participant was used to aggregate to various spatial scales:

- Government Office Region (GOR)
- North and South
- Urban and Rural
- The Office of National Statistics (ONS) Output Area Classification (OAC)

The Geoconvert application was used to complete this process (Census Dissemination Unit 2009). Geoconvert is an online tool provided by the Census Dissemination Unit at the University of Manchester, which will match one type of geography to another and provide a report of the target geographies. These target geographies can then be merged back into the source dataset.

These four spatial measures were chosen in order to investigate elements of both geographical context and composition. Two units incorporate solely context of where an individual resides: GOR and North-South. A GOR is a large geographical unit which is home to between two and nine million individuals. This unit is used by government and is known to most people. To report dietary consumption at this level means that substantial generalisation is made, but may be a useful unit in which to convey results to the public and policy makers. In the UK, research has shown that differences in health and mortality occur at an even larger geographic scale, the divide that is the North and the South of England. Economic differences also occur at this scale, with a higher cost of living in the South. This measure has been used in this research to investigate if any differences in health can be attributed to differences in dietary patterns or diet cost across the North South divide.

The Urban-Rural and OAC units can be pinpointed to a specific geographical context, but also include elements of geographic composition. The OAC does this to the greatest extent, incorporating 41 demographic variables from the census. Geodemographic classifications are developed such that they highlight groups of individuals with like characteristics and pinpoint them to specific small area geographies. This has been successfully used in marketing and also in other areas of research. This thesis will explore the extent to which these classifications are useful in diet and health research, while comparing to other common geographical scales.

It is recognised that there are many other ways in which geographical influences on diet can be investigated, including, but not limited to, investigating the area immediately surrounding an individual's home using buffer areas of a specified distance, or by using GPS tracking devices to record the activity space of an individual, these were not felt appropriate for this research as results are not necessarily generalisable to the wider population.

### *3.2.4.1 Government office regions*

England is divided into nine GORs. These are commonly used regions for spatial analysis in the UK. For the purpose of this thesis Wales and Scotland are also treated as if they are individual GORs (Figure 3.2). Northern Ireland has been excluded from analysis due to low numbers in the UKWCS. When considering the 9 English GORs plus Wales and Scotland in this manner, this effectively relates to the Nomenclature of Territorial Units for Statistics, first level codes (NUTS 1). However, GOR will be used in this thesis as this is a more widely recognised terminology in the UK.


Figure 3.2 - Regions in the UK

### *3.2.4.2 North-South divide in England*

The nine GORs of England can be dichotomised into North and South using the Severn-Wash line (Hacking et al., 2011) with The North consisting of GORs: North East, North West, Yorkshire and the Humber, West Midlands, East Midlands. The South comprising GORs: East of England, Greater London, South East and South West. This excludes Scotland and Wales so women from these regions were not included in any North South analysis.

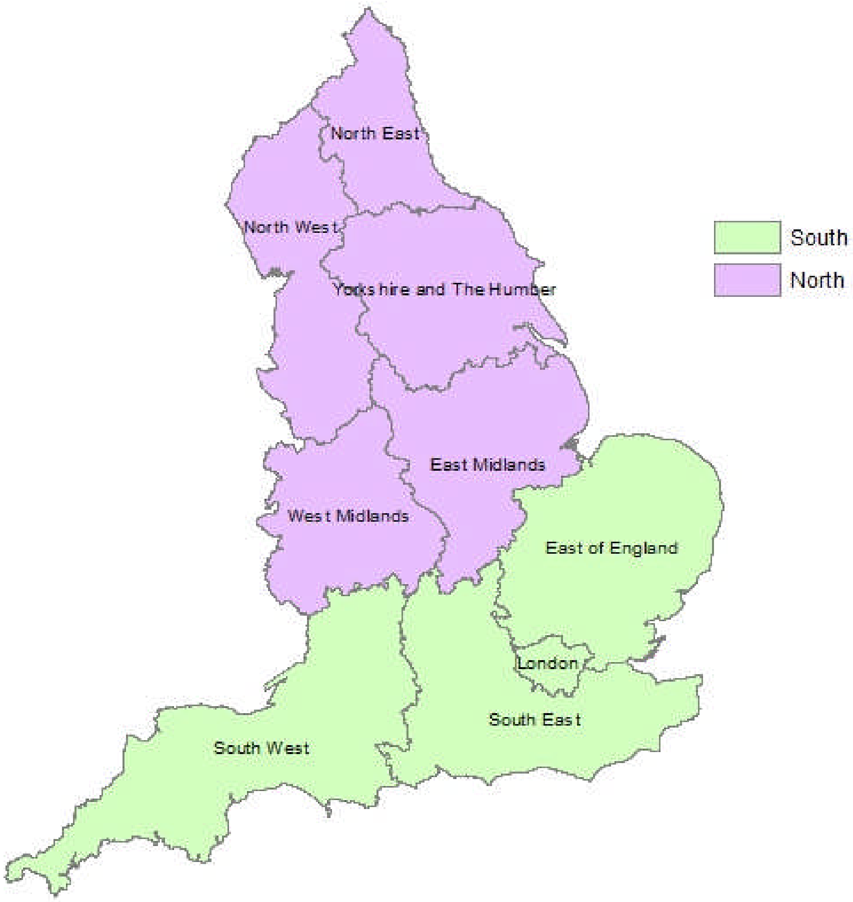



Figure 3.3 - North South divide in England

### *3.2.4.3 Urban-rural*

Using data on UK women and their Urban/Rural index from the 2001 census data it was possible to compare the distribution of UKWCS women with the population of women in the UK (England, Wales & Scotland). Urban/Rural indicators are determined at the smallest available geographical level, the Output Area. Eight specific types of Urban/Rural areas are described for England and Wales, two of which fall into an Urban category and six Rural (Office for National Statistics, 2001, Office for National Statistics, 2009). Scotland also has eight categories, two of which relate to Urban areas and six Rural (General Register Office for Scotland, 2001). The eight types of Urban or Rural area and their assignment to Urban or Rural are summarised in Table 3.3.

| | **England and Wales** | **Scotland** |
|---|---|---|
| Urban | Urban >10k - Sparse | Large Urban |
| | Urban >10k - Less Sparse | Other Urban |
| Rural | Town and Fringe - Sparse | Accessible Small Towns |
| | Village - Sparse | Remote Small Towns |
| | Hamlet & Isolated Dwelling - Sparse | Very Remote Small Towns |
| | Town and Fringe - Less Sparse | Accessible Rural |
| | Village - Less Sparse | Remote Rural |
| | Hamlet & Isolated Dwelling - Less Sparse | Very Remote Rural |

Table 3.3 - Urban Rural categories for England and Wales and Scotland

#### *3.2.4.4 Output Area Classification*

The Output Area Classification (OAC) is a geodemographic classification which has been created (Vickers et al., 2005) using the geographical unit OA and combinations of 41 variables from the 2001 Census. The OA is the smallest unit for which demographic information is readily available. It consists of a minimum of 40 households, and contains on average 250 people. Analysis was carried out to find clusters of like demographic characteristics, from the 41 census variables selected for inclusion in the analysis, at OA geographic level. A three tier classification was created. At the top level, 7 Supergroups were identified with like demographic characteristics at their corresponding OA unit. This effectively groups the whole UK population into 7 groups of similar characteristics who can be targeted at a small area geography level. More specific clusters are identified at the second level, splitting the population into 21 Groups. These are further broken down into 52 Subgroup units. Such a classification is useful for targeting groups in need of interventions, or for marketing purposes for companies to target specific products.

A summary of the characteristics of each OAC Supergroup (top level of the hierarchy) are presented in Table 3.4. The 21 Groups are summarised in Table 3.5. No sub group analysis is included in this thesis, due to small numbers in the Subgroups in the UKWCS. More information on all aspects of the OAC classification can be found on the OAC web page http://www.sasi.group.shef.ac.uk/area_classification/index.html (SASI Group).

Results are presented in this thesis at both Supergroup and Group level. The Supergroup results are included in regression models to investigate whether they can predict dietary patterns, diet cost, obesity or breast cancer outcomes. Results at the Group level are used to estimate small area dietary patterns.

| Supergroup | Distinctive variables - High | Distinctive variables - Low |
|---|---|---|
| 1 - Blue Collar Communities | Age 5-14<br>Lone parent households<br>Households with non-dependent children<br>Terraced housing<br>Routine/Semi-routine employment<br>Mining/Quarrying/Construction employment<br>Manufacturing employment<br>Retail trade employment | Indian, Pakistani and Bangladeshi<br>Black<br>Born outside the UK<br>Rent (Private)<br>Flats<br>Higher education qualifications<br>Financial intermediation employment |
| 2 - City Living | Age 25-44<br>Born outside the UK<br>Population density<br>Single person household<br>Rent (private)<br>Flats<br>No central heating<br>Higher education qualification<br>Students<br>Financial intermediation employment | Ages 0-4,5-14,25-44 and 65+<br>Single parent household<br>Households with non-dependent children<br>Rooms per household<br>Provide unpaid care<br>Economically inactive/looking after family<br>General employment |
| 3 - Countryside | Ages 45-64 and 65+<br>Detached housing<br>Rooms per household<br>2+car households<br>Work from home<br>Provide unpaid care<br>Agricultural employment | Indian, Pakistani and Bangladeshi<br>Black<br>Population density<br>Single person household<br>Flats<br>People per room<br>Public transport to work<br>Unemployment |
| 4 - Prospering Suburbs | Age 45-64<br>Two adults no children<br>Households with non-dependent children<br>Detached housing<br>Rooms per household<br>2+car households<br>Provide unpaid care | Indian, Pakistani and Bangladeshi<br>Black<br>Divorced/separated<br>Single person household<br>Single pensioner households<br>Renting public and private<br>Terraced housing<br>Flats<br>No central heating<br>Limiting long-term Illness<br>Unemployment |
| 5 - Constrained by Circumstance | Age 65+<br>Divorced/separated<br>Single pensioner households<br>Lone parent households<br>Rent (Public)<br>Flats<br>People per room<br>Routine/Semi routine employment<br>Limiting long-term Illness<br>Unemployment | Two adults no children<br>Rent (Private)<br>Detached housing<br>Rooms per household<br>Higher education qualification<br>2+ car households<br>Work from home |

| | | |
|---|---|---|
| 6 - Typical Traits | Work part time<br>Terraced housing | Age 65+<br>Rent (Public) |
| 7 - Multicultural | Ages 0-4 and 5-15<br>Indian, Pakistani and Bangladeshi<br>Black<br>Born outside the UK<br>Population density<br>No central heating<br>People per room<br>Public transport to work<br>Students<br>Unemployment | Ages 45-64 and 65+<br>Single pensioner households<br>Two adults no children<br>Economically inactive/looking after family or home |

Table 3.4 - Summary of OAC Supergroup characteristics, adapted from (Vickers et al., 2005)

| Group | Distinctive variables - High | Distinctive variables - Low |
|---|---|---|
| 1a - Terraced Blue Collar | Terraced Housing<br>Rent (public) | No central heating<br>Rent (private)<br>Detached housing<br>Higher education qualification<br>Live in a flat<br>Born outside UK |
| 1b - Younger Blue Collar | One parent household<br>No central heating<br>Terraced housing<br>Rent (public) | Detached housing<br>Higher education qualification<br>Lives in a flat |
| 1c - Older Blue Collar | Rent (public) | Living in a flat |
| 2a - Transient Communities | Financial intermediation employment<br>No central heating<br>Higher education qualification<br>Public transport to work<br>Single person household (not pensioner)<br>Born outside the UK<br>Rent (Private)<br>All Flats | Detached housing<br>Households with non-dependent children<br>Terraced housing<br>Age 5-14<br>2+ Car household<br>Working part-time<br>Economically inactive looking after family<br>Rooms per household<br>Mining/Quarrying/Construction employment<br>Age 0-4<br>Lone Parent household |
| 2b - Settled in City | Higher education qualification<br>Born outside the UK<br>Rent (Private)<br>All Flats | Detached housing<br>Households with non-dependent children |
| 3a - Village Life | Agriculture/fishing employment<br>Detached housing | Public transport to work<br>Population density<br>All Flats |
| 3b - Agricultural | 2+ Car household<br>Work from home<br>Detached Housing<br>Agriculture/fishing employment | Population Density<br>Terraced Housing<br>All Flats<br>Public Transport to work<br>Rent (Public) |
| 3c - Accessible Countryside | 2+ Car household<br>Agriculture/fishing employment<br>Detached housing | Rent (Public)<br>Population density<br>Public transport to work |
| 4a - Prospering Younger Families | 2+ Car household<br>Detached housing | All Flats<br>Rent (Public)<br>No central heating<br>Single pensioner household<br>Age 65+<br>Terraced housing |
| 4b - Prospering Older Families | 2+ Car household<br>Detached housing | Terraced housing<br>All flats<br>Rent (Public) |

| | | |
|---|---|---|
| | | No central heating<br>Rent (Private)<br>Single person household (not pensioner)<br>Lone parent household |
| 4c - Prospering Semis | - | All Flats<br>Rent (Public)<br>Terraced Housing<br>Rent (Private) |
| 4d - Thriving Suburbs | 2+ Car household<br>Detached Housing | Terraced housing<br>Rent (Public)<br>No central heating |
| 5a - Senior Communities | Age 65+<br>Single pensioner household<br>Rent (Public)<br>All Flats | 2+ Car household<br>Detached housing<br>Age 5-14<br>Age 0-4<br>Rooms per household<br>Economically inactive looking after family |
| 5b - Older Workers | All Flats<br>Rent (Public) | Detached housing |
| 5c - Public Housing | Public transport to work<br>Unemployed<br>Lone parent household<br>All Flats<br>Rent (Public) | Detached housing<br>2+ Car household<br>Higher education qualification<br>Rent (Private) |
| 6a - Settled Households | Terraced Housing | All Flats<br>Rent (Public) |
| 6b - Least Divergent | - | Age 5-14<br>Population density<br>Limiting long term illness (Age standardised illness ratio)<br>Agriculture/fishing employment<br>Households with non-dependent children<br>Age 25-44 |
| 6c - Young Families in Terraced Homes | Rent (Private)<br>No central heating<br>Terraced housing | Detached housing<br>Rent (Public) |
| 6d - Aspiring Households | - | Rent (Public) |
| 7a - Asian Communities | No central heating<br>Use public transport to travel to work<br>Privately renting usually terraced housing<br>Born outside the UK<br>Black African, Black Caribbean or other Black<br>Indian, Pakistani or Bangladeshi | Detached housing |

| | | |
|---|---|---|
| 7b - Afro-Caribbean | Privately or publically renting flats<br>Unemployed<br>Indian, Pakistani or Bangladeshi<br>Public transport to work<br>Born outside the UK<br>Black African, Black Caribbean or other Black | Detached housing or have more than 2 cars per household |

Table 3.5 - Summary of OAC Group characteristics, adapted from (Vickers et al., 2005)

#### *3.2.4.5 Intersection of spatial measures*

In order to place in context how the spatial measures intersect one another, for the UKWCS women, it is useful to use a series of maps which visually express this. Figure 3.4 shows the distribution of UKWCS women by OAC Supergroup within each of the GORs and Figure 3.5 shows the distribution of UKWCS women residing in urban and rural areas within each of the GORs. One example which stands out is that Greater London has the greatest proportion of women classified as Multicultural. Figure 3.4 also illustrates how geodemographic classifications do not account for the differences in typical characteristics between regions, (Petersen et al., 2010), one reason why it is relevant to use a range of geographies in this research. Figure 3.5 illustrates how Urban and Rural areas are distributed in each of the GORs. Unsurprisingly there are no UKWCS living in Rural areas in Greater London.

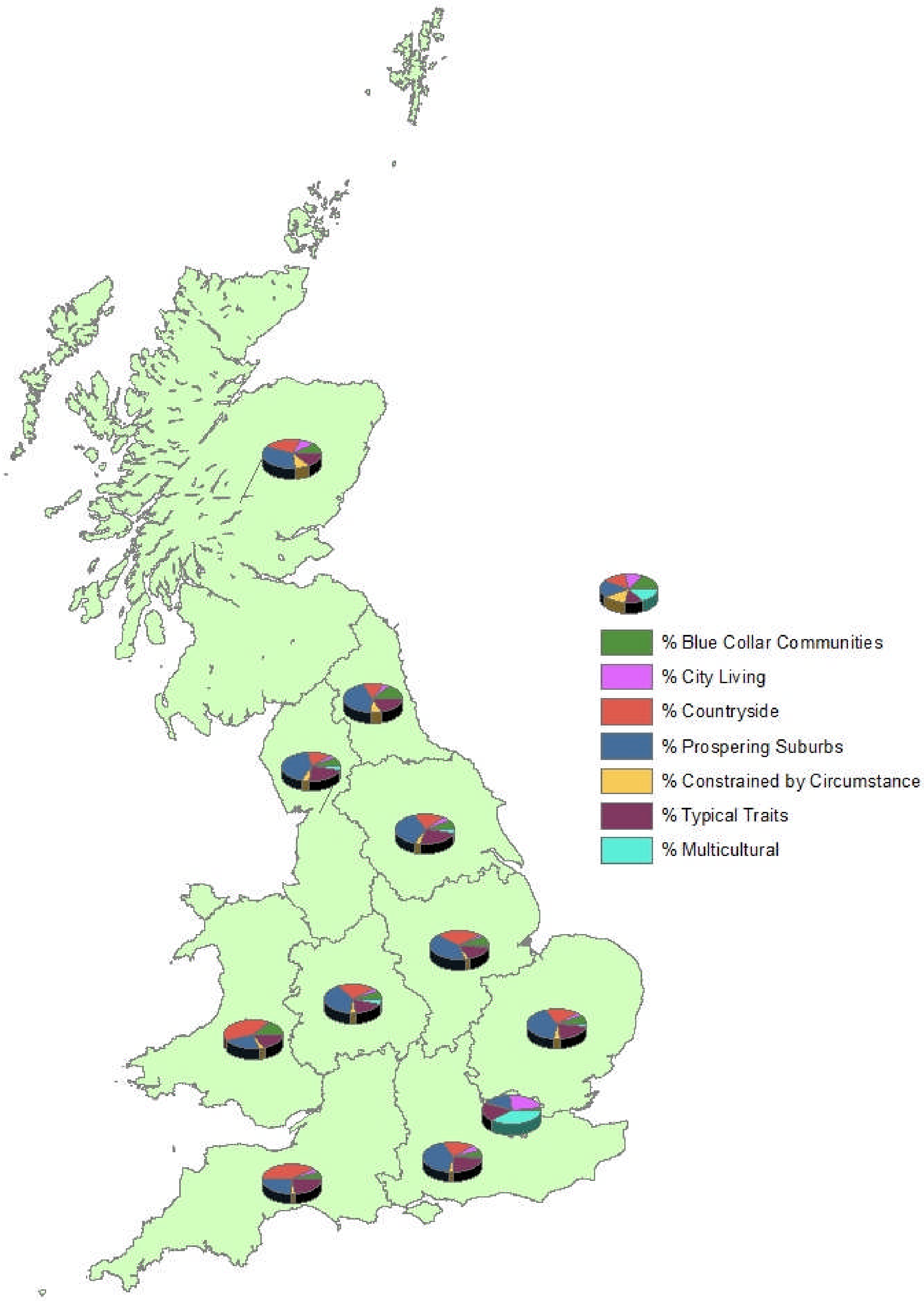



Figure 3.4 - Regions in the UK with the percentage of UKWCS women in that region comprising each OAC Supergroup

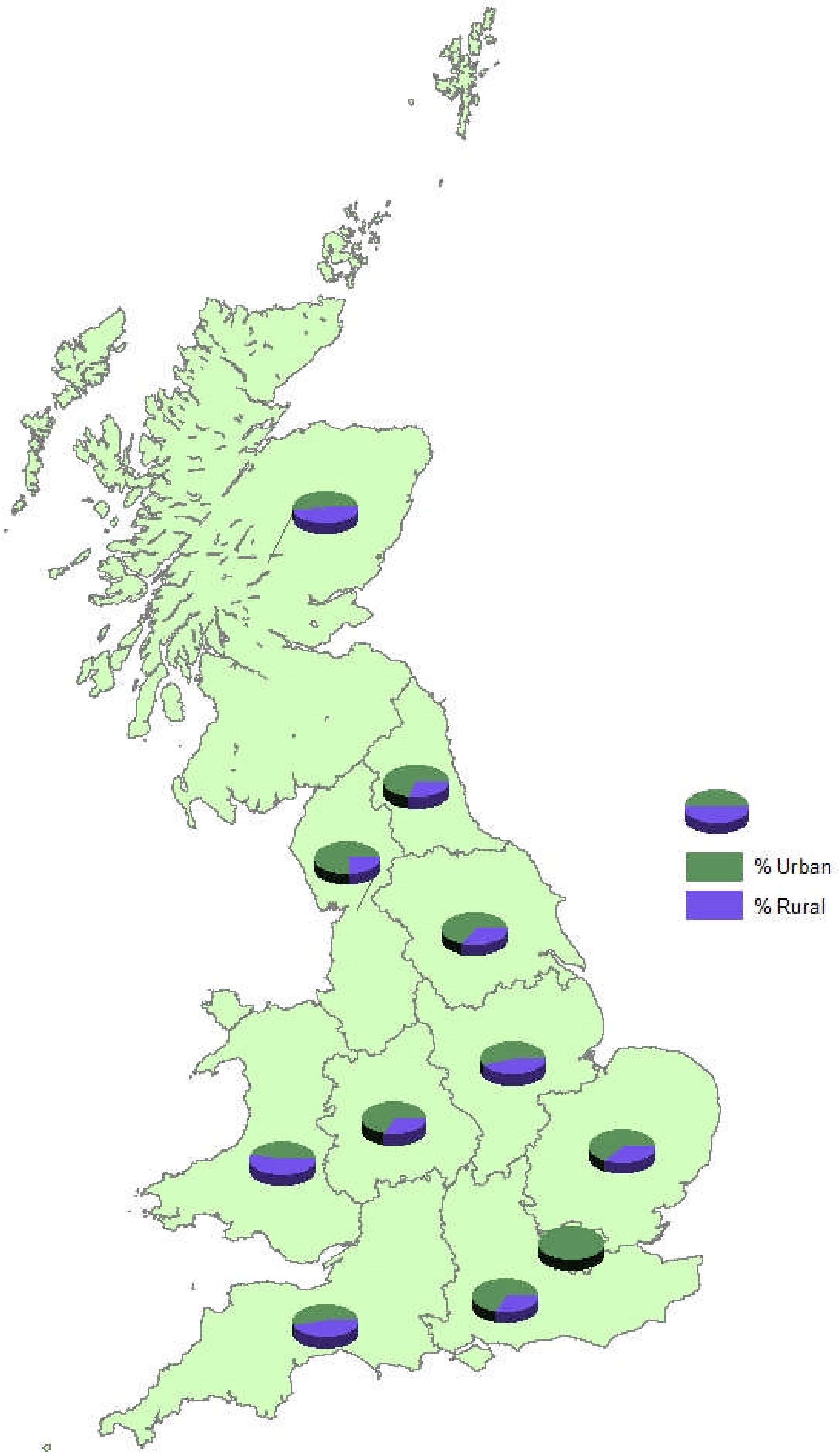



Figure 3.5 - Regions in the UK with percentage of UKWCS women in that region comprising Urban or Rural

### *3.2.4.6 Cameo - Australia and USA*

The OAC is a freely available UK specific geodemographic classification. As mentioned earlier, methods are published online which make it a preferable classification to use for research due to its transparency. For the international case studies in Chapter 8, such open source classifications were not available. In order to complete the international case studies it was necessary to use the Cameo Geodemographic Classification, which is available for 40 countries worldwide (Callcredit Information Group). Callcreditgroup in Leeds made this classification available for research purposes free of charge, for which a contract was drawn up (Appendix D). Cameo is similar to the OAC classification in that it utilises census demographic information and data from other sources and identifies clusters of similar characteristics, tied to a small area geography unit. Such classifications are used for the same purposes as the OAC.

*Australia*

This classification has been developed using data from the Australian census, the Australian household and expenditure survey (both run by the Australian Bureau of Statistics), along with transactional and proprietary data. The classification is hierarchal in structure with ten groups at the highest level. These are further subdivided into 51 sub groups, providing increased granularity. Only the ten highest level groups are used in this research. As the Cameo classification is a commercial product, the exact method used to develop it is not available (as is often the case using commercial geodemographic products as this data is their intellectual property). However a comprehensive overview of 35 different variables, by each Cameo group and sub-group are published online (Callcredit Information Group, 2013). Summary of the characteristics of Cameo Australia are included in Table 3.6.

| Cameo | Name | Summary |
|---|---|---|
| 1 | Affluent Urban Professionals | Accounts for 8% of households in Australia with household size between 1 and 5. Fewer children and over 65s than average in Australia. Very low indigenous population but otherwise mixed ethnicities. More likely than average to be urban with a wide variety of housing types which are likely to be owned outright/mortgaged. Mortgage and rent payments in these areas are very high but so is household income and level of qualifications. Employment is likely managerial, professional or white collar with unemployment rates lower than average. These areas have slightly higher than average voluntary workers and part time workers. The proportion of students is higher than average. |
| 2 | Wealthy Family Neighbourhoods | Accounts for 11% of households in Australia consisting of slightly higher than average numbers of children and much lower levels of over 65s. Household size varies between 1 and 4. Individuals are of mixed ethic origin with a higher than average proportion of indigenous populations. Homes are less likely than average to be in an urban area and likely to be detached homes owned outright/mortgaged. Mortgage and rents in these areas are above average and household income is high with above average level of qualifications. A variety of employment types reside here, with a lower than average number of students. Participation in voluntary and part time work is slightly above average. Unemployment is below average. |
| 3 | High Income Urbanites and Students | 6% of Australians reside in this type of area with household size between 1 and 6+. Child and elderly presence is lower than average. Below average indigenous population. Mixed ethic origins. Homes are more likely than average to be in an urban area with a variety of housing types with mixed tenure. Household income and qualification level are high with high mortgage repayment and above average rents. As the name suggest there is a very high proportion of students in these areas. Voluntary and part time work is around the average. A variety of employment types exists with average unemployment. |
| 4 | Comfortable Mixed Suburban Areas | Accounts for 17% of households in Australia with household size between 1 and 4. Average numbers of children and elderly. Slightly below average indigenous population. Primarily Australian or European origins. Homes have average likelihood of being in an Urban area with predominantly detached housing which is owned outright/mortgaged. Household income and mortgage repayments are above average with average rental payments and qualification levels. Voluntary and part time work is around the average. A variety of employment types exists with average unemployment. The student population is close to average. |
| 5 | Mixed Areas of Modest Detached Homes | Accounts for 14% of households in Australia with household size between 1 and 4. Presence of children is slightly higher than average with elderly below average. Slightly below average indigenous population. Australian and mixed ethnic origins. Homes have average likelihood of being in an urban area with predominantly detached housing which is owned outright/mortgaged. These areas have average mortgage and rental payments with average household income and qualification levels. Voluntary and part time work is around the average. A variety of employment types exists with average unemployment. The student population is below average. |
| 6 | Less Affluent Older Singles, Couples and Single Parents | Accounts for 7% of households in Australia with mixed household sizes. A high proportion of over 65s reside here with below average numbers of children. Average indigenous population. Australian, European and mixed ethnic origins. Homes have average likelihood of being in an Urban area with a variety of housing types with mixed tenure. Household income and mortgage repayments are low with below average rental payments and qualification levels. Voluntary and part time work are below average. A variety of employment types exists with average unemployment. The student population is below average. |

| | | |
|---|---|---|
| 7 | Less Affluent Mixed Family Neighbourhoods | Accounts for 15% of households in Australia with household size between 1 and 4. Average numbers of children and elderly. Above average indigenous population. Australian and mixed ethnic origins. Homes are less likely than average to be in an urban area with high proportion of detached and other housing types. Voluntary work is above average. Part time work is close to average. A variety of employment types exists with average unemployment. The student population is below average. |
| 8 | Low Income Rural and Suburban Neighbourhoods | Accounts for 12% of households in Australia with household size between 1 and 4. Average numbers of children and elderly. Above average indigenous population. Australian, other Oceania and European ethnic origins. Homes are less likely than average to be in an Urban area with high proportion of detached and other housing types of mixed tenure. Mortgage and rental payments are below average with low household income and qualification levels. Voluntary and part time work is around the average. A variety of employment types exists and unemployment is average. The student population is below average. |
| 9 | Diverse Low Income Urban Communities | Accounts for 5% of households in Australia with household size between 1 and 6+.Average children, high numbers of over 65s. Low indigenous population. Mixed ethnic origins. Homes are more likely than average to be in an Urban area with a variety of housing types of mixed tenure. Mortgage repayments in these areas are above average, with below average rental payments. Household income is very low with low education levels. Voluntary and part time work is below the average. A variety of employment types exists but unemployment is high. The student population is below average. |
| 10 | Very Low Income Rural Communities | Accounts for 5% of households in Australia with mixed household sizes. Above average numbers of children and below average elderly. Vast indigenous population. Primarily Australian ethnic origins. Homes are much less likely than average to be in an urban area with a variety of housing types of mixed tenure. Mortgage and rental payments are low with very low household income and qualification levels. Voluntary and part time work are around the average. A full spectrum of employment types exist here but unemployment is high. Student population is very low. |

Table 3.6 - Summary of the ten Cameo Australia groups, adapted from (Callcredit Information Group, 2013)

Cameo USA was developed using data from the 2010 USA census at census block level, the smallest geographic unit for which data is reported. The classification is a two tier hierarchy with 10 groups which are further subdivided into 58 categories. The 10 groups are summarised in Table 3.7.

| **Cameo** | **Name** | **Summary** |
|---|---|---|
| 1 | American Aristocracy | Residents are primarily Caucasian, living in detached homes with 5+ rooms, which they own, as husband and wife with their children. Homes are in urban areas and worth between $320,000 and $760,000, with few homes being vacant. They have very high household income. The majority of residents have a university education. There is below average unemployment and employment is commonly in information and financial services, property, science and technical and management. There are also some employed in public administration and arts and entertainment. These residents rarely use public transport |
| 2 | Exclusive Society | Residents are primarily Caucasian and Asian, living in detached homes or apartments with 1-7 rooms, which they own or rent. While most live as husband and wife with their children, there are some living alone in their apartments. Homes are in Urban areas and worth between $270,000 and $450,000, with few homes being vacant. They have high household income. The majority of residents have a university education. There is below average unemployment and employment is commonly in information and financial services, property, science and technical and management. There are also some employed in public administration and arts and entertainment. These residents used mixed methods of transport. |
| 3 | Prosperous Families | Residents are primarily Caucasian and Asian, living in detached homes or with 5-7 rooms, which they own or rent. While most live as husband and wife with their children, there are some living alone or with other family members. Homes are mostly in urban areas and worth between $160,000 and $280,000, with few homes being vacant. They have above average household income. Some residents have a university education. There is below average unemployment. There are a variety of employment types including: public administration and arts and entertainment and in agriculture and mining, construction and manufacturing, information and financial services, property, science and technical and management. These residents used mixed methods of transport. |
| 4 | Enterprising Households | Residents are Caucasian, Asian and African American, living in a mix of home types which are either owned or rented. Property values range from $140,000 to $210,000 with 5-7 rooms in mostly urban areas, with some homes vacant. There is an average presence of children of mixed ages contributing to mixed household sizes. High numbers live alone with few living as husband and wife. Other family members are also present. Residents have a range of education levels with some who did not graduate high school through to those who have higher university degrees. There is below average unemployment and the majority choose to use private transport. There are a variety of employment types including: public administration and arts and entertainment and in agriculture and mining, construction and manufacturing, information and financial services, property, science and technical, management and logistics and utilities - providing a broad range of household incomes. |

| | | |
|---|---|---|
| 5 | Comfortable Communities | Residents are Caucasian and Asian, living in a mix of home types, including detached, apartments and mobile homes which are either owned or rented. Property values range from $130,000 to $200,000 with 1-7 rooms in mostly Urban areas, with some homes vacant. There is a mixed likelihood of children living here with varied household composition, typically between 1 and 4. Residents have a range of education levels with some who did not graduate high school through to those who have higher university degrees. There is below average unemployment and the majority choose to use private transport. There are a variety of employment types including: arts and entertainment and in agriculture and mining, construction and manufacturing, information and financial services, property, science and technical, management and logistics and utilities - providing a broad range of household incomes. Some individuals in this type of area claim social security income. |
| 6 | Aspiring Consumers | Residents are Caucasian, Asian, African American and other ethnicities living with mixed household composition, with below average concentration of children, forming household sizes 1-4. There are a high proportion of under 40's and over 60's. The majority of these residents earn less than $25,000 per year and an average amount claim social security. Homes are populated with mixed tenures, comprising attached, detached and apartments worth between $70,000 and $150,000 with between 1 and 7 rooms in Urban and mixed areas. Some homes are vacant. Individuals are of varied educational backgrounds with a lot who have not graduated high school but some have university degrees. There is above average unemployment with those who are employed working mainly in the following areas: Agriculture, mining, hospitality, construction and manufacturing and administration. Some work in logistics, public administration, health and education and arts and entertainment. |
| 7 | Dynamic Neighbourhoods | Residents are Caucasian, Asian, African American and other ethnicities living with mixed household composition, with average concentration of children. Households are mixed in size, with few living as husband and wife, but 1 or 2 individuals per household is common, with a large proportion of residents being under 40 years old. Most earn less than $25,000 per year and some claim social security. Most accommodation is rented in properties worth between $70,000 and $120,000 with between 1 and 7 rooms in urban and mixed areas. Some are vacant. Individuals are of varied educational backgrounds with a lot who have not graduated high school but some have university degrees. Here there is above average unemployment with those who are employed working mainly in hospitality and administration jobs. A few are employed in agriculture and mining, construction and manufacturing, health and education, arts and entertainment and public administration. Transport types are mixed. |

| | | |
|---|---|---|
| 8 | Diverse Communities | Residents are Caucasian, Asian, African American and other ethnicities living with mixed household composition, with average concentration of children. There is a higher proportion of under 40s than average, with pockets of high concentration over 60s. Households are mixed in size, with few living as husband and wife, but 1 or 2 individuals per household is common, with living with non family members being extremely common. Most earn less than $25,000 per year and few claim social security. Most accommodation is rented apartments worth between $80,000 and $150,000 with between 1 and 7 rooms in mostly urban areas with high rate of vacant properties. Individuals are of varied educational backgrounds with a lot who have not graduated high school but some have university degrees. Here there is average unemployment with those who are employed working mainly in hospitality and administration jobs. A few are employed in agriculture and mining, construction and manufacturing, science and technology, property, information and financial services and arts and entertainment. Transport types are mixed. |
| 9 | Stretched Tenants | Residents are Caucasian, African American and other ethnicities with average concentration of children. There is a higher proportion of under 40s than average. Households are mixed in size, with few living as husband and wife but high numbers living alone, with other family members and non family members. Most earn less than $25,000 per year and average numbers claim social security. Most accommodation is rented apartments worth between $50,000 and $70,000 with between 1 and 7 rooms in mostly urban areas with high rate of vacant properties. The majority of individuals did not graduate high school, and very low numbers going on to study for a degree at university. There are high unemployment rates but those employed work mainly in hospitality and administration. In a pocket of higher employment jobs vary between agriculture and mining, construction and manufacturing, science and technology, property, information and financial services, management and arts and entertainment. Transport types are mixed. |
| 10 | Strained Society | Residents are Caucasian, African American and other ethnicities with high concentration of children. There is a higher proportion of under 40s than average. Households are mixed in size, with low numbers living as husband and wife but very high numbers living with other family members. Most earn less than $25,000 per year and average numbers claim social security. Most accommodation is rented of mixed type worth between $40,000 and $110,000 with between 1 and 7 rooms in mostly urban areas with above average rate of vacant properties. The majority of individuals did not graduate high school. There are very high unemployment rates but those employed work mainly in hospitality and administration. For those that are employed jobs vary between agriculture and mining, construction and manufacturing, hospital and education and public administration. Transport types are mixed. |

Table 3.7 - Summary of the ten Cameo USA groups, adapted from (Callcredit Information Group, 2013)

#### *3.2.4.7 Boundary data*

UK boundary data is freely available to academic institutions in the UK via the UK Data Service (UK Data Service, 2013). Boundary data was downloaded in ArcGIS shapefile format (ESRI Inc, 2010). For the international case studies, Callcreditgroup provided boundary files in MapInfo format (MapInfo, 2012), which were then converted to ArcGIS shapefiles.

## 3.3 Methods

In order to avoid replication in subsequent chapters, a detailed description of the methods used for analysis are included here.

### 3.3.1 Statistical analysis

All statistical analysis has been completed using Stata statistical software: version 11 for the early analysis and version 12 for later work (StataCorp, 2012, StataCorp, 2009).

Due to the multiple testing which occurs in this analysis, statistical significance is considered at $p<0.01$. Bonferroni correction was not used because this would have provided conservative estimates. 1% significance level and judgement using confidence intervals seems most appropriate given that the multiple testing consists of testing categorical variables against categorical variables, rather than multiple testing of the same variables.

### 3.3.2 Causal diagram

In order to adjust for confounders and mediators in statistical models a Causal diagram was used (Greenland et al., 1999). In this causal diagram, only variables for which data was available for are included. Confounders are variables which correlate with the dependent (outcome) and independent (exposure) variable. The independent variable may be associated with a mediator variable which then, in turn, influences the dependent variable. The direction of causation is identified with an arrow in a causal diagram. Key outcomes through the thesis (dependent variables) are highlighted in a red box, for example, obesity. Key exposures (independent variables) are highlighted in red dashed line, for example, place of residence (spatial measure). Possible confounders and mediators, are in black boxes, for example, smoking. These relationships were based on evidence in the literature. See Figure 3.6 for the diagram used to inform analysis in Chapters 4, 5, 6 and 7.

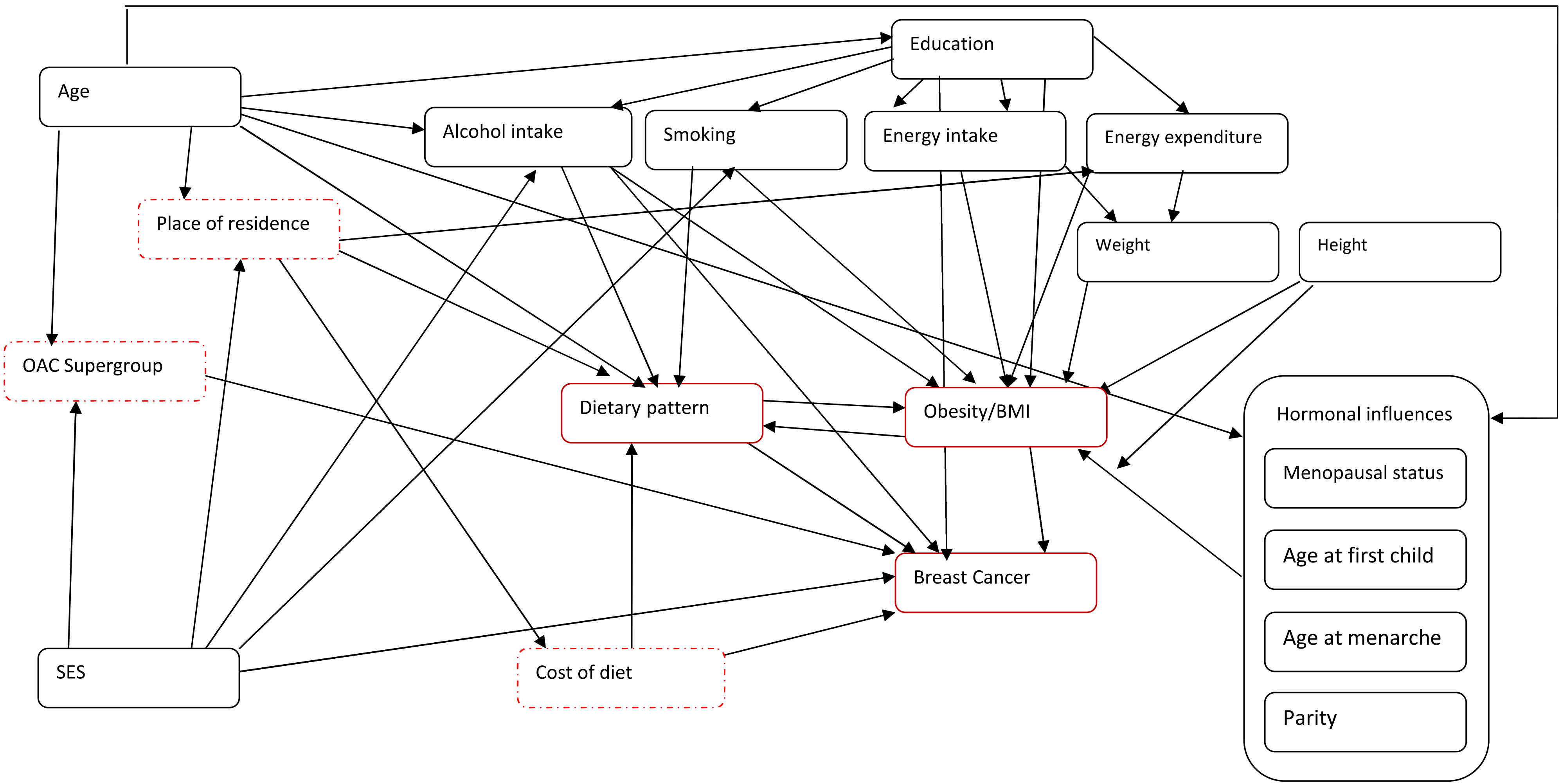



Figure 3.6 - Causal diagram - illustrating links between available data to inform statistical modelling

### 3.3.3 Logistic regression

Logistic regression is used when predicting an outcome which is a categorical dependent variable. A binomial logistic regression is appropriate when there are two categories. For more than two categories, which are not ordered, a multinomial logistic regression is used (Pevalin and Robson, 2009).

Binomial logistic regression is used in Chapter 6, in order to investigate geographic variation in diet for overweight and obese compared to those of a normal weight. In this case the UKWCS was dichotomised into normal weight and overweight/obese, with underweight women excluded. Chapters 4, 5, 6 and 7 investigate a dietary pattern outcome, for which there are 7 unordered categories. For this, multinomial logistic regression was performed.

### 3.3.4 Linear regression

When the dependent variable is continuous, a linear regression is most appropriate. This was used in Chapters 6 and 8, investigating predictors of BMI as a continuous measure. When the distribution of the continuous variable is skewed, a robust correction was used in the Stata syntax to use robust standard errors to account for the skewed data.

### 3.3.5 Survival analysis

Survival analysis is also referred to as a time to event analysis. This type of analysis accounts for the amount of time which an individual has been followed up without experiencing an event. In the context of this thesis, survival analysis has been used to investigate whether dietary patterns, diet cost or spatial measures can predict likelihood of breast cancer incidence. Survival analysis is preferred to other regression methods as it accounts for differences, due to time, over the duration follow up (Cleves et al., 2010).

Survival analysis is used in Chapter 7 of this thesis. Cox proportional hazards regression was used to perform time to event analysis, where the event is incidence of breast cancer. The censor date for breast cancer incidence was set to 1st October 2011. This includes updates from the NHS Information Centre (NHSIC) up to and including June 2012, but due to a time lag at the NHS in the processing of cancers the censor date is 8 months earlier. Any women who had died before the censor date were excluded. As the breast cancer incidence data is sent directly from the NHSIC, which the women agreed to at the time of consenting to the study, loss to follow up will be minimal. However, if any of the women have moved away from the UK and not informed the Nutritional Epidemiology Group (NEG), they will be treated as having not developed breast cancer.

In order to test the assumptions of a proportional hazard regression, i.e. that the hazard ratio does not vary with time, Kaplan-Meier graphs were created for each variable in the regression models (Cleves et al., 2010).

Log rank tests were used to test for difference between the survival curves (Peacock and Kerry, 2010) in different subgroups of the UKWCS women. In this test the difference between observed outcomes and expected outcomes (which are based on the assumption that the probability of the event occurring is the same in each group) are presented and difference identified with a p value less than 0.05.

### 3.3.6 Linking spatial data

Spatial identifiers were matched to the cohort spatial identifier - via postcode using Geoconvert for UKWCS, as described in section 3.2.4. For the ALSWH and SOS cohorts, the study participants were linked to the Cameo groups using a function in ArcGIS which matches via spatial location, using longitude and latitude coordinates for residential address of participants. The Cameo group id was then added to the demographic datasets in Stata and longitude and latitude values removed to maintain anonymity.

### 3.3.7 Spatial analysis and estimating small area diet

All maps presented were created using ArcGIS.

Geodemographic identifiers - OAC in UK analysis and Cameo in for Australian and USA analysis - were used as a method to estimate small area dietary patterns. Cohort values for specific dietary patterns, diet cost or overweight and obesity were assigned to the geodemographic category and then used to estimate patterns for a given city based on the assumption that individuals in that cohort are typical of the population in the target city. A table of values in Microsoft Excel were imported into ArcGIS and matched into the boundary file on the geodemographic id. Chloropleth maps, which display variation using different colours, were then created to map the variation of the variable of interest. These methods were used in Chapters 5 and 8.

## 3.4 Evaluation of the cost database

In order to meet the aims of the thesis relating to cost of diet, a food cost database was required. In order to assess the suitability of the Nutritional Epidemiology Group at the University of Leeds, food cost database for use in estimating cost of diet in population research, an evaluation study was carried out. This evaluation study is described here. This section differs in format from the data and methods sections in this chapter so far, as it details the complete evaluation study, including an introduction, methods, results and discussion. However, since it describes the evaluation of a dataset used in further analysis, is appropriate to include within this chapter.

### 3.4.1 Introduction: Cost database evaluation

A healthy diet can promote health and reduce risk of chronic disease (Key et al., 2002, Foresight, 2007). However, not everyone consumes a healthy diet (Department of Health and Food Standards Agency, 2012). One of the factors thought to contribute to dietary patterns is how much the food costs. A healthy diet has been shown to be more expensive than a less healthy diet (Cade et al., 1999, Darmon et al., 2004, Maillot et al., 2007b).

Effective measurement of dietary expenditure is challenging. Examples of the methods used include: till receipt collection or an expenditure diary (Defra, 2009); market sales data (Sun, 2005); retrospective expenditure questionnaires (Turrell and Kavanagh, 2005); and estimation using food price databases (Ryden and Hagfors, 2011). The first of these methods, as used by the Living Costs and Food Survey (Defra, 2009), requires households to collect till receipts or record all expenditure in a diary for a set period. Its ease of administration makes it desirable for large-scale surveys; but it does not directly assess individuals' dietary intake. The retrospective expenditure questionnaire also carries a low administrative burden, with the added advantage of a single time point of data collection. However, reliance on retrospective, self-reported information introduces a chance of recall bias, and specificity may be lost as foods are often aggregated into groups. Market sales data can be readily available from market research companies, but again dietary consumption must be inferred from purchase data and demographic or health information may be absent. The final method, estimation using food price databases, is increasingly common (Monsivais et al., 2010, Waterlander et al., 2010) in the food cost literature, and relies upon the assignment of an average food price to the foods reported as consumed in a diet diary or FFQ. Only this latter method makes use of established dietary assessment techniques. The approach of the other methods is to infer diet from purchase data, the main drawback of which is that not all foods purchased will necessarily be consumed by the purchaser. Conversely, using dietary data requires purchasing behaviour to be inferred from dietary consumption.

The validity of price databases is critical if these are to be used in research, yet there appears to be no literature evaluating them. Work has been done to compare how well diet cost calculations agree using different dietary assessment methods (Murakami et al., 2008).

### 3.4.2 Methods: Cost database evaluation

In order to investigate the effect of cost of diet, a food cost database was used. The analysis used a database developed by the Nutritional Epidemiology group at the University of Leeds using the McCance and Widdowson composition of food codes (Holland B, 1991) and supermarket food prices. In order to assess the reliability of using this database an evaluation was carried out in equal collaboration with fellow PhD student, Kate Timmins. A paper discussing the findings using the Supermarket Nutrition Information Project (SNIP) study has been published in the European Journal of Clinical Nutrition (Timmins et al., 2013). Results of the full evaluation are included below.

The validity of food costs applied to food intake using the Diet and Nutrition Tool for Evaluation (DANTE) cost database was assessed. This was carried out by comparing the daily expenditure recorded by till receipts collected over a number of weeks with the DANTE daily costs calculated from food diary dietary intake data. Till receipts and dietary intake data were collected concurrently.

A database containing food cost data was created in 2004 by the Nutritional Epidemiology Group at the University of Leeds. The costs were obtained online from supermarket websites: primarily Tesco, but where products were not available alternative stores and specialist shops were used. E.g. Sainsbury's, Goodness Direct and Gorton's of Gloucester were used. Prices were translated into cost per edible 100 grams. Foods were then mapped to food codes from the McCance and Widdowson composition of food tables (Holland B, 1991) in order to incorporate them into the in-house dietary assessment tool, DANTE. The food portion size handbook (Food Standards Agency, 1998) was used as a reference guide where food weights were not available. Special offer prices were not included. Three cost levels were assigned to most foods - low, medium and high - along with an average cost value. Approximately 3000 products were assigned costs. The cost database is referred to as the 'DANTE food cost database'. The DANTE food cost database was expanded in 2008, growing in size to 3192 products. The consumer price index (not food-group specific) was applied to the prices of more recently added items to bring them back in line with prices in 2004, the date when original items were entered into the database.

Till receipt and food diary data from two previous studies were used to evaluate the DANTE food cost database. The samples are described further below.

#### *3.4.2.1 Supermarket Nutrition Information Project sample population*

The Supermarket Nutrition Information Project (SNIP), collected till receipts alongside dietary surveys in 1998-99 (Ransley et al., 2003) with the main aim to assess the validity of using supermarket purchase information to estimate nutrient intake. This data offers a unique opportunity to assess the level of agreement between diet costs estimated from food diaries and the actual expenditure recorded from till receipts. Households (n = 284) were recruited from the Tesco Clubcard database held at the Roundhay store in Leeds. Participants were instructed to collect till receipts for all purchases of food for human consumption made over a 28-day period in 1998-1999. In addition, a weighed intake diet diary was completed for every member of the household over four days (three weekdays and one weekend day). The completion rate was 75%, with data available for 214 households, comprised of 522 individuals.

Following quality assurance checks (see 3.4.2.3 below), the sample was reduced to 325 individuals from 161 households after excluding individuals with missing household composition data (n = 28) or missing recipe information (n = 169). The final sample had a mean household size of two, and included adults (n=256, 79%) and children (n=69, 21%). Sample characteristics are presented in Table 3.8.

| Descriptor | Sample value |
| --- | --- |
| Mean household size | 2 (range 1-5) |
| % White | 94 |
| % Female | 53 |
| Age range (years) | 1-87 |
| % Adult | 79 |
| Social class of the majority | Intermediate and junior non-manual (50%) |
| BMI adults (kg/m$^2$) | 25.01 (24.45 to 25.57) |
| Mean[1] daily energy intake (MJ)[1] | 7.15 (6.88 to 7.43) |
| Mean daily energy intake - excluding top 5% (MJ) | 6.97 (6.69 to 7.26) |
| Mean daily energy intake - adults (MJ) | 7.32 (7.02 to 7.63) |
| Mean daily energy intake - children (MJ) | 6.54 (5.87 to 7.21) |
| Mean daily energy intake - males (MJ) | 7.93 (7.53 to 8.34) |
| Mean daily energy intake - females (MJ) | 6.48 (6.13 to 6.82) |

Table 3.8 - Characteristics of the SNIP sample. [1] 95% confidence intervals in brackets. [2] Energy intakes as calculated from diet diaries

Daily diet costs were generated from the diet diary information using the DANTE cost database, taking an average across the days. The participants also kept a ‘pocket book’ for all food eaten outside the home; but data from these were excluded from the analysis as they would not have appeared on the till receipts. Dietary data were not assessed for potential over- or under-reporting.

The SNIP data were collected approximately five years prior to the development of the DANTE cost database. To account for change in price over time, data from the Consumer Price Index (Office for National Statistics, 2011a) were used to calculate 27 food group-specific correction factors, which were then applied manually to the DANTE cost database to adjust prices to 1998-99 figures.

Using the till receipts, a daily diet cost can be calculated. To account for waste resulting from spoilage, inedible parts or discarding, a correction factor of -15% was applied (as per Department for Environment Food and Rural Affairs recommendations (Defra, 2010)).

Ethical approval for the SNIP study was obtained from the Leeds Health Authority, United Teaching Hospitals NHS Trust on the 13th March 1997 (Ransley et al., 2003).

### *3.4.2.2 Single Women from the UK Women's Cohort Study sample population*

In 2004, 200 single-living women, shopping for one person, randomly selected from the UKWCS cohort, were approached to participate in a food cost study. Fifty women agreed to take part. Of these, 36 returned till receipts and food diaries giving a completion rate of 72%. Sample characteristics are presented in Table 3.9.

| **Descriptor** | **Sample value** |
|---|---|
| Individuals (n) | 36 |
| Households (n) | 36 |
| Household size | 1 |
| % White | 89* |
| % Female | 100 |
| Age range (years) | 52-81 |
| Social class of the majority | Professional |

Table 3.9 - Characteristics of the Single Women from UKWCS sample

*The remaining 11% of the sample did not report ethnicity

Food diaries were coded using DANTE. A total cost spent on food for each woman was calculated by summing till receipts collected during the study period. Participants indicated in the diaries if foods were homegrown or bought outside the usual household purchases (for example, at a work canteen). These foods were not coded in DANTE and therefore excluded from this analysis, as there would have been no corresponding till receipt record.

As the data from this sample was collected at the same time that the DANTE cost database was developed, there was no need to apply a correction factor to adjust for change in cost over time. To account for waste resulting from spoilage, inedible parts or discarding, a correction factor of -15% was applied (as described above for the SNIP sample).

Ethical approval for the UKWCS was obtained from 174 research ethics committees across the UK between 1994 and 1995 (Woodhouse et al., 1997).

#### *3.4.2.3 Quality assurance of data entry - SNIP*

As the data from the two samples was collected for previous studies, data entry for the till receipt and diary data was carried out at an earlier date. To ensure the integrity of this data a quality assurance check was completed for a random sample of the data.

For SNIP a 5% sample of both till receipts and diary coding was checked. The till receipt values were summed and checked against the previously recorded values in the database and deemed satisfactory. However, for the diaries, the quality assurance identified a number of missing items, in particular recipe based foods. On investigation, it was found this was due to the fact that the data were originally coded using the Weighed Intake Software Program (WISP) for Windows v1.2, which used an earlier version of the McCance and Widdowson codes to those used in DANTE. To rectify this problem, nineteen food items (accounting for 73% of the missing values) were recoded for the purpose of this study to ensure compatibility. Participants with missing recipe information, which could not be matched during the recoding, for example pheasant casserole, were excluded from the validation.

#### *3.4.2.4 Quality assurance of data entry - Single Women from the UK Women's Cohort Study*

Following quality assurance of the Single women of the UKWCS sample, no data cleaning was required. Both the till receipt records and diary coding were deemed satisfactory for use.

#### *3.4.2.5 Statistical analyses for SNIP*

Daily mean values were calculated for both the DANTE food cost estimates and the till receipt totals, to reflect habitual dietary expenditure. An outlier was evident with respect to the DANTE food cost estimates: this was identified as anomalous (due to consumption of large quantities of bottled water which was not purchased during the collection period) and removed. The distributions of both variables were judged to deviate from normality. Therefore, the till receipt expenditures were tested for difference to the costs estimated by DANTE using Wilcoxon matched pairs. Significance levels were set at 5%. Pearson's product moment was used to test correlation between the methods. Agreement between the two methods was then assessed using Bland Altman difference plots.

A sensitivity analysis was undertaken, with the top 5% of spenders in each collection method removed. This excluded those who spent over £5.38 (n=17) as calculated by DANTE and over £5.84 (n=18) as estimated by the till receipts (total n=33).

These analyses were also repeated for subgroups: adults, children, males and females. Statistical analyses were performed using Stata IC 11 (StataCorp, 2009).

#### *3.4.2.6 Statistical analyses for Single Women from the UKWCS*

Daily mean values were calculated for both the DANTE food cost estimates and the till receipt totals, to reflect habitual dietary expenditure. The normally distributed till receipt expenditures were tested for difference to the costs estimated by DANTE using a paired t-test. Significance levels were set at 5%. Agreement between the two methods was then assessed for each sample using Bland Altman difference plots. Statistical analyses were performed using Stata IC 11 (StataCorp, 2009).

### 3.4.3 Results: Cost database evaluation

This section presents the results of evaluation for each study population, the SNIP and Single women from the UKWCS. Both samples show good agreement at the population level.

#### *3.4.3.1 Results from the SNIP sample*

Summary statistics for the daily cost estimates are presented in Table 3.10. Whole-sample analyses showed no significant differences between the two methods. However, a difference was observed when considering females alone or children alone (Table 3.10).

| | **DANTE cost database** | | **Till receipts** | | **Wilcoxon matched pairs p value** |
|---|---|---|---|---|---|
| | **Mean (95% CI)** | **Median (IQR)** | **Mean (95% CI)** | **Median (IQR)** | |
| Full sample (n=325) | 2.96 (2.82 to 3.10) | 2.88 (2.01 to 3.72) | 3.06 (2.91 to 3.20) | 2.71 (2.16 to 3.73) | 0.81 |
| Full sample excluding top 5% (n=292) | 2.76 (2.63 to 2.88) | 2.75 (1.88 to 3.55) | 2.78 (2.67 to 2.89) | 2.58 (2.09 to 3.45) | 0.80 |
| Males (n=152) | 3.17 (2.97 to 3.38) | 3.07 (2.15 to 3.89) | 3.11 (2.88 to 3.33) | 2.76 (2.15 to 3.75) | 0.23 |
| Females (n=172) | 2.75 (2.56 to 2.94) | 2.63 (1.78 to 3.51) | 3.02 (2.82 to 3.21) | 2.69 (2.16 to 3.72) | 0.04 |
| Adults (n=256) | 3.19 (3.04 to 3.35) | 3.06 (2.32 to 3.10) | 3.18 (3.01 to 3.36) | 2.77 (2.26 to 3.81) | 0.36 |
| Children (n=67) | 2.00 (1.79 to 2.20) | 1.83 (1.39 to 2.51) | 2.55 (2.30 to 2.80) | 2.31 (1.96 to 2.96) | 0.001 |

Table 3.10 - Estimated daily dietary costs (£) of the SNIP sample and subgroups using till receipts and the DANTE cost database, with results of Wilcoxon matched pairs tests for difference between the methods

Correlation coefficients are shown in Table 3.11. Overall, the cost estimates of the two methods correlated weakly but significantly (r=0.335, p<0.001, 95% CI 0.234, 0.428). No significant correlation was evident when children were analysed alone (Table 3.11).

| | Pearson's product moment r | 95% CI | p value |
|---|---|---|---|
| Full sample (n=325) | 0.335 | 0.234 to 0.428 | <0.001 |
| Full sample excluding top 5% (n=292) | 0.305 | 0.197 to 0.406 | <0.001 |
| Males (n=152) | 0.345 | 0.197 to 0.478 | <0.001 |
| Females (n=172) | 0.330 | 0.190 to 0.457 | <0.001 |
| Adults (n=256) | 0.298 | 0.182 to 0.406 | <0.001 |
| Children (n=67) | 0.229 | -0.012 to 0.445 | 0.06 |

Table 3.11 - Correlations between till receipt and DANTE cost database cost estimations for the full sample and subgroups

The Bland Altman plot of the differences can be seen in Figure 3.7, which shows 95% limits of agreement (±2σ) of £2.88 and -£3.08. The spread of scatterpoints widens as the mean difference between the methods increases. This was confirmed when a regression trend was fitted to the Bland Altman, as the 95% confidence limits widened along the x axis (Figure 3.9). As such, a sensitivity analysis was conducted, re-plotting the Bland Altman after excluding the top 5% of spenders in each of the diet cost estimation methods (Figure 3.8). This excludes those who spent over £5.38 (n=17) as calculated by DANTE and over £5.84 (n=18) as estimated by the till receipts (total n=33). The 95% limits of agreement (±2σ) are £2.31 and -£2.35. There were no significant differences between the two methods (t = -0.30, p =0.76). The mean difference for this reduced sample, was £0.02 (95% CI -0.15, 0.11).

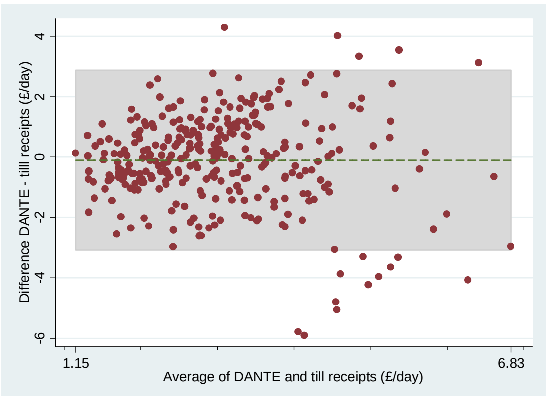



Figure 3.7- Differences between daily costs estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents 95% limits of agreement)

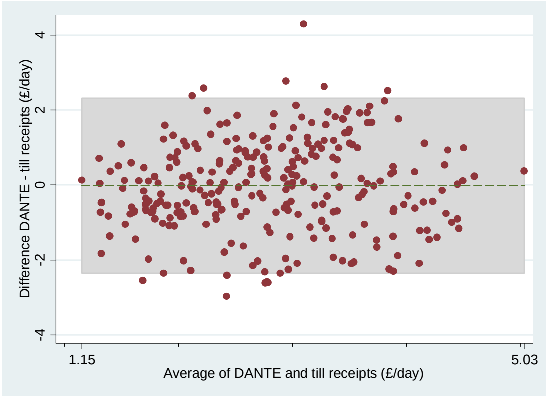



Figure 3.8 - Differences between daily costs estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents 95% limits of agreement)

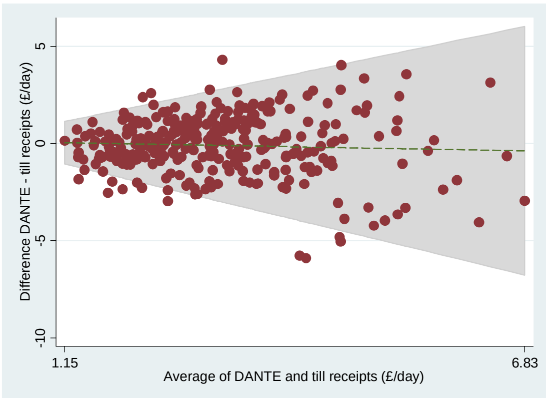



Figure 3.9 - Differences between daily costs - in the full sample - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents regression line 95% limits of agreement)

Results of the Bland Altman analyses for the full sample and subgroups are summarised in Table 3.12 (The related Bland Altman plots are included in Appendix B). The Bland Altman plots with a regression trend fitted display widening limits of agreement in all subgroups, indicating reduced agreement at higher diet costs. Males exhibited a similar pattern in agreement to the whole sample, both with (Appendix B, Figure 9) and without the top 5%. Females showed a reduction in the widening limits of agreement on exclusion of the top 5%, but not to the extent of the whole sample, or of males.

On excluding children, the mean difference was as small as £0.01, although limits of agreement remained similar to the whole sample estimates. Although the limits of agreement narrowed on excluding the top 5%, the mean difference between the methods increased when adults were analysed alone.

| | | | | **Excluding top 5%** | | |
|---|---|---|---|---|---|---|
| | **Mean difference**[1] | **95% limits of agreement** | | **Mean difference** | **95% limits of agreement** | |
| | | **Lower limit** | **Upper limit** | | **Lower limit** | **Upper limit** |
| Full sample | -0.10 | -3.08 | 2.88 | -0.02 | -2.35 | 2.31 |
| Males (n=152) | 0.07 | -2.95 | 3.09 | 0.16 | -2.21 | 2.52 |
| Females (n=172) | -0.27 | -3.16 | 2.63 | -0.19 | -2.42 | 2.04 |
| Adults (n=256) | 0.01 | -3.08 | 3.09 | 0.11 | -2.18 | 2.41 |
| Children (n=67) | -0.55 | -2.86 | 1.75 | -0.50 | -2.67 | 1.67 |

Table 3.12 - Results of the Bland Altman analyses comparing the agreement between the DANTE cost database and till receipt estimates. Figures presented for the full sample and for each of the subgroups, both with and without the top 5%.

[1] DANTE cost database minus till receipt estimates

### *3.4.3.2 Results from the Single Women from UKWCS sample*

An outlier, due to an incorrectly coded entry in a food diary, was detected and removed from the normally distributed UKWCS data. Following this, the mean daily cost given by DANTE was £3.96 (range: £1.97 to £6.28); and mean daily expenditure from till receipts was £3.75 (range: £0.25 to £7.95).

No significant difference between the means of the two methods of cost estimation were found (paired t-test: $t = -0.81$; 95% CI -0.74, 0.32; $p = 0.43$). The mean difference between the two estimates was £0.21 (range: -£2.90 to £2.90). Plots of the difference between the means indicated normal distribution, and a Bland Altman plot was generated which is displayed in Figure 3.10. The 95% limits of agreement (±2σ) were found to be £3.22 (upper) and -£2.80 (lower). No noteworthy bias toward over- or under-estimation was evident.

When a line of best fit is plotted, there appears to be a tendency for DANTE to underestimate the more expensive diets and overestimate those which cost less. See Figure 3.11.

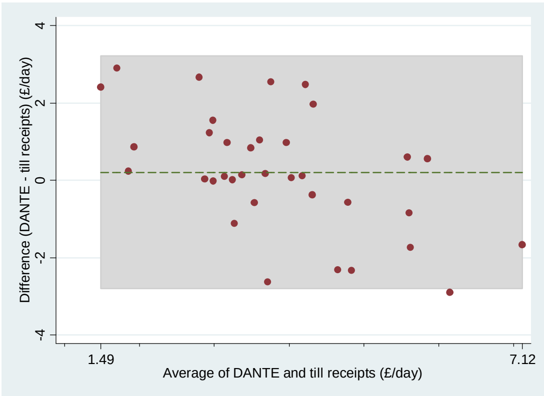



Figure 3.10 - Bland Altman plot of the difference between daily costs estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste), using the UKWCS data (£/day)

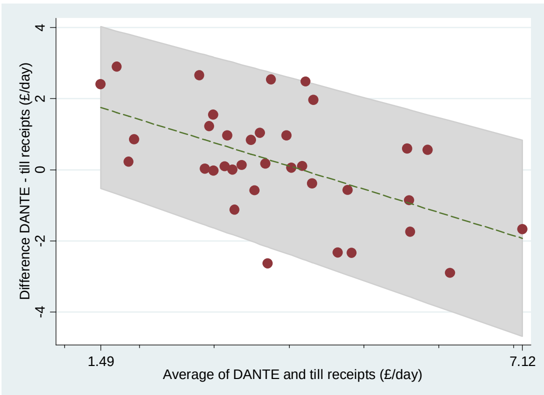



Figure 3.11 - Bland Altman plot of the difference between daily costs estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste), using the UKWCS data (£/day). Agreement plotted around a line of best fit

Due to the small numbers and homogeneity in this sample, it was not appropriate to perform any sensitivity or subgroup analysis.

### 3.4.4 Discussion: Cost database evaluation

Although food cost databases are a widely employed methodology in the literature (Monsivais et al., 2010, Waterlander et al., 2010), previous studies have neglected to assess their validity in estimating dietary expenditure. Some have compared food costs calculated using different dietary assessment methods (Murakami et al., 2008, Timmins et al., 2013). This study should assist researchers in interpreting estimates of diet cost provided by such databases.

The DANTE cost database estimated an average daily diet cost in the SNIP sample of £2.88. This compares well with the average daily expenditure for the GOR Yorkshire and the Humber (where the SNIP participants reside) of £2.96, as reported by the 2003-2004 Expenditure and Food Survey (Defra, 2004).

In our SNIP full-sample comparison, the mean difference between the DANTE cost database estimates and the till receipt calculations was £0.10. However, the limits of agreement suggest the DANTE cost database may mis-estimate daily diet costs for an individual by up to £3.00. This constitutes a potentially substantial mis-estimation, given that the mean daily expenditure calculated from till receipts was £3.05. The results of the sensitivity analysis imply that the database is most valid for the 95% of the sample spending less on their diets. When the more expensive diets in the sample were excluded, both the mean difference between the two methods and the limits of agreement were reduced. Although the limits of agreement suggest a potential mis-estimation at the individual level of about £2.30, the mean difference of £0.02 shows a reassuring level of accuracy at the group level. It is difficult to speculate why the cost of diet exhibits better agreement at the lower end of this cost spectrum, but perhaps lower costs foods vary less from the average price - used in the DANTE cost database - than the more expensive products do.

In the SNIP sample subgroup analyses, both methods revealed between-group differences in the same direction. These were greater when using DANTE to estimate costs, rather than till receipts. There was variation in the methods' agreement between males and females, and between adults and children. In particular, the DANTE cost database estimates for children varied noticeably from the till receipt values, on average exhibiting lower costs. This most likely reflects a drawback in the till receipt method, which assumed an equal consumption across members of the household. In actuality, both the quantity and composition of diet is likely to differ across the family unit (Department of Health and Food Standards Agency, 2012), patterns which are more likely to be captured using dietary assessment. The results of this study support this, showing decreased agreement in the subgroups likely to consume a smaller quantity of food.

In this study, no subgroup analysis was performed with respect to including/excluding alcohol or sugar sweetened beverages, as performed by Murukami et al 2008. This could be considered for future analysis.

The DANTE diet cost estimation for the single women of the UKWCS sample of £3.96 is slightly above the national average daily expenditure of £3.24, as reported by the 2003-2004 Expenditure and Food Survey (Defra, 2004). Difference between the two methods of estimating diet cost was £0.21 per day greater than that observed for the SNIP sample, whilst still not showing any statistically significant difference between the methods. With only 35 women in the analysis sample, the results are likely to be less reliable than they would be for a larger sample. However, results seem to support the findings demonstrated by the SNIP sample. These women have more expensive diets than the SNIP sample and the difference between the methods is greater, which supports the findings from the sensitivity analysis in using the SNIP sample, which showed that DANTE is least reliable at estimating diet cost in the most expensive diets. The upper and lower limits of agreement between the cost methods are very similar in both samples. Whilst the SNIP sample provides more confidence in the results, due to the larger sample size, it is reassuring that the results displayed from the single women of the UKWCS concur.

These results suggest that calculating the cost of food using dietary assessment data is useful in estimating the monetary value of a population's diets, and the cost estimates are comparable to recorded expenditure - in the SNIP sample particularly for the adult majority who spend less on their diets. It is for these people that diet cost research is arguably of most relevance, as it is those on a limited budget whose dietary patterns are most likely to be influenced by diet cost considerations.

These studies rely upon the assumption that the measurements of dietary consumption (using diet diaries) and expenditure (via till receipts) are reflective of habitual patterns. In order to gather a more precise measure of actual cost of daily intake, the foods recorded as eaten would need to be directly matched to the foods bought. This latter method is impractical for a number of reasons, largely due to the extensive time and participant costs involved. Therefore, usual daily expenditure and consumption were considered the best available measures to use.

While four-day collections of diet diaries are widely accepted as indicative of an individual's habitual behaviour in large-scale research (Bingham et al., 1994, Gibney, 2004), this may not hold true for short periods of till receipt data collection. In some households, for example, a fortnight's or month's worth of till receipts may not capture typical store cupboard purchases, the entertainment of guests or atypical eating out patterns. Dietary assessment is prone to measurement error (Freedman et al., 2011) and it is possible that, in addition to true variation in dietary intake, some misreporting has taken place.

The till receipt value in the SNIP sample is a household average, treating adults and children as equal in the mean daily expenditure calculations. If it is assumed that children and adults consume unequal proportions of the household expenditure, the daily individual estimate could have over-estimated children's expenditure and under-estimated that of adults.

The use of averaged food price databases, whilst widespread in the literature, has been acknowledged to have drawbacks (Cade et al., 1999, Waterlander et al., 2010, Aggarwal et al., 2011). The most commonly cited issue is that no consideration is given to food eaten outside the home (Cade et al., 1999). Recent UK data suggest that eating out accounts for 11% of total energy intake on average (Defra, 2009). In this study, foods bought and consumed outside the home were recorded and coded separately to the diary data. This means that the DANTE cost database, as it stands, is valid only for food consumed within the home.

A further common criticism of using a database of national average prices is that these prices may not be indicative of the prices faced by certain populations (Monsivais et al., 2010). Geographical variations, as well as retailer availability and access (Wrigley et al., 2002, Jiao et al., 2012), could affect costs encountered. Using mean prices could result in an overestimation of diet expenditure for groups which consistently purchase foods at lower than average prices (or vice versa), or those who make use of discounts and promotions to stretch their budget (Beatty, 2010). Neither of the two studies were designed to be nationally representative, and it is possible that the limits of agreement reflect a variation in product prices: only the mean costs of each food item from the DANTE cost database were employed, whereas the full range of costs may have been represented in the till receipts. In future applications of the DANTE cost database, there is the potential to use the low and high values within the database.

Within the DANTE cost database, some less common foods are without cost information - for example, some exotic fruits (rambutan) and offal (trotters and tails). Six of these foods were reported in the diaries of four participants from the sample. This may have resulted in an underestimation of expenditure for these participants. It is unlikely that the small amounts involved will have skewed the results.

The DANTE cost database boasts an important advantage: using dietary assessment methods - which reflect individual consumption - is likely to provide a clearer picture of cost of diet than relying upon household expenditure data. Of the methods used to gauge costs associated with diet, using a cost database linked to food intake remains the only one to utilise sound dietary assessment methods. This validation was therefore critical in assessing the accuracy of the estimated costs associated with habitual diet.

Cost of diet is likely to warrant an increasingly important role in public health research. The increasing economic pressures of recent years have elicited growing concern about the affordability of a healthy diet, and establishing whether diet costs contribute to inequalities in health could have far-reaching implications for public health policy.

### 3.4.5 Conclusion: Cost database evaluation

A food cost database linked to a dietary assessment tool agrees well with estimates from household expenditure at population level. Agreement was stronger for the 95% of the population spending less on their diets, and for adults. Estimating diet costs will always have limitations, but using a cost database linked to food composition tables remains a pragmatic method for large-scale dietary research. These results should help improve confidence in the interpretation of research assessing the monetary value of diets.

## 3.5 Summary and context

In this Chapter the data and methods used in Chapters 4 to 8 of this thesis have been explained. An evaluation of the food cost database has also been included which concludes that the DANTE food cost database is suitable for use in population research. Analysis using this food cost database is reported in Chapters 5, 6 and 7.

In Chapter 4 the cost of dietary patterns in the UKWCS will be reported, using DANTE food cost database and dietary patterns which were described in this chapter.

# Chapter 4: Cost of dietary patterns in the UKWCS

## 4.1 Overview

Chapter 1 described the aims and objectives and provided an overview of the thesis. Chapter 2 reviewed the existing literature in this area. Chapter 3 explained the data and methods to be used in the research. This Chapter uses food intake data from the UKWCS and assigns cost to the diet using the DANTE food cost database, as described in Chapter 3.

Dietary patterns are rated according to their healthiness in relation to the UK Department of Health's Eatwell Plate. The relationship between cost and healthiness is then explored. A reminder of how Chapter 4 fits into the thesis flow is shown in Figure 4.1.

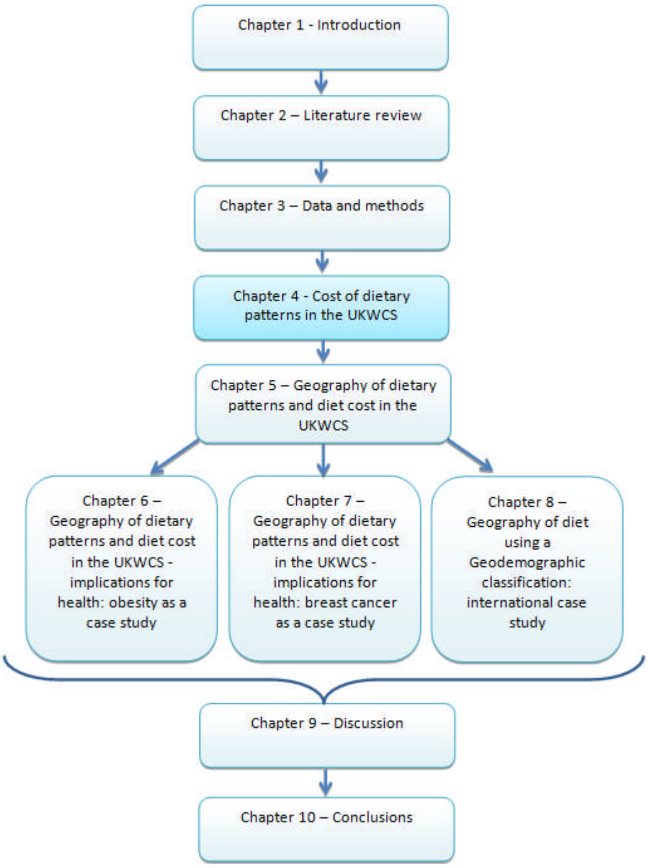



Figure 4.1- Overview showing how Chapter 4 fits into the overall thesis flow

## 4.2 Introduction

A healthy diet is important to promote health and wellbeing whilst preventing chronic disease. Diet is a well known modifiable risk factor for many chronic diseases such as obesity, cardiovascular disease and cancer (Key et al., 2002). However, consumption of a healthy diet can be challenging and gives rise to a number of questions. What constitutes a healthy diet? How do we measure a healthy diet? How much will it cost?

In order to answer these questions we need a robust indicator of a healthy diet. The presence of an individual food or nutrient in a diet provides little indication of whether that diet is healthy or not. Healthy eating guidelines may vary between developed countries but they tend to provide the same general message. In the UK, the Department of Health promote their dietary recommendations for optimum health using a pictorial illustration; the 'Eatwell Plate' (Department of Health, 2011), encouraging an overall healthy diet, rather than consumption of specific foods (Figure 4.2).


Figure 4.2 - The Eatwell Plate. Source https://www.gov.uk/government/publications/the-eatwell-plate-how-to-use-it-in-promotional-material

Data driven dietary patterns, created using techniques like factor analysis or cluster analysis are useful to identify patterns which exist in the dietary data of a specific study population (Hu, 2002). However, they do not necessarily offer an indicator of healthiness of a diet. Alternative methods measure healthfulness of diet according to predefined patterns. Some examples include: a diet quality index which assigns a score according to how well a diet conforms to components indicative of a healthy diet; a healthy eating index which assesses adherence to national/federal dietary recommendations; a dietary variety score considers the number of different food items consumed over a given time period. There are many derivatives of each of these methods used to assess the diets of different populations (Drewnowski et al., 1997, Kennedy et al., 1995, McNaughton et al., 2008, Kim et al., 2003). Combining dietary pattern methods with a healthy eating index could provide the best of both.

Diet choice may vary due to health, personal taste, income or cultural reasons, so while public health guidelines encourage consumption of a 'healthy diet', the choice and purchase of food is the responsibility of an individual or household. In the current economic climate, with rising unemployment and associated fall in income, combined with increased costs, people are making savings where they can. Food/grocery shopping is one of these places (Crossley et al., 2012, Office for National Statistics, 2012, United States Department of Agriculture, 2011). In the developed world, the choice of food is wide and varied so where cheaper food alternatives are available it could influence food purchasing. The increase in market share of 'discount' food retailers in the UK highlights this demand for cheaper food (Thompson et al., 2012).

In recent years there has been increased interest in how the price of food affects food consumed (Lee et al., 2011, Lee et al., 2013). The majority of this research shows that a healthy diet is a more expensive diet. It has been suggested that the least healthy, nutrient poor diets are consumed by the less affluent (Ryden and Hagfors, 2011) while those with more money can afford a more expensive diet, including options which are recommended to promote health.

Measuring diet accurately in a population is challenging and subject to measurement error (Freedman et al., 2011). Assigning a cost to a diet is also complex. Commonly used methods are till receipt collection - as used by the Family Food Survey in the UK - (Defra, 2009) or assigning prices from a food cost database (Ryden and Hagfors, 2011). As with nutrient analysis, it is not realistic to consider the cost of certain foods in isolation as that is not how they are consumed, so understanding dietary pattern consumption habits are not only important with respect to health but also in relation to cost.

Research carried out in the UK shows that the poor do spend less on food, mostly as a result of optimising on quantity discounting offers, such as 'buy one get one free' (Beatty, 2010) which may in turn promote monotony, or mean that a large amount of food storage space is required. So whilst variety is widely recommended for a healthy diet (Krauss et al., 1996, Department of Health, 2011, Australian Government, 2005, European Food Information Council, 2009), it has been shown to be more expensive (Ryden and Hagfors, 2011) which could amplify differences in dietary consumption by social class. Given that a healthy diet not only promotes health but reduces risk of chronic diseases it is possible that price of food contributes to inequalities in health which the UK government is keen to eradicate (Marmot, 2010). It is important, therefore, to understand more about how the cost of food impacts on the consumption of a healthy diet.

This Chapter investigates the cost of dietary patterns, derived by cluster analysis, consumed in the large UK Women's Cohort Study. The dietary patterns reflect both quantity and diversity of food and have been assigned a healthiness score according to how well they adhere to the Department of Health's Eatwell Plate. Diet cost is assigned from a food cost database. The main aim is to show whether there are any differences in cost between a healthy dietary pattern for UK women and a less healthy pattern.

### 4.2.1 Objectives

The objectives of this chapter, as described in Chapter 1, Table 1.1 are as follows:

- To categorise the healthiness of dietary patterns in the UKWCS
- To analyse the cost of dietary patterns in the UKWCS and how this cost is related to healthiness of the diet

## 4.3 Methods

### 4.3.1 Dietary pattern healthiness score

To make better sense of what the dietary patterns (summarised in Chapter 3 section 3.2.1.4) mean with respect to health, they were ranked in order of potential healthiness. To facilitate this, an index system has been developed, based in two indicators of healthy diet: (1) how closely the dietary pattern adheres to the Department of Health's Eatwell plate model and (2) the percentage of women in each dietary pattern adhering to recommended intake for fibre.

The Department of Health's Eatwell Plate is a "pictorial representation of the proportion that different food groups should make to the diet" (Department of Health, 2011). This illustrates the UK dietary guidelines: to consume plenty of starchy products like potatoes, bread, rice and pasta, choosing wholegrains where possible to increase fibre intake; at least 5 portions of fruit and vegetables daily; some high protein foods; some milk and dairy and only a small amount of saturated fat, sugar and salt. Using the contents and quantities of the UKWCS seven dietary patterns (summarised in Chapter 3, Table 3.2) a value, between negative one and plus two, is assigned for how well the dietary pattern achieves each component of the Eatwell Plate.

- A value of negative one is assigned if the dietary pattern falls short of the Eatwell Plate guidance, producing a negative effect on diet quality e.g. not consuming any fruit and vegetables. This value may also be assigned if the pattern exceeds the Eatwell Plate guidance such that it produces a negative effect on diet quality e.g. consuming too much saturated fat.
- A value of one is assigned if the pattern goes someway to meeting the Eatwell Plate guidance e.g. some fruit and vegetables are consumed, but not in excess of 5 portions a day.
- A value of 1.5 is assigned is the pattern meets the guideline, for example 5 portions of fruit and veg a day.
- A value of two is given if the pattern exceeds the Eatwell Plate guidance e.g. more than 5 portions of fruit and vegetables are consumed daily.

The individual component value is then weighted according to the proportion of the plate which that food constitutes.

- Fruit and vegetables = 33.3%
- Starchy foods = 33.3%
- Meat, fish and eggs = 12.5%
- Fatty and sugary foods = 8.4%
- Milk and dairy products = 12.5%

The weighted value for each component are then summed. See Table 4.1.

| Dietary pattern | Fruit and vegetables | Fruit and vegetables with weighting for proportion of plate (33.3%) | Starchy | Starchy weighting for proportion of plate (33.3%) | Meat, fish, eggs | Meat, fish, eggs weighting for proportion of plate (12.5%) | Eatwell Plate High Fat and Sugar | Fat and sugar weighting for proportion of plate (8.4%) | Milk and diary | Milk and dairy weighting for proportion of plate (12.5%) | Weighted value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Monotonous Low Quantity Omnivore** | -1.00 | -33.30 | 1.00 | 33.30 | 1.00 | 12.50 | -1.00 | -8.40 | 1.00 | 12.50 | 16.60 |
| **Traditional Meat, Chips and Pudding Eater** | -1.00 | -33.30 | 1.00 | 33.30 | 1.00 | 12.50 | -1.00 | -8.40 | 1.00 | 12.50 | 16.60 |
| **Conservative Omnivore** | 1.00 | 33.30 | 1.00 | 33.30 | 1.00 | 12.50 | 1.00 | 8.40 | 1.00 | 12.50 | 100.00 |
| **Low Diversity Vegetarian** | 1.00 | 33.30 | 1.00 | 33.30 | -1.00 | -12.50 | 1.00 | 8.40 | 1.00 | 12.50 | 75.00 |
| **Higher Diversity Traditional Omnivore** | 1.50 | 49.95 | 1.50 | 49.95 | 1.00 | 12.50 | 1.00 | 8.40 | 1.00 | 12.50 | 133.30 |
| **High Diversity Vegetarian** | 2.00 | 66.60 | 2.00 | 66.60 | -1.00 | -12.50 | 1.00 | 8.40 | 1.00 | 12.50 | 141.60 |
| **Health Conscious** | 2.00 | 66.60 | 2.00 | 66.60 | 1.00 | 12.50 | 1.00 | 8.40 | 1.00 | 12.50 | 166.60 |

Table 4.1 - Dietary patterns value derived from comparison to the Department of Health’s Eatwell Plate

Assigning values to the Eatwell Plate is somewhat subjective, as the Eatwell Plate does not come with weighed food values. Therefore, it was decided to explore combining the Eatwell Plate values with the percentage of women in each pattern meeting key dietary recommendations. Using the UK Dietary Reference Values (DRV) for energy intake, vitamin C, vitamin A, iron, zinc, fibre, carbohydrate, protein and fat (Committee on Medical Aspects of Food Policy and Department of Health, 1991), it is possible to deduce the percentage of women consuming each dietary pattern who meet these recommendations, as displayed in Table 4.2. There are a number of factors which can influence meeting such recommendations. For example, the recommended daily intake for an average woman in the UK is 2000kcal. This is dependent upon the energy expenditure of the women and also factors such as body size. Therefore, for simplicity, the best method seemed to be to use just one key nutrient which is often used as a proxy measure of healthiness: Fibre. Fibre has been shown to be effective in prevention of many chronic diseases (Threapleton et al., 2013, Cade et al., 2007, Hauner et al., 2012) and can be measured in the UKWCS using the FFQ dietary assessment method.

Using the Eatwell Plate score, combined with percentage of women meeting daily recommended intake for fibre, accounts for foods and a key nutrient whilst reflecting the Department of Health recommendation to consume wholegrains (to increase fibre consumption) where possible.

| Dietary pattern | Average daily kcal intake (SD) | % exceeding 2000 kcal/day | % meeting Vitamin C | % meeting Vitamin A | % meeting Iron | % meeting Zinc | % meeting Fibre (18g/day) | % meeting CHO (50% ) | % meeting Protein (53g/day) | % meeting Fat (<35%) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Monotonous Low Quantity Omnivore** | 1826 (569) | 32 | 100 | 74 | 60 | 73 | 46 | 58 | 81 | 73 |
| **Traditional Meat, Chips and Pudding Eater** | 2484 (657) | 79 | 100 | 97 | 80 | 98 | 72 | 45 | 99 | 57 |
| **Conservative Omnivore** | 1998 (494) | 46 | 100 | 89 | 79 | 87 | 78 | 71 | 91 | 83 |
| **Low Diversity Vegetarian** | 2180 (577) | 58 | 100 | 89 | 73 | 83 | 87 | 71 | 87 | 73 |
| **Higher Diversity Traditional Omnivore** | 2916 (756) | 95 | 100 | 100 | 95 | 100 | 97 | 56 | 100 | 62 |
| **High Diversity Vegetarian** | 2647 (716) | 85 | 100 | 98 | 92 | 96 | 99 | 80 | 96 | 73 |
| **Health Conscious** | 2873 (1046) | 88 | 100 | 99 | 96 | 97 | 99 | 85 | 97 | 86 |

Table 4.2 - Average daily calorie intake (including calories from alcohol) for each dietary pattern, percentage of women whose diet exceeds the daily recommended calorie intake of 2000kcal, along with the percentage of women consuming each dietary pattern whose diet meets the recommended nutrient intake

The Eatwell Plate summed weighted value (range 16.6 to 166.6) was added to the percentage of women meeting the daily fibre intake recommendations of 18g per day for each dietary pattern. The index score was derived according to quantiles of the weighted Eatwell Plate and fibre values (<65, 66-130, 131-195, 196-260 and >261) ensuring that the lowest value was assigned an index score equal to one and the highest equal to five. See Table 4.3.

| Dietary pattern | Healthiness score | Sum of Eatwell weighted value and % meeting fibre recommendations | High quantities | Moderate quantities | Low quantities | Healthiness explanation |
|---|---|---|---|---|---|---|
| **Monotonous Low Quantity Omnivore** | 1 | 62.60 | White bread, milk, sugar | Potatoes, meat | Most other foods | Nutrient poor diet promotes risk of obesity and related co-morbidities. Lacking in fruit and vegetables, with high amounts of sugar. |
| **Traditional Meat, Chips and Pudding Eater** | 2 | 88.60 | White bread, chips, meat, sugar, high-fat and creamy food, biscuits, cakes | Most other foods | Wholemeal food, soya products, vegetables, salad, fruit | An energy dense and nutrient poor diet promotes risk of obesity and related co-morbidities. Whilst this is a more varied diet than the Monotonous Low Quantity Omnivore, there is a limited consumption of healthful foods and too much high fat and sugary foods to match the Eatwell Plate. This does not provide all nutrients for recommended intake. |
| **Conservative Omnivore** | 3 | 178.00 | - | Most food, including potatoes, meat, fish, eggs, fruit, vegetables | Cereals, chips, wholemeal food, nuts, pulses, spreads and dressings, chocolate, crisps, biscuits. Less red meat, less chips and less puddings than the Traditional Meat Chips and Pudding Eater and the Higher Diversity Traditional Omnivore. | While this dietary pattern does not consume large amounts of any foods, it does follow the Eatwell Plate guidelines with lesser quantities. Nutrient intake falls short of the recommendations. |
| **Low Diversity Vegetarian** | 3 | 162.00 | Wholemeal bread, soya products, pulses, fruits (not exotic fruit), vegetables. | Cereals | Butter, eggs, meat, fish | With the exception of meat, fish and eggs this diet is close to the Eatwell Plate recommendations. It however does not meet the daily recommended nutrient intakes. |

| Dietary pattern | Healthiness score | Sum of Eatwell weighted value and % meeting fibre recommendations | High quantities | Moderate quantities | Low quantities | Healthiness explanation |
|---|---|---|---|---|---|---|
| **Higher Diversity Traditional Omnivore** | 4 | 230.30 | Chips, white pasta and rice, high-fat and creamy food, eggs, meat, fish, chocolate, biscuits, crisps. More fish and salad and general diversity than the Traditional Meat Chips and Pudding Eater. | Vegetables, fruit and alcohol. | Less cakes and puddings than the Traditional Meat Chips and Pudding Eater. | This dietary pattern contains good dietary diversity and is close to the Eatwell Plate guidelines. Recommended intakes of nutrients are met. More fruit and vegetables and less high fat food should be consumed to further promote health. |
| **High Diversity Vegetarian** | 4 | 240.60 | Wholemeal bread, cereals, wholemeal pasta and rice, soya products, spreads, nuts, pulses, vegetables, fruit, herbal tea (generally higher consumption of these products that the Low Diversity Vegetarian). | - | White bread, meat, fish | With the exception of meat, fish and eggs this diet is meets the Eatwell Plate recommendations and daily nutrient intakes. The high fibre content is likely associated with reduced obesity, CVD and some cancers. |
| **Health Conscious** | 5 | 265.60 | Bran, potatoes, wholemeal food, yoghurt, low-fat dairy products, pulses, fish, vegetables, salad, fruit | Most other foods | Chips, sugar | Rich in fruit, vegetables and wholemeal food, pulses and fish providing a range of essential nutrients. High fibre containing diet which protects against cardiovascular disease. This type of diet is likely to prevent against certain cancers. This diet meets the Eatwell Plate requirements well. |

Table 4.3 - Dietary pattern contents and suggested healthiness values

### 4.3.2 Cost of foods

The Nutritional Epidemiology Group at the University of Leeds have developed an in-house food cost database, based on the McCance and Widdowson food codes (Holland B, 1991). This database - the diet and nutrition tool for evaluation (DANTE) food cost database - has been evaluated for use at a population level (Timmins et al., 2013). See Chapter 3. A cost of the food can be assigned to dietary intake data in the same manner that nutrients are assigned from contributing food as a proportion of the total, via an in house Microsoft Access database. For example, when the participant records that they consumed white rice once per week on the FFQ, nutrient values, and in this case costs, are assigned equally from two commonly available white rice products - 'white rice, polished' and 'white rice, easy cook'. Therefore 50% of the nutrient composition/cost is assigned from each product to make a single price for rice, accounting for different types. The diet cost database contains prices from 2004 so a correction factor based on 27 food groups from the Consumer Price Index (Office for National Statistics, 2011a) has been applied to bring the prices back in line with those at the time of data collection (1998/99). Using the individuals daily diet cost derived from the FFQ a mean daily cost has been assigned to each of the seven dietary patterns. The mean daily diet cost is an estimated cost, as the actual money spent on food by these women was not collected.

### 4.3.3 Statistical analysis

Stata IC12 statistical software (StataCorp, 2012) has been used to perform the analysis.

A post hoc sample size calculation was carried out which showed that based on the numbers consuming each dietary pattern in the UKWCS, there is 95% power to detect a £0.07 difference in daily diet cost at the 5% significance level between any two of the dietary patterns. Given that the mean daily diet cost for the UKWCS (in 1998/9) was £4.47 this study is powered to detect a difference in cost of 2%.

In order to effectively adjust for confounders and mediators in statistical models, a causal diagram was constructed and is described in Chapter 3, section 3.3.2. Factors adjusted for are described below.

One-way analysis of variance was performed to test for difference between the daily costs of consuming each dietary pattern. The Kruskal-Wallis test was used when the data was non-parametric. Spearman's correlations were used to examine relationships between diet cost and diet quality (according to the Healthiness score) and diet cost and demographic variables, such as education, social class and age. To investigate how well dietary pattern consumption predicts the daily cost of diet, linear regression was used. The ‘Traditional Meat Chips and Pudding Eater’ dietary pattern was used as a reference group as this was the most commonly consumed dietary pattern in the UKWCS, with 18% of the women consuming this dietary pattern. Three models were created; unadjusted (model 1); adjusted for energy intake and physical activity (model 2) and adjusted for age, energy intake, physical activity, smoking, social class and education (model 3). Metabolic Equivalent of Tasks (METs) were used as a measure of physical activity, calculated by assigning a value from the Compendium of Physical Activities (Ainsworth et al., 2000) where the women reported hours per typical week spent in various common activities. Smoking is reported as a binary value which indicates if the woman was a current smoker or not. Total calorie intake (including calories from alcohol) was derived from the FFQ.

## 4.4 Results

### 4.4.1 Dietary pattern analysis

Results show that the Health Conscious dietary pattern, which as its name suggests is the healthiest diet in the UKWCS, has the highest estimated cost associated with it. The Monotonous Low Quantity Omnivore dietary pattern has the lowest estimated cost and this is also the least healthy. The results also show that diversity in a diet comes at a cost, with the estimated cost of the more diverse dietary patterns being  more expensive, for example the Health Conscious and the High Diversity Traditional Omnivore (Table 4.4). The vegetarian dietary patterns contain the highest percentages of women educated to A-level and above, with the High Diversity Vegetarian being more highly educated than the Low Diversity Vegetarian. The Health Conscious dietary pattern contains the next highest percentage of education above A-level, with the lowest percentage belonging to the poorest diet, the Monotonous Low Quantity Omnivore. The pattern is similar when considering occupation, with the highest percentage of women in the professional and managerial category consuming the healthiest diets. Interestingly the vegetarians are also the youngest women in the cohort.

A strong positive correlation exists between estimated diet cost and diet healthiness score (1-5) (Spearman’s rho = 0.6 $t<0.001$). Weak, but significant, positive correlations exist between estimated diet cost and increasing levels of education (Spearman’s rho=0.03 $t<0.001$), social class (Spearman’s rho=0.03 $t<0.001$) and age (Spearman’s rho=0.04 $t<0.001$).

In an unadjusted regression model examining whether the estimated daily cost of diet predicts the dietary pattern consumed, where the most commonly consumed dietary pattern (The

Traditional Meat Chips and Pudding Eater) is the reference, all p values are highly significant (Table 4.5) suggesting that the estimated daily diet cost can predict dietary pattern consumption. The coefficient value in the regression model equates to the estimated daily cost difference in pounds. The Monotonous Low Quantity Omnivore dietary pattern estimated cost is 25% (-£1.10, 95% CI -£1.15 to -£1.06) less per day than the Traditional Meat Chips and Pudding Eater pattern, whilst the Health Conscious dietary pattern estimated cost is most expensive being 51% (£2.24, 95% CI £2.19 to £2.30) more.

The unadjusted regression model (Table 4.5) explains 37% of variation in the model ($R^2$ = 0.37) showing that cost of food contributes to dietary pattern choice. Adjusting for total calorie intake, energy expenditure and age increases the $R^2$ to 0.69, with energy intake contributing most to this increase (Table 4.5). Inclusion of these variables also attenuate the regression coefficients, although all models are still highly statistically significant, showing that the Monotonous Low Quantity Omnivore dietary pattern is still the cheapest, with the estimated cost being 6% (-£0.24, CI -£0.27 to -£0.21) less per day than the Traditional Meat Chips and Pudding Eater, whilst the Health Conscious pattern remains the most expensive being 41% (£1.80, CI £1.76 to £1.84) more expensive per day. An interesting effect is observed in relation to the Conservative Omnivore dietary pattern where the direction of effect is swapped, now showing that the estimated cost of this pattern is in fact 9% (£0.39, CI £0.36 to £0.42) higher per day, where in the unadjusted model it was 6% (-£0.24, CI -£0.28 to -£0.20) per day cheaper. When socioeconomic status, education and smoking status are also added to the model, very little difference in the coefficients is observed.

| Dietary pattern | Diet Healthiness Score (1=lowest and 5=highest) | Mean daily diet cost in £ (95% CI) | Mean calorie intake (95% CI) | Mean Cost per calorie £ (95% CI) | Mean BMI (95% CI) | Median METS (IQR) | Age years (95% CI) | % educated above A level | % with professional/ managerial occupation |
|---|---|---|---|---|---|---|---|---|---|
| **Monotonous low quality omnivore** (n=5331) | 1 | 3.28 (3.26 to 3.31) | 1823 (1808 to 1838) | 0.19 (0.19 to 0.19) | 24.7 (24.6 to 24.9) | 12 (7 to 20) | 53.4 (53.1 to 53.7) | 37.3 | 53.7 |
| **Traditional meat chips and pudding eater** (n=5998) | 2 | 4.39 (4.36 to 4.41) | 2476 (2460 to 2492) | 0.18 (0.18 to 0.18) | 25.1 (25.0 to 25.2) | 14 (9 to 22) | 52.1 (51.9 to 52.4) | 43.9 | 55.8 |
| **Conservative omnivore** (n=5860) | 3 | 4.14 (4.12 to 4.17) | 1995 (1983 to 2008) | 0.21 (0.21 to 0.21) | 24.8 (24.7 to 24.9) | 14 (9 to 21) | 54.5 (54.3 to 54.8) | 48.7 | 61.9 |
| **Low diversity vegetarian** (n=5071) | 3 | 3.92 (3.90 to 3.95) | 2183 (2167 to 2199) | 0.18 (0.18 to 0.18) | 23.4 (23.3 to 23.5) | 13 (8 to 20) | 49.0 (48.8 to 49.3) | 62.5 | 69.0 |
| **Higher diversity traditional omnivore** (n=4733) | 4 | 5.50 (5.46 to 5.53) | 2892 2873 to 2912) | 0.19 (0.19 to 0.19) | 24.9 (24.7 to 25.0) | 16 (11 to 24) | 53.0 (52.7 to 53.2) | 54.5 | 64.2 |
| **High diversity vegetarian** (n=4273) | 4 | 5.01 (4.97 to 5.04) | 2637 (2617 to 2657) | 0.19 (0.19 to 0.19) | 23.2 (23.1 to 23.3) | 16 (10 to 23) | 49.7 (49.4 to 50.0) | 68.6 | 75.2 |
| **Health conscious** (n=2071) | 5 | 6.63 (6.55 to 6.71) | 2809 (2774 to 2843) | 0.24 (0.24 to 0.24) | 24.3 (24.1 to 24.5) | 17 (11 to 26) | 52.7 (52.3 to 53.1) | 57.7 | 71.5 |
| **Chi$^2$: p value** | - | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 |
| **All cohort** (n=33337) | - | 4.47 (4.46 to 4.49) | 2343 (2335 to 2351) | 0.19 (0.19 to 0.20) | 24.4 (24.4 to 24.4) | 14 (9 to 22) | 52.1 (52.0 to 52.2) | 52.3 | 63.2 |

Table 4.4 - Summary statistics for dietary patterns observed in the UKWCS, with energy intake <300 and >6000 kcal/day excluded

| Dietary Pattern | Unadjusted model (1) ($R^2$=0.37) | | Model (2) adjusted for age, energy intake and energy expenditure ($R^2$=0.69) | | Model (3) adjusted for age, energy intake, energy expenditure, smoking, social class and education ($R^2$=0.70) | |
|---|---|---|---|---|---|---|
| | Daily diet cost £ (95% CI) | P value | Daily diet cost £ (95% CI) | P value | Daily diet cost £ (95% CI) | P value |
| **Monotonous Low Quantity Omnivore** | -1.10 (-1.15 to -1.06) | <0.001 | -0.24 (-0.27 to -0.21) | <0.001 | -0.25 (-0.28 to -0.22) | <0.001 |
| **Traditional Meat Chips and Pudding Eater** | | | Reference | | | |
| **Conservative Omnivore** | -0.24 (-0.28 to -0.20) | <0.001 | 0.39 (0.36 to 0.42) | <0.001 | 0.39 (0.36 to 0.43) | <0.001 |
| **Low Diversity Vegetarian** | -0.46 (-0.51 to -0.42) | <0.001 | -0.06 (-0.09 to -0.03) | <0.001 | -0.05 (-0.08 to -0.02) | 0.002 |
| **Higher Diversity Traditional Omnivore** | 1.11 (1.07 to 1.16) | <0.001 | 0.55 (0.52 to 0.58) | <0.001 | 0.57 (0.54 to 0.61) | <0.001 |
| **High Diversity Vegetarian** | 0.62 (0.57 to 0.66) | <0.001 | 0.41 (0.38 to 0.45) | <0.001 | 0.43 (0.40 to 0.47) | <0.001 |
| **Health Conscious** | 2.24 (2.19 to 2.30) | <0.001 | 1.80 (1.76 to 1.84) | <0.001 | 1.81 (1.77 to 1.85) | <0.001 |

Table 4.5 - Regression model investigating the influence of dietary pattern consumption on daily diet cost

## 4.5 Discussion

This research is the first to assign costs to dietary pattern data in the UK. The strong positive correlation observed between the estimated diet cost and diet healthiness is consistent with other studies (Darmon et al., 2004, Drewnowski et al., 2007, Maillot et al., 2007a, Waterlander et al., 2010). Results show that those who have a higher socioeconomic status, indicated by both education and occupation, are also more likely to consume healthier and more expensive diets. The strength of correlation between demographic characteristics; age, education and occupation and the estimated cost of diet are statistically significant despite the homogeneity of the women in this cohort: in that they are typically middle aged and well educated. Healthier, more expensive diets and higher socioeconomic status markers also appear to be associated with increased physical activity levels, illustrated by highest median METS values for these women. It might be hypothesised that the increase in estimated diet cost is therefore due to an increase in total energy intake to balance increased energy expenditure through physical activity. However, controlling for these factors in regression analysis attenuates the difference in estimated daily diet cost between the most healthy and least healthy from £3.34 to £2.05, but a significantly increased cost of a healthier diet remains.

The dietary patterns in this study have been characterised according to both health promoting contents of the diet and the diversity of the diet, both of which contribute to a healthy diet (Drewnowski et al., 1997). The results suggest that both of these factors come at a financial cost. Another study has also observed that cost increases with diversity (Ryden and Hagfors, 2011). The dietary patterns in this study also include an aspect of quantity of the food consumed, as well as variety, defined by the number of different food types consumed in each pattern. This is something which has previously been omitted when considering diversity in diet (Drescher et al., 2007, Monsivais et al., 2012).

An interesting effect was observed relating to the Conservative Omnivore dietary pattern where it becomes more expensive in relation to the Traditional Meat Chips and Pudding Eater in the adjusted regression analysis, compared to being cheaper in the unadjusted analysis. This pattern is high in variety, but foods are consumed in low quantities. One explanation for this change in the direction of the effect could be that by controlling for energy intake the effect of the diversity becomes clearer; supporting the finding that diversity comes at a cost.

*Strengths and Limitations*

As with all studies involving dietary assessments there are limitations. FFQs have been shown to overestimate food intake in the UKWCS (Spence et al., 2002). For the purpose of this analysis though, the overestimation is likely to occur for all foods thus the ranking of the estimated cost of dietary patterns would be unaffected. Dietary assessment by FFQ, while cheap and convenient, is not the gold standard. Repeated 24 hour recall or weighed food diary would provide more reliable dietary data. However, these methods are challenging to deliver to large cohort studies such as the UKWCS. It may be possible in further work to investigate whether the same overestimation is observed with estimated cost of the foods assigned to weighed or recalled intake records. Whilst the FFQ does take into account food which has been eaten outside of the home, it does not differentiate in terms of the price difference of consuming food at home compared to in a restaurant. The DANTE food cost database contains low, medium and high prices for food items, but only an average price has been used in this study. These average prices do not account for regional, supermarket or brand variation in costs. As large savings can be made by purchasing cheaper, generic brands (Chapman et al., 2013), it may be expected brand purchasing would vary by socioeconomic status, so use of average prices may have attenuated differences in the estimated cost of dietary patterns.

Daily diet cost used in this study was the most appropriate method given that the dietary assessment method was a FFQ where weighed food intakes were not recorded. In order to take into account energy intake or expenditure these variables were adjusted for in the regression analysis. If this study had used weighed dietary intake records, cost per calorie or cost per daily recommended intake for a given dietary pattern may have been more appropriate.

The DANTE food cost database contains over 3000 food items. While there are only 217 food items in the FFQ there are many more individual foods contributing to these items (as described by the rice example in section 4.3.2 above). Other studies, for example the Family Food Survey in the UK, use diet cost data derived from till receipt totals over a two week period but no work was done in that study to link the till receipts to actual food consumption. In addition, the total cost per individual was estimated by dividing by the number of individuals in the household. The Family Food Survey also does not account for food consumed outside the home. Given that the DANTE cost estimates are for an individual's food consumption it could be argued to be more accurate.

The DANTE diet cost database was validated using a comparison of diet cost from till receipt collection and from a four day food diary with costs assigned by the database showing that at a population level, the difference was as little as £0.02, which is less than 1% of the mean daily diet cost ((Timmins et al., 2013) and Chapter 3). The costs in this study are also assigned at an individual level and averaged for the dietary patterns, further increasing reliability of the dietary pattern costs.

The UKWCS only includes women. These women were aged 35-69 at recruitment, thus limiting the generalisabilty of these findings. However, due to the large numbers in this study, the results are transferrable to such women throughout the UK.

Seasonal variation can be a problem when considering dietary consumption patterns and diet cost. Due to the phased rollout of recruitment in the UKWCS, and the FFQ assessment method recording frequency of consumption in the last 12 month, this problem is avoided. Dietary patterns identified in this cohort, using a cluster analysis, are derived from what the women actually ate, rather than trying to make their dietary consumption fit a predefined dietary pattern. So while the results are not directly comparable to other dietary pattern research they do reflect true dietary pattern consumption in this population.

The dietary data was collected between 1995 and 1998 in order to examine the relationship between diet and health. This study uses the cost of food from the time at which the data was collected. The food costs were not inflated to bring in line with today's prices. If the food group costs had changed at different rates it may have affected food choice, potentially altering dietary patterns; in which case it would have been incorrect to adjust for inflation to today's prices. Results are presented as a percentage of the mean diet cost to illustrate the proportion of difference, which would be comparable regardless of total cost. Further work will look at how the cost of the dietary pattern is related to the long term health of these women. The cost of these dietary patterns adds strong evidence supporting what is already known about the cost of a healthy diet.

No other study has been able to assigns costs from a cost database - which has been evaluated for use in population studies - to dietary data for such a large sample of women in the UK.

This analysis has the potential to influence public health policy as it conclusively shows that consuming foods which constitute a healthier dietary pattern does is more expensive. In order to promote health and well being in our society, and reduce health inequalities in addition to saving the National Health Service money by preventing future chronic disease incidence, interventions to promote healthy food choices which are accessible and affordable to all should be priority.

## 4.6 Summary and context

Chapter 4 has explored the healthiness of the data driven dietary patterns in the UKWCS, ranking the Health Conscious dietary pattern as most healthy, and the Monotonous Low Quantity Omnivore as least healthy. Costs were then assigned to the diets of individuals in the cohort in order that costs of the seven dietary patterns could be estimated. The analysis showed that the healthiest diets were indeed the most expensive (using average food costs), even once controlling for potential confounders. The estimated cost of the most healthy dietary pattern was over £2.00 per day more than the least healthy, in the fully adjusted model. This agrees with findings elsewhere in the literature.

Next, Chapter 5 will explore the geographical variations in dietary patterns and diet costs.

# Chapter 5: Geography of dietary patterns and diet cost in the UKWCS

## 5.1 Overview

The work in this chapter investigates the spatial variation in dietary patterns and diet cost and then goes on to explore whether it is possible to estimate small area dietary pattern consumption and cost for Leeds using the OAC Groups. As a reminder, this chapter will be using data from the UKWCS, profiled using four spatial scales. Figure 5.1 shows how this chapter fits into the overall thesis flow.

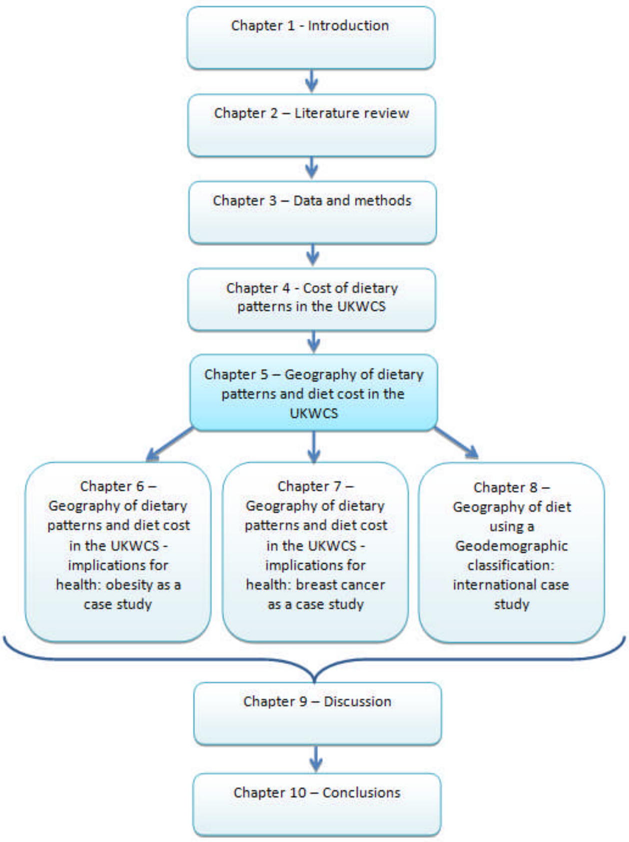



Figure 5.1 - Overview showing how Chapter 5 fits into the overall thesis flow

## 5.2 Introduction

It is well known that dietary consumption can have an impact on long term health in the UK and elsewhere (Doll and Peto, 1981, Key et al., 2002, Foresight, 2007). Diet is a modifiable risk factor in a number of chronic diseases: for example, type II diabetes, coronary heart disease, cardiovascular disease, hypertension and obesity (Foresight, 2007). Diet is a complex phenomenon. A single food is not eaten in isolation, but in combination with others. Interaction between foods can affect how they are absorbed and processed by the body which may subsequently affect health. Analysis of common dietary patterns is therefore particularly relevant to explore links between diet and health. The contents of a dietary pattern can vary widely from the differentiation between two main types of diet; omnivores and vegetarians (Appleby et al., 2002) to patterns such as a Mediterranean diet (Sofi et al., 2008, Trichopoulou and Vasilopoulou, 2000) and data driven patterns specific to a particular population, assigned by a cluster analysis (Greenwood et al., 2000).

Measurement of diet is subject to a range of potential bias including under or over reporting. Collecting information from a sample which is large enough to be generalisable to a given population is also challenging due to temporal and financial constraints. The UK Women's Cohort Study (UKWCS) is a large sample which used a validated food frequency questionnaire to collected diet data for approximately 35000 women (Cade et al., 2004) at baseline. Dietary information was collected in a subset of these women (approximately 14000) five years later. Many dietary components and health outcomes have been investigated using this cohort (Hutchinson et al., 2010, Burley et al., 2010), along with some dietary pattern analysis (Cade et al., 2004, Cade et al., 2011, Cade et al., 2010).

Reports such as the National Diet and Nutrition Survey show regional variations in diet. Whilst the participants in this survey are recruited such that the sample is deemed geographically representative, the sample size is limited with 3073 individuals from years one, two and three of the rolling programme combined (Department of Health and Food Standards Agency, 2012). Using the Government Office Region (GOR), a large geographical unit which is home to between two and nine million individuals, to report dietary consumption means that substantial generalisation is made. Often literature will refer to a North-South divide in health (Hacking et al., 2011). Another common scale for analysis is a Urban Rural classification (Smith et al., 2010). Some reports also use a geodemographic classification to present dietary habits, for example, the Family Food Survey (Defra, 2009).

Influences on dietary patterns span a broad spectrum entwining social, economic, demographic, environmental and individual factors (Swinburn et al., 2011, Monsivais et al., 2010, Drewnowski et al., 2007, Cummins, 2005a, Wrigley et al., 2002). An ecological model framework to investigate how individuals and their environments interact is well documented, and described in Chapter 1 of this thesis (Olsen, 2009, Bronfenbrenner, 1977, Gibney, 2004) however, unpicking such complex relationships is challenging. Geodemographic classifications combine a small spatial scale geography (Output Area) combined with demographic characteristics taken from the national census and sometimes lifestyle surveys, to create profiles for similar types of areas which exist all over the country (CACI Ltd, Vickers et al., 2005). Such types of classifications are widely used in marketing, but used to a lesser extent in health (Abbas et al., 2009). The use of a geodemographic classification goes some way towards accounting for social (compositional) and environmental (context) interactions by grouping people with similar demographic and neighbourhood characteristics residing in small geographical units together. Applying such a classification to a large cohort such as the UKWCS, to describe dietary patterns could provide some important information about the type of people, living in types of neighbourhoods, consuming different dietary patterns which could be linked to spatial variations in future health outcomes, generalisable to women in the UK population. Geodemographic classifications have also been used in diet research (Hughes et al., 2012, Edwards et al., 2009) but not as a method to predict small area diet patterns, so this will be a first.

If a better understanding of dietary habits in certain groups of people can be developed then there is the potential to provide dietary interventions which could benefit many in terms of health and wellbeing in addition to prevention of chronic diseases.

The over-arching purpose of this chapter is to explore dietary pattern according to different geographical scales. Some evaluation of these methods and their relevance in translating results to public health will be carried out.

### 5.2.1 Objectives

The objectives for this Chapter, as outlined in Chapter 1, Table 1.1 are to:

- Explore variations in dietary patterns across the UK according to four spatial scales: at GOR level, by North-South divide, by Urban/Rural classification and using geodemographic Supergroups (which are created at Output Area level).
- Explore variations in diet cost across the UK according to four spatial scales: at GOR level, by North-South divide, by Urban/Rural classification and using geodemographic Supergroups.
- Use the results generated, by geodemographic classification, to estimate small area dietary patterns and diet cost for a large UK city.
- Consider the application of results at different spatial scales in a public health context.

## 5.3 Methods

### 5.3.1 Spatial scale

Four spatial scales are used for analysis in this chapter. Please refer to Chapter 3, section 3.2.2 for further explanation. Firstly this Chapter reports dietary pattern according to the nine government office regions (GOR) of England and Scotland and Wales as entire countries, and then diet cost by the same spatial measures. Secondly, an analysis of North and South of England is performed (Scotland and Wales are not included in this analysis). The third spatial scale is an Urban/Rural classification, which is applied to all regions of England, Wales and Scotland. For the final spatial method the Output Area Classification (OAC) (Vickers et al., 2005) is used for the geodemographic analysis, which is also applied to all regions of England, Wales and Scotland.

### 5.3.2 Statistical analysis

Stata IC12 statistical software has been used for the analysis (StataCorp, 2012). $Chi^2$ statistics are used to test for difference between groups in cross tabulated data.

*Comparison to the UK population*

In order to assess how the UKWCS women compare to the female population of the UK, population figures are presented for each spatial scale.

*Dietary Pattern Analysis*

Descriptive statistics were carried out for each spatial scale. $Chi^2$ and t-test, tested for difference between the spatial scales ($chi^2$ for GOR and OAC and t-tests for North South and Urban/ Rural) and p values presented. Maps are used to present variation in dietary pattern by GOR and North South and bar charts present differences according to Urban/Rural and OAC Supergroups.

Logistic regression was carried out to test whether spatial measures predict dietary pattern outcomes. The regression analysis looked at whether place of residence predicts the likelihood of consuming a particular dietary pattern compared to the reference category - the Traditional Meat Chips and Pudding Eater pattern (the most commonly consumed in the UKWCS). In the GOR models, living in the South East was the reference category as this region is home to the largest proportion of UKWCS women. The OAC analysis used the Typical Traits Supergroup as reference category as those residing in these areas are most average with respect to the demographic variables. For North South analysis the South is the reference category and Urban is the reference category for Urban/Rural analysis, as these categories contain the larger numbers of women. Results are presented as Relative Risk Ratios with standard errors and p values.

A Causal Diagram was used to choose relevant confounders and mediators to adjust for in the regression models (see Chapter 3, section 3.3.2). The adjusted regression model includes physical activity, smoking, total calorie intake including energy from alcohol, age, social class and education. Metabolic Equivalent of Tasks (METs) were used as a measure of physical activity, calculated by assigning a value from the Compendium of Physical Activities (Ainsworth et al., 2000) to the results of questions asked at baseline where the women reported hours per typical week spent in various common activities. Smoking is reported as a binary value which indicates if the woman was a current smoker at baseline. Total calorie intake, including calories from alcohol, is derived from the FFQ. Age, social class (derived from occupation) and highest education level were collected in the UKWCS baseline lifestyle questionnaire.

*Diet cost analysis*

Mean daily diet cost of the women is presented for each spatial scale with confidence intervals. Linear regression analysis is used to investigate whether spatial measures: GOR, North South, Urban/ Rural and OAC, predict diet costs. Model 1 is unadjusted analysis. Model 2 adjusts for total calorie intake, physical activity, smoking, social class, education level and age. Model 3 also adjusts for dietary pattern. For the models investigating OAC as a predictor of diet cost, social class, education level and age were not included in the model as they are components of the OAC classification.

### 5.3.3 Estimated dietary patterns

The percentages of women within each OAC Group (the second layer of the OAC hierarchy) consuming each of the seven dietary patterns were calculated and described for the UKWCS.

Assuming that dietary patterns by OAC Group in the cohort are typical for all women in these Groups throughout the UK, due to their similar demographic characteristics, dietary profiles from the UKWCS were used to estimate dietary patterns of women in a specific UK city. Dietary pattern and diet cost values were assigned to the equivalent OAC Groups in the Leeds area. This pattern could be applied to any area, but Leeds was selected as this is an ‘average’ UK city, which has been used in other studies relating to geography and health (Tomintz et al., 2008, Edwards et al., 2009). A map specific to consumption of each dietary pattern and diet cost has been produced for Leeds.

## 5.4 Results

### 5.4.1 Regional analysis

The distribution of the UKWCS as a percentage of the total population, by region, ranges from 0.04% for the North East, North West and Scotland and 0.08% for the South East and South West (with other regions falling in-between) (Table 5.1). So whilst the lowest number of cohort women reside in the North East, this is also the region which has the lowest population.

| GOR | UK population 2001 - Females (%) | UKWCS population at baseline (%) | UKWCS population as a % of UK females |
|---|---|---|---|
| North East | 1,296,863 (4) | 974 (3) | 0.08 |
| North West | 3,470,810 (12) | 3038 (9) | 0.09 |
| Yorkshire and the Humber | 2,552,889 (9) | 2561 (8) | 0.10 |
| East Midlands | 2,123,316 (7) | 2405 (7) | 0.11 |
| West Midlands | 2,692,197 (9) | 2534 (8) | 0.09 |
| East of England | 2,749,805 (9) | 3001 (9) | 0.11 |
| Greater London | 3,703,298 (13) | 3709 (11) | 0.10 |
| South East | 4,095,490 (14) | 6789 (21) | 0.17 |
| South West | 2,532,019 (9) | 4155 (13) | 0.16 |
| Scotland | 2,629,517 (9) | 2199 (7) | 0.08 |
| Wales | 1,499,303 (5) | 1419 (4) | 0.09 |
| Total | 29,345,507 | 32,784 | 0.11 |

Table 5.1 - UKWCS population compared to the female population of the UK by GOR (Office for National Statistics, 2001) (General Register Office for Scotland, 2001)

Variation in the age of the UKWCS women by GOR varies by levels approaching statistical significance. Physical activity, total calorie intake, occupation and education level all vary significantly (Table 5.2). Highest physical activity levels are in the South West (17.7 METS), whilst highest daily calorie intake (2482 kcal/day) occurs in Scotland. London and Scotland have the highest percentage of women employed in professional or managerial occupation (67%). Scotland also has the highest level of women educated above A-level (65%).

| | **Mean Age (years)** | **Physical activity (METS)** | **Total calorie intake/day** | **% professional/managerial occupation** | **% educated A level and above** |
|---|---|---|---|---|---|
| North East | 51.5 | 17.3 | 2452 | 65 | 52 |
| North West | 52.2 | 16.4 | 2373 | 63 | 51 |
| Yorkshire and the Humber | 52.0 | 17.1 | 2383 | 64 | 53 |
| East Midlands | 51.9 | 17.0 | 2382 | 65 | 49 |
| West Midlands | 52.3 | 16.6 | 2336 | 62 | 52 |
| East of England | 51.9 | 17.3 | 2335 | 58 | 46 |
| Greater London | 51.4 | 15.3 | 2259 | 67 | 61 |
| South East | 52.4 | 17.1 | 2326 | 61 | 49 |
| South West | 52.89 | 17.8 | 2366 | 62 | 50 |
| Scotland | 52.1 | 17.0 | 2482 | 67 | 65 |
| Wales | 52.4 | 17.3 | 2395 | 66 | 53 |
| $Chi^2$ (p) | 0.02 | <0.001 | <0.001 | <0.001 | <0.001 |
| Total | 52.1 | 14 | 2343 | 63.2 | 52.3 |

Table 5.2 - Descriptive statistics of UKWCS according to the GOR in which they live

Consumption of dietary patterns, with the exception of the Health Conscious dietary pattern, vary by more than just chance (p<0.01) between regions, measured using $chi^2$ statistic (Figure 5.2), with the highest consumption of Traditional Meat Chips and Pudding Eaters in the North East and highest proportion of High and Low Diversity Vegetarians in Greater London.

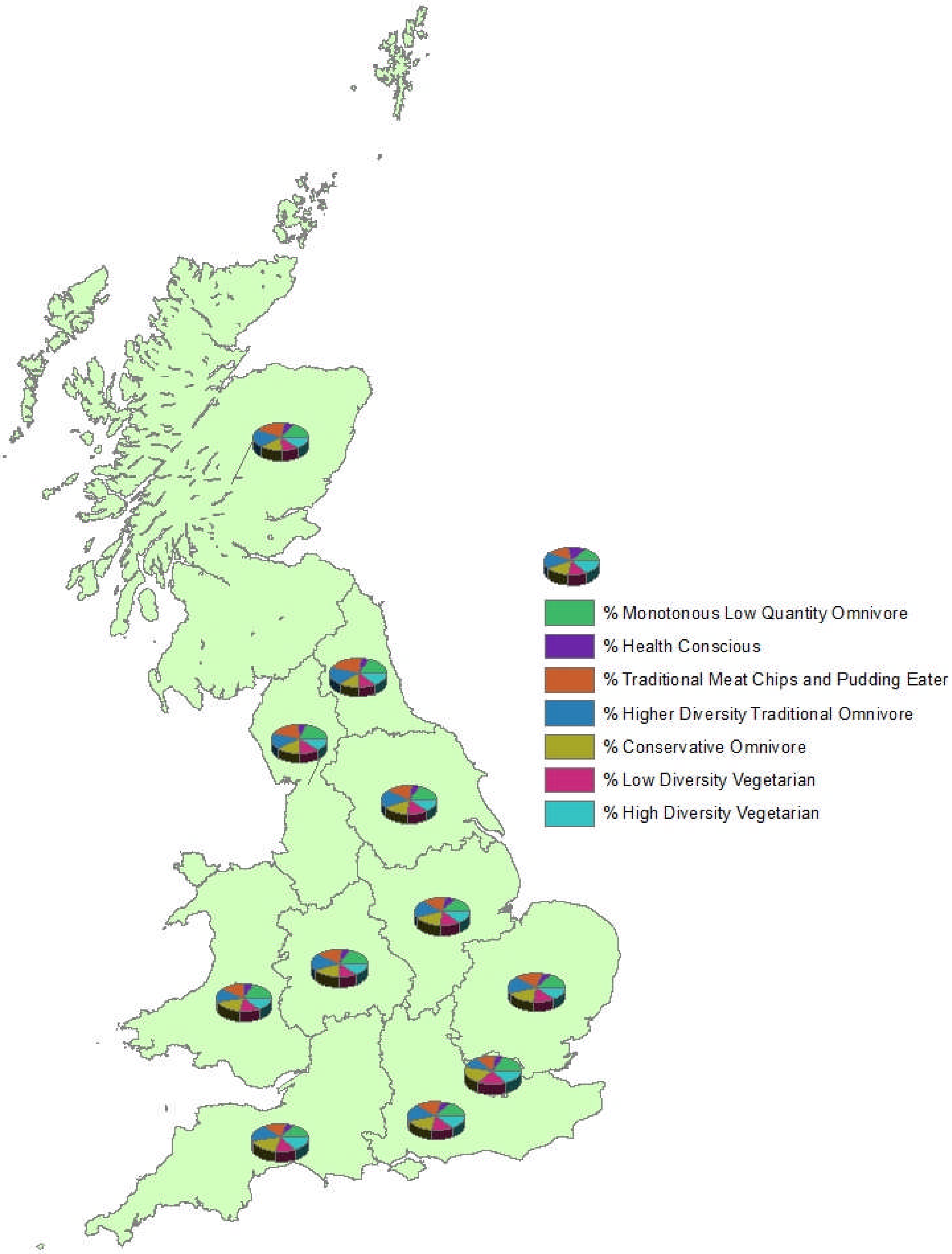



Figure 5.2 - Percentage of UKWCS consuming each dietary pattern by GOR

Results from the unadjusted multinomial regression analysis, for dietary pattern consumption by region are presented in Table 5.3. These identify that Greater London is most likely to have dietary patterns which differ from the majority with significantly increased RRR of consuming the least healthy Monotonous Low Quantity Omnivore pattern (RRR 1.42) but also the healthiest Health Conscious dietary pattern (RRR 1.38). Greater London also has significantly higher RRR of consuming the vegetarian dietary patterns (RRR 1.77 and 1.64 for the Low and High Diversity Vegetarian patterns respectively). Scotland and the North East are significantly less likely to consume the Conservative Omnivore (RRR 0.62) and Low Diversity Vegetarian (RRR 0.64) dietary patterns whilst those in the North West have increased RRR of consuming the poorest quality dietary pattern, the Monotonous Low Quantity Omnivore (RRR 1.29).

Greater London highlights how one region can exhibit statistical significance in RRR of consuming both a Health Conscious and also of a Monotonous Low Quantity Omnivore diet showing that both extremes of dietary patterns reside in the same region.

These observations remain true when the model is adjusted for total calorie intake, physical activity and smoking and for demographic characteristics; age, social class and education and show much the same results, in some cases slightly accentuated (Table 5.3). The North West, Yorkshire and the Humber, Greater London and Wales have higher RRR of consuming the least healthy Monotonous Low Quantity Omnivore dietary pattern (RRR 1.48, 1.27, 1.23 and 1.49 respectively) while only Greater London is more likely to also consume the healthiest, Health Conscious, pattern. Greater London see significantly increased RRR of the vegetarian dietary patterns (RRR 1.46 and 1.45 for Low Diversity and High Diversity Vegetarians respectively), while Scotland have significantly decreased RRR of these patterns (RRR 0.63 and 0.68 for Low Diversity and High Diversity Vegetarians respectively).

The pseudo $R^2$ value shows that the adjusted model explains 12% of variation in dietary pattern, compared to less than 1% in the unadjusted model. Most of this variation is explained by the energy intake adjustment.

| | Monotonous Low Quantity Omnivore | Health Conscious | Traditional Meat Chips and Pudding Eater | Higher Diversity Traditional Omnivore | Conservative Omnivore | Low Diversity Vegetarian | High Diversity Vegetarian |
|---|---|---|---|---|---|---|---|
| | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** |
| Unadjusted model (pseudo $R^2$=0.003 ) | | | | | | | |
| North East | 0.87 (0.70 to 1.08) p=0.211 | 0.82 (0.61 to 1.12) p=0.218 | 1.00 | 0.79 (0.63 to 0.99) p=0.043 | **0.62 (0.49 to 0.77) p<0.001** | **0.64 (0.51 to 0.81) p<0.001** | 0.84 (0.67 to 1.06) p=0.149 |
| North West | 1.20 (1.04 to 1.38) p=0.013 | 0.91 (0.75 to 1.12) p=0.368 | 1.00 | 0.81 (0.69 to 0.94) p=0.006 | 0.85 (0.73 to 0.98) p=0.023 | 0.84 (0.73 to 0.98) p=0.025 | 0.83 (0.71 to 0.98) p=0.024 |
| Yorkshire and the Humber | 1.11 (0.95 to 1.30) p=0.179 | 0.97 (0.78 to 1.20) p=0.776 | 1.00 | 1.07 (0.92 to 1.25) p=0.392 | 0.96 (0.83 to 1.12) p=0.646 | 0.90 (0.77 1.06) p=0.199 | 0.99 (0.84 to 1.17) p=0.873 |
| East Midlands | 1.01 (0.85 to 1.18) p=0.944 | 1.15 (0.93 to 1.42) p=0.209 | 1.00 | 1.07 (0.91 to 1.25) p=0.444 | 1.13 (0.97 to 1.32) p=0.123 | 0.92 (0.78 to 1.09) p=0.340 | 1.00 (0.84 to 1.19) p=0.997 |
| West Midlands | 1.07 (0.92 to 1.25) p=0.384 | 1.05 (0.85 to 1.30) p=0.652 | 1.00 | 0.88 (0.75 to 1.04) p=0.135 | 1.03 (0.89 to 1.19) p=0.713 | 0.82 (0.70 to 0.97) p=0.019 | 0.96 (0.81 to 1.13) p=0.610 |
| East of England | 0.89 (0.76 to 1.03) p=0.115 | 0.92 (0.75 to 1.12) p=0.407 | 1.00 | 0.94 (0.81 to 1.09) p=0.400 | 0.98 (0.85 to 1.12) p=0.736 | 0.89 (0.77 to 1.03) p=0.124 | 0.89 (0.76 to 1.05) p=0.169 |
| Greater London | **1.42 (1.23 to 1.64) p<0.001** | **1.38 (1.14 to 1.67) p=0.001** | 1.00 | 0.89 (0.76 to 1.05) p=0.165 | **1.28 (1.12 to 1.48) p<0.001** | **1.77 (1.54 to 2.04) p<0.001** | **1.64 (1.41 to 1.90) p<0.001** |
| South East | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| South West | 1.01 (0.89 to 1.16) p=0.837 | 1.15 (0.96 to 1.38) p=0.120 | 1.00 | 1.11 (0.97 to 1.27) p=0.142 | 1.04 (0.92 to 1.19) p=0.528 | 0.99 (0.86 to 1.13) p=0.847 | 1.20 (1.04 to 1.38) p=0.011 |
| Scotland | 0.84 (0.71 to 0.99) p=0.039 | 1.11 (0.90 to 1.37) p=0.329 | 1.00 | 0.97 (0.82 to 1.14) p=0.688 | **0.77 (0.65 to 0.90) p=0.001** | **0.66 (0.56 to 0.78) p<0.001** | 0.84 (0.70 to 1.00) p=0.046 |
| Wales | 1.25 (1.03 to 1.51) p=0.027 | 1.19 (0.92 to 1.55) p=0.187 | 1.00 | 0.89 (0.72 to 1.10) p=0.300 | 1.08 (0.89 to 1.31) p=0.417 | 0.98 (0.80 to 1.20) p=0.817 | 1.11 (0.90 to 1.37) p=0.327 |

*Table continued on page 133*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Adjusted model (adjusting for smoking, total calorie intake including alcohol, typical daily physical activity (METs), age, social class, education) (pseudo $R^2$= 0.12)** | | | | | | | |
| North East | 1.14 (0.89 to 1.47) p=0.303 | 0.77 (0.56 to 1.06) p=0.107 | 1.00 | 0.71 (0.56 to 91) p=0.006 | 0.71 (0.55 to 0.91) p=0.007 | 0.67 (0.53 to 0.86) p=0.002 | 0.74 (0.58 to 0.95) p=0.019 |
| North West | 1.48 (1.25 to 1.74) p<0.001 | 0.91 (0.73 to 1.12) p=0.368 | 1.00 | 0.76 (0.64 to 0.90) p=0.001 | 0.95 (0.81 to 1.11) p=0.525 | 0.90 (0.77 to 1.06) p=0.206 | 0.81 (0.68 to 0.96) p=0.016 |
| Yorkshire and the Humber | 1.27 (1.06 to 1.53) p=0.009 | 0.95 (0.75 to 1.20) p=0.676 | 1.00 | 1.03 (0.87 to 1.23) p=0.712 | 1.05 (0.89 to 1.25) p=0.571 | 0.92 (0.77 to 1.09) p=0.323 | 0.96 (0.80 to 1.15) p=0.636 |
| East Midlands | 1.09 (0.90 to 1.32) p=0.363 | 1.10 (0.87 to 1.38) p=0.429 | 1.00 | 1.07 (0.90 to 1.28) p=0.433 | 1.19 (1.00 to 1.41) p=0.046 | 0.95 (0.80 to 1.14) p=0.590 | 0.96 (0.80 to 1.15) p=0.658 |
| West Midlands | 1.16 (0.97 to 1.38) p=0.105 | 1.02 (0.81 to 1.27) p=0.879 | 1.00 | 0.86 (0.72 to 1.03) p=0.094 | 1.01 (0.85 to 1.19) p=0.949 | 0.79 (0.66 to 0.94) p=0.008 | 0.91 (0.76 to 1.09) p=0.316 |
| East of England | 0.90 (0.76 to 1.07) p=0.229 | 0.96 (0.77 to 1.20) p=0.721 | 1.00 | 1.00 (0.85 to 1.17) p=0.987 | 0.98 (0.84 to 1.14) p=0.764 | 0.84 (0.07 to 0.99) p=0.038 | 0.91 (0.77 to 1.08) p=0.270 |
| Greater London | 1.23 (1.04 to 1.45) p=0.015 | 1.37 (1.11 to 1.68) p=0.003 | 1.00 | 0.89 (0.75 to 1.05) p=0.178 | 1.14 (0.97 to 1.33) p=0.107 | 1.46 (1.26 to 1.70) p<0.001 | 1.45 (1.23 to 1.70) p<0.001 |
| South East | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| South West | 1.07 (0.92 to 1.25) p=0.392 | 1.09 (0.90 to 1.32) p=0.381 | 1.00 | 1.11 (0.96 to 1.29) p=0.157 | 1.01 (0.87 to 1.17) p=0.909 | 0.99 (0.85 to 1.14) p=0.860 | 1.18 (1.01 to 1.37) p=0.032 |
| Scotland | 1.06 (0.88 to 1.28) p=0.550 | 0.89 (0.71 to 1.11) p=0.305 | 1.00 | 0.80 (0.67 to 0.96) p=0.014 | 0.83 (0.70 to 0.99) p=0.044 | 0.63 (0.53 to 0.76) p<0.001 | 0.68 (0.56 to 0.81) p<0.001 |
| Wales | 1.49 (1.31 to 1.69) p=0.002 | 1.06 (0.80 to 1.41) p=0.670 | 1.00 | 0.77 (0.61 to 0.97) p=0.026 | 1.15 (0.93 to 1.43) p=0.194 | 1.00 (0.81 to 1.25) p=0.988 | 0.97 (0.78 to 1.22) p=0.823 |

Table 5.3 - Regression models investigating whether GOR predicts dietary patterns displaying RRR (95% CI) and p value

### 5.4.2 North - South analysis

| | UK women | UKWCS |
|---|---|---|
| The North | 48% (n=12136075) | 39% (n=11512) |
| The South | 52% (n=13080612) | 61% (n=17654) |

Table 5.4 - UKWCS population compared to the population of UK women by North South

The distribution of the UKWCS women compared to the women in the UK is presented in Table 5.4. In the North of England (Scotland and Wales are not included in this spatial measure) the UKWCS women do not vary significantly by age, physical activity level or education. However, they do consume more calories per day and a higher percentage work in a professional or managerial occupation (Table 5.5).

| | Age (years) | Physical activity (METS) | Total calorie intake | % professional/managerial occupation | % educated A level and above |
|---|---|---|---|---|---|
| The North (n=11512) | 52.07 | 16.80 | 2376 | 64 | 52 |
| The South (n=17654) | 52.19 | 16.91 | 2310 | 62 | 51 |
| $Chi^2$ & t-test$^t$ (p) | 0.2907$^t$ | 0.4451$^t$ | <0.001$^t$ | 0.009 | 0.152 |

Table 5.5 - North South descriptive statistics

As illustrated in Figure 5.3, the proportion of each dietary pattern consumed in the North compared to the South varies. The Monotonous Low Quantity Omnivore and Traditional Meat Chips and Pudding Eater are significantly more likely to be consumed n the North than the South (t test p<0.001). Conversely, both vegetarian dietary patterns are significantly more likely to be consumed in the South (t test p<0.001). The other patterns do not vary significantly between North and South.

In unadjusted regression analysis, where the Traditional Meat Chips and Pudding Eater is the reference category, the North is significantly less likely to consume the Conservative Omnivore or Vegetarian dietary patterns. Once the model is adjusted for smoking, total calorie intake, physical activity, age, social class and education, results show that the North is significantly more likely to consume the least healthy Monotonous Low Quantity Omnivore dietary pattern, and significantly less likely to consume the healthiest Health Conscious, Higher Diversity Traditional Omnivore, Low Diversity Vegetarian and High Diversity Vegetarian dietary patterns. No difference is observed with the Conservative Omnivore pattern (see Table 5.6).

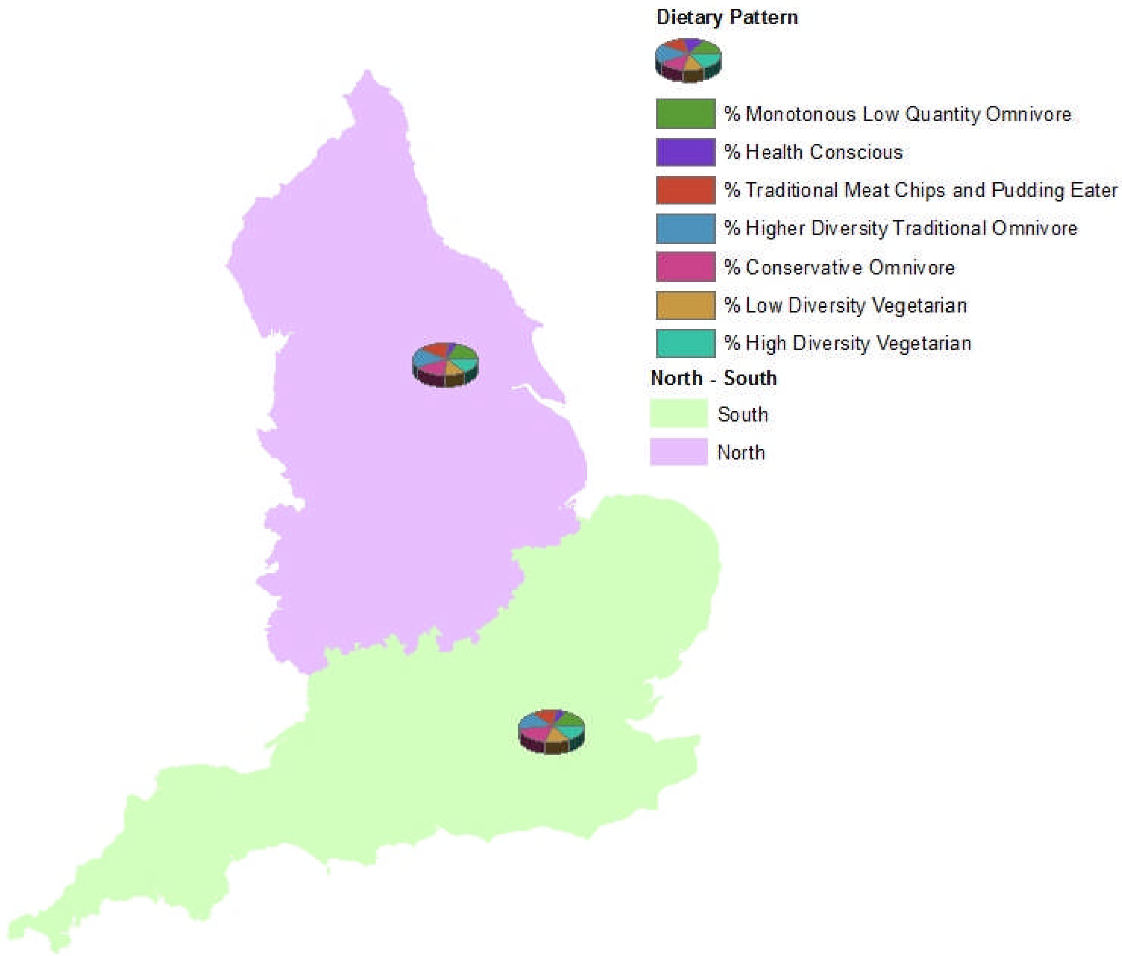



Figure 5.3 - Percentage of UKWCS consuming each dietary pattern in the North and South of England

| | Monotonous Low Quantity Omnivore | Health Conscious | Traditional Meat Chips and Pudding Eater | Higher Diversity Traditional Omnivore | Conservative Omnivore | Low Diversity Vegetarian | High Diversity Vegetarian |
|---|---|---|---|---|---|---|---|
| | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **Ref.** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** |
| Unadjusted model (pseudo $R^2$= 0.0006) | | | | | | | |
| North | 1.03 (0.95 to 1.11) p=0.514 | 0.91 (0.82 to 1.02) p=0.107 | 1.00 | 0.93 (0.86 to 1.01) p=0.102 | **0.90 (0.83 to 0.97) p=0.006** | **0.77 (0.71 to 0.83) p<0.001** | **0.82 (0.75 to 0.89) p<0.001** |
| South | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Adjusted model (Adjusted for smoking, total calorie intake including alcohol, typical daily physical activity (METs), age, social class, education) (pseudo $R^2$= 0.1080) | | | | | | | |
| North | **1.21 (1.10 to 1.32) p<0.001** | 0.89 (0.79 to 1.00) p=0.045 | 1.00 | 0.88 (0.81 to 0.97) p=0.006 | 0.99 (0.90 to 1.08) p=0.756 | **0.82 (0.75 to 0.90) p<0.001** | **0.81 (0.73 to 0.88) p<0.001** |
| South | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 5.6 - Regression models investigating whether living in the north or south of England predicts dietary patterns displaying Relative Risk Ratio (RRR) (95% confidence interval) and p value

### 5.4.3 Urban - Rural analysis

In the UKWCS 67% of the women live in an urban area. In the UK general population (women only) 80% live in an urban area (Office for National Statistics, 2011b). This means that there is some underrepresentation of urban areas and overrepresentation of women residing in rural areas within the UKWCS.

Characteristics of the UKWCS vary significantly according to whether they live in a Rural or Urban area (Table 5.7). With regards to difference in dietary pattern consumption (Figure 5.4), all differences are statistically significant (p<0.001), using a t-test, except for the Traditional Meat Chips and Pudding Eaters and High Diversity Vegetarians which show no difference at all between Urban and Rural areas.

Unadjusted regression models (Table 5.8) show that women livening in Rural areas are significantly more likely to consume Health Conscious, Higher Diversity Traditional Omnivore and Conservative Omnivore dietary patterns, compared to Urban Areas. They are also significantly less likely to consume Monotonous Low Quantity Omnivore and Low Diversity Vegetarian dietary patterns. No difference is observed for the High Diversity Vegetarian pattern. These relationships hold true in the adjusted model.

| | **Mean Age (years)** | **Physical activity (METS)** | **Total calorie intake/day** | **% professional/managerial occupation** | **% educated A level and above** |
|---|---|---|---|---|---|
| Urban (n=21913) | 52.05 | 16.34 | 2334 | 62 | 51 |
| Rural (n=10867) | 52.37 | 17.97 | 2396 | 66 | 54 |
| $Chi^2$ & T test$^t$ (p) | 0.003 $^t$ | <0.001 $^t$ | <0.001 $^t$ | <0.001 | <0.001 |

Table 5.7 - Urban Rural descriptive statistics

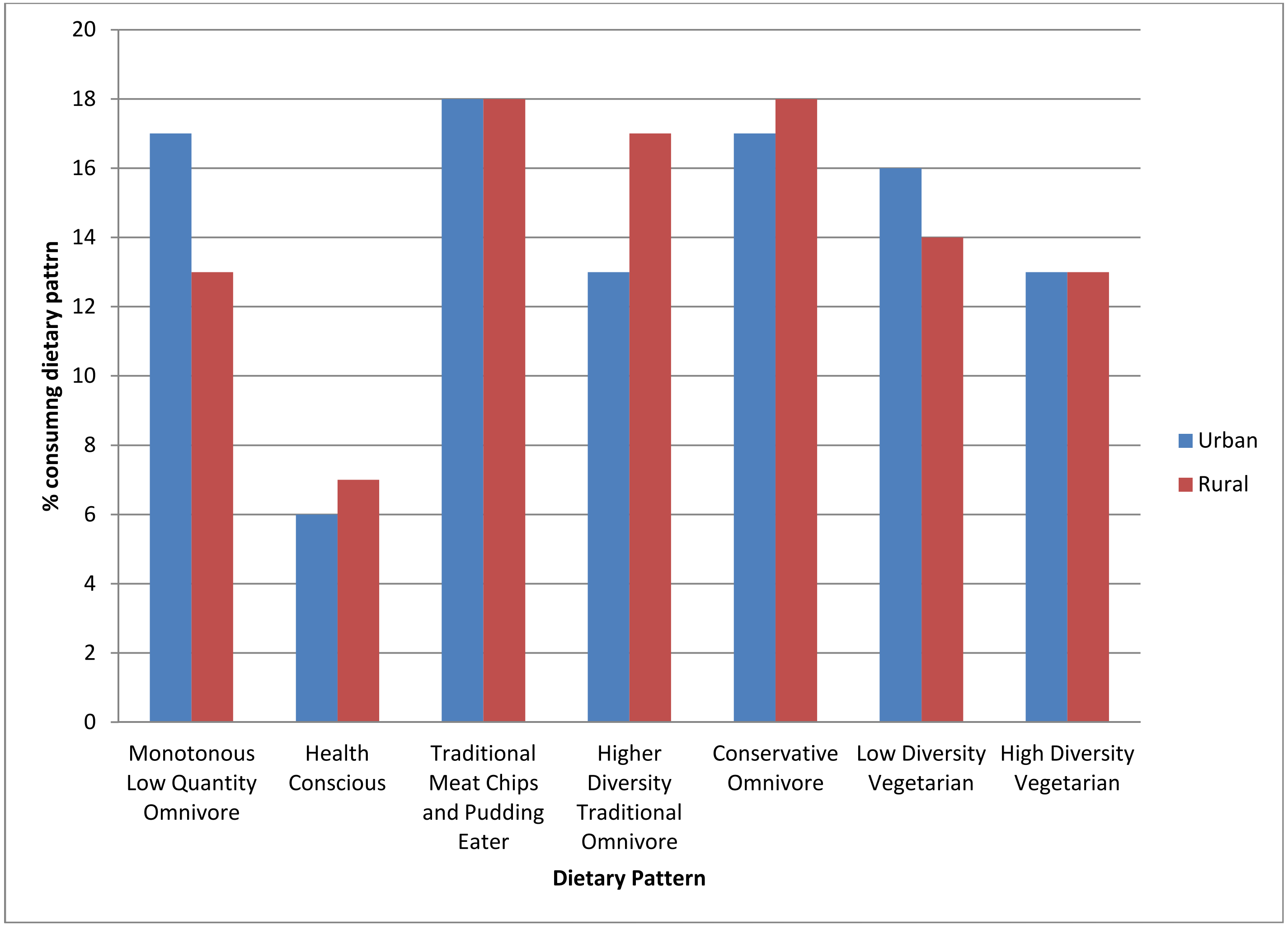



Figure 5.4 - Percentage of UKWCS women consuming each dietary pattern in Urban and Rural areas

| | Monotonous Low Quantity Omnivore | Health Conscious | Traditional Meat Chips and Pudding Eater | Higher Diversity Traditional Omnivore | Conservative Omnivore | Low Diversity Vegetarian | High Diversity Vegetarian |
|---|---|---|---|---|---|---|---|
| | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **Ref.** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** |
| Unadjusted model (pseudo $R^2$=0.0019) | | | | | | | |
| Urban | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Rural | **0.76 (0.70 to 0.82) p<0.001** | **1.27 (1.14 to 1.41) p<0.001** | 1.00 | **1.33 (1.23 to 1.44) p<0.001** | **1.11 (1.03 to 1.20) p=0.006** | **0.87 (0.80 to 0.94) p=0.001** | 1.05 (0.96 to 1.14) p=0.302 |
| Adjusted model (Adjusted for smoking, total calorie intake including alcohol, typical daily physical activity (METs), age, social class, education) (pseudo $R^2$=0.1092) | | | | | | | |
| Urban | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Rural | **0.79 (0.72 to 0.87) p<0.001** | **1.17 (1.05 to 1.31) p=0.006** | 1.00 | **1.29 (1.18 to 1.40) p<0.001** | **1.12 (0.03 to 1.22) p=0.008** | **0.84 (0.77 to 0.92) p<0.001** | 0.98 (0.90 to 1.07) p=0.680 |

Table 5.8 - Regression models investigating whether urban or rural environment predicts dietary patterns displaying relative risk ratio (95% confidence interval) and p value

### 5.4.4 Geodemographic analysis

The UKWCS comprises women from each of the seven OAC Supergroups, with some over-representation in the Prospering Suburbs and Countryside and under-representation in the Constrained by Circumstance, Blue Collar Communities and Multicultural Supergroups. This is as expected in a predominantly middle class cohort of women. That said, there are still large numbers in each of the Supergroups (Figure 5.5).

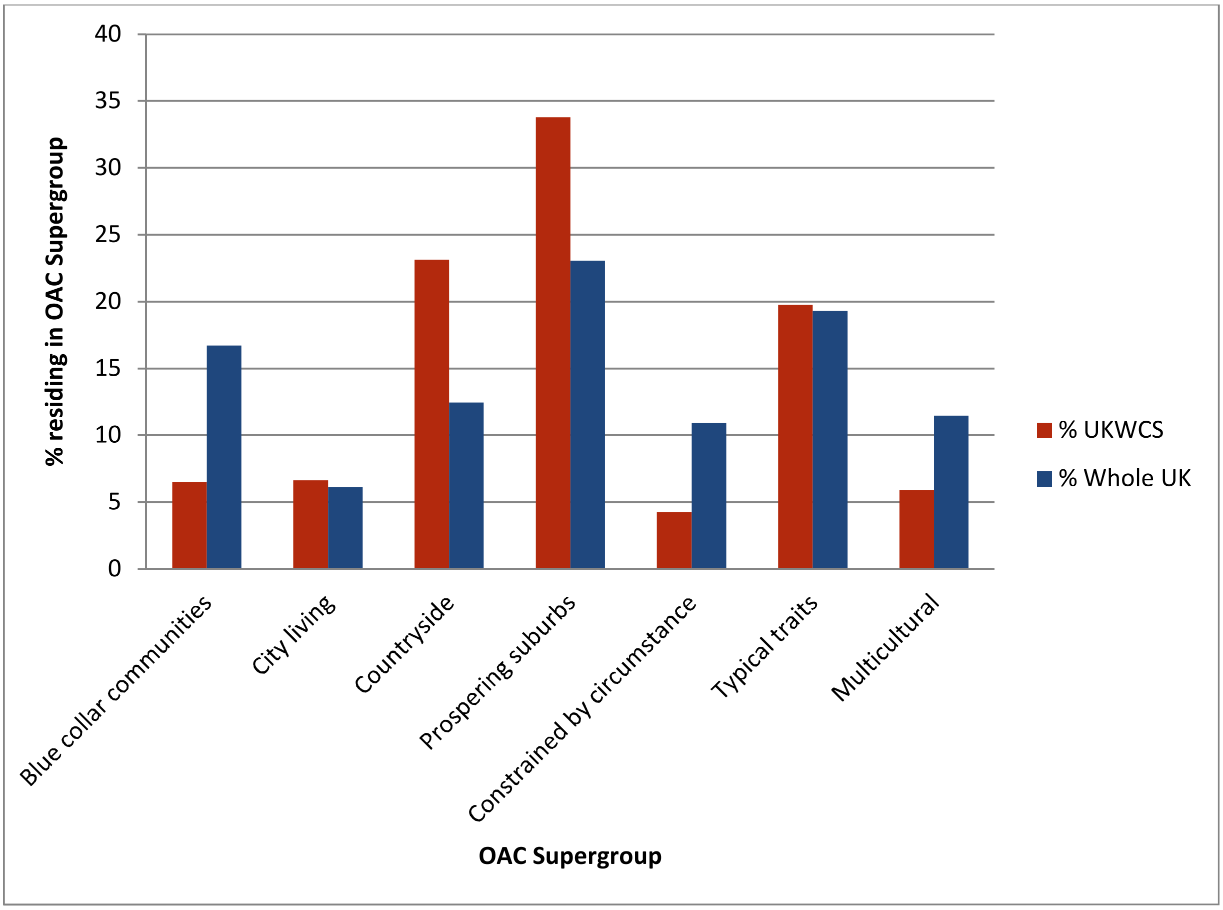



Figure 5.5 - UKWCS compared to the UK population, by OAC Supergroup

Variations in age, physical activity, calorie intake, occupation and education across OAC Supergroups are all statistically significant. See Table 5.9.

| | Mean Age (years) | Physical activity (METS) | Total calorie intake/day | % professional/managerial occupation | % educated A level and above |
|---|---|---|---|---|---|
| Blue Collar Communities | 52.1 | 18.0 | 2392 | 54 | 37 |
| City Living | 51.7 | 15.0 | 2303 | 73 | 70 |
| Countryside | 52.7 | 18.2 | 2392 | 66 | 54 |
| Prospering Suburbs | 52.4 | 16.7 | 2351 | 61 | 51 |
| Constrained by Circumstance | 53.3 | 16.3 | 2343 | 55 | 37 |
| Typical Traits | 51.6 | 16.8 | 2346 | 63 | 52 |
| Multicultural | 50.0 | 15.2 | 2287 | 68 | 61 |
| $Chi^2$ (p) | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

Table 5.9 - OAC Supergroup descriptive statistics

All differences are statistically significant ($chi^2$ $p<0.001$) between each OAC Supergroup for all dietary patterns (Figure 5.6).

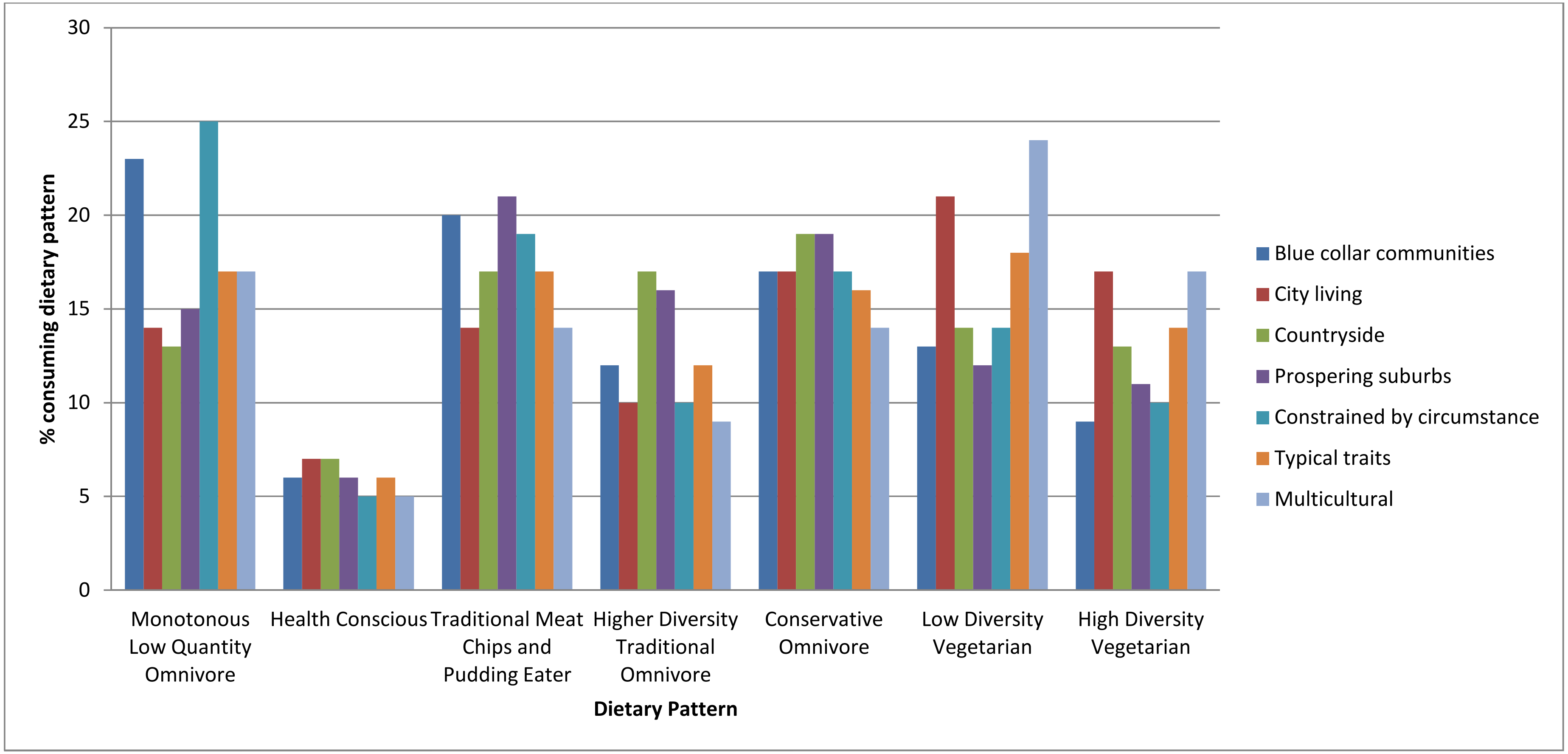



Figure 5.6 - Percentage of UKWCS women consuming each dietary pattern by OAC Supergroup

Unadjusted multinomial logistic regression models show that the Constrained by Circumstance group have a significantly elevated RRR of 1.35 for consuming the least healthy, Monotonous Low Quantity Omnivore dietary pattern whilst also having a significantly lower RRR for consuming the two most healthy dietary patterns, the Health Conscious (RRR=0.72) and the Higher Diversity Traditional Omnivore (RRR=0.77). The opposite is observed for the Countryside Supergroups who have an increased likelihood of consuming the healthy patterns (Health Conscious RRR=1.33 and High Diversity Traditional Omnivore RRR=1.36) and decreased likelihood of consuming the unhealthiest, Monotonous Low Quantity Omnivore pattern (RRR=0.76).

The adjusted model accounts for total calorie intake and physical activity in order that observed effects can be assumed to be dietary pattern related and not due to the volume of energy intake or expenditure. Other variables, such as age, education and social class, which could be considered as confounders have not been adjusted for, as these variables are included in assignment of the OAC.

In the adjusted model, these relationships hold true and are in fact accentuated for the Constrained by Circumstance group (Monotonous Low Quantity Omnivore diet RRR=1.43, Health Conscious RRR=0.68, Higher Diversity Traditional Omnivore RRR=0.71). The Blue Collar Communities also show the same convincing pattern (Monotonous Low Quantity Omnivore diet RRR=1.24, Health Conscious RRR=0.77, Higher Diversity Traditional Omnivore RRR=0.73).

Interestingly, in both the unadjusted and adjusted models, the City Living and Multicultural Supergroups are approximately 50% more likely to consume a vegetarian diet, with other Supergroups less likely, compared to the Typical Traits Supergroup.

The pseudo $R^2$ value indicates 10% of variation in dietary pattern is explained by the adjusted model. This is similar to that shown in model using GOR and adjusting for key demographic characteristics.

| | Monotonous Low Quantity Omnivore | Health Conscious | Traditional Meat Chips and Pudding Eater | Higher Diversity Traditional Omnivore | Conservative Omnivore | Low Diversity Vegetarian | High Diversity Vegetarian |
|---|---|---|---|---|---|---|---|
| | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **Ref.** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI) p value** | **RRR (95% CI p value** |
| Unadjusted model (pseudo $R^2$=0.007) | | | | | | | |
| Blue Collar Communities | 1.16 (1.00 to 1.36) p=0.054 | 0.87 (0.69 to 1.10) p=0.206 | 1.00 | 0.81 (0.68 to 0.97) p=0.022 | 0.87 (0.74 to 1.02) p=0.095 | **0.65 (0.55 to 0.78) p<0.001** | **0.57 (0.47 to 0.69) p<0.001** |
| City Living | 1.05 (0.88 to 1.26) p=0.587 | **1.55 (1.23 to 1.95) p<0.001** | 1.00 | 1.05 (0.87 to 1.28) p=0.611 | **1.36 (1.14 to 1.61) p=0.001** | **1.56 (1.32 to 1.84) p<0.001** | **1.59 (1.33 to 1.89) p<0.001** |
| Countryside | **0.76 (0.68 to 0.86) p<0.001** | **1.33 (1.14 to 1.55) p<0.001** | 1.00 | **1.37 (1.21 to 1.54) p<0.001** | 1.15 (1.03 to 1.29) p=0.016 | **0.79 (0.71 to 0.89) p<0.001** | 0.97 (0.86 to 1.09) p=0.582 |
| Prospering Suburbs | **0.76 (0.69 to 0.84) p<0.001** | 0.90 (0.78 to 1.04) p=0.144 | 1.00 | 1.10 (0.99 to 1.23) p=0.083 | 0.96 (0.87 to 1.06) p=0.444 | **0.59 (0.53 to 0.65) p<0.001** | **0.66 (0.59 to 0.73) p<0.001** |
| Constrained by Circumstance | **1.33 (0.11 to 1.59) p=0.002** | 0.69 (0.51 to 0.93) p=0.015 | 1.00 | **0.74 (0.59 to 92) p=0.008** | 0.91 (0.75 to 1.10) p=0.330 | **0.71 (0.58 to 0.86) p=0.001** | **0.67 (0.54 to 0.83) p<0.001** |
| Typical Traits | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | **1.00** | **1.00** |
| Multicultural | 1.23 (1.02 to 1.47) p=0.027 | 1.16 (0.90 to 1.49) p=0.260 | 1.00 | 0.93 (0.76 to 1.15) p=0.524 | 1.09 (0.91 to 1.32) p=0.360 | **1.70 (1.43 to 2.01) p<0.001** | **1.55 (1.29 to 1.86) p<0.001** |
| Adjusted model (adjusting for smoking, total calorie intake including alcohol and typical daily physical activity (METs) (pseudo $R^2$=0.10) | | | | | | | |
| Blue Collar Communities | **1.25 (1.06 to 1.48) p=0.009** | 0.79 (0.62 to 1.00) p=0.050 | 1.00 | **0.72 (0.60 to 87) p=0.001** | 0.95 (0.80 to 1.12) p=0.526 | **0.70 (0.59 to 0.83) p<0.001** | **0.55 (0.05) p<0.001** |
| City Living | 0.88 (0.73 to 1.07) p=0.208 | **1.66 (1.32 to 2.10) p<0.001** | 1.00 | 1.10 (0.90 to 1.34) p=0.349 | 1.24 (1.04 to 1.48) p=0.019 | **1.48 (1.25 to 1.75) p<0.001** | **1.64 (0.15) p<0.001** |
| Countryside | **0.79 (0.69 to 0.89) p<0.001** | **1.31 (1.12 to 1.53) p=0.001** | 1.00 | **1.35 (1.20 to 1.53) p<0.001** | 1.14 (1.01 to 1.28) p=0.029 | **0.79 (0.70 to 0.89) p<0.001** | 0.95 (0.06) p=0.386 |
| Prospering Suburbs | **0.77 (0.69 to 0.86) p<0.001** | 0.91 (0.79 to 1.06) p=0.215 | 1.00 | 1.11 (1.00 to 1.25) p=0.052 | 0.94 (0.85 to 1.05) p=0.259 | **0.58 (0.52 to 0.64) p<0.001** | **0.65 (0.04) p<0.001** |
| Constrained by Circumstance | **1.40 (1.16 to 1.71) p=0.001** | **0.66 (0.49 to 0.90) p=0.008** | 1.00 | **0.69 (0.55 to 0.86) p=0.001** | 0.98 (0.80 to 1.20) p=0.859 | **0.75 (0.61 to 0.92) p=0.005** | **0.66 (0.08) p<0.001** |
| Typical Traits | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Multicultural | 0.99 (0.82 to 1.21) p=0.951 | 1.24 (0.96 to 1.60) p=0.107 | 1.00 | 0.98 (0.79 to 1.21) p=0.839 | 0.99 (0.82 to 1.20) p=0.938 | **1.62 (1.36 to 1.92) p<0.001** | **1.65 (0.15) p<0.001** |

Table 5.10 - Regression models investigating whether geodemographic Supergroup predicts dietary patterns displaying RRR (95% CI) and p value

### 5.4.5 Differences by spatial scale

Significant variation for consumption of all dietary patterns by OAC Supergroup is observed (Figure 5.6). In general, variation is wider when considering differences by geodemographic Supergroup compared to by GOR (Figure 5.2), North South (Figure 5.3) and Urban/Rural (Figure 5.4) - suggesting that the inclusion of demographic variables with small area geography could tell us something more interesting about dietary patterns.

The UKWCS mean age was 52.2 years (SD 9.3) with means ranging between 51.4 to 53.0 years by GOR and 50.1 to 53.4 years by OAC Supergroup. Age is more homogenous by North South and Urban Rural, with all women aged approximately 52 years.

Physical activity, measured used METS calculated from a 24 hour physical activity diary had a mean value of 16.0 METS (SD 11.7); regional range 16.4-17.8; OAC Supergroup range 15.0 - 18.2. Again, results are similar by North South split (16.8 compared to 16.9) but physical activity is higher in Rural areas (17.97 METS) compared to Urban (16.34 METS).

65% of the cohort had a professional/managerial occupation with a GOR range of 58-67% and OAC Supergroup range of 53-73%, with highest numbers in the City Living Supergroup and lowest in the Blue Collar Communities. Values for North South and Urban/Rural were all in the 62-66% range.

52% of the cohort were educated to A-level or above with a regional range of 46-65% and OAC Supergroup range 37-70% with lowest values in the Constrained by Circumstance and Blue Collar Communities and highest again in City Living. North South and Urban/ Rural were again very similar with percentages 51-54%.

A wider variation in the mean characteristics of the women is evident when grouped by OAC Supergroup compared to grouping by region or North South or Urban/ Rural. This supports the expectation that incorporating demographic variables into a geographical classification tells us something about specific populations, rather than seeing averaged out characteristics of a geographical region.

### 5.4.6 Diet cost analysis

Daily cost of diet varies significantly according to where the women live (Figure 5.7). Differences were observed within each spatial measure. When comparing mean diet cost according to GOR, the cheapest diet is observed in Greater London, £4.26 per day and the most expensive in Scotland at £4.65 per day. The diet in the North of England is more expensive than in the South, £4.51 compared to £4.42 per day and Urban areas have a cheaper diet than Rural areas, £4.41 compared to £4.58 per day.

When grouped according to which OAC Supergroup the women live, the variation in diet cost, whilst statistically significant, is less pronounced than by GOR, with costs ranging between £4.33 and £4.57. Interestingly, the cheapest diets are for those living in Multicultural areas, with the most expensive residing in the Countryside Supergroups.

Unadjusted regression analysis (Model 1 in Table 5.11 and Table 5.12) for all the spatial measures explains very little variation in diet cost, with $R^2$ values less than 1%, although some statistically significant differences are observed. Scotland and the East Midlands have significantly higher diet costs, as much as £0.20 and £0.13 per day, compared to the reference group, the South East. Greater London has significantly lower daily diet cost which is £0.18 cheaper per day. The North consume a significantly more expensive diet than the South, by £0.08 per day and diet costs are £0.16 per day more expensive in Rural areas compared to Urban. Compared to the Typical Traits super group, the Countryside group and the Prospering Suburbs consume a diet costing £0.15 and £0.09 respectively, more per day, whist those in Multicultural areas spend £0.16 less per day.

Controlling for energy intake, energy expenditure (in METS for physical activity) age, social class and education (Model 2 Table 5.11) explains 59-60% variation in the models for GOR, North South and Urban/ Rural. Significant differences in diet cost are only observed between Urban and Rural environments, with those in Rural areas consuming a diet costing £0.05 more per day.

For analysis with OAC as a predictor, model 2 controls for energy intake, energy expenditure and smoking (See Table 5.12). Social class, education and age are not included as these are components of the OAC classification. In this model, significant differences remain in diets consumed by those in Countryside and Prospering Suburb groups who consume a diet which is £0.07 more expensive per day. $R^2$ values for this model are very similar to those in model 2 for GOR, North South and Urban/ Rural models, indicating that OAC Supergroup explains 59% of the variance. The Multicultural diet is still cheaper by £0.05, but this result is no longer statistically significant.

In all four fully adjusted models, also accounting for dietary pattern consumed, the explained variance in the model increased to 70%.

No significant differences are seen for GOR, North South or Urban/ Rural spatial measures (Table 5.11). In the OAC Supergroup analysis significant differences are observed. Interestingly, the diet consumed by those in City Living Supergroups is significantly less expensive now by £0.06 per day. The diet in Prospering Suburbs remains significantly more expensive by £0.04 per day (Table 5.12).

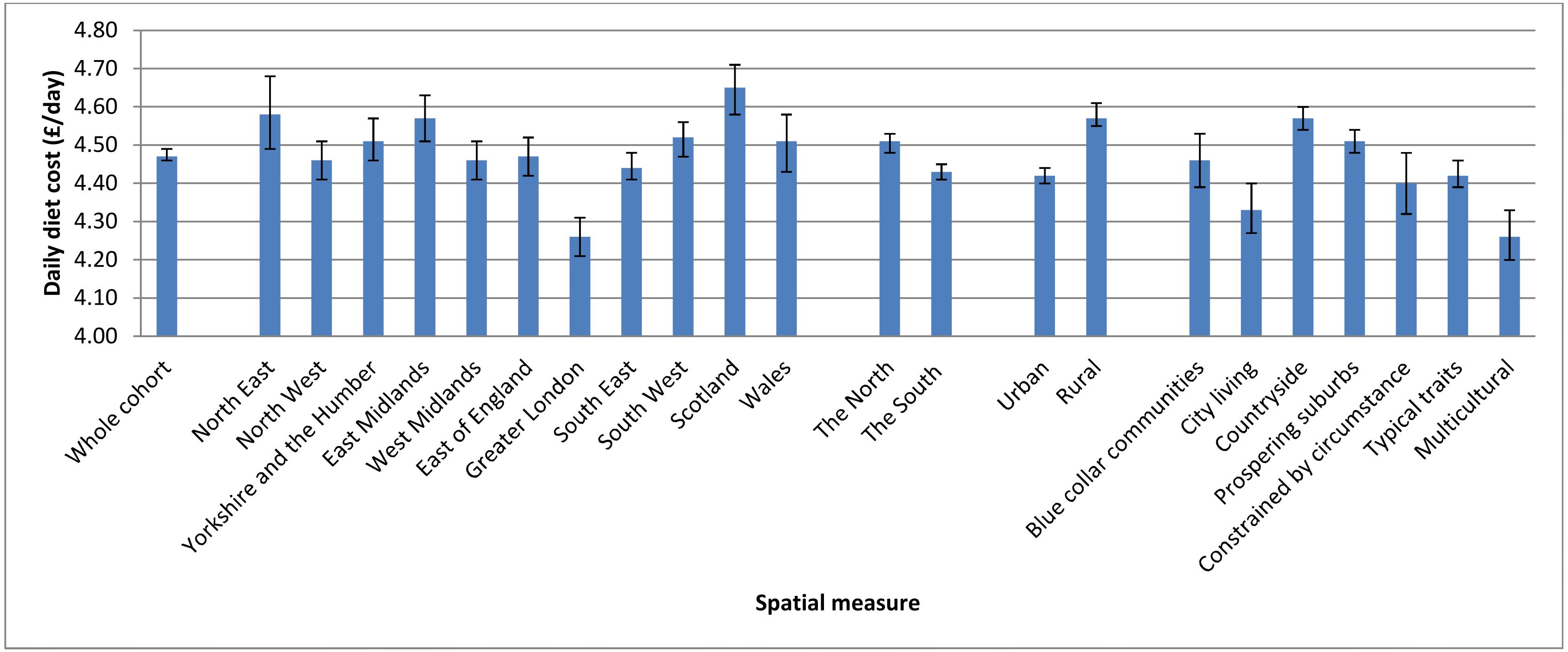



Figure 5.7 - Mean daily cost of diet (£/day) in the UKWCS, with 95% confidence interval error bars, by spatial measure

| | **Model 1 - Unadjusted** | **Model 2 - adjusted for smoking, energy intake, energy expenditure, age, social class and education** | **Model 3 - adjusted for dietary pattern, smoking, energy intake, energy expenditure, age, social class and education** |
|---|---|---|---|
| | **Diet cost £/day (95% CI) p value** | **Diet cost £/day (95% CI) p value** | **Diet cost £/day (95% CI) p value** |
| **Government office region** | $R^2$=0.0043 | $R^2$= 0.5962 | $R^2$=0.6979 |
| North East | **0.14 (0.04 to 0.24) 0.005** | -0.03 (-0.09 to 0.03) 0.355 | 0.01 (-0.05 to 0.06) 0.840 |
| North West | 0.01 (-0.05 to 0.08) 0.642 | -0.04 (-0.08 to 0.00) 0.073 | 0.00 (-0.03 to 0.04) 0.910 |
| Yorkshire and the Humber | 0.07 (0.01 to 0.14) 0.033 | 0.01 (-0.03 to 0.06) 0.530 | 0.03 (-0.01 to 0.06) 0.171 |
| East Midlands | **0.13 (0.06 to 0.19) <0.001** | 0.05 (0.00 to 0.09) 0.047 | 0.03 (0.00 to 0.07) 0.094 |
| West Midlands | 0.02 (-0.05 to 0.08) 0.616 | 0.01 (-0.03 to 0.05) 0.653 | 0.01 (-0.03 to 0.05)0.565 |
| East of England | 0.03 (-0.03 to 0.09) 0.361 | 0.04 (0.00 to 0.08) 0.045 | 0.03 (0.00 to 0.07) 0.067 |
| Greater London | **-0.18 (-0.24 to -0.13) <0.001** | -0.03 (-0.07 to 0.00) 0.086 | -0.03 (-0.07 to 0.00) 0.043 |
| South East | Ref. | Ref. | Ref. |
| South West | 0.07 (0.02 to 0.13) 0.011 | 0.01 (-0.02 to 0.05) 0.498 | 0.00 (-0.03 to 0.04) 0.851 |
| Scotland | **0.20 (0.13 to 0.27) <0.001** | -0.03 (-0.07 to 0.02) 0.289 | -0.02 (-0.06 to 0.02) 0.414 |
| Wales | 0.06 (-0.02 to 0.15) 0.130 | -0.03 (-0.07 to 0.02) 0.638 | 0.01 (-0.04 to 0.05) 0.825 |
| **North South** | $R^2$=0.0007 | $R^2$=0.5937 | $R^2$=0.6950 |
| The North | **0.08 (0.04 to 0.11) <0.001** | 0.00 (-0.02 to 0.02) 0.944 | 0.02 (0.00 to 0.04) 0.083 |
| The South | Ref. | Ref. | Ref. |
| **Urban Rural** | $R^2$=0.0027 | $R^2$=0.5962 | $R^2$=0.6977 |
| Urban | Ref. | Ref. | Ref. |
| Rural | **0.16 (0.13 to 0.19)<0.001** | **0.05 (0.03 to 0.08) <0.001** | 0.01 (-0.01 to 0.03) 0.217 |

Table 5.11 - Regression model investigating the influence of GOR, North South and Urban/ Rural on cost of diet

| | **Model 1 - Unadjusted** | **Model 2 - adjusted for smoking, energy intake and energy expenditure** | **Model 3 - adjusted for dietary pattern, smoking, energy intake and energy expenditure** |
|---|---|---|---|
| | **Diet cost £/day (95% CI) p value** | **Diet cost £/day (95% CI) p value** | **Diet cost £/day (95% CI) p value** |
| **OAC Supergroup** | $R^2$=0.0035 | $R^2$=0.5940 | $R^2$=0.6957 |
| Blue Collar Communities | 0.03 (-0.04 to 0.11) 0.334 | -0.01 (-0.06 to 0.03) 0.589 | 0.03 (-0.01 to 0.06) 0.195 |
| City Living | -0.09 (-0.16 to 0.02) 0.011 | -0.01 (-0.05 to 0.03) 0.686 | **-0.06 (-0.09 to -0.02) 0.004** |
| Countryside | **0.15 (0.10 to 0.19) <0.001** | **0.07 (0.04 to 0.10) <0.001** | 0.01 (-0.02 to 0.04) 0.404 |
| Prospering Suburbs | **0.09 (0.04 to 0.13) <0.001** | **0.07(0.04 to 0.10) <0.001** | **0.04 (0.02 to 0.07) 0.001** |
| Constrained by Circumstance | -0.03 (-0.11 to 0.06) 0.523 | -0.01 (-0.07 to 0.04) 0.618 | 0.04 (0.00 to 0.09) 0.058 |
| Typical Traits | Ref. | Ref. | Ref. |
| Multicultural | **-0.16 (-0.23 to 0.09) <0.001** | -0.05 (-0.10 to -0.003) 0.036 | -0.05 (-0.09 to -0.01) 0.019 |

Table 5.12 - Regression model investigating the influence of OAC Supergroup on cost of diet

### 5.4.7 Applying geodemographic analysis results to a new area

In order to estimate dietary patterns for a new area using OAC Group diet must first be profiled by OAC Group (Table 5.13). Group is used for this instead of Supergroup as it is a more granular unit, being the second layer of the OAC hierarchy.

Dietary profiles, consisting of the percentage of women consuming each dietary pattern, were applied to the OAC Groups for Leeds (see Figure 5.8 to Figure 5.14). Three visible bands appear in these maps of estimated dietary patterns:

- Band A - A broad band across the north and east of the city containing more affluent wards (Aireborough, Otley and Wharfedale, North; Wetherby, Barwick and Kippax) This band is largely comprised of geodemographic Supergroups such as Countryside and Prospering Suburbs;
- Band B - An inner city cluster encompassing wards University, Headingley, Weetwood, Chapel Allerton and parts of Armley, Holbeck, Kirkstall and Moortown). There are high concentrations of students in at least three of these wards and all contain multicultural communities;
- Band C - A corridor to the south east of the city centre area, including fewer superstores and poorer wards (Seacroft, Burmantofts, Richmond Hill, Hunslet, and Middleton), consisting of geodemographic Supergroups such as; Blue Collar Communities and those Constrained by Circumstances.

Figure 5.8 to Figure 5.15 show that Leeds - Band A, are estimated to have a higher than average consumption of the Health Conscious, Higher Diversity Traditional Omnivore and the High Diversity Vegetarian dietary patterns and a low consumption of the Monotonous Low Quantity Omnivore. The Conservative Omnivore dietary pattern is moderately consumed here, along with that of the High Diversity Vegetarian. The inner city wards - Band B are estimated to have high concentration of vegetarian dietary patterns, both high and low diversity. Moderate Conservative Omnivore dietary pattern consumption exists here, but low concentrations of all the other patterns. Poorer wards to the south and east - Band C, are estimated to have a higher than average estimated percentage of the Monotonous Low Quantity Omnivore and Traditional Meat Chips and Pudding Eater dietary patterns. It has not been possible at this stage to perform any validation of these patterns using actual dietary data for Leeds residents. With this in mind a visual comparison with deprivation in Leeds has been included in order to compare the results in this paper with findings in other research. Figure 5.15 shows the Index of Multiple Deprivation (IMD) for Leeds. This serves as an aid to better understand the social geography of the City for those not familiar with the area. This map shows that the estimated poorer dietary patterns correlate with the poorer areas of Leeds. Daily diet cost also appears to correlate well with IMD for Leeds, with the most expensive diets appearing in Band A (Figure 5.16). The OAC and IMD share some demographic variables so some correlation could be expected.

| Group | % Monotonous low quantity omnivore | % Health conscious | %Traditional Meat Chips and Pudding Eaters | % Higher diversity traditional omnivore | % Conservative Omnivore | % Low Diversity Vegetarian | % High Diversity Vegetarian |
|---|---|---|---|---|---|---|---|
| 1a - Terraced Blue Collar (n=114) | 26 | 4 | 22 | 11 | 18 | 11 | 8 |
| 1b - Younger Blue Collar (n=204) | 24 | 6 | 21 | 10 | 16 | 14 | 9 |
| 1c - Older Blue Collar (n=472) | 22 | 6 | 19 | 13 | 16 | 13 | 10 |
| 2a - Transient Communities (n=210) | 15 | 6 | 14 | 5 | 17 | 27 | 16 |
| 2b - Settled in City (n=660) | 14 | 7 | 13 | 12 | 17 | 19 | 17 |
| 3a - Village Life (n=1109) | 14 | 6 | 18 | 17 | 18 | 13 | 14 |
| 3b - Agricultural (n=914) | 12 | 8 | 18 | 17 | 18 | 14 | 13 |
| 3c - Accessible Countryside (n=1258) | 12 | 8 | 16 | 17 | 20 | 14 | 13 |
| 4a - Prospering Younger Families (n=459) | 17 | 5 | 23 | 15 | 16 | 15 | 9 |
| 4b - Prospering Older Families (n=1844) | 14 | 7 | 21 | 17 | 19 | 12 | 12 |
| 4c - Prospering Semis (n=1145) | 19 | 5 | 21 | 14 | 20 | 11 | 9 |
| 4d - Thriving Suburbs (n=1374) | 14 | 7 | 19 | 18 | 18 | 13 | 12 |
| 5a - Senior Communities (n=49) | 29 | 5 | 13 | 13 | 18 | 15 | 7 |
| 5b - Older Workers (n=410) | 23 | 5 | 19 | 11 | 17 | 14 | 12 |
| 5c - Public Housing (n=51) | 35 | 5 | 25 | 5 | 14 | 11 | 5 |
| 6a - Settled Households (n=654) | 19 | 4 | 19 | 12 | 17 | 18 | 12 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 6b - Least Divergent (n=805) | 16 | 6 | 17 | 13 | 17 | 17 | 13 |
| 6c - Young Families in Terraced Homes (n=407) | 20 | 4 | 18 | 11 | 14 | 19 | 14 |
| 6d - Aspiring Households (n=832) | 14 | 7 | 16 | 13 | 16 | 17 | 16 |
| 7a - Asian Communities (n=503) | 16 | 5 | 15 | 10 | 15 | 23 | 16 |
| 7b - Afro-Caribbean (n=277) | 18 | 5 | 11 | 8 | 14 | 25 | 19 |
| Chi$^2$ test: p value for dietary pattern by Group | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 | p<0.001 |
| 16 | 6 | 18 | 14 | 18 | 15 | 13 | |

Table 5.13 - Percentage of OAC group consuming each dietary pattern, in the UKWCS

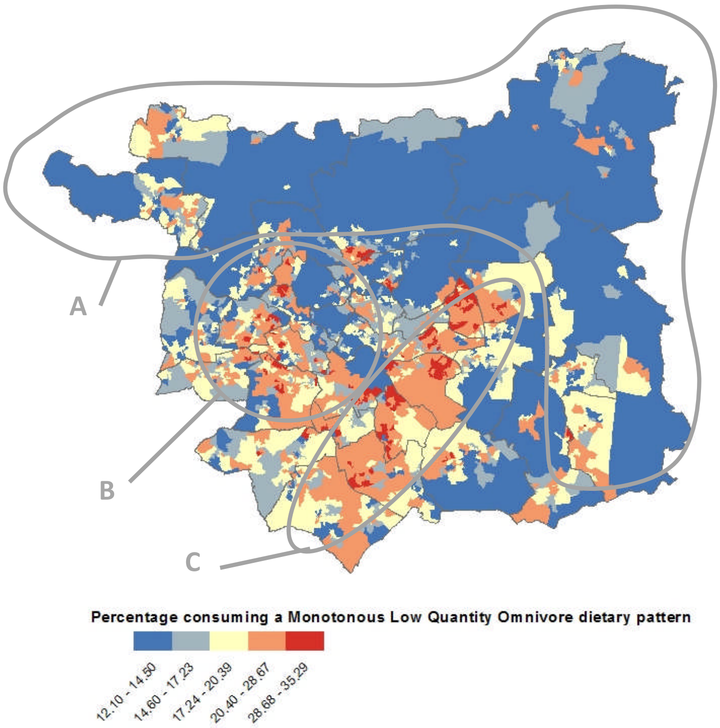



Figure 5.8 - Estimated Monotonous Low Quantity Omnivore Dietary Pattern Consumption at OAC Group Level for Leeds, with wards indicated and three observed bands highlighted

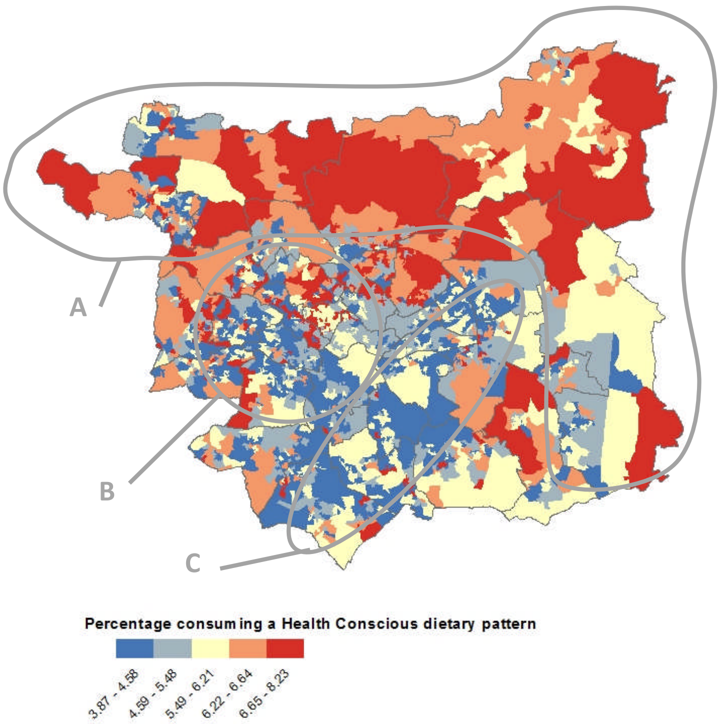



Figure 5.9 - Estimated Health Conscious dietary pattern for Leeds at OAC group level, with wards indicated and three observed bands highlighted

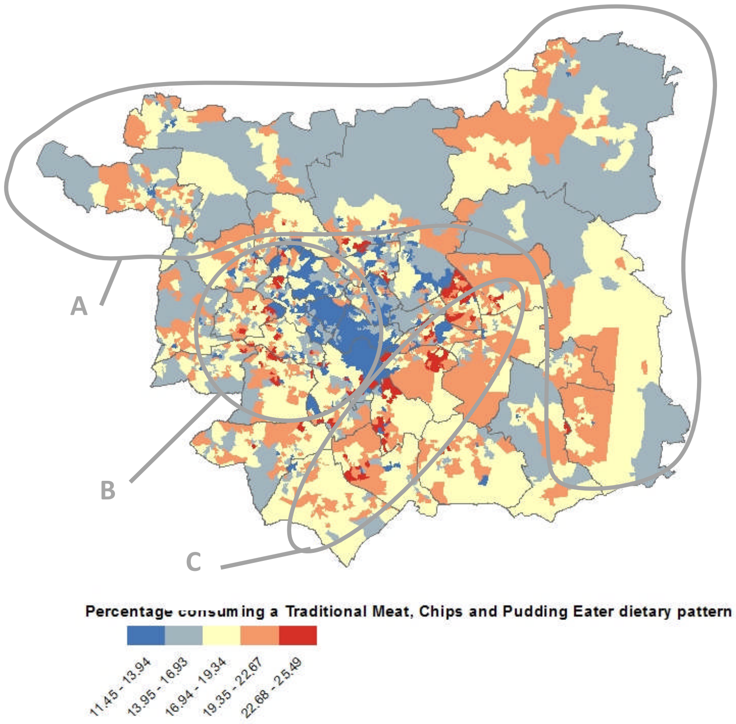



Figure 5.10 - Estimated Traditional Meat, Chips and Pudding Eater dietary pattern consumption for Leeds at OAC group level, with wards indicated and three observed bands highlighted

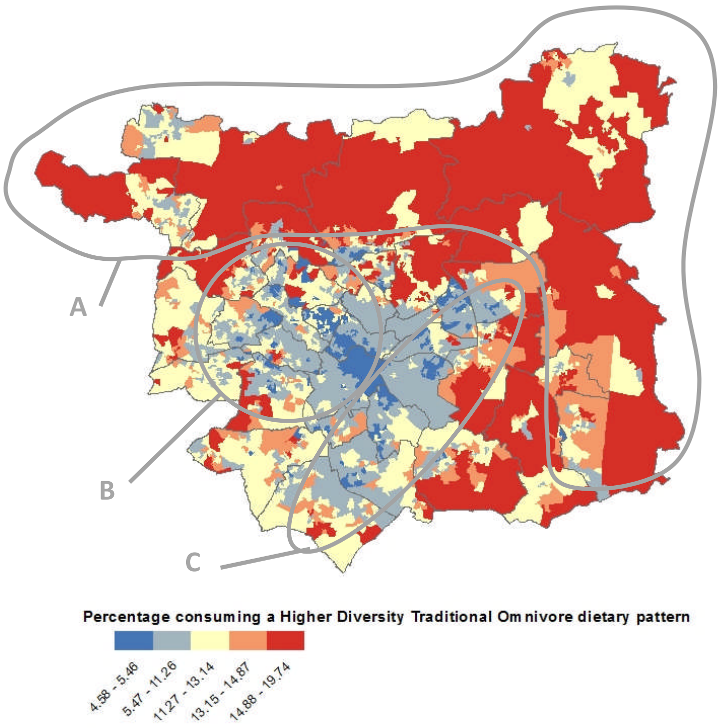



Figure 5.11 - Estimated Higher Diversity Traditional Omnivore dietary pattern consumption at OAC Group Level for Leeds, with wards indicated and three observed bands highlighted

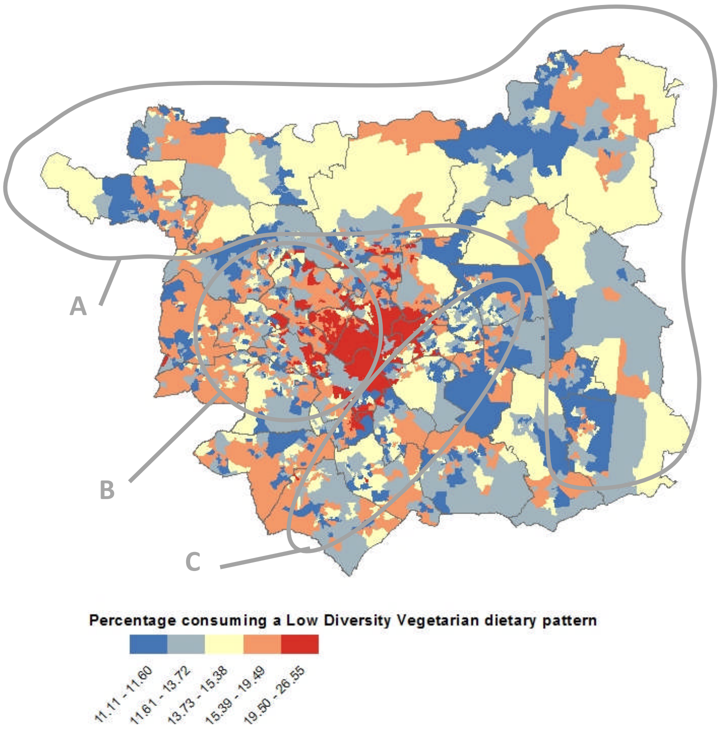



Figure 5.12 - Estimated Low Diversity Vegetarian dietary pattern consumption at OAC Group Level for Leeds, with wards indicated and three observed bands highlighted

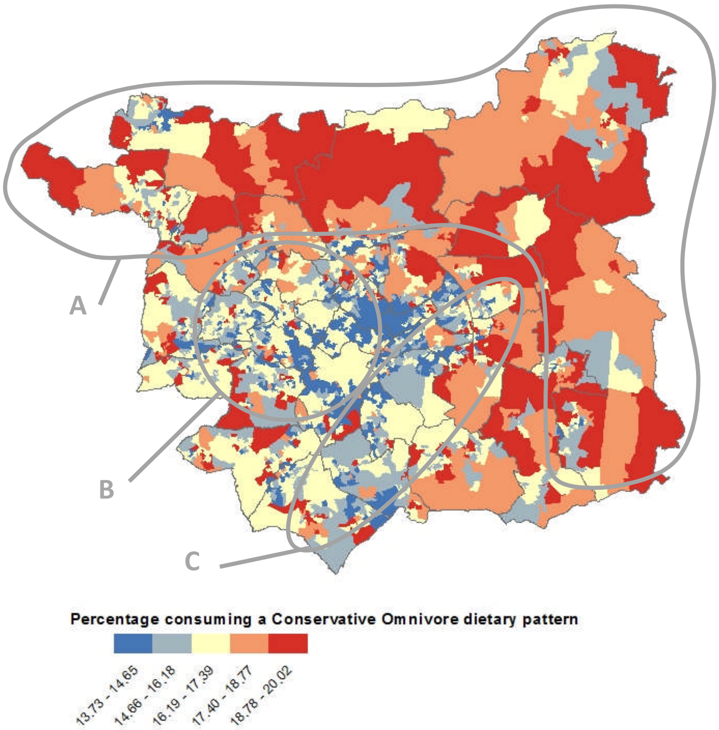



Figure 5.13 - Estimated Conservative Omnivore dietary pattern consumption at OAC Group Level for Leeds, with wards indicated and three observed bands highlighted

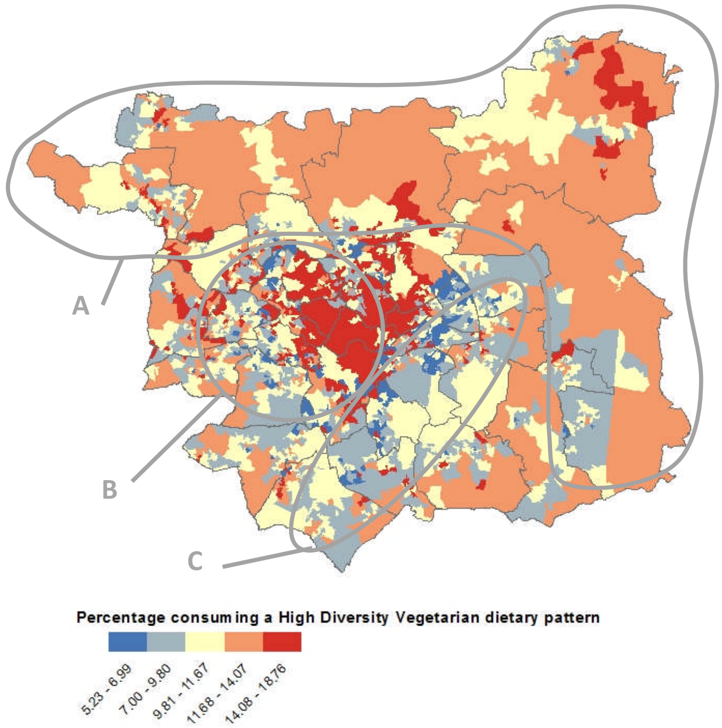



Figure 5.14 - Estimated High Diversity Vegetarian dietary pattern consumption at OAC Group Level for Leeds, with wards indicated and three observed bands highlighted

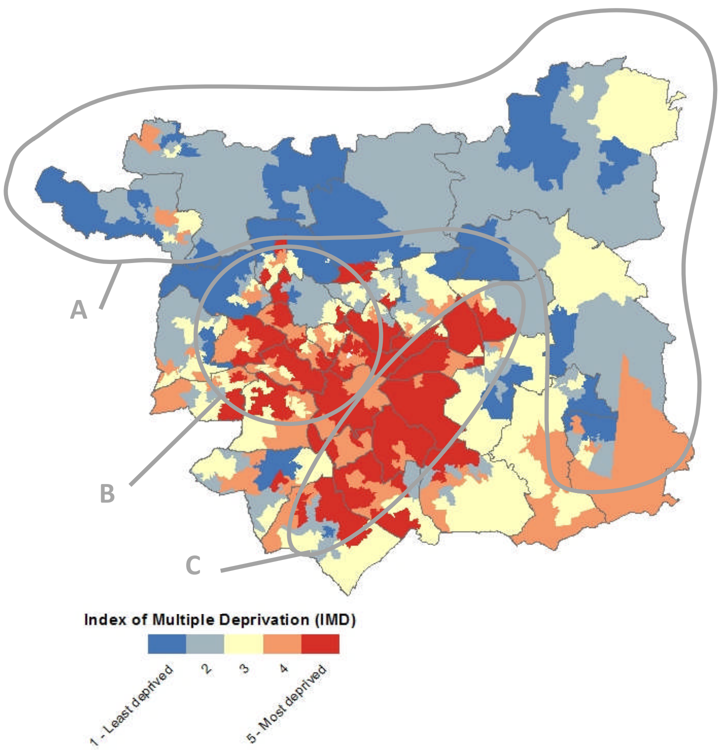



Figure 5.15 - Index of Multiple Deprivation (IMD) for Leeds, with wards indicated and three observed bands highlighted

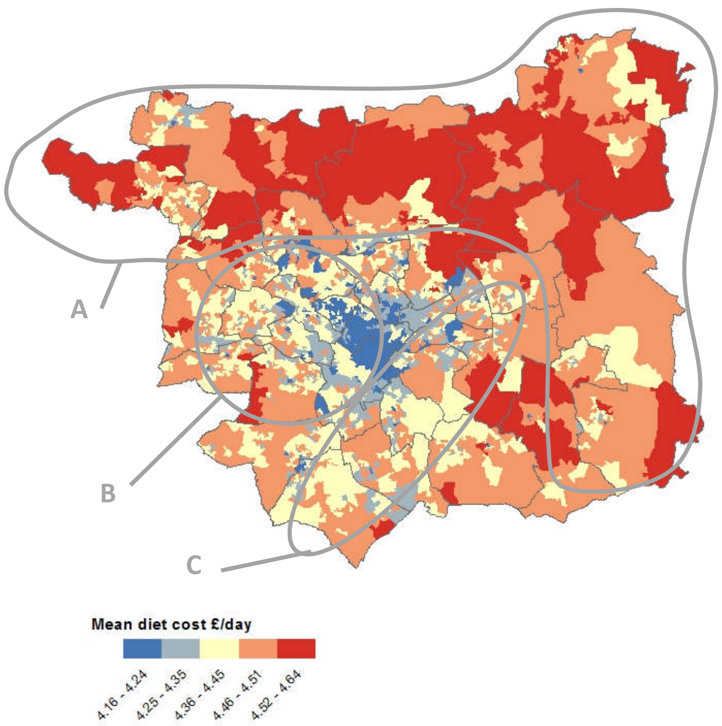



Figure 5.16 - Estimated diet cost at OAC Group Level for Leeds, with wards indicated and three observed bands highlighted

## 5.5 Discussion

Four different spatial measures are used to describe the UKWCS and the dietary patterns and cost of diet consumed by these women. Interesting variations in dietary pattern consumption are observed by each spatial measure.

### 5.5.1 Summary characteristics

Summary characteristics such as education and occupation for the UKWCS vary only a little by region from the mean values for the cohort as a whole. However, when the same characteristics are presented by OAC Supergroup variation is greater. The OAC uses variables such as age, occupation and education, as collected in the census. No information relating to diet is included in the census or the OAC, yet increased variation in dietary patterns by OAC rather than GOR is observed supporting evidence that socio-demographic characteristics influence diet (Aggarwal et al., 2011), but also that small area geography - the immediate local environment - influences diet (Fraser and Edwards, 2010). Using the small geographic unit of output area (incorporating a local environment as small as 100 individuals) compared to the large geographical unit of a government office region (containing millions of individuals) produces results which will be much more relevant at a local area level.

### 5.5.2 Dietary patterns

The geodemographic classification generates a more detailed picture of the dietary patterns than regression with government office region, controlling for a number of demographic characteristics, although regression models explain comparable levels of variation. Results are also suggestive of an association between a healthy diet - illustrated by the Health Conscious dietary pattern and dietary diversity - and being a member of the more affluent OAC groups. This is supportive of other research suggesting that a healthy diet is more expensive and therefore restricted by financial ability (Maillot et al., 2007b). The Family Spending report from the Living Costs and Food Survey (Office for National Statistics, 2010) indicates that the OAC Supergroups Countryside and Prospering Suburbs spend the most on food per week, which is also in line with these findings based on the assumption that a healthier diet is a more expensive one.

Unlike earlier studies using fruit and vegetable intake as a proxy measure of a healthy diet (Wrigley et al., 2002), in this study the Health Conscious dietary pattern incorporates more dietary components than just fruit and vegetable intake to represent dietary healthiness. A full spectrum of dietary diversity is also represented by the dietary patterns (as indicated by the pattern names). High dietary diversity appears to be associated with living in more affluent OAC Groups, and conversely dietary monotony correlates with the poorer areas. A conservative dietary pattern and the traditional dietary pattern have slightly higher concentrations in the deprived areas.

### 5.5.3 Diet cost

Variation in diet cost does not necessarily fit with what might have been expected, Scotland is the region with the most expensive diet, yet higher consumption of the healthiest dietary patterns are not seen here. It is widely accepted that cost of living is more in Greater London than areas such as the North of England, yet Greater Londoners are consuming the least expensive diet and significant variations in dietary patterns are observed. The cost of diet is less in the South of England and in Urban areas, which is further supported by City Living and Multicultural areas consuming the cheapest diets. These areas have higher proportion of vegetarians, but only the Low Diversity Vegetarian patterns is a cheaper dietary option (as shown in Chapter 4). Adding a geocoded location to dietary cost patterns could be useful to policy makers in trying to combat the unhealthy dietary behaviours. Using Geodemographics in this way is a meaningful way to examine spatial variations in diet cost.

When taking into account results from Chapter 4 relating to the cost of dietary patterns it can be seen that the most average dietary pattern - the Traditional Meat Chips and Pudding Eater - has a very similar diet cost to those living in the most average geodemographic group - Typical Traits - with only £0.03 per day (<1%) daily cost difference in diet. The Typical Traits Supergroup are also those with the closest to average values for BMI, physical activity, age, education and occupation. This could be expected for age, education and occupation as these variables are included in the development of the OAC classification, but is interesting to see for BMI and physical activity. Using this data it seems reasonable that dietary patterns could be included in geodemographic classification profiles, in a similar way to which the Acorn Health classification does (CACI Ltd).

Differences between Urban and Rural living can be seen in the UKWCS, with the Countryside Supergroup spending more on food than those in City Living, which is consistent with recent findings in the 2012 Living costs and food survey (Office for National Statistics, 2013). While the difference in costs between groups appears greater in the Family Spending survey results, this may be due to it being household expenditure at 2012 prices compared with the earlier prices for individual’s food in the UKWCS.

### 5.5.4 Estimated dietary patterns

This study is the first of its kind to take the dietary patterns based on a large national sample of women and using a geodemographic classification profile estimated patterns for a specific city. This method allows dietary cost and dietary patterns to be visualised for neighbourhoods within a specific UK city, which may be applicable in other public health contexts.

While estimated patterns seem reasonable as they are so similar to patterns of deprivation in the city, validation of the estimates would be required in order to be confident of their reliability.

### 5.5.5 Policy implications

Dietary pattern variation between regions, with the exception of the Health Conscious dietary pattern, is statistically significant. It is interesting to see that such variation occurs at a large geographical scale, suggestive that there are regional influences on eating habits, not just that of the local surrounding area. Further investigation could be carried out into this regional variation; however it may be erroneous to ascribe specific factors to a particular region, if certain areas within that region have a dominating influence on the dietary pattern. The regression analysis using region, including demographic variables, shows that some regions have significant relative risk ratios of consuming a particular dietary pattern. For example, Greater London has significantly increased risk, compared to those living in the South East, of consuming both the Health Conscious (most healthy) and the Monotonous Low Quantity Omnivore (least healthy) diet (compared to consuming the traditional diet). While this is an interesting observation, it means that implementing a cost effective nutritional intervention at regional level would be extremely difficult.

However, when we consider dietary pattern variation by geodemographic Supergroup the results present a clearer picture. Those living in a Blue Collar Community, for example, have a significantly increased risk of consuming the Monotonous Low Quantity Omnivore (least healthy) dietary pattern and a significantly reduced risk of consuming the Health Conscious and High Diversity Traditional Omnivore (two most healthy) dietary patterns, suggesting that it would be worthwhile implementing a healthy diet promotion in these types of areas. The difference in geodemographic results are likely to be due to more than just age or education, as including these variables in a regional regression model did not influence the results in the same way as geodemographic Supergroups. A combination of both the regional and Supergroup results could help to target interventions to certain types of areas within the region most at risk of consuming a poor dietary pattern.

Being able to estimate dietary patterns at a small area level using a classification such as OAC, allows for smarter targeting of public health interventions, to improve diet and subsequent health. For example, to provide a specific intervention to individuals living in Public Housing and Senior Communities OAC Groups (who consume the highest percentage of Monotonous Low Quantity Omnivore diets) which would encourage them to introduce more variety into their diets with the best addition being fruit and vegetables. This could be done through social services interventions in specific communities, or at a GP practice level in local communities.

### 5.5.6 Strengths and limitations

The UKWCS, specifically designed to investigate the effect of dietary patterns on health outcomes, provides quality dietary data for analysis. As a result, there are a high proportion of vegetarians in the cohort. This does not bias the results when considering whether women are more or less likely to consume a vegetarian dietary pattern but provides higher reliability due to the larger numbers.

Geographic location of participants was not a design factor for the UKWCS, so despite large numbers in each of the nine government office regions of the UK and Scotland and Wales these regions are not equally represented across the cohort, with over-sampling or under-sampling in some regions compared to the distribution of women in the whole of the UK. The lowest numbers (n=957) were observed residing in the North East. However, when considering this in the context of other dietary survey information, this is a large number of women on which to base robust analysis. The National Diet and Nutrition Survey (NDNS), for example, whilst it was designed to be geographically representative, only includes 3073 individuals in total (from the three year rollout) of which only about one quarter are women (Department of Health and Food Standards Agency, 2012). The UKWCS sample is approximately 40 times the size of that national sample of women.

Recruitment of the UKWCS was on a volunteer basis from a World Cancer Research Fund (WCRF) mailing list of previous questionnaire participants so it may be expected that there is some volunteer bias. The women are predominantly middle age, middle class and white. This may account for the over representation observed when comparing the UKWCS to the whole UK population, in the Countryside and Prospering Suburbs Supergroups which are characterised by middle age, more middle class, white individuals with larger detached houses. Under representation is observed in Blue Collar Communities, Constrained by Circumstance and Multicultural Supergroups. Despite this there are large numbers of women in each of the seven OAC Supergroups, sufficient to provide confidence in the associations observed in this study.

The dietary patterns used, whilst data driven, are not necessarily comparable to dietary patterns used in other surveys. That said, they provide a comprehensive illustration of dietary patterns consumed in the UK. Comparing these to another data driven but not international classification, such as the OAC, could be considered subjective. However, this chapter aims to illustrate how geodemographic classifications can be useful in public health and specifically dietary research, rather than critique dietary patterns or geodemographic classifications.

The OAC groups include all of the UK population who completed the Census questionnaire, so incorporate men and children in addition to women. Whilst the NDNS reports statistical differences between the food consumed by men and women (stratified both by region and whether or not the individuals are in receipt of benefits) (Food Standards Agency and Department of Health, 2002) it has not specifically reported whether there was a difference in the diets of men and women within the same household in the UK. With this in mind, the results of this study can only reliably be applied to women.

The dietary data used in this study were collected in the late 1990s. It is feasible that dietary habits could have changed since this time. However, it is rare that dietary information of this quality is collected in such a large sample in the UK. Dietary data with a time lag such as this is essential when considering influence on diseases with a latent development period, such as cancer (which will be explored in Chapter 7). Therefore the application of these results with respect to the effect of diet on health is relevant, despite possible dietary change since the data was collected.

Given that the women reside within specific geographical units, forming a hierarchal structure to the data, it could have been possible to perform analysis using multilevel modelling. The advantage of this method would have been to model the clustering effects within a geographical unit, for example within GORs. This could be included in future research.

## 5.6 Summary and context

This Chapter has summarised how dietary patterns and diet cost vary by four spatial measures and whether the OAC Group measure can be used to predict dietary pattern consumption.

Dietary pattern consumption is associated with where individuals reside. The type of area, using a small scale geographical unit combined with demographic characteristics, provides a richer understanding of dietary consumption than the large regional unit. Healthier, more expensive dietary patterns are more common in geodemographic groups Countryside or Prospering Suburbs. Less healthy dietary patterns are consumed in areas such as Constrained by Circumstance and Blue Collar Communities, however, these are not the areas consuming cheapest diets. It is Multicultural areas who consume the cheapest diets. With this in mind it may be beneficial to use such classifications in the application of dietary advice to encourage healthy eating in order to promote long term health. It may also be useful to estimate dietary patterns for a given city, in this case Leeds, in order to assist implementation of dietary interventions to benefit public health.

In Chapter 6, the impact of spatial variations in dietary patterns and diet cost on obesity will be explored before going on to investigate impact on breast cancer incidence in Chapter 7.

# Chapter 6: Geography of dietary patterns and diet cost in the UKWCS - implications for health: obesity as a case study

## 6.1 Overview

Chapter 6 will explore the impact of spatial variations in the cost of dietary patterns on obesity in the UKWCS. Figure 6.1 is a reminder of where this chapter fits into the thesis flow.

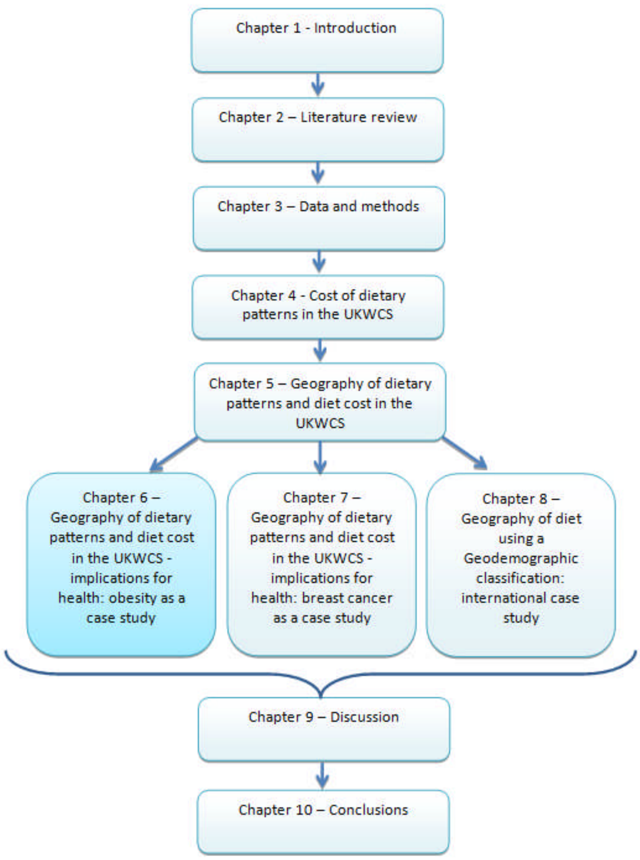



Figure 6.1 - Overview showing how Chapter 6 fits into the overall thesis flow

## 6.2 Introduction

Obesity is a huge public health concern, which poses both personal and societal health issues and is an economic burden. Prevalence of obesity is at a record high in the UK, with 26 % of women and 24% of men in England obese in 2011 (Health & Social Care Information Centre, 2013c). Prevalence has been increasing rapidly since the 1970s. The Foresight report, in 2007, predicted that obesity was likely to continue to increase unless serious changes are made at both an individual and population level (Foresight, 2007). While the rate at which obesity is increasing is beginning to decelerate (Rokholm et al., 2010), obesity is an enormous problem which needs drastic action to reverse.

Obesity is defined by the World Health Organisation (WHO) as: "abnormal or excessive fat accumulation that may impair health"(World Health Organisation, 2013). It can be measured using a number of techniques, of which Body Mass Index (BMI) is the most widely used (Hu, 2008). BMI is calculated by dividing an individual's weight (kg) by the square of their height (m), which accounts for differences in weight by height, but is unable to take into account body composition of an individual. However, for large scale epidemiological studies, the use of BMI is cost effective and convenient (Hu, 2008). The WHO has published cut points to determine categories of BMI in order to facilitate consistent use of terms such as overweight and obese: underweight <18.5 $kg/m^2$, normal weight 18.5 $kg/m^2$ to 24.9 $kg/m^2$, overweight 25 $kg/m^2$ to 30 $kg/m^2$ and obese >30 $kg/m^2$ (World Health Organisation, 2013).

Causes of obesity are a network of complex individual, social, cultural and environmental interactions which require multifaceted change in order to reverse the problem. The complexity of the issue is summarised by Figure 6.2, the Obesity system map, from the 2007 Foresight report. The obesity system map goes further than the ecological model to display how many factors interact to influence obesity prevalence. The main message to take from this diagram is that influences on obesity are complex. Research in this Chapter will touch on aspects from a number of areas of this map.

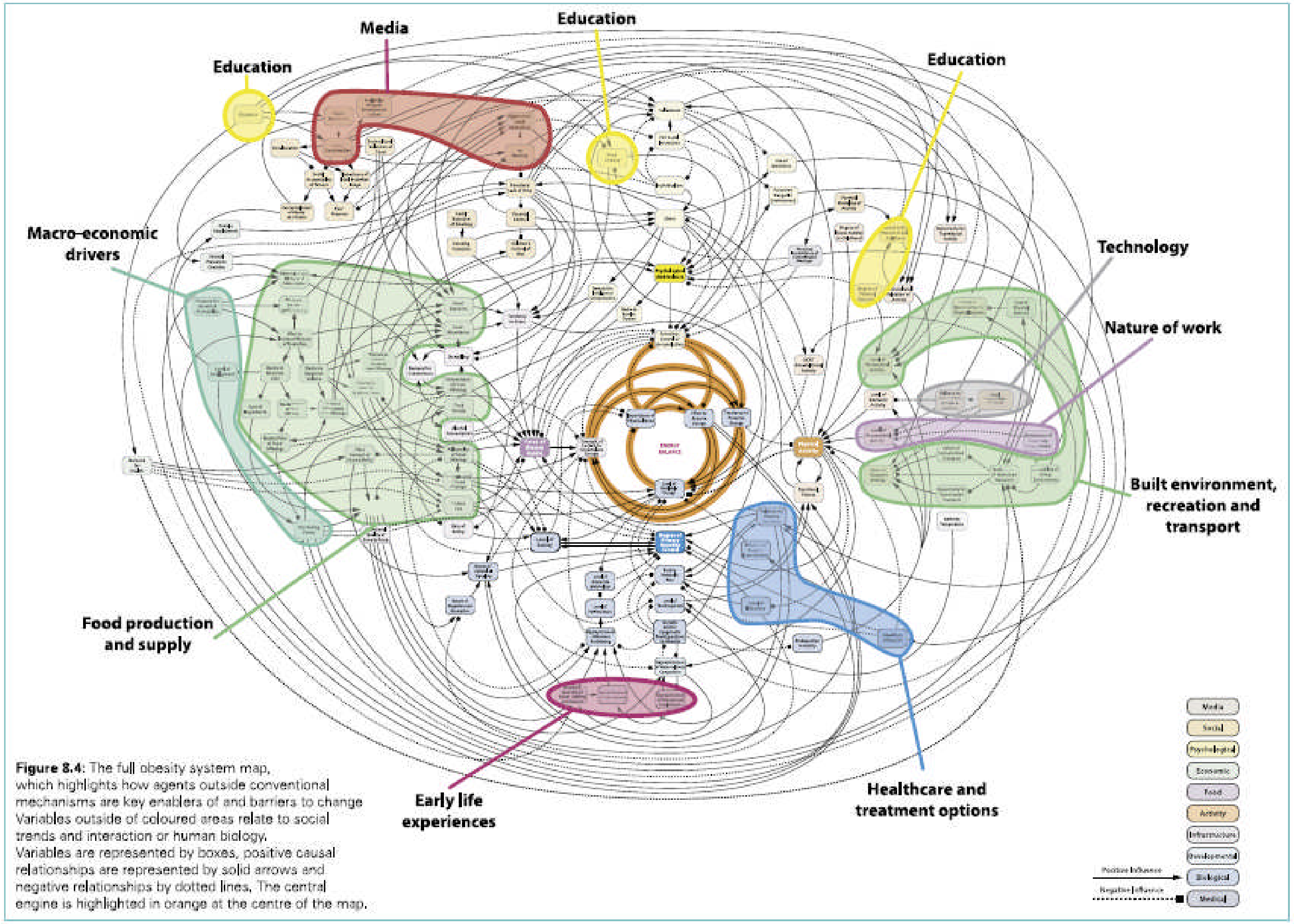



Figure 6.2 - The full system obesity map highlighting complexity of causes of obesity (Foresight, 2007)

Most simplistically, obesity is a result of being in a state of positive energy balance for a prolonged period of time, where energy consumed is greater than energy expended. Therefore, obesity is often associated with an energy dense diet and a sedentary lifestyle (Prentice and Jebb, 1995), although this can be to different extremes depending upon the individual. The dramatic increase in overweight and obesity in recent years cannot be due to genetic changes alone, due to too few generations having passed since the 'outbreak' of obesity. While ultimately an individual has a choice over what to eat and whether or not to exercise, there is a complex interplay between individual, social, cultural and environmental factors contributing to behaviour choice. The reasons that obesity rates vary, in that not everyone has responded to the environment in such a way that they become obese, are well understood to be "people's latent biological susceptibility interacting with a changing environment" (Foresight, 2007).

Swinburn et al (2011), discuss an "energy balance flipping point" at which time energy intake and expenditure were knocked out of balance around the early 1970s. It is not surprising, given the changes to the environment, that an increase in overweight and obesity is widespread. With the development of an infrastructure conducive of cars, but not necessarily of active transport, a decrease in manual labour employment and technology advances making television and computers available to everyone, the result has been an environment which promotes sedentary behaviours and energy dense food consumption (Hill and Peters, 1998). With many of these changes, time has also become a restricting factor in that people will often choose convenience foods which involve less preparation time (Jabs and Devine, 2006) and are often more energy dense.

Obesity occurs along a socioeconomic gradient, with those of higher socioeconomic status displaying less obesity. This is contrary to patterns we have seen historically, where only the most affluent had the money to purchase enough food to eat and become obese. Research has shown that once GDP reaches USA $5000 per head, there is no longer a positive linear relationship, but one which is flat (Swinburn et al., 2011). In fact the opposite is seen, where obesity is more prevalent in those of lower socioeconomic status (Wilkinson and Pickett, 2010, Offer et al., 2012, Drewnowski and Specter, 2004). In developing countries, a problem of under nutrition is also accompanied by over nutrition, seeing instances of both malnutrition and obesity in the same population, and worldwide there are more obese than underweight (World Health Organisation, 2000, Delpeuch et al., 2009).

Considering the diversity of residential neighbourhoods in the UK, and the fact that obesogenic environments are common, it is not surprising that obesity rates can vary according to where an individual lives. National statistics for obesity are published by GOR in the UK (Health & Social Care Information Centre, 2011). These are updated annually with results from the Health Survey for England (HSE) (See Table 6.3). For obesity data at a more granular geographic scale than Strategic Health Authority (SHA) (which are almost congruent with GOR, with the exception of 2 SHAs which together make up one GOR), synthetic estimates must be used which are based on the HSE 2006-2008 synthetic estimates. HSE has collected measured height and weight data for adults annually since 1991. This is considered to be the most robust measure of adult obesity in the UK. However, the HSE data displays the caveat “Data not sufficiently robust to measure geographical boundaries smaller than SHA” (Public Health England, 2013b). The annual sample size in recent years has included approximately 8000 adults, which are representative at the GOR level (Health & Social Care Information Centre, 2013a). For more granular geographic data, the National Obesity Observatory uses synthetic estimates from the 2006-2008 HSE statistics, simulated by combining with demographic data from the 2001 census.

Following initiation of the National Child Measurement Programme in 2006 (Health & Social Care Information Centre, 2013b) detailed and more current data for child obesity rates, with geographical location, are also now available. However, such data are not available for adults. There are plans to use self reported data from the Active People Survey (Sport England, 2013) - but this is subject to methodological testing to assess the reliability of self reported height and weight estimates.

Often obesity is associated with other chronic diseases such as: hypertension, type 2 diabetes, some cancers, cardiovascular disease, depression, sleep apnoea, osteoarthritis and others - through an accumulation of excess body fat which can lead to organ specific pathological consequences through various mechanisms. Being overweight increases risk of such diseases, risk that continues to increase with increasing BMI (Crawford et al., 2010). Despite the risks and potential causes of obesity being well documented (Foresight, 2007, Gibney, 2004, Swinburn et al., 2011, Wang et al., 2011, Hu, 2008, Gortmaker et al., 2011, Hall et al., 2011, Pearce and Witten, 2010), obesity remains a disease with no quick fix. More interdisciplinary research is required in order to cover combined aspects of influence on obesity. To understand the determinants of (and possible interventions to reduce) the prevalence of obesity and how risk factors interact also needs to be better understood so that they can be more effectively addressed by policy makers.

Using the available data to present descriptive statistics means that some generalisation occurs due to the large geographical scale. This may average out potential differences in areas of varying socioeconomic status and rurality. Synthetic estimates are a beneficial alternative where data are missing, but there could be better use made of data which are collected for large cohort studies such as the UKWCS. While such data are also subject to reporting bias, the large sample and wide geographical spread of women provides some confidence in the sample. The UKWCS is an appropriate population to use to answer the following research question: does overweight/obesity vary according to where women live in the UK Women's Cohort? It could be hypothesised that there will be a spatial variation in overweight/obesity in the UKWCS, with those in the South, Rural, Countryside and City Living Areas less likely to be overweight/obese.

### 6.2.1 Objectives

The objectives for this Chapter, as described in Chapter 1, Table 1.1 are to:

- Describe overweight/obesity in the UKWCS using GOR, North South divide, Urban and Rural and OAC Supergroup.
- Describe cost of diet by weight status and also by spatial measures.
- Investigate associations between overweight/obesity, diet cost and GOR, North South divide, Urban and Rural and OAC in the UKWCS, using both a continuous measure of BMI and by dichotomising the sample as normal weight or overweight/obese.

## 6.3 Methods

### 6.3.1 Weight status

BMI in this cohort is calculated from self reported height and weight in the baseline survey, collected between 1995 and 1998. The effect of BMI was investigated using BMI as a continuous measure and also using the WHO BMI categories, as described in Chapter 3, section 3.2.1.2.

### 6.3.2 Diet cost

Cost of diet was calculated by adding average prices from the DANTE food cost database to FFQ diet records for the UKWCS women, as described in Chapter 4, section 4.3.2.

### 6.3.3 Spatial analysis

Four spatial measures are used for this analysis (described in Chapter 3, section 3.2.4):

1. Government office region (GOR) - includes the nine GORs of England plus Scotland and Wales
2. North - South divide - includes data for England only
3. Urban - Rural living - includes England, Scotland and Wales
4. Output Area Classification (OAC) geodemographic classification - Supergroup (top layer of the classification hierarchy) and Group (middle layer) - For England, Scotland and Wales

### 6.3.4 Statistical analysis

All statistical analysis have been completed using Stata statistical software: version 12 (StataCorp, 2012, StataCorp, 2009). Due to the multiple testing which occurs in this analysis, statistical significance is considered at $p<0.01$.

#### *6.3.4.1 Descriptive statistics*

Descriptive statistics are presented by spatial measure for the percentage of women falling within the four BMI categories and mean BMI. Difference in variation by spatial measures was tested using $chi^2$ or a t-test as appropriate. Descriptive statistics for OAC Group (the middle layer of OAC hierarchy) are presented in a separate table.

UKWCS obesity prevalence results are presented alongside statistics for all women in the UK for comparison purposes.

#### *6.3.4.2 Diet cost and BMI*

Correlation between daily diet cost and BMI - as a continuous measure and using WHO categories - was tested using Pearson product moment and Spearman's rank correlation respectively. A 'test for trend' was conducted across the four BMI categories.

#### *6.3.4.3 Spatial measures and overweight/obese*

The UKWCS women were categorised into two groups: (1) normal weight and (2) overweight/obese. Underweight women were excluded from the analysis (n=726). Daily diet cost was presented by spatial measure for these two categories and difference between these values tested using $chi^2$ or a t-test as appropriate.

Logistic regression assessed whether the spatial measures predict relative risk of being overweight/obese compared to normal weight at each spatial measure.

Reference categories for each spatial measure were assigned to the area with the largest number of women residing within, with the exception of OAC Supergroup, where the Typical Traits category is used as reference as these have the most average demographic characteristics (see Table 6.1).

| **Spatial measure** | **Reference category** |
|---|---|
| GOR | South East |
| North - South | South |
| Urban - Rural | Urban |
| OAC Supergroup | Typical Traits |

Table 6.1 - Reference categories for regression models

The reference category for GOR analysis, the South East, has the largest population within the cohort and also, the South East Strategic Health Authority has the lowest obesity prevalence in the UK.

There are three regression models for each spatial measure. Model 1 looks at how well the spatial measure predicts overweight/obesity with no adjustments. Model 2 adjusts for dietary patterns (described in Chapter 3, section 3.2.1.4) and daily cost of diet (described in Chapter 4, section 4.3.2). Model 3 adjusts for energy intake, physical activity, dietary pattern, daily cost of diet, smoking, menopause status, education and social class. For models where OAC Supergroup is the predictor, education and social class are not included as these variables are already accounted for in the OAC and could cause colinearity problems.

#### *6.3.4.4 Spatial measures and BMI*

In order to quantify whether a dose response relationship exists for spatial measures and BMI, linear regression was carried out with the same spatial measures as predictor variables (GOR, North South, Urban/Rural and OAC) and a continuous measure of BMI as the outcome. Three models were produced with adjustments as described above.

## 6.4 Results

### 6.4.1 UKWCS compared to UK women

Less overweight and obesity is observed in the UKWCS compared to women of the same age during the same time period for the whole of the UK. There was 9% less overweight and 6% less obesity in the UKWCS. Prevalence of underweight was the same, with a higher proportion of UKWCS women being of normal weight (Table 6.2). This is also the case when looking at variation in obesity by GOR (Table 6.3).

| **BMI Category** | **Weight status for UK women 1995-1998 (% )** (Public Health England, 2013b) | **Weight status for UKWCS (%)** |
|---|---|---|
| Underweight | 2.1 | 2.2 |
| Normal weight | 45.9 | 60.7 |
| Overweight | 32.9 | 24.2 |
| Obese | 19.2 | 13.0 |

Table 6.2 - BMI category variation in UK women compared to UKWCS

| **Government office region** | **Obesity prevalence for UK Women 1998-2000 (%)** | **Obesity prevalence for UKWCS (%)** |
|---|---|---|
| North East | 19.8 | 13.7 |
| North West | 21.0 | 13.9 |
| Yorkshire and the Humber | 20.5 | 13.8 |
| East Midlands | 24.3 | 13.6 |
| West Midlands | 23.5 | 14.0 |
| East of England | 19.6 | 12.3 |
| Greater London | 20.2 | 12.6 |
| South East | 17.7 | 12.3 |
| South West | 18.8 | 12.3 |

Table 6.3 - Obesity prevalence from Public Health England Adult Obesity Prevalence data for Women (Public Health England, 2013a)

Results suggest that there is a spatial variation in prevalence of overweight and obesity, but less obesity variation exists in the UKWCS by GOR than compared to the whole UK, suggesting more homogeneity in UKWCS women.

### 6.4.2 Spatial measures and weight status

Weight status by BMI category varies by spatial measure. Figure 6.3 shows the percentage of women in each BMI category by four spatial measures: GOR, North South, Urban/Rural and OAC Supergroup. Narrow confidence intervals exist around the percentages in each BMI category providing assurance in the estimates, presented in Table 6.4.

Prevalence varies significantly ($chi^2$ test $p<0.01$) by GOR in all weight status categories, except obese (Table 6.4). The lowest percentage of overweight exists in Greater London (21.1%) and highest in East of England (26.6).The prevalence of obesity is homogenous across GORs in the UKWCS, with range only between 12.3% and 14% throughout GORs of England and Wales and Scotland in the UKWCS, which is statistically non-significant.

There is little variation seen for mean BMI by spatial measure, although analysis of variance tests show this is significant between groups. On closer examination, using ANOVA with Scheffe post hoc tests, significant differences appear between Southern and Northern GORs, suggestive of a North South divide with BMI being significantly ($p<0.01$) higher in the North East, North West, Yorkshire and the Humber, East Midlands, West Midlands and Scotland compared to Greater London. There is also a significantly higher BMI in the North West and Yorkshire and the Humber compared to the South East.

When the women are categorised as North or South, there is a significant difference, for each BMI category and also for mean BMI (Table 6.4).

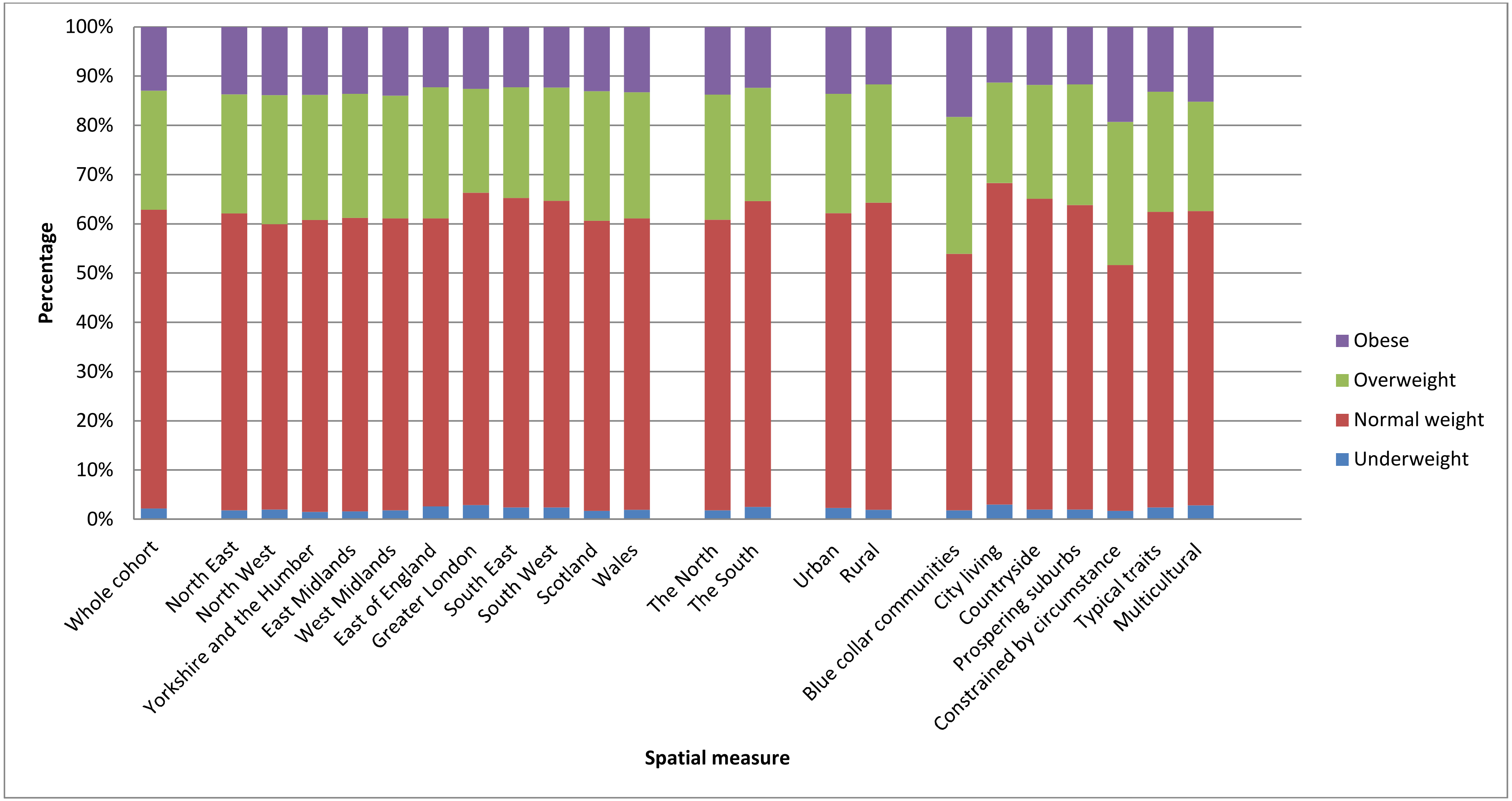



Figure 6.3 - Distribution of women in each BMI category according to spatial measure. Equivalent results with confidence intervals are displayed in Table 6.4

| | Underweight % (95% CI) | Normal weight % (95% CI) | Overweight % (95% CI) | Obese % (95% CI) | Mean BMI (95% CI) |
|---|---|---|---|---|---|
| Whole cohort (n=32784) | 2.2 (2.0 to 2.3) | 60.7 (60.2 to 61.3) | 24.2 (23.7 to 24.6) | 13.0 (12.6 to 13.3) | 24.4 (24.4 to 24.4) |
| **Government office region** | $Chi^2$ p=0.002 | $Chi^2$ p<0.001 | $Chi^2$ p<0.001 | $Chi^2$ p=0.202 | $Chi^2$ p<0.001 |
| North East (n=974) | 1.8 (1.0 to 2.7) | 60.3 (57.2 to 63.3) | 24.2 (21.5 to 26.9) | 13.7 (11.5 to 15.8) | 24.8 (24.5 to 25.1) |
| North West (n=3038) | 2.0 (1.5 to 2.5) | 57.9 (56.2 to 59.7) | 26.2 (24.6 to 27.8) | 13.9 (12.6 to 15.1) | 24.6 (24.5 to 24.8) |
| Yorkshire and the Humber (n=2561) | 1.5 (1.0 to 2.0) | 59.3 (23.7 to 27.1) | 25.4 (23.7 to 27.1) | 13.8 (12.4 to 15.1) | 24.7 (24.5 to 24.9) |
| East Midlands (n=2405) | 1.6 (1.0 to 2.1) | 59.6 (57.6 to 61.5) | 25.2 (23.5 to 27.0) | 13.6 (12.2 to 15.0) | 24.6 (24.4 to 24.8) |
| West Midlands (n=2534) | 1.8 (1.3 to 2.3) | 59.3 (57.4 to 61.1) | 24.9 (23.2 to 26.5) | 14.0 (12.7 to 15.4) | 24.6 (24.4 to 24.7) |
| East of England (n=3001) | 2.6 (2.0 to 3.1) | 58.5 (56.8 to 60.3) | 26.6 (25.0 to 28.2) | 12.3 (11.1 to 13.5) | 24.4 (24.3 to 24.6) |
| Greater London (n=3709) | 2.9 (2.3 to 3.4) | 63.4 (61.9 to 65.0) | 21.1 (19.7 to 22.3) | 12.6 (11.5 to 13.7) | 24.1 (23.9 to 24.2) |
| South East (n=6789) | 2.4 (2.0 to 2.7) | 62.9 (61.7 to 64.0) | 22.5 (21.5 to 23.5) | 12.3 (11.5 to 13.1) | 24.2 (24.1 to 24.3) |
| South West (n=4155) | 2.4 (1.9 to 2.9) | 62.2 (60.7 to 63.7) | 23.0 (21.8 to 24.3) | 12.3 (11.3 to 13.3) | 24.2 (24.1 to 24.4) |
| Scotland (n=2199) | 1.7 (1.1 to 2.2) | 58.9 (56.9 to 61.0) | 26.3 (24.5 to 28.2) | 13.1 (11.6 to 14.5) | 24.6 (24.4 to 24.8) |
| Wales (n=1419) | 1.9 (1.2 to 2.6) | 59.2 (56.6 to 61.8) | 25.6 (23.3 to 27.9) | 13.3 (11.5 to 15.1) | 24.6 (24.4 to 24.8) |
| **North South** | T test p<0.001 | T test p<0.001 | T test p<0.001 | T test p<0.001 | T test p<0.001 |
| The North (n=11512) | 1.8 (1.5 to 2.0) | 59.1 (58.2 to 60.0) | 25.4 (24.6 to 26.2) | 13.8 (13.2 to 14.4) | 24.6 (24.6 to 24.7) |
| The South (n=17654) | 2.5 (2.3 to 2.7) | 62.1 (61.4 to 62.8) | 23.0 (22.4 to 23.6) | 12.4 (11.9 to 12.9) | 24.2 (24.1 to 24.3) |
| **Urban Rural** | T test p=0.014 | T test p <0.001 | T test p=0.567 | T test p<0.001 | T test p<0.001 |
| Urban (n=21866) | 2.3 (2.1 to 2.5) | 59.9 (59.2 to 60.5) | 24.3 (23.7 to 24.8) | 13.6 (13.1 to 14.0) | 24.5 (24.4 to 24.6) |
| Rural (n=10843) | 1.9 (1.6 to 2.1) | 62.4 (61.5 to 63.3) | 24.0 (23.2 to 24.8) | 11.7 (11.1 to 12.3) | 24.2 (24.1 to 24.3) |
| **OAC Supergroup** | $Chi^2$ p=0.003 | $Chi^2$ p<0.001 | $Chi^2$ p<0.001 | $Chi^2$ p<0.001 | $Chi^2$ p<0.001 |
| Blue Collar Communities (n=2114) | 1.8 (1.2 to 2.4) | 52.1 (50.0 to 54.2) | 27.8 (25.9 to 29.7) | 18.3 (16.7 to 20.0) | 25.3 (25.1 to 25.6) |
| City Living (n=2216) | 3.0 (2.3 to 3.7) | 65.2 (63.2 to 67.2) | 20.4 (18.8 to 22.1) | 11.3 (10.0 to 12.6) | 23.8 (23.7 to 24.0) |
| Countryside (n=7518) | 2.0 (1.7 to 2.3) | 63.1 (62.0 to 64.2) | 23.1 (22.2 to 24.1) | 11.8 (11.0 to 12.5) | 24.2 (24.1 to 24.3) |
| Prospering Suburbs (n=11029) | 2.0 (1.7 to 2.2) | 61.8 (60.9 to 62.8) | 24.5 (23.7 to 25.3) | 11.7 (11.1 to 12.3) | 24.3 (24.3 to 24.4) |
| Constrained by Circumstance (n=1411) | 1.7 (1.0 to 2.4) | 49.9 (47.3 to 52.5) | 29.1 (26.8 to 31.5) | 19.3 (17.2 to 21.3) | 25.5 (25.2 to 25.8) |
| Typical Traits (n=6576) | 2.4 (2.1 to 2.8) | 60.0 (58.8 to 61.2) | 24.4 (23.3 to 25.4) | 13.2 (12.4 to 14.0) | 24.4 (24.3 to 24.5) |
| Multicultural (n=1981) | 2.8 (2.1 to 3.6) | 59.8 (57.6 to 61.9) | 22.2 (20.4 to 24.0) | 15.2 (13.6 to 16.8) | 24.4 (24.2 to 24.6) |

Table 6.4 - Percentage of UKWCS women in each of the WHO BMI categories by spatial measure

There is a significant difference in BMI by Urban/Rural with Urban areas having higher BMI (Table 6.4). Urban areas have a higher percentage of obese and a lower percentage of normal weight women. Underweight and overweight groups do not differ significantly.

Those in areas Constrained by Circumstance and of Blue Collar Communities have a higher percentage of overweight and obese than other types of areas with prevalence being lowest in City Living areas. When the OAC Supergroups are further subdivided into OAC Groups (see Figure 6.4), it can be seen that, the high percentage of overweight and obese comes mainly from the Terraced Blue Collar and Younger Blue Collar Groups, but the Older Blue Collar women are also above average for percentage of overweight and obese. All groups within the Constrained by Circumstance Supergroups have a high percentage of overweight/obese. Within the Prospering Suburbs Supergroup, the Prospering Semis Group have a much higher prevalence of both overweight and obese, a difference most prominent when compared to the Thriving Suburbs Group. In general, confidence intervals about the percentage of women in each BMI category are slightly wider in the granular OAC Groups than when looking at the Supergroup level (see Table 6.5). Mean BMI is lowest in Transient Communities (23.7 $kg/m^2$). The highest BMI can be seen in Younger Blue Collar and Public Housing (25.9 $kg/m^2$).

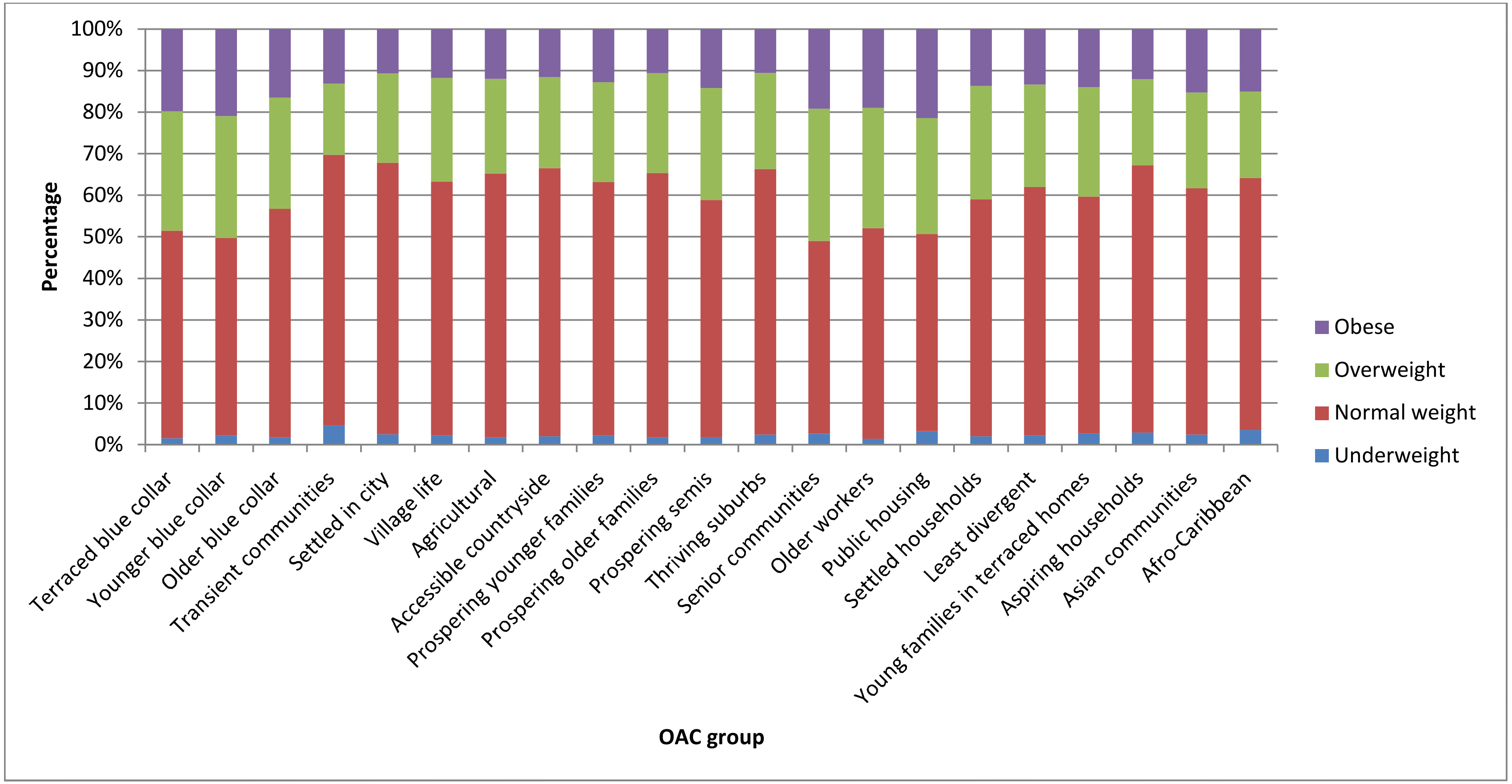



Figure 6.4 - Distribution of women in each BMI category according to OAC Group. Equivalent results with confidence intervals are displayed in Table 6.5.

| Group name | Underweight (%) (95% CI) | Normal weight (%) (95% CI) | Overweight (%) (95% CI) | Obese (%) (95% CI) | Mean BMI kg/m$^2$ (95% CI) | Cost of diet in £/day (95% CI) |
|---|---|---|---|---|---|---|
| 1A - Terraced Blue Collar (n=344) | 1.5 (0.2 to 2.7) | 50.0 (44.7 to 55.3) | 28.8 (24.0 to 33.6) | 19.8 (15.5 to 24.0) | 25.5 (25.0 to 26.0) | 4.39 (4.23 to 4.55) |
| 1B - Younger Blue Collar (n=600) | 2.2 (1.0 to 0.3) | 47.5 (43.5 to 51.5) | 29.3 (25.7 to 33.0) | 21.0 (17.7 to 24.3) | 25.9 (25.3 to 26.4) | 4.47 (4.34 to 4.60) |
| 1C - Older Blue Collar (n=1170) | 1.7 (1.0 to 2.5) | 55.0 (52.2 to 57.9) | 26.8 (24.2 to 29.3) | 16.5 (14.4 to 18.6) | 25.0 (24.7 to 25.3) | 4.47 (4.39 to 4.56) |
| 2A - Transient Communities (n=544) | 4.6 (2.8 to 6.4) | 65.1 (61.1 to 69.1) | 17.1 (13.9 to 20.3) | 13.2 (10.4 to 16.1) | 23.7 (23.3 to 24.0) | 4.16 (4.04 to 4.29) |
| 2B - Settled in City (n=1672) | 2.5 (1.8 to 3.3) | 65.3 (63.0 to 67.5) | 21.5 (19.6 to 23.5) | 10.7 (9.2 to 12.2) | 23.9 (23.7 to 24.1) | 4.39 (4.32 to 4.46) |
| 3A - Village Life (n=2457) | 2.2 (1.6 to 2.8) | 61.1 (59.2 to 63.0) | 24.9 (23.2 to 26.6) | 11.8 (10.6 to 13.1) | 24.3 (24.1 to 24.5) | 4.51 (4.45 to 4.56) |
| 3B - Agricultural (n=2136) | 1.7 (1.2 to 2.3) | 63.5 (61.5 to 65.6) | 22.8 (21.0 to 24.5) | 12.0 (10.6 to 13.4) | 24.1 (24.0 to 24.3) | 4.64 (4.57 to 4.70) |
| 3C - Accessible Countryside (n=2925) | 2.0 (1.5 to 2.5) | 64.5 (62.8 to 66.3) | 21.9 (20.4 to 23.4) | 11.6 (10.4 to 12.7) | 24.1 (23.9 to 24.2) | 4.57 (4.52 to 4.63) |
| 4A - Prospering Younger Families (n=1124) | 2.2 (1.4 to 3.1) | 60.9 (58.1 to 63.8) | 24.0 (21.5 to 26.5) | 12.8 (10.9 to 14.8) | 24.3 (24.0 to 24.5) | 4.45 (4.37 to 4.53) |
| 4B - Prospering Older Families (n=4196) | 1.7 (1.3 to 2.1) | 63.6 (62.1 to 65.0) | 24.0 (22.8 to 25.3) | 10.6 (9.7 to 11.6) | 24.3 (24.1 to 24.4) | 4.54 (4.50 to 4.58) |
| 4C - Prospering Semis (n=2632) | 1.8 (1.3 to 2.3) | 57.1 (55.2 to 59.0) | 27.0 (25.3 to 28.7) | 14.2 (12.8 to 15.5) | 24.8 (24.7 to 25.0) | 4.50 (4.45 to 4.56) |
| 4D - Thriving Suburbs (n=3077) | 2.4 (1.8 to 2.9) | 63.9 (62.2 to 65.6) | 23.1 (21.6 to 25.6) | 10.6 (9.5 to 11.7) | 24.1 (23.9 to 24.2) | 4.50 (4.45 to 4.55) |
| 5A - Senior Communities (n=151) | 2.6 (0.0 to 5.2) | 46.4 (38.3 to 54.4) | 31.8 (24.3 to 39.3) | 19.2 (12.8 to 25.5) | 25.8 (24.9 to 26.7) | 4.24 (4.00 to 4.48) |
| 5B - Older Workers (n=1106) | 1.4 (0.7 to 2.0) | 50.7 (47.8 to 53.7) | 28.9 (26.3 to 31.6) | 19.0 (16.7 to 21.3) | 25.4 (25.1 to 25.7) | 4.43 (4.34 to 4.52) |
| 5C - Public Housing (n=154) | 3.2 (0.4 to 6.0) | 47.4 (39.4 to 55.4) | 27.9 (20.8 to 35.1) | 21.4 (14.9 to 28.0) | 25.9 (25.0 to 26.7) | 4.31 (4.07 to 4.54) |
| 6A - Settled Households (n=1593) | 2.0 (1.3 to 2.7) | 57.0 (54.6 to 59.4) | 27.3 (25.1 to 29.5) | 13.7 (12.0 to 15.4) | 24.7 (24.5 to 24.9) | 4.42 (4.35 to 4.49) |
| 6B - Least Divergent (n=1994) | 2.2 (1.6 to 2.9) | 59.8 (57.6 to 61.9) | 24.6 (22.7 to 26.5) | 13.4 (11.9 to 14.9) | 24.5 (24.3 to 24.7) | 4.46 (4.39 to 4.52) |
| 6C - Young Families in Terraced Homes (n=1060) | 2.7 (1.8 to 3.7) | 57.0 (54.0 to 60.0) | 26.3 (23.7 to 29.0) | 14.0 (11.9 to 16.0) | 24.6 (24.3 to 24.8) | 4.35 (4.26 to 4.43) |
| 6D - Aspiring Households (n=1929) | 2.9 (2.2 to 3.7) | 64.3 (62.1 to 66.4) | 20.7 (18.9 to 22.5) | 12.1 (10.7 to 13.6) | 24.0 (23.9 to 24.2) | 4.44 (4.37 to 4.50) |
| 7A - Asian Communities (n=1245) | 2.4 (1.6 to 3.3) | 59.3 (56.5 to 62.0) | 23.1 (20.7 to 25.4) | 15.3 (13.3 to 17.3) | 24.5 (24.3 to 24.8) | 4.31 (4.23 to 4.39) |
| 7B - Afro-Caribbean (n=736) | 3.5 (2.2 to 4.9) | 60.6 (57.1 to 64.1) | 20.8 (17.8 to 23.7) | 15.1 (12.5 to 17.7) | 24.2 (23.9 to 24.6) | 4.19 (4.09 to 4.29) |

Table 6.5 - Percentage of UKWCS women in each of the WHO BMI categories by OAC Group.

In logistic regression analysis (see Table 6.6) odds ratios (ORs) show a statistically significant increase in risk of being overweight/obese in all areas - except the North East, Greater London and the South West - compared to those in the South East of England. When the women are dichotomised as living in the North or South, those in the North have a significantly increased risk of being overweight/obese compared to the South, a difference which remains in the adjusted models. Urban living is also associated with an increased risk of overweight/obese with those in rural areas having a 10% reduced risk of overweight/obesity compared to the urban reference group. For the OAC Supergroup analysis ORs show a 42% increased risk of being overweight/obese in Blue Collar Communities compared to areas of Typical Traits and 56% increased risk of being overweight/obese in Constrained by Circumstance. When the model is adjusted for dietary pattern and diet cost, the explained variance in the model increases and the effects of OAC Supergroup on prevalence of overweight/obesity relationships remain although slightly attenuated. Including further possible confounders and mediators in the model further increases the explained variance, with little effect on the ORs.

Interestingly social class was the only variable which had no significant effect on the regression models but was retained in the model for completeness (stepwise methods were not employed).

Linear regression analysis using a continuous measure of BMI concurs with the findings reported above, quantifying the unit increase in BMI which in most cases is only a small, but statistically significant difference (see Table 6.7). The largest unit increase in BMI is 0.61 kg/m$^2$ (95% CI 0.31 to 0.90) observed in the North East compared to the South East in the fully adjusted model. The only areas where BMI does not vary significantly from the South East are the other Southern areas; East of England, Greater London and the South West.

The regression models in Table 6.6 and Table 6.7, the $R^2$ values are relatively small suggesting that they all explain less than 5% of variation in overweight/obesity and BMI respectively.

| | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | **OR (95% CI)** | **P value** | **OR (95% CI)** | **P value** | **OR (95% CI)** | **P value** |
| **Government office region** | Pseudo $R^2$= 0.002 | | Pseudo $R^2$=0.018 | | Pseudo $R^2$= 0.035 | |
| North East | 1.14 (0.99 to 1.31) | 0.074 | 1.11 (0.97 to 1.28) | 0.139 | 1.12 (0.96 to 1.31) | 0.139 |
| North West | **1.25 (1.14 to 1.37)** | **<0.001** | **1.23 (1.12 to 1.34)** | **<0.001** | **1.27 (1.15 to 1.40)** | **<0.001** |
| Yorkshire and the Humber | **1.19 (1.09 to 1.31)** | **<0.001** | **1.18 (1.07 to 1.29)** | **0.001** | **1.19 (1.07 to 1.32)** | **0.001** |
| East Midlands | **1.18 (1.07 to 1.30)** | **0.001** | **1.16 (1.05 to 1.28)** | **0.002** | **1.17 (1.05 to 1.30)** | **0.004** |
| West Midlands | **1.19 (1.08 to 1.30)** | **<0.001** | **1.17 (1.06 to 1.29)** | **0.001** | **1.20 (1.09 to 1.34)** | **<0.001** |
| East of England | **1.20 (1.10 to 1.31)** | **<0.001** | **1.19 (1.09 to 1.30)** | **<0.001** | **1.19 (1.08 to1.32)** | **<0.001** |
| Greater London | 0.96 (0.88 to 1.04) | 0.334 | 1.02 (0.94 to 1.11) | 0.651 | 1.05 (0.96 to 1.16) | 0.268 |
| South East | 1.00 | - | 1.00 | | 1.00 | |
| South West | 1.03 (0.95 to 1.11) | 0.533 | 1.03 (0.95 to 1.12) | 0.483 | 1.01 (0.92 to 1.11) | 0.829 |
| Scotland | **1.21 (1.09 to 1.33)** | **<0.001** | **1.18 (1.06 to 1.30)** | **0.002** | **1.27 (1.14 to 1.42)** | **<0.001** |
| Wales | **1.19 (1.05 to 1.34)** | **0.005** | **1.18 (1.05 to 1.34)** | **0.006** | **1.20 (1.05 to 1.37)** | **0.006** |
| **North South** | Pseudo $R^2$= 0.001 | | Pseudo $R^2$= 0.017 | | Pseudo $R^2$= 0.035 | |
| The North | **1.16 (1.11 to 1.22)** | **<0.001** | **1.13 (1.08 to 1.19)** | **<0.001** | **1.15 (1.09 to 1.22)** | **<0.001** |
| The South | 1 | - | 1 | - | 1 | - |
| **Urban Rural** | Pseudo $R^2$= 0.0004 | | Pseudo $R^2$= 0.017 | | Pseudo $R^2$= 0.036 | |
| Urban | 1 | - | 1 | - | | - |
| Rural | **0.91 (0.86 to 0.95)** | **<0.001** | **0.89 (0.85 to 0.94)** | **<0.001** | **0.90 (0.85 to 0.95)** | **<0.001** |
| **OAC Supergroup** | Pseudo $R^2$= 0.005 | | Pseudo $R^2$= 0.021 | | Pseudo $R^2$= 0.033 | |
| Blue Collar Communities | **1.41 (1.28 to 1.56)** | **<0.001** | **1.33 (1.20 to 1.47)** | **<0.001** | **1.36 (1.23 to 1.52)** | **<0.001** |
| City Living | **0.78 (0.70 to 0.86)** | **<0.001** | **0.82 (0.74 to 0.91)** | **<0.001** | **0.81 (0.73 to 0.90)** | **<0.001** |
| Countryside | **0.88 (0.82 to 0.95)** | **<0.001** | **0.86 (0.80 to 0.92)** | **<0.001** | **0.85 (0.79 to 0.92)** | **<0.001** |
| Prospering Suburbs | 0.93 (0.88 to 1.00) | 0.035 | **0.88 (0.83 to 0.94)** | **<0.001** | **0.86 (0.80 to 0.92)** | **<0.001** |
| Constrained by Circumstance | **1.55 (1.38 to 1.74)** | **<0.001** | **1.47 (1.31 to 1.66)** | **<0.001** | **1.47 (1.30 to 1.66)** | **<0.001** |
| Typical Traits | 1.00 | - | 1.00 | - | 1.00 | - |
| Multicultural | 1.00 (0.90 to 1.11) | 0.984 | 1.08 (0.97 to 1.20) | 0.176 | 1.09 (0.97 to 1.21) | 0.136 |

Table 6.6 - Logistic regression investigating the odds of being overweight/obese compared to normal weight by spatial measure. Model 1 = unadjusted. Model 2 = adjusted for diet cost and dietary pattern. Model 3= adjusted for Energy intake, energy expenditure, diet cost, dietary pattern, smoking, menopause, education and class. OR=Odds ratio CI=Confidence Interval

| | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | **Coef (95% CI)** | **P value** | **Coef (95% CI)** | **P value** | **Coef (95% CI)** | **P value** |
| **Government office region** | $R^2$ = 0.0027 | | $R^2$ =0.0323 | | $R^2$ =0.0491 | |
| North East | **0.60 (0.30 to 0.89)** | **<0.001** | **0.53 (0.24 to 0.82)** | **<0.001** | **0.61 (0.31 to 0.90)** | **<0.001** |
| North West | **0.45 (0.26 to 0.64)** | **<0.001** | **0.39 (0.20 to 0.58)** | **<0.001** | **0.43 (0.24 to 0.62)** | **<0.001** |
| Yorkshire and the Humber | **0.52 (0.32 to 0.72)** | **<0.001** | **0.47 (0.27 to 0.67)** | **<0.001** | **0.48 (0.28 to 0.68)** | **<0.001** |
| East Midlands | **0.41 (0.20 to 0.61)** | **<0.001** | **0.36 (0.16 to 0.56)** | **0.001** | **0.38 (0.18 to 0.59)** | **<0.001** |
| West Midlands | **0.37 (0.17 to 0.57)** | **<0.001** | **0.33 (0.13 to 0.53)** | **0.001** | **0.33 (0.13 to 0.53)** | **0.001** |
| East of England | 0.22 (0.03 to 0.41) | 0.022 | 0.19 (0.00 to 0.38) | 0.047 | 0.16 (-0.02 to 0.35) | 0.089 |
| Greater London | -0.12 (-0.30 to 0.05) | 0.171 | 0.03 (-0.14 to 0.23) | 0.711 | 0.05 (-0.12 to 0.23) | 0.549 |
| South East | 0 | - | 0 | - | 0 | - |
| South West | 0.06 (-0.11 to 0.22) | 0.515 | 0.06 (-0.11 to 0.23) | 0.473 | 0.06 (-0.11 to 0.23) | 0.488 |
| Scotland | **0.40 (0.19 to 0.62)** | **<0.001** | **0.31 (0.10 to 0.52)** | **0.004** | **0.39 (0.18 to 0.61)** | **<0.001** |
| Wales | **0.42 (0.17 to 0.67)** | **0.001** | **0.40 (0.15 to 0.65)** | **0.002** | **0.44 (0.18 to 0.69)** | **0.001** |
| **North South** | $R^2$ = 0.0023 | | $R^2$ =0.0319 | | $R^2$ =0.0489 | |
| The North | **0.43 (0.32 to 0.53)** | **<0.001** | **0.35 (0.24 to 0.45)** | **<0.001** | **0.37 (0.27 to 0.48)** | **<0.001** |
| The South | 0 | - | 0 | - | 0 | - |
| **Urban Rural** | $R^2$= 0.0076 | | $R^2$= 0.0317 | | $R^2$= 0.0481 | |
| Urban | 0 | - | 0 | - | 0 | - |
| Rural | **-0.27 (-0.38 to -0.17)** | **<0.001** | **-0.30 (-0.40 to -0.20)** | **<0.001** | **-0.30 (-0.40 to -0.20)** | **<0.001** |
| **OAC Supergroup** | $R^2$= 0.0076 | | $R^2$= 0.0370 | | $R^2$= 0.0544 | |
| Blue Collar Communities | **0.90 (0.68 to 1.11)** | **<0.001** | **0.73 (0.52 to 0.95)** | **<0.001** | **0.80 (0.58 to 1.02)** | **<0.001** |
| City Living | **-0.58 (-0.79 to -0.67)** | **<0.001** | **-0.43 (-0.64 to -0.23)** | **<0.001** | **-0.44 (-0.65 to -0.23)** | **<0.001** |
| Countryside | **-0.26 (-0.41 to -0.12)** | **<0.001** | **-0.32 (-0.46 to -0.18)** | **<0.001** | **-0.35 (-0.50 to -0.21)** | **<0.001** |
| Prospering Suburbs | -0.07 (-0.21 to 0.06) | 0.280 | **-0.21 (-0.35 to -0.08)** | **0.002** | **-0.30 (-0.43 to -0.16)** | **<0.001** |
| Constrained by Circumstance | **1.09 (0.84 to 1.35)** | **<0.001** | **0.96 (0.71 to 1.21)** | **<0.001** | **0.94 (0.69 to 1.20)** | **<0.001** |
| Typical Traits | 0 | - | 0 | - | 0 | - |
| Multicultural | -0.11 (-0.23 to 0.21) | 0.922 | 0.16 (-0.06 to 0.38) | 0.147 | 0.18 (-0.05 to 0.40) | 0.119 |

Table 6.7 - Linear regression showing how spatial measures predict BMI, with coefficients quantifying the unit change in BMI. Model 1 = unadjusted. Model 2 = adjusted for diet cost and dietary pattern. Model 3= adjusted for Energy intake, energy expenditure, diet cost, dietary pattern, smoking, menopause, education and class. OR=Odds ratio CI=Confidence Interval

### 6.4.3 Spatial measure, diet cost and weight status

The mean daily diet cost for: underweight, normal weight, overweight and obese women were £4.33, £4.45, £4.52 and £4.50 respectively. These differences were statistically significant ($p<0.001$ using ANOVA). Post hoc testing identified that the significant differences lay between overweight and underweight; overweight and normal weight; obese and underweight. There is a significant ($p<0.001$) trend of increasing diet cost with increasing BMI category. A weak but significant positive correlation exists between BMI and daily diet cost ($r=0.042$ $p<0.001$). This is also the case with BMI categories ($rho=0.0196$ $p<0.001$). Cost of diet also varies significantly ($p<0.001$) between categories of each spatial measure.

When considering differences in the cost of diet between those who are normal weight compared to the overweight/obese by spatial measure, in most cases the overweight/obese have a more expensive diet (see Table 6.8). These differences are, however, only significant in: Urban (£0.08/day more expensive) compared to Rural areas; in the South (£0.07/day more expensive) compared to the North; and in Greater London (£0.18/day more expensive) and the East Midlands (£0.16/day more expensive) compared to the South East. When considering difference in cost by OAC Supergroup none of the differences are statistically significant ($p<0.01$).

| | Cost of diet in £/day (95% CI) | Diet cost for Normal Weight £/day (95% CI) | Diet cost for Overweight/Obese £/day (95% CI) | Mean Difference (£/day) in cost for normal weight and overweight/obese | T-test P value for difference in cost for normal weight and overweight/obese |
|---|---|---|---|---|---|
| **Whole cohort** | 4.47 (4.46 to 4.49) | 4.45 (4.43 to 4.47) | 4.51 (4.49 to 4.54) | 0.06 | <0.001 |
| **Government office region** | | | | | |
| North East | 4.58 (4.49 to 4.68) | 4.58 (4.46 to 4.70) | 4.59 (4.45 to 4.74) | 0.01 | 0.885 |
| North West | 4.46 (4.41 to 4.51) | 4.48 (4.41 to 4.53) | 4.44 (4.36 to 4.53) | -0.04 | 0.521 |
| Yorkshire and the Humber | 4.51 (4.46 to 4.57) | 4.49 (4.42 to 4.56) | 4.53 (4.44 to 4.62) | 0.04 | 0.483 |
| East Midlands | 4.57 (4.51 to 4.63) | 4.51 (4.44 to 4.58) | 4.67 (4.57 to 4.77) | 0.16 | **0.009** |
| West Midlands | 4.46 (4.41 to 4.51) | 4.45 (4.38 to 4.52) | 4.50 (4.41 to 4.58) | 0.05 | 0.387 |
| East of England | 4.47 (4.42 to 4.52) | 4.46 (4.40 to 4.52) | 4.49 (4.40 to 4.58) | 0.03 | 0.529 |
| Greater London | 4.26 (4.21 to 4.31) | 4.20 (4.15 to 4.26) | 4.38 (4.29 to 4.46) | 0.18 | **<0.001** |
| South East | 4.44 (4.41 to 4.48) | 4.44 (4.39 to 4.48) | 4.47 (4.42 to 4.53) | 0.04 | 0.287 |
| South West | 4.52 (4.47 to 4.56) | 4.50 (4.45 to 4.56) | 4.55 (4.47 to 4.56) | 0.05 | 0.358 |
| Scotland | 4.65 (4.58 to 4.71) | 4.62 (4.54 to 4.70) | 4.68 (4.57 to 4.79) | 0.06 | 0.358 |
| Wales | 4.51 (4.43 to 4.58) | 4.52 (4.42 to 4.61) | 4.50 (4.38 to 4.62) | -0.02 | 0.800 |
| **North South** | | | | | |
| The North | 4.51 (4.48 to 4.53) | 4.49 (4.46 to 4.52) | 4.53 (4.49 to 4.58) | 0.04 | 0.106 |
| The South | 4.43 (4.41 to 4.45 | 4.41 (4.38 to 4.43) | 4.48 (4.44 to 4.51) | 0.07 | **0.002** |
| **Urban Rural** | | | | | |
| Urban | 4.42 (4.40 to 4.44) | 4.39 (4.37 to 4.42) | 4.47 (4.44 to 4.50) | 0.08 | **<0.001** |
| Rural | 4.57 (4.55 to 4.61) | 4.57 (4.53 to 4.60) | 4.61 (4.56 to 4.66) | 0.04 | 0.143 |
| **OAC Supergroup** | | | | | |
| Blue Collar Communities | 4.46 (4.39 to 4.53) | 4.39 (4.30 to 4.48) | 4.53 (4.43 to 4.64) | 0.14 | 0.042 |
| City Living | 4.33 (4.27 to 4.40) | 4.32 (4.24 to 4.39) | 4.39 (4.28 to 4.40) | 0.07 | 0.281 |
| Countryside | 4.57 (4.54 to 4.60) | 4.57 (4.53 to 4.61) | 4.58 (4.53 to 4.64) | 0.01 | 0.694 |
| Prospering Suburbs | 4.51 (4.48 to 4.54) | 4.49 (4.46 to 4.52) | 4.55 (4.51 to 4.60) | 0.06 | 0.023 |
| Constrained by Circumstance | 4.40 (4.32 to 4.48) | 4.38 (4.27 to 4.48) | 4.42 (4.31 to 4.54) | 0.04 | 0.563 |
| Typical Traits | 4.42 (4.39 to 4.46) | 4.40 (4.36 to 4.45) | 4.47 (4.41 to 4.53) | 0.07 | 0.060 |
| Multicultural | 4.26 (4.20 to 4.33) | 4.20 (4.12 to 4.28) | 4.36 (4.26 to 4.47) | 0.16 | 0.016 |

Table 6.8 - Diet cost differences between normal weight and the overweight/obese by spatial measure

## 6.5 Discussion

### 6.5.1 Spatial variation in obesity

Differences exist despite less overweight and obesity prevalence in the cohort compared to the UK as a whole, which may reflect the above average social class of the women in the UKWCS (Cade et al., 2004). Overweight and obesity are significantly higher in the North than the South of the UK, which is as hypothesised. It is interesting to see this difference remains even when controlling for all confounders and mediators - including social class and education - especially when variation was less evident when grouping by GOR. This is as hypothesised and in line with other spatial obesity research (Moon et al., 2007, Scarborough and Allender, 2008). Scarborough and Allender (2008), from an analysis of HSE data, also showed increased odds of being overweight or obese in the North compared to South for middle aged women. It may be that obesity and related conditions, contribute to the observed difference in mortality between North and South too (Shaw et al., 1998). Obesity varies significantly between Urban and Rural areas with higher percentage of obese in the Urban environments, consistent with existing literature in the UK (Riva et al., 2009), although no variation is seen for overweight women in the UKWCS. When the women are grouped by OAC Supergroup, there are large differences observed in percentages of both overweight and obese women, supporting evidence that inclusion of demographic characteristics aids the understanding of groups more or less likely to be overweight/obese. These results are in line with results of earlier Chapters. As observed in Chapter 5, $R^2$ and pseudo $R^2$ values, as an indicator of model fit, of OAC models compared to GOR models adjusting for key demographic variables are very similar, yet the OAC models highlight greater variation.

Considering BMI as a continuous measure, rather than by category, more significant variation is observed by GOR, showing that while the women may not be classified as overweight or obese, they do have higher BMIs in Northern GORs, as seen when the sample is dichotomised to North or South. We still see that Rural areas are less likely to be overweight or obese, with a reduced BMI coefficient compared to the Urban reference group. These results show slightly different patterns compared to the categorical results, as they take into account those who fall into the normal weight category but have higher BMI.

Use of either categorisation or BMI as a continuous measure shows a strong relationship that those in Blue Collar Communities and Constrained by Circumstance are significantly more likely to be overweight/obese than those in areas of Typical Traits. The City Living, Countryside and

Prospering Suburb groups are significantly less likely to be overweight/obese with negative coefficients for BMI. No difference observed between the Typical Traits and Multicultural areas may be a result of the fact that while the UKWCS have women living in Multicultural areas, 99% of the UKWCS women report ethnicity as White. The ethnic characteristics which typify Multicultural areas do not exist in the UKWCS which may explain why less difference is observed than would be expected.

### 6.5.2 Diet cost

In a similar vein to the dietary pattern analysis in Chapter 5, the largest variation in diet cost by weight status is observed in Greater London, with normal weight women spending £0.18 per day less than overweight/obese women. This difference is indicative of the diverse types of dietary pattern consumed in this region. In the East Midlands the overweight/obese women also pay £0.16 per day more, which is a statistically significant difference. However, in the East Midlands there are no significant differences in dietary pattern consumption. Chapter 5 showed in unadjusted regression analysis, women in Greater London paid significantly less for their diets while the East Midlands paid significantly more. It is interesting to see that despite the differences in daily diet cost, the overweight/obese women in each of these regions pay significantly more than the normal weight women.

In the UKWCS, the North of England has a higher percentage of overweight and obese and a more expensive diet. Urban areas have higher overweight and obese than rural, yet rural areas consume a more expensive diet. Differences in diet cost between normal weight and overweight/obese are higher and significantly different in Urban and Southern areas indicating that patterns are not consistent in different geographical regions. No significant difference in cost of diet between normal weight and overweight/obese women is observed by OAC group, but the direction of difference is always that the overweight/obese pay more. These findings are contrary to the existing evidence that energy dense, nutrient poor diets, which promote obesity are most expensive as in this study the overweight/obese spend the most on their food.

While differences in overweight/obesity by spatial measure do not necessarily correlate with diet cost, women who are overweight/obese generally pay more for their diet. Only in the North West is a cheaper diet recorded for overweight/obese women, although this difference is not statistically significant.

In regression analysis controlling for diet cost and dietary pattern, this does very little to alter the odds of being overweight/obese for any spatial measure, although does increase the model fit to explain 3.5% of the variation. This may suggest that the area in which you live is a stronger influence on obesity status than diet pattern or cost.

### 6.5.3 Strengths and limitations

There is strength in the size of the sample and wide geographic spread of the women's locations, despite the sample not being designed to be geographically representative. The sample size far outweighs the sample sizes of other national surveys which collect height and weight data such as the NDNS, which has data on ~3000 adults over three years of data collection (Department of Health and Food Standards Agency, 2012) and the Health Survey for England which collects data on ~8000 per year (Health & Social Care Information Centre, 2013a). No synthetic estimates were required to increase the sample size. However, using self reported height and weight to calculate BMI - as the UKWCS does - can introduce systematic bias into data as heavier individuals are more likely to underestimate their weight and likewise short individuals are more likely to overestimate their height (Hu, 2008).

Underweight individuals were excluded from the logistic regression analysis and normal weight used as the reference group. These underweight women constitute only a small proportion (~2%) of the study population, but may provide important information if underweight and food insecurity were the research question of interest. In future work it would be interesting to look into higher prevalence of underweight observed in Transient Communities, Public Housing and Afro-Caribbean OAC Groups.

While results at each of the geographical scales add to understanding the issue of overweight and obesity in the UK, their use in application for public health policy needs some consideration. When planning how best to implement an intervention to reach those most in need in a cost effective manner, a combination of approaches may be most beneficial. For example, based on the results of this research, to reach middle aged women with the highest odds of being overweight or obese in the UK it would be most beneficial to target those in Constrained by Circumstance groups in the North West of England and in Scotland.

### 6.5.4 Policy implications

Adding identifiers such as North South, Urban/Rural and OAC to survey datasets, like the NDNS (Department of Health and Food Standards Agency, 2012) as standard could be a useful tool for researchers. The Family Spending Survey presents results using a range of identifiers, including Urban/Rural, OAC, GOR and some reference to the North South divide.

## 6.6 Summary and context

This chapter has shown that a spatial variation in overweight and obesity exist in the UK and within the UKWCS.

- Areas with the highest prevalence of overweight/obese do not necessarily have the most expensive diet, but within a spatial measure, the overweight/obese pay more for their diet.
- Urban areas have higher BMI and are more likely to be overweight/obese than Rural areas.
- Northern areas have higher BMI and are more likely to be overweight/obese than Southern areas.
- Those in areas Constrained by Circumstance and of Blue Collar Communities have a higher percentage of overweight and obese than other types of areas with prevalence being lowest in City Living areas.
- It would seem sensible to consider geographic measures and combinations of spatial measures to best target public health interventions to prevent and reduce overweight and obesity.

In Chapter 7, variation in breast cancer incidence by spatial measure, dietary pattern, diet cost and weight status will be explored.

# Chapter 7: Geography of dietary patterns and diet cost in the UKWCS and implications for health: breast cancer as a case study

## 7.1 Overview

This Chapter will explore the impact of spatial variations in the cost of dietary patterns on breast cancer incidence in the UKWCS. Figure 7.1 is a reminder of where this chapter fits into the thesis flow.

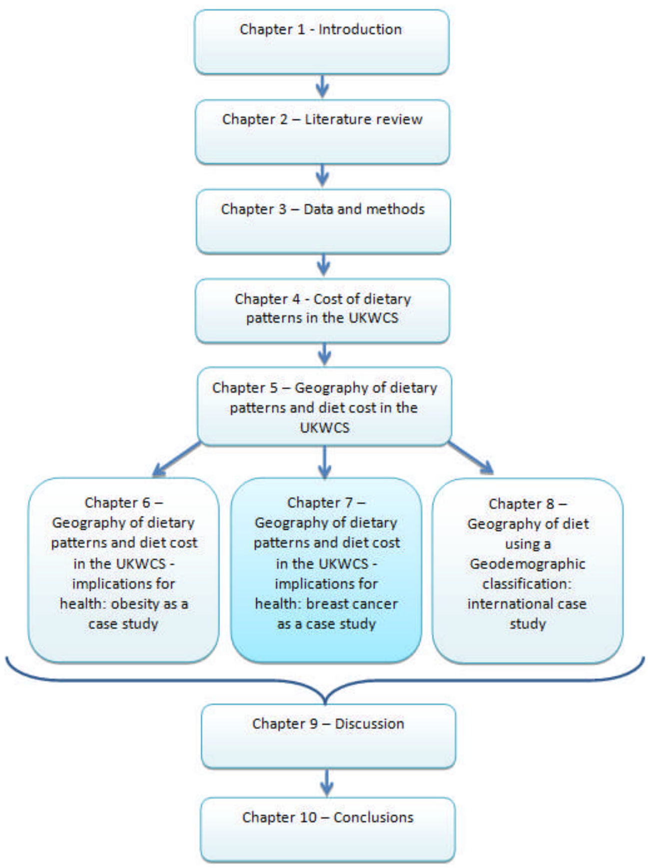



Figure 7.1 - Overview showing how Chapter 7 fits into the overall thesis flow

## 7.2 Introduction

Breast cancer is the most common type of cancer in the UK (Cancer Research UK, 2012, Office for National Statistics, 2007), despite the fact that it predominantly affects women. There were over 49564 new cases in women in the UK during 2010 and 397 in men. These women range in age, ethnicity, socioeconomic status (SES) and menopausal state, but patterns of association exist, for example: breast cancer incidence and age are strongly associated, with 80% of 2008-2010 diagnoses being in the over 50s and 45% over 65 (Cancer Research UK, 2012). This may be linked to menopausal status which changes around this age. Other hormonal factors are also associated with breast cancer incidence, including age at menarche (Baanders and de Waard, 1992), parity, breast feeding (Akbari et al., 2010), mothers age at birth of first child, use of hormonal contraceptives (Calle et al., 1996) and hormone replacement therapy (Collaborative Group on Hormonal Factors in Breast Cancer, 1997).

Evidence shows that affluent women are more likely to develop breast cancer (Brown et al., 2007, Schrijvers et al., 1995) than those of a lower SES. These women are also more likely to present at an earlier stage of breast cancer than women from more deprived areas (Adams et al., 2004). Suggested reasons why affluent women are more likely to develop breast cancer can be linked to better educational level (Heck and Pamuk, 1997) and also hormonal related factors such as breast feeding and mothers' age at the birth of first child. Research also suggests that taller women are at an increased risk (Waard, 1975) and that this may also be linked to a higher SES, through improved nutritional status in infanthood (Baanders and de Waard, 1992).

Lifestyle factors such as alcohol consumption, smoking status, physical activity level, diet and obesity are all risk factors in the development of breast cancer (Danaei et al., 2005). These unhealthy lifestyle factors are typically more prevalent in the least affluent. This introduces an interesting pattern with highest incidence in both the poorest and most affluent, with a 'gap' of lower incidence in those of average affluence. Research in Scotland suggests that breast cancer incidence is increasing in all socioeconomic groups, but that lower incidence in those of average affluence remains (Brown et al., 2007). However, other research in Denmark, using a longitudinal cohort, shows that breast cancer incidence is rising at a greater pace in the least affluent and subsequently closing the gap in incidence created by SES (Dano et al., 2003).

Breast cancer also provides an interesting example of risk related to obesity. Some research suggests that obesity prior to menopause can be protective against breast cancer, whilst post menopausally obesity poses increased risk (Key et al., 2002) (WCRF, 2007). This relationship is likely mediated by increased conversion of androgen to oestradiol in adipose tissue which is more abundant in the obese.

Many studies have investigated relationships between specific aspects of diet and breast cancer. For example; high fibre intake (Cade et al., 2007) and high folate intake (Ericson et al., 2007) decrease the risk of breast cancer, whilst red meat consumption (Taylor et al., 2007) and alcohol consumption increase risk (Burley et al., 2010). Often, very little association is found with dietary components and breast cancer risk (Hutchinson et al., 2012, Key et al., 2011). Some studies have investigated complete diet or diet patterns and their association with breast cancer. A meta-analysis up to 2009 has found some evidence of decreased breast cancer risk with an increasingly healthy diet (Brennan et al., 2010), but this is not the case in all studies (Cade et al., 2011). The relationship with alcohol consumption and breast cancer risk is the most convincing (Key et al., 2002) .

Geographic variation in breast cancer incidence is observed at regional level in the UK and Ireland (Figure 7.2). The bars denoting the confidence intervals do not often overlap, demonstrating that there is a real difference for 95% of the population, although only small. This is suggestive of the fact that there could be environmental risk factors for the development of breast cancer. Differences in environment may relate to context (geographic location) or composition (social makeup). An example of the latter is Ireland - where there is a lower incidence of breast cancer - there are a high proportion of women practicing the Catholic religion. Being pregnant is protective against breast cancer and given that Catholics typically have large families due to beliefs relating to contraception; religion seems a feasible explanation for this difference, thus highlighting the influence of environmental composition.

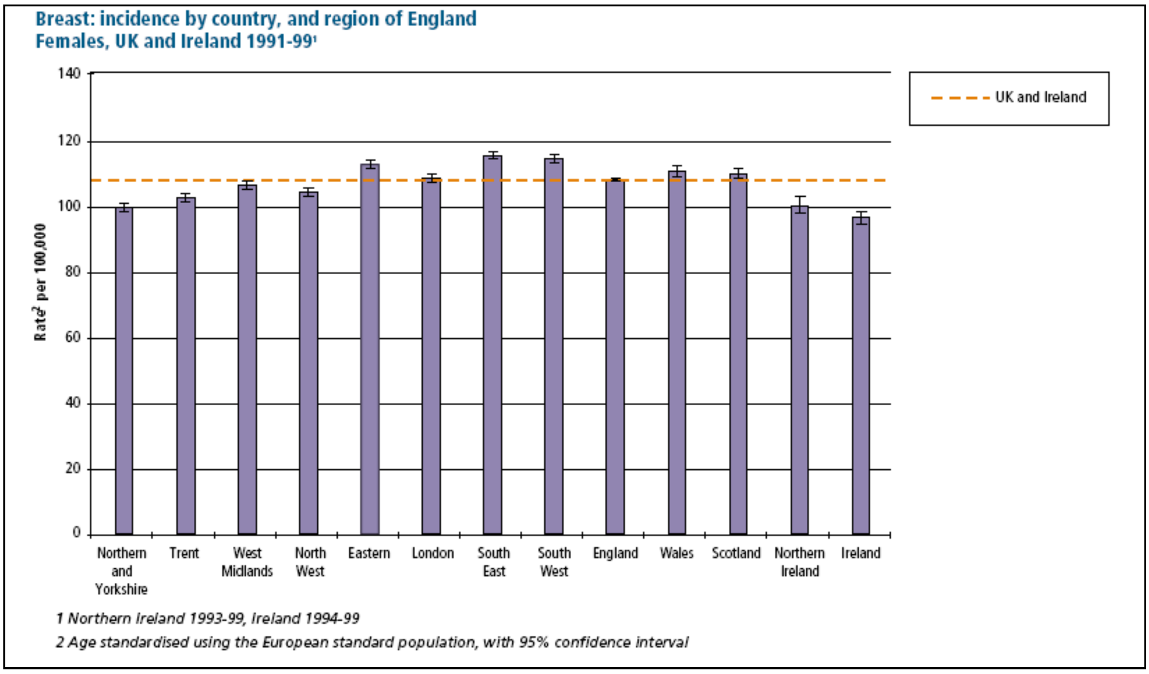



Figure 7.2 - Variation in UK Breast Cancer Incidence by region (Office for National Statistics, 2007)

When drilling down the geographical scale to health authority (a smaller geographical unit) variation exists even within a region. Incidence of breast cancer in Ireland and Northern Ireland remains the lowest, with some health authorities much lower than others. Breast cancer incidence according to Geodemographic classification, or specifically, the type of area in which a person lives, is not readily available in peer reviewed journals. Some evidence can be found in health authority publications which profile age adjusted cancer incidence by Geodemographic group (Robinson et al., 2010).

Research including cost of diet is a relatively new area, but evidence has shown that a healthy diet is a more expensive diet (see Chapter 4), so we may infer that consuming such a diet would decrease risk. However, evidence also shows that breast cancer is more prevalent in more affluent women, who are more likely to have the means to pay for a more expensive diet leading us to suggest that perhaps a more expensive diet would be associated with increased risk of breast cancer. To date, there is no research in the UK tying together spatial variations, dietary patterns and cost of diet with breast cancer incidence.

### 7.2.1 Objectives

The objectives for this Chapter, as described in Chapter 1, Table 1.1 are to:

- To describe demographic and diet characteristics of UKWCS women according to whether they have developed breast cancer or not.
- To investigate whether there is a difference in the observed and expected breast cancer incidence in a time to event analysis by predictor variables: GOR, North South, Urban/Rural, dietary pattern, diet cost and weight status.
- To investigate the hazard ratio for breast cancer incidence by predictor variables: GOR, North South, Urban/Rural, dietary pattern, diet cost and weight status in order to quantify risk.

## 7.3 Methods

Cancer incidence for the women within the UKWCS is reported from the NHSIC to the Nutritional Epidemiology group on an approximately quarterly basis. As at October 2012, 1445 of the women in the cohort had developed breast cancer since the baseline (between 1995 and 1998). Descriptive statistics are presented for the women according to whether or not they developed breast cancer, and t-tests and $chi^2$ tests are used to detect differences between groups.

Time to event analysis is used to investigate whether there is an association between spatial diet cost patterns and breast cancer incidence. This methodology is explained in Chapter 3, section 3.3.3. Observed verses expected outcomes were compared using log rank test for difference. Kaplan Meier survival curves were used to display the probability of remaining breast cancer free in different categories of exposure and to test the assumptions of Cox regression. The four spatial scales defined in Chapter 3, section 3.2.4, and used in Chapters 5 and 6 were also used in this analysis. Dietary patterns were described in Chapters 3, section 3.2.1.4. Diet cost has been described in Chapter 4.

Cox regression (described in Chapter 3, section 3.3.3) was performed to calculate hazard ratios for breast cancer risk. This allows breast cancer incidence risk to be quantified. Three models were run for each main predictor variable: GOR, North South, Urban/Rural, dietary pattern, diet cost and weight status. Model 1 includes the main predictor variable and controls for age. Model 2 is minimally adjusted including age, spatial, diet cost and dietary pattern variables. Model 3 includes all variables detailed in the causal diagram in Chapter 3, section 3.3.2. Subgroup analysis was performed using menopausal status as reported by the women at baseline.

Statistical significance is considered when $p<0.01$ to account for effects of multiple testing.

## 7.4 Results

In this study 1445 out of 35372 women developed breast cancer during a mean time to follow up of 14.6 years. Cases prevalent at baseline were excluded. There are 14 incident breast cancer cases which occur after the censor date and 1962 women who did not fully complete the FFQ and were excluded from all analysis leaving a sample with 1344 incident breast cancers for survival analysis. Of the 1962 women who did not complete the FFQ, 87 developed breast cancer.

### 7.4.1 Characteristics of the UKWCS by breast cancer incidence

Menopausal status, BMI category and dietary pattern consumption are significantly different between cases and non-cases, with no significant difference in educational level and daily diet cost. BMI as a continuous measure and social class difference are approaching levels of significance (see Table 7.1).

| | **Case** | **Non-case** | **Test for difference** |
|---|---|---|---|
| Social class | % (95% CI) | % (95% CI) | p=0.061 ($Chi^2$) |
| Professional | 61 (58 to 63) | 63 (63 to 64) | |
| Intermediate | 30 (27 to 32) | 27 (27 to 28) | |
| Routine/Manual | 9 (7 to 10) | 9 (9 to 9) | |
| NK | 1 | 1 | |
| Education | % (95% CI) | % (95% CI) | p=0.163 ($Chi^2$) |
| No education | 18 (16 to 20) | 17 (16 to 17) | |
| O-level | 30 (28 to 33) | 31 (31 to 32) | |
| A-level | 26 (24 to 29) | 24 (24 to 25) | |
| Degree | 25 (23 to 28) | 28 (27 to 28) | |
| Mean Daily diet cost £ (95% CI) | 4.43 (4.36 to 4.50) | 4.48 (4.46 to 4.49) | p=0.268 (t test) |
| Dietary pattern | % (95% CI) | % (95% CI) | p=**0.008** ($Chi^2$) |
| Monotonous Low Quantity Omnivore | 15 (13 to 17) | 16 (16 to 16) | |
| Health Conscious | 6 (5 to 7) | 6 (6 to 7) | |
| Traditional Meat, Chips and Pudding Eater | 18 (16 to 21) | 18 (18 to 18) | |
| Higher Diversity Traditional Omnivore | 15 (14 to 17) | 14 (14 to 15) | |
| Conservative Omnivore | 20 (18 to 23) | 17 (17 to 18) | |
| Low Diversity Vegetarian | 13 (11 to 15) | 15 (15 to 16) | |
| High Diversity Vegetarian | 11 (10 to 13) | 13 (13 to 13) | |
| Menopause status | % (95% CI) | % (95% CI) | p**<0.001** ($Chi^2$) |
| Postmenopausal | 57 (54 to 60) | 52 (51 to 52) | |
| Premenopausal | 43 (40 to 46) | 48 (48 to 49) | |
| Mean BMI $kg/m^2$ (95% CI) | 24.7 (24.5 to 24.9) | 24.4 (24.3 to 24.4) | p=0.012 (t test) |
| BMI category | % (95% CI) | % (95% CI) | p=**0.001** ($Chi^2$) |
| Underweight | 1 (1 to 2) | 2 (2 to 2) | |
| Normal weight | 57 (24 to 59) | 61 (60 to 61) | |
| Overweight | 28 (25 to 30) | 24 (24 to 24) | |
| Obese | 14 (14 to 16) | 13 (13 to 13) | |

Table 7.1 - Characteristics of breast cancer cases and non-cases

#### *7.4.1.1 Log rank tests for difference*

Log rank tests for difference for time to event analysis show significant differences in the observed verses expected cases in GOR, dietary pattern and BMI category. The South West and East Midlands have more observed cases than expected, compared to less observed than expected cases in the South East and Scotland.

No differences were observed when investigating breast cancer risk by quintiles of diet cost. For dietary pattern, there are less observed than expected cases in the Low Diversity and High Diversity Vegetarian patterns. Conversely, most notably, more cases were observed than expected in women consuming a Conservative Omnivore dietary pattern. In underweight and normal weight women fewer cases were observed than expected, yet in the overweight and obese the opposite was true (see Table 7.2).

| Predictor | | Observed/Expected | Log rank test for difference |
|---|---|---|---|
| GOR | | | p=0.060 |
| | North East | 34/39 | |
| | North West | 102/122 | |
| | Yorkshire and the Humber | 110/102 | |
| | East Midlands | 110/96 | |
| | West Midlands | 103/101 | |
| | East of England | 123/122 | |
| | Greater London | 161/150 | |
| | South East | 257/273 | |
| | South West | 191/167 | |
| | Scotland | 67/89 | |
| | Wales | 60/57 | |
| North South | | | p=0.612 |
| | The North | 459/468 | |
| | The South | 732/723 | |
| Urban Rural | | | p=0.843 |
| | Urban | 877/880 | |
| | Rural | 440/437 | |
| OAC | | | p=0.218 |
| | Blue Collar Communities | 66/85 | |
| | City Living | 91/90 | |
| | Countryside | 310/301 | |
| | Prospering Suburbs | 465/442 | |
| | Constrained by Circumstance | 53/55 | |
| | Typical Traits | 243/264 | |
| | Multicultural | 89/80 | |
| Quintiles of daily diet cost | | | p=0.372 |
| | 1 - Lowest cost | 265/271 | |
| | 2 | 256/272 | |
| | 3 | 283/263 | |
| | 4 | 284/268 | |
| | 5 – Highest cost | 253/267 | |
| Dietary pattern | | | p=0.003 |
| | Monotonous Low Quantity Omnivore | 205/210 | |
| | Health Conscious | 80/85 | |
| | Traditional Meat, Chips and Pudding Eater | 248/238 | |
| | Higher Diversity Traditional Omnivore | 208/189 | |
| | Conservative Omnivore | 274/232 | |
| | Low Diversity Vegetarian | 176/212 | |
| | High Diversity Vegetarian | 153/178 | |
| BMI Category | | | p<0.001 |
| | Underweight | 20/29 | |
| | Normal weight | 761/825 | |
| | Overweight | 370/319 | |
| | Obese | 193/171 | |

Table 7.2 - Log rank test results for predictor variables showing observed verses expected outcomes and test for difference between these measures

### 7.4.2 Breast cancer incidence by spatial measure

There is some variation in breast cancer incidence by spatial measure. These differences are significant by GOR and OAC Supergroup, but not for North South or Urban/Rural. For example, In the South West incident breast cancer is seen in 4.7% of the population, compared to only 3% in Scotland. Blue Collar Communities show 3.2% incident cases compared to 4.5 % in Multicultural Supergroups (see Table 7.3).

| **Spatial measure (n)** | **Incident breast cancer as % of spatial measure (95% CI)** |
|---|---|
| Whole cohort (33410) | 4.1 (3.9 to 4.3) |
| Government office region | p<**0.001** (Chi$^2$) |
| North East (n=974) | 3.6 (2.4 to 4.8) |
| North West (n=3038) | 3.4 (2.7 to 4.0) |
| Yorkshire and the Humber (n=2561) | 4.3 (3.5 to 5.0) |
| East Midlands (n=2405) | 4.6 (3.7 to 5.4) |
| West Midlands (n=2534) | 4.1 (3.4 to 4.9) |
| East of England (n=3001) | 4.1 (3.4 to 4.8) |
| Greater London (n=3709) | 4.3 (3.7 to 5.0) |
| South East (n=6789) | 3.8 (3.4 to 4.3) |
| South West (n=4155) | 4.7 (4.1 to 5.4) |
| Scotland (n=2199) | 3.0 (2.3 to 3.8) |
| Wales (n=1419) | 4.0 (3.0 to 5.0) |
| North South | p=0.476 (t test) |
| The North (n=11512) | 4.0 (3.7 to 4.4) |
| The South (n=17654) | 4.2 (3.9 to 4.5) |
| Urban Rural | p=0.688 (t test) |
| Urban (n=21913) | 4.0 (3.8 to 4.3) |
| Rural (n=10867) | 4.1 (3.7 to 4.5) |
| OAC Supergroup | p<**0.001** (Chi$^2$) |
| Blue Collar Communities (n=2114) | 3.2 (2.5 to 4.0) |
| City Living (n=2216) | 4.2 (3.3 to 5.0) |
| Countryside (n=7518) | 4.2 (3.7 to 4.6) |
| Prospering Suburbs (n=11029) | 4.2 (3.9 to 4.6) |
| Constrained by Circumstance (n=1411) | 3.8 (2.8 to 4.7) |
| Typical Traits (n=6576) | 3.8 (3.3 to 4.2) |
| Multicultural (n=1981) | 4.5 (3.6 to 5.5) |

Table 7.3 - Percentage of incident breast cancer cases at each spatial measure

The OAC Supergroup in which the women reside has less effect than weight status on their probability of staying breast cancer free, with lines closer together (Figure 7.3). However, it does illustrate the differences observed, that Multicultural groups have the lowest probability of staying breast cancer free, and the Blue Collar Communities the highest.

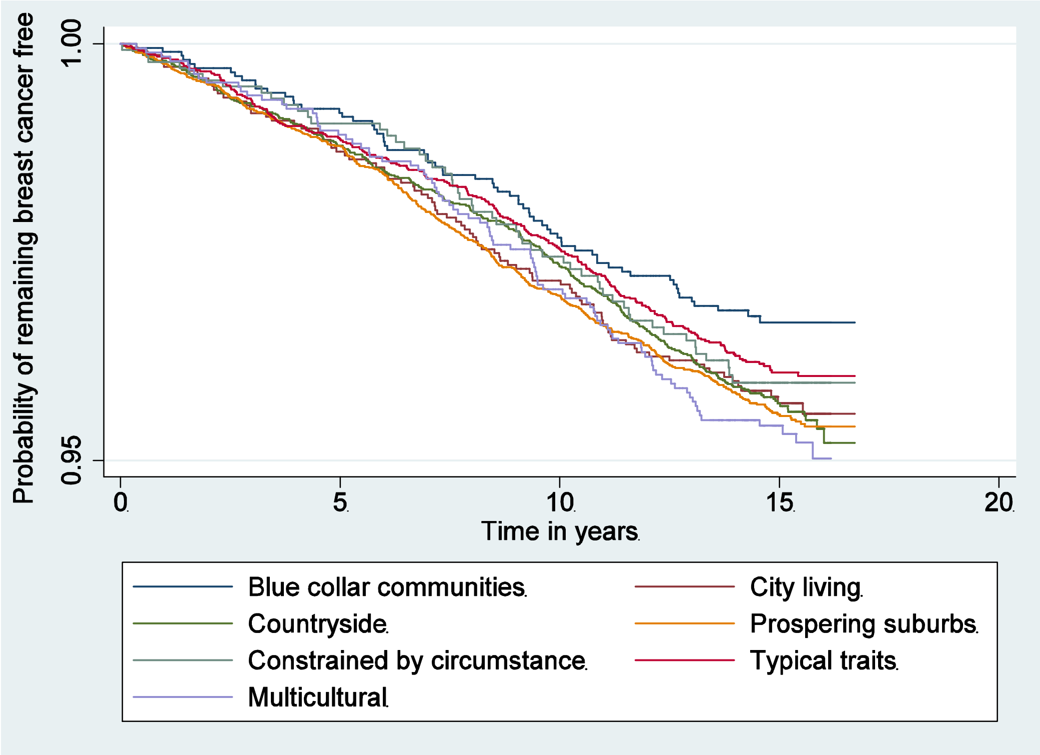



Figure 7.3 - Kaplan-Meier survival curve for probability of remaining breast cancer free by OAC Supergroup for all UKWCS women

In Cox regression models, adjusted for age only, no significant variation in breast cancer incidence is observed by any spatial measure. Increased risk of breast cancer, at a level approaching significance, is observed in the East of England and South West compared to women in the reference category, the South East (Model 1 Table 7.4). This pattern remains in the adjusted models (Models 2 and 3 Table 7.4).

In menopausal status subgroup analysis, increased risk, at levels approaching significance, is observed in the East of England, which is observed in all models. Pre-menopausal women are at increased risk, at levels approaching significance when they reside in Prospering Suburbs and Multicultural areas compared to the reference category Typical Traits. These differences remain in the adjusted models Table 7.5.

Postmenopausally, there is an increased risk of developing breast cancer, at levels approaching significance, living in Yorkshire and Humber, the South West and Wales compared to the South East reference group. These differences remain in Model 2, adjusting for age, dietary pattern and diet cost for all these regions, but only for the South West and Wales in the fully adjusted model (see Table 7.6).

| | Cases/Non-cases | Model 1<br>HR (95% CI) | P value | Model 2<br>HR (95% CI) | P value | Model 3<br>HR (95% CI) | P value |
|---|---|---|---|---|---|---|---|
| Government office region | | | | | | | |
| North East (n=974) | 34/940 | 0.93 (0.65 to 1.33) | 0.689 | 0.94 (0.66 to 1.34) | 0.724 | 0.78 (0.48 to 1.25) | 0.305 |
| North West (n=3038) | 102/2936 | 0.89 (0.71 to 1.12) | 0.338 | 0.89 (0.71 to 1.12) | 0.333 | 0.96 (0.72 to 1.27) | 0.769 |
| Yorkshire and the Humber (n=2561) | 110/2451 | 1.15 (0.92 to 1.44) | 0.217 | 1.16 (0.93 to 1.45) | 0.195 | 1.12 (0.84 to 1.49) | 0.433 |
| East Midlands (n=2405) | 110/2295 | 1.24 (0.99 to 1.55) | 0.062 | 1.24 (0.99 to 1.55) | 0.058 | 1.26 (0.95 to 1.66) | 0.105 |
| West Midlands (n=2534) | 103/2431 | 1.09 (0.86 to 1.36) | 0.479 | 1.09 (0.87 to 1.37) | 0.470 | 1.14 (0.85 to 1.51) | 0.384 |
| East of England (n=3001) | 123/2878 | 1.08 (0.87 to 1.34) | 0.480 | 1.08 (0.87 to 1.34) | 0.473 | 1.21 (0.94 to 1.57) | 0.139 |
| Greater London (n=3709) | 161/3548 | 1.16 (0.95 to1.41) | 0.149 | 1.17 (0.96 to 1.42) | 0.125 | 1.08 (0.82 to 1.41) | 0.593 |
| South East (n=6789) | 257/6532 | 1 | | 1 | | 1 | |
| South West (n=4155) | 191/3964 | 1.20 (1.00 to 1.45) | **0.054** | 1.21 (1.00 to 1.46) | **0.044** | 1.33 (1.05 to 1.67) | **0.016** |
| Scotland (n=2199) | 67/2132 | 0.80 (0.61 to 1.04) | 0.099 | 0.80 (0.61 to 1.05) | 0.107 | 0.99 (0.72 to 1.36) | 0.932 |
| Wales (n=1419) | 60/1359 | 1.13 (0.85 to 1.50) | 0.396 | 1.14 (0.86 to 1.51) | 0.350 | 1.05 (0.73 to 1.51) | 0.802 |
| North South | | | | | | | |
| The North (n=11512) | 459/11053 | 0.98 (0.87 to 1.10) | 0.681 | 0.98 (0.87 to 1.10) | 0.692 | 0.95 (0.82 to 1.10) | 0.522 |
| The South (n=17654) | 732/16922 | 1 | | | | | |
| Urban Rural | | | | | | | |
| Urban (n=21036) | 877/21036 | 1 | | | | | |
| Rural (n=10867) | 440/10427 | 1.00 (0.90 to 1.13) | 0.928 | 0.99 (0.89 to 1.12) | 0.932 | 0.94 (0.82 to 1.09) | 0.422 |
| OAC Supergroup | | | | | | | |
| Blue Collar Communities (n=2144) | 66/2048 | 0.85 (0.64 to 1.11) | 0.227 | 0.83 (0.63 to 1.09) | 0.184 | 0.83 (0.62 to 1.11) | 0.210 |
| City Living (n=2216) | 91/2125 | 1.10 (0.86 to 1.40) | 0.438 | 1.11 (0.87 to 1.41) | 0.406 | 1.09 (0.85 to 1.40) | 0.512 |
| Countryside (n=7518) | 310/7208 | 1.10 (0.93 to 1.30) | 0.290 | 1.08 (0.91 to 1.30) | 0.368 | 1.07 (0.90 to 1.28) | 0.438 |
| Prospering Suburbs (n=11029) | 465/10564 | 1.12 (0.96 to 1.31) | 0.141 | 1.11 (0.95 to 1.29) | 0.205 | 1.13 (0.96 to 1.33) | 0.130 |
| Constrained by Circumstance (n=1411) | 53/1358 | 1.01 (0.75 to 1.35) | 0.974 | 1.01 (0.75 to 1.36) | 0.968 | 1.07 (0.79 to 1.45) | 0.669 |
| Typical Traits (n=6576) | 243/6333 | 1 | | 1 | | 1 | |
| Multicultural (n=1981) | 89/1892 | 1.26 (0.99 to 1.60) | **0.065** | 1.26 (0.99 to 1.61) | **0.065** | 1.17 (0.90 to 1.52) | 0.237 |

Table 7.4 - Cases/non-cases and results of Cox proportional hazards regression models for the UKWCS at each spatial scale. Model 1 is age adjusted only. Model 2 is adjusted for age, dietary pattern and diet cost. Model 3 is adjusted for age, dietary pattern, diet cost, energy intake, energy expenditure, smoking status, parity, education*, social class* age at menarche, age at first child, ethanol consumption, height and weight.*Education and social class are not included in the OAC Supergroup models

| | Cases/Non-cases | Model 1<br>HR (95% CI) | P value | Model 2<br>HR (95% CI) | P value | Model 3<br>HR (95% CI) | P value |
|---|---|---|---|---|---|---|---|
| Government office region | | | | | | | |
| North East (n=465) | 16/449 | 0.95 (0.56 to 1.60) | 0.850 | 0.95 (0.56 to 1.60) | 0.842 | 0.74 (0.36 to 1.55) | 0.430 |
| North West (n=1361) | 41/1320 | 0.82 (0.58 to 1.18) | 0.290 | 0.82 (0.58 to 1.18) | 0.284 | 1.03 (0.68 to 1.58) | 0.880 |
| Yorkshire and the Humber (n=1210) | 1169/41 | 0.93 (0.65 to 1.33) | 0.693 | 0.93 (0.65 to 1.33) | 0.706 | 1.06 (0.68 to 1.64) | 0.803 |
| East Midlands (n=1110) | 56/1054 | 1.40 (1.02 to 1.93) | 0.039 | 1.41 (1.02 to 1.94) | 0.036 | 1.45 (0.97 to 2.16) | 0.069 |
| West Midlands (n=1176) | 47/1129 | 1.10 (0.78 to 1.55) | 0.583 | 1.10 (0.78 to 1.54) | 0.599 | 1.31 (0.86 to 1.99) | 0.210 |
| East of England (n=1384) | 55/1329 | 1.09 (0.79 to 1.51) | 0.595 | 1.09 (0.79 to 1.50) | 0.603 | 1.26 (0.85 to 1.85) | 0.248 |
| Greater London (n=1839) | 70/1769 | 1.05 (0.78 to 1.42) | 0.734 | 1.06 (0.79 to 1.44) | 0.703 | 1.09 (0.73 to 1.63) | 0.689 |
| South East (n=3088) | 113/2975 | 1 | | 1 | | | |
| South West (n=1832) | 72/1760 | 1.06 (0.79 to 1.43) | 0.678 | 1.07 (0.80 to 1.44) | 0.649 | 1.27 (0.89 to 1.81) | 0.182 |
| Scotland (n=961) | 28/933 | 0.80 (0.53 to 1.21) | 0.293 | 0.80 (0.53 to 1.21) | 0.290 | 0.87 (0.52 to 1.47) | 0.605 |
| Wales (n=650) | 18/632 | 0.76 (0.46 to 1.26) | 0.288 | 0.76 (0.46 to 1.26) | 0.287 | 0.88 (0.49 to 1.59) | 0.675 |
| North South | | | | | | | |
| The North (n=5322) | 201/5121 | 1.00 (0.84 to 1.19) | 0.985 | 1.00 (0.84 to 1.19) | 0.983 | 1.03 (0.83 to 1.28) | 0.801 |
| The South (n=8143) | 310/7833 | 1 | | 1 | | 1 | |
| Urban Rural | | | | | | | |
| Urban (n=10164) | 386/9778 | 1 | | 1 | | 1 | |
| Rural (n=4910) | 171/4739 | 0.90 (0.75 to 1.08) | 0.243 | 0.90 (0.75 to 1.08) | 0.256 | 0.86 (0.69 to 1.07) | 0.173 |
| OAC Supergroup | | | | | | | |
| Blue Collar Communities (n=946) | 24/922 | 0.80 (0.51 to 1.25) | 0.335 | 0.80 (0.51 to 1.24) | 0.314 | 0.86 (0.54 to 1.36) | 0.510 |
| City Living (n=1070) | 37/1031 | 1.17 (0.81 to 1.69) | 0.408 | 1.17 (0.81 to 1.70) | 0.396 | 1.11 (0.75 to 1.64) | 0.604 |
| Countryside (n=3300) | 118/3182 | 1.10 (0.84 to 1.44) | 0.480 | 1.10 (0.84 to 1.44) | 0.480 | 1.09 (0.83 to 1.44) | 0.540 |
| Prospering Suburbs (n=4901) | 201/4700 | 1.27 (1.00 to 1.61) | **0.050** | 1.26 (0.99 to 1.60) | **0.060** | 1.30 (1.01 to 1.66) | **0.040** |
| Constrained by Circumstance (n=581) | 25/556 | 1.38 (0.89 to 2.14) | 0.147 | 1.37 (0.88 to 2.12) | 0.160 | 1.45 (0.92 to 2.30) | 0.109 |
| Typical Traits (n=3190) | 101/3089 | 1 | | 1 | | | |
| Multicultural (n=1095) | 49/1046 | 1.46 (1.04 to 2.06) | **0.029** | 1.47 (1.04 to 2.07) | **0.027** | 1.37 (0.95 to 1.98) | **0.097** |

Table 7.5 - Pre menopausal cases/non-cases and results of Cox proportional hazards regression models for the UKWCS at each spatial scale. Model 1 is age adjusted only. Model 2 is adjusted for age, dietary pattern and diet cost. Model 3 is adjusted for age, dietary pattern, diet cost, energy intake, energy expenditure, smoking status, parity, education*, social class* age at menarche, age at first child, ethanol consumption, height and weight.*Education and social class are not included in the OAC Supergroup models

| | Cases/ Non-cases | Model 1<br>HR (95% CI) | P value | Model 2<br>HR (95% CI) | P value | Model 3<br>HR (95% CI) | P value |
|---|---|---|---|---|---|---|---|
| Government office region | | | | | | | |
| North East (n=472) | 18/454 | 0.94 (0.57 to 1.53) | 0.795 | 0.94 (0.58 to 1.54) | 0.813 | 0.82 (0.44 to 1.53) | 0.527 |
| North West (n=1556) | 59/1497 | 0.94 (0.70 to 1.28) | 0.715 | 0.95 (0.70 to 1.28) | 0.724 | 0.92 (0.63 to 1.33) | 0.644 |
| Yorkshire and the Humber (n=1243) | 67/1176 | 1.34 (1.00 to 1.79) | 0.049 | 1.33 (0.99 to 1.78) | 0.054 | 1.18 (0.81 to 1.72) | 0.400 |
| East Midlands (n=1187) | 54/1133 | 1.13 (0.82 to 1.54) | 0.451 | 1.12 (0.82 to 1.54) | 0.474 | 1.11 (0.75 to 1.63) | 0.610 |
| West Midlands (n=1250) | 56/1194 | 1.10 (0.80 to 1.49) | 0.563 | 1.09 (0.80 to 1.49) | 0.586 | 1.01 (0.69 to 1.50) | 0.942 |
| East of England (n=1512) | 68/1444 | 1.10 (0.82 to 1.47) | 0.521 | 1.10 (0.82 to 1.47) | 0.527 | 1.18 (0.83 to 1.66) | 0.354 |
| Greater London (n=1686) | 87/1599 | 1.25 (0.96 to 1.64) | 0.102 | 1.27 (0.97 to 1.67) | 0.079 | 1.07 (0.74 to 1.54) | 0.723 |
| South East (n=3427) | 140/3287 | 1 | | 1 | | 1 | |
| South West (n=2158) | 119/2039 | 1.35 (1.05 to 1.72) | **0.017** | 1.35 (1.06 to 1.73) | **0.016** | 1.37 (1.01 to 1.85) | **0.042** |
| Scotland (n=1084) | 39/1045 | 0.88 (0.62 to 1.26) | 0.487 | 0.88 (0.62 to 1.26) | 0.496 | 1.09 (0.72 to 1.64) | 0.679 |
| Wales (n=700) | 40/660 | 1.41 (1.00 to 2.01) | 0.053 | 1.43 (1.01 to 2.03) | 0.046 | 1.17 (0.74 to 1.87) | 0.500 |
| North South | | | | | | | |
| The North (n=5708) | 254/5454 | 0.96 (0.82 to 1.12) | 0.592 | 0.95 (0.82 to 1.12) | 0.553 | 0.90 (0.74 to 1.10) | 0.317 |
| The South (n=8783) | 414/8369 | 1 | | 1 | | 1 | |
| Urban Rural | | | | | | | |
| Urban (n=10813) | 481/10332 | 1 | | 1 | | 1 | |
| Rural (n=5460) | 265/5195 | 1.09 (0.94 to 1.26) | 0.266 | 1.08 (0.93 to 1.25) | 0.323 | 1.00 (0.83 to 1.21) | 0.992 |
| OAC Supergroup | | | | | | | |
| Blue Collar Communities (n=1066) | 38/1028 | 0.79 (0.55 to 1.14) | 0.207 | 0.79 (0.55 to 1.13) | 0.197 | 0.81 (0.56 to 1.17) | 0.257 |
| City Living (n=1046) | 51/995 | 1.07 (0.77 to 1.47) | 0.699 | 1.08 (0.78 to 1.49) | 0.645 | 1.09 (0.78 to 1.52) | 0.605 |
| Countryside (n=3869) | 191/3678 | 1.08 (0.87 to 1.34) | 0.498 | 1.06 (0.86 to 1.33) | 0.573 | 1.04 (0.83 to 1.31) | 0.734 |
| Prospering Suburbs (n=5651) | 259/5392 | 1.00 (0.82 to 1.23) | 0.992 | 0.98 (0.80 to 1.21) | 0.879 | 1.01 (0.82 to 1.25) | 0.895 |
| Constrained by Circumstance (n=764) | 28/736 | 0.81 (0.54 to 1.21) | 0.308 | 0.81 (0.54 to 1.22) | 0.310 | 0.87 (0.58 to 1.31) | 0.503 |
| Typical Traits (n=3087) | 141/2946 | 1 | | 1 | | 1 | |
| Multicultural (n=800) | 38/762 | 1.06 (0.74 to 1.52) | 0.742 | 1.08 (0.76 to 1.55) | 0.662 | 1.04 (0.72 to 1.52) | 0.820 |

Table 7.6 - Post menopausal cases/non-cases and results of Cox proportional hazards regression models for the UKWCS at each spatial scale. Model 1 is age adjusted only. Model 2 is adjusted for age, dietary pattern and diet cost. Model 3 is adjusted for age, dietary pattern, diet cost, energy intake, energy expenditure, smoking status, parity, education*, social class* age at menarche, age at first child, ethanol consumption, height and weight.*Education and social class are not included in the OAC Supergroup models

### 7.4.3 Breast cancer and weight status

Kaplan-Meier survival curves show that the assumptions for Cox regression are met. They also display how the probability of staying breast cancer free decreases with increasing weight status in the UKWCS women (menopausal status not considered). Underweight and normal weight women have noticeably higher probability of staying breast cancer free (Figure 7.4).

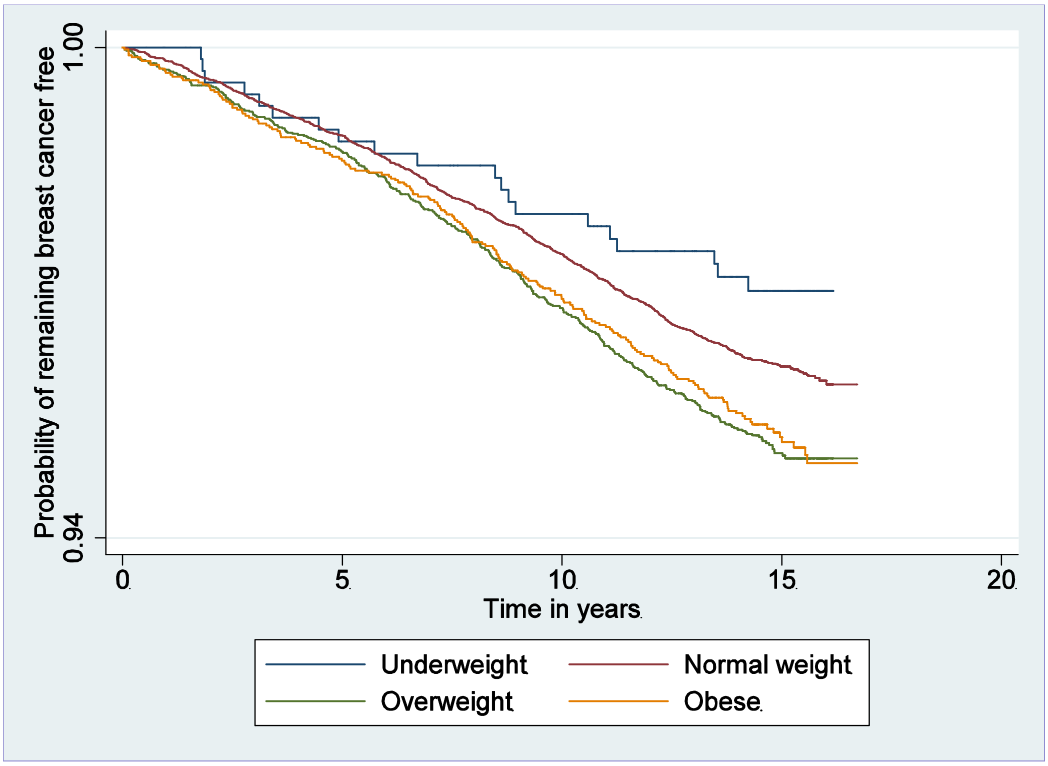



Figure 7.4 - Kaplan-Meier survival curve for probability of remaining breast cancer free by weight status for all UKWCS women

While regression results for weight status are statistically non-significant, there is a clear significant trend that with increasing BMI category, risk of breast cancer increases. This can be seen in the whole sample and post menopausal subgroup. The relationship holds true when adjusting for age and spatial diet cost pattern exposures (Model 2) but disappears when all confounders and mediators are included, see Table 7.7. A ‘test for trend’ shows a statistically significant increase in breast cancer incidence with increasing weight status for the whole sample and postmenopausal women.

### 7.4.4 Breast cancer and dietary pattern

No clear trends are observed when considering dietary pattern as the primary exposure. There appears to be some reduced risk in women consuming a low diversity vegetarian dietary pattern in the whole cohort, but this is not statistically significant (Table 7.8).

| | | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|---|
| | **Cases/Non-cases** | **HR (95% CI)** | **P value** | **HR (95% CI)** | **P value** | **HR (95% CI)** | **P value** |
| Whole cohort - breast cancer incidence test for trend p<0.001 | | | | | | | |
| Underweight | 20/706 | 1 | | 1 | | 1 | |
| Normal weight | 761/19528 | 1.30 (0.83 to 2.02) | 0.254 | 1.31 (0.83 to 2.06) | 0.248 | 1.22 (0.69 to 2.17) | 0.492 |
| Overweight | 370/7687 | 1.55 (0.99 to 2.43) | **0.057** | 1.55 (0.97 to 2.46) | **0.065** | 1.47 (0.82 to 2.62) | 0.196 |
| Obese | 193/4145 | 1.53 (0.97 to 2.43) | **0.069** | 1.55 (0.97 to 2.49) | **0.070** | 1.40 (0.77 to 2.53) | 0.272 |
| Pre-menopausal- breast cancer incidence test for trend p=0.740 | | | | | | | |
| Underweight | 12/346 | 1 | | 1 | | 1 | |
| Normal weight | 374/9828 | 1.03 (0.58 to 1.84) | 0.909 | 1.09 (0.60 to 1.99) | 0.781 | 1.09 (0.60 to 1.99) | 0.781 |
| Overweight | 126/2969 | 1.10 (0.61 to 2.00) | 0.746 | 1.14 (0.61 to 2.12) | 0.673 | 1.14 (0.61 to 2.12) | 0.673 |
| Obese | 60/1636 | 0.98 (0.53 to 1.83) | 0.957 | 1.03 (0.54 to 1.97) | 0.925 | 1.03 (0.54 to 1.97) | 0.925 |
| Post-menopausal - breast cancer incidence test for trend p<0.001 | | | | | | | |
| Underweight | 8/318 | 1 | | 1 | | 1 | |
| Normal weight | 379/8848 | 1.57 (0.78 to 3.17) | 0.205 | 1.50 (0.74 to 3.03) | 0.255 | 1.53 (0.63 to 3.72) | 0.345 |
| Overweight | 241/4326 | 2.03 (1.00 to 4.10) | **0.049** | 1.91 (0.94 to 3.87) | **0.074** | 1.94 (0.79 to 4.74) | 0.146 |
| Obese | 130/2279 | 2.11 (1.03 to 4.32) | **0.040** | 2.03 (0.99 to 4.16) | **0.053** | 1.89 (0.76 to 4.68) | 0.168 |

Table 7.7 - Cases/non-cases and results of Cox proportional hazards regression models for the UKWCS at each BMI category. Model 1 is only age adjusted. Model 2 is adjusted for age, dietary pattern, diet cost, Supergroup and GOR. Model 3 is adjusted for age, dietary pattern, Supergroup, GOR, energy intake, energy expenditure, smoking status, ethanol consumption, age of first child, age at menarche and parity

| | | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|---|
| | **Cases/Non-cases** | **HR (95% CI)** | **P value** | **HR (95% CI)** | **P value** | **HR (95% CI)** | **P value** |
| Whole cohort | | | | | | | |
| Monotonous Low Quantity Omnivore | 205/5130 | 0.92 (0.77 to 1.11) | 0.395 | 0.90 (0.74 to 1.09) | 0.289 | 0.87 (0.68 to 1.10) | 0.243 |
| Health Conscious | 80/2014 | 0.90 (0.70 to 1.16) | 0.407 | 0.96 (0.72 to 1.26) | 0.747 | 1.01 (0.73 to 1.41) | 0.944 |
| Traditional Meat, Chips and Pudding Eater | 248/5757 | 1 | | 1 | | 1 | |
| Higher Diversity Traditional Omnivore | 208/4550 | 1.04 (0.87 to 1.25) | 0.669 | 1.07 (0.88 to 1.29) | 0.521 | 1.02 (0.82 to 1.28) | 0.835 |
| Conservative Omnivore | 274/5588 | 1.08 (0.91 to 1.29) | 0.358 | 1.06 (0.89 to 1.26) | 0.510 | 1.09 (0.88 to 1.35) | 0.413 |
| Low Diversity Vegetarian | 176/4895 | 0.85 (0.70 to 1.03) | **0.099** | 0.83 (0.68 to 1.01) | **0.062** | 0.77 (0.60 to 0.98) | **0.033** |
| High Diversity Vegetarian | 153/4132 | 0.86 (0.70 to 1.05) | 0.130 | 0.86 (0.70 to 1.06) | 0.154 | 0.89 (0.70 to 1.14) | 0.370 |
| Pre-menopausal | | | | | | | |
| Monotonous Low Quantity Omnivore | 80/2071 | 0.91 (0.69 to 1.22) | 0.539 | 0.89 (0.66 to 1.20) | 0.446 | 0.97 (0.71 to 1.32) | 0.828 |
| Health Conscious | 33/832 | 0.90 (0.61 to 1.32) | 0.588 | 0.92 (0.60 to 1.42) | 0.720 | 1.04 (0.66 to 1.64) | 0.880 |
| Traditional Meat, Chips and Pudding Eater | 112/2641 | 1 | | 1 | | 1 | |
| Higher Diversity Traditional Omnivore | 71/1934 | 0.84 (0.63 to 1.14) | 0.263 | 0.83 (0.61 to 1.14) | 0.261 | 0.84 (0.61 to 1.16) | 0.290 |
| Conservative Omnivore | 83/1981 | 0.93 (0.70 to 1.23) | 0.611 | 0.89 (0.67 to 1.19) | 0.432 | 0.94 (0.69 to 1.28) | 0.693 |
| Low Diversity Vegetarian | 103/2972 | 0.82 (0.63 to 1.08) | 0.156 | 0.80 (0.61 to 1.05) | 0.108 | 0.78 (0.58 to 1.04) | **0.089** |
| High Diversity Vegetarian | 90/2343 | 0.90 (0.68 to 1.18) | 0.436 | 0.89 (0.66 to 1.18) | 0.405 | 0.84 (0.62 to 1.14) | 0.268 |
| Post-menopausal | | | | | | | |
| Monotonous Low Quantity Omnivore | 124/2787 | 0.96 (0.75 to 1.22) | 0.723 | 0.93 (0.72 to 1.21) | 0.606 | 0.90 (0.66 to 1.23) | 0.498 |
| Health Conscious | 46/1058 | 0.91 (0.65 to 1.27) | 0.563 | 0.98 (0.68 to 1.41) | 0.919 | 0.94 (0.61 to 1.45) | 0.775 |
| Traditional Meat, Chips and Pudding Eater | 134/2870 | 1 | | 1 | | 1 | |
| Higher Diversity Traditional Omnivore | 135/2391 | 1.19 (0.93 to 1.51) | 0.161 | 1.23 (0.96 to 1.58) | 0.105 | 1.12 (0.84 to 1.50) | 0.429 |
| Conservative Omnivore | 187/3322 | 1.17 (0.94 to 1.46) | 0.166 | 1.15 (0.92 to 1.45) | 0.210 | 1.19 (0.91 to 1.55) | 0.210 |
| Low Diversity Vegetarian | 71/1743 | 0.86 (0.64 to 1.14) | 0.290 | 0.84 (0.62 to 1.12) | 0.228 | 0.81 (0.57 to 1.15) | 0.244 |
| High Diversity Vegetarian | 61/1600 | 0.78 (0.57 to 1.05) | 0.104 | 0.78 (0.57 to 1.06) | 0.116 | 0.87 (0.61 to 1.24) | 0.438 |

Table 7.8 - Cases/non-cases and results of Cox proportional hazards regression models for the UKWCS for dietary patterns. Model 1 is only age adjusted. Model 2 is adjusted for age, diet cost, Supergroup and GOR. Model 3 is adjusted for age diet cost, Supergroup, GOR, energy intake, energy expenditure, smoking status, ethanol consumption, age of first child, age at menarche, parity, height and weight

## 7.5 Discussion

This chapter shows little association between spatial variation, diet cost or dietary patterns and breast cancer risk in the UKWCS. This could be due to the homogeneity of the women in the cohort being primarily middle class, thus the likelihood of developing health outcomes such as breast cancer would also be homogenous. Differences which we have seen in diet cost, dietary pattern consumption and obesity in earlier chapters appear to be influenced by the geographical and geodemographic environment, but this is not seen in breast cancer incidence. This suggests that it may be that dietary cost patterns and geographical environment are not the drivers for this health outcome. Nationally, variation in age standardised breast cancer incidence is observed by region, but this is not by more than 10% from the average.

### 7.5.1 Spatial variation and breast cancer incidence

Initial descriptive statistics suggested that there were significant differences in incidence rates by GOR and OAC spatial measures. The Kaplan-Meier plot illustrates that those in Blue Collar Communities had the highest probability of remaining breast cancer free after 15 years. This is in line with other research that suggests, that those of lower SES (and these areas are typical of the working class) are at lower breast cancer risk. However, once these spatial measures were investigated as observed verses expected time to event analysis, no statistical difference by OAC Supergroup was seen. Cox regression results show an increased risk, approaching levels of statistical significance, for premenopausal women residing in a Prospering Suburbs Supergroup, in line with evidence demonstrating an inverse association between breast cancer incidence and deprivation. Increased risk is also seen in the Multicultural Supergroup for the whole cohort and pre menopausal women. This increase is approaching levels of statistical significance but significance diminished in the fully adjusted models. This is contrary to results reported by Robinson et al, that decreased age standardised incidence rates of breast cancer in 'Multicultural Centres', a group in the People and Places geodemographic classification, in the North West of England (Robinson et al., 2010).

Difference by GOR was approaching significance in observed verses expected time to event analysis. Increased breast cancer risk, approaching levels of significance, are seen in the South West, compared to the reference category - the South East - in the whole cohort and post menopausally, and remain in the fully adjusted models. A difference in risk at such a large geographical scale suggests that perhaps there is a spatial diet cost pattern related risk or it could be related to difference in regional breast cancer services in this area which it has not been possible to control for. National statistics show similar rates of breast cancer incidence to this study for the South East and South West, with both above average (Office for National Statistics, 2007).

These results contradict results from earlier spatial diet cost analysis where the OAC Supergroups have presented higher variations in outcome measure, suggesting that inclusion of the demographic characteristics are a valuable tool. Interestingly, in this sample of women, there is no significant difference in breast cancer incidence by education or diet cost. Crude social class differences are at levels approaching significance, but not as pronounced as might have been expected given that this is a disease more highly associated with being affluent than poor.

### 7.5.2 Dietary patterns and breast cancer incidence

Significant differences were observed by breast cancer incidence for, dietary pattern consumption and BMI, as a continuous measure and with higher significance by BMI category. Difference by dietary pattern is interesting as this is something which has not been conclusively shown in other studies (Cade et al., 2010, Cade et al., 2011, Brennan et al., 2010). Once time to event analysis was applied, the difference remained. In Cox regression models, adjusting for confounders and mediators, no association between dietary patterns and breast cancer risk was observed, either in the full sample or the pre and post menopause sub group analysis. This agrees with earlier research completed on this cohort looking at Mediterranean dietary patterns and the healthy diet index, when the UKWCS follow up time was only 9 years (Cade et al., 2011), some trend towards reduced risk with healthier diets was observed although non-significant. It might have been expected that a similar trend would be seen in this analysis also, with a significantly reduced risk in breast cancer for those consuming a healthier dietary pattern such as the Health Conscious or High Diversity Vegetarian. Instead, some reduced risk is shown for these patterns, although not significant, and is similar to the reduced risk observed for those who consume the least healthy dietary pattern - the Monotonous Low Quantity Omnivore. The only pattern showing levels approaching significance for risk reduction is the Low Diversity Vegetarian diet, which is high in fibre - a known protective factor - yet the other patterns with higher fibre content do not show this significant reduction. The Low Diversity Vegetarian dietary pattern combines high fibre content with low energy intake, so perhaps this combination contributes to a protective effect, although further investigation would be required to confirm this.

### 7.5.3 Weight status and breast cancer incidence

The literature is clear that weight status has a significant impact on breast cancer incidence (Key et al., 2002, WCRF, 2007). This is also observed in the UKWCS women, in both the descriptive statistics and in the observed versus expected time to event analysis. In Cox regression models, the difference only remains for an increased risk in overweight and obese women in minimally adjusted models for the whole cohort and post menopausal subgroup. In the fully adjusted model, including hormonal risk factors, no difference is observed.

So, while dietary pattern and spatial environment can influence obesity risk (Chapter 6), this Chapter shows obesity has some association with breast cancer risk, but there is no spatial or dietary pattern association with breast cancer risk, suggesting that this is not the mechanism through which breast cancer risk is facilitated. The fact that the fully adjusted model (Model 3) shows no breast cancer association with weight status also detracts from this. In the UKWCS, mean BMI is in the healthy range, so perhaps there is not enough variation in weight status for an association to remain when fully adjusted, or perhaps results indicate that the association between weight status and breast cancer risk is in fact a weak one. One suggested mechanism could be that hormonal factors influence both weight status and breast cancer so that weight status can be used as a proxy for breast cancer risk, but does not have a causal link.

### 7.5.4 Strengths and limitations

Given that 14.6 years have passed since menopausal status was collected, and the mean age at baseline was 52, the majority of women in the UKWCS women will probably now be postmenopausal. Therefore, the baseline measure of menopausal status in the UKWCS is unlikely to be a reliable estimation of current menopausal status. Despite this, differences are still observed between different menopausal states. This may be due to the amount of time women have spent in menopause or that their diet and weight status prior to menopause affects longer term health outcomes.

This study does not account for changes to diet or place of residence since baseline data was collected, so this research can only look at the effect of such measures at baseline. Time to event analysis is dependent to some extent on all factors staying equal. We do not know whether the women have moved home during this time to a different area or type of area, nor do we know whether they made substantial changes to the diet and/or weight status since baseline which could impact outcomes. In future research it would be interesting to include analysis of the estimated social mobility of the UKWCS using flow data from the Centre for Interaction Data Estimation and Research (Centre for Interaction Data Estimation and Research, 2011). However, the large numbers of women in this study and their geographic spread should give some confidence in the reliability of the results.

It is always difficult to have absolute confidence in self reported height and weight due to bias, but this is something which is widely accepted as a weakness in epidemiological research (Hu, 2008). FFQs also have their pitfalls, but in large studies such as this they are a valid method of dietary assessment (Willett, 2013). Follow up analysis could use the subset of UKWCS women (n~14000) who completed a four day food diary at follow up, approximately 5 years after the baseline data was collected, which would provide more detailed dietary intake data.

### 7.5.5 Future work

Clear geographical variations - such as North South divide, and by OAC Supergroup (including the significant differences for Blue Collar Communities and Constrained by Circumstance), exist in dietary pattern and obesity outcomes. However, it is reassuring that these do not translate to inequality in health, when breast cancer is the outcome. Weight and diet are known to introduce inequalities in health for other outcomes such as diabetes or cardiovascular disease (Foresight, 2007, Marmot, 2010). It would be worth while exploring these health outcomes by spatial measure in this cohort in future research.

## 7.6 Summary and context

There is no significant spatial variation in breast cancer incidence in the UKWCS. However, there appears to be some increased risk, approaching levels of statistical significance, in:

- South West for whole cohort analysis and post menopause
- Multicultural area for whole cohort and pre menopause
- Prospering Suburbs pre menopause.

A positive association exists between weight status risk of breast cancer incidence, but not in premenopausal women.

Quintiles of diet cost are not associated with breast cancer incidence.

In Chapter 8, exploratory analysis will be carried out using a geodemographic classification in an Australian and American population to profile diet and health, using methods described in Chapter 3.

# Chapter 8: Geography of diet using a Geodemographic classification: international case studies

## 8.1 Overview

This Chapter investigates whether the methods discussed so far, using geodemographic identifiers to profile diet and health and then estimate small area patterns are applicable to other developed populations. A Cameo geodemographic identifier is added to a cohort from Australia and the USA to investigate fruit and vegetable consumption and weight status by Cameo group. Research was carried out during short research visits, therefore analysis is exploratory and not as extensive as work seen in other chapters of this thesis. Figure 8.1 is a reminder of where this fits into the thesis flow.

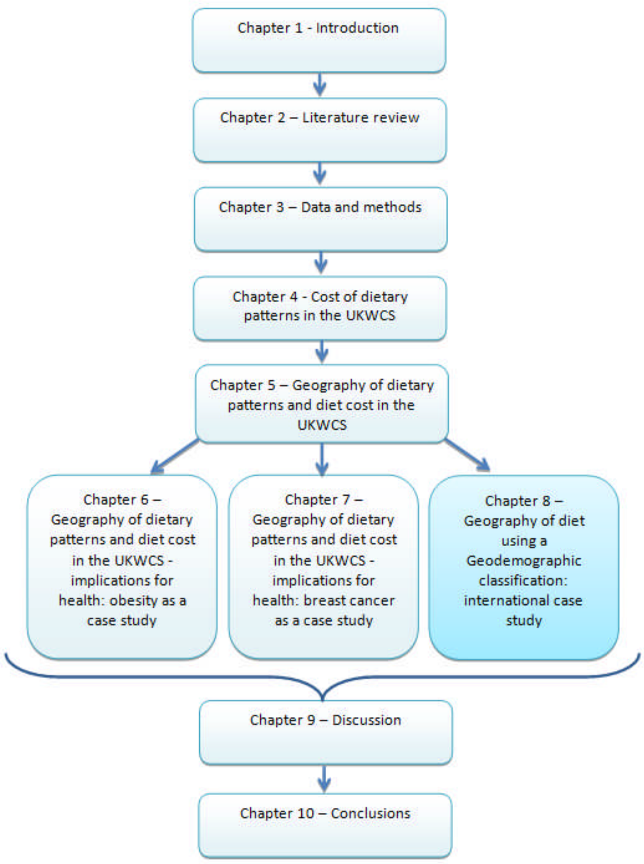



Figure 8.1 - Overview showing how Chapter 8 fits into the overall thesis flow

## 8.2 Introduction

Diet is important for health and prevention of non-communicable diseases worldwide, not only the UK. The effect of energy dense diets and sedentary activity manifesting as what some describe as an 'obesity epidemic' are being observed across the developed world (World Health Organisation, 2000). Australia and the USA join the UK in the top seven most obese countries in the world (OECD, 2012).

Influences on dietary consumption in Australia do not differ greatly from those in other developed countries. A number of studies have investigated how dietary consumption varies according to socioeconomic status (SES) in Australia. Typically, lower SES is associated with poorer diet quality (Thornton et al., 2010, Thornton et al., 2011b, Mishra et al., 2002, Ball et al., 2006, Mishra et al., 2005). The situation is similar in the USA (Diez Roux et al., 1999, Larson et al., 2009, Aggarwal et al., 2011). Various indicators of SES are commonly used in order to control for confounding effects. The most common are: education level, age, income, occupation, country of birth and presence/number of dependent children. In Australia the Socio-Economic Areas for Indexes (SEIFA)(Australian Bureau of Statistics, 2008) are widely used as a measure of deprivation. In the USA an equivalent index does not exist. Measures of SES in health studies are inconsistent (Braveman et al., 2005) and education and income or poverty to income ratios are widely used. There are also some interesting recent results showing how property value is a strong indicator of obesity in the USA (Rehm et al., 2012).

Australia is a large and diverse country. Whilst there are similarities to the UK, the geography of the country varies greatly. In urban areas, segregation is observed by SES, but this is less pronounced than countries such as the UK and the USA (Turrell et al., 2004). Environments, including proximity to different types of food premises, including fast food restaurants, along with supermarket access in Australia have been shown to have differing effects on diet quality (Thornton et al., 2011a); although multilevel analysis has shown that place of residence is not a more powerful predictor of dietary consumption than your own personal characteristics (Turrell et al., 2004). There are many USA studies, some included in the review by Larson et al (Larson et al., 2009) discussing influence of neighbourhood environment on food consumption showing that access to food, geographically and economically, can effect food choice with subsequent effects on health.

Geodemographic classifications can act to combine both where someone lives and their demographics. Geodemographic classifications are not commonly used as a tool to explore diet and weight status in Australia and the USA. In the absence of neighbourhood level data on health, it is increasingly common that synthetic data will be generated, using techniques such as spatial microsimulation. Such estimated data is beneficial for understanding local communities and where best to place public health interventions. If it were possible to profile diet and health using a small area geographic system like a geodemographic classification, it may be possible, in turn, to use this to estimate diet and health at a small area geography level - a novel method to obtain neighbourhood level estimates in these populations.

### 8.2.1 Objectives

The objectives for this chapter, as described in Chapter 1, Table 1.1 are to:

- Use Cameo geodemographic classifications to describe diet (fruit and vegetable consumption) and weight status in the Australian Longitudinal Study for Women's Health (ALSWH) and the Seattle Obesity Study (SOS1).
- Explore use of Cameo geodemographic classification to estimate small area diet or health patterns in an Australian and USA city.

### 8.2.2 Background

In order to explore the diets of Australians and Americans it is important to place their dietary guidelines in context to aid interpretation of the results.

#### *8.2.2.1 Australian dietary guidelines*

Dietary guidelines in Australia are subtly different from those recommended in the UK (see Table 8.1). The fundamental values are the same, promoting variety and high fruit and vegetable intake with lower high fat and sugary products (Australian Government, 2005).

| Food | Recommended number of servings per day |
|---|---|
| Cereals | 4-9 |
| Vegetables, legumes (including potatoes) | 5 |
| Fruit | 2 |
| Milk, yoghurt, cheese | 2 |
| Lean meat, fish, poultry & alternatives | 1 |
| Extra foods (e.g. cakes, pies, soft drinks, lollies) | 0-2.5 |

Table 8.1 - Summary of Australian dietary guidelines for women aged 19-60 years

It is important to note that the vegetable recommendation in Australia includes potatoes, which is not the case in the UK and also, Australia recommend a total of seven portions of fruit and vegetables per day (two more than in the UK). The Department of Health's Eatwell Plate pictorially dictates the UK recommendations. Potatoes are included with bread, rice, pasta, cereals: wholegrains are recommended to be chosen where possible. The Eatwell Plate recommends lean meat, fish, poultry and beans in the same amounts as dairy; while in Australia, dairy products are recommended to be double that of meats. In Australia, the guidelines for 'Extra foods' are more lenient, advising between 0 and 2.5, while the Eatwell Plate suggests this is the smallest proportion of food consumed, without a specific value assigned. As in the UK, few women meet dietary recommendations in Australia (Ball et al., 2004).

#### *8.2.2.2 USA dietary guidelines*

The MyPyramid (Figure 8.2) recommendation was released in 2005 jointly by the USDA and the Department for Health and Human Services, to pictorially illustrate dietary recommendation in the USA. In 2011 these were replaced by the MyPlate pictorial following an update to USDA guidelines. The daily recommended intake of fruit and vegetables are 2 portions and 3 portions respectively, for someone consuming 2000 kcal/day.

The US guidelines are simpler than those for the UK and Australia in that they contain fruits, vegetables, proteins, grains and dairy, but do not include any reference to the unhealthy foods such as high fat and sugar products. Like the UK and Australia, typical dietary consumption falls short of the recommendations (United States Department of Agriculture, 2010).

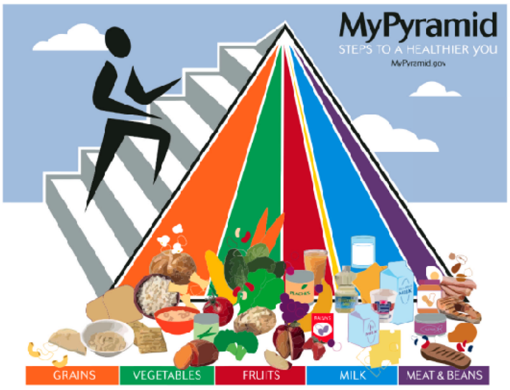



Figure 8.2 - MyPyramid. Source: www.mypyramid.gov


Figure 8.3 - MyPlate. Source: www.choosemyplate.gov

## 8.3 Methods

### 8.3.1 Research visits

Two research visits were made in order to complete the analysis in the chapter:

- November 2012 - 1 week visit to the University of Queensland, School of Population Health, Working with Professor Gita Mishra and Dr Caroline Jackson.
- August 2013 - 3 week visit to the University of Washington, Seattle in the Centre for Public Health Nutrition, working with Professor Adam Drewnowski and Dr Anju Aggarwal.

### 8.3.2 Study design and sample

*ALSWH*

This is a cross sectional analysis using data from the ALSWH cohort which was established in 1995 to focus on women's health and well being across the life course (Lee et al., 2005, Brown et al., 1998). A cohort profile was published in 2005 by Lee et al describing recruitment and response rate, as well as details on the data collected.

ALSWH contains three different age groups of women, but the research described here specifically looks at the middle aged cohort of women, born between 1946 and 1951 (see Chapter 3, section 3.2.2, for more information).

*SOS1*

Participants for the initial wave of the Seattle Obesity Study (SOS1) were recruited between October 2008 and March 2009 from King County, Washington to participate in a study designed to investigate social disparities, diet and health. Adult residents from a stratified random sample of 2001 completed a twenty minute telephone questionnaire, including FFQ. Geocoded addresses in Seattle King County are available for 1992 of these participants (see Chapter 3, section 3.2.2, for more information).

### 8.3.3 Ethics and data access

*ALSWH*

Ethical approval for ALSWH was provided by the Human Research Ethics Committees of the Universities of Newcastle and Queensland.

In order for this research to be completed, an Expression of Interest Form application was submitted in September 2012 which was approved shortly after. Prior to being given access to the data an ALSWH a Memorandum of understanding and an ALSWH Privacy Protocol (not included in this thesis) were signed (see Appendix C and D).

*SOS1*

Ethical approval to conduct this study was obtained from the Institutional Review Board at the University of Washington.

Agreement for using the SOS1 data was that analysis had to be completed at the University of Washington and no data was to be retained at the end of the visit.

### 8.3.4 Cameo geodemographic classification

CallCreditGroup in the UK supplied the Cameo geodemographic classification for Australia and the USA. In order to be given access to this data, a contract was put in place (see Appendix E). CallCreditGroup also provided training on using the data. Chapter 3, section 3.2.4.6 describes the classifications.

### 8.3.5 Fruit and vegetable consumption

Fruit and vegetable consumption has been recorded in both the ALSWH and SOS1 cohorts. For consistency and comparison purposes, adherence to national recommendations for fruit and vegetables will be investigated in this analysis, as a measure of a healthy diet.

ALWSH

Fruit and vegetable consumption was recorded in the ALSWH cohort through specific questions: "How many pieces of FRESH fruit do you usually eat per day? (Count ½ cup of diced fruit, berries or grapes as one piece) and "How many different vegetables do you usually eat per day? (Count all types, fresh, frozen or tinned). The results of this compare well to the fruit and vegetable intake derived from the results of the FFQ (results not reported in this thesis). New binomial variables were created: (1) to indicate adherence to fruit guidelines - 2 portions per day; (2) to indicate adherence to vegetable guidelines - 5 portions per day; (3) to indicate adherence to combined fruit and vegetable guidleines.

*SOS1*

Fruit and vegetable consumption in SOS1 was derived from the results of the FFQ through analysis by a third party group at Fred Hutchinson Cancer Research Centre. Using this data new binomial variables were created to determine whether the participant met the recommended dietary guidelines for the USA: (1) to indicate adherence to fruit guidelines - 2 portions per day; (2) to indicate adherence to vegetable guidelines - 3 portions per day; (3) to indicate adherence to combined fruit and vegetable guidleines.

### 8.3.6 Weight status

Both cohorts contain self reported height and weight which has been used to generate Body Mass Index (BMI). Using BMI the participants were categorised according to the World Health Organisation (WHO) categories, reported in Chapter 3, section 3.2.1.2.

### 8.3.7 Data linking

*ALSWH*

In order to maintain anonymity of the ALSWH women and to respect the intellectual property rights associated with the Cameo data, linking the Cameo ID to ALSWH dietary and demographic data was performed in collaboration with the ALSWH data manager. This involved a two stage process using dummy ID numbers. First, a Cameo ID was assigned to a dummy ID for each woman by matching via spatial location (specifically longitude and latitude) using ArcGIS 10 software (ESRI Inc, 2010). Second, the dietary and demographic data was merged into the dataset containing Cameo ID for each women, via the dummy ID using Stata IC12 (StataCorp, 2012).

*SOS1*

Each participant is assigned one of ten Cameo ID's which are determined at census block. Census block is the smallest geographical unit used in the USA census. It is usually determined by physical boundaries such as roads, and often correspond to city blocks. These may vary in population size depending on their location. SOS1 participants were matched by spatial location, using the longitude and latitude coordinates of their home address, in ArcGIS (ESRI Inc, 2010). All participants matched, but 15 matched to blank Cameo records so were excluded, leaving a sample of 1977 participants.

### 8.3.8 Statistical analysis

Stata IC12 statistical software has been used to perform the analysis (StataCorp, 2012).

Descriptive statistics are presented. Test for difference between Cameo groups was carried out using $chi^2$ and Kruskal Wallis for parametric and non-parametric data as appropriate. Linear regression was used to test how well Cameo groups predicted the continuous outcome - BMI. Logistic regression was used to generate odds ratios for categorical outcomes: meeting fruit and vegetable guidelines; and BMI category. Stacked bar charts were created using Microsoft Excel to illustrate the distribution of BMI category by Cameo group. Robust standard errors were applied to regression models where non-normal distribution exists. Regression models are unadjusted, as the Cameo classification is created such that all demographic and socio-economic confounders and mediators are already accounted for. P values are considered significant at 99% confidence level, $p<0.01$, due to the multiple testing nature of this analysis.

In the ALSWH analysis Cameo Australia Affluent Urban Professionals, group 1 was used as a reference value as this group is the most affluent. In Cameo USA the American Aristocracy, group 1 was used as reference. This will allow for some comparison between the Australian and USA population behaviours relative to the most affluent Cameo groups.

Not being obese is used as a reference in regression investigating differences in obesity status and not adhering to fruit and vegetable guidelines was reference category for adherence to dietary guidelines analysis.

### 8.3.9 Estimating obesity

By making an assumption that obesity status for each Cameo group in the cohort are typical for residents in these groups throughout the country it is possible to estimate obesity status for any city, using Cameo codes. Methods for estimating small area patterns follow the same principle as methods described in Chapter 3, section 3.3.7.

*ALSWH*

Boundary files for Statistical local areas were downloaded from the Australian Standard Geographical Classification (ASGC) Digital Boundaries, published by the Australian bureau of statistics in 2011 (Australian Bureau of Statistics, 2011) in ArcGIS  shape file format (ESRI Inc, 2010).

*SOS1*

A boundary file for King County was provided by CallCreditGroup (Callcredit Information Group, 2013). An Urban Growth Boundary file, downloaded from King County GIS Centre (King County, 2013) was used to clip the boundary file in order to present data for this specific area. Geographic coordinates were converted to match the King County file.

## 8.4 Results

### 8.4.1 ALWSH - Descriptive statistics

It was not possible to match 1689 of the 11154 women in ALWSH to a Cameo identifier, by matching the longitude and latitude coordinate to inside a mesh block with a Cameo ID, or by matching to the closest mesh block with a Cameo ID, leaving a sample of 9465 for analysis. Due to the sensitivity of the two datasets, it was not possible to investigate further (in the available timeframe) the reasons for this unmatched data, but likely causes are; erroneous longitude and latitude data in the ALSWH, missing data in Cameo or differing geographic coordinate systems in the two datasets. Data on the unmatched women was not retained, so it is not possible to present a table of demographic differences in the matched and unmatched women. A map showing the locations of matched and women are available in Figure 8.4.

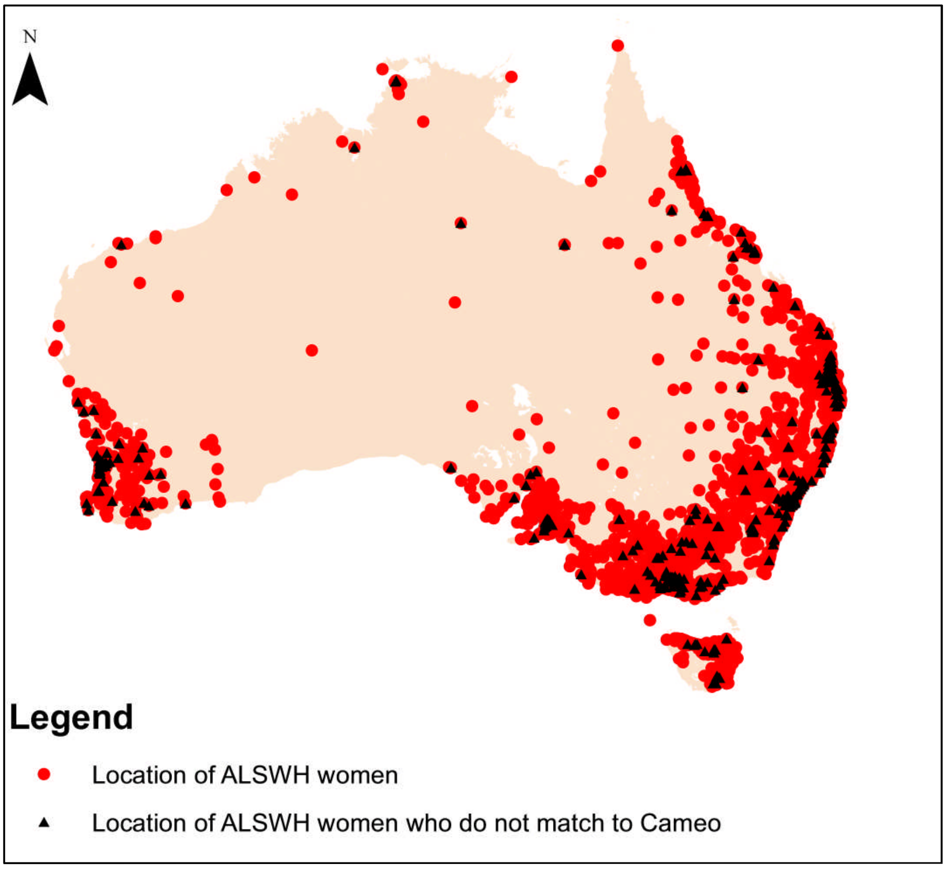



Figure 8.4 - A map showing the location of matched and unmatched women in the ALSWH

ALSWH women are distributed across all ten Cameo groups. Compared to Cameo distribution for the Australian population, some under representation exists in the Affluent Urban Professionals, High Income Urbanities and Students and Diverse Low Income Urban

Communities Cameo groups. There is also some over representation in the Comfortable Mixed Suburban Areas, Low Income Rural and Suburban Neighbourhoods and Very Low Income Rural Communities (see Figure 8.5).

In order to roughly evaluate the Cameo classification, demographic variables of the ALSWH cohort are presented in Table 8.2. The Affluent Urban Professionals earn the highest annual household salaries; have the highest percentage of managerial or professional occupations for both the women and their partners. They are the most educated group with 31% having completed a university higher degree. The numbers of women who do not have a partner in this group are close to the average value for the cohort. The Very Low Income Rural Communities have only 8% with an annual household salary income of over $78000 Australian and only 10% of this group are educated at university level. The group of High Income Urbanities and Students have a high percentage of those on the highest earnings, and also 19% educated to university level. Interestingly, this group has nearly one quarter of the women living with no partner. The other Cameo groups show subtle differences in these SES markers.

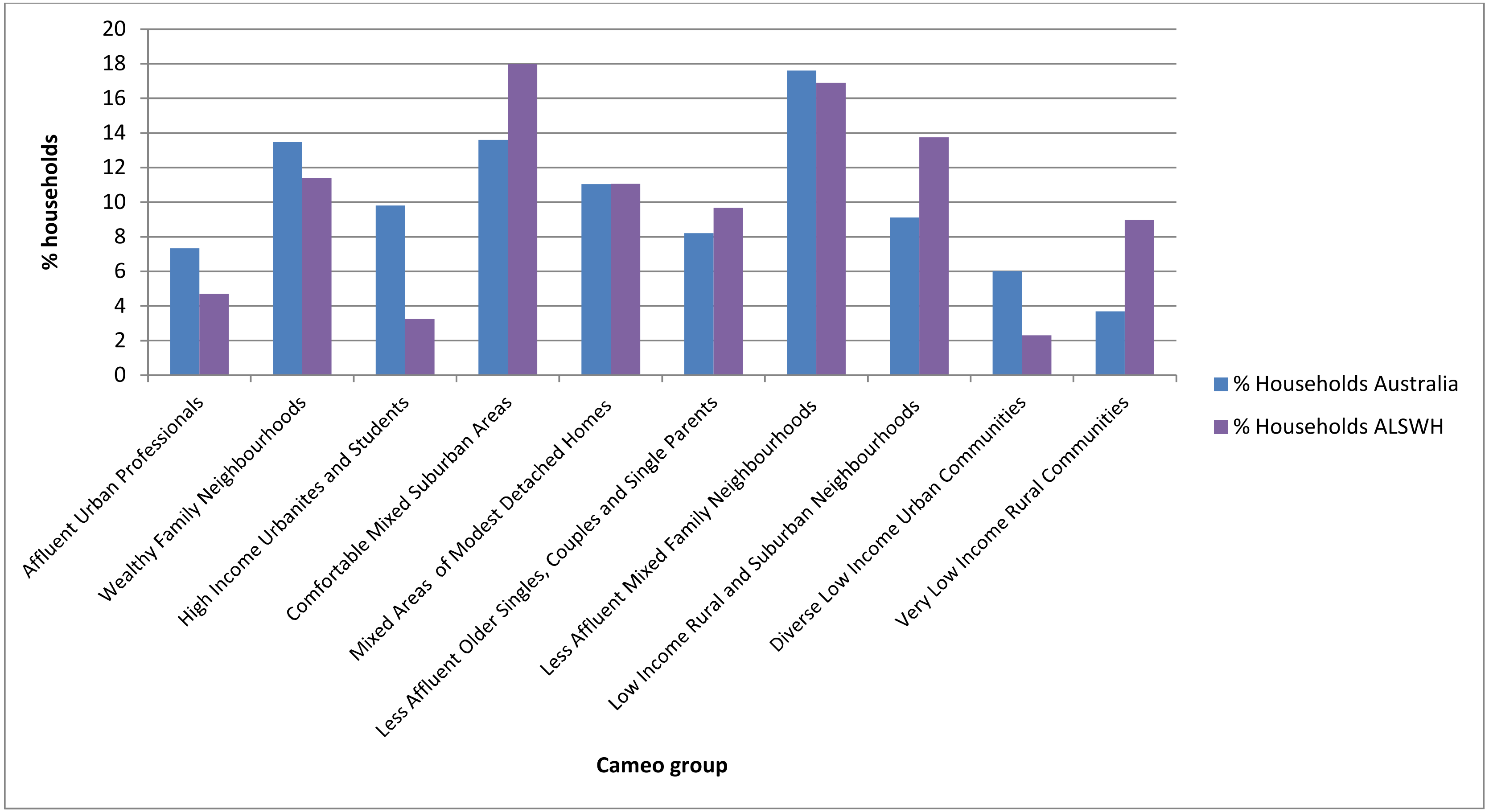



Figure 8.5 - Distribution of ALSWH participants compared to Australian population by Cameo geodemographic group

| Name | % Salary >$78,000 (AUD) per year ALSWH Household (95% CI) | % Managerial or professional - ALSWH woman (95% CI) | % Managerial or professional - ALSWH partner (95% CI) | % ALSWH women with no partner (95% CI) | % with university degree or higher (95% CI) |
|---|---|---|---|---|---|
| Affluent Urban Professionals (n=445) | 36 (31 to 41) | 42 (37 to 47) | 49 (44 to 54) | 13 (10 to 17) | 31 (26 to 74) |
| Wealthy Family Neighbourhoods (n=1079) | 18 (15 to 20) | 28 (25 to 31) | 28 (26 to 31) | 13 (11 to 15) | 16 (14 to 18) |
| High Income Urbanites and Students (n=308) | 22 (17 to 27) | 30 (24 to 35) | 31 (26 to 37) | 23 (18 to 28) | 19 (14 to 23) |
| Comfortable Mixed Suburban Areas (n=1704) | 15 (13 to 17) | 28 (26 to 31) | 31 (28 to 33) | 14 (12 to 15) | 15 (13 to 17) |
| Mixed Areas of Modest Detached Homes (n=1046) | 18 (15 to 20) | 26 (23 to 30) | 27 (24 to 30) | 12 (10 to 14) | 14 (12 to 16) |
| Less Affluent Older Singles, Couples and Single Parents (n=916) | 14 (12 to 17) | 29 (26 to 32) | 26 (23 to 29) | 16 (13 to 18) | 16 (13 to 18) |
| Less Affluent Mixed Family Neighbourhoods (n=1599) | 13 (11 to 15) | 29 (27 to 32) | 33 (31 to 35) | 12 (10 to 13) | 13 (11 to 15) |
| Low Income Rural and Suburban Neighbourhoods (n=1301) | 15 (13 to 17) | 24 (21 to 26) | 27 (24 to 29) | 14 (12 to 16) | 11 (9 to 13) |
| Diverse Low Income Urban Communities (n=219) | 11 (7 to 16) | 28 (22 to 35) | 25 (19 to 31) | 14 (10 to 19) | 16 (11 to 20) |
| Very Low Income Rural Communities (n=848) | 8 (6 to 10) | 26 (23 to 29) | 26 (23 to 30) | 12 (9 to 14) | 10 (8 to 12) |
| Whole cohort | 16 (15 to 16) | 28 (27 to 29) | 30 (29 to 31) | 13 (13 to 14) | 15 (14 to 15) |

Table 8.2 - Summary of ALSWH demographic characteristics

### 8.4.2 SOS1 - Descriptive statistics

The SOS1 participants are all from within the Seattle Urban Growth Boundary, which constitutes 90% of the population in King County, Washington. In order to assess how well Cameo groups are represented in King County and the SOS1 study population, compared to the whole USA, Figure 8.6 was created. SOS1 appears to track well with King County although both King County and SOS1 over represent groups such as American Aristocracy, Exclusive Society, and Prospering Families. Aspiring Consumers and Strained Society are under-represented.

Table 8.3 shows the demographic profile of SOS1 participants in each of the Cameo groups. The variables presented are all included in the Cameo classification, taken from the 2010 census, and appear to match well to the description of Cameo groups provided by CallCreditGroup, summarised in Chapter 3, Table 3.7. Due to the low number of participants falling into the Strained Society category, these results will be interpreted with caution as are likely to be unreliable.

The American Aristocracy group contains the highest percentage of those with annual household income over $100,000 (50%), the highest percentage of married couples (70%) and college graduates (71%), those in employment (66%) and the highest property values. They also have the lowest percentage of children under 18 living at home (60%). Interestingly it is those in Exclusive Society group who have the highest percentage who own their own homes (94%). The Diverse Communities have a similar percentage of those employed to the American Aristocracy (65% compared to 66%), however, have less than half who own their own homes (44%) and are married (30%), only 10% have household income over $100,000. They have the lowest percentage of college graduates (37%), lowest percentage of non-Hispanics (66%) and much lower property values (Table 8.3).

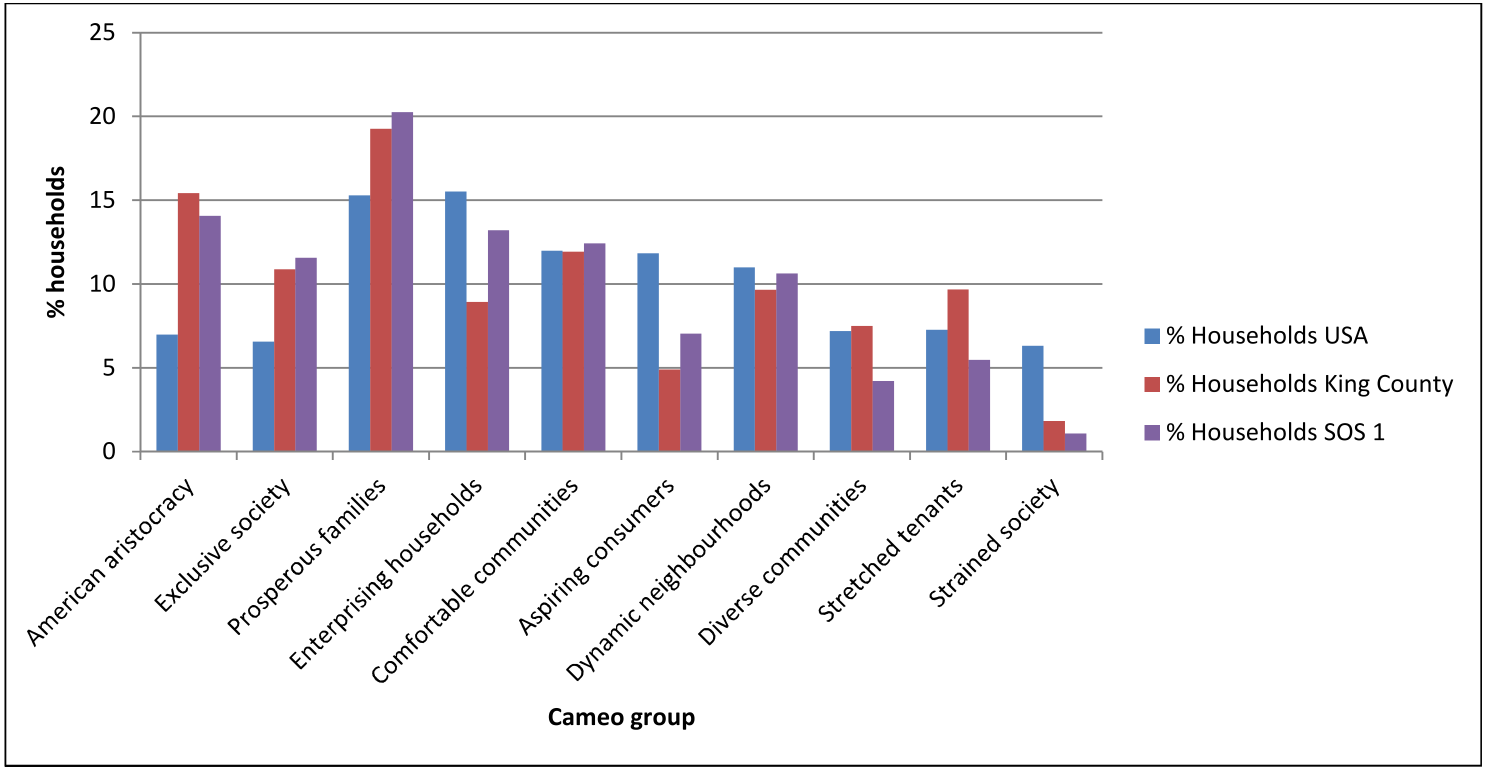



Figure 8.6 - Distribution of SOS participants compared to USA and King County citizens by Cameo geodemographic group

| Name | % who own their home (95% CI) | % married (95% CI) | % no children <18 at home (95% CI) | % over $100K household income (USD) (95% CI) | % employed & self employed (95% CI) | % college graduates (95% CI) | % Males (95% CI) | % non-Hispanic white (95% CI) | Mean Age (years) (95% CI) | Median Property value ($) (95% CI) |
|---|---|---|---|---|---|---|---|---|---|---|
| American Aristocracy (n=270) | 91 (88 to 95) | 70 (65 to 76) | 60 (54 to 65) | 50 (43 to 56) | 66 (60 to 72) | 71 (65 to 76) | 40 (34 to 46) | 84 (79 to 88) | 53.7 (52.1 to 55.4) | 453000 (351000 to 627000) |
| Exclusive Society (n=200) | 94 (91 to 98) | 64 (57 to 71) | 68 (61 to 75) | 41 (33 to 48) | 63 (56 to 69) | 67 (60 to 73) | 34 (27 to 40) | 84 (79 to 89) | 54.4 (52.3 to 56.4) | 427911 (259500 to 518500) |
| Prosperous Families (n=392) | 90 (87 to 93) | 58 (53 to 63) | 68 (64 to 73) | 29 (24 to 34) | 60 (55 to 65) | 58 (53 to 62) | 39 (34 to 43) | 83 (80 to 87) | 54.8 (53.3 to 56.3) | 327542 (243959 to 369000) |
| Enterprising Households (n=280) | 85 (81 to 89) | 50 (44 to 56) | 69 (63 to 74) | 18 (13 to 23) | 62 (56 to 67) | 48 (41 to 54) | 40 (34 to 45) | 76 (71 to 81) | 54.9 (53.1 to 56.7) | 283917 (225000 to 317000) |
| Comfortable Communities (n=228) | 79 (74 to 84) | 42 (36 to 49) | 78 (73 to 83) | 19 (14 to 25) | 62 (56 to 69) | 48 (41 to 54) | 38 (31 to 44) | 79 (74 to 85) | 55.4 (53.4 to 57.4) | 286022 (196000 to 324000) |
| Aspiring Consumers (n=137) | 83 (77 to 90) | 48 (40 to 57) | 66 (58 to 74) | 20 (13 to 27) | 59 (51 to 67) | 50 (42 to 59) | 32 (24 to 40) | 80 (73 to 87) | 55.7 (53.0 to 58.4) | 250797 (182000 to 292000) |
| Dynamic Neighbourhoods (n=217) | 64 (58 to 70) | 34 (27 to 40) | 80 (74 to 85) | 11 (6 to 15) | 56 (49 to 63) | 44 (38 to 51) | 39 (33 to 46) | 73 (67 to 79) | 55.7 (53.5 to 57.9) | 227674 (124156 to 279000) |
| Diverse Communities (n=107) | 44 (35 to 54) | 30 (22 to 39) | 71 (63 to 80) | 10 (4 to 16) | 65 (56 to 74) | 37 (27 to 46) | 38 (29 to 48) | 66 (57 to 76) | 49.6 (46.7 to 52.6) | 186594 (83228 to 247000) |
| Stretched Tenants (n=121) | 32 (24 to 41) | 24 (16 to 32) | 88 (83 to 94) | 15 (8 to 22) | 59 (50 to 68) | 53 (43 to 62) | 44 (35 to 53) | 82 (75 to 89) | 54.2 (51.2 to 57.1) | 173198 (69583 to 229193) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Strained Society (n=25) | 40 (19 to 61) | 28 (9 to 47) | 88 (74 to 100) | 8 (0 to 20) | 48 (27 to 69) | 32 (12 to 52) | 48 (27 to 69) | 83 (67 to 99) | 53.1 (47.9 to 58.3) | 135774 (101470 to 274000) |
| Whole cohort (n=1977) | 79 (77 to 81) | 50 (48 to 52) | 71 (69 to 73) | 25 (23 to 27) | 61 (59 to 63) | 54 (52 to 56) | 38 (36 to 40) | 80 (78 to 81) | 54.5 (53.8 to 55.2) | 269000 (206000 to 377000) |
| $Chi^2$ p | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.600 | 0.002 | 0.165 | <0.001^ |

Table 8.3 - Descriptive statistics for SOS by Cameo group

### 8.4.3 ALWSH - Fruit and vegetable consumption

The highest percentage of those adhering to fruit consumption recommendations, compared to average households, occurs in the Affluent Urban Professionals Cameo group, with 63% of the group meeting the guidelines. The lowest was observed in the Very Low Income Rural Communities, with only 46% meeting the guidelines. Results for the unadjusted logistic regression analysis show that all other groups consume significantly less fruit than the Affluent Urban Professionals, with the exception of the High Income Urbanities and Students, whose lower odds of meeting the guidelines is non-significant. The highest difference is observed in the Very Low Income Rural Communities whose odds of meeting the recommended fruit guidelines are 50% less.

Results for vegetable consumption are very different. The highest consumption of vegetables occurs in the Very Low Income Rural Communities of whom 46% meet the guidelines, compared to the lowest in the High Income Urbanities and Students of whom only 36% meet the guidelines. However, following logistic regression, all differences are statistically non-significant.

When the fruit and vegetable adherence is combined no statistical differences are observed, but the Affluent Urban Professional have the highest odds of meeting the guidelines.

| Cameo Name | Adherence to recommended fruit intake | | Adherence to recommended vegetable intake | | Fruit and vegetable intake combined as reported in general questionnaire | |
|---|---|---|---|---|---|---|
| | % meeting recommended fruit intake | Odds ratio (95% CI) p value | % meeting recommended vegetable intake | Odds ratio (95% CI) p value | % meeting recommended fruit and vegetable intake | Odds ratio (95% CI) p value |
| 1 - Affluent Urban Professionals (n=425) | 63 | 1 | 40 | 1 | 29 | 1 |
| 2 - Wealthy Family Neighbourhoods (n=1037) | 55 | 0.71 (0.57 to 0.90) **0.004** | 39 | 0.95 (0.75 to 1.21) 0.674 | 24 | 0.78 (0.60 to 1.01) 0.063 |
| 3 - High Income Urbanites and Students (n=293) | 56 | 0.75 (0.55 to 1.01) 0.060 | 36 | 0.87 (0.63 to 1.19) 0.381 | 23 | 0.72 (0.50 to 1.03) 0.069 |
| 4 - Comfortable Mixed Suburban Areas (n=1613) | 53 | 0.66 (0.53 to 0.82) **<0.001** | 39 | 0.96 (0.77 to 1.20) 0.711 | 24 | 0.79 (0.62 to 1.01) 0.059 |
| 5 - Mixed areas of modest detached homes (n=989) | 53 | 0.65 (0.52 to 0.82) **<0.001** | 42 | 1.09 (0.86 to 1.38) 0.474 | 24 | 0.77 (0.59 to 1.00) 0.049 |
| 6 - Less Affluent Older Singles, Couples and Single Parents (n=861) | 55 | 0.70 (0.55 to 0.89) **0.004** | 39 | 0.98 (0.77 to 1.26) 0.889 | 25 | 0.84 (0.64 to 1.10) 0.198 |
| 7 - Less Affluent Mixed Family Neighbourhoods (n=1521) | 52 | 0.63 (0.50 to 0.78) **<0.001** | 41 | 1.06 (0.85 to 1.34) 0.565 | 25 | 0.81 (0.63 to 1.04) 0.098 |
| 8 - Low Income Rural and Suburban Neighbourhoods (n=1233) | 55 | 0.70 (0.56 to 0.88) **0.002** | 40 | 1.00 (0.79 to 1.26) 0.974 | 27 | 0.91 (0.70 to 1.16) 0.439 |
| 9 - Diverse Low Income Urban Communities (n=210) | 52 | 0.64 (0.46 to 0.89) **0.009** | 40 | 1.07 (0.73 to 1.45) 0.861 | 22 | 0.70 (0.47 to 1.03) 0.074 |
| 10 - Very Low Income Rural Communities (n=798) | 46 | 0.50 (0.39 to 0.64) **<0.001** | 46 | 1.27 (0.99 to 1.62) 0.061 | 24 | 0.80 (0.61 to 1.05) 0.104 |
| Pseudo $R^2$ | | 0.0031 | | 0.0013 | | 0.0009 |
| $Chi^2$ p | <0.001 | | 0.084 | | 0.436 | |
| Mean value for the whole cohort | 53 | | 40 | | 25 | |

Table 8.4 - Fruit and Vegetable consumption in ALSWH; percentage meeting guidelines by Cameo group and unadjusted logistic regression analysis - displaying odds ratio (95% CI) p value (adherence reference group = 0)

### 8.4.4 SOS1 - Fruit and vegetable consumption

The American Aristocracy are the Cameo group with the highest proportion of individuals meeting the daily recommended intake of fruit with 42% meeting the guidelines (Table 8.5). The lowest percentage meeting the recommendations are in the Strained Society group, with only 32% consuming enough fruit to meet the guidelines. Following logistic regression analysis, all Cameo groups have a lower odds than the Affluent Urban Professionals of meeting the guidelines, but these differences were statistically non-significant.

The highest proportion meeting vegetable consumption recommendations were are the American Aristocracy with 41% meeting the guidelines. The lowest intakes were seen in the Strained Society (20%) and Enterprising Households (30%). Logistic regression analysis showed statistically significant differences with Enterprising Households being nearly 40% less likely to meet the recommendations with an odds ratio of 0.61 and the Strained Society group being over 60% less likely with odds ratio 0.36.

When adherence to the fruit and vegetable guidelines were combined, no significant difference was observed across the Cameo Groups.

| Cameo Name | Fruit intake (2 portions) | | Vegetable intake (3 portions) | | Fruit and vegetable intake combined (5 portions) | |
|---|---|---|---|---|---|---|
| | % meeting recommended fruit intake | Odds ratio (95% CI) p value | % meeting recommended vegetable intake | Odds ratio (95% CI) p value | % meeting recommended fruit and vegetable intake | Odds ratio (95% CI) p value |
| American Aristocracy | 42 | 1 | 41 | 1 | 33 | 1 |
| Exclusive Society | 41 | 0.93 (0.64 to 1.35) p=0.708 | 34 | 0.75 (0.51 to 1.10) p=0.137 | 32 | 0.95 (0.64 to 1.41) p=0.802 |
| Prosperous Families | 42 | 0.97 (0.71 to 1.33) p=0.870 | 32 | 0.69 (0.50 to 0.95) p=0.023 | 31 | 0.93 (0.67 to 1.30) p=0.690 |
| Enterprising Households | 38 | 0.82 (0.58 to 1.16) p=0.258 | 30 | **0.61 (0.43 to 0.87) p=0.007** | 30 | 0.89 (0.62 to 1.27) p=0.512 |
| Comfortable Communities | 44 | 1.07 (0.75 to 1.53) p=0.713 | 39 | 0.95 (0.66 to 1.36) p=0.774 | 39 | 1.30 (0.90 to 1.88) p=0.163 |
| Aspiring Consumers | 37 | 0.79 (0.51 to 1.20) p=0.266 | 33 | 0.71 (0.46 to 1.10) p=0.122 | 30 | 0.88 (0.57 to 1.38) p=0.585 |
| Dynamic Neighbourhoods | 37 | 0.82 (0.57 to 1.18) p=0.273 | 33 | 0.71 (0.49 to 1.03) p=0.069 | 31 | 0.94 (0.64 to 1.38) p=0.768 |
| Diverse Communities | 34 | 0.69 (0.43 to 1.11) p=0.126 | 35 | 0.77 (0.48 to 1.23) p=0.269 | 33 | 1.00 (0.62 to 1.62) p=0.982 |
| Stretched Tenants | 36 | 0.75 (0.48 to 1.18) p=0.213 | 35 | 0.77 (0.49 to 1.21) p=0.259 | 36 | 1.14 (0.73 to 1.79) p=0.569 |
| Strained Society | 32 | 0.64 (0.27 to 1.54) p=0.324 | 20 | **0.36 (0.13 to 1.00) p=0.050** | 24 | 0.65 (0.25 to 1.69) p=0.381 |
| $R^2$ | | 0.5529 | | 0.1244 | | 0.6501 |
| $Chi^2$ | 0.557 | Pseudo $R^2$ 0.0029 | 0.128 | 0.0055 | 0.642 | 0.0028 |

Table 8.5 - Compliance to fruit and vegetable recommendations in SOS1

### 8.4.5 ALWSH - Weight status

In the ALWSH cohort there is a significant difference in mean BMI across the Cameo groups ($chi^2$ $p<0.001$). On closer inspection by BMI category, the differences lie in the normal weight and the obese (Table 8.6). Differences in weight status in each category, compared to the whole cohort, can be seen in Figure 8.7.

Using linear regression, compared to the Affluent Urban Professional, all groups have an increased BMI. With the exception of the High Income Urbanities and Students group, these differences are statistically significant. The Diverse Low Income Urban Communities group have an increase of 2.43 $kg/m^2$ compared to the Affluent Urban Professionals. When considering the odds of being obese, the Diverse Low Income Urban Communities and the Very Low Income Rural Communities are twice as likely to be obese than the Affluent Urban Professionals (OR 2.34 and 1.99 respectively) (see Table 8.6).

| Cameo group and name | % Underweight (95% CI) | % Normal weight (95% CI) | % Overweight (95% CI) | % Obese (95% CI) | Mean BMI kg/m$^2$ (95%CI) | BMI - unadjusted Coef. (95% CI) p value | Obese compared to not obese - unadjusted OR (95% CI) p value |
|---|---|---|---|---|---|---|---|
| 1 - Affluent Urban Professionals (n=445) | 1 (0 to 2) | 51 (47 to 56 ) | 29 (25 to 33) | 19 (15 to 23) | 25.5 (25.1 to 26.0) | Ref. 0.00 | Ref. 1.00 |
| 2 - Wealthy Family Neighbourhoods (n=1079) | 2 (1 to 2) | 41 (38 to 44) | 31 (28 to 34) | 27 (24 to 29) | 26.7 (26.3 to 27.0) | 1.14 (0.52 to 1.76) **<0.001** | 1.55 (1.18 to 2.04) **0.002** |
| 3 - High Income Urbanites and Students (n=308) | 2 (0 to 3) | 45 (39 to 50) | 26 (21 to 31) | 28 (23 to 33) | 26.5 (25.8 to 27.1) | 0.95 (0.12 to 1.77) 0.024 | 1.64 (1.16 to 2.31) **0.005** |
| 4 - Comfortable Mixed Suburban Areas (n=1704) | 1 (1 to 2) | 40 (38 to 43) | 29 (27 to 31) | 29 (27 to 31) | 26.7 (26.4 to 27.0) | 1.17 (0.58 to 1.76) **<0.001** | 1.77 (1.37 to 2.29) **<0.001** |
| 5 - Mixed areas  of modest detached homes (n=1046) | 1 (1 to 2) | 39 (36 to 42) | 32 (29 to 34) | 28 (26 to 31) | 26.8 (26.5 to 27.2) | 1.34 (0.72 to 1.96) **<0.001** | 1.70 (1.30 to 2.24) **<0.001** |
| 6 - Less Affluent Older Singles, Couples and Single Parents (n=916) | 2 (1 to 3) | 39 (35 to 42) | 30 (27 to 33) | 30 (27 to 33) | 27.0 (26.6 to 27.4) | 1.49 (0.85 to 2.13) **<0.001** | 1.81 (1.37 to 2.24) **<0.001** |
| 7 - Less Affluent Mixed Family Neighbourhoods (n=1599) | 1 (1 to 2) | 38 (36 to 40) | 29 (27 to 32) | 31 (29 to 34) | 27.0 (26.7 to 27.3) | 1.48 (0.89 to 2.07) **<0.001** | 1.97 (1.52 to 2.55) **<0.001** |
| 8 - Low Income Rural and Suburban Neighbourhoods (n=1301) | 1 (0 to 2) | 40 (38 to 43) | 32 (30 to 35) | 27 (24 to 29) | 26.6 (26.3 to 26.9) | 1.07 (0.46 to 1.68) **0.001** | 1.56 (1.19 to 2.03) **0.001** |
| 9 - Diverse Low Income Urban Communities (n=219) | 1 (0 to 2) | 37 (30 to 43) | 28 (22 to 34) | 34 (28 to 41) | 27.9 (27.1 to 28.7) | 2.43 (1.53 to 3.34) <**0.001** | 2.24 (1.55 to 3.23) **<0.001** |
| 10 -Very Low Income Rural Communities (n=848) | 1 (0 to 2) | 35 (32 to 38) | 32 (29 to 35) | 32 (28 to 35) | 27.7 (27.3 to 28.1) | 2.18 (1.53 to 2.82) **<0.001** | 1.99 (1.50 to 2.62) **<0.001** |
| Whole cohort | 1 (1 to 2) | 40 (39 to 41) | 30 (29 to 31) | 29 (28 to 30) | 26.8 (26.7 to 26.9) | | |
| Chi$^2$ p value | 0.904 | **<0.001** | 0.301 | **<0.001** | <0.001 | **<0.001** | **<0.001** |
| $R^2$ or Pseudo $R^2$ | | | | | | $R^2$ = 0.0069 | Pseudo $R^2$ = 0.0036 |

Table 8.6 - Weight status by Cameo group in ALWSH

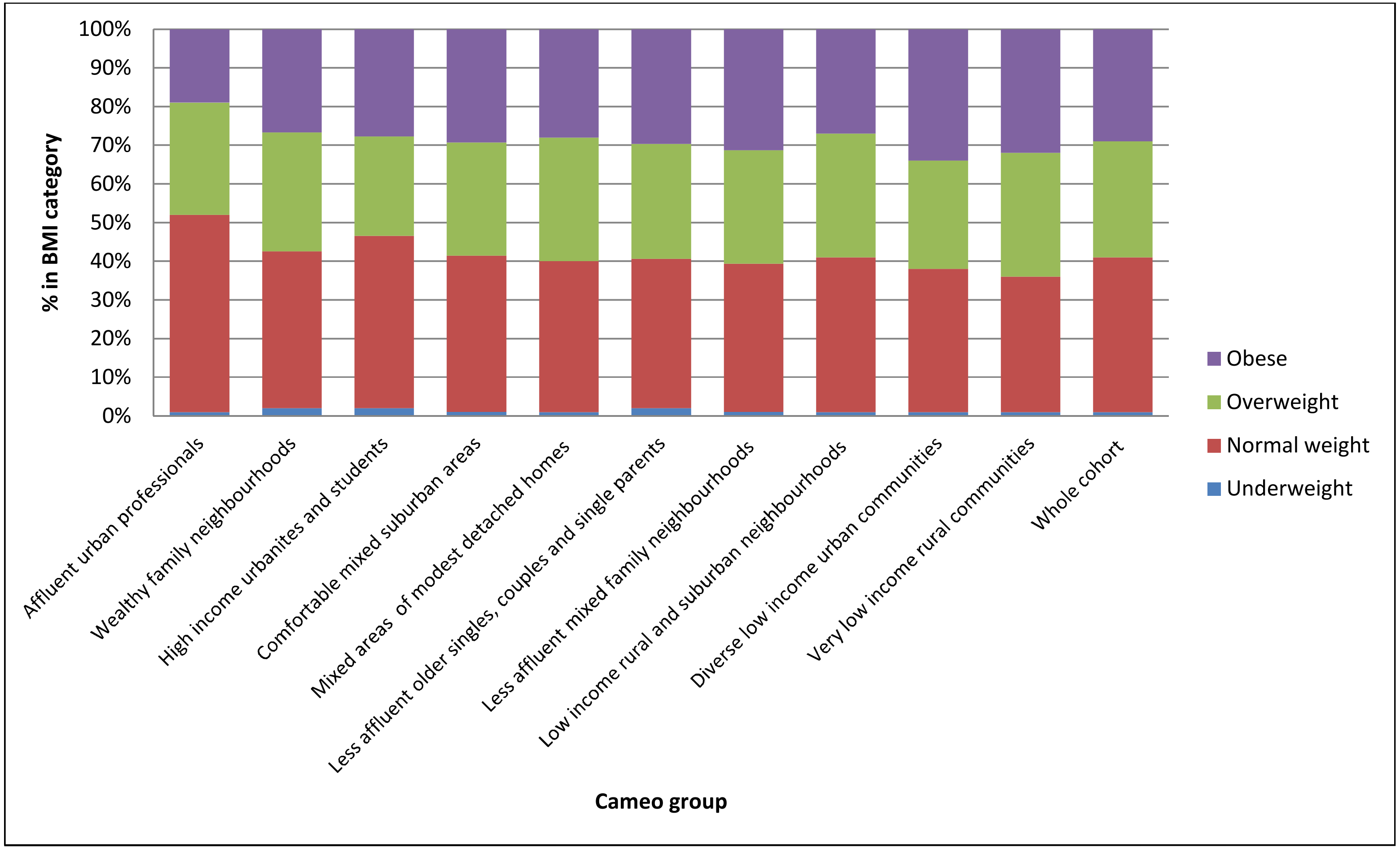



Figure 8.7 - Weight status by Cameo group in ALWSH women

### 8.4.6 SOS1 - Weight status

BMI was skewed in this sample so median BMI was presented and non-parametric tests for difference carried out. Robust standard errors were applied to the regression analysis. Statistically significant differences in median BMI exist between Cameo groups (Kruskal Wallis $p<0.001$). The highest BMI was observed in the Enterprising Households (26.3 $kg/m^2$) and lowest in the American Aristocracy (24.4 $kg/m^2$).

When investigating according to BMI category, these differences exist in the normal weight and obese categories. Dynamic Neighbourhoods have the highest percentage of obese (26%) and Exclusive Society the lowest (12%). The lowest percentage of normal weight is seen in the Aspiring Consumers (38%) compared to the highest in the American Aristocracy and Exclusive Society (both 51%) (See Table 8.7). These differences in weight status by Cameo group compared to the whole cohort can be seen in Figure 8.8.

Linear regression results show that compared to the American Aristocracy, all except those in Exclusive Society are significantly more likely to have a higher BMI. The Stretched Tenants have an increase in BMI which is nearing significance. This difference is not statistically significant for all groups. In fact, the following three show no statistical difference from the American Aristocracy: Excusive Society, Stretched Tenants and Strained Society.

When considering those who are obese, compared to those who are not, it can be seen that only Enterprising Households, Comfortable Communities and Dynamic Neighbourhoods have a significant increase in odds of being obese compared to American Aristocracy.  For these groups the odds are nearly double (ORs 1.97, 2.01 and 2.09 respectively). Interestingly, the Aspiring Consumers do not have significantly increased odds of being obese, despite such a difference in the BMI coefficient.

| **Cameo group and name** | **% Underweight (95% CI)** | **% Normal weight (95% CI)** | **Overweight (95% CI)** | **Obese (95% CI)** | **Median BMI kg/m$^2$ (95% CI)** | **BMI - unadjusted Coef. (95% CI) p value** | **Obese compared to not obese - unadjusted OR (95%CI) p value** |
|---|---|---|---|---|---|---|---|
| 1 - American Aristocracy (n=256) | 2 (0 to 4) | 51 (45 to 57) | 33 (27 to 39) | 14 (10 to 18) | 24.4 (22.0 to 27.5) | Ref 0.00 | Ref 1.00 |
| 2 - Exclusive Society (n=190) | 1 (0 to 3) | 51 (43 to 58) | 37 (30 to 44) | 12 (7 to 16) | 24.8 (22.0 to 28.2) | 0.49 (-0.41 to 1.40) p=0.286 | 0.80 (0.45 to 1.41) p=0.441 |
| 3 - Prosperous Families (n=392) | 2 (0 to 3) | 45 (40 to 50) | 33 (28 to 37) | 20 (16 to 25) | 25.1 (22.9 to 28.9) | 1.23 (0.46 to 2.00) **p=0.002** | 1.56 (1.01 to 2.41) p=0.043 |
| 4 - Enterprising Households (n=263) | 1 (0 to 2) | 37 (31 to 42) | 38 (32 to 44) | 24 (19 to 30) | 26.3 (23.2 to 29.6) | 1.89 (1.03 to 2.75) **p<0.001** | 1.97 (1.25 to 3.09) **p=0.003** |
| 5 - Comfortable Communities (n=214) | 1 (0 to 3) | 39 (32 to 45) | 35 (29 to 41) | 25 (19 to 31) | 25.7 (23.0 to 30.0) | 1.73 (0.77 to 2.70) **p<0.001** | 2.01 (1.26 to 3.22) **p=0.004** |
| 6 - Aspiring Consumers (n=130) | 0 | 38 (29 to 46) | 40 (31 to 49) | 22 (15 to 30) | 26.1 (22.9 to 29.5) | 2.22 (0.97 to 3.46) **p<0.001** | 1.75 (1.02 to 3.02) p=0.042 |
| 7 - Dynamic Neighbourhoods (n=200) | 3 (0 to 5) | 42 (35 to 49) | 30 (24 to 36) | 26 (19 to 32) | 25.8 (23.0 to 30.2) | 1.90 (0.92 to 2.88) **p<0.001** | 2.09 (1.30 to 3.36) p**=0.002** |
| 8 - Diverse Communities (n=99) | 0 | 35 (26 to 45) | 40 (31 to 50) | 24 (16 to 33) | 26.5 (23.3 to 29.9) | 1.88 (0.81 to 2.95) **p=0.001** | 1.96 (1.10 to 3.49) p=0.023 |
| 9 - Stretched Tenants (n=113) | 4 (1 to 8) | 41 (32 to 50) | 34 (25 to 42) | 21 (14 to 29) | 25.1 (23.6 to 28.6) | 1.36 (0.11 to 2.61) p=0.033 | 1.65 (0.93 to 2.92) p=0.087 |
| 10 - Strained Society (n=23) | 0 | 43 (22 to 65) | 43 (22 to 65) | 13 (0 to 28) | 25.8 (23.4 to 28.7) | 1.14 (-0.77 to 3.05) p=0.243 | 0.92 (0.26 to 3.24) p=0.893 |
| Whole cohort (1856) | 2 (1 to 2) | 43 (41 to 45) | 35 (33 to 37) | 21 (21 to 22) | 25.7 (22.9 to 29.0) | | |
| Chi$^2$ p value | 0.17 | 0.007 | 0.482 | 0.002 | <0.001* | | |
| $R^2$ and pseudo $R^2$ | | | | | | $R^2$ 0.0162 | Pseudo $R^2$ 0.0144 |

Table 8.7 - Weight status by Cameo group in SOS1. *Kruskal wallis for difference in BMI between groups

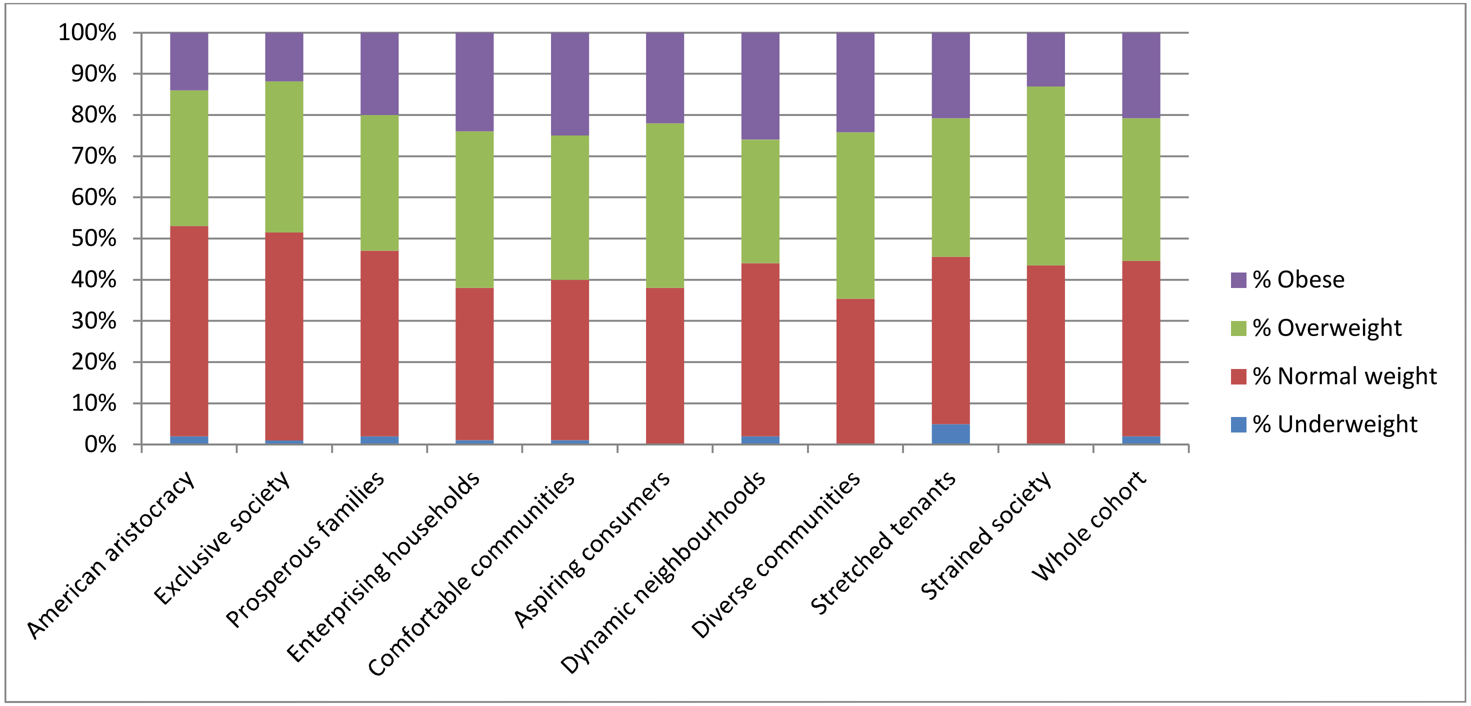



Figure 8.8 - Weight status by Cameo group in SOS1

### 8.4.7 ALSWH - Estimated obesity for Newcastle, Australia

Using results of the logistic regression, it is possible to map odds of being obese, compared to Affluent Urban Professionals for an Australian City. Newcastle has been chosen as an example. Newcastle is a medium sized city on the east coast, north of Sydney and is home to one of the universities hosting the ALSWH data.

Figure 8.9 shows the odds of being obese at a mesh block unit of geography (the unit of geography used by the Cameo classification). Areas which are white inside of the boundary lines are a result of unmatched data. Small pockets of high odds of obesity exist, highlighted by the red colour.

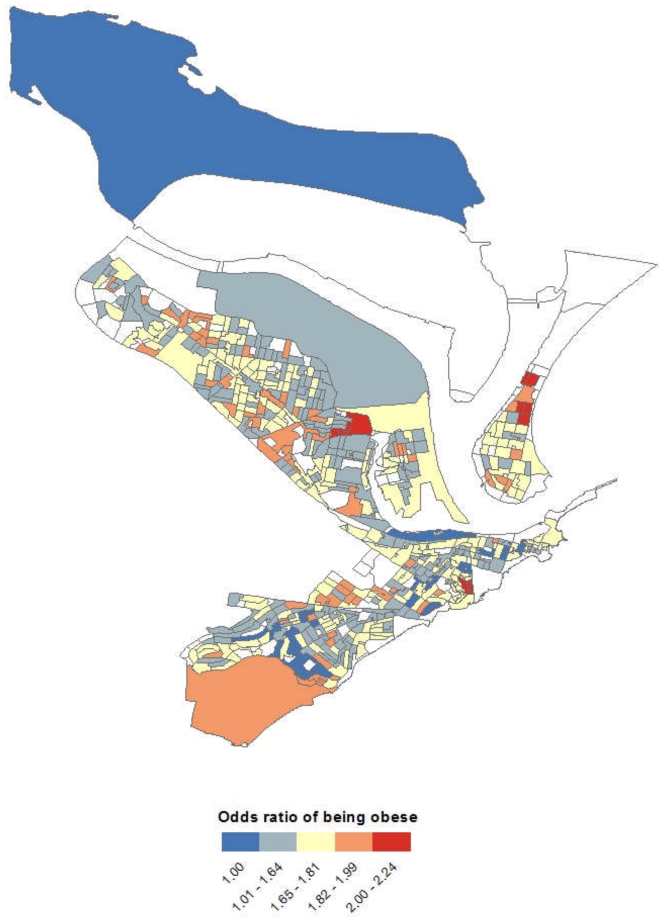



Figure 8.9 - Odds of being obese for Newcastle inner city

### 8.4.8 SOS1 - Estimated obesity for Seattle, WA, USA

Using the results of the logistic regression, it is possible to map odds of being obese, compared to American Aristocracy for Seattle at a census block level (the unit of geography used by the Cameo classification). Results are illustrated in Figure 8.10. Lowest odds of obesity are shown in blue. These highlight the more affluent areas along the waterfront (water shown in white) and on Mercer Island (the large island in the middle of the map).

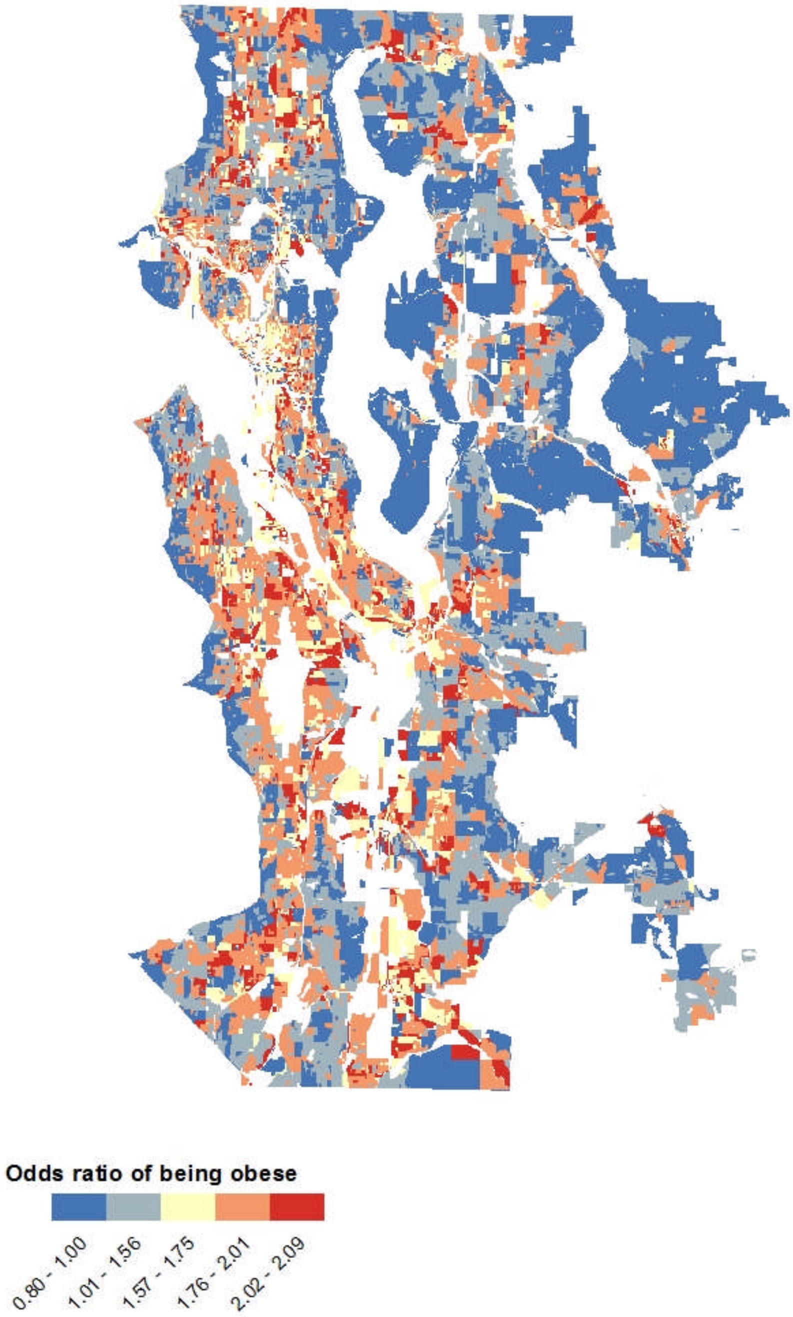



Figure 8.10 - Odds of being obese for Seattle urban growth boundary.

Seattle urban growth boundary file:"Data provided by permission of King County".

## 8.5 Discussion

This chapter demonstrates that using geodemographic classifications may be useful in diet and health research internationally. While results are of an exploratory nature, it seems worthwhile to invest more time in the future completing further analysis.

*Cameo classification*

This chapter was not intended to critique the Cameo classification, but the descriptive statistics show that the classifications match well with the socioeconomic characteristics of the cohort samples. The Australian Cameo classification used a number of sources of data to construct the classification, whereas the USA classification was created solely using census variables. This will mean that the two classifications are not comparable, but this research is to explore whether using 'off the peg' resources are valuable in diet and health research, so these differences are not directly relevant.

Summarising the characteristics of the ALSWH women by their Cameo group provides a crude evaluation of the Cameo classification showing expected gradients in income, occupation and education. The same is seen in the SOS1 sample. Interestingly, the categories appear to correlate well with property value in SOS1.

*Generalisability of sample to general population by Cameo group*

The SOS1 sample comprises men and women from Seattle, King County. As shown in Figure 8.6 the distribution of Cameo groups in King County is very different from other areas of the USA. It seems feasible to suggest that the behaviours of Seattle residents may vary to those from other areas in the USA. Using Seattle obesity estimates from SOS1 data to populate estimated obesity for each census block is reasonable. Further analysis would need to be carried out using data from other parts of the USA in order to assess whether results are transferrable across counties or even states.

Comparing the ALWSH women to Australian population by the Cameo groups in which they reside shows some over and under representation of certain groups. However, there are sufficient numbers in each group to have confidence in the analysis. The ALSWH women reside throughout Australia, which means that there is no large concentration in any one state. The difference in climates and access to foods in different regions of Australia may vary greatly. This could impact on the reliability of estimates for a specific city. Perhaps a bigger sample would provide greater confidence in the estimates in the USA. However, having 2000 participants from Seattle to estimate King County patterns may be more reliable that having 9500 women from across Australia to estimate patterns in one Australian city.

*Comparability of Australian and American analysis*

While the Australian and USA analysis both use the Cameo classification to profile dietary components and weight status, they are not directly comparable. Both classifications have a link to SES gradient and compared to the Affluent Urban Professionals in Australia and the American Aristocracy in the USA most other groups have increased odds of being obese. This may not appear to offer any further evidence than using other markers of SES such as income or education. However, the Cameo classification can identify areas at a small geographic unit (Mesh block in Australia and census block in the USA) which could be beneficial to policy makers for targeting interventions.

*Fruit and vegetable consumption*

Difference in the national recommendations for fruit and vegetable consumption, between countries, prevent results from being directly comparable. However, the findings show stark differences. In Australia the Affluent Urban Professional group contains the highest proportion of women meeting the recommended guidelines. They also have significantly higher odds compared to almost all other groups of meeting these recommendations. In the Seattle sample, the American Aristocracy also have the highest percentage of participants meeting the guidelines, but there is no significant difference in the odds of meeting the guidelines between the groups. The sample is far more homogenous in their fruit consumption. From this research it is not possible to tell whether this is due to differences in availability of fruit, cultural attitudes or even just differences in gender. Other studies have shown that association between diet healthiness and diet cost have been stronger in women (Monsivais and Drewnowski, 2009, Rehm et al., 2011, Lopez et al., 2009a). Perhaps the heterogeneity in fruit consumption in the ALWSH cohort is a result of all participants being women, rather than them being in Australia. Without a male comparison dataset, it is not possible to draw any further conclusion on this.

Differences in vegetable consumption - by Cameo group - in both samples are different. In ALWSH the Very Low Income Rural Communities women had the highest vegetable consumption, but the odds of meeting the guidelines were not significantly different in any of the other groups. In the SOS1 sample, the Enterprising Households were significantly less likely to meet the guidelines than the American Aristocracy, suggesting there is something quite different about this group which is influencing intake. The Strained Society group were also significantly less likely to meet the recommendations, however results from this group should be treated with caution as the sample size is small (n=25) and the confidence interval about the odds ratio wide, incorporating 1. The highest percentage of all groups meeting the recommended guidelines for fruit and vegetables was only 39%, suggesting there is some way to go to meeting the recommendations. This is recognised by the USDA who encourage, more fruit and vegetable consumption in their Dietary Guidelines for Americans, Food and Nutrients to Increase Chapter (United States Department of Agriculture, 2010).

Combining the adherence to fruit and vegetable guidelines does not add to the results. The lesser effect seen in the vegetable adherence dilutes the association seen with the fruit guidelines. In the 1995 National Nutrition Survey in Australia it was shown that ethnic vegetable consumption (e.g. courgettes and peppers) significantly increases with increased SES (measured by employment status) but traditional vegetable consumption decreases with increased SES (Mishra et al., 2002). This could be supportive of patterns observed in the Cameo groups with the Very Low Income Rural Communities having the highest proportion of women meeting vegetable intake guidelines. Further analysis of specific vegetable consumption would be required to be certain. The same survey reports how likely individuals are to have consumed fruit or vegetables on the previous day. Adult women in the lowest quintile of income were 2.5 times more likely not to have consumed fruit the previous day, compared to being 1.6 times more likely not to have consumed vegetables (Giskes et al., 2002). While in this national sample there is a reduced consumption of vegetables with reduced income, it is not as steep an association as that associated with fruit consumption, but is opposite to the findings in the ALSWH using Cameo groups. This might be explained by the fact that Cameo groups contain more demographic information than just income.

*Weight status*

Difference in odds of being obese are as large as two, indicating that some groups are twice as likely to be obese than the affluent reference categories; Affluent Urban Professionals (Australia) and American Aristocracy (USA). The logistic regression models for this analysis were unadjusted, so as not to over adjust demographic variables included in the Cameo classification. In Australia the Diverse Low Income Urban Communities and the Very Low Income Rural Communities are twice as likely to be obese than the Affluent Urban Professionals (OR 2.34 and 1.99 respectively). This is supportive of other research which shows that those of lower SES have increased likelihood of being obese (Drewnowski and Specter, 2004).

The same is seen in the SOS1 study, but this time the groups who have highest odds of being obese are those whose SES is somewhere in the middle of the range. The Enterprising Households, Comfortable Communities and Dynamic Neighbourhoods have odds of about double that of the American Aristocracy and are statistically significant. Further analysis to try and investigate what is so different about these groups could aid understanding about the determinants of health behaviours.

*Predicting small area patterns*

Making the assumption that obesity risk is the same for all people in a specified Cameo group means that small area patterns can be predicted. This is a large assumption, but one which seems reasonable. Care needs to be taken with regards to generalisability of the cohort sample to the general population. Further analysis to validate the estimations is essential in order to test how well estimate agree with real data. However, this method is a low cost tool which could be applied to any large cohort or national survey in order to better understand behaviours of health outcomes of a specific population at local level unit geography. It seems intuitive to make the best use of classifications already developed to target like populations, even though it was developed for marketing purposes, not health needs.

*Public health relevance of results*

The Cameo classification is a useful tool in order to target obesity interventions in both Australia and the USA as certain groups have been identified as having an increased likelihood of being obese. This could allow neighbourhood level interventions to be initiated.

*Strengths and limitations*

While this research has been worthwhile, there are limits to its usefulness in public health. Not knowing the exact methods for developing the Cameo classification means that it cannot be critiqued with respect to its use in predicting dietary consumption patterns or health outcomes; however this is common place for developers of such classifications not to share methods as they are intellectual property. In the UK, the Output Area Classification (OAC) is open source, so perhaps open source Australian and USA alternatives could be sought or funded to facilitate future research.

The sample size in both sets of analysis limits the transferability of findings. While the SOS1 sample only included 2000 participants (men and women), the ALSWH contained 9500 women. Larger sample sizes may have provided more confidence in the results. However, the quality of the dietary data in these samples was excellent. There is scope to investigate more complex dietary components and health outcomes within the existing data.

*Future work*

It would be possible to combine all three age cohorts of the ALSWH cohorts to increase sample size for further analysis and estimation of small area patterns. In the USA, if it were possible to add geodemographic identifiers to a cohort such as NHANES, there would be scope for some interesting work.

The Cameo classification is available in 40 countries internationally. This classification could be applied to the UKWCS and cross country comparison carried out in the UK, Australia and the USA. It would also be interesting to compare dietary and weight status variations when using the Cameo classification compared to the OAC classification. The cameo classification was not used for the UK analysis as the methods used to create it are not transparent as they were for the OAC.

CallCreditGroup who author the Cameo classification also supply an international classification. This is a 25 category classification taking into account life stage and wealth and is comparable for all countries. With it only using two demographic variables, it is not as rich as each county classification in its own right (which use dozens of variables) and would not be as useful for predicting small area patterns, but may be one tool for consistent cross country comparisons.

## 8.6 Summary and context

This chapter has explored the use of a geodemographic classification as a method to profile diet and weight status in Australia and the USA.

- Results show that the demographic characteristics in the ALSWH and SOS1 cohorts correspond well with the demographic characteristics of the country specific Cameo classification.
- In the ALSWH cohort all groups are significantly less likely to adhere to recommendations for fruit intake, compared to the Affluent Urban Professionals. However, no significant variation is seen for vegetable intake.
- In the SOS1 cohort no significant variation is seen in adherence to fruit recommendations. The Enterprising Households groups are less likely to adhere to vegetable intake recommendations compared to the American Aristocracy.
- Significant variations in odds of being obese exist in both the ALSWH and SOS1 cohort. Affluent Urban Professionals have the lowest odds of being obese in the ALSWH and the American Aristocracy and Exclusive Society groups have lowest odds in SOS1.
- Estimating odds of obesity using this type of data could have impact in potential public health interventions.

Chapter 9 will now discuss the key messages found throughout this thesis, the strengths and limitations of the work, its relevance to public health and potential future work in this area.

# Chapter 9: Discussion

## 9.1 Overview

This Chapter will discuss the findings throughout the thesis, considering strengths and limitations. Suggestions for future work will be made along with messages for policy makers.

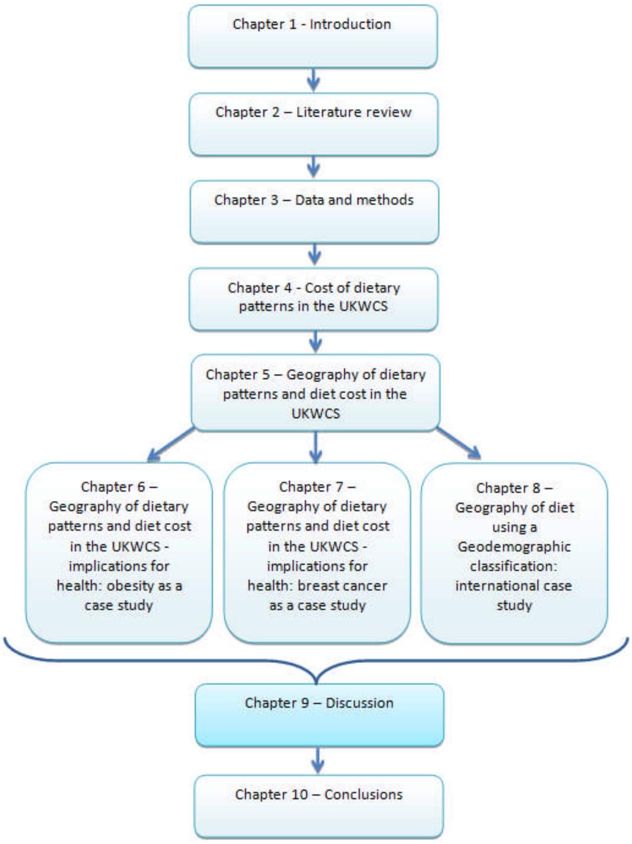



Figure 9.1 - Overview showing how Chapter 9 fits into the overall thesis flow

## 9.2 Key messages

This thesis has demonstrated (as illustrated in Figure 9.2) that:

- Spatial measures influence diet cost, dietary pattern consumption, weight status and breast cancer incidence, as shown by the blue arrows.
- Diet cost influences dietary pattern, in that a healthy diet does cost more than a less healthy one. The diet cost increases with increasing BMI category. These relationships are illustrated by the red arrows.
- Dietary pattern consumption influences breast cancer incidence, although this seems to be independent of diet cost, illustrated by the green arrow.
- A positive association exists between weight status risk of breast cancer incidence, but not in premenopausal women, illustrated by the purple arrow. This also appears to be independent of diet cost.

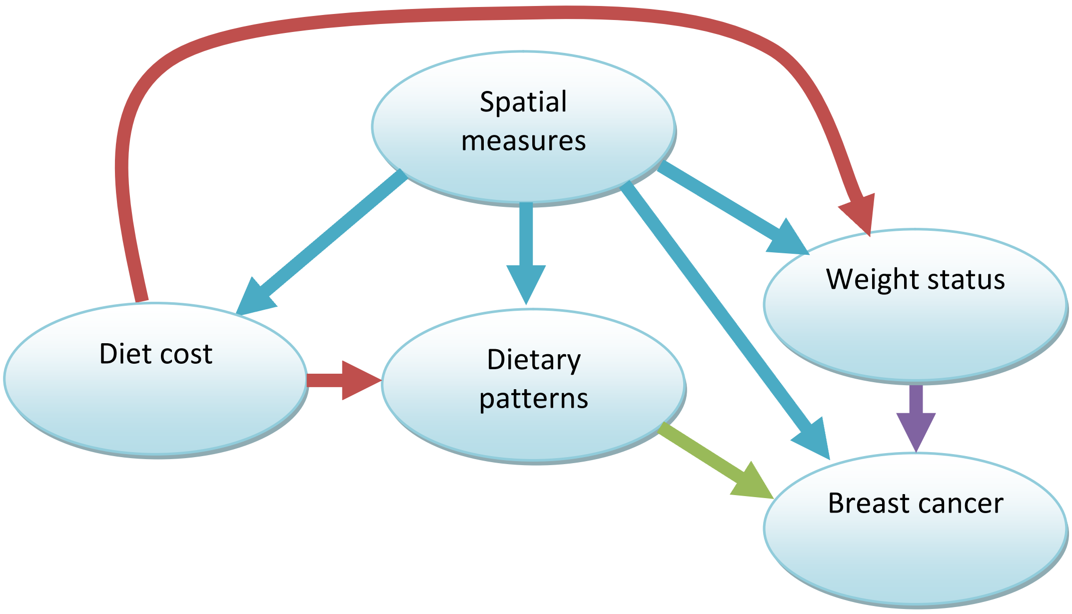



Figure 9.2 - Key messages

The work also shows that there is potential to use geodemographic classifications as a tool in diet and public health research, to:

- Profile diet and health behaviours so interventions are more effectively targeted
- Estimate small area diet and health patterns in the UK and other developed countries, such as Australia and the USA.

### 9.2.1 Importance of the research area

Investigation of determinants of dietary behaviour is essential in attempt to combat the rise in diet related chronic diseases. Cost of diet, dietary patterns consumed and the environment in which an individual resides are three important aspects to be considered.

Most previous studies have found that a healthier diet does in fact cost more. This is particularly evident in areas of low income populations, whose economic and physical access to healthy foods is limited. However, studies have rarely combined valid methods of dietary assessment with application of costs from a reliable cost database, with small scale geographic units as this thesis does. In Chapter 4, results concur with those reported elsewhere that a healthy diet does indeed cost more, and diet cost does appear to vary depending upon where the women live. Along with diet cost, consumption of different dietary patterns appears to be influenced by where the women live (Chapter 5).

Using this large cohort of women, it has been possible to profile dietary patterns and diet cost by OAC Supergroup and estimate small area dietary patterns for Leeds in the UK. This could prove to be a valuable tool for local health authorities to target diet and health interventions. Exploratory work in Australia and the USA also show that this method of estimating small area diet and health could be beneficial there too.

Diet constitutes one half of the energy balance equation. The imbalance in energy intake and expenditure over recent decades has led to the high prevalence of obesity which is causing such a personal and economic burden on society in the UK and worldwide (Foresight, 2007) (World Health Organisation, 2000). Chapter 6 has explored the spatial variations in obesity and the influence of dietary pattern and diet cost. Results suggest that combining a large geographical unit such as GOR with smaller scale geography like the OAC could add a more meaningful understanding of diet and health of specific populations which could benefit public health

Breast cancer affects more people in the UK than any other cancer despite it predominantly only affecting women. In the spatial analysis, little association between incidence and space is observed. The South West has a higher Incidence of breast cancer than other regions, but it is difficult to tease out what it is about this region which is different. The composition, measured by the distribution of OAC groups does not differ in extremes from Wales, except perhaps fewer Blue Collar Communities, yet Wales does not have higher incidence of breast cancer. Prospering Suburbs and Multicultural areas have a higher incidence, although this result is non-significant.

### 9.2.2 Target population

This research carried out in the UK only includes women. While the methods are transferrable to populations including men, the results cannot be assumed for men. Other studies show that stronger associations in diet cost and a healthy diet are more strongly associated in women (Monsivais and Drewnowski, 2009, Rehm et al., 2011), which suggests that the behaviour of men towards diet differs. Difference in obesity prevalence in the UK differs between men and women, again suggesting that the genders may respond differently to drivers of dietary behaviour.

### 9.2.3 Mechanisms of associations

The work in this thesis assumes a link between cost of diet, dietary consumption and health behaviours. The influences of place of residence, including contextual environment, incorporating demographic variables is the suggested mechanism for differences in the cost-consumption-outcome pathway.

## 9.3 Strengths and Limitations

### 9.3.1 Data sources

*Cohort data*

This research is a secondary data analysis of existing prospective cohort data. The UKWCS dietary data was collected prior to development of breast cancer so eliminates recall bias and reverse causality possibilities which may be associated with case control designs. The large sample size adds power to the research.

*Diet data*

Dietary data was collected in the 1990s, so while prices were also from that time, it may not reflect dietary patterns of women in 2013. This could be considered a substantial limitation of applicability to determinants of dietary behaviour today. However, it is still worthwhile to carry out the cross sectional analysis on this sample as it is such a large sample with quality dietary data on UK women. Other research does not typically use geo-coded data from such a large sample. Since inequalities in health remain a problem in the UK, analysis investigating variation in diet and subsequent health, using geographic identifiers and socioeconomic markers is well timed.

*Diet cost data*

Use of a diet cost database, which has been validated, is unique. The size of the database is also a great strength as other research has used costs from databases containing far fewer foods. Often research which uses spatial variables to investigate cost of dietary patterns tend to use measures such as the cost of a 'healthy basket' in order to infer healthiness of diet, which is not the same as cost of diet as consumed (Cummins et al., 2010, Larsen and Gilliland, 2009).

*Spatial measures*

As with much research using spatial measures, there is some risk that applying individual level dietary data and discussing results at a more aggregated geographical scale there may be an element of discrepancy as a result of ecological fallacy. A relationship observed at an area level - even a small area such as output area - may not be typical of an individual, or conversely the effect of the individual may infer a relationship to an area, which is actually driven by the individual. In the UKWCS where the women are typically middle class, there is a chance that they are not typical of the area they live in, even though they are well distributed across different types of area. Results therefore are only relevant at the area scale which they were studied.

The Modifiable Areal Unit Problem (MAUP) should also be considered as the location of the women's postcode may vary depending on the timeframe at which that postcode was aggregated to a larger geographical scale. If boundaries of output areas change with time, then these results which were determined by boundaries at a different set time may no longer remain applicable (Openshaw, 1984).

Using the range of spatial measures covered in this thesis, the effect of context verses composition of the geographical environment can be considered. It seems that in the majority of cases combining context and composition, using the Output Areas Classification is the most effective method of identifying variation in dietary behaviour.

*Obesity*

Measurement of weight status using BMI is accepted worldwide. It does, however, have limitations with respect to body composition and can place athletic individuals with a high proportion of lean muscle mass at a higher weight status than they would be should their body composition have been measured by alternative methods. It is also reliant on accurate measurements of height and weight, both of which are self reported in the UKWCS. This may introduce potential reporting bias. It is well documented that the obese under-report both their weight and dietary consumption (Willett, 2013).

Physical activity is taken into account in this research by adjusting for daily METS values for physical activity. This data was derived from a 24 hour physical activity diary. This research did not investigate physical activity as a determinant of dietary behaviours and subsequent obesity or breast cancer outcome, but it was included in the regression analysis to control for confounding effects. It cannot be ruled out that physical activity is a stronger driver for weight status and further analysis would be required to investigate this.

*Breast cancer*

Data on breast cancer incidence in this sample is reported directly from the registry data to the Nutritional Epidemiology Group. This does not rely on the participant in any way so increases reliability of the data. Diagnoses of breast cancer are recorded using the International Classification of Diseases and health related problems. The morphology of the neoplasm is not included in this analysis, only incidence of any type of breast cancer. The long term outcome of the breast cancer is also not considered, but it would be possible to investigate deaths caused by breast cancer in the UKWCS in future analysis. There is some lag time in receiving notification of cancer incidence, despite quarterly updates from the NHSIC. No statistical comparison to incidence of breast cancer in the whole UK population by different spatial measures has been included in this analysis, so it is difficult to conclude whether the UKWCS is different to those in the whole of the UK.

## 9.4 Relevance to Public Health Policy

*Diet cost*

Understanding the barriers to consumption of a healthy diet is important in order to effectively promote such a diet. While this research shows that a healthy dietary pattern is most expensive, it seems that this is not related to weight status or breast cancer incidence. In fact, those with a higher BMI spend more on their diet, a relationship which is not significant in this research, but has been observed as significant elsewhere (Lopez et al., 2009b).

*Spatial measures*

Using a combination of spatial measures may be the best approach for targeting public health interventions. For example: Greater London contains pockets of individuals consuming the healthiest dietary pattern who need less help improving their diet. Using the OAC classification to identify the specific neighbourhood areas within Greater London which do not need assistance, money can be saved by not rolling out interventions in these areas. Conversely, Greater London also has pockets of individuals at high risk of consuming the poorest, Monotonous Low Quantity Omnivore dietary pattern. Their neighbourhoods can be identified and targeted to improve education and access to healthy foods. This could be by making healthier option available through physical access, or by making them affordable.

An Output Area Classification specific to London (Petersen et al., 2010) has been developed as the diversity of individuals and areas in Greater London is recognised in other areas, but this research shows the dietary diversity in this region too, so it would be interesting to see how dietary patterns vary according to the London Output Area Classification in future research.

*Estimating data*

Collecting data at a neighbourhood level geography is expensive and time consuming and synthetic estimates are often carried out using national data to provide such estimates (Health & Social Care Information Centre, 2013a, Edwards and Clarke, 2009). Methods described in this thesis could provide an alternative means for estimating data, through use of geodemographic identifiers. Others have also suggested incorporation of geodemographic measure into methods to estimate neighbourhood level data (Birkin and Clarke, 2012). Using data from large cohort studies, it may mean that specific populations can be targeted, using cohort data from similar populations.

## 9.5 Future Work

*Longitudinal analysis*

With the exception of Chapter 7, the analysis in this thesis is cross sectional. It would be possible to perform longitudinal analysis on the diet and cost of diets in the UKWCS, using the four day food diaries at follow up (approximately 5 years later). While only a sample of these diaries have been coded, it would be possible to perform a nested case control analysis of the effect of the cost of diet on breast cancer incidence, or death caused by breast cancer.

It would also be of particular interest to investigate the effect of dietary patterns and diet cost on cardiovascular disease (CVD) outcomes. Quality data is available in the UKWCS to allow this research to be carried out, using either baseline FFQ data or follow up four day food diaries. The mechanism by which diet is associated with CVD outcomes is better understood than the relationship between diet and breast cancer incidence, so it would be particularly interesting to investigate whether diet cost and spatial variations in diet are determinants of dietary consumption associated with CVD outcomes.

*International studies*

The international case studies in this thesis are of an exploratory nature. The cohorts used have potential for further analysis of geographic patterns in dietary consumption, weight status and other health outcomes, however, these were out of the scope and time frames of this thesis. One example would include spending more time analysing data from the Australian Longitudinal Survey of Women's Health with respect to dietary patterns consumed. Dietary patterns have already been identified by factor analysis in the cohort (Mishra et al., 2010). It would also be possible to combine the three age cohorts in this data, to increase the sample size for analysis. With more time, investigation of unmatched women to Cameo identifiers would also increase the sample size.

In the Seattle Obesity Study data, it was observed that the Cameo groups correlated strongly with property values. Work completed in Seattle has identified property value as a strong predictor of weight status (Rehm et al., 2012). As property value is available at a tax parcel unit, it could be possible to aggregate property value to a census block geography level and use odds of obesity - or other health outcomes - according to deciles of property value to estimate small areas patterns in Seattle. Developing simple but effective measures to generate neighbourhood level data which accurately reflect health status could be of benefit to many.

Assigning the Cameo classification to the UKWCS would allow for an international comparison between the three countries and also allow for some general comparison between the OAC and Cameo.

*Validation of estimated data*

For all the estimated small area patterns validation of the output is essential. In future work, collecting data from a sample in an estimated city and comparing this to the estimated values would allow agreement between the methods to be assessed and thus help to validate this method. It may be possible to perform some analysis using the Health Survey for England data which is available at lower super output area level (Public Health England, 2013b). This would require some aggregation of estimates, which are at the smaller geographical unit of output area.

## 9.6 Summary and Context

This Chapter has discussed the findings of this thesis in relation to key messages, the strengths and limitations, potential policy implications and future work. Chapter 10 will conclude the research findings.

# Chapter 10: Conclusion

## 10.1 Overview

This chapter briefly concludes the research reported in this thesis.

A reminder of the thesis flow is illustrated in Figure 10.1.

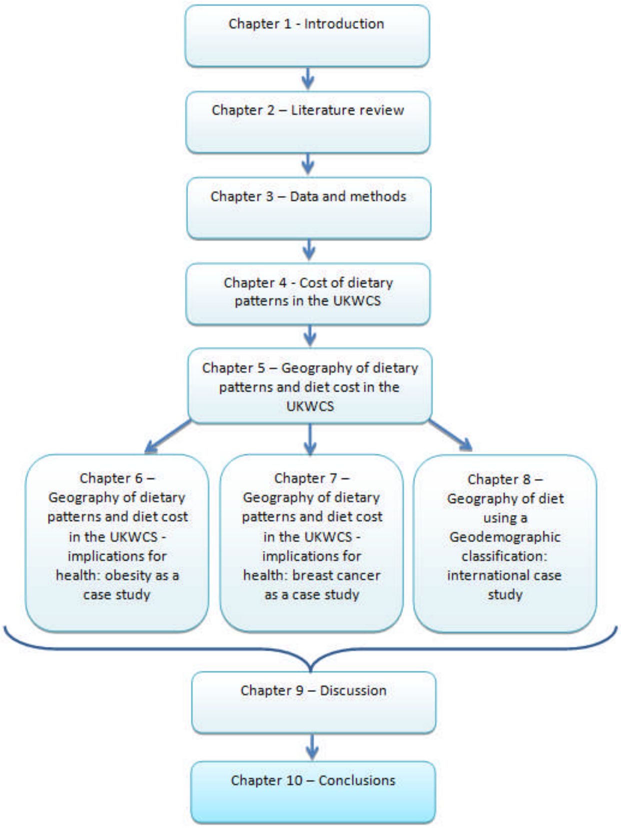



Figure 10.1 - Overview showing how Chapter 10 fits into the overall thesis flow

## 10.2 A healthy diet is more expensive

A healthy diet is estimated to be more expensive to the consumer than a less healthy one, as described in Chapter 4, using data from the UKWCS. This relationship between the estimated cost of food consumed and healthiness is independent of socioeconomic status. However, those who consume a healthier dietary pattern are more likely to be better educated and in a better paid profession.

Using a food cost database, applied to dietary record data is a reliable method of estimating the cost of diet in population research.

## 10.3 Spatial variations in diet cost and dietary patterns exist in UK women

Dietary pattern consumption is associated with where individuals reside. The type of area, using a small scale geographical unit, combined with demographic characteristics, provides a richer understanding of dietary consumption than the large regional unit. Healthier, more expensive dietary patterns are more common in geodemographic groups such as Countryside or Prospering Suburbs. Less healthy dietary patterns are consumed in areas such as Constrained by Circumstance and Blue Collar Communities. However, these are not the areas consuming the cheapest diets. It is Multicultural areas who consume the cheapest diets. With this in mind it may be beneficial to use such classifications to tailor application of dietary advice to encourage healthy eating in order to promote long term health. It may also be useful to estimate dietary patterns for a given city, in this case Leeds, in order to assist in a more focussed implementation of dietary interventions to benefit public health.

## 10.4 Spatial variations in obesity exist in UK women

Areas with the highest prevalence of overweight/obese do not necessarily have the most expensive diet, but within a spatial measure, the overweight/obese pay more for their diet.

It would seem sensible to consider geographic measures and combinations of spatial measures to best target public health interventions to prevent and reduce overweight and obesity.

## 10.5 Breast cancer incidence

There is no significant spatial variation in breast cancer incidence in the UKWCS. However, there appears to be some increased risk, approaching levels of statistical significance, in: (1) South West for whole cohort analysis and post menopause; (2) Multicultural area for whole cohort and pre menopause and (3) Prospering Suburbs pre menopause.

A positive association exists between weight status risk of breast cancer incidence, but not in premenopausal women. There appears to be no association between diet cost and breast cancer incidence.

## 10.6 Spatial scale

While variations in dietary patterns, diet cost and health outcomes exist at varying spatial scales, those derived from small area geographic units, such as the OAC may be most relevant for use by policy makers.

## 10.7 Geodemographic classifications may be a useful tool for estimating small area dietary and health patterns in the UK

Making use of geodemographic classifications to estimate small area diet and health patterns shows potential and could be a cost effective tool for producing synthetic estimates at a neighbourhood level. However, validation work is first required to test the outputs.

## 10.8 Potential uses in other developed populations

There appears to be scope to use geodemographic classifications as a tool to profile diet and health in other developed populations. Careful consideration of the source data for estimating target populations is required due to the diversity within larger land masses.

## 10.9 Summary

Understanding determinants of dietary patterns remains important for public health and making healthy diets accessible to all is important. However, while diets higher estimated cost reflect  healthier dietary patterns, they do not appear to be the mechanism for which obesity prevalence and breast cancer incidence occur. Geodemographic classifications, combined with other spatial measures could aid more effective targeting of public health nutrition policy.

# Conference presentations

- August 2013: International Geographic Union: Regional conference, Kyoto (Japan)

Oral Presentation: Spatial analysis of overweight and obese in the UK Women's Cohort Study

- May 2013: International Society for Behavioural Nutrition and Physical Activity, Ghent (Belgium)

Poster presentation: Spatial analysis of overweight and obese in the UK Women's cohort Study

- May 2013: International Geographic Union: Applied GIS and Spatial Modelling Conference, Leeds (UK)

Oral Presentation: Spatial analysis of overweight and obese in the UK Women's Cohort Study

- December 2012: The Australia New Zealand Regional Science Association International Conference, Wollongong (Australia)

Oral presentation: Exploring regional dietary patterns of UK women: Estimated dietary patterns using a geodemographic classification

- May 2012: International Conference on Diet and Activity Methods, Rome (Italy)

Oral presentation: Is a healthy diet the most expensive type of diet?; Using dietary data from the UK Women's Cohort Study

Poster presentation: Can a price database linked to food intake produce an accurate estimated cost of food purchased?: A validation study

- March 2012: White Rose Doctoral Training Centre Economics PhD Conference, Leeds (UK)

Oral presentation: Is a healthy diet the most expensive type of diet?; Using dietary data from the UK Women's Cohort Study

- September 2011: European Colloquium on Quantitative and Theoretical Geography, Athens (Greece)

Oral presentation: Spatial Analysis of Breast Cancer Incidence, Dietary Patterns and Diet Cost in the UK

# Publication and pending publications

- Comparability of methods assigning monetary costs to diets: derivation from household till receipts versus cost database estimation from four-day food diaries.

Timmins K, Morris M, Hulme C, Edwards K, Clarke G, Cade J. - *European Journal of Clinical Nutrition,* 67**,** 1072-1076.

- Geography of diet in the UK Women's Cohort Study: Benefits of using a geodemographic classification.

Morris M, Clarke G, Edwards K, Hulme C, Cade J - In submission process

- What is the cost of a healthy diet? Using diet data from the UK Women's Cohort Study

Morris M, Clarke G, Edwards K, Hulme C, Cade J- In submission process

- Using a geodemographic classification to estimate neighbourhood level health in Seattle

Morris M, Clarke G, Edwards K, Hulme C, Cade J, Aggarwal A, Drewnowski A - In preparation

- The role of dietary patterns and diet cost in breast cancer incidence in the UK Women's Cohort Study

Morris M, Clarke G, Edwards K, Hulme C, Cade J - In preparation

- Are geodemographic classifications a useful tool for profiling diet and health? An international comparison

Morris M, Clarke G, Edwards K, Hulme C, Cade J, Aggarwal A, Drewnowski A, Jackson C, Mishra G - In preparation

# Appendices

## Appendix A - Procedure for Updating Cancers and Deaths

### Protocol & log file to import death and cancer registry data to the cohort access file

Example document, red text is where this should be filled in

Date data received/ data period: ____________________

Stages to import all data:

Database: Food/Cohort/Cohort97_New (links to Cohorttl.mdb)

Press [Database window] to open table and query list

1) Check MR511 files (English and Welsh):
    a. Save csv files as xls files.
    b. Remove 'Cancelled cancer' entries out of 'cancer and deaths' and create a separate file of cancelled cancers.
    c. Delete deaths out of 'cancer and deaths'.xls file.
    d. Add in headings for both cancer and death files from 'cen data\ MR511 headings.xlsx'. Headings are based on document : MRIS File Formats 01Dec2011.pdf.
    e. In both cancer and death files, ensure event date is in format dd/mm/yyyy (ensures when Access table created the field is created as a date field). Can use formula: =MID(E2,7,2)&"/"&MID(E2,5,2)&"/"&MID(E2,1,4)
    f. check in excel 'MR number' is populated and numeric
2) IMPORT DATA: Open [frmUpdateDeathsCancer]
    a. Click 'Import Excel data from ONS' (check for import errors)
    b. Select the file MR511*date*.xls (from cen data folder ) [import deaths first (more columns)]. Process deaths to step h) below, then repeat from this step, but *append* cancers.
    c. Ensure header row selected [NEXT], select no primary key, [NEXT]...
    d. Import to table: call the file same name as from cen data MR511_*MM_YYYY*
    *e.* [Finish] The English death and cancer info is now in a table you just created called MR511_*MM_YYYY*
    f. Change format of 'Event date' to Date/Time with a format of 'short date'.
    g. Open your new file MR511date to check import numbers

| Type of | Number of rows from original check of excel file | Number of rows in this imported file | Notes if numbers don't match up- explain discrepancy | Your name | Date of action |
|---|---|---|---|---|---|
| Deaths | ? | ? | | ? | |
| Misc | ? | ? | | ? | |
| Cancers | ? | ? | | ? | |

h. Check your new file for duplicate records (ie duplicate MR number) - sort by ID then name to scan the file twice. When two rows refer to the same participant, check if duplicate by looking at data type (D/CA) and at cancer location and morphology (use event details column)

| Type of data | Number duplicates | IDs | Duplicate data deleted from import file? | Number to import after removal of dups/ errors | Your name | Date of action |
|---|---|---|---|---|---|---|
| Deaths | 0 | | | | ? | ? |
| Cancers | 0 | | | | ? | ? |

3) UPDATE DATA: Open [frmUpdateDeathsCancer]

a. Open the 'Cancer Details' table and 'Death_flagging' table. Record the row count in both, 'Cancer Details' : .............. 'Death_flagging' :...........

b. Select the table you just imported and checked, which is free of any duplicates

c. Click 'update cancer information' *(note #of records appending to check they all imported-fill in below table 3)*

d. Click 'update death information' *(note how many records it will append-fill in below table 3)*

e. Click 'update multiple underlying causes' (click ok to the many messages that pop up)

f. The deaths go to 'death_flagging' table. Open this table up, sort by date added. Check: are the new ones there -are all the important fields occupied in the same way as the records previously added? fill in below table 3

g. Cancers go to 'cancer details' table. Open table, check has import worked? Fill in table 3

*Table 3: Check that the numbers imported match expected*

| | Number expected to import from table above | Number the append said would be added | Notes to explain any missing | Look in death_ flagging table for # added in this period | Look in cancer details table for # added in this period | Notes to explain any missing | Your name & date |
|---|---|---|---|---|---|---|---|
| Death | ? | ? | | ? | / | | ? |
| Cancer | ? | ? | | / | ? | | ? |

4) **Scottish** registry **deaths** (paper format)
   a. Print out death certificates (single sided so can file later).
   b. Open up COHORT table to identify participant ID numbers using date of birth and names. Write the number on the certificate
   c. Use 'death_flagging_form' to enter deaths:
      i. Press 'add new record' and type 'Subject ID' into box-check and select 'pending pencil' icon so that the record is updated and the name and DOB appear. Ensure name and DOB match the certificate [if ICD codes appear already check if this certificate is just a duplicate]
      ii. Select yourself from the dropdown list on the left and check ICD version says 10
      iii. Fill in **date of death** and **NHS number** (found at the bottom of the form) if this is absent in our records (new format is a 9 digit all numerical code, old format includes letters and numbers)
      iv. ICD10 code for underlying cause (often in bold) must be entered into 'original underlying cause' blue box
      v. If nothing is listed as a secondary underlying cause (II) then copy all remaining ICD10 cause of death codes into the green box, remembering to enter numbers 1, 2, 3, 4, etc into the first column of this green table for each code you enter
      vi. If there are conditions under primary and secondary underlying causes (I(a, b, c, d) and II) then use the ICD10 code book (or http://apps.who.int/classifications/icd10/browse/2010/en#/C80 - if 4 digits may need to insert a full stop before last digit) to identify those that are secondary and these must be entered into the red box, put the rest into the green. Again, remember to manually add numbers in the first column for each ICD code you enter.
      vii. Note that the primary original code (often in bold) may also appear as a primary underlying cause-it is ok to enter this twice if that is how it is presented on the form.
   d. Write your name and the date on the paper form. **File securely** with rest of death records in Victoria's office-stored alphabetically by surname.

| Death certificates entered? | Your name | Date | Notes- were any duplicates found? |
|---|---|---|---|
| ? | ? | ? | |

5) *Fixing secondary underlying cause ICD codes for deaths & multiples*
   *[*During the automated import, the secondary codes cannot be separated from multiple underlying cause codes (the ICD codes come mixed and the description for the secondary code is free text). You must view the text in the secondary field and column to the right from the excel doc and identify the correct ICD codes to match to this]:
   a. Open the excel file of English deaths. Copy the data onto a new tab and cut out the cancer data and all records where there is no text in the columns

labelled Cause of Death text II and the one to the right of this. So you just have the data for the IDs where there is some text in these columns.

b. Open 'death_flagging_form': On this form the ICD10 codes from the multiples list are matched to their proper description at the bottom of the form. Note that the codes with three characters don't appear in the grey box at the bottom-use the ICD10 codebook to look these up.

c. Using the free text descriptions from the excel doc identify those multiple codes to move into the red box-as they actually relate to secondary underlying causes and not multiples. Add them to the red box, remembering to number the lines 1, 2, 3 etc and then cut the lines from the green box.

| | Your name | Date |
|---|---|---|
| Secondary underlying codes sorted out? | ? | ? |

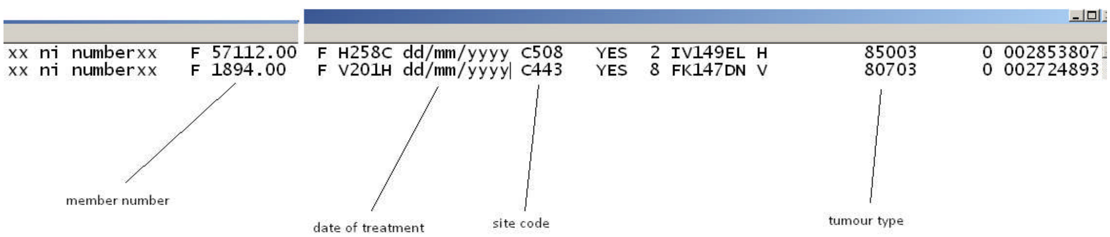
6) **Scottish** registry **cancers** (paper format)/electronic

<u>New format text files</u>

a. Open the text files in excel.

b. Merge into one text document (one line may have more than one entry - these need separating onto separate line)

c. Replace tabs with a single space, then make align by inserting further spaces




d. Paste the information into Excel (use fixed length delimiting) - Excel allows you to pick where the field delimiters are.

e. In Excel format the information so that it matches the layout on the 'Cancer details' table.

f. Check that the UKWCS id provided in the files matches to the correct woman using the COHORT table .

g. Open the 'Cancer Details' table in the Cohort97New database

h. Check that the new cancer does not already exist (ID, Cancer site, cancer data and cancer type all match)

i. If the Cancer Type is blank, use the literal 'unkown'

j. On a new line, populate the ID, Cancer site, cancer date and cancer type field with the new cancer details (by pasting from Excel)

k. Add a note to include the month of the Scottish Cancer update provided and that this was manually added and by whom. E.g. Dec 2009 Scottish cancer manually added MM

Old format text files

a. Print the cancer records
b. Some of the women in these files will need to be matched to obtain their ID number, using the COHORT table from COHORTtbl database.
c. Check that the new cancer does not already exist (ID, Cancer site, cancer date and cancer type all match)
d. On a new line, populate the ID, Cancer site, cancer data and cancer type field with the new cancer details.
e. Add a note to include the month of the Scottish Cancer update provided and that this was manually added and by whom. E.g. Dec 2009_Scottish cancer_Manually added MM

7) **Cancelled cancers**:
   a. Open the new cancelled cancer file.
   b. Copy the cancelled cancers into the Cancelled cancers tab in the log file : <LOG FILE Death and cancer import to access.xlsx>
   c. Open the 'Cancer details' table in Access.
   d. Locate the cancer entry to be cancelled (match on Member No., Cancer date, Cancer site, Cancer type)
   e. Populate the 'cancelled date' field with the 1st day of the month that the update relates to.
   f. Update the 'notes' field to record when the cancer was cancelled and who performed the cancellation.
   g. Complete the LOG FILE using the colours in the key, as per earlier updates.

8) **Other documents to process ?**
   a. Other documents may be present in the folder eg a pdf document containing surname change. These need reviewing and actioning. If there has been a name change, this can be recorded in the notes field in the Cohort table.

9) EXPORT DATA [Export] on frmUpdateDeathsCancer
   a. Output death and cancers in excel format to send to the database manager

| | Export done? | Sent to database manager? | Your name and date |
|---|---|---|---|
| Cancers | ? | ? | ? |
| Deaths | ? | ? | ? |

**10) To finish:**

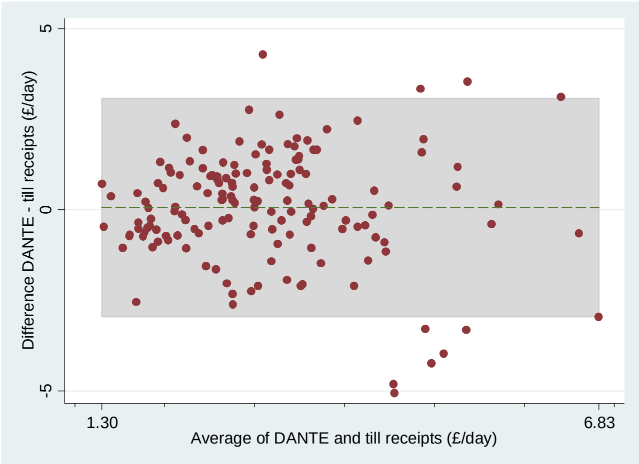
Once have added deaths, cancers, cancelled cancers and Scottish data from the data folder- update the **LOG FILE** < LOG FILE Death and cancer import to access.xlsx> in the cohort/cen data file

11) CHECKS

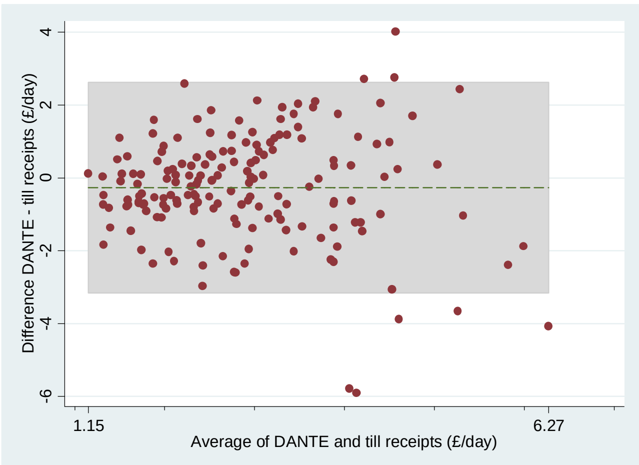
Table 1: Numbers log file (initial count and final check of the import)

| **BEFORE** | | | | | **AFTER** | | |
|---|---|---|---|---|---|---|---|
| **List all files located within this cen data time period (xxxxxxxxxxx)** | **Data type** | **Row count/ record# using summary page info and counting rows** | **Notes** | **Name of person doing data count initially & date** | **Number of rows from this file now in access** | **Notes for any for any IDs not in access** | **Name of person doing data count to check import** |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

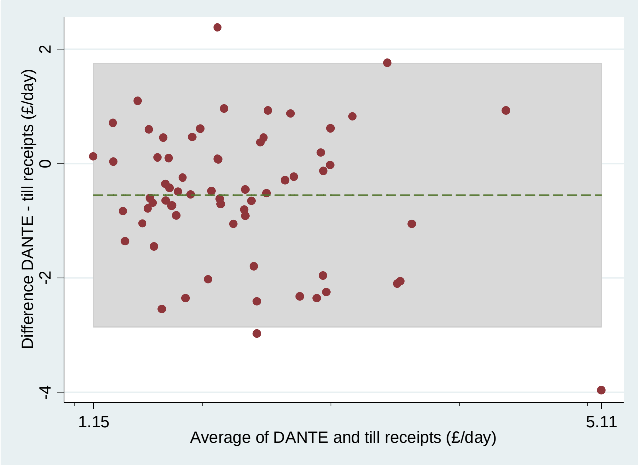
# Appendix B - Supplementary Bland Altman Plots for Chapter 3, Table 3.12



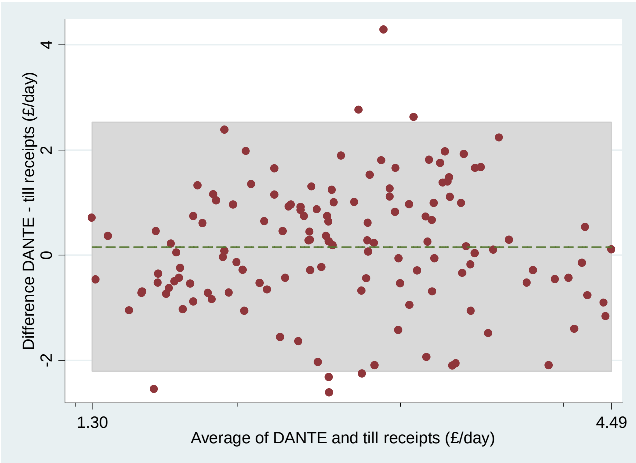

Appendix B, Figure 1 - Differences between daily costs - in the male subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents 95% limits of agreement)



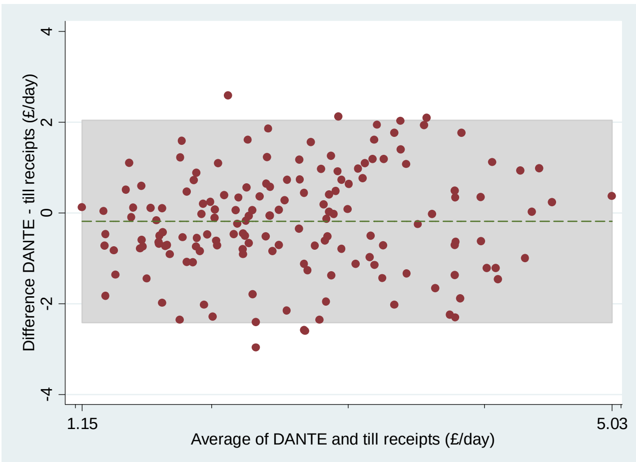

Appendix B, Figure 2 - Differences between daily costs - in the female subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents 95% limits of agreement)



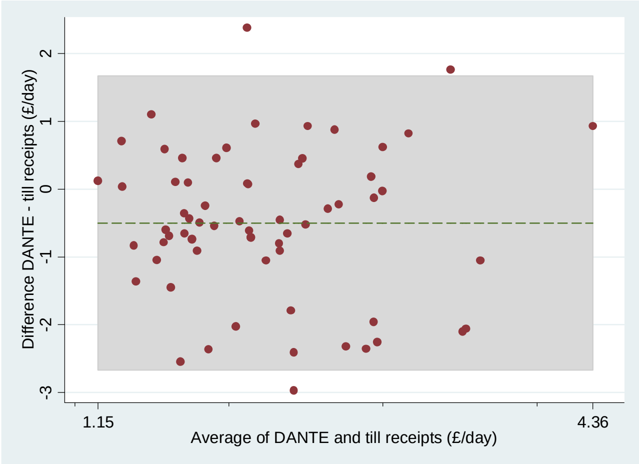

Appendix B, Figure 3 - Differences between daily costs - in the child subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents 95% limits of agreement)



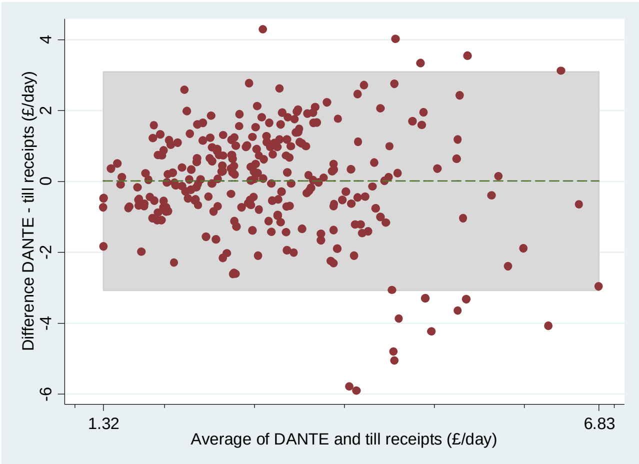

Appendix B, Figure 4 - Differences between daily costs - in the male subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents 95% limits of agreement)



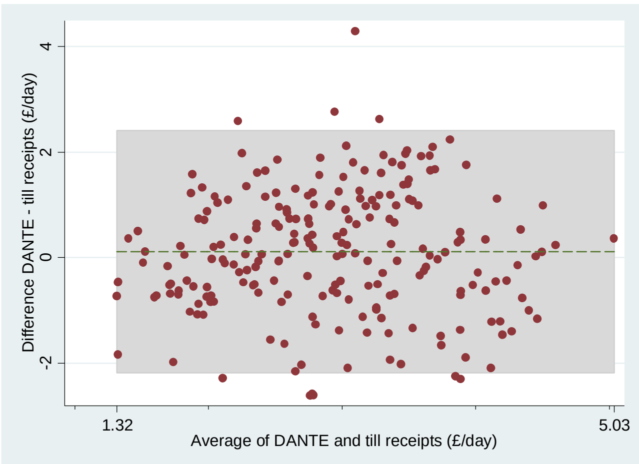

Appendix B, Figure 5 - Differences between daily costs - in the female subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents 95% limits of agreement)



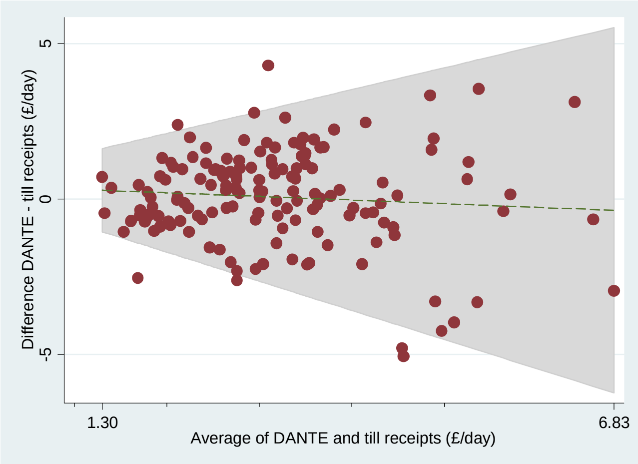

Appendix B, Figure 6 - Differences between daily costs - in the child subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents 95% limits of agreement)



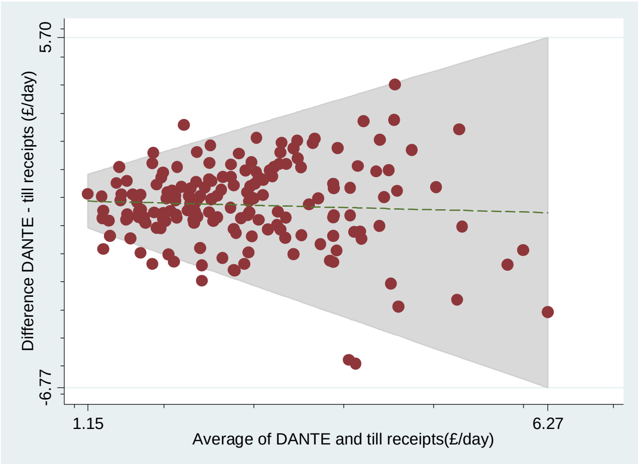

Appendix B, Figure 7 - Differences between daily costs - in the adults subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents 95% limits of agreement)



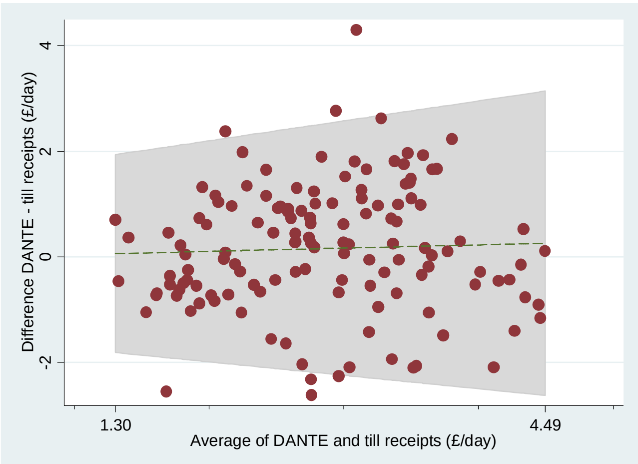

Appendix B, Figure 8 - Differences between daily costs - in the adults subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents 95% limits of agreement)



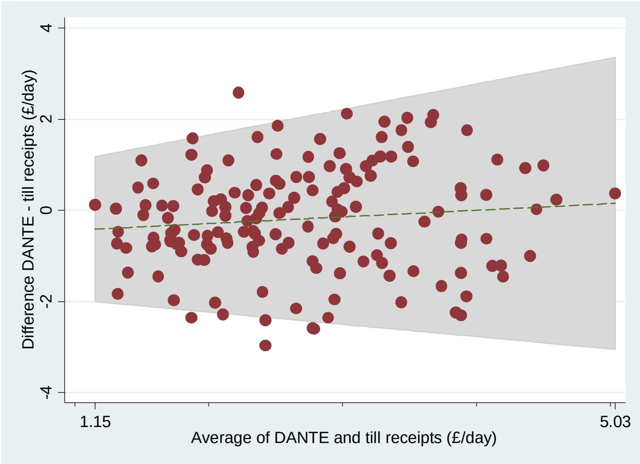

Appendix B, Figure 9 - Differences between daily costs - in male subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents regression line for 95% limits of agreement)



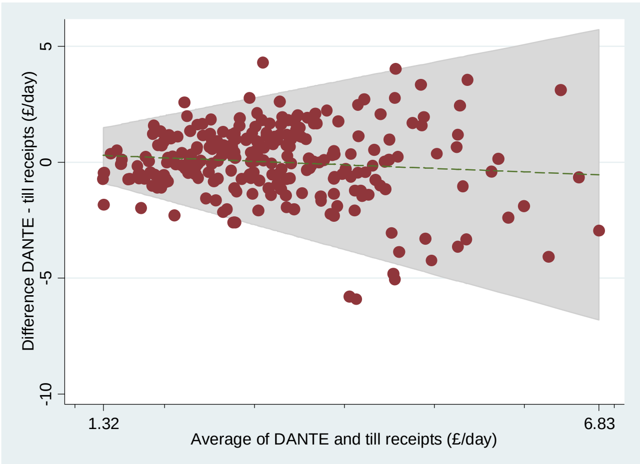

Appendix B, Figure 10 - Differences between daily costs - in female subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents regression line for 95% limits of agreement)



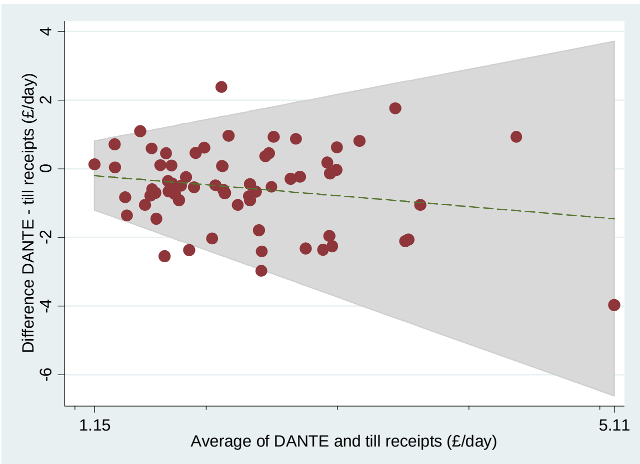

Appendix B, Figure 11 - Differences between daily costs - in male subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents regression line 95% limits of agreement)



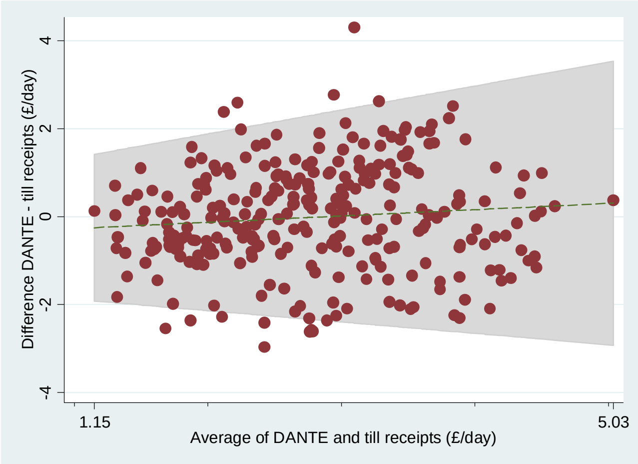

Appendix B, Figure 12 - Differences between daily costs- in female subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents regression line 95% limits of agreement)



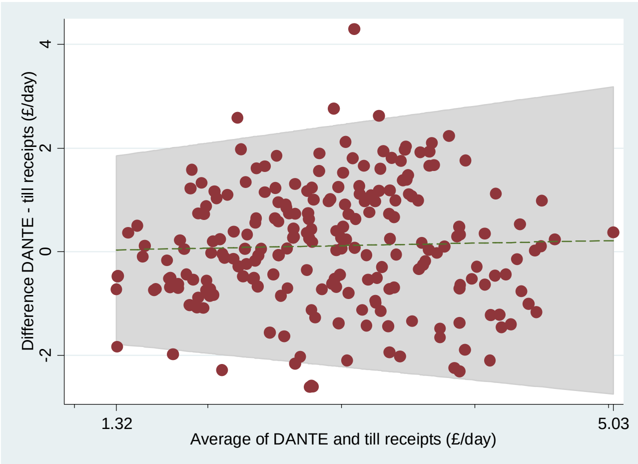

Appendix B, Figure 13 - Differences between daily costs - in adult subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents regression line 95% limits of agreement)



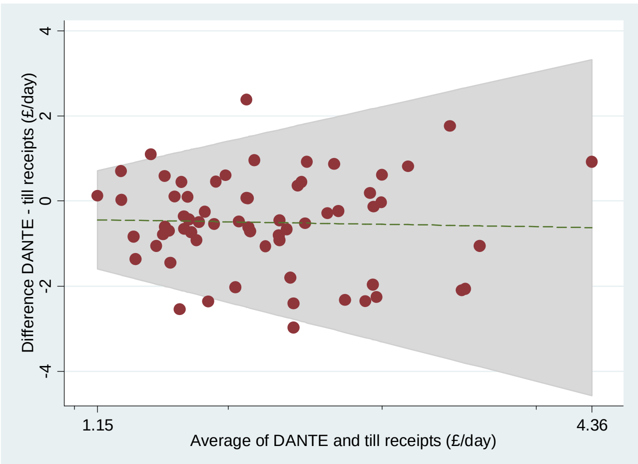

Appendix B, Figure 14 - Differences between daily costs - in child subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) (shaded area represents regression line 95% limits of agreement)




Appendix B, Figure 15 - Differences between daily costs - in the full sample - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents 95% limits of agreement




Appendix B, Figure 16 - Differences between daily costs - in adult subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents regression line 95% limits of agreement)




Appendix B, Figure 17 - Differences between daily costs - in child subgroup - estimated by the DANTE cost database and daily expenditure calculated from till receipts (adjusted for waste) (£/day) when the highest 5% of spenders are excluded (shaded area represents regression line 95% limits of agreement)

# Appendix C - ALSWH Expression of Interest Form

Document F
May 2011

women's health australia

| Office use only |
|---|
| EoI no: _______ |
| Date received: _______ |

**Expression of Interest – ALSWH analysis**

**Section A**

| | |
|---|---|
| Name, title, email address and institution of Project Leader (lead person) | Michelle Morris, PhD Student.<br>Email: bms9mam@leeds.ac.uk<br>University of Leeds, UK. |
| Corresponding ALSWH liaison & contact details (If ALSWH liaison is also a collaborator please list below) | Professor Gita Mishra<br>Email: g.mishra@uq.edu.au<br>University of Queensland. |
| Other Collaborators (include title, email address and institution) | Dr Caroline Jackson<br>Email: caroline.jackson@uq.edu.au<br>University of Queensland.<br><br>Professor Annette Dobson<br>Email: a.dobson@sph.uq.edu.au<br>University of Queensland. |
| Brief title of potential project | Exploring the dietary patterns of Australian women: Estimated dietary patterns for Sydney using a geodemographic classification |
| Is a student involved in the project (will the project contribute to their award)? | Yes ☒ No ☐ |
| If yes what percentage will this project contribute to their final award? (For example if their project is based entirely on ALSWH data this would be 100%) | 5% |
| If yes, please provide student's name, course, institution, course start date and expected completion date | Michelle Morris, PhD, University of Leeds, UK.<br>Start date October 2010. Expected completion date October 2013. |
| Please list supervisors | Principal Supervisor: Professor Janet Cade, University of Leeds, UK. |
| | Associate Supervisor/s:<br>Professor Graham Clarke, University of Leeds, UK.<br>Professor Claire Hulme, University of Leeds, UK.<br>Dr Kimberley Edwards, University of Nottingham, UK |
| Does this new EOI supersede a current or previous EOI? | Yes ☐ No ☒ |
| If yes, please provide details of which EOI it supersedes and confirm whether the EOI it supersedes can now be made inactive | |

| |
|---|
| Please provide a lay synopsis of your proposed project (75-100 words) that can be published on the ALSWH website if the project is approved. (This synopsis may also be included in the ALSWH annual newsletter to participants.) |
| Using dietary patterns derived from the ALSWH food frequency questionnaires and the post codes of these women, a dietary profile for specific geodemographic groups can be created. Based on the assumption that women living in the same geodemographic group consume similar dietary patterns it is possible to use these dietary profiles to estimate dietary patterns for a specific city, in this case Sydney.<br>Geodemographic classifications are widely used in marketing, but currently underused in health research. Using census and survey demographic information for small area geographies a geodemographic classification groups like types of people. |

Document F
May 2011

| | |
|---|---|
| Brief outline of what you are interested in. Include:<br>- Hypothesis/research questions<br>Research Question: Can geodemographic dietary profiles be used to predict dietary patterns in an Australian city?<br>Hypothesis: Dietary patterns will likely vary by geodemographic group. Therefore using dietary patterns from a large cohort it will be possible to create a geodemographic dietary profile which can be used to estimate dietary patterns for a given city.<br><br>- Main variables – outcome response/dependent variables and explanatory/ predictor/ independent variables<br>• Dietary pattern (from FFQ at surveys 3 and 5)<br>• Geodemographic identifiers (assigned via postcodes at survey 3 and 5)<br><br>- Main method of analysis<br>• Each woman in ALSWH will be assigned a geodemographic identifier (via her residential post code) from an Australian geodemographic classification for example GeoSmart or GMAP.<br>• Dietary patterns in the cohort will be identified using a cluster analysis. If dietary patterns already identified in this cohort are available, then it may be possible to use these patterns in this analysis. A dietary profile for each geodemographic group will be created. This profile will not contain any identifiable information about the ALSWH participants.<br>• Descriptive statistics and multinomial regression analysis will be performed using Stata statistical software. The variables which I will require for this analysis include; age, height, weight, waist circumference, smoking status, occupation, highest qualification, physical activity, income and dietary information. These relate to the following survey questions:<br>-Survey 3 (1973-78) - Q40, Q41, Q46, Q52, Q59, Q61-Q78 inclusive, Q103, Q104, Q105, Q106, Q107a, Q108a and Q108b.<br>-Survey 5 (1973-78) – Q14, Q15a, Q52, Q53, Q54, Q58-68 inclusive, Q70, Q72-Q90 inclusive, Q113, Q114, Q115, Q116a, Q116b.<br>-Survey 3 (1946-51) – Q32, Q41, Q46a, Q46b, Q48a, Q49-Q69 inclusive, Q80a, Q80b, Q87a, Q87b.<br>-Survey 1 (1946-51) – Q90.<br>• Estimated dietary patterns for Sydney will be mapped using a geographic information system (ArcGIS).<br>• This method has already been used in a study of dietary pattern in UK women. | |
| Expected outcomes & likely target audience (eg. paper to be submitted to a psychology journal; paper to be submitted to National Rural health Conference) | Paper to be submitted to peer reviewed journal in one of the following areas: health geography, public health or nutrition.<br>Results will be included in Michelle's PhD thesis.<br>Some early findings will be included in a presentation at the Australia and New Zealand Regional Science Association International conference in Wollongong Dec 2012. |
| Time frame of the project (start/finish dates; expected paper submission dates) | Project start date: 26th November 2012.<br>Expected completion and paper submission June 2013. |
| How do you expect to fund this work? (Describe source and amount) | Flights to Australia in November 2012 are funded by the Centre for Spatial Analysis and Policy, University of Leeds. During this visit Michelle plans to perform the analysis at the University of Queensland.<br>Michelle is funded via an ESRC/MRC PhD studentship. This work will be completed to test a method developed in Michelle's PhD. |
| Names and email addresses of all people who will have access to the raw data | Michelle Morris<br>Nutritional Epidemiology Group<br>School of Food Science and Nutrition<br>University of Leeds<br>Leeds |

Document F
May 2011

| | |
|---|---|
| | LS2 9JT<br>UK |
| Who will provide the substantive expertise and input? | Michelle Morris |
| Who will provide statistical expertise and input? | Michelle Morris, Gita Mishra |
| Who will provide qualitative analysis expertise if required? | N/A |
| Keywords | Dietary patterns<br>Geodemographic<br>Estimated<br>Women |
| Applicable themes – ***tick all that apply***<br>Chronic conditions ☐<br>Health service use & systems including Medicare Australia analyses ☐<br>Social factors in health & wellbeing ☐<br>Health in rural and remote areas ☐<br>Roles & relationships ☐<br>Intergenerational issues ☐<br>Methodology ☒<br>Independent living / Aged Care ☐<br>Abuse ☐ | Tobacco, alcohol & other drugs ☐<br>Weight nutrition & physical activity ☒<br>Mental health ☐<br>Ageing ☐<br>Medications ☐<br>Formal & informal work patterns & work-family balance ☐<br>Caring ☐<br>Reproductive health ☐ |
| ***Please complete section B of this document on the following page*** | |

Document F
May 2011

**Section B - ALSWH Dataset Instructions**

| | |
|---|---|
| **Collaborator name & address for posting:** | Michelle Morris |
| | Nutritional Epidemiology Group |
| | University of Leeds |
| | LS2 9JT. UK |
| **ALSWH Liaison person for the project:** | Gita Mishra |

The data sets on this form are specific to data collected by ALSWH. Access to other data sets such as MBS or PBS should be discussed with your ALSWH Liaison person. Please list other data sets required in the comments box below. A separate form is required for use with MBS / PBS data.

***Please indicate preferences for the following by checking the applicable boxes:***

| | SAS user | SPSS user | STATA user | Other (text) |
|---|---|---|---|---|
| **Type of data files** | ☐ | ☐ | ☒ | ☐ |

When datasets are requested the accompanying formats and labels will also be sent if they exist.

| | 1973-78 | | | 1946-51 | | 1921-26 | |
|---|---|---|---|---|---|---|---|
| | Survey data | Qualitative data | Child data | Survey data | Qualitative data | Survey data | Qualitative data |
| **Survey 1** | ☐ | ☐ | N/A | ☒ | ☐ | ☐ | ☐ |
| **Survey 2** | ☐ | ☐ | N/A | ☐ | ☐ | ☐ | ☐ |
| **Survey 3** | ☒ | ☐ | N/A | ☒ | ☐ | ☐ | ☐ |
| **Survey 4** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **Survey 5** | ☒ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **Survey 6** | N/A | N/A | N/A | ☐ | ☐ | N/A | N/A |

| ***OTHER DATA SETS*** | 1973-78 | 1946-51 | 1921-26 |
|---|---|---|---|
| **Participant Status** | ☐ | ☐ | ☐ |
| **Cause of Death** | ☐ | ☐ | ☐ |
| **Medications** | N/A | 5th Survey ☐ | 4th Survey ☐ |

| **Food Frequency Questionnaire** | 1973-78 | 1946-51 |
|---|---|---|
| **Survey 3** | ☒ | ☒ |
| **Survey 5** | ☒ | N/A |

**Comments:** Food frequency questionnaire data required, along with corresponding demographic and lifestyle questionnaire information.

***Office use only***

| | | | |
|---|---|---|---|
| **EOI No:** | | **Email address** | |
| **Date EOI approved:** | __/__/___ | | |
| **Dataset created by:** | | | |
| **Date dataset created:** | __/__/___ | | |

# Appendix D - ALSWH Memorandum of understanding

**Document D**
**September 2009**

women's health australia

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

**Memorandum of Understanding**

*between*

GITA MISHRA
**The University of Queensland**

*and*

MICHELLE MORRIS
*of*
**University of Leeds**

*and*

CAROLINE JACKSON and ANNETTE DOBSON
*of*
**The University of Queensland**

*Stating terms and conditions for collaboration on the Australian Longitudinal Study on Women's Health (ALSWH)*

1 The research covered by this memorandum includes research into:

*Exploring the dietary patterns of Australian women: Estimated dietary patterns for Sydney using a geodemographic classification* as detailed in Proposal ID A440.

2 MICHELLE MORRIS, CAROLINE JACKSON and ANNETTE DOBSON will be accorded Collaborator status on the Australian Longitudinal Study on Women's Health (ALSWH) project, limited to the research described in Clause 1, in this instance.

Researchers are encouraged to collaborate on the ALSWH project when they have particular expertise in a subject area.

In this instance, GITA MISHRA will liaise regularly on behalf of the ALSWH researchers.

3 All Collaborators will abide by the "PSA Policy and Procedures for Data Access, Analysis and Publication" (Document B), "PSA Policy and Procedures for Substudies" (Document C) and 'Privacy Protocol' (Document E), that govern all ALSWH projects.

**Document D**
**September 2009**

4 All publications arising from the studies specified in clause 1 will include the standard acknowledgment as in the "PSA Policy and Procedures for Data Access, Analysis and Publication" (Document B). Where Medicare or Pharmaceutical Benefits Scheme (PBS) data are used, 'Medicare Australia' must be acknowledged with the statement 'We acknowledge Medicare Australia for providing the PBS and MBS data'.

5. All parties are to notify each other before presenting any data at conferences, seminars and fora, and must always consult with their ALSWH Liaison prior to submitting an abstract. The ALSWH Liaison person reserves the right to refer presentations to the PSA Committee.

6. The results of the research specified in clause 1 will not be used to seek funding for any further research without a new proposal to the PSA.

7. Copies of all data files and statistical codes will be provided to the ALSWH Data Manager on completion of the studies.

8. All quantitative and qualitative data collected as part of a substudy will be provided to ALSWH for secure electronic archiving. These data may be used in future projects, and the project leader who collected the data will always be invited to participate in these future analyses. Hard copies of data will be the responsibility of the project leader and must be stored and disposed of in accordance with HREC and NHMRC guidelines http://www.nhmrc.gov.au/publications/synopses/r39syn.htm

I have read the above terms and conditions and I agree to them

..................................................

Gita Mishra
The University of Queensland

Date ..............................

..................................................

Michelle Morris
University of Leeds

Date......................................

..................................................

Caroline Jackson
The University of Queensland

Date ..............................

..................................................

Annette Dobson
The University of Queensland

Date......................................

**Document D**
**September 2009**

As project leader I accept the following responsibilities:

- Completion of 6 monthly progress reports as requested; AND
- That if I do not provide progress reports when requested that the permission to conduct the project may be withdrawn; AND
- Ensuring the investigators/collaborators or students involved have adequate facilities and resources to enable the project to progress in a reasonable manner to its conclusion; OR
- That in the event of unforseen circumstances that if a project can not proceed that the PSA is notified so that the project can be terminated.

..................................................

Michelle Morris
University of Leeds

Date......................................

# Appendix E - CallCreditGroup contract

Callcredit
Information Group
Marketing Solutions

**CAMEO INTERNATIONAL DATA LICENCE & SERVICES AGREEMENT**
**(Client as End User)**

This Agreement is made between:

(1) Callcredit Marketing Limited (registered in England and Wales with company number 2733070) whose registered office is One Park Lane, Leeds, West Yorkshire, LS3 1EP (“**Callcredit**”); and

(2) [Professor Janet Cade] (registered in England and Wales with company number [RC000658]) whose registered office is [Nutritional Epidemiology Group, School of Food Science & Nutrition, University of Leeds]
(the “**Client**”).

The Client wishes Callcredit to provide certain data services to the Client and Callcredit wishes to provide such services. By signing this front signature sheet the parties agree to be bound by the terms of this Agreement.

This Agreement comprises this front signature sheet and the Schedules attached to it.

| Signed for and on behalf of Callcredit Marketing Limited | Signed for and on behalf of [Professor Janet Cade] |
|---|---|
| ............................................ Signature | ............................................ Signature |
| ............................................ Name of authorised signatory | ............................................ Name of authorised signatory |
| ............................................ Position | ............................................ Position |
| ............................................ Date | ............................................ Date |

SCHEDULE 1
AGREEMENT INFORMATION

**AGREEMENT TYPE/REFERENCE** [N/A]
**PROJECT CODE** [N/A]
**JOB REFERENCE** [EDA20742]

**CALLCREDIT INFORMATION**

| | |
|---|---|
| Full Name: | Callcredit Marketing Limited |
| Company Number: | 2733070 |
| Registered Office Address: | One Park Lane, Leeds, West Yorkshire, LS3 1EP |
| Telephone Number: | 0113 242 4747 |
| Fax Number: | 0113 242 4646 (for notices) |

**AUTHORISED DATA PROCESSOR INFORMATION**
[N/A]

**SERVICES & DELIVERABLES**
Callcredit agrees to supply the Client with the following services (the "**Services**") and deliverables (the "**Deliverables**") in accordance with the terms and conditions of this Agreement:

a. [access to Callcredit's online append service at www.cameo-online.com (the "**Website**") (the "**Online Append Services**");]

[
• ]

The CAMEO Directories [via CAMEO Online] to be [provided/in respect of which Data Client Services are to be provided] are each as set out below:

- [CAMEO UK, CAMEO USA and CAMEO Australia (limited by volume)]

The Deliverables shall be provided to the Client in the following format and manner:

[N/A]
Additional services and deliverables can be provided from time to time by agreement in writing between the Client and Callcredit.

**SCOPE OF LICENCE:**
The Client is permitted to use the Deliverables for its own internal business purposes only, including without limitation for the following purposes:

- appending the CAMEO Directories and the CAMEO Analyses Directories to the details of the Client's existing internal customer portfolio; and
- appending the CAMEO Directories and the CAMEO Analyses Directories to the details of the Client's prospects acquired and to be acquired from third parties.

The Client's use of the Deliverables shall be limited as follows:

[(i) Where the Client employs the Website to access the Deliverables the scope of licence granted under this Agreement shall be limited to the number of transactions in any Year of this Agreement as is set out in section 7, paragraph a below (if any).

]

**DESIGNATED SITE(S):**
The Deliverables (and any copies made thereof as permitted by this Agreement) may only be held at:

(a) the following site of the Client:

[Nutritional Epidemiology Group, School of Food Science & Nutrition, University of Leeds]
(b) the following site of the Authorised Data Processor (as defined below):

[Nutritional Epidemiology Group, School of Food Science & Nutrition, University of Leeds]

The Client may otherwise access the Deliverables itself or via the Authorised Data Processor via the Website from (i) the above addresses and (ii) from any other corporate site of the Client and (iii) the address of the Authorised Data Processor.

Callcredit
Information Group
Marketing Solutions

**COMMENCEMENT AND TERM**

This Agreement shall be deemed to have commenced on the second signatory date of this Agreement ("**Commencement Date**"). If any Services or Deliverables have been delivered to the Client prior to the Commencement Date the Parties hereby agree that such Services and Deliverables will be deemed to have been provided pursuant to and in accordance with the terms of this Agreement.

This Agreement shall (subject to earlier termination under the provisions of clause 10 of Schedule 2) continue from the Commencement Date as follows:

[This Agreement shall terminate automatically at 23.59 hours on the day immediately preceding the first anniversary date of the Commencement Date of this Agreement.]

**PRICE (excluding any value added, sales or other applicable tax, custom or excise)**

The Client agrees to pay to Callcredit the following annual fees which cover [the Online Append Service] each as described in section 1 above (subject as hereinafter provided):

| | |
|---|---|
| | **Licence Fee per Year** (in [Pounds Sterling]) |
| Total Yearly Licence Fee or "**Price**": | [£0] |

The Price includes supply of the following services to Client by Callcredit:

[(a) Online Append Services

The Price entitles the Client to Online Append Services not exceeding for up to 250,000 transactions in total in any Year of this Agreement in any combination save that such Online Append Services must be used in respect of the CAMEO Directories listed in paragraph (a) above only. Over and above such limit, Callcredit shall be entitled to charge the Client and the Client shall be required to pay an amount as agreed between the Parties at such time.]

If the Client requires Callcredit to undertake any additional service(s), including any training, installation, data cleansing service(s) and any data verification services, Callcredit and the Client will discuss on any additional fees payable in respect of such additional service(s).

**SPECIAL CONDITIONS:**

The General Terms shall be amended, supplemented or interpreted as follows:

(i) [The CAMEO Codes and/or any information derived thereof may be used solely for the purposes of this academic study.

(ii) The CAMEO Codes and/or information derived thereof must not be used for commercial purposes.

(iii) CAMEO Codes and/or information derived thereof may not be used in comparative research against competitor systems unless prior written permission has been granted by Callcredit.]

**9. Entire Agreement**

This agreement shall comprise the following documents:

- the front sheet to this document
- this Schedule 1; and
- the general terms and conditions attached at Schedule 2 (the "**General Terms**");

collectively referred to as this "**Agreement**".

This Agreement shall prevail over any terms or conditions contained in, or referred to in, any correspondence, own purchase order, confirmation of order, or Services specification, or implied by law, trade, custom, practice or course of dealing as between the Client and Callcredit in respect of the matters contained herein and such terms or conditions shall not vary the terms of this Agreement.

Without prejudice to the foregoing, but only to the extent relevant, the terms and conditions relating to the use of the online append service at www.cameo-online.com (the "**Website**") which the Client in registering for or in using the Online Append Services purports or is purported to accept shall not apply to the Client to the extent the Client accesses the Website to obtain the benefit of the Online Append Services. The Client acknowledges and agrees that to the extent it accesses the Website and takes the benefit of services other than the services expressly set out in section 1 above (for example, any data cleansing services), such additional services will be subject to the separate applicable terms and conditions set out on the Website or as otherwise settled with or accepted by the Client.

In the event of any conflict between the different elements of this Agreement, the following order of priority shall apply:

1. the General Terms
2. Schedule 1.

Callcredit
Information Group
Marketing Solutions

## SCHEDULE 2
## GENERAL TERMS

1. **DEFINITIONS**

1.1 In this Agreement the terms here specifically defined have the meanings assigned:

**'Affiliate'** means, in relation to a Party, any undertaking which is such Party's parent undertaking or its subsidiary undertaking or a subsidiary undertaking of any such parent undertaking from time to time ("**undertaking**", "**parent undertaking**" and "**subsidiary undertaking**" having the meanings given to them by sections 1161 and 1162 of the Companies Act 2006) it being acknowledged that such undertaking shall be deemed as an Affiliate of a Party only so long as such definitions continue to satisfied.

**'Applicable Obligation'** means any statute, statutory instrument, bylaw, law, order of a court of competent jurisdiction and any requirement of any regulatory, fiscal or governmental body to which the relevant Party is subject and submits, in all cases which applies to the relevant Party in undertaking any relevant activity pursuant to or in connection with this Agreement, in all cases to the extent in force from time to time.

**'Authorised Data Processor'** means the third party bureau specified in Schedule 1 (if any) which may hold and use the Deliverables on behalf of the Client subject to the terms of this Agreement.

**'CAMEO Analyses Directories'** means classification typology systems maintained and developed by Callcredit the main deliverable of which is statistical measures for a limited number of variables, including any updates to such systems.

**'CAMEO Codes'** means the geodemographic classification codes owned and developed by Callcredit, constituting a database.

**'CAMEO Data'** means the output area reference code and other raw data, constituting a database, as identified in Schedule 1.

**'CAMEO Directories'** means the combination of the CAMEO Data and the associated CAMEO Codes, constituting a database.

**'Client Materials'** means any analysis works, reports or other materials produced by the Client which are derived from the Deliverables but which do not replicate or incorporate (absolutely or to any material extent) any element of the CAMEO Directories or CAMEO Analyses Directories.

**'Commencement Date'** means (unless otherwise stated in Schedule 1) the earlier of (i) the date that the second signatory signs this Agreement, (ii) the date that Callcredit begins performance of the Services and (iii) the date the Deliverables are first made available to the Client.

**'Confidential Information'** means for any Party, all trade secret and confidential or proprietary information of such Party or such Party's Affiliates including information concerning its products, services, customers, suppliers, business accounts, financial or contractual arrangements or other dealings, computer systems, data, software, source and object code, business methods and development plans, contained in any format and whether or not communicated orally and whether or not marked "confidential", including such information obtained before the Commencement Date. Without limiting the above, in the case of the Client's obligations the term Confidential Information shall be deemed to include the Deliverables, and, in the case of Callcredit's obligations the term Confidential Information shall be deemed to include the Client Materials.

**'Deliverables'** means the CAMEO Directories and any CAMEO Analyses Directories listed in Schedule 1 and any other deliverables as may be listed in Schedule 1 or which are delivered pursuant to or in connection with this Agreement, including other variables, scores, results, propensity scores, reports, documents, analyses, flags and other information (and all updates thereof) in whatever form to be made available to the Client by or on behalf of Callcredit;

**'Designated Site'** means the premises identified in Schedule 1.

**'Extended Term'** shall bear the meaning given to such term in Schedule 1

**'General Terms'** means these terms and conditions.

**'Input'** shall bear the meaning given to such term in Schedule 1.

**'Minimum Term'** means any period of time identified as such in Schedule 1.

**'Online Append Services'** shall bear the meaning given to such term in Schedule 1.

**'Output'** shall bear the meaning given to such term in Schedule 1.

**'Party'** means either party to this Agreement.

**'Price'** means Callcredit's charges described in Schedule 1.

**'Schedule'** means the schedule attached to these General Terms and signed by the Parties containing details of the Services and Deliverables.

**'Scope of Licence'** means the scope of licence as expressly set out in Schedule 1.

**'Services'** means the services described in Schedule 1.

**'Special Conditions'** means special conditions identified as such in Schedule 1.

**'Website'** shall bear the meaning given to such term in Schedule 1.

**'Year'** means each period of twelve consecutive months commencing on the Commencement Date or any anniversary of that date.

1.2 The Schedule contents (including Special Conditions) form part of this Agreement.

1.3 To the extent of any inconsistency between Schedule 1 contents and these General Terms, these General Terms shall take precedence.

2. **OBLIGATIONS OF THE PARTIES**

2.1 In consideration of the Client paying to Callcredit the Price, Callcredit agrees to perform the Services and provide the Deliverables using reasonable care and skill.

2.2 Callcredit does not warrant that the Deliverables will be entirely error free.

2.3 Callcredit will use up to date anti-virus systems in accordance with good industry practice as appropriate in providing the Deliverables. The Client is responsible for using up to date anti-virus systems in accordance with good industry practice on its own systems as appropriate in respect of receipt of the Services.

2.4 The Client shall remain at all times responsible for the security, integrity and control of the all CAMEO Directories and CAMEO Analyses Directories provided to the Client pursuant to this Agreement and any backup copies thereof (such backups as permitted by this Agreement) and shall not allow access by or disclosure of these to any third party, except as expressly permitted by this Agreement.

3. **LICENCE**

3.1 Callcredit grants to the Client a personal, non-exclusive, non-transferable licence to use the Deliverables at the Designated Site in accordance with the Scope of Licence only for the duration of this Agreement.

3.2 The Client shall not sell, transfer, distribute or otherwise make the Deliverables themselves available to, or use the Deliverables on behalf of, any third party or create derivative works based on the Deliverables (except as permitted under section 50B of the Copyright, Designs and Patents Act 1988).

3.3 Save as expressly required by Applicable Obligation, the Client may make one copy of the Deliverables for back-up purposes but otherwise shall not copy the Deliverables.

3.4 Save as expressly required by Applicable Obligation, the Client shall not replicate or incorporate (absolutely or to any material extent) any element of the CAMEO Directories or CAMEO Analyses Directories in any Client Materials and shall not adapt, alter, modify, reverse engineer, de-compile or otherwise interfere with any element of the CAMEO Directories or CAMEO Analyses Directories.

3.5 If, at any time during the term of this Agreement, a new revision, update or upgrade of the Deliverables becomes available for general release, any such new revision, update or upgrade shall be made available and implemented to the Client.

3.6 Notwithstanding the provisions of clauses 3.1, 3.2, and 3.3, but without prejudice to the provisions of clauses 3.3, 3.4 and 3.6, if a Authorised Data Processor is identified in Schedule 1, the Client may hold its licensed copy of the Deliverables at the premises of the Authorised Data Processor provided that the Client shall procure that the Authorised Data Processor (i) only uses the Deliverables on behalf of the Client for the Scope of Licence in accordance with the terms and conditions of this Agreement; (ii) does not under any circumstances use the Deliverables on its own behalf or on behalf of a third party; and (iii) complies in full with all terms and conditions of this Agreement (as applicable) to the same extent and to the same standard as if the Authorised Data Processor had executed this Agreement in its own right in place of the Client.

3.7 Where the Authorised Data Processor delivers the Deliverables to the Client the Client acknowledges that the Authorised Data Processor is responsible for ensuring that the Services are not affected by the fact that the Deliverables are utilised or delivered via the Authorised Data Processor and that Callcredit cannot be responsible for any defects or delay in the Services as a result of the Deliverables being delivered via the Authorised Data Processor rather than being delivered direct from Callcredit to the Client.

3.8 The Client agrees that Callcredit shall be permitted to use the Client's name and company logo for reference purposes (both on its website and in its publicity materials) to describe the Client as a recipient of its services. This permission shall continue for a reasonable period after completion of the Services unless otherwise notified to Callcredit in writing.

Callcredit
Information Group
Marketing Solutions

4. INPUT

4.1 This clause 4 shall only apply to the extent the Client provides input in connection with the subject matter of this Agreement.

4.2 The Client grants to Callcredit a non-transferable, non-exclusive royalty-free licence irrevocable during the term of this Agreement to use and copy the input solely for the purpose of enabling Callcredit to carry out its obligations under the Agreement. For the avoidance of doubt, Callcredit will not, subject as hereinafter provided, disclose to any third party any input of the Client, or use the Client's input except in the proper performance of its obligations under this Agreement. Subject to as hereinafter provided, Callcredit will only disclose the Client's input to its personnel on a 'need to know' basis, and must give its personnel (including but not limited to employees, officers, agents and contractors) notice of the confidentiality obligations under this Agreement. Callcredit may disclose the Client's input to AddressDoctor GmbH, to the extent necessary for AddressDoctor GmbH to host and run the Website.

4.3 In order to allow Callcredit to discharge its duties under this Agreement and deliver the Services in accordance with this Agreement, the Client shall provide Callcredit with input which is free from corruption or defect which would prevent or materially limit Callcredit's ability to access and use the input.

4.4 Upon receiving the input in a form which does not in all respects comply with clause 4.3 above, Callcredit shall immediately notify the Client together with details of such non-compliance and the Parties shall agree to one of the following: (i) Callcredit shall correct the input at its own expense; (ii) Callcredit shall correct the input at the Client's expense; or (iii) the Client will supply replacement input.

4.5 The Client shall clearly define to Callcredit any instructions which it has regarding the processing of the input. If the Client fails to do so then Callcredit shall immediately seek clarification from the Client regarding such instructions. If the Client's instructions would require Callcredit to perform work which is outside the scope of Services to be provided by Callcredit under this Agreement, such instructions shall be treated as a request for additional services and Callcredit shall not be required to perform such work until all appropriate changes to this Agreement (including any changes regarding the Price) have been agreed between the Parties. For the avoidance of doubt, Callcredit shall not (and shall not be required to) perform any services (or incur any costs) on behalf of the Client outside the explicit scope of this Agreement without obtaining prior written consent from the Client.

5. OWNERSHIP

5.1 Title, copyright and all other intellectual property rights in the Deliverables and the Services shall at all times remain vested in Callcredit (or its third party licensors) and the Client shall acquire no rights whatsoever therein save as expressly provided in this Agreement.

5.2 Callcredit warrants that it has the right to make the Deliverables and Services available to the Client for the purposes of this Agreement and it is not actually aware of any allegations of infringement or notices of misappropriation issued by any person or any claims that the Services and Deliverables infringe or will infringe any rights, including any intellectual property rights, of any third party.

5.3 Title, copyright and all other intellectual property rights in the Client Materials and any input shall at all times remain vested in the Client (or its third party licensors) and Callcredit shall acquire no rights whatsoever therein save as expressly provided in this Agreement.

5.4 To the extent input is to be provided by the Client to Callcredit pursuant to this Agreement, the Client warrants that it has the right to make any such input available to Callcredit for the purposes of this Agreement and it is not aware of any allegations of infringement or notices of misappropriation issued by any person or any claims that the input infringe or will infringe any rights, including any intellectual property rights, of any third party.

6. COMPLIANCE WITH LAWS AND AUDIT

6.1 Each Party shall at all times in respect of the subject matter of this Agreement comply with all Applicable Obligations.

6.2 The Client shall ensure that all necessary licences, permissions and consents have been obtained and remain in force to allow at all times Callcredit to undertake its obligations under this Agreement, including in respect of Callcredit's processing of any Input.

6.3 The Client agrees that, subject to providing the Client with reasonable prior written notice, it shall allow or otherwise procure the right for Callcredit and its agents to have reasonable access during the Client's normal business hours to the Client's premises and operations for the purpose of ensuring that the Client is complying with its obligations under this Agreement.

6.4 Callcredit agrees to use reasonable endeavours to procure that during any access to the Client's premises under clause 6.2 above, Callcredit's personnel observe and comply with all of the Client's reasonable directions, security procedures, rules, regulations, policies, working hours and holiday schedules and will use reasonable endeavours to minimise any disruption to the Client's normal business operations. Callcredit acknowledges and agrees that the Client may supervise Callcredit's personnel at all times while such personnel are at the Client's premises.

6.5 Callcredit shall use reasonable endeavours to comply with the Client's information security policies, standards and guidelines which, at a reasonably advance time, the Client makes expressly known to Callcredit before Callcredit uses the Client's systems, networks and applications, and when communicating with the Client via email and/or over the Internet in the course of performing Services provided in doing so Callcredit is not put to any material additional expenditure. Callcredit shall use reasonable endeavours to notify the Client of any situation of which it is actually aware and over which it has some control that will or is reasonably likely to put the Client's systems, networks or applications at risk.

7. CONFIDENTIALITY

7.1 Each Party shall in respect of the other Party's Confidential Information keep the Confidential Information in strictest confidence and not make the same available to any third party, or use the other Party's Confidential Information except in the proper performance of its obligations under this Agreement (or, in the case of the Client, in the use of the Services performed under this Agreement).

The obligation contained in clause 7.1 does not arise in connection with Confidential Information which is:

7.1.1 required to be disclosed by Applicable Obligation, provided that the disclosing Party discloses the minimum amount of information required to satisfy the relevant Applicable Obligation;

7.1.2 in the public domain otherwise than as a result of a breach of this Agreement or another obligation of confidence; or

7.1.3 rightfully known by the recipient independently of this Agreement or interaction with the other Party and free of any obligation of confidence.

7.2 It is expressly agreed that each Party may disclose to its Affiliates the Confidential Information of the other Party for the purposes of such former Party discharging its duties or exercising its rights under this Agreement provided that this clause 7.2 shall not grant to the Client any right to derogate from the provisions of clause 5.

8. LIABILITY

8.1 The Client acknowledges that the Deliverables may be based on information provided to Callcredit by third parties over whom Callcredit has no control. Therefore, Callcredit can give no warranties as to the accuracy of the Deliverables nor the suitability of the Deliverables for any specific purposes.

8.2 The Client accepts and agrees that Callcredit does not warrant any decisions taken by the Client as a result of use of the Deliverables to any extent whatsoever and that Callcredit will not be liable for any loss or damage howsoever that may arise from any such decisions made.

8.3 Subject to clause 8.7, Callcredit shall not be liable for any special, indirect or consequential loss arising out of or in connection with this Agreement or its subject matter even if Callcredit had notice of the possibility of such loss.

8.4 Subject to clause 8.7, Callcredit shall not liable for any loss of business, loss of profits, loss of anticipated savings, loss of reputation, loss of goodwill, business interruption or increase in bad debt arising out of or in connection with this Agreement or its subject matter even if had notice of the possibility of such loss.

8.5 Subject to clause 8.7, Callcredit's entire aggregate liability in respect of all claims, losses, damages and costs arising out of or in connection with this Agreement or its subject matter (whether in contract, tort including negligence, breach of statutory duty or otherwise) in any Year shall not exceed an amount equal to the sums paid or payable by Client under this Agreement in respect of that Year or £200 (whichever is the greater).

8.6 Except as expressly provided in this Agreement and subject to clause 8.7, all conditions and warranties or terms of equivalent effect whether express or implied (by statute or otherwise) are excluded to the fullest extent permitted by law.

8.7 Notwithstanding any other term of this Agreement, neither Party limits or excludes liability for death or personal injury arising from its negligence.

9. PAYMENT AND COSTS

9.1 The Client shall pay the Price to Callcredit, together with all travel, accommodation and subsistence expenses reasonably incurred by or on behalf of Callcredit during performance of any of the Services away from Callcredit's premises.

9.2 Callcredit may increase the Price payable hereunder on one occasion each Year during the term of this Agreement. Any such increase shall not exceed

Callcredit Information Group
Marketing Solutions

the increase (expressed as a percentage) in the Retail Prices (all items) index since the Commencement Date or since the date of the last increase (if any) in Callcredit's charges, whichever is the later. If that index ceases to be published then Callcredit and the Client shall agree another comparable replacement index (such agreement not to be unreasonably withheld or delayed or made subject to unreasonable conditions).

9.3 Any value added, sales or other applicable tax, custom or excise ("Applicable Tax") is to be paid by the Client at the prevailing rate on all sums due under this Agreement. All sums due are quoted exclusive of any such Applicable Tax.

9.4 All sums due must be paid within 30 days of the date of Callcredit's invoice. The Client shall pay interest on any overdue payment at a compound rate of 5% per annum above Barclays Bank plc base rate from time to time from the due date until the date of payment, such interest to accrue on a daily basis both before and after judgment.

**10. TERMINATION FOR CAUSE**

Either Callcredit or the Client may terminate this Agreement (including all licences granted hereunder) immediately on written notice if:

10.1 the other Party commits any material breach of this Agreement and such breach (where capable of remedy) is not remedied to the non-defaulting Party's reasonable satisfaction within 30 days of written notice specifying the breach and requiring its remedy; or

10.2 in respect of the other Party a resolution is passed or an order is made for winding up (save for the purpose of a bona fide reconstruction or amalgamation); or

10.3 in respect of the other Party an administration order is made, or a receiver or administrative receiver is appointed over any of its property or assets;

10.4 the other Party is dissolved or is insolvent or would be taken to be insolvent under section 123 of the Insolvency Act 1986; or

10.5 an event analogous to any of the events listed in clauses 10.1 to 10.4 inclusive occurs to the other Party.

**11. CONSEQUENCES OF TERMINATION**

11.1 On termination of this Agreement or any licence for any reason, the Client shall:

11.1.1 immediately return to Callcredit all hard copies of all Deliverables (and all copies thereof) including the Original Directories; and

11.1.2 delete all Deliverables (including all flags and other data and information appended to or forming part of any of the Client's databases which are derived from the Deliverables, save for the Client's Materials) from the Client's computer systems and other storage media; and

11.1.3 provide Callcredit with a certificate of compliance with the provisions of this clause 11 signed by a duly authorised officer.

11.2 Nothing in clause 11 shall require the Client to surrender to Callcredit or destroy any Client Materials.

11.3 Upon termination of this Agreement or any licence for any reason, Callcredit shall delete any information, data or materials including any Client's Materials which were provided by the Client to Callcredit in relation to this Agreement and shall provide to the Client, upon request, a certificate of compliance with the provisions of this clause 11.3 signed by a duly authorised officer. Upon termination of this Agreement by Client for Callcredit's material breach pursuant to Clause 10.1, Callcredit shall refund to the Client a proportion of the Price that reflects a pre-payment for such Services or Deliverables, such proportion to reflect the unexpired part of the period in respect of which the Price was paid.

11.4 Callcredit may terminate:

11.4.1 any or all of the licences granted hereunder; or

11.4.2 this Agreement

if as a result of any act or default of any third party Callcredit is no longer permitted or licensed to provide the Services and/or to make the Deliverables available pursuant to this Agreement or if in Callcredit's reasonable opinion it is necessary to do so to comply with any law, regulation or applicable code of practice. In the event of such termination, Callcredit shall refund to the Client a proportion of the Price that reflects a pre-payment for such Services or Deliverables, such proportion to reflect the unexpired part of the period in respect of which the Price was paid.

11.5 Save for the refund in respect of matters referred to in clause 11.4:

11.5.1 Callcredit shall have no liability whatsoever to the Client; and

11.5.2 Callcredit shall not be deemed to be in breach of any of its obligations under this Agreement.

11.6 Termination is without prejudice to any antecedent breach or to any continuing obligation. The provisions of clauses 1, 3, 4, 7, 8, 11, this 11.6 and 10 shall continue after termination or expiry of this Agreement.

11.7 The Client shall pay all amounts due and owing up to the date of termination of this Agreement.

**12. GENERAL**

12.1 This Agreement sets out the entire agreement between the Parties in relation to its subject matter and supersedes all previous written or oral agreements, representations, undertakings, warranties or arrangements of any nature whatsoever between the Parties.

12.2 The Parties acknowledge and agree that in entering into this Agreement they have not relied on any statement, representation, assurance or warranty (whether made negligently or innocently) other than as expressly set out in this Agreement.

12.3 Nothing in clauses 12.1 or 12.2 will have the effect of limiting or restricting any liability of the Parties arising as a result of any fraud or fraudulent misrepresentation.

12.4 No variation of this Agreement shall be binding on either Party unless made in writing and signed by an authorised signatory of each of the Client and Callcredit.

12.5 Neither Party shall be liable to the other for any delay or non-performance of its obligations under this Agreement (except for any obligation to make payment) arising from any cause beyond its reasonable control including any of the following: act of God, governmental act, war, fire, flood, explosion or civil commotion, industrial action, epidemics, any changes in Applicable Obligation or failure in telecommunications services.

12.6 The Client may not assign, subcontract, sublicense or otherwise transfer any of its rights or obligations under this Agreement without Callcredit's prior written consent signed by an authorised signatory.

12.7 Any concession or indulgence made by either Party shall not be considered as a waiver of its rights.

12.8 The headings in this Agreement are for convenience only and do not affect its meaning or construction.

12.9 Any use of "including" shall be deemed to mean "including without limitation".

12.10 This Agreement shall be construed in accordance with the laws of England and the English Courts shall have exclusive jurisdiction.

12.11 A person who is not a party to this Agreement has no right under the Contracts (Rights of Third Parties) Act 1999 to rely on or enforce any term of this Agreement.

12.12 Any reference to a statutory provision includes a reference to any modification or re-enactment of it from time to time.

12.13 If any provision of this Agreement is found illegal or unenforceable by a court of competent jurisdiction, the rest of this Agreement will remain valid; and the illegal or unenforceable provision will be automatically modified as needed to make that provision enforceable.

12.14 Any notice given under this Agreement shall be in writing and shall be effective when delivered personally to the address set forth in this Agreement, or five (5) days following mailing by registered post. Either Party may designate a different address by notice to the other Party given in accordance herewith.

# References

ABBAS, J., OJO, A. & ORANGE, S. 2009. Geodemographics – a tool for health intelligence? *Public Health,* 123**,** e35-e39.

ADAMS, J., WHITE, M. & FORMAN, D. 2004. Are there socioeconomic gradients in stage and grade of breast cancer at diagnosis? Cross sectional analysis of UK cancer registry data. *British Medical Journal,* 329**,** 142.

AGGARWAL, A., MONSIVAIS, P., COOK, A. J. & DREWNOWSKI, A. 2011. Does diet cost mediate the relation between socioeconomic position and diet quality? *European Journal of Clinical Nutrition,* 65**,** 1059-1066.

AGGARWAL, A., MONSIVAIS, P. & DREWNOWSKI, A. 2012. Nutrient intakes linked to better health outcomes are associated with higher diet costs in the US. *PLoS ONE [Electronic Resource],* 7.

AINSWORTH, B. E., HASKELL, W. L., WHITT, M. C., IRWIN, M. L., SWARTZ, A. M., STRATH, S. J., O'BRIEN, W. L., BASSETT, D. R., JR., SCHMITZ, K. H., EMPLAINCOURT, P. O., JACOBS, D. R., JR. & LEON, A. S. 2000. Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise,* 32**,** S498-504.

AKBARI, M. E., AKBARI, A., HOMAEE, F., TABATABAEE, M. & KHAYAMZADEH, M. 2010. Parity and breast feeding are preventive measures for breast cancer in Iranian women. *European Journal of Cancer, Supplement,* Conference: European Breast Cancer Conference, EBCC7 Barcelona Spain. Conference Start: 20100324 Conference End: 20100327. Conference Publication: (var.pagings). 8 (3)**,** 212.

ANDRIEU, E., DARMON, N. & DREWNOWSKI, A. 2006. Low-cost diets: more energy, fewer nutrients. *European Journal of Clinical Nutrition,* 60**,** 434-6.

APPLEBY, P. N., DAVEY, G. K. & KEY, T. J. 2002. Hypertension and blood pressure among meat eaters, fish eaters, vegetarians and vegans in EPIC-Oxford. *Public Health Nutrition,* 5**,** 645-654.

AUSTRALIAN BUREAU OF STATISTICS. 2008. *SEIFA: Socio-Economic Indexes for Areas* [Online]. Available: http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Seifa_entry_page [Accessed February 2013].

AUSTRALIAN BUREAU OF STATISTICS. 2011. *Australian Standard Geographical Classification (ASGC) Digital Boundaries, Australia* [Online]. Available: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1259.0.30.001July%202011?OpenDocument [Accessed November 2012].

AUSTRALIAN GOVERNMENT 2005. Dietary guidelines for Australians: A Guide to healthy eating.

BAANDERS, A. N. & DE WAARD, F. 1992. Breast cancer in Europe: the importance of factors operating at an early age. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP),* 1**,** 285-291.

BALL, K., CRAWFORD, D. & MISHRA, G. 2006. Socio-economic inequalities in women's fruit and vegetable intakes: a multilevel study of individual, social and environmental mediators. *Public Health Nutrition,* 9**,** 623-30.

BALL, K., MISHRA, G. D., THANE, C. W. & HODGE, A. 2004. How well do Australian women comply with dietary guidelines? *Public Health Nutrition,* 7**,** 443-52.

BALL, K. & THORNTON, L. 2013. Food environments: measuring, mapping, monitoring and modifying. *Public Health Nutrition,* 16**,** 1147-50.

BEATTY, T. K. M. 2010. Do the Poor Pay More for Food?: Evidence from the United Kingdom. *American Journal of Agricultural Economics,* 92**,** 608-621.

BERNSTEIN, A. M., BLOOM, D. E., ROSNER, B. A., FRANZ, M. & WILLETT, W. C. 2010. Relation of food cost to healthfulness of diet among US women. *American Journal of Clinical Nutrition,* 92**,** 1197-203.

BEYDOUN, M. A., POWELL, L. M. & WANG, Y. F. 2008. The association of fast food, fruit and vegetable prices with dietary intakes among US adults: Is there modification by family income? *Social Science & Medicine,* 66**,** 2218-2229.

BINGHAM, S. A., GILL, C., WELCH, A., DAY, K., CASSIDY, A., KHAW, K. T., SNEYD, M. J., KEY, T. J. A., ROE, L. & DAY, N. E. 1994. Comparison of Dietary Assessment Methods in Nutritional Epidemiology - Weighed Records V 24-H Recalls, Food-Frequency Questionnaires and Estimated-Diet Records. *British Journal of Nutrition,* 72**,** 619-643.

BIRKIN, M. & CLARKE, G. 2012. The enhancement of spatial microsimulation models using geodemographics. *The Annals of Regional Science*.

BLACK, C., NTANI, G., KENNY, R., TINATI, T., JARMAN, M., LAWRENCE, W., BARKER, M., INSKIP, H., COOPER, C., MOON, G. & BAIRD, J. 2012. Variety and quality of healthy foods differ according to neighbourhood deprivation. *Health Place,* 18**,** 1292-1299.

BRAVEMAN, P. A., CUBBIN, C., EGERTER, S., CHIDEYA, S., MARCHI, K. S., METZLER, M. & POSNER, S. 2005. Socioeconomic status in health research: one size does not fit all. *JAMA,* 294**,** 2879-88.

BRENNAN, S. F., CANTWELL, M. M., CARDWELL, C. R., VELENTZIS, L. S. & WOODSIDE, J. V. 2010. Dietary patterns and breast cancer risk: a systematic review and meta-analysis. *American Journal of Clinical Nutrition,* 91**,** 1294-302.

BRINSDEN, H., LOBSTEIN, T., LANDON, J., KRAAK, V., SACKS, G., KUMANYIKA, S., SWINBURN, B., BARQUERA, S., FRIEL, S., HAWKES, C., KELLY, B., L'ABBE, M., LEE, A., MA, J., MACMULLEN, J., MOHAN, S., MONTEIRO, C., NEAL, B., RAYNER, M., SANDERS, D., SNOWDON, W., VANDEVIJVERE, S. & WALKER, C. 2013. Monitoring policy and actions on food environments: rationale and outline of the INFORMAS policy engagement and communication strategies. *Obesity Reviews,* 14 Suppl 1**,** 13-23.

BRONFENBRENNER, U. 1977. Toward an Experimental Ecology of Human Development. *American Psychologist,* 32**,** 513-531.

BROWN, S. B., HOLE, D. J. & COOKE, T. G. 2007. Breast cancer incidence trends in deprived and affluent Scottish women. *Breast Cancer Research and Treatment,* 103**,** 233-8.

BROWN, W. J., BRYSON, L., BYLES, J. E., DOBSON, A. J., LEE, C., MISHRA, G. & SCHOFIELD, M. 1998. Women's Health Australia: recruitment for a national longitudinal cohort study. *Women Health,* 28**,** 23-40.

BURLEY, V. J., GREENWOOD, D. C., HEPWORTH, S. J., FRASER, L. K., DE KOK, T. M., VAN BREDA, S. G., KYRTOPOULOS, S. A., BOTSIVALI, M., KLEINJANS, J., MCKINNEY, P. A. & CADE, J. E. 2010. Dietary acrylamide intake and risk of breast cancer in the UK women's cohort. *British Journal of Cancer,* 103**,** 1749-54.

CACI LTD. *Acorn geodemographic classification* [Online]. Available: http://www.caci.co.uk/acorn-classification.aspx.

CADE, J., UPMEIER, H., CALVERT, C. & GREENWOOD, D. 1999. Costs of a healthy diet: analysis from the UK Women's Cohort Study. *Public Health Nutrition,* 2**,** 505-12.

CADE, J. E., BURLEY, V. J. & GREENWOOD, D. C. 2007. Dietary fibre and risk of breast cancer in the UK Women's Cohort Study. *International Journal of Epidemiology,* 36**,** 431-8.

CADE, J. E., BURLEY, V. J., GREENWOOD, D. C. & GROUP, U. K. W. S. C. S. S. 2004. The UK Women's Cohort Study: comparison of vegetarians, fish-eaters and meat-eaters. *Public Health Nutrition,* 7**,** 871-8.

CADE, J. E., TAYLOR, E. F., BURLEY, V. J. & GREENWOOD, D. C. 2010. Common dietary patterns and risk of breast cancer: analysis from the United Kingdom Women's Cohort Study. *Nutrition & Cancer,* 62**,** 300-6.

CADE, J. E., TAYLOR, E. F., BURLEY, V. J. & GREENWOOD, D. C. 2011. Does the Mediterranean dietary pattern or the Healthy Diet Index influence the risk of breast cancer in a large British cohort of women? *European Journal of Clinical Nutrition,* 65**,** 920-8.

CALLCREDIT INFORMATION GROUP. 2013. *CAMEO customer segmentation & analysis for the UK & overseas* [Online]. Available: http://www.callcredit.co.uk/products-and-services/consumer-marketing-data/segmentation-analysis.

CALLE, E. E., HEATH, C. W., JR., MIRACLE-MCMAHILL, H. L., COATES, R. J., LIFF, J. M., FRANCESCHI, S., TALAMINI, R., CHANTARAKUL, N., KOETSAWANG, S., RACHAWAT, D., MORABIA, A., SCHUMAN, L., STEWART, W., SZKLO, M., BAIN, C., SCHOFIELD, F., SISKIND, V., BAND, P., COLDMAN, A. J. & ET AL. 1996. Breast cancer and hormonal contraceptives: Further results Collaborative group on hormonal factors in breast cancer. *Contraception,* 54 (3 SUPPL.)**,** 1S-106S.

CANCER RESEARCH UK. 2012. *Breast cancer Key Facts* [Online]. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/breast-cancer/.

CENSUS DISSEMINATION UNIT 2009. *GeoConvert* [Online]. Available: http://geoconvert.mimas.ac.uk/ [Accessed August 2010].

CENTRE FOR INTERACTION DATA ESTIMATION AND RESEARCH. 2011. *Flow Data* [Online]. Available: http://cider.census.ac.uk/cider/wicid/wicid_library.php [Accessed July 2011].

CHAPMAN, K., INNES-HUGHES, C., GOLDSBURY, D., KELLY, B., BAUMAN, A. & ALLMAN-FARINELLI, M. 2013. A comparison of the cost of generic and branded food products in Australian supermarkets. *Public Health Nutrition,* 16**,** 894-900.

CLEVES, M., GOULD, W., GUTIERREZ, R. & MARCHENKO, Y. 2010. *An Introduction to Survival Analysis Using Stata,* US, Stata Press.

COLE, S., FILOMENA, S. & MORLAND, K. 2010. Analysis of fruit and vegetable cost and quality among racially segregated neighborhoods in Brooklyn, New York. *Journal of Hunger and Environmental Nutrition,* 5**,** 202-215.

COLLABORATIVE GROUP ON HORMONAL FACTORS IN BREAST CANCER 1997. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. *Lancet,* 350**,** 1047-1059.

COMMITTEE ON MEDICAL ASPECTS OF FOOD POLICY & DEPARTMENT OF HEALTH 1991. *Dietary Reference Values of Food Energy and Nutrients for the United Kingdom*, Stationery Office Books (9 July 1991).

CRAWFORD, D., JEFFERY, R., BALL, K. & BRUG, J. (eds.) 2010. *Obesity Epidemiology,* United States: Oxford University Press.

CROSSLEY, T., LOW, H. & O'DEA, C. 2012. Household Consumption Through Recent Recessions. Institute for fiscal studies - working paper W11/18.

CUMMINS, S. 2005a. Food environments and obesity--neighbourhood or nation? *International Journal of Epidemiology,* 35**,** 100-104.

CUMMINS, S. 2005b. Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *Journal of Epidemiology & Community Health,* 59**,** 1035-1040.

CUMMINS, S. 2007. Commentary: Investigating neighbourhood effects on health - avoiding the 'Local Trap'. *International Journal of Epidemiology,* 36**,** 355-357.

CUMMINS, S. & MACINTYRE, S. 2002a. "Food deserts" - evidence and assumption in health policy making. *British Medical Journal,* 325**,** 436-438.

CUMMINS, S. & MACINTYRE, S. 2002b. A Systematic Study of an Urban Foodscape: The Price and Availability of Food in Greater Glasgow. *Urban Studies,* 39**,** 2115-2130.

CUMMINS, S., SMITH, D. M., AITKEN, Z., DAWSON, J., MARSHALL, D., SPARKS, L. & ANDERSON, A. S. 2010. Neighbourhood deprivation and the price and availability of fruit and vegetables in Scotland. *Journal of Human Nutrition & Dietetics,* 23**,** 494-501.

DANAEI, G., VANDER HOORN, S., LOPEZ, A. D., MURRAY, C. J. & EZZATI, M. 2005. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet,* 366**,** 1784-93.

DANO, H., ANDERSEN, O., EWERTZ, M., PETERSEN, J. H. & LYNGE, E. 2003. Socioeconomic status and breast cancer in Denmark. *International Journal of Epidemiology,* 32**,** 218-24.

DARMON, N., BRIEND, A. & DREWNOWSKI, A. 2004. Energy-dense diets are associated with lower diet costs: a community study of French adults. *Public Health Nutrition,* 7**,** 21-7.

DATA.GOV.UK 2010. English Indices of Deprivation. *In:* GOVERNMENT, D. F. C. A. L. (ed.).

DAVIES, S. 2011. Annual report of the Chief Medical Officer - On the state of the public's health. *In:* HEALTH, D. O. (ed.). UK.

DEFRA 2004. Family Food: A report on the 2003-04 Expenditure and Food Survey. The Office for National Statistics: Defra.

DEFRA 2009. Family Food: A report on the 2009 Family Food Module of the Living Costs and Food Survey. The Office for National Statistics: Defra.

DEFRA 2010. Household Food and Drink Waste linked to Food and Drink Purchases. Department for Environment Food and Rural Affairs.

DELPEUCH, F., MAIRE, B., MONNIER, E. & HOLDSWORTH, M. 2009. *Globesity,* UK, Earthscan.

DEPARTMENT FOR ENVIRONMENT FOOD AND RURAL AFFAIRS. 2004. Family Food: A report on the 2003-04 Expenditure and Food Survey. Available: http://www.defra.gov.uk/statistics/files/defra-stats-food-family-annual-2004.pdf [Accessed October 2011].

DEPARTMENT OF HEALTH 2011. The Eatwell Plate.

DEPARTMENT OF HEALTH & FOOD STANDARDS AGENCY 2012. National Diet and Nutrition Survey: Headline results from Years 1, 2 and 3 (combined) of the Rolling Programme (2008/2009 – 2010/11). *In:* BEVERLEY BATES, ALISON LENNOX, ANN PRENTICE, CHRIS BATES & SWAN, G. (eds.).

DIEZ ROUX, A. V., NIETO, J. F., CAULFIELD, L., TYROLER, H. A., WATSON, R. L. & SZKLO, M. 1999. Neighbourhood differences in diet: the Atherosclerosis Risk in Communities (ARIC) Study. *Journal of Epidemiology & Community Health,* 53**,** 55-69.

DOLL, R. & PETO, R. 1981. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *Journal of the National Cancer Institute,* 66**,** 1191-308.

DRESCHER, L. S., THIELE, S. & MENSINK, G. B. 2007. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *Journal of Nutrition,* 137**,** 647-51.

DREWNOWSKI, A., HENDERSON, S. A., DRISCOLL, A. & ROLLS, B. J. 1997. The Dietary Variety Score: assessing diet quality in healthy young and older adults. *Journal of the American Dietetic Association,* 97**,** 266-71.

DREWNOWSKI, A., MONSIVAIS, P., MAILLOT, M. & DARMON, N. 2007. Low-energy-density diets are associated with higher diet quality and higher diet costs in French adults. *Journal of the American Dietetic Association,* 107**,** 1028-32.

DREWNOWSKI, A. & SPECTER, S. E. 2004. Poverty and obesity: the role of energy density and energy costs. *American Journal of Clinical Nutrition,* 79**,** 6-16.

EDWARDS, K. L., CADE, J. E., RANSLEY, J. K. & CLARKE, G. P. 2009. A cross-sectional study examining the pattern of childhood obesity in Leeds: affluence is not protective. *Archives of Disease in Childhood,* 95**,** 94-99.

EDWARDS, K. L. & CLARKE, G. P. 2009. The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity. *Soc Sci Med,* 69**,** 1127-34.

ERICSON, U., SONESTEDT, E., GULLBERG, B., OLSSON, H. & WIRFALT, E. 2007. High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmo Diet and Cancer cohort. *American Journal of Clinical Nutrition,* 86**,** 434-43.

ESRI INC 2010. ArcMap 10.

EUROPEAN FOOD INFORMATION COUNCIL. 2009. *Food-Based Dietary Guidelines in Europe* [Online]. Available: http://www.eufic.org/article/en/expid/food-based-dietary-guidelines-in-europe/ [Accessed April 2013].

FOOD STANDARDS AGENCY (ed.) 1998. *Food Portion Sizes,* London: The Stationary Office.

FOOD STANDARDS AGENCY & DEPARTMENT OF HEALTH 2002. National Diet & Nutrition Survey: adults aged 19 to 64 years.

FORESIGHT. 2007. Tackling Obesities: Future Choices – Project report.

FRASER, L. K. & EDWARDS, K. L. 2010. The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK. *Health & Place,* 16**,** 1124-1128.

FREEDMAN, L. S., SCHATZKIN, A., MIDTHUNE, D. & KIPNIS, V. 2011. Dealing with dietary measurement error in nutritional cohort studies. *Journal of the National Cancer Institute,* 103**,** 1086-92.

GENERAL REGISTER OFFICE FOR SCOTLAND 2001. Population Estimates.

GIBNEY, M. (ed.) 2004. *Public Health Nutrition*: Blackwell Science Ltd.

GISKES, K., TURRELL, G., PATTERSON, C. & NEWMAN, B. 2002. Socio-economic differences in fruit and vegetable consumption among Australian adolescents and adults. *Public Health Nutrition,* 5**,** 663-9.

GLANZ, K., BASIL, M., MAIBACH, E., GOLDBERG, J. & SNYDER, D. 1998. Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *Journal of the American Dietetic Association,* 98**,** 1118-26.

GORTMAKER, S. L., SWINBURN, B. A., LEVY, D., CARTER, R., MABRY, P. L., FINEGOOD, D. T., HUANG, T., MARSH, T. & MOODIE, M. L. 2011. Changing the future of obesity: science, policy, and action. *Lancet,* 378**,** 838-47.

GREENLAND, S., PEARL, J. & ROBINS, J. M. 1999. Causal diagrams for epidemiologic research. *Epidemiology,* 10**,** 37-48.

GREENWOOD, D., GILTHORPE, M., GOLDING, C. & CADE, J. 2003. Stability over time of dietary patterns in the UK Women's Cohort Study. *Proceedings of the Nutrition Society,* 62.

GREENWOOD, D. C., CADE, J. E., DRAPER, A., BARRETT, J. H., CALVERT, C. & GREENHALGH, A. 2000. Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *European Journal of Clinical Nutrition,* 54**,** 314-20.

GUSTAFSON, A., CHRISTIAN, J. W., LEWIS, S., MOORE, K. & JILCOTT, S. 2013. Food venue choice, consumer food environment, but not food venue availability within daily travel patterns are associated with dietary intake among adults, Lexington Kentucky 2011. *Nutrition Journal,* 12.

HACKING, J. M., MULLER, S. & BUCHAN, I. E. 2011. Trends in mortality from 1965 to 2008 across the English north-south divide: comparative observational study. *British Medical Journal,* 342**,** d508.

HALL, K. D., SACKS, G., CHANDRAMOHAN, D., CHOW, C. C., WANG, Y. C., GORTMAKER, S. L. & SWINBURN, B. A. 2011. Quantification of the effect of energy imbalance on bodyweight. *The Lancet,* 378**,** 826-837.

HAUNER, H., BECHTHOLD, A., BOEING, H., BRONSTRUP, A., BUYKEN, A., LESCHIK-BONNET, E., LINSEISEN, J., SCHULZE, M., STROHM, D. & WOLFRAM, G. 2012. Evidence-based guideline of the German Nutrition Society: carbohydrate intake and prevention of nutrition-related diseases. *Annals of Nutrition and Metabolism,* 60 Suppl 1**,** 1-58.

HEALTH & SOCIAL CARE INFORMATION CENTRE 2011. Health Survey for England.

HEALTH & SOCIAL CARE INFORMATION CENTRE. 2013a. *Methodological Changes to Health Survey for England (HSE) 2013* [Online]. Available: http://www.hscic.gov.uk/media/11361/Methodological-Changes-to-Health-Survey-for-England-HSE-2013-NS/pdf/MethChange201312_HSE_Sample_Design.pdf

HEALTH & SOCIAL CARE INFORMATION CENTRE 2013b. National Child Measurement Programme.

HEALTH & SOCIAL CARE INFORMATION CENTRE 2013c. Statistics on Obesity, Physical Activity and Diet - England 2013.

HECK, K. E. & PAMUK, E. R. 1997. Explaining the relation between education and postmenopausal breast cancer. *American Journal of Epidemiology,* 145**,** 366-72.

HILL, J. O. & PETERS, J. C. 1998. Environmental contributions to the obesity epidemic. *Science,* 280**,** 1371-1374.

HOLLAND B, W. A., UNWIN ID, BUSS DH, PAUL AA, SOUTHGATE DAT 1991. *McCance and Widdowson's the Composition of Foods,* London.

HU, F. B. 2002. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol,* 13**,** 3-9.

HU, F. B. 2008. *Obesity Epidemiology,* USA, Oxford University Press.

HUGHES, R. J., EDWARDS, K. L., CLARKE, G. P., EVANS, C. E., CADE, J. E. & RANSLEY, J. K. 2012. Childhood consumption of fruit and vegetables across England: a study of 2306 6-7-year-olds in 2007. *British Journal of Nutrition***,** 1-10.

HUTCHINSON, J., BURLEY, V. J., GREENWOOD, D. C., THOMAS, J. D. & CADE, J. E. 2010. High-dose vitamin C supplement use is associated with self-reported histories of breast cancer and other illnesses in the UK Women's Cohort Study. *Public Health Nutrition,* 14**,** 768-777.

HUTCHINSON, J., LENTJES, M. A., GREENWOOD, D. C., BURLEY, V. J., CADE, J. E., CLEGHORN, C. L., THREAPLETON, D. E., KEY, T. J., CAIRNS, B. J., KEOGH, R. H., DAHM, C. C., BRUNNER, E. J., SHIPLEY, M. J., KUH, D., MISHRA, G., STEPHEN, A. M., BHANIANI, A., BORGULYA, G. & KHAW, K. T. 2012. Vitamin C intake from diary recordings and risk of breast cancer in the UK Dietary Cohort Consortium. *European Journal of Clinical Nutrition,* 66**,** 561-8.

INSTITUTE FOR HEALTH METRICS AND EVALUATION 2013. The Global Burden of Disease: Generating Evidence, Guiding Policy – European Union and Free Trade Association Regional Edition.

JABS, J. & DEVINE, C. M. 2006. Time scarcity and food choices: an overview. *Appetite,* 47**,** 196-204.

JIAO, J., MOUDON, A. V., ULMER, J., HURVITZ, P. M. & DREWNOWSKI, A. 2012. How to identify food deserts: measuring physical and economic access to supermarkets in King County, Washington. *Am J Public Health,* 102**,** e32-9.

KAMPHUIS, C. B. M., GISKES, K., DE BRUIJN, G. J., WENDEL-VOS, W., BRUG, J. & VAN LENTHE, F. J. 2006. Environmental determinants of fruit and vegetable consumption among adults: a systematic review. *British Journal of Nutrition,* 96**,** 620-635.

KENNEDY, E. T., OHLS, J., CARLSON, S. & FLEMING, K. 1995. The Healthy Eating Index: design and applications. *Journal of the American Dietetic Association,* 95**,** 1103-8.

KEY, T., ALLEN, N., SPENCER, E. & TRAVIS, R. 2002. The effect of diet on risk of cancer. *The Lancet,* 360**,** 861-868.

KEY, T. J., APPLEBY, P. N., CAIRNS, B. J., LUBEN, R., DAHM, C. C., AKBARALY, T., BRUNNER, E. J., BURLEY, V., CADE, J. E., GREENWOOD, D. C., STEPHEN, A. M., MISHRA, G., KUH, D., KEOGH, R. H., WHITE, I. R., BHANIANI, A., BORGULYA, G., MULLIGAN, A. A. & KHAW, K. T. 2011. Dietary fat and breast cancer: comparison of results from food diaries and food-frequency questionnaires in the UK Dietary Cohort Consortium. *American Journal of Clinical Nutrition,* 94**,** 1043-52.

KIM, S., HAINES, P. S., SIEGA-RIZ, A. M. & POPKIN, B. M. 2003. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. *Journal of Nutrition,* 133**,** 3476-84.

KING COUNTY. 2013. *KC GIS Cente* [Online]. Available: ftp://ftp.kingcounty.gov/gis/Web/GISData/urban_growth_SHP.zip [Accessed August 2013].

KRAUSS, R. M., DECKELBAUM, R. J., ERNST, N., FISHER, E., HOWARD, B. V., KNOPP, R. H., KOTCHEN, T., LICHTENSTEIN, A. H., MCGILL, H. C., PEARSON, T. A., PREWITT, T. E., STONE, N. J., HORN, L. V. & WEINBERG, R. 1996. Dietary guidelines for healthy American adults. A statement for health professionals from the Nutrition Committee, American Heart Association. *Circulation,* 94**,** 1795-800.

LARSEN, K. & GILLILAND, J. 2009. A farmers' market in a food desert: Evaluating impacts on the price and availability of healthy food. *Health & Place,* 15**,** 1158-1162.

LARSON, N. I., STORY, M. T. & NELSON, M. C. 2009. Neighborhood environments: disparities in access to healthy foods in the U.S. *American Journal of Preventive Medicine,* 36**,** 74-81.

LEE, A., MHURCHU, C. N., SACKS, G., SWINBURN, B., SNOWDON, W., VANDEVIJVERE, S., HAWKES, C., L'ABBE, M., RAYNER, M., SANDERS, D., BARQUERA, S., FRIEL, S., KELLY, B.,

KUMANYIKA, S., LOBSTEIN, T., MA, J., MACMULLAN, J., MOHAN, S., MONTEIRO, C., NEAL, B. & WALKER, C. 2013. Monitoring the price and affordability of foods and diets globally. *Obesity Reviews,* 14 (Suppl. 1)**,** 82-95.

LEE, C., DOBSON, A. J., BROWN, W. J., BRYSON, L., BYLES, J., WARNER-SMITH, P. & YOUNG, A. F. 2005. Cohort profile: The Australian Longitudinal Study on Women's Health. *International Journal of Epidemiology,* 34**,** 987-991.

LEE, J., RALSTON, R. A. & TRUBY, H. 2011. Influence of food cost on diet quality and risk factors for chronic disease: A systematic review. *Nutrition and Dietetics,* 68**,** 248-261.

LENNERNAS, M., FJELLSTROM, C., BECKER, W., GIACHETTI, I., SCHMITT, A., REMAUT DE WINTER, A. & KEARNEY, M. 1997. Influences on food choice perceived to be important by nationally-representative samples of adults in the European Union. *European Journal of Clinical Nutrition,* 51 Suppl 2**,** S8-15.

LIESE, A. D., WEIS, K. E., PLUTO, D., SMITH, E. & LAWSON, A. 2007. Food Store Types, Availability, and Cost of Foods in a Rural Environment. *Journal of the American Dietetic Association,* 107**,** 1916-1923.

LOPEZ, C. N., MARTINEZ-GONZALEZ, M. A., ALONSO, A., SANCHEZ-VILLEGAS, A., DE LA FUENTE, C. & BES-RASTROLLO, M. 2009a. Cost of compliance with daily recommended values of micronutrients among a cohort of Spanish university graduates: the SUN (Seguimiento Universidad de Navarra) Study. *Public Health Nutrition,* 12**,** 2092-2096.

LOPEZ, C. N., MARTINEZ-GONZALEZ, M. A., SANCHEZ-VILLEGAS, A., ALONSO, A., PIMENTA, A. M. & BES-RASTROLLO, M. 2009b. Costs of Mediterranean and western dietary patterns in a Spanish cohort and their relationship with prospective weight change. *Journal of Epidemiology and Community Health,* 63**,** 920-927.

MACINTYRE, S. 2007. Deprivation amplification revisited: or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *International Journal of Behavioral Nutrition and Physical Activity,* 4.

MACINTYRE, S., ELLAWAY, A. & CUMMINS, S. 2002. Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine,* 55**,** 125-39.

MAILLOT, M., DARMON, N., DARMON, M., LAFAY, L. & DREWNOWSKI, A. 2007a. Nutrient-dense food groups have high energy costs: an econometric approach to nutrient profiling. *Journal of Nutrition,* 137**,** 1815-20.

MAILLOT, M., DARMON, N., VIEUX, F. & DREWNOWSKI, A. 2007b. Low energy density and high nutritional quality are each associated with higher diet costs in French adults. *American Journal of Clinical Nutrition,* 86**,** 690-6.

MAPINFO 2012. MapInfo Professional 11.0.4.

MARMOT, M. 2010. *Fair Society, Healthy Lives - The Marmot Review* [Online]. Available: http://www.marmotreview.org/AssetLibrary/pdfs/Reports/FairSocietyHealthyLives.pdf [Accessed August 2010].

MCNAUGHTON, S. A., BALL, K., CRAWFORD, D. & MISHRA, G. D. 2008. An index of diet and eating patterns is a valid measure of diet quality in an Australian population. *Journal of Nutrition,* 138**,** 86-93.

MISHRA, G., BALL, K., ARBUCKLE, J. & CRAWFORD, D. 2002. Dietary patterns of Australian adults and their association with socioeconomic status: results from the 1995 National Nutrition Survey. *European Journal of Clinical Nutrition,* 56**,** 687-93.

MISHRA, G., BALL, K., PATTERSON, A., BROWN, W., HODGE, A. & DOBSON, A. 2005. Socio-demographic inequalities in the diets of mid-aged Australian women. *European Journal of Clinical Nutrition,* 59**,** 185-95.

MISHRA, G. D., MCNAUGHTON, S. A., BALL, K., BROWN, W. J., GILES, G. G. & DOBSON, A. J. 2010. Major dietary patterns of young and middle aged women: results from a prospective Australian cohort study. *European Journal of Clinical Nutrition,* 64**,** 1125-33.

MONSIVAIS, P., AGGARWAL, A. & DREWNOWSKI, A. 2011. Following federal guidelines to increase nutrient consumption may lead to higher food costs for consumers. *Health affairs (Project Hope),* 30**,** 1471-1477.

MONSIVAIS, P., AGGARWAL, A. & DREWNOWSKI, A. 2012. Are socio-economic disparities in diet quality explained by diet cost? *Journal of Epidemiology and Community Health,* 66**,** 530-535.

MONSIVAIS, P. & DREWNOWSKI, A. 2009. Lower-energy-density diets are associated with higher monetary costs per kilocalorie and are consumed by women of higher socioeconomic status. *Journal of the American Dietetic Association,* 109**,** 814-22.

MONSIVAIS, P., MCLAIN, J. & DREWNOWSKI, A. 2010. The rising disparity in the price of healthful foods: 2004–2008. *Food Policy,* 35**,** 514-520.

MONSIVAIS, P., PERRIGUE, M. M., ADAMS, S. L. & DREWNOWSKI, A. 2013. Measuring diet cost at the individual level: a comparison of three methods. *European Journal of Clinical Nutrition,* 67**,** 1220-5.

MOON, G., QUARENDON, G., BARNARD, S., TWIGG, L. & BLYTH, B. 2007. Fat nation: Deciphering the distinctive geographies of obesity in England. *Social Science & Medicine,* 65**,** 20-31.

MOONEY, C. 1990. Cost and availability of healthy food choices in a London health district. *Journal of Human Nutrition and Dietetics,* 3**,** 111-120.

MURAKAMI, K., SASAKI, S., OKUBO, H., TAKAHASHI, Y., HOSOI, Y. & ITABASHI, M. 2007. Monetary costs of dietary energy reported by young Japanese women: association with food and nutrient intake and body mass index. *Public Health Nutrition,* 10**,** 1430-9.

MURAKAMI, K., SASAKI, S., TAKAHASHI, Y., OKUBO, H., HIROTA, N., NOTSU, A., FUKUI, M. & DATE, C. 2008. Comparability of weighed dietary records and a self-administered diet history questionnaire for estimating monetary cost of dietary energy. *Environ Health Insights,* 1**,** 35-43.

NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE 2012. Obesity: working with local communities.

OECD. 2012. Obesity update 2012. Available: www.oecd.org/health/49716427.pdf [Accessed Nov 2013].

OFFER, A., PECHEY, R. & ULIJASZEK, S. (eds.) 2012. *Insecurity, inequality, and obesity in affluent societies,* United States: Oxford University Press.

OFFICE FOR NATIONAL STATISTICS 2001. 2001 Key statistics.

OFFICE FOR NATIONAL STATISTICS. 2007. *Cancer Atlas of the UK and Ireland 1991-2000* [Online]. Available: http://www.statistics.gov.uk/statbase/product.asp?vlnk=14059 [Accessed June 2011].

OFFICE FOR NATIONAL STATISTICS 2009. Comparing lives using Rural-Urban classifications.

OFFICE FOR NATIONAL STATISTICS 2010. Family Spending - A report on the 2009 Living Costs and Food Survey. *In:* GILES HORSFIELD (ed.).

OFFICE FOR NATIONAL STATISTICS. 2011a. *RE: Consumer Price Indices 2011.*

OFFICE FOR NATIONAL STATISTICS. 2011b. *Regional Trends, No. 43, 2011 Edition* [Online]. Available: http://www.ons.gov.uk/ons/rel/regional-trends/regional-trends/no--43--2011-edition/rural-and-urban-areas--comparing-lives-using-rural-urban-classifications---supporting-data.xls

OFFICE FOR NATIONAL STATISTICS. 2012. Impact of the recession on household spending. Available: http://www.ons.gov.uk/ons/dcp171766_256980.pdf.

OFFICE FOR NATIONAL STATISTICS 2013. Family Spending 2012 - Chapter 5: Weekly household expenditure, an analysis of the regions of England and countries of the United Kingdom.

OLSEN, J. (ed.) 2009. *Principles in Health Economics,* Oxford.

OPENSHAW, S. 1984. Ecological Fallacies and the Analysis of Areal Census-Data. *Environment and Planning A,* 16**,** 17-31.

PEACOCK, J. & KERRY, S. 2010. *Presenting Medical Statistics from Proposal to Publication*, Oxford University Press.
PEARCE, J., HISCOCK, R., BLAKELY, T. & WITTEN, K. 2008. The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption. *Journal of Epidemiology and Community Health,* 62**,** 198-201.
PEARCE, J. & WITTEN, K. (eds.) 2010. *Geographies of Obesity,* UK: Ashgate publishing.
PEARCE, J., WITTEN, K., HISCOCK, R. & BLAKELY, T. 2007. Are socially disadvantaged neighbourhoods deprived of health-related community resources? *International Journal of Epidemiology,* 36**,** 348-55.
PETERSEN, J. H., GIBIN, M., MATEOS, P., ATKINSON, P. & ASHBY, D. 2010. Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems,* 13**,** 173-192.
PEVALIN, D. & ROBSON, K. 2009. *The Stata Survival Manual*, Open University Press.
PRENTICE, A. M. & JEBB, S. A. 1995. Obesity in Britain: gluttony or sloth? *British Medical Journal,* 311**,** 437-9.
PUBLIC HEALTH ENGLAND 2013a. Adult Obesity.
PUBLIC HEALTH ENGLAND. 2013b. *Health Survey for England* [Online]. Available: http://www.noo.org.uk/data_sources/adult/health_survey_for_england [Accessed September 2013].
RANSLEY, J. K., DONNELLY, J. K., BOTHAM, H., KHARA, T. N., GREENWOOD, D. C. & CADE, J. E. 2003. Use of supermarket receipts to estimate energy and fat content of food purchased by lean and overweight families. *Appetite,* 41**,** 141-148.
RAO, M., AFSHIN, A., SINGH, G. & MOZAFFARIAN, D. 2013. Do healthier foods or diet patterns cost more than less healthy options? A systematic review and meta-analysis. *Circulation,* 127 (12 Meeting Abstracts).
REHM, C. D., MONSIVAIS, P. & DREWNOWSKI, A. 2011. The quality and monetary value of diets consumed by adults in the United States. *American Journal of Clinical Nutrition,* 94**,** 1333-9.
REHM, C. D., MOUDON, A. V., HURVITZ, P. M. & DREWNOWSKI, A. 2012. Residential property values are associated with obesity among women in King County, WA, USA. *Social Science & Medicine,* 75**,** 491-5.
RIVA, M., CURTIS, S., GAUVIN, L. & FAGG, J. 2009. Unravelling the extent of inequalities in health across urban and rural areas: Evidence from a national sample in England. *Social Science & Medicine,* 68**,** 654-663.
ROBINSON, M., LOWEY, H., DEDMAN, D., TOCQUE, K., PERKINS, C. & BELLIS, M. 2010. Using geodemographic analysis to develop our understanding of inequalities in cancer incidence in the North West.
ROKHOLM, B., BAKER, J. L. & SORENSEN, T. I. 2010. The levelling off of the obesity epidemic since the year 1999--a review of evidence and perspectives. *Obes Rev,* 11**,** 835-46.
RYDEN, P. & HAGFORS, L. 2011. Diet cost, diet quality and socio-economic position: how are they related and what contributes to differences in diet costs? *Public Health Nutrition,* 24**,** 1-13.
SASI GROUP The National Classification of Census Output Areas.
SCARBOROUGH, P. & ALLENDER, S. 2008. The North-South gap in overweight and obesity in England. *British Journal of Nutrition,* 100**,** 677-684.
SCHRIJVERS, C. T., MACKENBACH, J. P., LUTZ, J. M., QUINN, M. J. & COLEMAN, M. P. 1995. Deprivation and survival from breast cancer. *British Journal of Cancer,* 72**,** 738-43.
SCHRODER, H., MARRUGAT, J. & COVAS, M. I. 2006. High monetary costs of dietary patterns associated with lower body mass index: a population-based study. *International Journal of Obesity,* 30**,** 1574-9.
SHAW, M., DORLING, D. & BRIMBLECOMBE, N. 1998. Changing the map: health in Britain 1951-91. *Sociology of Health & Illness,* 20**,** 694-709.

SHOHAIMI, S. 2004. 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 & Community Health,* 58**,** 686-691.

SLATER, M. D. & FLORA, J. A. 1991. Health lifestyles: audience segmentation analysis for public health interventions. *Health Educ Q,* 18**,** 221-33.

SMITH, D. M., CUMMINS, S., TAYLOR, M., DAWSON, J., MARSHALL, D., SPARKS, L. & ANDERSON, A. S. 2010. Neighbourhood food environment and area deprivation: spatial accessibility to grocery stores selling fresh fruit and vegetables in urban and rural settings. *International Journal of Epidemiology,* 39**,** 277-84.

SOFI, F., CESARI, F., ABBATE, R., GENSINI, G. & CASINI, A. 2008. Adherence to Mediterranean diet and health status: meta-analysis. *British Medical Journal,* 337.

SOOMAN, A., MACINTYRE, S. & ANDERSON, A. 1993. Scotland's health--a more difficult challenge for some? The price and availability of healthy foods in socially contrasting localities in the west of Scotland. *Health Bull (Edinb),* 51**,** 276-84.

SPENCE, M., CADE, J., BURLEY, V. & GREENWOOD, D. 2002. Ability of the UK Women's Cohort Study food frequency questionnaire to rank dietary intakes: a preliminary valdation study. *Proceedings of the Nutrition Society,* 61**,** 117A.

SPORT ENGLAND 2013. Active People Survey. National Obesity Observatory.

STATACORP 2009. Stata Statistical Software: Release 11. *In:* COLLEGE STATION, T. S. L. (ed.).

STATACORP 2012. Stata/IC 12.1. *In:* COLLEGE STATION, T. S. L. (ed.).

STEAD, M., HASTINGS, G. & MCDERMOTT, L. 2007. The meaning, effectiveness and future of social marketing. *Obesity Reviews,* 8 Suppl 1**,** 189-93.

SUN, B. 2005. Promotion Effect on Endogenous Consumption. *Marketing Science,* 24**,** 430-443.

SWINBURN, B., EGGER, G. & RAZA, F. 1999. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Preventive Medicine,* 29**,** 563-70.

SWINBURN, B., SACKS, G., VANDEVIJVERE, S., KUMANYIKA, S., LOBSTEIN, T., NEAL, B., BARQUERA, S., FRIEL, S., HAWKES, C., KELLY, B., L'ABBE, M., LEE, A., MA, J., MACMULLAN, J., MOHAN, S., MONTEIRO, C., RAYNER, M., SANDERS, D., SNOWDON, W. & WALKER, C. 2013. INFORMAS (International Network for Food and Obesity/non-communicable diseases Research, Monitoring and Action Support): overview and key principles. *Obesity Reviews,* 14 Suppl 1**,** 1-12.

SWINBURN, B. A., SACKS, G., HALL, K. D., MCPHERSON, K., FINEGOOD, D. T., MOODIE, M. L. & GORTMAKER, S. L. 2011. The global obesity pandemic: shaped by global drivers and local environments. *The Lancet,* 378**,** 804-814.

TAYLOR, E. F., BURLEY, V. J., GREENWOOD, D. C. & CADE, J. E. 2007. Meat consumption and risk of breast cancer in the UK Women's Cohort Study. *British Journal of Cancer,* 96**,** 1139-1146.

THOMPSON, C., CLARKE, G. P., CLARKE, M. & STILLWELL, J. 2012. Modelling the future opportunities for deep discount food retailing in the UK. *The International Review of Retail, Distribution and Consumer Research,* 22**,** 143-170.

THORNTON, L. E., BENTLEY, R. J. & KAVANAGH, A. M. 2011a. Individual and area-level socioeconomic associations with fast food purchasing. *Journal of Epidemiology and Community Health,* 65**,** 873-80.

THORNTON, L. E., CRAWFORD, D. A. & BALL, K. 2010. Neighbourhood-socioeconomic variation in women's diet: the role of nutrition environments. *European Journal of Clinical Nutrition,* 64**,** 1423-32.

THORNTON, L. E., CRAWFORD, D. A. & BALL, K. 2011b. Who is eating where? Findings from the SocioEconomic Status and Activity in Women (SESAW) study. *Public Health Nutrition,* 14**,** 523-31.

THREAPLETON, D. E., GREENWOOD, D. C., EVANS, C. E., CLEGHORN, C. L., NYKJAER, C., WOODHEAD, C., CADE, J. E., GALE, C. P. & BURLEY, V. J. 2013. Dietary Fiber Intake and Risk of First Stroke: A Systematic Review and Meta-Analysis. *Stroke*.

TIMMINS, K. A., MORRIS, M. A., HULME, C., EDWARDS, K. L., CLARKE, G. P. & CADE, J. E. 2013. Comparability of methods assigning monetary costs to diets: derivation from household till receipts versus cost database estimation using 4-day food diaries. *European Journal of Clinical Nutrition,* 67**,** 1072-1076.

TOMINTZ, M. N., CLARKE, G. P. & RIGBY, J. E. 2008. The geography of smoking in Leeds: estimating individual smoking rates and the implications for the location of stop smoking services. *Area,* 40**,** 341-353.

TOWNSEND, M. S., AARON, G. J., MONSIVAIS, P., KEIM, N. L. & DREWNOWSKI, A. 2009. Less-energy-dense diets of low-income women in California are associated with higher energy-adjusted diet costs. *American Journal of Clinical Nutrition,* 89**,** 1220-6.

TRICHOPOULOU, A. & VASILOPOULOU, E. 2000. Mediterranean diet and longevity. *British Journal of Nutrition,* 84 Suppl 2**,** S205-9.

TURRELL, G., BLAKELY, T., PATTERSON, C. & OLDENBURG, B. 2004. A multilevel analysis of socioeconomic (small area) differences in household food purchasing behaviour. *Journal of Epidemiology and Community Health,* 58**,** 208-15.

TURRELL, G. & KAVANAGH, A. 2005. Socio-economic pathways to diet: modelling the association between socio-economic position and food purchasing behaviour. *Public Health Nutrition,* 9**,** 375-383.

UK DATA SERVICE. 2013. *Census Geography* [Online]. Available: http://ukbsrv-at.edina.ac.uk/ukborders/action/restricted/index.

UNITED STATES DEPARTMENT OF AGRICULTURE 2010. Dietary Guidelines for Americans. *In:* PROMOTION, C. F. N. P. A. (ed.).

UNITED STATES DEPARTMENT OF AGRICULTURE. 2011. *Food Spending Adjustments During Recessionary Times* [Online]. Available: http://www.ers.usda.gov/amber-waves/2011-september/food-spending.aspx.

UNIVERSITY OF OTAGO 2013. Socioeconomic Deprivation Indexes: NZDep and NZiDep. *In:* UNIVERSITY OF OTAGO, W., DEPARTMENT OF PUBLIC HEALTH (ed.).

US BURDEN OF DISEASE COLLABORATORS 2013. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *The Journal of the American Medical Association,* 310**,** 591-608.

VICKERS, D., REES, P. & BIRKIN, M. 2005. *Creating the national classification of census output areas: data, methods and results* [Online]. University of Leeds. Available: http://www.geog.leeds.ac.uk/fileadmin/downloads/school/research/wpapers/05-2.pdf.

WAARD, F. D. 1975. Breast-Cancer Incidence and Nutritional-Status with Particular Reference to Body-Weight and Height. *Cancer Research,* 35**,** 3351-3356.

WALKER, R. E., KEANE, C. R. & BURKE, J. G. 2010. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & Place,* 16**,** 876-884.

WANG, Y. C., MCPHERSON, K., MARSH, T., GORTMAKER, S. L. & BROWN, M. 2011. Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet,* 378**,** 815-825.

WATERLANDER, W. E., DE HAAS, W. E., VAN AMSTEL, I., SCHUIT, A. J., TWISK, J. W., VISSER, M., SEIDELL, J. C. & STEENHUIS, I. H. 2010. Energy density, energy costs and income - how are they related? *Public Health Nutrition,* 13**,** 1599-608.

WCRF (ed.) 2007. *Food, nutrition and the prevention of cancer: a global perspective,* Washington DC.

WILKINSON, R. & PICKETT, K. 2010. *The Spirit Level,* London, Penguin group.

WILLETT, W. C. (ed.) 2013. *Nutritional Epidemiology*.

WOODHOUSE, A., CALVERT, C. & CADE, J. 1997. The UK Women's Cohort Study: background and obtaining local ethical approval. *Proceedings of the Nutrition Society,* 56**,** 64A.

WORLD HEALTH ORGANISATION. 2000. Obesity: Preventing and managing the Global Epidemic. *Report of a WHO consultation, WHO technical report series 894* [Online]. Available: http://libdoc.who.int/trs/WHO_TRS_894.pdf.

WORLD HEALTH ORGANISATION. 2006. *BMI classification* [Online]. Available: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html [Accessed November 2013].

WORLD HEALTH ORGANISATION. 2013. *Obesity and overweight fact sheet (number 311)* [Online]. Available: http://www.who.int/mediacentre/factsheets/fs311/en/ [Accessed September 2013].

WRIGLEY, N. 2002. 'Food deserts' in British cities: Policy context and research priorities. *Urban Studies,* 39**,** 2029-2040.

WRIGLEY, N., WARM, D., MARGETTS, B. & WHELAN, A. 2002. Assessing the Impact of Improved Retail Access on Diet in a 'Food Desert': A Preliminary Report. *Urban Studies,* 39**,** 2061-2082.

WYNDELS, K., DALLONGEVILLE, J., SIMON, C., BONGARD, V., WAGNER, A., RUIDAVETS, J. B., ARVEILER, D., FERRIERES, J., AMOUYEL, P. & DAUCHET, L. 2011. Regional factors interact with educational and income tax levels to influence food intake in France. *European Journal of Clinical Nutrition,* 65**,** 1067-75.