University of Southampton

Faculty of Environmental and Life Sciences

Geography and Environmental Science

Regional variations in diet in England

by

Monique Alayne Campbell

ORCID ID 0000-0002-6123-7170

Thesis for the degree of Doctor of Philosophy

[July 2021]

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

Geography and Environmental Science

Doctor of Philosophy

Regional variations in diet in England

by

Monique Alayne Campbell

Despite countless studies and interventions globally, sub-optimal diet continues to account for more death and disease than physical inactivity, excessive alcohol consumption and smoking combined. Although the relationship between unhealthy diets and chronic diseases is a well-established fact globally, very few studies have truly investigated possible differences in dietary patterns across and within the regions of England. Research in this area has tended to focus on childhood obesity, the individual determinants of dietary choice, geographical access to healthy food choices, and general public health promotion campaigns such as '5-A-Day'. However, rather less attention has been placed on how diets vary geographically, beyond generally negative assessments of the 'Scottish diet'. This project aimed to contribute to the literature by examining the extent to which fruit and vegetable consumption and dietary patterns vary at the regional and sub-regional level amongst persons aged 16-64 years in England.

Data from the National Diet and Nutrition Survey (NDNS), 2008-2016 and the Health Survey for England (HSE), 2008-2016 were used to explore whether 'regional diets' are evident within England and to uncover small area level variations in diet. Two-way cross-classified multi-level modelling was used to model and explore the effect of region on fruit and vegetable consumption. Principal Components Analysis (PCA) was used to derive dietary patterns at the national level and for each of the nine regions of England. Thereafter, two-way cross-classified multi-level modelling was used to examine the effect of region on dietary patterns identified within the national sample. Small area microsimulation, specifically the Iterative Proportional Fitting (IPF) method, was used to explore possible variations in fruit and vegetable consumption across populations in Middle Layer Super Output Areas (MSOAs) in England. Overall, this project found no statistically significant differences in fruit and vegetable consumption and dietary patterns across the regions of England, but more variation across smaller areas (MSOAs), particularly in and around urban centres within the North of England and Midlands. It is expected that the results obtained from this project will contribute to ongoing research by enhancing the understanding of the dietary challenges facing the UK and more specifically, England. The findings presented could help to inform social and health policy as well as more localised interventions (such as voucher schemes and subsidies) aimed at improving the diets of individuals, especially among more disadvantaged groups.

Table of Contents

Table of Tables

Table of Figures

Table of Equations

Research Thesis: Declaration of Authorship

Print name: MONIQUE ALAYNE CAMPBELL

Title of thesis: Regional variations in diet in England

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Campbell, M., Smith, D., Baird, J., Vogel, C. and Moon, E. G. (2020) 'A critical review of diet-related surveys in England, 1970-2018', *Archives of Public Health*, 78(1), pp. 66. doi:10.1186/s13690-020-00447-6

Smith, D. M., Vogel, C., Campbell, M., Alwan, N. and Moon, G. (2021) 'Adult diet in England: Where is more support needed to achieve dietary recommendations?', *PLOS ONE,* 16(6), pp. e0252877. doi:10.1371/journal.pone.0252877

Signature: ....Date: July 14, 2021...........................

Acknowledgements

I thank God for strength, endurance and for all the persons who were a part of this journey. I especially want to thank my amazing Supervisory Team. Dianna, for being there from literally day one and who went beyond the call of duty to assist me in every way possible. Graham, for his words of wisdom, encouragement and his commitment to the process. Janis, for her encouraging words and support. Christina, for her support and willingness to assist where possible. My family, especially my mother, who was constantly motivating me, despite the thousands of miles. Last but definitely not least, I would like to thank all the Administrative staff, including Julie, Glenn and all the staff from the SCDTP.

Definitions and Abbreviations

Adult .... A person belonging to the traditional working age group, defined by the Office of National Statistics (ONS) as persons aged 16-64 years (ONS, 2017a)

Chapter 1 Introduction

"Fish and chips", "Sausage and mash", "Toad in the Hole", "the traditional English breakfast", "meat and two veg" and "Sunday roast" are dishes generally associated with the traditional English diet. However, from as early as the 19th century there have been noted dietary and health differences between the regions of England (Bambra, 2016). Evidence of this can be seen in regionally distinct foods such as the Yorkshire pudding from the Yorkshire and Humber region, Chicken Parmo, synonymous with the North East region, jellied eels associated with London and the Cornish pasty, traditionally aligned to Cornwall in the South West region (Vogler, 2020). Despite the existence of these and other traditional "regional foods", studies have shown that the English diet has transitioned from a diet of mostly meat pies, cooked potatoes and green vegetables (Riley, 2010), to a more "Western" type dietary pattern, high in sugar sweetened beverages (SSBs), saturated fats, free sugars, salt and ultra-processed foods and low in fruits and vegetables (Newton et al, 2015). Major historical events such as World War II, rationing, changes in domestic food technology (e.g. refrigeration), deindustrialisation and post-colonial migration have helped to shape the current English diet (Otter, 2012), to the extent that dishes such as "Chicken Tikka Masala" have been incorporated and acknowledged as being uniquely English (Buettner, 2008).

In light of the drastic change in global dietary patterns (Popkin, 1993), the World Health Organisation (WHO, 2015a) has recommended the adoption of a healthy diet, high in fresh fruits, vegetables and legumes, and low in saturated fats, salt and free sugars. Several studies, interventions, strategies and initiatives have been carried out over the years, all with the aim of encouraging the maintenance of a healthy diet to protect against malnutrition and reduce the risk of overweight, obesity and related chronic diseases such as cancer and diabetes (Afshin et al, 2019; Gakidou et al, 2017). Despite these attempts, sub-optimal diet continues to be the leading cause of rising obesity rates, death and disability in England (Jebb, 2018; Newton et al, 2015). Although physical inactivity, excessive alcohol consumption and smoking have been acknowledged as risk factors for obesity and other diet-related chronic diseases, sub-optimal diet accounts for approximately 10.8% of the overall burden of disease in England (Newton et al, 2015). In 2017, only 29% of adults in England consumed the recommended daily five portions of fruits and vegetables (referred to as 5-A-Day) (NHS Digital, 2019). Adults in England consume approximately 3.8 portions of fruits and vegetables daily, which is 1.2 portions below the recommended five portions. It has been estimated that more than 55 million additional portions of fruits and vegetables would need to be consumed daily in England if all adults were to meet the national 5-A-Day target (Pinho-Gomes, Knight, Critchley and Pennington, 2021). Moreover, sub-optimal fruit and vegetable consumption accounts for approximately 16,000 deaths in England (Pinho-Gomes, Knight, Critchley and Pennington, 2021) and diet-related ill-health is estimated to cost the National Health Service (NHS) approximately £5.8 billion annually (Scarborough et al, 2011).

It has been widely acknowledged that there are differences in diet according to age, sex and socio-economic position (SEP) which result in disparities in obesity and related chronic diseases, which translate to inequalities in health (Tiffin and Salois, 2012). Far less attention has been placed on how dietary patterns vary across the regions and smaller areas of England, beyond "what everyone knows" (Macintyre, Maciver and Sooman, 1993 p. 230). To date, most academic and non-academic diet-related studies undertaken in the UK tend to focus either on diet as a risk factor for health outcomes such as obesity and cardiovascular diseases or the prevalence of childhood obesity (Scarborough, Morgan, Webster and Rayner, 2011). Studies which explore diet (and not obesity), tend not to disaggregate data by English region but instead focus on the UK as whole or make generally negative assessments of the Scottish diet in comparison to the English diet (Walsh et al, 2016; Wrieden et al, 2015). In addition, the few studies (Morris et al, 2016; Billson, Pryer and Nichols, 1999; Whichelow, Erzinclioglu and Cox, 1991) which have attempted to examine dietary patterns in England, (some of which were conducted over 20 years ago) were: i) not representative of the English population, ii) restricted to certain groups/cohorts in the population (e.g. middle-aged white women) or iii) reliant on past observations such as the stereotypical North and South divide in England.

With respect to the North-South divide, worse health outcomes (such as lower life expectancy) and higher inequalities have been observed in the North of England compared to the remaining regions of England, particularly the South of England (Corris et al, 2020; Newton et al, 2015). Moreover, there is evidence to suggest that the gap between the regions, especially between the North East and the rest of England is widening, with persons in the North East more likely to spend a larger proportion of their shorter lifespan in poor health and more likely to die prematurely from largely preventable diseases (Corris et al, 2020). Whilst it is widely accepted that there are traditional diet and health differences between the regions, particularly the North and South of England (Bambra, 2016), very few empirical studies have systematically assessed the overall geography of diets in England specifically. Thus, the current English diet and possible differences at the regional and small area level remains a relatively unexplored topic.

Following the passing of the Health and Social Care Act (2012), local authorities gained ultimate responsibility for public health and increased oversight for the achievement of national diet-related targets such as the 5-A-Day target (defined as the daily consumption of five or more portions of fruits and vegetables). Diet-related survey data are therefore a major source of information used by researchers and local authorities to monitor and assess dietary patterns, evaluate the success or failure of interventions and identify potential inequalities. However, despite the availability of several large-scale repeated cross-sectional and longitudinal diet-related surveys, a major challenge is that most of the diet-related surveys conducted in England are not powered to capture more local variation (Pryce et al, 2020). Additionally, sample size issues, the different area boundaries used by data collectors and the different ways in which data are captured or formatted are some of the other noted data challenges which have hampered the assessment of dietary patterns in England and made it more difficult for local authorities/policymakers to plan and identify the areas (populations) most in need of dietary change and intervention (Brug, 2008; Campbell et al, 2020; Macintyre, Maciver and Sooman, 1993; Moretti and Whitworth, 2021; Pryce et al, 2020; Whitworth, 2013). Thus, inequalities in diet and health in England could go undetected due to the poor availability of data, especially at the small area level (Twigg, Moon and Jones, 2000). These findings further highlight the urgent need for more geographic studies which explore the foods consumed by persons in England and whether there are variations in dietary patterns regionally and/or across the smaller areas of England.

This project aims to add to the current literature by exploring potential regional and small area differences in dietary patterns in England amongst the traditional working age adult population (persons aged 16-64 years). Gaining insight into the types of foods consumed by persons in England will provide interested researchers and policymakers with a more updated and empirical snapshot of aspects of the "English diet" and the possible existence of region/area-specific dietary patterns (or none whatsoever), beyond traditionally held stereotypes. This type of information could assist local authorities in identifying the regions or small areas most in need of dietary change, prioritisation and more targeted interventions. Policies/interventions that do not acknowledge or deviate greatly from current consumption patterns are less likely to succeed at the population level (De Ridder et al, 2017; Fulgoni and Drewnowski, 2019). Moreover, it is widely acknowledged that the characteristics of individuals such as age, sex, ethnicity and SEP are key (compositional) factors associated with diet. However, this project will highlight the significance of acknowledging "place" (context) or geography in the identification of potential diet-related inequalities in England, which as far as is known, has not been fully explored previously. The project will make the best use of current diet-related secondary survey data (despite the many noted challenges) and demonstrate how a combination of quantitative statistical techniques can be utilised to assess dietary patterns in England and provide researchers and policymakers with useful information which may not be readily available otherwise.

Chapter 1

The current project has been organised into seven main chapters. The following chapter (Chapter 2) presents a review of the current literature, highlights the major gaps/themes identified and outlines the three main research questions which emerged from the review. Chapter 3 provides a detailed assessment (data landscape) of the major repeated cross-sectional and longitudinal surveys conducted in the UK over the last 48 years, their major advantages and disadvantages and the surveys assessed as being the most suitable for further analysis in the current project. Chapters 4, 5 and 6 are the standalone analytical chapters pertaining to the project's three main research questions. Finally, Chapter 7 provides an overarching summary of the project and its major findings, the overall strengths and limitations as well as possible policy and research implications. References cited, a glossary of terms and other supplemental results are presented after Chapter 7.

Chapter 2 Literature Review

This chapter has been organised into eight sections and begins by discussing the generally accepted characteristics of a healthy diet, the issue of sub-optimal diets globally and the major determinants of diet. Literature pertaining to global transitions and variations in diet is then reviewed, followed by the discussion of the Scottish and English diet, with the focus thereafter being on regional differences in diets in England. The chapter concludes with a summary of the limitations of existing literature, the major gaps the project will attempt to fill and the three main research questions which emerged from the literature review.

2.1 What is a healthy diet?

According to the WHO (2015a), a healthy diet can be achieved by: i) consuming no more than five grams of salt per day, ii) reducing free sugars and total fat (especially saturated fat) to less than 5% and 30% of total daily energy intake (respectively), iii) consuming at least 400 grams or five portions of fruits and vegetables daily (not including potatoes, sweet potatoes, cassava and other starchy roots) and iv) incorporating nuts, legumes and whole grains. Particular attention tends to be placed on the daily consumption of fruits and vegetables (also known as the "5-A-Day" target), due to their high fibre, vitamin and mineral content, low energy density and their association with the reduced risk of chronic diseases (Dowler, 2008; Hawkesworth et al, 2010). It is therefore not strange to find a "healthy diet" in literature and policy being generally limited to, or defined, as a diet high in fruits and vegetables (Roos et al, 2000).

There is still no global consensus on diet and the optimal number of fruits and vegetables to be consumed daily (Oyebode, Gordon-Dseagu, Walker and Mindell, 2014). Countries such as Canada, Australia and Denmark recommend the consumption of at least six to ten portions per day, whereas England/the UK recommends the consumption of five portions daily or 5-A-Day (Pem and Jeewon, 2015). Despite these inter-country differences, the WHO highly recommends the monitoring of fruit and vegetable consumption or the "5-A-Day" target as it is thought to be the most convenient, easy, cost-effective and standardised means by which inter and intra-country comparisons of dietary consumption can be made over time (Agudo, 2005). The WHO also recommends adherence to overall dietary guidelines, as this should translate to what is known as a healthy or high-quality diet. Such a diet eliminates hunger, reduces all forms of malnutrition and reduces the risk of diet-related chronic diseases such as heart disease and stroke (Global Panel on Agriculture and Food Systems for Nutrition, 2016).

Although there is still no universally accepted definition, index or measure of diet and diet quality (Global Panel on Agriculture and Food Systems for Nutrition, 2016), WHO recommendations are generally accepted and used by member states to develop national dietary guidelines and policy targets (WHO, 2015a). These recommendations have resulted in the development of policy tools and interventions such as England's Eatwell Guide and national "5-A-Day" campaign which both seek to highlight dietary guidelines/targets, visually and otherwise, and promote dietary change (especially fruit and vegetable consumption) at the population level.

2.2 The significance of diet and the Global Burden of Disease

Besides WHO recommendations, several studies, interventions and initiatives have been carried out globally, all with the aim of encouraging the maintenance of a healthy diet. It is well established that good dietary practices throughout the lifecourse can significantly reduce the risk of overweight/obesity and related chronic diseases, as well as improve the overall health status of individuals (Walsh et al, 2016). However, despite countless studies and interventions globally, sub-optimal diet continues to account for more death and disease than physical inactivity, excessive alcohol consumption and smoking combined (Afshin et al, 2019; BMA, 2016; Forouzanfar et al, 2015; Malhotra, Noakes and Phinney, 2015).

The Global Burden of Disease (GBD) project, established in 1990, is currently the most comprehensive, annual, peer-reviewed programme, which investigates the burden of disease attributable to diet across 195 countries worldwide. In 2016, poor/sub-optimal diet was the second highest ranked risk factor for deaths worldwide, accounting for 18.8% of deaths (Gakidou et al, 2017). Similarly, the 2017 wave of the study found dietary risk factors to be responsible for approximately 11 million deaths (amongst adults) worldwide, with cardiovascular disease being the leading cause (Afshin et al, 2019). Approximately 45% of diet-related deaths occurred amongst adults less than 70 years old and it was further estimated that one in every five deaths worldwide could be prevented with improvements to diet. Diets high in sodium, low in whole grains, low in fruits, vegetables, nuts and seeds and those low in omega-3 fatty acids were the leading dietary risk factors, each accounting for more than 2% of global deaths. These findings demonstrate the increasing prevalence of sub-optimal diet, the net effect on death and disability, and the urgent need for the improvement in diets worldwide (Afshin et al, 2019; Hawkes, Harris and Gillespie, 2017).

2.3 Factors associated with diet consumption

Undoubtedly, hunger, satiety, personal taste, culture and physiological/biological factors (e.g. hormones, genetics), all play a part in food choice and the overall quality of a person's diet. However, diets vary and are influenced by a myriad of factors at the individual level (compositional factors) as well as at the broader socio-cultural and environmental level (contextual factors) (Shepherd, 1999).

2.3.1 Compositional Factors

Compositional factors refer to the socio-demographic characteristics of individuals within populations such as their age, sex, ethnicity, health status, nutritional knowledge and their individual socio-economic position (SEP), primarily measured by income, education, and occupation (Berzofsky et al, 2014). Although no definite or significant relationship has been found between ethnicity, diet and related illnesses such as obesity in countries such as the UK (Higgins, Nazroo and Brown, 2019; National Obesity Observatory, 2011), significant associations have been found between diet, age, sex and SEP (Imamura et al, 2015; Roberts, Cavill, Hancock and Rutter, 2013; Tiffin and Salois, 2012).

With regard to age, younger adults generally consume less healthy foods compared to older adults (Imamura et al, 2015). The 16-24 age group has been highlighted as the least likely to consume the recommended five portions of fruits and vegetables daily (NHS Digital, 2017). Thus, there is an age gradient in diet, with older adults more likely to have better dietary patterns compared to younger adults. In terms of sex, women tend to have a healthier diet than men by their consumption of foods such as semi-skimmed milk, wholemeal bread and fruits and vegetables at least once per day (Acheson, 1998; Imamura et al, 2015).

Regardless of age and sex, it has been widely acknowledged that a social gradient exists in diet, with persons from low SEP groups having comparatively worse diets and health than those in higher SEP groups (Acheson, 1998; Roberts, Cavill, Hancock and Rutter, 2013; Tiffin and Salois, 2012). Those within lower SEP groups tend to eat fewer fruits, vegetables and high fibre foods and more table sugar and ultra-processed foods such as soft drinks, pizza and processed meats, compared to those in higher SEP groups (Acheson, 1998; Roos et al, 2000; Tiffin and Salois, 2012).

The cost of food has been noted to be an important mediating factor between SEP and diet (Darmon and Drewnowski, 2008; Monsivais, Aggarwal and Drewnowski, 2012). Data from the 2003-2005 Low Income Diet and Nutrition Survey showed that 36% of respondents from low-income households reportedly could not afford to maintain a healthy diet (Roberts, Cavill, Hancock and Rutter, 2013). Whilst incomes have remained fairly stable since the 2007-2008 financial crisis, the minimum cost to achieve a relatively healthy diet or one aligned with dietary guidelines, has increased in both real terms and as a proportion of household income (O'Connell et al, 2018). The additional cost of food may seem negligible to some persons with higher income but can be a very high proportion of household income for lower SEP persons. Persons from low-income households or those with low levels of education, though possibly motivated to adopt healthier diets, may not have the financial resources to consistently purchase and consume nutritious foods. Therefore, research, interventions and initiatives which promote healthy diets but fail to consider demographic characteristics as well as cost (actual or perceived) and/or individual level SEP variables such as income, are not realistic and are likely to be unsuccessful (Darmon and Drewnowski, 2008). Likewise, interventions (such as the UK's Change4Life, 5-A-Day campaign) which focus on the distribution of information and tools to help people to make better lifestyle choices, including diet, are valuable. However, these types of information-based interventions (often referred to as downstream or high agency interventions) require that individuals be motivated to engage with the information received and be able to use their personal resources to make the recommended changes (Adams, Mytton, White and Monsivais, 2016). Persons with higher levels of education may find it easier to understand and relate to the information provided via public health messages (Maguire and Monsivais, 2015). In addition, persons within high income groups may have more time to invest in preparing high quality meals and are more likely to have the material resources to purchase recommended healthy foods (Adams, Mytton, White and Monsivais, 2016).

2.3.2 Contextual factors

Contextual factors relate to the conditions in which individuals live and work, their access to food and essential services and the wider physical/structural, socio-economic, cultural and political factors which impact on diet. The Dahlgren Whitehead Model (Figure 2.1) best illustrates this concept by presenting contextual and compositional factors as layers of influence on health; collectively termed the social determinants of health (Bambra et al, 2010; Dahlgren and Whitehead, 1991).

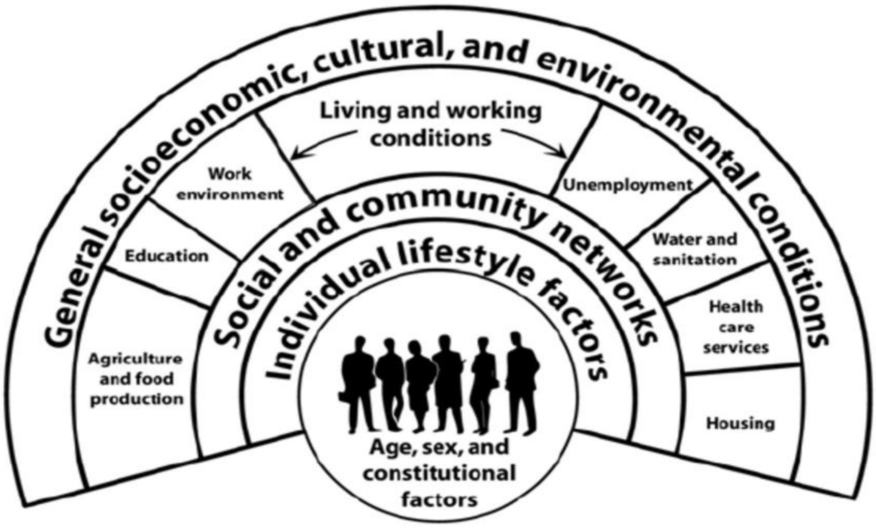


Figure 2.1: The social determinants of health, Dahlgren and Whitehead, 1991

The placement of individuals at the heart of the model recognises that previously discussed compositional factors such as age, sex, ethnicity, health status and individual SEP, significantly influence diet. However, higher level or "upstream" contextual factors (such as physical environment, geographic location, access to food, neighbourhood deprivation, cultural and environmental conditions) represented in layers 3 and 4, also have a tremendous influence on an individual's ability to maintain a healthy diet and lifestyle (Acheson, 1998). The model highlights the significance of contextual factors, which if not considered in conjunction with the former, can introduce inequalities in diet and health or widen existing ones (Bambra, 2016; Mozaffarian, Angell, Lang and Rivera, 2018).

Diet-related studies have the tendency to focus more on compositional factors, which emphasise who individuals are, rather than where they live (contextual factors), in relation to their dietary quality (Brug, 2008). Although geographic differences/inequalities in diet have been observed across countries worldwide, less is known about how diets vary within countries (Imamura et al 2015). Recent studies and interventions which consider within country geographic differences in diet, tend to focus on the physical food environment; particularly food access, cost and availability and their potential influence on the diets of individuals living within these areas (Williamson et al, 2017). For instance, the rise in the number of fast-food outlets in the UK over the last decade, notably in deprived areas and in close proximity to schools, has been linked to the increasingly poor diet and health outcomes observed amongst low-income persons (Blow, Gregg, Davies and Patel, 2019; Maguire, Burgoine, and Monsivais, 2015; PHE, 2018c; Saunders, Saunders and Middleton, 2015; Tinkler and Ritchie, 2016). These general assumptions have led to the development of concepts such as "obesogenic environments"-defined as the "sum of influences that the surroundings, opportunities, or conditions of life have on promoting obesity in individuals or populations" (Swinburn, Egger and Raza, 1999, p. 564). One challenge in this area of contextual food environment research is the over-reliance on neighbourhood food outlet exposures, due to data constraints. In reality, people can spend vast amounts of time away from their residential neighbourhood and there is increasing recognition that the activity spaces people experience across each day should be taken into account when assessing the potential impact of context on diet (Perchoux, Chaix, Cummins and Kestens, 2013). In addition, contextual food environment research/interventions which focus on the school environment assume that children have limited mobility, spend the majority of their time at school, and are only influenced by the school environment. Studies have found that parents exert much control over their young children's behaviour and therefore have a stronger influence on children's food choices than outside influences such as their school or peers (European Food Information Council, 2012; Scaglioni et al, 2018). It has been found that once children begin school, most would have already developed their food preferences, which makes behaviour change even more difficult at this stage (Savage, Fisher and Birch, 2007).

Although there is increasing literature related to the contextual food environment and the dietary pattern of children globally (Black, Moon and Baird, 2014; Narzisi and Simons, 2021), much less is known about how diets vary geographically within the adult population (Imamura et al 2015; Smith et al, 2021). Such findings highlight the need for more studies which explore: i) the diet of adults, who may have more of an influence on the diet and health of the younger generation, especially at the early stages of life and ii) possible geographic differences in dietary patterns (Macintyre, Maciver and Sooman, 1993; Musingarimi, 2009).

Besides fast food availability and obesogenic environments previously discussed, the existence of "food deserts" is another widely discussed topic within food environment research. Although no universal definition exists (Widener, 2018), food deserts can be defined as areas in which fresh, healthy and affordable food is limited or non-existent. The concept focuses on the accessibility (physical access), availability (diversity of food choices) and affordability (cost) of foods, primarily fruits and vegetables, in the food retail economy of a particular neighbourhood (Cummins, 2014). Popularization of the term ensued after researchers in the United States of America (USA) observed that healthy foods were limited in supply and more expensive in retail stores in deprived areas than in less deprived areas (Cummins, 2014; Williamson et al, 2017). Similarly, in the UK, "the Independent Inquiry into Inequalities in Health" (also known as the "Acheson Report"), published in 1998, highlighted the differential and worsening access to food experienced by persons in lower SEP groups (Acheson, 1998) and its contribution to diet and health inequalities (Cummins, 2014). The report recommended that households in poorer neighbourhoods receive better access to retail services. This sparked several heated debates about social exclusion and the high cost and limited healthy food options available to persons in low-income areas. These recommendations eventually resulted in the acceptance of “food deserts” without any evidence or research to support their existence (Cummins and Macintyre, 2002; Wrigley, 2002).

Evidence for the existence of food deserts was primarily based on general assumptions or small-scale observational studies which failed to systematically assess the diet and socio-demographic characteristics (age, sex, SEP) of persons living within these areas (Cummins and Macintyre, 2002; Cummins, 2014). Studies, (mostly conducted in the USA) have investigated how access to a new neighbourhood supermarket affects the diet of individuals living in food deserts. Researchers in the USA found that increased access to supermarkets significantly improved the consumption of fruits and vegetables and the overall diet of low-income and minority groups living in deprived areas (Cummins, 2014). On the contrary, results from two small-scale studies carried out in Leeds, England and Glasgow, Scotland, over a decade ago, showed mixed results; with supermarket access having little or no effect on fruit and vegetable consumption of residents living in deprived areas (Cummins et al, 2005; Wrigley, Warm and Margetts, 2003). More recent studies in the UK (though not nationally representative), also found that the availability, cost, quality and the variety of healthy foods was more related to the size and type of food retail store, rather than neighbourhood deprivation alone (Black et al, 2012; Williamson et al, 2017). These findings were inconsistent to those observed in the USA and thus illustrate that the relationship between area deprivation, the physical food environment and dietary patterns is complex and likely to be influenced by a myriad of inter-related demographic, social, cultural, structural and socio-economic factors (Roberts, Cavill, Hancock, and Rutter, 2013).

This project focuses on regional/geographic variations in diet in England. As such, the effect of retail food store/supermarket access on the diets of individuals living in deprived areas is not within the project’s immediate scope. The findings presented in this review simply highlight that researchers and policymakers are acknowledging the wider contextual factors and their contribution to diet and inequalities. Although laudable, critics, especially those outside of the USA, have called for the term “food deserts” to be officially retired (Widener, 2018), as it fails to recognise the characteristics of individuals (such as age, sex, individual SEP, household characteristics, transportation options, time availability) living within deprived areas and the combined influence both compositional and contextual factors have on food choice, diet and associated inequalities (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Macintyre, Maciver and Sooman, 1993; Widener, 2018). On the other hand, it has been suggested that dietary patterns and inequalities in diet and health, could have more to do with who persons are (compositional factors) versus where they reside (contextual factors) (Bambra, 2016). This could explain the tendency for health researchers to either focus solely on compositional factors (Brug, 2008) or rely heavily on past assumptions regarding the physical food environment and the existence of food deserts, without systematically assessing the diets of individuals and possible geographic differences (Cummins and Macintyre, 2002).

2.3.3 Further considerations in health and place research

Section 2.3.1 and 2.3.2 highlighted that "place"/context and the characteristics of individuals (composition) play a significant role as it relates to diet, health and related inequalities, as evidenced by the numerous studies conducted globally over the years. However, there is a tendency for empirical research to rely on more conventional definitions/representations of space and place, which in turn has the potential to create a false dualism of context and composition (Twigg, 2014). Instead, it should be recognised that there is a mutually reinforcing and reciprocal relationship between people and place, which suggests that people cannot be separated from place (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Diez-Roux, 2002). Public health geographers have been urged to be more contextually sensitive and consider adopting a more "collective" or "relational" view/approach (rather than the traditional/conventional view) to health and place-based research. Section 2.3.3.1-2.3.3.2 will elaborate on these considerations by outlining the key theoretical concepts associated with the collective and relational geographies view/approach (respectively) to health and place-based research. Section 2.3.3.3-2.3.3.4 then goes on to present some of the key considerations (mainly the importance of acknowledging geographic scale in the analysis of diet and overall health) and the associated data-related/methodological challenges identified within the existing literature (respectively).

2.3.3.1 The Collective Lifestyles approach to health and place

It has been widely acknowledged that studies which explore context are an improvement upon what would have otherwise been an individualistic perspective of diet, health and associated inequalities. However, it has been suggested that traditional definitions/views and the classic methods used to study context are usually limited in that they: i) typically do not seek to answer the question of how or why place influences diet and health, ii) typically oversimplify composition and context by viewing them as distinct or mutually exclusive and iii) tend to analyse what geographers consider to be geographic "space"(an objectively defined, relatively static/fixed location or area with geographic boundaries drawn at a specific scale e.g. countries, regions) instead of "place" (an objectively and subjectively defined location which is dynamic, fluid, multi-scale and described variably by different individuals and groups) (Bernard et al, 2007; Cummins, Curtis, Diez-Roux and Macintyre, 2007; Frohlich, Corin and Potvin, 2001). Macintyre, Ellaway and

Cummins (2002) highlighted the tendency for context to be treated as a "black box" or miasma of unspecified influences, which although known to have an effect on diet and health, remain relatively unknown and underexplored. Diez-Roux (2002) noted the limited studies which have been able to fully examine the actual processes which connect the specific characteristics of neighbourhoods with the health of individuals residing these areas. It is further argued that there is a tendency for studies to assume that "place" can be ignored once individual-level factors have been accounted for (Diez-Roux 2002). Moreover, it has been suggested that the seeming reliance on the more conventional approach to space and place could help to explain why most quantitative studies find the proportion of variation attributable to context to be smaller in comparison to individual-level risk factors (Pickett and Pearl, 2001).

In an attempt to improve on these noted limitations, Frohlich and Potvin (1999) recommended the adoption of what has been termed a collective lifestyles approach. This approach seeks to merge context with behaviour by considering the relationship between the social conditions (factors associated with an individual's relationship to others) of individuals and their behavioural patterns. Collective lifestyles can be thought of as a shared way of relating and acting within a given environment. It takes into account the social, environmental and cultural environments in which people live and operate. It is a concept which: i) acknowledges that context is created by relationships with people, ii) considers context to be reflective of both place and the characteristics of individuals within the place, iii) acknowledges that there is a relationship between agency (the ability for people to deploy a range of causal powers), practices (activities carried out) and social structure (societal rules and resources) and iv) recognises that the relationship between people and place is recursive and influences diet, health and inequalities (Frohlich, Corin and Potvin, 2001).

Having endorsed the collective lifestyles approach, Macintyre, Ellaway and Cummins (2002) attempted to unpack the "black box" of area effects through the development of a conceptual framework. This framework consisted of five categories of features/measures which can be used to gain insight into how the local physical and social environment is likely to influence diet and health directly or indirectly. These five categories included: i) The physical features of the shared environment such as air and water quality, ii) The availability of healthy environments at home, work and play e.g., the built environment (man-made structures created by people for people), quality of housing, the home food environment, safe play areas, non-hazardous working environments, parks, recreational areas etc, iii) The provision of public or private services to support people in their daily lives e.g. education, transportation, street cleaning, lighting, health and welfare services, iv) Socio-cultural features of a neighbourhood e.g. community norms and values, the political, economic, religious and ethnic history of a community, level of crime, level of community integration, community support networks, social capital and v) The reputation of an area e.g. how areas are perceived by residents, service/amenity planners and providers, investors etc. The first three categories refer to material or infrastructural resources and the remainder relate to collective social functioning and practices.

Although a useful tool, the conceptual framework suggested by Macintyre, Ellaway and Cummins (2002) is not exhaustive and somewhat limited, because it does not detail which aspects within each of the five domains need to be studied to understand the exact features of the physical and social environment most likely to influence diet and health. However, the lack of specificity within the framework is understandable, given that the features selected for analysis and inclusion in any given study will vary based on the pathways researchers hypothesise as linking area to diet/health. Thus, researchers are not expected to assess every aspect of the physical and social context which could possibly influence diet/health but rather to test the key features they theorise to be the most relevant (Macintyre, Ellaway and Cummins, 2002). For instance, studies interested in diet and related outcomes could consider exploring contextual features of physical and social environment such as aspects of the built environment, social capital, local norms and values, the price and availability of healthy and unhealthy foods and the location of takeaway outlets, among others. Diez-Roux and Mair (2010) highlighted social norms, social stressors and social cohesion/social capital to be some of the features of the neighbourhood social environment likely to influence health. Stafford et al (2007) explored the inter-relationships between ten selected environmental domains and their relationship with diet-related outcomes such as obesity and body fat distribution. The ten domains included in that study were: Crime, Policing, Physical dereliction, Supermarkets, Fast-food outlets, Leisure centres, Physical activity, Social disorder, Urban sprawl (measured as population density per square km) and High street services. Each domain included at least one indicator which measured specific features of the local physical and social environment. For example, the number of McDonalds' restaurants in areas was the sole indicator within the "Fast-food outlets" domain, whereas the perception (score) of safety walking alone in the area after dark was one of five indicators included within the "Social Disorder" domain. Overall, the study found neighbourhood disorder and access to local high street facilities (local shops, health-related stores typically found in a small UK town and financial services) to be the aspects of the local environment most significantly associated with obesity. Additionally, the study found social capital (the network of relationships between persons who live, work and operate in a particular place) to be a potential mediating factor, linking aspects of the built environment and public services to obesity. Although noting major limitations such as the poor availability of local level data and the study's use of cross-sectional (instead of longitudinal) data, the findings from Stafford et al (2007) can be used as a guide for future studies (e.g. features/measures of the local social and physical environment to consider) given that it was one of the first and largest studies which acknowledged the collective lifestyles approach and attempted to model direct pathways which linked the local residential environment to diet-related outcomes such as obesity.

2.3.3.2 Relational views of health and place

In keeping with the collective lifestyles approach, Cummins, Curtis, Diez-Roux and Macintyre (2007) also called for researchers to move way from conventional views of "place" and to adopt a more "relational" view which recognises place as nodes in networks which are dynamic, fluid, multi-scale and which change with time. Similar to the collective lifestyles approach, the relational view recognises that the characteristics of individuals (compositional factors) and the contexts (places) in which they live and operate daily and over the lifecourse, are tightly inter-related, complex, vary over time and space and operate at various geographic scales (from the global, national, regional to the local geographic level) (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Dearden, Lloyd and Green, 2020; Frohlich, Corin and Potvin, 2001). The relational view also acknowledges that individuals and their attitudes are likely to be influenced by different types of people and several "contextual" environments (household/family, social, school, neighbourhood, workplace, leisure), all operating simultaneously (Twigg, 2014). In order to gain a deeper understanding of how place affects diet and overall population health, researchers have been strongly urged (as far as possible) to adopt a relational or collective view/approach (Bernard et al, 2007; Clary, Matthews and Kestens, 2017).

The incorporation of a relational view into health and place related research can be achieved in three basic steps (Cummins, Curtis, Diez-Roux and Macintyre, 2007). Firstly, (as previously mentioned in Section 2.3.3.1) compositional and contextual factors must not be viewed as mutually exclusive and/or competing explanations for health inequalities. Although oftentimes done unintentionally, the conventional viewing of context and composition as distinct or mutually exclusive is problematic because it fails to recognise that the characteristics of individuals are themselves shaped by the characteristics of their locality (Macintyre, Ellaway and Cummins, 2002). Dearden, Lloyd and Catney (2019) noted that individuals and their diet/health help to shape and are shaped by the places they live and inhabit regularly. This is partially because individuals with similar socio-demographic characteristics tend to cluster in space (e.g. ethnic minority areas) and because individuals in the same area typically share common contextual influences (Dearden, Lloyd and Catney, 2019, Diez-Roux, 2001). For instance, an individual's social class (measured by occupation/occupational grade) is likely to be influenced by the local labour market and similarly, educational attainment is likely to be shaped by local school standards.

Individuals living in ethnic minority areas (sometimes referred to as ethnic enclaves or areas in which a particular ethnic group(s) is spatially clustered) may have dietary habits similar to that of the general population but also unique to their ethnic (sub) group (Stephenson, Herring and Olshan, 2020). Studies have also shown that areas with higher levels of deprivation and those which are more ethnically diverse tend to have greater access to takeaway (fast food) outlets than more affluent white areas (Black, Moon and Baird, 2014). These findings could suggest that differences in diet and health vary geographically as well as between population sub-groups (Dearden, Lloyd and Catney, 2019). Moreover, the findings also highlight the wide range of factors which influence the diet/health of individuals (factors within and sometimes outside the control of individuals) and the multi-dimensional, multi-faceted, dynamic and complex (causal) pathways which link area to diet (Chen and Antonelli, 2020; Dearden, Lloyd and Green, 2020; Macintyre, Ellaway and Cummins, 2002).

The second step necessary for adopting a relational approach is the acknowledgement that context and place vary in time and space (Cummins, Curtis, Diez-Roux and Macintyre, 2007). Thus, areas are likely to change over time and individuals are not static or constrained to their residential location. Individuals are exposed to multiple contexts/activity spaces (which vary from person to person), which are likely to change in time and space (e.g. changes in movements over a given day, week, month, year or over the lifecourse) and are associated with socio-demographic factors (such as age, sex, ethnicity, occupation, income, level of education), cultural factors (e.g. religious affiliation) and household level factors (e.g. partner/children in the household) (Kestens et al, 2010). For instance, older retired persons are more likely to spend most of their day/time at home or in their neighbourhood, whereas younger working age adults may spend the majority of their day at work or outside the confines of their residential area. In cautioning researchers against falling into a local or residential "trap" (where contextual measures are limited to the residential area), several scholars have suggested that monitoring the movements of individuals or tracking their changing "contexts" over time could help to improve measures of exposure and provide insight into the location of possible healthy promoting/damaging environments and how individual-level characteristics mediate this relationship (Clary, Matthews and Kestens, 2017; Cummins, Curtis, Diez-Roux and Macintyre, 2007; Kestens et al, 2010).

The final step towards a relational view is the incorporation of geographic scale in the analysis of context. This step involves the acknowledgment that diet and related health outcomes are associated with processes operating at a myriad of spatial scales/levels besides the local or neighbourhood level (Curtis and Rees Jones, 1998). National/international regulatory policies, foreign direct investment, trade agreements, trends in regional economies and the actions of private sector trans-national organisations are just some of the exogenous processes likely to influence dietary patterns and health outcomes at the lower local level geographic scale (Chen and Antonelli, 2020; Cummins, Curtis, Diez-Roux and Macintyre, 2007). Thus, when adopting a relational view, it must be recognised that the analysis of context cannot be constrained to a single spatial scale (such as the local/neighbourhood level) but should include a variety of scales (e.g. macro-scales such as the national, regional, international scale). The following section (Section 2.3.3.3) expounds on this key consideration by discussing the importance of scale in diet and health related research and policy.

2.3.3.3 The importance of geographic scale in health and place research

In expanding on the relational view, Cummins, Curtis, Diez-Roux and Macintyre (2007) noted that besides features of the local area (such as the built environment, social capital, local takeaway and retail outlets previously discussed in Section 2.3.3.1), the availability and consumption of (un)healthy foods was also influenced by factors operating at a higher supra-national/global, national and regional scale. Several studies have highlighted the tremendous impact that international trade agreements/trade liberalisation has had on the diet and health of individuals worldwide (Hawkes, 2006; Friel et al, 2013; Rayner, Hawkes, Lang and Bello, 2006). Trade liberalisation (the removal or reduction of trade barriers/restrictions between countries) operates primarily through the negotiation and signing of international (bi-lateral, regional or multi-lateral) trade and investment agreements or treaties. The European Union (EU)-UK Trade Cooperation Agreement (a new post-Brexit free trade agreement between the EU and the UK which took effect on January 1, 2021) is a recent demonstration of how events which occur at the supra-natural/global scale can possibly impact local food systems, in this instance, food systems within EU member countries, EU partner countries, the UK and the world at large (Freund and Springmann, 2021). Since trade liberalisation directly affects food imports and exports, a major concern is the UK's current dependence on food imports (especially fruits and vegetables) from the EU, the uncertainty around the domestic policies which are to replace those formerly instituted by the EU and the likely impact (positively or negatively) this new agreement will have on the UK's economy, agricultural trade and production, food supply as well as diet and public health going forward (Freund and Springmann, 2021).

Besides the effect on imports and exports (as highlighted with the EU-UK Trade Cooperation Agreement example), trade liberalisation also affects foreign direct investment (FDI), the internal dynamics of the food chain and the retail and commercial promotion/marketing of food within countries (Rayner, Hawkes, Lang and Bello, 2006). FDI (defined as long-term investments made by trans-national enterprises from one country into other overseas countries) has been acknowledged as a major contributor to diet and health, primarily through the increased investment and availability of processed foods (Rayner, Hawkes, Lang and Bello, 2006). As liberalisation increased in the 1980s, major trans-national corporations such as PepsiCo and Nestle increased their investments in overseas manufacturing facilities, which meant that FDI gradually shifted from the export of raw materials to highly/ultra-processed foods such as soft drinks, processed baked goods and snacks (Hawkes, 2006). In 1998, USA-based supermarket chains and major fast-food restaurants such as KFC and McDonalds invested more than US$5 billion in overseas food establishments, and in 2000, US$ FDI in overseas food processing companies was estimated at US$36 billion compared to US$9 billion in 1980 (Hawkes, 2006). The increase in FDI resulted in the increased availability, marketing/advertising and consumption of processed and fast foods globally, the reduced cost for these food items and increased sales for trans-national corporations (Friel et al, 2013). With increased sales, global advertising budgets further increased over the years, with companies in countries such as the UK expending more than £143 million on advertising/marketing in 2016 alone (D'Angelo, Gloinson, Draper and Guthrie, 2020). The McDonalds Dollar Menu in the USA and the (Pound) Saver Menu in the UK are examples of some of the popular global advertising and marketing campaigns and price promotion strategies which primarily target groups such as low-income families, children and ethnic minorities. However, the relationship between advertising and diet-related practices is complex and appears to be more evident for children and adolescents than adults (D'Angelo, Gloinson, Draper and Guthrie, 2020).

The increased availability and consumption of ultra-processed foods and the reduction in the consumption of minimally processed foods (such as fruits and vegetables) globally, has been associated with changes in global food supply chains/systems, increased urbanisation, advances in technology (inclusive of developments in food technology) and global marketing strategies and overall changes in the natural, physical and social environment (Chen and Antonelli, 2020). It has been further argued that the increased availability, popularity and demand for relatively cheap, yet mostly high-fat, high calorie, ultra-processed foods/meals, has helped to change the cultural identity and the expectations of populations worldwide (Cummins, Curtis, Diez-Roux and Macintyre, 2007). Furthermore, the changes in food preferences and global food systems have been associated with the increasing prevalence of obesity (usually measured by the Body Mass Index-BMI) and the tendency for overweight and obese individuals to show a liking for more energy-dense foods (Chen and Antonelli, 2020; Mela, 2001). Trans-national corporations have been acknowledged as major contributors to world food markets and changing dietary patterns, based on their role as producers, manufacturers, retailers and promoters of a variety of foods consumed globally (Friel et al, 2013).

Besides the role of trans-national corporations, national level policy decisions can also have a direct or indirect influence on diet and health. On March 20, 2020, the UK Government announced that standard allowances for Universal Credit (a benefit payment provided to working-age adults who are out of work, unable to work or on low-incomes) were to be increased by £20/week or £1,000 annually (Winchester, 2021). This short-term policy change was designed to provide financial support to persons (especially recently unemployed persons and those new to the benefits system) during the heights of the COVID-19 pandemic. Prior to the end of the uplift, several governmental officials, civil society groups, non-governmental organisations (NGOs) and charities recommended that the uplift be made permanent, to prevent and ease the financial burden faced by the more than five million families affected by the pandemic (Mackley, Hobson and McInnes, 2021; Winchester, 2021). The Trussell Trust (a major food bank NGO in the UK) further noted the positive impact the uplift had on the diets of low-income families (Spoor, 2021). A study conducted by the NGO (through YouGov) found that 72% of persons on Universal Credit since early 2020 said that the uplift was beneficial and facilitated the purchase of basic food items and other necessities such as heating (Spoor, 2021). A major concern is that when the Universal Credit uplift is removed after September 2021, without a suitable replacement financial support scheme, the diet and health of persons (such as disabled persons, those within low-income families) could be severely affected in the long-run (Mackley, Hobson and McInnes, 2021; Winchester, 2021). These findings further highlight how actions or decisions taken by governments, private sector institutions and trans-national corporations (which operate at the national, supra-national and global scale), can help to shape the diet and health of individuals at the local level and vice versa.

As discussed previously in Section 2.3.2, the association between diet and the availability, access and cost of food within the local environment (inclusive of the concept of food deserts and obesogenic environments) has been widely acknowledged by several scholars in the UK and other parts of the world, especially the USA (Acheson, 1998; Black et al, 2012; Blow, Gregg, Davies and Patel, 2019; Cummins, 2014; Cummins and Macintyre, 2002; Cummins et al, 2005; Maguire, Burgoine, and Monsivais, 2015; PHE, 2018c; Roberts, Cavill, Hancock, and Rutter, 2013; Saunders, Saunders and Middleton, 2015; Swinburn, Egger and Raza, 1999; Tinkler and Ritchie, 2016; Williamson et al, 2017; Wrigley, 2002; Wrigley, Warm and Margetts, 2003). However, local/regional culture, norms, perceptions and traditions as well as the influence of the home, family and friends also help to shape individual and local level food consumption. There is relatively strong evidence (from over 50 cross-sectional studies) to support the association between shared family meals and favourable dietary patterns (including the consumption of fruits and vegetables) amongst children (Glanz et al, 2021). However, the influence of the home food environment on diet quality is inconsistent and more complex for working-age adults (the age-group of focus within the current project), who are influenced by multiple environments outside of the home, such as the workplace (Fulkerson, Larson, Horning and Neumark-Stzainer, 2014; Kegler, Hermstad and Haardorfer, 2021; Tyrell et al, 2017). Although the family and the overall home environment has been generally recognised for its positive influence on diet quality and health outcomes, there is still no consensus and very limited data/research regarding the features of the home food environment most associated with adult diet quality (Glanz et al, 2021; Kegler, Hermstad and Haardorfer, 2021). Although not within the scope of the current project, going forward, it is clear that there is need for: i) additional data, especially longitudinal survey data which capture various features of the family and the home food environment (e.g. how food is prepared and served in the home, family dynamics, perceptions of the home etc) and ii) more future studies which explore the mechanisms/pathways which link diet to the home food environment (Fulkerson, Larson, Horning and Neumark-Stzainer, 2014; Glanz et al, 2021; Kegler, Hermstad and Haardorfer, 2021; Tyrell et al, 2017).

Despite liberalisation, globalisation and urbanisation, traditional local/regional foods (foods typically associated with a region or locality), local perceptions, norms and cultural practices form a strong part of national heritage and have the potential to influence the food choices of individuals at the local level (Holt and Amilien, 2007; Monterrosa et al, 2020; Sproesser et al, 2019; Trichopoulou, Soukara and Vasilopoulou, 2007). Perceptions about the geographic region or country of origin of food products have been associated with food preferences (D'Angelo, Gloinson, Draper and Guthrie, 2020). Data from wave five of the UK's Food and You Survey showed that 58% of survey respondents were more confident in eating nationally or locally sourced food products, than those procured from overseas (Fuller et al, 2019). The increased confidence of consumers in UK sourced foods and produce has been partially attributed to the perceived higher quality and stricter food hygiene compliance standards within the UK compared to overseas countries (D'Angelo, Gloinson, Draper and Guthrie, 2020). The importance of geographic food identity and the preservation of national, regional and local food tradition has become even more evident with the establishment of the UK Government's Geographical Indication (GI) scheme on January 1, 2021. The GI (which replaced the former EU scheme) outlines the new post-Brexit rules and legislation (inclusive of intellectual property rights) which govern and protect the authenticity of regional, traditional or geographically connected foods, drinks and agricultural products within the UK (DEFRA, 2020). The famous Cornish Pasty (which consists of potatoes, minced beef and vegetables) is an example of a food item which has been traditionally associated with the county of Cornwall, located in the South West region of England. The meal which became a staple for Cornish miners, working-class men and their families during the 18th-20th century, continues to be produced on a large-scale throughout the county of Cornwall (Cornish Pasty Association, 2021). So integral is this dish to the county that it was formally recognised in 2011 and now under the new GI scheme as being a Protected Geographical Indication (PGI) (Cornish Pasty Association, 2021; DEFRA, 2021). With this designation, a pasty can only be referred to as a “Cornish pasty” if it is prepared and manufactured within the geographical county of Cornwall. Any pasty produced outside of Cornwall cannot be legally referred to as “Cornish” but simply a pasty. The fact that residents of Cornwall, the Cornwall Council and the Cornish Pasty Association collaborated to protect this local tradition and their local economy, helps to further demonstrate the importance of acknowledging regional/local food culture, perceptions, traditions and norms and their likely influence on the diet and health of individuals.

Overall, the findings of this section (Section 2.3.3.3) highlight the importance of geographic scale in the analysis of context and how events/actions at the supra-national, global, national, regional and local scale collectively impact diet and health. The findings also underscore the fact that different health relevant environmental influences are likely to vary at different geographic scales (Weigand, Wurm, Dech and Taubenbock, 2019). Openshaw (1984) further noted the issue of the modifiable area unit problem (MAUP), in which statistical results obtained within health and geography related studies may vary depending on the geographic scale at which data are measured and analysed. A major consequence of the MAUP is that statistical estimates obtained and conclusions inferred may depend on the aggregation unit selected and analysed (Tuson et al, 2019). Hence, the choice of geographic scale and the MAUP (although a relatively common issue) are potential issues which need to be considered when analysing spatially aggregated data. A recommended solution to the MAUP is the selection and use of various levels or multiple geographic scales (such as the national, regional and local geographic scales analysed in the current project) (Wong, 2009). The analysis of more than one geographic scale (to be conducted in the current project) could reveal a range of possible results and could demonstrate how the use of different geographic scales may impact overall study findings (Arsenault et al, 2013; Lee et al, 2014; Wong, 2009). Although the use of a multi-level/scale geographic approach can help to solve the MAUP, data-related and methodological challenges are some of the major issues identified within the current literature for consideration. These challenges will be discussed in the following section, Section 2.3.3.4.

2.3.3.4 Data-related challenges and considerations in health and place research

The association between health and “place” has been well-established. However, there is still a tendency for research to be focused on individuals rather than the physical and social environments to which they are exposed (Macintyre, Ellaway and Cummins, 2002). The poor availability of multi-scale contextual environmental data has been said to be one of the major limitations which prevent researchers from fully exploring the exact features of the contextual environment which most influence diet and health (Gong and Hassink, 2020; Stafford et al, 2007). Potvin and Hayes (2007) noted that researchers' ability to empirically measure and assess the effects of "place" on health is oftentimes constrained by the availability of mostly static, cross-sectional data which lack the dynamic and relational qualities necessary for gaining further insight on this topic. Whilst highlighting the importance of context, Gong and Hassink (2020) found data-related challenges to be partly responsible for the relatively limited studies which have carefully theorised and re-theorised the role of context on health. Other researchers have brought attention to the challenges faced as it relates to the collection, comparison, analysis and interpretation of contextual/area level data (Cummins, Macintyre, Davidson and Ellaway, 2005). More recent studies have highlighted unavoidable practical and data-related challenges (such as the poor availability of true area level data) which prevent researchers from establishing causal links between place and health (Mollborn, Lawrence and Saint Onge, 2021).

Despite continuous calls for more "area" data, the reality is that this type of (usually aggregated) data are mostly pre-defined geographic units (e.g. regions, municipalities, census tracts) captured within "off the shelf" routine secondary cross-sectional surveys and censuses (Cummins, Macintyre, Davidson and Ellaway, 2005; Frohlich, 2000; Macintyre, Ellaway and Cummins, 2002). Increasingly limited research budgets and the lack of consensus on the exact features of the physical and social environment which influence diet and health, makes it difficult for secondary survey data collectors to capture "true" area data, especially at the national level (Cummins, Macintyre, Davidson and Ellaway, 2005, Dearden, Lloyd and Green, 2020; Frohlich, 2000; Mollborn, Lawrence and Saint Onge, 2021). This would explain why true area data are not usually available in routine nationally representative surveys such as the NDNS and HSE analysed in this project. For this reason (and due to the significant financial, human and physical resources required to collect primary data), researchers are very often forced to make the best use of the secondary data currently available and adjust research questions accordingly (Campbell et al, 2020, WHO, 2015b). As a result, very few studies have been able to explore the specific features of the local contextual environment, as context is usually constrained to existing "area" level data captured within secondary administrative, survey and census datasets (Diez-Roux, 2002; Macintyre, Ellaway and Cummins, 2002; Mollborn, Lawrence and Saint Onge, 2021; Stafford et al, 2007).

Cummins, Macintyre, Davidson and Ellaway (2005) investigated the influence of social capital and transportation on health. The authors of that study conceptualised the features of the social and physical environment and the corresponding indicators which they believed were the most relevant to health and the objectives of their study. Following the conceptualisation process, a wish list of 22 indicators were proposed, which consisted of 10 indicators/measures for social capital and 12 indicators for transportation. However, data were only available for six (2 indicators for social capital and 4 indicators for transportation) of the 22 indicators initially proposed. Among the major data-related challenges cited by the authors were: the type and format in which data were captured, the survey design (e.g. cross-sectional, longitudinal), the accuracy and completeness of datasets, the geographic scale at which data were readily available, differences in how area boundaries were defined/measured, the type(s) of variables/questions captured within surveys and the lag time between data collection, processing and public dissemination. The cost to access data, especially from commercial institutions or large brokering firms, was another barrier with the potential to restrict studies to routine cross-sectional or administrative data, which may not be aligned to pre-conceived constructs/hypotheses (Cummins, Macintyre, Davidson and Ellaway, 2005). Data collectors/secondary data repositories may not be able to facilitate special requests for additional area level data, due to confidentiality or the sensitivity of geographic data. Also, institutions which house these types of data may not have the time, interest, staff or resources to cater to the unique data needs of researchers. Arsenault et al (2013) noted that researchers are often limited in terms of the features of the contextual environment and the geographic scale they can realistically investigate, due to the: i) the generally poor availability of area-level data in routine surveys/databases, ii) the geographic scale at which data are available, which may not be in keeping with research interests and hypotheses, iii) privacy and data dissemination restrictions which prevent data repositories/data from sharing data and iv) the high cost associated with collecting primary data. Moreover, a lot of the “true area data” on social and material context are not available in routine nationally representative cross-sectional surveys, due to the complexity and difficulty in collecting, generating and interpreting this type of information at the national level (Cummins, Macintyre, Davidson and Ellaway, 2005). This poses an additional challenge to projects/studies (such as the current project) interested in assessing diet and health at the national level.

As mentioned in Section 2.3.3.1, Stafford et al (2007) examined various contextual environmental domains (such as crime, policing, fast-food outlets, social disorder etc) and their relationship with diet-related outcomes such as obesity and body fat distribution. However, one of the major limitations noted within that study was the unavailability of data for several domains initially conceptualised. The unavailability of data at the geographic scale of interest to the authors (e.g. postcode sector), the lack of longitudinal data which linked the contextual environment to health and the inability to investigate behavioural change based on the availability of mostly cross-sectional data were some of the other data-related challenges noted within that study.

Frohlich (2000) noted the importance of acknowledging research interests, the scope of research projects and the data/method which best suits research objectives. For instance, a researcher's decision to explore context as pre-defined geographic units (such as geographic region, municipality, census tract etc) could be due to their interest in examining *where* there is variance rather than *why* there is variance at different geographic levels/scales (Frohlich, 2000). Similarly, the role of the family and the overall home environment (as discussed in Section 2.3.3.3) has been recognised for its positive influence on diet quality and health outcomes, especially amongst children (Glanz et al, 2021). Whilst some diet-related data are captured in household expenditure surveys, these types of surveys are primarily designed to measure household food acquisition and not individual-level dietary intake (Micha et al, 2018). Thus, researchers more interested in assessing the types of foods actually consumed by individuals are advised to use individual-level diet surveys instead of household expenditure surveys (Campbell et al, 2020; Micha et al, 2018). However, individual-level diet surveys tend to have very limited data on the dietary pattern of family members or other persons within the household, which essentially limits the examination of the influence of the household/family on individual dietary patterns (Micha et al, 2018). The influence of the family/household on diet and the need for more studies which explore the features of the social and physical environment which influence diet and health is acknowledged. However, it must be recognised that a researcher's decision to explore a specific topic and their choice of research question(s)/scale, is also based on identified research needs/gaps, research interests, and other practical issues such as the cost and availability of data (Boo and Froelicher, 2013; Frohlich, 2000). Arsenault et al (2013) further noted that there may be instances where the selection of the research topic and geographic scale may have to be purposively selected in order to address a particular issue/research need, a specific gap(s) in the literature and/or to meet the unique needs of targeted end-users. In addition, findings may also have to be communicated in a format and scale that end-users can easily understand and use for planning, policy formulation and targeted intervention.

Based on the findings of this section (Section 2.3.3), it is evident that there are still several fundamental issues which need to be addressed, chief of which include how context is to be theorised and measured and what are the specific features of the local environment most likely to influence diet and overall health (Cummins, Macintyre, Davidson and Ellaway, 2005). Another lacuna which exists in diet and public health literature, is that despite countless diet-related studies, the dietary pattern of individuals and how patterns vary across and within countries remains relatively unexplored (Imamura et al 2015; Micha et al, 2018). Thus, little is known about geographic differences in diet and health, which is partly due to the poor availability of diet-related data (Imamura et al 2015; Micha et al, 2018; Mollborn, Lawrence and Saint Onge, 2021).

Sproesser et al (2019) highlighted the rapid changes/transitions in what and how people eat globally, commonly referred to as the nutrition transition (discussed in Section 2.4). Global nutritional changes have been partially attributed to globalisation, urbanisation, liberalisation and the modernisation of food eating (through access to new technologies, modern food markets/supermarkets etc) (Popkin, 2006; Sproesser et al, 2019). Foods and dietary patterns once considered to be traditional or typical for a region/country/locality, have been gradually replaced by more Western dietary patterns, which consist of a high proportion of saturated fats and free sugars (Sproesser et al, 2019; Trichopoulou and Lagiou, 2004). The rapid change in how the world eats threatens the extinction of traditional country/region-specific dietary patterns and begs the question, what are people currently eating and do region-specific/traditional dietary patterns still exist? (Trichopoulou and Lagiou, 2004). These findings underscore the need for more current, rigorous and evidence-based studies, especially in countries like England where the study of possible region-specific dietary patterns remains unclear and underexplored (Emmett, Tweney, Golding and Taylor, 2017). There is a genuine need to examine the dietary pattern of individuals across and within countries, in order to: i) detect the type of foods people are actually eating (rather than what they should be eating), ii) identify possible foods to be flagged for further nutritional intervention, iii) assist in preserving local and regional food cultures where possible, iv) highlight possible geographic and/or socio-demographic variations in diet consumption and v) provide researchers and policymakers with the information necessary to systematically plan and execute targeted evidence-based interventions aimed at improving population diet, rather than those based on inaccurate assumptions (Cummins, 2014; Sproesser et al, 2019; Trichopoulou and Lagiou, 2004).

Before delving into literature on the English diet (the focus of this study), Section 2.4-2.6 will highlight the significant transitions in diet and nutrition which have occurred at the global, national, regional and sub-regional level and expound upon key conceptual/theoretical models (such as the Nutrition Transition Model to be discussed in Section 2.4) which have been used to explain these changes.

2.4 The Nutrition Transition Model

Most public health scholars accept that rapid and significant transitions in diet have occurred across most, if not all regions of the world (Hawkes, Harris and Gillespie, 2017; Popkin, 2006; Shetty, 2013). Popkin (1993) theorised that demographic and epidemiological transitions usually precede or occur simultaneously with changes in the type, quantity and quality of food consumed worldwide. The "Nutrition Transition" (Popkin, 1993; Popkin, 2002; Popkin and Ng, 2007), generally models the significant changes in dietary patterns within developed and developing countries worldwide.

According to Popkin (2006), there are five broad nutrition patterns which summarise the shift in global dietary patterns (Figure 2.2).

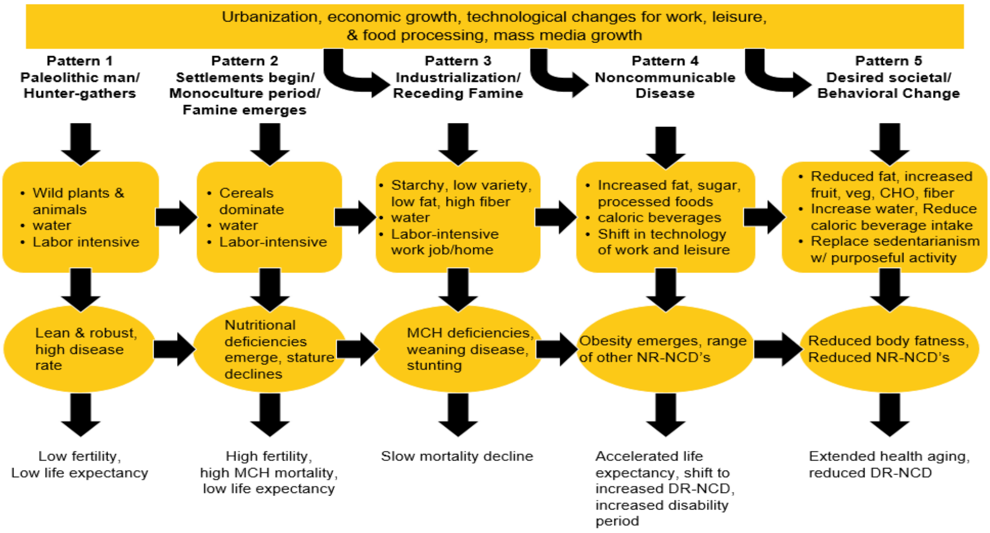


Figure 2.2: The stages of the Nutrition Transition, Popkin (2006)

Pattern one (Collecting Food) is characteristic of the hunter-gatherer society, in which diets tend to be high in carbohydrates and fibre, and low in saturated fat. Pattern two (Famine) describes periods of food scarcity and an inevitable shift to diets with little or no variety. During the latter stage of this pattern, differences in diet by sex and SEP increase and social stratification intensifies. By pattern three, (Receding Famine), chronic hunger begins to decrease as famines recede; the consumption of fruits, vegetables and animal protein increases, the consumption of starchy staples decreases, and physical activity begins to decrease. Pattern four (nutrition-related non-communicable disease) stands in complete contrast to its predecessors, as sedentary lifestyles become the norm and diets become high in total fat, cholesterol, sugar, refined carbohydrates, processed foods and low in polyunsaturated fatty acids and fibre (Popkin, 2002; Shetty, 2013). At this stage, low physical activity and poor diets result in a high prevalence of obesity and diet-related chronic diseases. Pattern five (Behavioural Change) on the other hand, describes the emergence of a new dietary pattern intended to prevent or delay the development of obesity and chronic diseases and increase disability-free life expectancy. At this final stage, it is expected that diets will include lower levels of saturated fat and sugar; and the reliance on processed foods will begin to decline. To date, no country has fully reached pattern five of the nutrition transition, but most have transitioned at least to pattern three. Pattern four is characteristic of high income developed countries such as England. However, many low-income developing countries have begun to experience the double burden of infectious and diet-related chronic diseases (Dye, 2014), as they gradually move to pattern four of the nutrition transition.

The nutrition transition cannot be limited to any particular period of human history, as the "pace of dietary changes appears to have occurred to varying degrees in different regions of the world" (Popkin, 2006, p. 289). The unavailability of routine standardised diet-related data also makes it difficult to identify when transitions actually began and how changes over time compare across and within countries (Popkin, Adair and Ng, 2012). Nevertheless, Popkin (2006) generally acknowledges that over time, besides demographic and epidemiological changes, globalisation coupled with changes in food supply, transport, agricultural systems, policies, technology and mass media have significantly affected food demand, usage and overall dietary patterns, as evidenced by rising obesity rates in developing and especially developed countries worldwide. However, despite these paralleled changes, the model's focus (similar to this project), is primarily on the significant shift in the structure and composition of diets and more specifically, geographic/regional and socio-demographic variations in dietary patterns.

2.5 Global and regional variations in dietary consumption

2.5.1 Evidence of global transitions and variations in diet

Several decades ago, it was the North American region, specifically the USA which was said to have a problem with diet and obesity, and not the rest of the world (Popkin, Adair and Ng, 2012). At that time, it was becoming clear that diets in the USA and, to a lesser extent, sections of Europe, were worsening and obesity was rapidly increasing in these regions. However, the significant rise in obesity rates globally and the transition in how the *world* eats today is undeniable. The nutrition transition clearly shows that global dietary patterns are moving away from diets high in legumes, fruits, vegetables and coarse grains and converging to what is known as a more "Western diet." The Western diet can be broadly defined as a diet high in refined carbohydrates, ultra-processed foods, saturated fats, salt, free sugars and low in fruits and vegetables (Bere and Brug, 2010; Kyriacou, Evans, Economides and Kyriacou, 2015; Ogce, Ceber, Ekti and Oran, 2008; Popkin, Adair and Ng, 2012). Given that various sub-populations within a country may manifest different dietary patterns, Popkin, Adair and Ng (2012) further underscore the importance of exploring how vastly diets have changed over time and the extent to which regional diets differ or have converged to the stereotypical Western diet.

As previously mentioned, there are challenges related to the availability of consistent and standardised diet data, which limits the exploration of diets across and within countries over time (Popkin, Adair and Ng, 2012; Walls, Johnston, Mazalale and Chirwa, 2018). Nevertheless, Afshin et al (2019), besides estimating the contribution of diet to the global burden of disease (GBD), attempted to explore regional differences in diet across 195 countries, broadly classified into 21 study regions. Overall, daily consumption of healthy foods such as fruits, vegetables, nuts and whole grains was sub-optimal in all 21 study regions. Unsurprisingly, unhealthy foods, particularly sodium and SSBs exceeded the optimal level in almost all of the regions investigated. Optimal consumption was measured as the level of exposure that minimised the risk from all causes of death.

Although all 21 regions investigated had diets which were below par, regional differences were observed when data were further disaggregated. For instance, fruit consumption was below the optimal level in all 21 regions. However, Central Asia reportedly had the highest consumption of vegetables in 2017. Seafood/omega-3 fatty acids were consumed highly in the Asia Pacific region whereas legumes were mostly consumed in the Caribbean, tropical Latin America, South Asia and Western and Eastern sub-Saharan Africa. In terms of unhealthy foods, red meat consumption was highest in Australasia, followed by Southern Latin America. North America, followed by Central and Southern Latin America, consumed the most trans fats and SSBs. The consumption of processed meats (foods that are key characteristics of the Western diet and stage four of the nutrition transition) was highest in North America, followed by Asia Pacific and Western Europe.

The high consumption of unhealthy foods in North America was not surprising given that the USA falls within this region and has had long standing issues with sub-optimal diet, which date back to the 1970s (Popkin, Adair and Ng, 2012). Similarly, the relatively high consumption of processed meats/foods observed in regions such as Western Europe (Afshin et al, 2019) appeared to be consistent with other studies which noted the low consumption of minimally processed food (such as fruits and vegetables) and the higher consumption of ultra-processed foods in countries such as the UK (Monteiro et al, 2018). These findings not only help to confirm the 2017 GBD study (Afshin et al, 2019), but the global nutrition transition and the seeming convergence of countries worldwide towards the consumption of more unhealthy foods such as ultra-processed foods.

2.5.2 Transitions and variations in diet in Europe

Almost all foods and drinks consumed today, are processed in some way. Therefore, traditional and "back to nature" approaches, may not always be the most practical solution for today's dietary challenges (Monteiro, 2009). According to Monteiro (2009), although current food guidelines and classifications (such as the WHO guidelines) recommend the reduced consumption of highly (ultra) processed foods, very little guidance has been provided beyond this.

In an attempt to move beyond the traditional assessment of dietary patterns, Monteiro et al (2018) analysed the association between the national availability of ultra-processed foods and the prevalence of obesity among adults in 19 European countries. The study spanned a wide range of countries across Europe; from the UK to Cyprus/Mediterranean sections. The exploration of these countries was in line with other studies which noted significant diet and health inequalities across Europe and the need for more research on this region (Bambra, 2016).

Expenditure on ultra-processed foods varied across the countries, ranging from 10.2% in Portugal (lowest expenditure), 13.4% in Italy, 46.2% in Germany, to 50.7% in the UK (highest expenditure). Packaged breads, cakes, cookies and other baked products, reconstituted meat products and SSBs were the most commonly consumed ultra-processed foods across all countries studied. Malta and Portugal, located within the South and Mediterranean sections of Europe, had the lowest consumption of packaged breads compared to Ireland (North of Europe) which consumed the most. Croatia (South Eastern section of Europe) reportedly had the lowest consumption of cakes, cookies and other baked products, unlike Belgium (West Europe) which had the highest. Reconstituted meats were least consumed in Greece (South East of Europe) and mostly consumed in Germany (West Europe). SSBs were least consumed in Italy (located within the South and Mediterranean sections of Europe) and mostly consumed in Latvia (Central and Eastern Europe) (Monteiro et al, 2018).

With the release of the study's findings in 2018, the UK was immediately branded by the media as having the worst diet in Europe, with subsequent studies showing that approximately 19,000 deaths due to heart disease could be averted if healthier diets (such as those consumed in the Mediterranean region) were to be adopted in the UK (Tong et al, 2016). The UK's high expenditure on ultra-processed foods, indirectly points to the nutrition transition and the shift in diets from what Monteiro et al (2016) refers to as minimally processed foods (such as vegetables, grains, fish and plain unsweetened yogurt), to unhealthy ultra-processed foods (such as soft ("fizzy") drinks, ice-cream, burgers, ready-to-eat or heat foods and pizza dishes).

The findings from Monteiro et al (2018) although useful, should be interpreted with caution, as the study, like most diet-related studies, was more focused on the association between diet (in this instance ultra-processed food consumption) and the prevalence of obesity (health outcome). Therefore, dietary patterns and regional differences in diet were not within the study's immediate scope. Another limitation of the Monteiro et al (2018) study, was the use of survey estimates from as far back as 1991. Estimates of ultra-processed food consumption were said to be the highest in the UK and the lowest for Portugal, followed by Italy. However, UK consumption estimates, although captured over a decade ago (in 2008), were the most recent and perhaps most reflective of current consumption; unlike Portugal and Italy where surveys used were conducted in 1996 and 2000 respectively. It is therefore possible that the differences observed between these countries were more reflective of the time period in which survey data were collected. In light of the nutrition transition, more recent food consumption survey data (though limited or unavailable), could possibly reveal little or no variation between the regions or a reversal of previously observed trends.

Despite the limitations and considering the difficulty posed in assessing dietary consumption, the findings from Monteiro et al (2018), appeared to be line with previous research. Studies which have alluded to a North-South differentiation, with persons from the South/Mediterranean sections of Europe generally consuming greater amounts of fruits, vegetables, cereals, fish and seafood, and vegetable oils (olive oil), compared to persons in the North (Roos et al, 2000; Slimani et al, 2002; Trichopoulou, Naska and Costacou, 2002). Popkin (2006) noted the consistently high consumption of fruits and vegetables in Spain and Greece, despite drastic reductions in rest of the world. Rumm-Kreuter (2001) also made note of substantial socio-economic changes throughout Europe over the past four decades, which are likely to have contributed to the worsening of diets in North Europe and to a lesser extent, South Europe/Mediterranean areas, once again suggesting a North-South differentiation in eating habits. Such findings help to explain the tremendous focus and critical acclaim South Europe/the Mediterranean region and the Mediterranean diet (MD) has received over the past several decades.

2.5.3 South Europe/Mediterranean region and the Mediterranean Diet

The MD (which has been likened to the Japanese diet) has been said to be one of the healthiest diets in the world (Grimaldi and Paoli, 2015). This diet was first introduced in the 1950s after the overall lifestyle practices and the foods consumed within the Mediterranean region were investigated and found to be responsible for the relatively good health of the people in that region (Ogce, Ceber, Ekti and Oran, 2008; Rumm-Kreuter, 2001). Since then, numerous studies have been conducted, most highlighting the benefits of the MD on blood pressure, BMI, dementia, heart disease, obesity, all-cause mortality, cancer and type 2 diabetes (Ogce, Ceber, Ekti and Oran, 2008; Renzella et al, 2018; Sánchez-Villegas and Martínez-Lapiscina, 2018). The diet originated in the olive oil growing areas of South Europe and the Mediterranean region, primarily in Cyprus, Croatia, Greece, Italy, Morocco, Portugal and Spain. The traditional MD consists of a high consumption of fruits, vegetables, nuts, legumes, whole grains, olive oil;

moderate consumption of fish, poultry and wine and a low intake of dairy products (yogurt and cheese), processed/red meat and sweets. The replacement of sweets for fresh fruit, the moderate consumption of wine with meals and the heavy use of olive oil are some of the unique features of the diet. Although widely touted as a model diet to be adopted by countries such as the UK, public health scholars have recently debated whether countries in South Europe/the Mediterranean region have truly held fast to the traditional MD.

D'Innocenzo, Biagi and Lanari (2019) noted the seeming disappearance of the traditional MD and its replacement with more of a Western diet. The consumption of olive oil, fresh fruits and vegetables is still relatively prominent in countries such as Portugal, Spain, Italy and Greece (Cruz, 2000). However, persons from higher SEP groups are more likely to adhere to the traditional MD compared to lower SEP groups (Bonaccio et al, 2017). Additionally, obesity rates, especially amongst youth in the Mediterranean region are reportedly higher than Nordic countries such as Sweden (D'Innocenzo, Biagi and Lanari, 2019). The inexorable increase in the incidence of obesity (particularly among youth) within the Mediterranean region over the years, has been partly attributed to the nutrition transition and the higher consumption of meat, dairy products, confectionery and “junk” food; and the reduced consumption of fruits and vegetables (Cruz, 2000; Denoth et al, 2016; Ogce, Ceber, Ekti and Oran, 2008). These findings could indicate a reversal of previously observed trends (D'Innocenzo, Biagi and Lanari, 2019), a closing of the gap between the North and South of Europe and convergence to a more universal unhealthy dietary pattern (e.g. the Western diet).

Studies such as Afshin et al (2019) and Monteiro et al (2018) are evidence of the steps being taken by researchers to acknowledge the nutrition transition and investigate regional differences in dietary patterns beyond the conventional approach. However, a gap remains. There are very few studies which have been able to systematically examine current geographic/regional and socio-demographic differences in diet (Bonaccio et al, 2017). The poor availability of dietary consumption data is a major challenge which prevents researchers from fully exploring this topic. These data are usually limited, scattered or not within the public domain, and where publicly available, data are inconsistently collected across and within countries (National Academies of Sciences, Engineering & Medicine, 2019). Household consumption and expenditure surveys (HCES) are frequently used to estimate individual dietary patterns, since they are conducted regularly (e.g. annually) in many countries worldwide and provide excellent geographical coverage over time (Micha et al, 2018). However, socio-demographic/compositional factors such as age, sex and SEP cannot be fully explored as data are collected at the household and not the individual level. Individual level dietary recall surveys (e.g. 24-hour diet recall surveys and Food Frequency Questionnaires) can be used to examine socio-demographic patterns in food and beverage consumption. However, these types of surveys are usually limited in terms of geographical coverage and oftentimes focused on the consumption of single food groups such as fruits and vegetables rather than overall diet. These data challenges once again help to explain the tendency for researchers to focus solely on health outcomes such as obesity or on compositional factors related to diet (Brug, 2008).

Despite data challenges, countries in the South/Mediterranean region continue to be lauded for the MD and the healthy lifestyle practices supposedly exhibited in this region. These findings implicitly suggest a North-South divide in dietary patterns in Europe, whilst explicitly recommending the universal adoption of "model" dietary patterns (like the MD), without the requisite data and research to validate these claims. Undoubtedly, there are still inherent traditional, social and cultural differences between the UK and the South/Mediterranean region. However, with the nutrition transition, it is becoming more evident that global and regional diets are slowly converging to a more widespread unhealthy dietary pattern, synonymous with the Western diet (Afshin et al, 2019; Popkin, Adair and Ng, 2012). Therefore, the wholesale adoption and consumption of diets like the MD or region-specific foods, may not be the most culturally/regionally appropriate solution to the poor-quality diets observed in the UK and globally (Bere and Brug, 2010). Instead, it may be more appropriate at this stage for Mediterranean and non-Mediterranean countries (like the UK/England) to return to or adopt (respectively) the general principles of the MD. In particular, consuming fewer ultra-processed foods and more minimally processed, fresh, nutrient-dense foods with health benefits (Bere and Brug, 2010; Renzella et al, 2018).

The exploration of regional differences between the UK and the Mediterranean region (South Europe) is not within the immediate scope of this project. Instead, this project aims to contribute to the literature by critically analysing geographic and socio-demographic differences in individual dietary patterns across the regions of England, a constituent country within the UK. This type of assessment will add to the limited studies which investigate how diets vary within countries at the national/population level (Imamura et al 2015), especially within the UK, which is now said to have the worst diet in Europe. Findings garnered from this project could also provide the evidence needed by researchers and policymakers to plan and execute more appropriate, country-specific and effective localised interventions, aimed at improving individual diets and reducing the prevalence of diet-related health outcomes (such as obesity) and health inequalities in general.

2.6 Regional Variations in dietary consumption across the UK

2.6.1 England versus Scotland

For decades, inequalities in diet and health have been observed across small and larger areas of the UK (Scotland, Wales, Northern Ireland and England) (Barton, Chambers, Anderson and Wrieden, 2018; Shelton, 2009). For instance, the Scottish diet has been found to be higher in saturated fat, salt, alcohol, red and processed meat, confectionery and soft drinks, and lower in fruits, vegetables and fibre, compared to the average English diet (Barton, Chambers, Anderson and Wrieden, 2018; Scarborough, Morgan, Webster and Rayner, 2011; Shelton, 2009). Similarly, persons in Northern Ireland reportedly consume more saturated fat per day and less fruits and vegetables per week, compared to England (Scarborough, Morgan, Webster and Rayner, 2011). On the contrary, fruits and vegetables are consumed more in Wales compared to England, Scotland and Northern Ireland. Although no significant difference has been found between the English and Welsh diet, Scarborough, Morgan, Webster and Rayner (2011) noted that overall improvements to the average diet in Wales, Scotland and Northern Ireland, to a level already achieved in England, could significantly reduce geographic variations in diet and health in the UK.

The Scottish diet, and more specifically, dietary inequalities between Scotland and England has been a subject of particular interest for many years. Despite several dietary policy initiatives and interventions, Scotland's alleged acceptance and consumption of energy-dense, nutrient poor foods (e.g. deep fat fried Mars Bars and processed meats such as Haggis and Black Pudding), coupled with high levels of socio-economic deprivation, are factors said to explain the relatively slow pace of dietary change in Scotland compared to England (Barton, Chambers, Anderson and Wrieden, 2018; Shelton, 2009; Wrieden et al, 2013). As a result, most diet-related studies in the UK tend to make generally negative assessments of Scotland, whilst painting England in a more positive light. It has been suggested that if Scotland were to adopt a diet equivalent to that of the English diet, approximately 40% of deaths due to heart disease, stroke and diet-related cancers could be delayed or averted (Scarborough, Morgan, Webster and Rayner, 2011).

In a review of over 50 diet-related studies, Wrieden et al (2015) revealed that sample sizes for Scottish studies are usually small (less than 200) and less than optimum for a robust comparison with England. Most studies that have investigated dietary patterns in Scotland and England, were not conducted at the national level, but instead made comparisons between Glasgow (a city in Scotland), Manchester and Liverpool (two cities in the North of England). The focus on urban areas and the small sample sizes used in Scottish studies, coupled with the tendency for studies to focus on single food groups (such as fruits and vegetables), are major limitations which need to be considered when comparing diets in Scotland and England. Overall, very few studies have investigated England as a whole or differences in dietary patterns across the regions of England (Walsh et al, 2016; Wrieden et al, 2015).

Another limitation observed is the tendency for studies to focus on the relationship between diet and health outcomes such as cardiovascular disease within certain cohorts/groups (e.g. older women) as opposed to assessing geographic variations in diet at the national level (Scarborough, Morgan, Webster and Rayner, 2011). Household consumption and expenditure surveys (HCES) are oftentimes used to estimate individual consumption. These data, however, do not accurately reflect individual dietary consumption (Micha et al, 2018). Thus, more detailed diet-related surveys such as the annual UK National Diet and Nutrition Survey (NDNS) should be further utilised to assess actual and individual level (rather than household level) dietary patterns within England and Scotland (Scarborough, Morgan, Webster and Rayner, 2011; Wrieden et al, 2015).

Citing Riva, Gauvin and Barnett (2007), Shelton (2009) noted the necessity of considering and analysing local deprivation in addition to a regional or national context. Based on how data are structured and collected, individuals are naturally nested within regions as well as within mutually exclusive neighbourhoods across England and Scotland. As such, measures of relative deprivation (although not perfect) such as the Index of Multiple Deprivation (IMD) should ideally be taken into consideration when assessing regional variations in diet (Shelton, 2009). Acknowledging the geographic scale of measurement therefore supports the need for multi-level modelling of dietary patterns across the regions and deprivation areas of Scotland and in particular, England (Shelton, 2009).

Although Scotland has been portrayed by the media as the “sick man of Europe” and a country in a worse position than England in terms of diet, the studies upon which this claim is based, have a number of limitations as highlighted in this section. The focus on urban areas in Scotland, the small sample sizes used and other data-related challenges, highlight the need for more within country studies which investigate differences in diet across the regions of Scotland and England before inter-country comparisons can be made.

2.7 The English Diet: Past and Present

Although diets in England were said to vary by “social class”, historical accounts (though limited) have revealed that the typical English diet during the 19th century was predominantly a working-class diet of sugary tea, white bread, margarine, condensed milk and fried fish and chips (Otter, 2012). By the turn of the century, the diet gradually transitioned to what was oftentimes referred to as “meat and two veg.” This traditional meal usually consisted of cooked meat (mostly beef or mutton), potatoes, and two types of green vegetables, or a meat dish in the form of a casserole/pie (Riley, 2010). "Sunday Dinner/Roast" (typically served on a Sunday) is a traditional form of the meat and two veg dish, as it consists of roasted meat/beef, roasted potatoes, at least two types of vegetables (such as peas and carrots) and accompaniments such as Yorkshire pudding and gravy (Otter, 2012). Although meals such as the Sunday Roast are (nationally recognised) features of the traditional English diet (Vogler, 2020), historically, there have been noted dietary and health differences between the regions of England (Bambra, 2016). In fact, several traditionally English dishes have either been named after or associated with the region or area in which they are said to have originated. For instance, the Yorkshire pudding is traditionally associated with Yorkshire and the Humber region, likewise, Staffordshire oats cakes from Staffordshire in the West Midlands; the Cornish pasty from Cornwall in the South West region; the Lancashire hot pot from Lancashire in the North West region; the Bedfordshire clanger from Bedfordshire in the East of England and jellied eels traditionally associated with London (Vogler, 2020). Although these and other regionally distinct foods have been in existence for decades, studies have shown that the English diet has transitioned from a diet of mostly meat pies, cooked potatoes and green vegetables (Riley, 2010), to a more "Western" type dietary pattern, high in SSBs, saturated fats, free sugars, salt and ultra-processed foods and low in fruits and vegetables (Newton et al, 2015).

Historians are unable to establish the exact timing and impact of the English nutrition transition or the extent to which current diets in England vary regionally (Otter, 2012; Popkin, 2002). However, the English diet cannot be objectively discussed without considering wider themes in English history such as World War II, rationing, domestic food technology (e.g. refrigeration), migration, deindustrialisation, the "North-South health divide", health inequalities and the social determinants of health (Figure 2.3). For instance, the increased immigration of persons from post-colonial territories in the Caribbean and particularly the South Asian region had a significant impact on the English diet as the gradual injection of ethnic cuisine made items such as "Chicken Tikka Masala" a new English favourite (Buettner, 2008). Besides post-colonial migration and the end of rationing in 1954, the increase in foreign holidays/travel and the introduction of more foreign style restaurants in England (inclusive of Indian curry houses and Italian, French and Thai restaurants) were other events which helped to change the English diet to what could be considered a melting pot of cultures (Buettner, 2008; Burnett, 1989) (Figure 2.3). Although these and other major historical events are not discussed in great detail within this project, it must be recognised that these events have helped to shape the current English diet and have significantly impacted how dietary patterns and health outcomes, at the national and sub-national level, have evolved over time (Burnett, 1989; Buettner, 2008; Oddy, 2003; Vogler, 2020). Figure 2.3 was therefore designed* and prepared specifically for this project and provides a general overview of some of the of major historical events and diet-related thematic areas which help to explain the changes in the English diet, inequalities in diet/health and the economic, social, political and cultural factors associated with these changes.

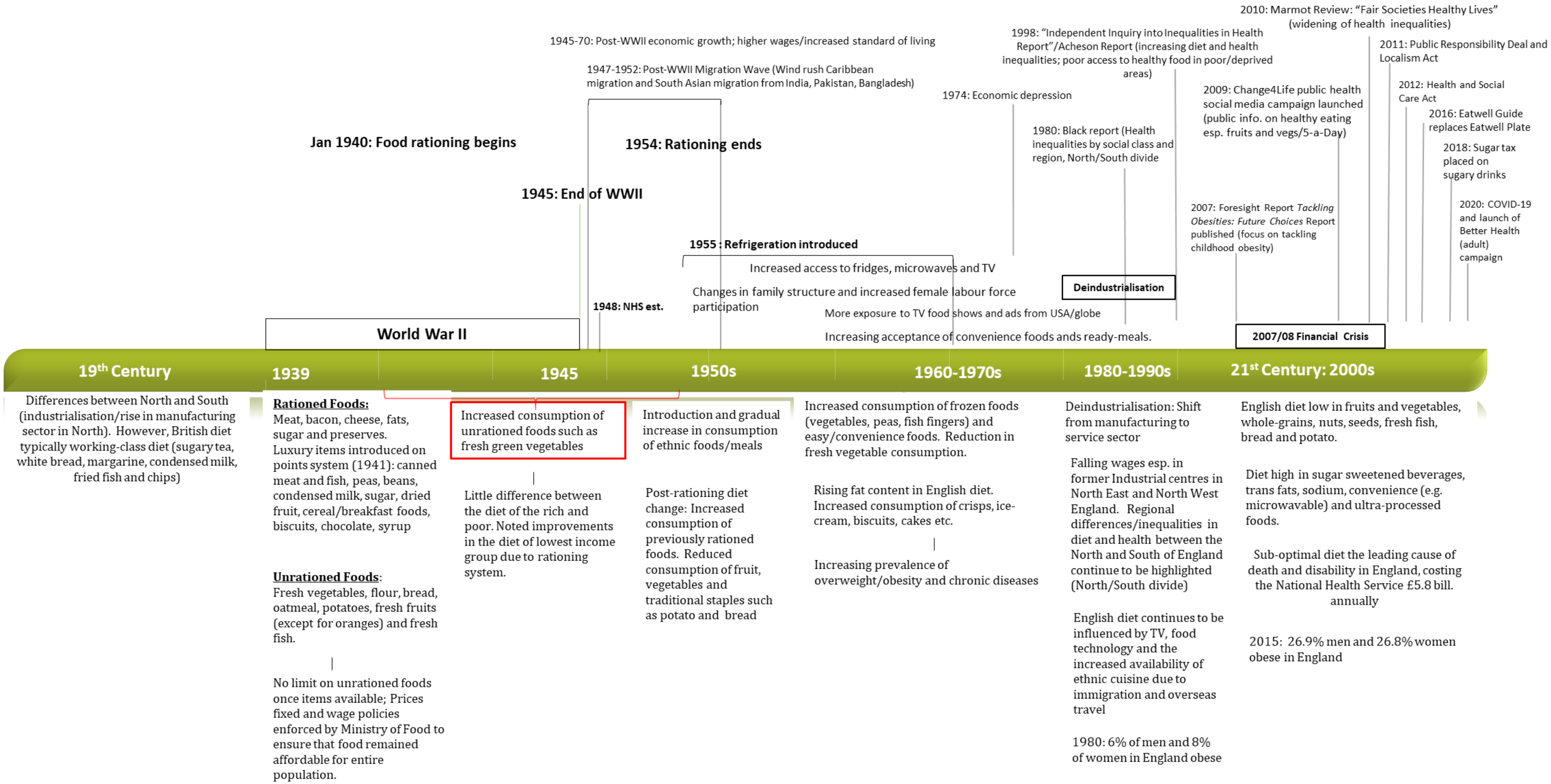


Figure 2.3: Transitions in the English diet, 19th century to the 21st century

*Designed and prepared by the author, Monique Alayne Campbell

2.7.1 Dietary recommendations and the current English diet

The Eatwell Guide (which replaced the former Eatwell Plate in 2016) (Figure 2.3) is currently the English Government's major policy document on diet (not obesity) and provides recommendations on eating a healthy and balanced diet (NHS Digital, 2019). The Eatwell Guide (which is aligned to WHO recommendations mentioned in Section 2.1) generally recommends: the consumption of at least five portions of fruits and vegetables daily for persons aged 11 and over; no more than 70 grams/day of red and processed meat; at least one portion (20 grams per day or 140 grams per week) of oily fish per week; the consumption of more whole-grain, high fibre starchy carbohydrates and protein rich foods (e.g. beans, eggs, pulses) and the reduced consumption of free sugars and saturated fat to no more than 5% and 11% (respectively) of daily calorie intake (NHS Digital, 2019). Despite these recommendations, studies have found that individuals in England consume 30% fewer fruits and vegetables and double the recommended amount of fat (especially saturated fat) and free sugars (Global Food Security, 2016; NHS Digital, 2019). In 2017, only 29% of adults in England consumed the recommended five portions of fruits and vegetables daily, with women and persons over the age of 25 more likely to do so than men and persons aged 16-24 years. Over the 2009-2016 period, adult fruit and vegetable consumption in England ranged from 3.5 to 3.6 portions per day and although consumption increased slightly to 3.8 in 2017, this was still below current recommendations (NHS Digital, 2019).

An analysis of data from the 2016/17 Living Costs and Food Survey (the UK's premier household consumption and expenditure survey), also revealed that bread (white, brown and wholemeal) and potato, once traditional staples in the English diet, as well as fresh fruits and green vegetables continued a long-term downward trend. The trends were first observed after World War II and the end of food rationing in 1954 (DEFRA, 2018) (Figure 2.3). However, ultra-processed foods, characteristic of the Western diet, continued an upward trend, peaking at 170 grams per person per week. These statistics again highlight the shift away from consuming more minimally processed foods such as fruits and vegetables, to eating more ultra-processed and unhealthy foods (DEFRA, 2018; PHE, 2018a).

Although the English diet tends to be placed in a more positive light compared to diets in the rest of the UK, particularly Scotland, the previous findings add credence to the fact that dietary patterns in England are far from perfect and not in alignment with current dietary recommendations. Several interventions, policy reports, papers, strategies and academic research papers, have ensued over the last 20 years, in response to dietary changes and the rising rate of obesity and health inequalities in England. However, sub-optimal diet continues to be the leading cause of death, disability and rising obesity rates in England; ultimately contributing to increasing diet and health inequalities and financial costs to the society (Jebb, 2018; Newton et al, 2015). It has been estimated that approximately 30,000 deaths per year in England, are due to diet-related diseases such as obesity (PHE, 2017). Additionally, sub-optimal diet continues to be the largest economic burden to the National Health Service (NHS), with diet-related ill-health costs estimated at £5.8 billion annually (Scarborough et al, 2011). Based on these findings, it is clear that more evidence-based diet specific (not only obesity) policies are needed. The development and implementation of effective diet policies is dependent on solid research and evidence. Research which goes beyond the traditional focus on individual level (compositional) socio-demographic factors and considers geographic/regional differences in dietary patterns in England may aid in the development of such novel, localised policies.

2.7.2 Geographic/Regional differences in dietary consumption in England

Inequalities in diet and obesity have been observed across the regions of England (DEFRA, 2017; Moon et al, 2007; Morris et al, 2016; NHS Digital, 2017; Roberts, 2014; Scarborough and Allender, 2008; Tiffin and Salois, 2012). However, very few studies have systematically explored socio-demographic and geographic differences in dietary patterns in England (Macintyre, Maciver and Sooman, 1993; Tweney et al, 2017). Although not focused on dietary patterns, Scarborough and Allender (2008) found that over the 1993-2004 period, the prevalence of overweight and obesity plateaued for women in the South of England, whereas the prevalence (of overweight and obesity) increased significantly for both sexes in the North of England. Moon et al (2007) found the prevalence of obesity to be highest in the West Midlands and in Southern Yorkshire, a finding which was fairly similar to recent national statistics which identified that obesity is more prevalent in the Midlands and the North of England than in the South of England (NHS Digital, 2017).

Oddy (2003) (through the use of household expenditure data), found that “Londoners” in the South of England purchased more ethnic meals (e.g. Indian, Chinese), fish, fruits, sandwiches, ice-cream, biscuits and cakes than any other region in England. In 1999, the purchase of ethnic meals in London was said to be approximately 90 grams per week compared to only 17 grams in Yorkshire and the Humber, in the North of England. Although no clear or statistically significant relationship has been found between ethnicity, diet and related illnesses such as obesity (Higgins, Nazroo and Brown, 2019; Moon et al, 2007; National Obesity Observatory, 2011); the higher purchase of ethnic meals in London is plausible, given that approximately 40% of Londoners identify with the Asian, Black, Mixed or Other ethnic group (ONS, 2018a). Roberts (2014) noted that besides variations in individual SEP, mean daily consumption of fresh fruits and vegetables varied by region, with adults in London and the South of England more likely to meet the national

5-A-Day target compared to those in the North of England. In addition, adults in the least deprived quintile were more likely to achieve the 5-A-Day target compared to those in the most deprived quintiles. These findings from Roberts (2014), highlight a possible association between fruit and vegetable consumption, region and deprivation quintiles where relatively poorer diets occurred in the more deprived areas of the North compared to the South of England. Further disaggregation of the data revealed that men in the North West, Yorkshire and the Humber, and women in the West Midlands were least likely to consume the recommended daily portions of fruits and vegetables. Although Roberts (2014) only assessed fruit and vegetable consumption, the study's findings could still help to explain the high prevalence of obesity observed in the Midlands/West Midlands and the North of England (Moon et al, 2007; NHS Digital, 2017; Scarborough and Allender, 2008). These findings also point to the English North-South divide (observed from as early as the 19th century), in which persons in the North of England are said to have relatively poorer diets and worse health outcomes than those in the South of England (Bambra, 2016).

Whichelow, Erzinclioglu and Cox (1991) noted the existence of a clear North-South divide in diet in England/UK. The study (published over 20 years ago) found that more chips, processed meats and fried foods, and fewer fruits and vegetables were consumed in the northern part of England, Scotland and Wales, than in the South East of England. Similarly, Billson, Pryer and Nichols (1999) noted that fewer fruits and vegetables were consumed in the North of England than in London and the South East of England. More recently, Morris et al (2016) seemed to add further credence to the long-standing North-South divide in health observed in England. The study explored the dietary pattern of women using data from the UK Women's Cohort Study and found that the least healthy and monotonous diets were consumed in the North West of England, whereas the highest proportion of vegetarian diets were consumed in Greater London (Morris et al, 2016). However, the findings from this cohort study should be interpreted with caution because the sample was not representative of the English population. The sample consisted of predominantly vegetarian, middle aged, middle class, white women who volunteered to be a part of the study during the late 1990s. Nonetheless, Morris et al (2016) is currently one of the few studies which has investigated the differences in dietary patterns specifically across regions of England as opposed to the wider UK.

As previously mentioned, diet-related studies tend to focus on the more compositional factors, which emphasise who individuals are and their relationship to diet, rather than where persons live (contextual factors) and their association with diet (Brug, 2008; Moon et al, 2007). Macintyre, Maciver and Sooman (1993) highlighted a number of reasons that could be hindering researchers from considering contextual factors. These include: i) the unavailability of data which are of interest to researchers, ii) the geographic scale at which data are available, iii) the different area boundaries used by data collectors and iv) the way in which data are formatted. The authors also noted the tendency for researchers to rely on past observations (e.g. North-South divide in England) and make general assumptions about the diet and general health of persons living in particular areas, without conducting the research necessary to substantiate these claims. Moreover, Smith et al (2021) noted that although public health is ultimately the responsibility of local authorities (underneath the Health and Social Care Act, 2012), very little research has gone into the modelling of local level variations in dietary patterns in England. The major challenge faced by researchers interested in conducting this type of research is that diet-related data are primarily obtained from national Government funded surveys such as the Health Survey for England (HSE), which are not powered to capture more local variation (Pryce et al, 2020). Thus, the poor availability of diet-related data, especially below the regional level, could possibly mask local inequalities in diet and health (Twigg, Moon and Jones, 2000). It has therefore been recommended that researchers (as far as possible) make the best use of the data currently available to systematically compare the health of persons in different areas/regions, rather than rely on “what everyone knows” (Macintyre, Maciver and Sooman, 1993 p. 230). The issue of scale (or the modifiable area unit problem, MAUP) must also be acknowledged, as statistical results are likely to differ depending on the geographical scale of the data (Openshaw, 1984; Prouse, Ramos, Grant and Radice, 2014). Thus, the investigation of regional and sub-regional variations in diet could reveal fewer differences than expected or different patterns of diet to what was initially assumed. Such findings could go undetected if not properly researched (Macintyre, Maciver and Sooman, 1993).

Despite data-related, and other challenges, it is clear that more nationally representative studies which explore possible socio-demographic (compositional factors) and geographic/regional (contextual factors) differences in dietary patterns (not obesity) in England are needed. This type of assessment could provide policy makers with more detailed, empirical information about possible diet and health needs of persons from different SEP groups in different sections of the country (Curtis and Rees Jones, 1998). Additionally, such information could help to inform social policy as well as more localised interventions aimed at improving the diets of individuals, especially those in lower SEP groups (Smith, Heppenstall and Campbell, 2021).

2.8 Final remarks and summary

Despite countless studies and interventions worldwide, there has been a drastic shift in global dietary patterns (commonly referred to as the nutrition transition), as evidenced by the reduced consumption of minimally processed foods such as fruits and vegetables, in favour for more ultra-processed fats (Popkin, 1993; Newton et al, 2015). Sub-optimal diet now accounts for more death and disease than physical inactivity, excessive alcohol consumption and smoking combined, ultimately contributing to rising obesity rates and increasing inequalities in diet and health (Afshin et al, 2019; BMA, 2016; Forouzanfar et al, 2015; Malhotra, Noakes and Phinney, 2015; Newton et al, 2015).

The diet of individuals is influenced by a host of demographic, social, cultural, economic and political factors, as well as multiple contextual environments, beyond the residential area (such as family, friends, school, the home food environment, the workplace, places of leisure etc) (Clary, Matthews and Kestens, 2017). Inequalities in diet/health are a result of a complex mix of individual-level socio-demographic and wider environmental factors, which vary over time and space and at various geographic scales (Cummins, Curtis, Diez-Roux and Macintyre, 2007). There are many factors operating at the higher global, national and regional scale which influence the diet of individuals at the local level. Some of these include, trade liberalisation, increased urbanisation, increased FDI, changes in global food supply chains/systems, technological advances, increased global marketing strategies, local food cultures and norms and overall changes in the natural, physical and social environment (Sproesser et al, 2019). The complex and collective inter-play between compositional and contextual factors makes it difficult for people to be separated from place and makes it impossible for any particular study/survey to capture all the potential pathways/mechanisms which connect area to diet, especially at the national level (Dearden, Lloyd and Green, 2020; Macintyre, Ellaway and Cummins, 2002). It is therefore recommended that studies systematically and theoretically conceptualise the specific features of physical and social contextual environment likely to influence individual dietary patterns and the complex causal pathways which link area to diet (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Kestens et al, 2010). The adoption of this collective or relational mindset/approach to the study of diet could provide greater insight into the processes and reasons for diet inequalities, changes over time and the areas or population sub-groups in which to focus more targeted interventions. However, one of the major findings garnered from the literature reviewed, was the relatively poor availability of diet-related and true area-level data and the dearth of studies which have investigated dietary patterns in England specifically. In terms of data-related challenges, there is still no formal consensus on the exact features of the physical and social environment which influence diet and health, which makes it difficult for the collectors of secondary survey data to capture "true" area data, especially at the national level (Cummins, Macintyre, Davidson and Ellaway, 2005, Dearden, Lloyd and Green, 2020; Frohlich, 2000; Mollborn, Lawrence and Saint Onge, 2021). As a result, researchers are usually constrained by the lack of true area data and are oftentimes forced to use pre-defined geographic units (e.g. regions, municipalities, census tracts) captured within “off the shelf” routine secondary cross-sectional surveys and censuses, which may not be in keeping with research interests and initial hypotheses (Cummins, Macintyre, Davidson and Ellaway, 2005; Frohlich, 2000; Macintyre, Ellaway and Cummins, 2002). In addition, diet-related policy as well as academic and non-academic diet-related studies in the UK to date, have tended to focus on health outcomes such as obesity and its prevalence and association with diet, especially amongst children. Cross-sectional surveys such as the HSE have been used by policy makers to annually assess the health of the nation. However, emphasis is usually placed on the prevalence of obesity rather than individual dietary patterns. Whilst it has been widely acknowledged that diet varies by age, sex and SEP, most diet-related studies tend not to disaggregate data by English region but instead focus on the UK or make generally negative assessments of the Scottish diet in comparison to the English diet.

Besides world famous dishes such as English Fish and Chips, there have been noted dietary and health differences between the administrative regions of England from as far back as the 19th century (Bambra, 2016). Some of these regionally distinct foods include the Yorkshire pudding from the Yorkshire and Humber region and the Cornish pasty, traditionally aligned to Cornwall in the South West region (Cornish Pasty Association, 2021; Vogler, 2020). Although England boasts several distinctly traditional “regional foods”, the English diet has transitioned from a diet of mostly meat pies, cooked potatoes and green vegetables (Riley, 2010), to a more “Western” type dietary pattern, high in sugar sweetened beverages (SSBs), saturated fats, free sugars, salt and ultra-processed foods and low in fruits and vegetables (Newton et al, 2015). Such changes could potentially threaten the existence of traditional country/region-specific dietary patterns and highlights the need for more empirical studies, especially in England where the study of possible region-specific and small area dietary patterns remains unclear and underexplored (Emmett, Tweney, Golding and Taylor, 2017; Smith et al, 2021). The few studies which have investigated geographic variations in diets in England were not nationally representative, constrained to specific cohorts/groups within the population (e.g. middle-aged women) or used past observations (e.g. North and South divide in England) to make general assumptions, without systematically comparing the diets of persons in the different geographic regions of England. Additional studies which explore the dietary pattern of individuals (at the national level) and model regional and small area variations in dietary patterns in England, could help to verify already established trends or could yield unexpected differences which would not have been detected if not investigated. This type of exploratory analysis could help to identify the types of foods people are actually eating (rather than what they should be eating), highlight possible foods to be flagged for further nutritional intervention, identify possible administrative regions/small areas for more detailed investigation in future studies and provide policymakers with information (which they can visualise and easily comprehend) which could go towards the creation of more impactful policy actions (Smith, Heppenstall and Campbell, 2021).

Based on the identified gaps in the literature, it is clear that more studies are needed to assess the overall dietary patterns of adults in England and possible differences at the regional and small area level. This project aims to improve upon current evidence by assessing the dietary patterns (not obesity or related health outcomes) of persons aged 16-64 years, covering both sexes, across the administrative regions and small areas of England. This will be achieved through three main research questions:

1. To what extent does fruit and vegetable consumption (fruit and vegetable portions and 5-A-Day achievement) vary by region amongst persons aged 16-64 years in England?

2. What are the dietary patterns of persons aged 16-64 years in England and to what extent do dietary patterns vary regionally?

3. To what extent does estimated achievement of 5-A-Day vary across smaller areas of England?

The project's focus on persons aged 16-64 years (the traditional working age adult population, ONS (2017a)), was based on the limited studies which have investigated how diets vary geographically amongst adults in England (Smith et al, 2021) and the identification of working age adults aged 16-24 years, as the age group least likely to consume the recommended five portions of fruits and vegetables daily (NHS Digital, 2017). In addition, the project recognises the importance of adults, who are potential agents of change with the ability to positively influence the younger generation (Musingarimi, 2009).

In order to adopt a more relational view to the study of diet, the project acknowledges that diet is associated with many complex mechanisms operating at several spatial levels/scales (Curtis and Rees Jones, 1998). In light of the MAUP, the project recognises that statistical results obtained are likely to vary, depending on the geographic scale at which data are measured and analysed (Openshaw, 1984; Tuson et al, 2019). As such, a multi-scale/level approach will be used within the current project to assess diets in England at the higher national, regional and finally the local level, using aggregate pre-defined geographic measures (GOR, MSOAs). The project's selection and sole use of GOR and MSOA as geographic measures, could be viewed as a limitation, because the exploration of additional area attributes/features such as the family and home food environment, social capital, the built environment (among others) and the examination and the causal processes which link place and diet/health are not within the scope of the current project. However, based on the dearth of studies in England and the generally poor availability of true area-level data, this project's primary aim is to fill a current research gap by assessing the dietary patterns of individuals (not households) and possible geographic variations in England (using pre-defined geographic units, GOR and MSOA). The focus at this stage is therefore to assess *where* there is variance rather than *why* there is variance at different geographic levels/scales (Frohlich, 2000). Gaining insight into what persons are eating and how these dietary patterns are spatially structured is a necessary first step, which should assist future researchers in further understanding and conceptualising the reason(s) for these patterns and the mechanisms/causal pathways which connect area to diet (Dearden, Lloyd and Catney, 2019).

The project's first research question (to be addressed in the first empirical chapter, Chapter 4) will focus on regional variations in fruit and vegetable consumption amongst persons aged 16-64 years in England. The initial focus on fruit and vegetable consumption is because it is a relatively simple and well-established national and global summary measure of a healthy diet, which policymakers, local authorities and members of the public can easily understand and relate to. Subsequently, the second research question (to be presented in Chapter 5) aims to add to the current body of literature by examining the overall dietary pattern of working age adults in England and investigating whether there are region-specific dietary patterns or no regional variations whatsoever. With the exception of Smith et al (2021), the modelling of small area level variations in diet in England is a relatively unexplored topic. This is partially due to the poor availability of diet-related data at the small area level (Pryce et al, 2020; Smith, Heppenstall and Campbell, 2021; Twigg, Moon and Jones, 2000). The project's third research question (to be addressed in the final empirical chapter, Chapter 6) aims to fill this specific research gap by utilising a combination of advanced quantitative statistical techniques in order to assess possible variations in diet (specifically 5-A-Day achievement) at a more local level. This type of analysis could provide local authorities (ultimately responsible for public health) with insight into food consumption patterns across smaller administrative areas of England and could help to identify areas possibly in need of prioritisation and more targeted interventions. Overall, it is expected that this project will add to the limited information currently available on the topic and could possibly help to inform more culturally appropriate, localised interventions aimed at improving diets and the long-term health of individuals across England, especially those in lower SEP groups.

Chapter 3 Data Landscape

Sections of this chapter have been published as Campbell et al (2020).
Campbell, M., Smith, D., Baird, J., Vogel, C. and Moon, E. G. (2020) 'A critical review of diet-related surveys in England, 1970-2018', *Archives of Public Health*, 78(1), pp. 66. doi:10.1186/s13690-020-00447-6

This chapter (Chapter 3) is a secondary data assessment or landscape which details the strengths and weaknesses of the major repeated cross-sectional and longitudinal (quantitative) surveys conducted in England over the last 48 years (1970–2018) and the surveys (the National Diet and Nutrition Survey-NDNS and the Health Survey for England -HSE) most aligned to the project's aim. This chapter has been organised into three main sections. Section 3.1 details the two-stage review process used to identify, assess and select the surveys most suitable for use within the current project. Section 3.2 gives the rationale for selecting the NDNS and HSE over the other surveys assessed. Finally, Section 3.3 provides a background to the NDNS and HSE inclusive of the survey design, methodology and the dietary assessment methods used in each.

3.1 Two-stage data landscape and survey assessment process

A two-stage review process, similar to that used by Campbell et al (2020), was used to identify, assess and select the surveys most suitable for analysis within the current project. Stage one (Figure 3.1) involved the identification of all major repeated cross-sectional and longitudinal health, diet-related surveys, conducted in England over the period from January 1970 to December 2018. This 48-year assessment period was thought to be an adequate time span in which a sufficient number of longstanding and current survey datasets (especially longitudinal surveys) could be captured. The UK Data Service, the Medical Research Council (MRC) Cohort Directory, Cohort and Longitudinal Studies Enhancement Resources (CLOSER) and the Consumer Data Research Centre's (CDRC) online directory were used to identify surveys. These four online directories were selected because they provided a comprehensive list of all surveys (i.e. surveys open for secondary data analysis) conducted within the UK over time, a summary of the survey design, variables captured within datasets, links to survey documentation and where relevant, the institutions (academic and research) ultimately responsible for managing and disseminating data.

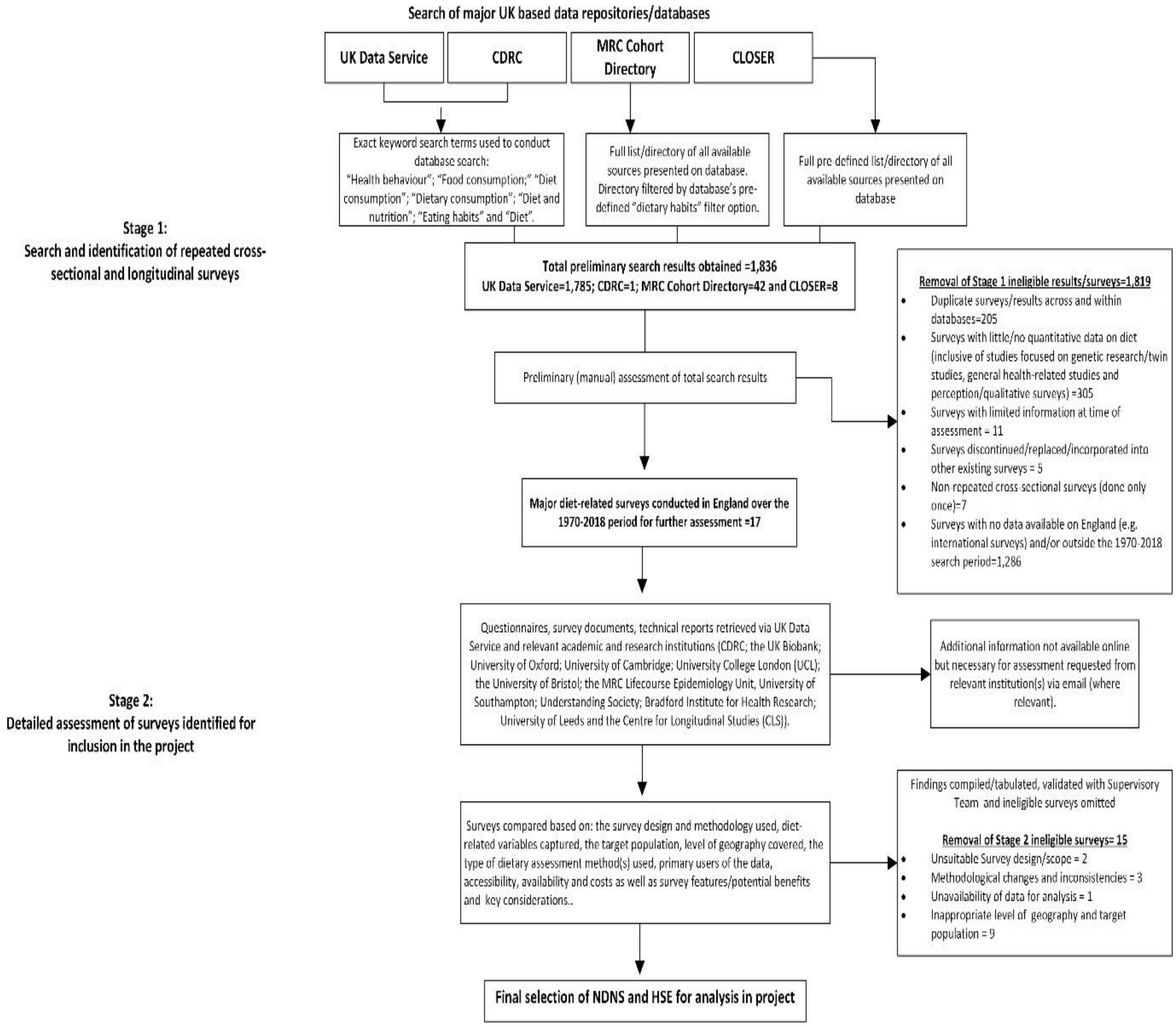


Figure 3.1: Two-stage research and review process used to identify and assess surveys

The search strategy used to identify initial survey results varied, based on the inherent structure of each of the four databases (Figure 3.1). An exact keyword search for “Health behaviour”; “Food consumption;” “Diet consumption”; “Dietary consumption”; “Diet and nutrition”; “Eating habits” and “Diet” was conducted within the UK Data Service and the CDRC databases. This was done to ensure that a wide variety of surveys, especially those not directly associated with diet, but which captured aspects of diet-related behaviours, would have been initially identified. The MRC Cohort Directory presented a full list of all major cohort (longitudinal) studies conducted in the UK, from which diet-related surveys relevant to this review were identified using the database’s pre-defined “Dietary Habits” topic filter option. CLOSER was strictly focused on eight longitudinal surveys (the Hertfordshire Cohort Study, 1946 and 1970 British Cohort Study, 1958 Child Development Study, Avon Longitudinal Study of Parents and Children, Southampton Women’s Survey, Understanding Society and Millennium Cohort Study) which captured persons born throughout the 20th and 21st centuries. All eight affiliated surveys were listed in the “Our

Studies" section of the CLOSER database, which meant there was no need to filter or conduct any keyword searches. In total, 1,836 preliminary results were obtained across the four databases, of which 97% (1,785 results) were from the UK Data Service.

Preliminary search results obtained during Stage one were manually assessed to filter out duplicates (205 of the 1,836 total preliminary results) and ineligible surveys (1,614 out of the 1,836 total preliminary results) (Figure 3.1). Ineligible surveys included: surveys which were replaced/incorporated into other existing surveys (such as the Expenditure and Food Survey (EFS) which was replaced in 2008 by the current Living Costs and Food Survey (LCFS)); non-repeated cross-sectional surveys conducted only once; surveys which although diet-related, had no data for England (e.g. international studies or studies focused on a particular UK constituent country such as Scotland only); surveys which fell outside the 1970–2018 search period; surveys which had little or no quantitative diet-related data (e.g. qualitative/perception studies, gene/twin studies, general health studies with no diet-related data) and surveys which could not have been properly assessed due to limited documentation at the time of assessment. These processes left 17 potentially useful surveys for further screening in Stage two.

In Stage two (Figure 3.1), questionnaires, documents and technical reports for the 17 surveys were retrieved online from the UK Data Service and the official websites of the responsible academic and research institutions. Academic and research institutions included: the CDRC; UK Biobank; University of Oxford; University of Cambridge; University College London (UCL); University of Bristol; the MRC Lifecourse Epidemiology Unit, University of Southampton; Understanding Society; the Bradford Institute for Health Research; University of Leeds and the Centre for Longitudinal Studies (CLS). Where necessary, follow-up emails were sent directly to the UK Data Service and institutions to garner additional information not available on official websites. Documents (inclusive of questionnaires used across survey waves/periods) received either from websites or via email were thoroughly reviewed in order to identify and assess the survey(s) most suitable for use within the current project. Surveys were assessed based on: the survey design and methodology used, diet-related questions/variables captured, the target population, level of geography covered, the type of dietary assessment method(s) used, primary users of the data, accessibility, availability, data costs (where applicable), as well as the key survey features/potential benefits and key considerations for each survey. This information was compiled, tabulated and presented to the full Supervisory team for preliminary review and feedback (detailed assessment presented in Appendix A).

3.2 Rationale for final selection of NDNS and HSE

Of the 17 surveys assessed, the British Cohort Study 1970 (BCS70), Millennium Cohort Study (MCS), Understanding Society, Southampton Women's Survey (SWS), Born in Bradford (BiB), Avon Longitudinal Study of Parents and Children (ALSPAC), UK Women's Cohort Study (UKWCS), Whitehall II, European Prospective Investigation into Cancer and Nutrition **(**EPIC), UK Biobank, British Regional Heart Study (BRHS), British Women's Heart and Health Study (BWHHS), Living Costs and Food Survey (LCFS), Food and You and the Active Lives Survey (ALS) were the 15 surveys excluded from the current project.

Prior to the assessment, it was initially assumed that longitudinal surveys would have been ideal, based on their large sample sizes, the inclusion of diet-related data and the ability to use such data to track changes over time and possibly across the lifecourse. However, all 12 longitudinal studies identified were excluded due to the level of geography covered, the type of questions posed, the unavailability of data for some survey waves and the overall target population captured (Appendix A). A key feature of the BCS70 was its large sample size and the availability of detailed food diary data for the 1986 and 2016 survey wave. However, food diary data for both these waves were still being processed at the time of assessment and the expected date of release is yet to be determined. Diet-related questions posed in the MCS and Understanding Society were limited and inconsistent across survey waves, and SWS, BiB, ALSPAC, UKWCS, Whitehall II, EPIC, UK Biobank, BRHS and BWHHS were not representative of the general English population. BiB focused on Bradford in the North of England and SWS and ALSPAC were limited to Southampton and Bristol in South East and South West England, respectively. Besides the study's focus on middle-aged persons, EPIC Norfolk/Oxford was also limited in terms of its focus on the geographical areas of Norwich, Greater Manchester, Oxfordshire and Buckinghamshire. Data captured in BRHS and BWHHS were not representative of the English population and were limited to middle-aged males and females from only 24 and 23 towns (respectively) across Scotland, Wales and England. Although the UK Biobank followed 500,000 persons across the UK, the survey was focused on middle-aged persons aged 40-69 years.

With regards to the repeated cross-sectional surveys assessed, the LCFS was excluded because it captured household food purchasing and expenditure data, and not foods actually consumed by individuals (the focus of this project). The Food and You survey was omitted as significant changes were made to diet-related questions posed over the survey waves, which would have hindered the pooling of data and the assessment of diet consumption in England. Initially, the ALS' large sample size (approximately 198,000 persons) and it being the only survey assessed which captured fruit and vegetable consumption data at the local authority level, were key features of relevance to the current project. However, the survey was eventually excluded as it overestimated fruit and vegetable consumption/5-A-Day achievement (55% achievement), when compared to the NDNS and HSE (PHE, 2020a). This could be because the ALS' focus was not on diet, but on sport and physical activity in England.

Despite their limitations/key considerations (outlined in Appendix A), the NDNS and HSE were the surveys finally assessed as being the most suitable for use within the current project. The detailed diet-related data captured in the NDNS and the HSE's large annual sample were some of the major benefits to the project. The NDNS' consistent/annual use of the food diary assessment method across the eight survey waves, was a feature which set it apart from the remainder of the surveys which used FFQ, shortened dietary screener instruments as in the HSE and ALS, 24-hour diet recalls or a combination of these methods across the different survey waves. The food diary method facilitated the collection of detailed information, including nutrient content and portion size of all foods and beverages actually consumed by individuals, over a four-day period. Unlike the HSE (which captures fruit and vegetable consumption), data from the NDNS was especially useful to the project's second research question, which examines regional variations in dietary patterns in England. However, a noteworthy limitation was the NDNS' annual sample size of approximately 1,000 persons annually, which although customary for surveys which use the food diary assessment method (Campbell et al, 2020), was much smaller than the HSE's annual target sample size of approximately 10,000 persons. The pooling of data across the eight NDNS survey waves (to be discussed in Chapter 4, Section 4.2.1.1) was a possible workaround to this limitation.

The project's selection of both surveys was also beneficial as it provided the opportunity for a descriptive comparison to be made, to assess similarities or possible discrepancies in fruit and vegetable data captured, given the NDNS' much lower sample size (compared to the HSE) and the different dietary assessment method used in each. In addition, because the NDNS and HSE are currently used by the Government to monitor the overall diet and nutritional status of the population and key diet-related targets such as the national 5-A-Day target (Bates et al, 2012), the project's use of both surveys instead of one, could help to strengthen project findings and provide a more comprehensive picture of fruit and vegetable consumption, dietary patterns and possible geographic/regional inequalities in England.

The following sections (Section 3.3.1 and Section 3.3.2) provide a narrative overview of the NDNS and HSE (respectively), inclusive of the survey design, methodology and the dietary assessment methods used in each.

3.3 Overview of the NDNS and HSE

3.3.1 NDNS

The NDNS was formally established in 1992 and initially consisted of four separate, age-specific cross-sectional surveys, which captured food consumption amongst persons aged 18 months and older (across the 1992–2001 period) and 19–64 years (in 2000–2001). With the introduction of the NDNS rolling programme (NDNS RP) in 2008, the survey changed from a series of ad-hoc age-group specific surveys, to an annual repeated cross-sectional survey for all age groups. The survey aimed to achieve a representative sample of individuals (1,000 persons: 500 children and 500 adults annually), aged 18 months and over, living within private households (exclusive of pregnant/lactating women) in the UK. Although nine waves of the NDNS rolling programme survey have been conducted to date, only waves 1-8 (2008-16) were available at the time of this project.

For all eight waves, a multi-stage stratified random sample of addresses was drawn from the Postcode Address File (PAF). The PAF was an electronic list (generated by the Royal Mail) of all addresses or delivery points in the UK. Addresses were clustered into primary sampling units (PSUs), which are small geographical areas, based on postcode sectors, randomly selected across the UK. Addresses were randomly selected from each PSU over the April 2008 -March 2011 (waves 1-4), April 2012 - March 2014 (waves 5-6) and April 2014-March 2016 period (waves 7-8). GOR and PSU are the primary geographic identifiers available in the NDNS.

Demographic details (e.g. age, sex, ethnicity of participants), information on shopping and food preparation practices, household facilities and the Household Reference Person (HRP) were captured via Computer Assisted Personal Interviewing (CAPI). The HRP was defined as the person (in whose name the property was owned or rented) with the highest income within the household. The eldest person in the household was selected as the HRP if there was more than householder, and they had an equal income. The identification of the HRP was primarily to ascertain housing tenure and to collect information regarding the HRP's work status at the time of the interview. This was done in order determine the socio-economic classification (the National Statistics Socio-economic Classification (NS-SEC)) grouping of the household.

The (unweighted) food diary dietary assessment method was used to capture detailed information on all foods and beverages actually consumed by participants, over a four-day period. A key feature of the food diary method used was that recording of data was done at the time of consumption, which helped to reduce recall bias or the reliance on memory and improved the quality and accuracy of data collected (Shim, Oh and Kim, 2014). Participants were trained to complete food diaries and estimated portion sizes using generic household measures (e.g. two thick slices of bread or four tablespoons of peas) or weights from labels (e.g. 420g tin of baked beans or 330ml can of lemonade). Diaries included pictures of 10 of the most frequently consumed foods, which made it easier for persons to describe portion sizes in a user-friendly manner and helped to further reduce participant burden. Survey interviewers made follow-up checks (in person or via telephone) on the second or third day of recording to review diaries, check for and fill in missing details, provide guidance on how to improve recording and motivate participants to continue recording.

In order to capture both weekday and weekend consumption, the food diary recording period during wave one of the survey was set to start on either a Thursday, Friday or Saturday, and included both weekend days (Saturday and Sunday). However, this resulted in an over-representation of weekend days and an under-representation of weekdays (especially Wednesday). This also introduced bias in reported food consumption and nutrient intake, due to possible variations in the type of foods consumed by day of the week. In wave 2 (and onwards), the four-day recording period was generated using a software programme on the interviewer's laptop and based on the day of the first CAPI. This was done to ensure that all days of the week were equally represented, as far as possible. Persons who completed at least three diary recording days were classified as fully productive and eligible for inclusion in the final sample.

Food diaries from fully productive persons were coded and entered into the Medical Research Council Elsie Widdowson Laboratory's (MRC EWL) dietary assessment system, DINO (Diet In Nutrients Out), an all-in-one dietary recording and analysis written in Microsoft Access. All foods and beverages recorded in diaries were matched with corresponding food and portion codes from DINO. Composite food items which had several components (e.g. sandwiches) and homemade recipes were assigned codes for each individual component to properly estimate the total amount of individual foods consumed. Additional details on the survey design and the methodology used in the NDNS have been described elsewhere (MRC Elsie Widdowson Laboratory and NatCen Social Research, 2018)

3.3.2 HSE

The Health Survey for England (HSE) is an annual, repeated cross-sectional survey which captures the general health status of approximately 10,000 individuals (8,000 adults and 2,000 children), living in private residential households (exclusive of persons living in institutions) in England. Since its inception in 1991, the survey has collected information on the adult population, which the survey (like this project) defines as persons aged 16 years and older. The HSE mirrors the

NDNS in terms of survey design in that each year, a multi-stage stratified sampling strategy is used. Addresses are drawn from the PAF (over the January to December period of each survey year), clustered, and then randomly selected from each PSU. Selected addresses are sent background information related to the survey, subsequent checks are made to assess the number of households/dwelling units and the necessary household selections made. Similar to the NDNS, GOR and PSU are the primary geographic identifiers available in the HSE.

Survey data were captured via face-to-face interviews and nurse visits. Core topics covered annually typically include: general health, longstanding illness, social care and lifestyle/ behavioural factors associated with health (e.g. fruit and vegetable consumption and BMI/height and weight measurements). However, depending on the topic(s) prioritised for a particular survey year, additional questions may be included, or previously asked questions may be omitted entirely or for a particular population subgroup (e.g. adults 16 years and over).

The HSE is currently the primary survey used to monitor the Government's national 5-A-Day target (Bates et al, 2012). Prior to 2009, a Food Frequency Questionnaire (FFQ) was the dietary assessment method used to capture the frequency of consumption for at least 12 food items. Food categories in the FFQ included: fruits and vegetables, cheese, red and white meat, fried food, sweets, soft/fizzy drinks, among others. However, since 2009, the HSE only captures fruit and vegetable consumption via a (single) 24-hour shortened dietary instrument/screener. Additional details on the survey design and the methodology used in the HSE have been described elsewhere (NatCen Social Research and University College of London, 2017).

Chapter 4 To what extent does fruit and vegetable consumption (fruit and vegetable portions and 5-A-Day achievement) vary by region amongst persons aged 16-64 years in England?

4.1 Introduction/Context

Fruits and vegetables are a major part of what typically constitutes a “healthy diet” (WHO, 2015a). Although there is still no universally accepted definition of diet (Global Panel on Agriculture and Food Systems for Nutrition, 2016; Pem and Jeewon, 2015), the Eatwell Guide (England’s national dietary guidelines) recommends the daily consumption of at least five portions (400 grams or 5-A-Day) of fruits and vegetables, in keeping with WHO international recommendations (PHE, 2018b). Despite these recommendations and numerous diet-related policies and interventions over the years, such as the national Change4Life (5-A-Day) campaign and the recently launched Better Health campaign in 2020, sub-optimal diet continues to be the leading cause of death, disability and increasing health inequalities in England (Newton et al, 2015).

Approximately 16,000 deaths in England have been attributed to sub-optimal fruit and vegetable consumption, with costs to the healthcare sector estimated at £706 million in 2017 (Pinho-Gomes, Knight, Critchley and Pennington, 2021). In 2017, only 29% of adults in England consumed the recommended five portions of fruits and vegetables daily (NHS Digital, 2019). Thus, over 70% of adults in England fail to meet the national 5-A-Day target. In addition, it is estimated that more than 55 million additional portions of fruits and vegetables would need to be consumed daily in England if all adults were to meet the national 5-A-Day target (Pinho-Gomes, Knight, Critchley and Pennington, 2021). These findings add credence to the severity of sub-optimal diet, specifically fruit and vegetable consumption in England and the urgent need for more evidence-based diet-related policies and interventions to tackle this issue.

It has been widely accepted that there are differences in diet by age, sex and SEP (Imamura et al, 2015; Roberts, Cavill, Hancock and Rutter, 2013). However, the findings which emerged from the literature review (presented in Chapter 2) highlighted the limited empirical studies which have explored regional/geographic differences in diet in England. Morris et al (2016) was one of the few studies which attempted to explore GOR variations in diet (inclusive of fruit and vegetable consumption) in England/UK. However, the study was not representative of the English population as participants were mostly vegetarian, middle-aged women. Given the dearth of literature on the geography of diets in England, researchers have been implored to systematically compare the diet/health of persons in different regions, instead of over-relying on past assumptions (such as the stereotypical North-South divide in England) without substantial evidence (Macintyre, Maciver and Sooman, 1993). This chapter (the project's first analytical chapter), aims to contribute to the literature by exploring fruit and vegetable consumption/5-A-day achievement in England and potential differences at the regional level. In light of the recent association made between diet/fruit and vegetable consumption and COVID-19 (Butler and Barrientos, 2020), the findings of this project could be used to identify regions or specific groups within the population, possibly in need of more targeted diet-related policies and interventions. Also, the official launch of the national Better Health campaign in July 2020, signalled the Government's renewed focus on not only childhood obesity, but on improving the diet/health of adults in England, in response to the COVID-19 pandemic (PHE, 2020b). Thus, the project's focus on fruit and vegetable consumption/5-A-Day (a simple yet major diet and nutrition-related target in England) within the working age adult population is timely, beneficial and aligned to current national priorities. The following section (Section 4.2) details the sources of data and the method of analysis used within this chapter to address the first research question.

4.2 Method

4.2.1 Data sources, access and preparation

The NDNS and HSE were the two data sources used within this chapter. Background information, inclusive of the survey methodology and the dietary assessment methods used in each, was previously discussed in Section 3.3.1 and 3.3.2. Secondary data ethical approval was received from the University of Southampton's Research Ethics Committee, after which registration with the UK Data Service was completed, and End User Licenses granted for use of the NDNS and HSE data within the current project. Nine waves of the NDNS rolling programme survey have been conducted to date, but only waves 1-8 (2008-16) were available at the time of this project. In order to make a descriptive comparison between the NDNS and the HSE, the corresponding 2008-16 period was also selected for the HSE. Thus, NDNS and HSE data collected prior to 2008, and after 2016, were not analysed in this project.

Finalised SPSS datasets for waves 1-8 (2008-16 period) of the NDNS (MRC Elsie Widdowson Laboratory and NatCen Social Research, 2018) and for the 2008-16 survey years of the HSE

(NatCen Social Research and University College London Department of Epidemiology and Public Health, 2008-2016), were downloaded from the UK Data Service's website in December 2018 and July 2019 respectively. "Individual level" NDNS datasets contained demographic characteristics (e.g. age, sex, ethnicity and marital status), measures of SEP (e.g. highest level of education attained) and geographical region of residence (Government office region-GOR). "Personal level food" NDNS datasets contained information related to the national "5-A-Day" target in addition to all foods, beverages and nutrients consumed by respondents, averaged over the four days in which food diaries were kept. Individual and personal level food datasets for NDNS waves 1-8 were merged (in SPSS version 24, (IBM, 2018)) to create a single dataset for analysis.

"Individual level" datasets for each survey year of the HSE contained information on the socio-demographic characteristics of individuals sampled (age, sex, ethnicity, marital status and SEP measures such as education), region of residence (GOR) and portions of fruits and vegetables consumed over a 24-hour period. Individual datasets for the 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 and 2016 survey years of the HSE were merged (in SPSS version 24 (IBM, 2018)) into a single dataset for analysis.

4.2.1.1 NDNS 2008-16 sample size and preliminary exclusions

Due to the NDNS' relatively small annual sample size (approximately 1,000 persons annually), data from the survey's eight survey waves (2008-16) were pooled to help increase the overall sample size. The total or pooled NDNS sample for waves 1-8, which comprised of only fully productive persons (defined as those who completed three or four diary recording days), was 12,097 persons. Persons who did not complete at least three diary days (2% of persons) were omitted from the final sample/datasets downloaded from the UK Data Service. The sample for waves 1-4 consisted of 6,828 fully productive persons, whereas waves 5-6 and 7-8 consisted of 2,546 and 2,723 fully productive persons, respectively. The overall response rate for fully productive persons was 56% for waves 1-4, compared to 53% in waves 5-6 and a corresponding 53% in waves 7-8.

Overall, 57% of respondents resided in England, compared to 17% in Scotland, 13% in Northern Ireland and 13% in Wales. Forty-one percent (41%) of respondents were children under the age of 16, 48% were aged 16-64 years old (classified by the ONS as being "traditional working age" adults) and the remainder were 65 years and older. The current project focuses on persons (males and females) living in England, aged 16-64 years old (traditional working age adults) (ONS, 2017a). Therefore, persons under the age of 16, those 65 years and older, and persons living outside of England were not included in this project. The exclusion of these groups from the NDNS, reduced the sample size from 12,097 to 3,286 persons (Figure 4.1).

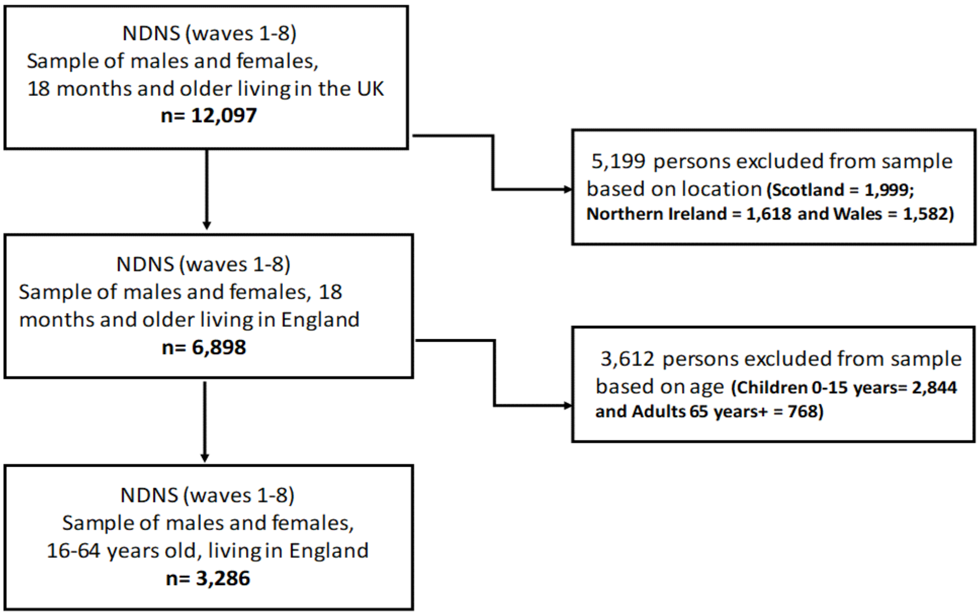


Figure 4.1: Final NDNS sample size for waves 1-8 after age and location-based exclusions

4.2.1.2 HSE 2008-16 sample size and preliminary exclusions

The total/pooled HSE sample for the full 2008-16 investigative period was 111,158 persons, 70% of whom were adults, classified by the survey as persons 16 years and older. The response rate for adults varied across the years, ranging from 58% in 2008 to 55% in 2016. The project's focus on the 16-64 age group meant that similar to the NDNS, persons under the age of 16 years and those 65 years and older were excluded. Additionally, questions related to fruit and vegetable consumption were completely omitted (across all age groups) in the 2012 survey year and were only asked of children aged 2-15 years old in 2014. As a result, data for both years could not have been analysed. Based on the unavailability of fruit and vegetable consumption data for the 2012 and 2014 period, data were analysed for the 2008-11, 2013 and 2015-16 survey years. The exclusion of children and older adults as well as the 2012 and 2014 survey years reduced the HSE's sample size (before the removal of missing cases) from 111,158 to 46,353 persons (Figure 4.2). The exclusion of missing cases as shown in Figure 4.2 will be discussed in Section 4.2.2.

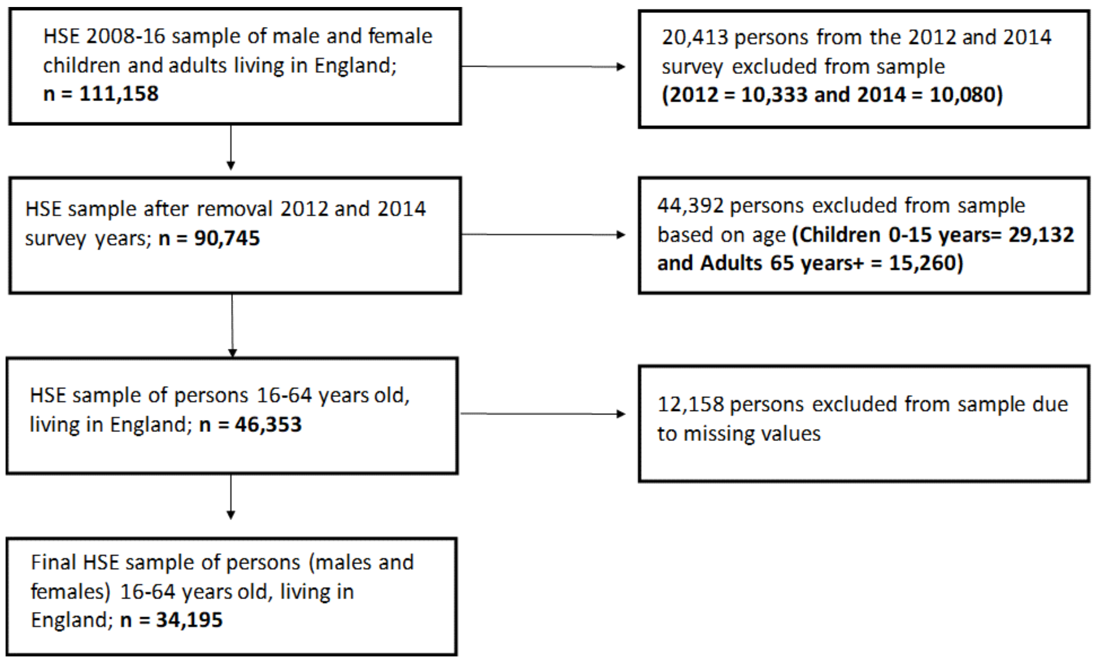


Figure 4.2: Final HSE sample size for the 2008-11, 2013 and 2015-16 period, following the removal of the 2012 and 2014 sample, children and older adults and the removal of missing cases

4.2.1.3 Variable selection and description

Fourteen variables were assessed and deemed relevant to the project's first research question. Given this chapter's focus on fruit and vegetable consumption, the number of portions of fruits; vegetables; and fruits and vegetables consumed by individuals, and the achievement of the national "5-A-Day" target were the four separate outcome variables selected. Independent variables selected included: age, sex, ethnicity, marital status, highest level of education attained, equivalised household income, occupational class (National Statistics Socio-economic Classification (NS-SEC)), Government Office Region (GOR), the Index of Multiple Deprivation (IMD) and Body Mass Index (BMI). A detailed description of these variables has been provided in the following sections (Section 4.2.1.4 – 4.2.1.8).

4.2.1.4 Diet outcome variables

Diet-related variables captured in NDNS and HSE datasets were derived from four-day food diaries and responses to 24-hour fruit and vegetable screener (interview) questions, respectively. In keeping with the Eatwell Guide (see Section 2.7.1), both surveys defined one portion of fruit/vegetable as 80 grams and the total achievement of the "5-A-Day" national fruit and vegetable target as the consumption of at least five portions or 400 grams of fruits and vegetables daily. Unsweetened, 100% fruit juice and smoothies were included in fruit portion calculations and counted as one portion but were not to exceed 150 millilitres (ml) daily. Therefore, consumption of say 300 ml of fruit juice in a day would still only be counted as one portion (150 ml) towards the overall 5-A-Day target. One portion of dried fruit was 30 grams, which was the equivalent of 80 grams of fresh fruit. Beans and pulses (e.g. lentils, kidney beans) counted towards one portion of vegetables per day, regardless of the amount consumed in a day.

Diet-related variables in the HSE included total fruit portions (only), vegetable portions (only) and the combined fruit and vegetable portions consumed by persons over the 24-hour survey period. These variables were captured as three separate continuous variables, labelled as "frtpor", "vegpor" and porftvg" (respectively). Fruit portions ("frtpor") comprised of portions of fruit juice, all sizes of whole fruit (small to large sized), dried fruit, frozen/canned fruit and fruit in composite dishes. "Vegpor" consisted of beans/pulses, salad and vegetables, including those consumed in composite dishes. Total fruit and vegetable portions ("porftvg") was the summation of "frtpor" and "vegpor". All three variables were recoded from continuous into categorical (ordinal) variables (i) 0/Under 1 portion; ii) 1-2 portions; iii) 3-4 portions and iv) 5 portions or more) for further descriptive analysis and comparison with the NDNS. There was no variable in the HSE which specifically captured achievement of the 5-A-Day target. Therefore, the "porftvg" (total fruit and vegetable portion) variable was used to create a new binary (Yes; No) variable of achievement of 5-A-Day, for comparison with the NDNS.

The NDNS captured the achievement of the 5-A-Day target in a binary (Yes; No) variable entitled "Achieve 5." The total number of fruit and vegetable portions (combined) consumed by individuals was captured as a continuous variable called "Totfruitvegportions." The NDNS' "totalfruit" continuous variable was initially thought to be representative of the total number of fruit (only) portions consumed by individuals. However, fruit juice was not included. To adjust for this, NDNS syntax and data documentation guidelines were used to create a total fruit (only) portions variable (MRC Elsie Widdowson Laboratory and NatCen Social Research, 2018). Fruits and dried fruits, (whether raw, cooked, frozen or in composite dishes) were all measured in grams and disaggregated in the original NDNS dataset as "Fruitsg" and "Driedfruitx3", respectively. Both variables were summed and then divided by 80 (grams) to receive portions of fruit/dried fruit consumed. These portions were then added to total fruit juice and fruit smoothie portions (captured as "Fruitjuiceportions" and "SmoothieFruitPortions", respectively) to create the new fruit portions (only) continuous variable.

Similarly, NDNS documentation and data guidelines (MRC Elsie Widdowson Laboratory and NatCen Social Research, 2018) were used to create a variable which captured total portions of vegetables only (inclusive of vegetables in composite dishes or those in raw, cooked, frozen or canned form). Tomato puree, beans/pulses, brassicaceae vegetables (e.g. cabbage, broccoli, cauliflower), yellow, red and green vegetables, other vegetables and tomatoes were all measured in grams and disaggregated in datasets as “TompureeX5”, “Beansmax”, “Brassicaeag,” “YellowRedGreeng”, “Othervegg” and “Tomatoesg”, respectively. All six of these variables were summed and then divided by 80 (grams) to receive total portions of vegetables (only) consumed. Newly created fruit (only) portions and vegetable (only) portions variables were summed, and proved to be accurate, as values received corresponded with values from the NDNS’ original “Totfruitvegportions” variable. Fruit (only), vegetable (only), and fruit and vegetable portions were recoded from continuous to categorical (ordinal) variables (i) 0/Under 1 portion; ii) 1-2 portions; iii) 3-4 portions and iv) 5 portions or more), for comparison with the HSE and in preparation for the modelling of data.

These newly created ordinal variables also helped to further disaggregate “achievers” from “non-achievers” of the 5-A-Day target, by capturing the different levels of consumption, ranging from zero, low to moderate to high consumption. This level of disaggregation was possible due to the ready availability of data on fruit and vegetable portions in the HSE and NDNS. This level of detail facilitated a more in-depth exploration of regional differences in the achievement of the 5-A-Day target, portions of fruit and vegetable (combined) consumed, in addition to portions of fruits (only) and vegetables (only) consumed.

4.2.1.5 Demographic variables

Demographic (independent) variables selected were age, sex, ethnicity and marital status. Sex (Male; Female) and ethnicity (White; Non-white) were the only demographic variables used as originally captured within the NDNS and HSE datasets. Over 80% of respondents in the NDNS and HSE were “White”, which meant that the selection of a more disaggregated ethnicity variable (available in both surveys) was unwarranted.

Although persons aged 16 and 17 years are not legally considered adults, the HSE defined adults as persons aged 16 years and over. In addition, persons aged 16-64 years old (the project’s target age group) are classified by the ONS as the “traditional working age adult population” (ONS, 2017a). Since this project aims to contribute to existing literature by focusing on the diets of adults, its focus on the traditional working age adult population and the exclusion of children, meant that age variables in the NDNS and HSE now had a lower age limit of 16 years. Age, which was originally a continuous variable in the NDNS and a categorical variable in the HSE (initially 22 categories), was recoded into a categorical variable with five categories (i) 16-24 years; ii) 25-29 years; iii) 30-39 years; iv) 40-49 years and v) 50-64 years). The recoding of age in both surveys was done to highlight persons at different stages of life (from the youth to pre-retirement stage) and to easily explore possible non-linear differences between younger and older adult groups as it relates to fruit and vegetable consumption. The selection of the 16-24 year age group as the first category was purposefully done and based on previous literature which found this group to be the least likely to achieve the "5-A-Day" target (NHS Digital, 2017; NHS Digital, 2019).

Marital status which was initially a seven-category variable in the NDNS, was recoded into four categories (i) Single/never married; ii) Married and living with your husband or wife/civil partner in a legally recognised civil partnership; iii) Married and separated from your husband or wife/divorced; and iv) Widowed/other). The original NDNS marital status variable was recoded such that married persons were combined with those in legally recognised civil partnerships and divorced persons were combined with legally separated persons. Initially, the survey had a category called "Formerly in a legally recognised civil partnership", in which one individual was identified. This category was merged with widowed and renamed, "Widowed/other" as it was unclear whether this partnership dissolved due to separation, legal dissolution or death. In the HSE, marital status, initially a six-category variable, was recoded into five categories (i) Single; ii) Married, including civil partnership; iii) Separated/Divorced, including from a civil partnership; iv) Widowed, including civil partnership and v) Cohabitees). The reduction from six to five categories was due to the merger of separated and divorced persons. Despite attempts to align (as far as possible) the HSE and NDNS recoded marital status variables, differences in how the data were captured in both surveys hindered this. For instance, the initial HSE variable had a category called "Cohabitees" (persons who lived with a partner but not legally married), which was not included in the NDNS. The unavailability of this category in the NDNS meant that persons sampled who were in this type of relationship, were presumably included in the survey's "Single, never married" category.

4.2.1.6 Socio-economic position (SEP) variables

Socio-economic position (SEP) refers to the social and economic factors that influence a person's social standing within a society (Galobardes et al, 2006). Besides commonly used measures of SEP such as highest level of education attained, equivalised household income and occupational grade, the number of children in the household and the total number of persons in the household, were two additional SEP variables initially considered for inclusion. However, based on current literature, there appeared to be more evidence for the use of equivalised household income, education and occupation as conventional measures of SEP, especially as it relates to their association with fruit and vegetable consumption and overall diet (Acheson, 1998; Berzofsky et al, 2014; Darmon and Drewnowski, 2008; Roos et al, 2000; Tiffin and Salois, 2012). Other studies have found education, occupation and household income to be distinct socio-economic constructs, which need to be analysed collectively (not interchangeably), in order to truly assess the association between SEP and diet (Maguire and Monsivais, 2015; Turrell, Hewitt, Patterson and Oldenburg, 2003). It was based on these findings that all three (equivalised household income, education and occupation) SEP variables were selected for this analysis.

4.2.1.6.1 Highest level of education

Highest level of education, which was initially an eight-category variable in both the NDNS and HSE, was recoded into a six-category variable, respectively (i) No qualifications; ii) GCSE and equivalent qualifications; iii) GCE A level or equivalent qualifications; iv) Higher education, below degree level; v) Degree or equivalent; vi) Still in full-time education/Other foreign qualification/Other). General Certificate of Secondary Education (GCSE) and equivalent qualifications are generally accepted school-leaving qualifications usually completed at age 16. Qualifications in this category included GCSE qualifications, discontinued qualifications equivalent to GCSE such as the General Certificate of Education (GCE) O level and the Certificate of Secondary Education (CSE) qualifications (discontinued in 1988), and other technical/vocational qualifications of an equivalent level to the GCSE, such as Levels 1 and 2 of the National Vocational Qualification (NVQ Level 1 and 2)). Based on how data were captured in the NDNS, persons who were initially categorised as having attained "GCSE grades A-C" were merged with persons who reportedly attained "GCSE grades D-G/Commercial qualifications/apprenticeship" qualifications. Similarly, respondents in the HSE who attained "NVQ1/CSE other grade equivalent" qualifications were combined with those who attained "NVQ2/GCE O Level equivalent" qualifications. The merger of these categories in both surveys was done to form one category representative of GCSE and equivalent qualifications (as described previously) and to ensure that survey categories were aligned as far as possible.

GCE A level or equivalent qualifications are more advanced than GCSE qualifications, usually completed by persons 16-18 years of age and generally recognised as the basis for assessing the suitability of applicants for admission into universities in the UK. Qualifications in this category included GCE A level and other technical/vocational qualifications of an equivalent level such as NVQ Level 3. Higher education, below degree qualifications represented academic and vocational courses considered to be at the foundation level in higher education. Degree or equivalent qualifications represented higher level tertiary academic and vocational qualifications, such as Bachelor's, Master's, Doctoral degrees and equivalent NVQ Level 4 and 5 qualifications. For the purposes of this project, persons "Still in full-time education" and those with "Foreign or other qualifications" were merged and reclassified as "Still in full-time education/Other foreign qualification/Other". This was done as persons still in full-time education had not yet attained their highest level of education and there was not enough information to properly assign the educational level of persons with foreign or other qualifications (Maguire and Monsivais, 2015).

4.2.1.6.2 Equivalised household income

Equivalisation is the process of adjusting household income by taking into account differences in household size, composition and the financial resources of different household types (Galobardes et al, 2006). With this process complete, households with the same equivalised household income are assumed to have a comparable standard of living. Equivalised household income (a continuous variable) was removed from the HSE in 2015 (onwards) following a disclosure review by NHS Digital. As such, HSE data related to this variable could not be included or analysed within this project, based on the inconsistency across the 2008-16 period. On the contrary, equivalised household income was captured as a continuous variable in all eight waves of the NDNS.

The initial continuous variable captured in the NDNS was recoded into a six-category variable (i) Under £20,000; ii) £20,000-£29,999; iii) £30,000-£39,999; iv) £40,000-£49,999; v) £50,000-£99,999 and vi) £100,000 or more) for further analysis. The selection of "Under £20,000" as the first category was based on recent income inequality statistics published by the UK Government, for the 2017-18 period (McGuinness and Harari, 2019). In 2017-18, individuals in the bottom 10% of households had a weekly household income (equivalised gross income) which ranged from approximately £291 (for a couple with no children) to a maximum of £407 (for a couple with two children under age 14). Overall, this amounted to an annual household income of below £20,000. A similar strategy was employed for the calculation of the remaining categories, which were based on the weekly income for households in the middle and top 10%.

A noted advantage of equivalised household income (as captured in the NDNS) was its acknowledgement of the number and age(s) of children and the combined effect this has on the standard of living of households (Anyaegbu, 2010). Thus, the project's exclusion of variables such as the number of children in the household was thought to be further justified, as this has been accounted for, with the inclusion and analysis of equivalised household income.

4.2.1.6.3 Occupational class: National Statistics Socio-economic Classification (NS-SEC)

The NS-SEC is a conceptual measure of employment relations and conditions of occupations (not a measure of skill levels in employment) which has been used in official statistics and surveys since 2001. Each NS-SEC class represents occupational groups with similar employment relations, which are usually captured in eight main groups/categories (Never worked/long term unemployed; Routine; Semi-routine; Lower supervisory and technical; Small employers/own accounts; Intermediate; Lower managerial and professional; Higher managerial and professional)

(Maguire and Monsivais, 2015). Conceptually, the NS-SEC is a nominal rather than an ordinal measure. Nevertheless, it shows the structure of socio-economic positions/class in England/UK, helps to explain possible differences in social behaviour and other social phenomena, and is generally considered to be a good predictor of health and related outcomes (ONS, 2019).

In keeping with guidelines from the Office of National Statistics (ONS, 2019), the NS-SEC occupational grouping assigned to a household in the NDNS and HSE was based on the Household Reference Person (HRP) (see Section 3.3 for the definition of HRP). The NS-SEC was initially eight categories in the NDNS and nine categories in the HSE (the ninth category in the HSE was "Other"). The "NS-SEC8" variable in both surveys was reduced/reclassified to 5 categories: i) Never worked/long-term unemployed; ii) Routine manual and other occupations; iii) Intermediate occupations; iv) Small employers and own account workers and v) Higher managerial, administrative and professional occupations) based on guidelines from the Office of National Statistics (ONS, 2019). Categories were then re-ordered to be directionally consistent, ranging from lowest to highest. "Never worked/long-term unemployed" consisted of full-time students, persons who never worked or those with occupations which were not stated. Persons in the newly created "Routine manual and other occupations" category were those in routine, semi-routine, other and lower supervisory and technical occupations. "Intermediate occupations" was a merger of intermediate, and lower managerial administrative and professional occupations. "Small employers and own account workers" and "Higher managerial, administrative and professional occupations" were the two remaining occupational categories used, based on ONS (2019) guidelines.

4.2.1.7 Body Mass Index (BMI)

Obesity is not within this project's immediate scope. Nevertheless, overweight/obesity (frequently measured by BMI) continues to be the greatest diet-related challenge faced by England and the world at large (WHO, 2021). Approximately 1.9 billion adults worldwide were classified as being overweight/obese in 2016, with global obesity figures almost tripling since 1975 (WHO, 2021). Two-thirds (63%) of adults in England were reportedly overweight/obese in 2018 (NHS Digital, 2020). Obesity is a complex metabolic disease associated with a combination of social, economic, demographic, genetic and environmental factors (Jia et al, 2020). Obesity has been found to increase with age, be more common amongst women than men and obesity rates in England are highest in the most deprived areas (Baker, 2021; NHS Digital, 2020). Obesity rates in England have increased faster than most OECD countries, with rates said to increase by a further 10% over the next 10 years (OECD, 2021). The rise in obesity rates in England and worldwide has been said to parallel the global nutrition transition and the drastic shift from mostly traditional diets (primarily based on fresh minimally processed foods) to Western dietary patterns, comprised of mostly ultra-processed foods, high in saturated fats and free sugars (Sung et al, 2021; Monteiro et al, 2018; Popkin, 2006; WHO, 2021). In response to the COVID-19 pandemic, nutritional changes and rising obesity rates, the English Government launched the Better Health Campaign in July 2020, with the primary aim of getting *currently* overweight and obese adults in England to lose weight by adopting a healthier diet (Theis and White, 2021). The adoption of a diet high in fruits and vegetables and whole grains and low in saturated fats, free sugars and ultra-processed foods (e.g. fizzy drinks, packaged snacks, ready to eat products etc) has been generally recommended to reduce the risk of obesity, additional weight gain and the development of other chronic diseases, (NIH, 2019; Monteiro et al, 2018; WHO, 2021), especially amongst working age adults (the focus of this project) who are at high risk for obesity (Jia et al, 2020). It has been postulated that with increased intakes of fruits and vegetables, their high fibre and satiating properties can potentially contribute to reduced caloric intake, the reduced consumption of energy dense, nutrient poor foods and possible weight management (Nour, Lutze, Grech and Allman-Farinelli, 2018). On the contrary, an unhealthy diet, especially one high in ultra-processed foods, has been significantly associated with increased BMI and the increased likelihood of being obese (Chen and Antonelli, 2020). Studies (cross-sectional and prospective studies) conducted worldwide, inclusive of those carried out in the UK, have noted a significant association between higher BMI and the consumption of ultra-processed foods (Rauber et al, 2021; Sung et al, 2021). Furthermore, there is increasing evidence which suggests that overweight and obese persons (persons with increased BMI) could have the tendency to consume less fibre rich foods such as fruits and vegetables (the measure of diet used in this chapter) and exhibit a liking for more energy dense foods, which creates a vicious cycle between food consumption and the consequences of food consumption (Chen and Antonelli, 2020; Howarth, Saltzman and Roberts, 2001; Mela, 2001). Overweight and obese groups have been found to consume significantly less fruits and vegetables compared to their normal weight counterparts, after adjusting for socio-demographic factors (Heo et al, 2011). However, the complex association between fruit and vegetable consumption and BMI/weight status is still unclear and warrants further investigation in future studies (Charlton et al, 2014).

As good practice and in recognition of the noted association between BMI and diet (inclusive of fruit and vegetable consumption), BMI was considered for inclusion in the current project, because it remains the most widely used indicator of nutritional risk/status and health worldwide (Barao and Forones, 2012, FAO et al, 2021). BMI is calculated by dividing a person's weight in kilograms by their height in metres squared (kg/m^2). Adults with a BMI of 18.5–24.9 kg/m^2 are considered to have a healthy weight, whereas those with a BMI below 18.5 kg/m^2, in the range of

25-29 kg/m ^{2}, 30-40 kg/m ^{2} and over 40 kg/m ^{2} are classified as being underweight, overweight, obese and morbidly obese, respectively. Although BMI is a good measure of excess weight, it is not the most accurate measure of body fat/fat distribution, which is one of its major limitations. Despite this, it continues to be widely used worldwide, as it is a simple, non-invasive and fairly straightforward measure (reliant solely on height and weight measurements), typically used to indicate the weight and nutritional status of adults (WHO, 2021). Based on the current literature, BMI was deemed beneficial to the current project, as it could be used as a proxy (indicator) of diet quality/the nutritional health and risk status of adults (FAO et al, 2021). However, it was recognised that additional variable checks, exploratory analyses and model testing assessments would need to be conducted before including this variable in the final analysis. BMI was initially captured in the NDNS and HSE as a continuous variable but was recoded into a four-category variable (i) Underweight; ii) Normal weight; iii) Overweight; iv) Obese/morbidly obese for further analysis. Section 4.3.7 and Section 4.3.10 presents the results of the bivariate analysis and the models tested with and without the BMI variable (respectively) and provides additional justification for the inclusion of BMI in all final models tested.

4.2.1.8 Geographical variables

Geographical variables of interest included region of residence (measured by GOR) and IMD. The selection of these variables was crucial in order to investigate possible regional differences in fruit and vegetable consumption in England (Research Question 1). In both surveys, GOR was captured as a nine-category variable (i) North East; ii) North West; iii) Yorkshire and the Humber; iv) East Midlands; v) West Midlands; vi) East of England; vii) London; viii) South East; ix) South West), which required no recoding or manipulation (map of GORs presented in Appendix E.1). Similarly, the IMD variable was kept as a five-category variable (i) Least deprived; ii) 2; iii) 3; iv) 4 and v) Most deprived), as originally captured in the NDNS and HSE.

GOR is primarily an administrative, sub-national measure of the geographical regions across England. IMD however, is a measure of relative deprivation used to rank lower layer super output areas (LSOAs) (with an average population of 1,500) in England, from the least to the most deprived areas. The IMD score is calculated by combining census and other administrative data, across seven ‘domains’ of local social and material deprivation: income deprivation; employment deprivation; education, skills and training deprivation; health deprivation and disability; crime; barriers to housing and services and living environment deprivation (Department for Communities and Local Government, 2016). Although individual domains (e.g. the health or income domain) can be disaggregated and analysed separately, this could not have been done in the current project, as only the composite index (expressed as quintiles the respondent’s LSOA was located in) was included in the NDNS and HSE datasets. IMD 2010 was used across waves 1-8 of the NDNS, whereas IMD 2007, IMD 2010 and IMD 2015 were used for the 2008-10, 2011-13 and 2015-16 HSE survey periods, respectively.

Like most statistical measures, there are limitations which should be noted when using the IMD. The IMD is limited in that it can only be used to compare/rank areas but cannot be used to measure the level of deprivation amongst individuals living in these areas. The methodology used to derive the IMD has remained fairly consistent across the different releases. This makes it possible to compare rankings between newer and older versions, such as the different versions used across the HSE survey years. However, the analysis of change over time should be done with caution, as the IMD is not a direct measure of whether an area is improving or worsening (over time), against the average. Though noteworthy, this should not pose a challenge, as the analysis of change over time is not within this project's scope.

Another source of contention surrounding the IMD is the level or geographic scale at which the index was conceived. The IMD is often described in Governmental documents as the official small area/neighbourhood level measure of relative deprivation in England (Department for Communities and Local Government, 2016). Despite this, some researchers still view the IMD as predominantly an individual/household (compositional) characteristic, rather than a neighbourhood level (contextual) measure (Deas, Robson, Wong and Bradford, 2003). In response to this, Deas, Robson, Wong and Bradford (2003, pp.898) have cautioned researchers and policymakers not to "restrict the conceptualisation of deprivation to the individual and household levels and ignore the effect of the neighbourhood." Instead, a multi-level approach (to be discussed in Section 4.2.3.3) has been recommended by the authors, as it acknowledges the hierarchical or non-hierarchical structure of data and adequately estimates the individual as well as neighbourhood level effects on outcomes of interest.

4.2.2 Missing data handling

Prior to the analysis of data, preliminary checks were carried out in the HSE and NDNS to check for missing data and to assess whether data were missing completely at random (MCAR) or missing at random (MAR). MCAR assumes that missing values are randomly distributed across all observations, whereas MAR assumes that missing values are not random, but can be explained by other variables of interest with complete information (Garson, 2015).

4.2.2.1 HSE

Preliminary missing value analyses for the HSE were conducted in SPSS to check for missing data for the 13 variables discussed in Section 4.2.1.4 – 4.2.1.8 (age, sex, ethnicity, marital status, highest level of education, NS-SEC occupational grade, fruit portions, vegetable portions, fruit and vegetable portions, achievement of the 5-A-Day target, GOR, IMD and BMI). As previously noted, (see Section 4.2.1.6.2), equivalised household income was removed from the HSE dataset in 2015 and could not be analysed within this project. There were no missing values for age, sex, GOR and IMD. However, missing values were observed for the nine remaining variables: NS-SEC occupational grade (1.5%), highest level of education (0.4%), ethnicity (0.3%), marital status (0.00%), fruit portions (0.00%), vegetable portions (0.00%), fruit and vegetable portions (0.00%), achievement of the 5-A-Day target (0.00%) and BMI (25.4%). With the exception of BMI (which had the most noteworthy missing percentage of 25%), the remaining variables of interest had under 5% percent missing.

Overall, there was no apparent pattern to the missing data (non-monotone missing pattern) and in most instances, "item not applicable" was the primary reason for missing values. According to HSE missing value coding conventions, "Item not applicable" refers to instances in which a particular variable did not apply to a given respondent, usually due to internal routing. For instance, only two of the 30 individuals with missing values for highest level of education variable, gave no response or refused to respond. The remaining 28 individuals were coded with "item not applicable." For the BMI variable, height and/or weight measurements were not taken for adult participants who were pregnant, unable to stand or unsteady on their feet. Also, measurements deemed unreliable (e.g. participants wearing bulky/excessive clothing) by survey interviewers were not included in the final HSE dataset. The unavailability of height/weight measurements in these instances meant that BMI could not be calculated and was therefore coded as "item not applicable" in final HSE datasets.

In keeping with good statistical practice, the Pearson chi-square test of association was used in the current project to test the relationship between missing variables and other fully observed variables of interest (such as age and sex) and to assess whether data were MCAR or MAR (Garson, 2015). Missing variables were recoded into categorical variables, with a category for missing and observed values. The null hypothesis (H_0) in this instance was that data were MCAR and the alternative hypothesis (H_1), that data were not MCAR. Tests which yielded a significant association at the 5% (0.05) significance level, implied that data were not MCAR, but more consistent with the MAR assumption. Garson (2015) previously noted that if a statistically significant association is found between any missing variable and at least one other variable of interest in the dataset, it is safe to assume that data are MAR.

A weak but statistically significant association was found between BMI and region (χ^2 (40) = 362.2, p-value < 0.05, Cramer's V= 0.04 and BMI and ethnicity (χ^2 (25) = 539.0, p-value < 0.05, Cramer's V= 0.05). A moderate and statistically significant association was found between BMI and sex (χ^2 (5) = 635.7, p-value < 0.05, Cramer's V= 0.12), BMI and age (χ^2 (45) = 3,136, p-value < 0.05, Cramer's V= 0.12), BMI and highest level of education (χ^2 (6) = 406.9, p-value < 0.05, Cramer's V= 0.10) and BMI and NS-SEC occupational grade (χ^2 (5) = 369.1, p-value < 0.05, Cramer's V= 0.10). The results of the Pearson chi-square test showed that data could be assumed to be MAR and also highlighted the importance of carrying out the necessary data checks before making any assumptions.

Complete case analysis is the most frequently used method of dealing with missing data in the behavioural and social sciences (McNeish, 2017). This method restricts data analysis to only individuals with complete data. Alternatively, multiple imputation (MI) predicts plausible values for missing data based on other observed data, and then imputes/fills in missing data with these predicted values multiple times. The major drawback with complete case analysis, compared to MI, is that the deletion of incomplete or missing cases from the dataset can drastically reduce the sample size, which poses a challenge for studies with initially small sample sizes (McNeish, 2017). Besides this limitation, there is also currently no established cut-off regarding the percentage of missing data acceptable for valid inferences (Dong and Peng, 2013). Despite the lack of consensus, it has been suggested that complete case analysis may be appropriate for studies with large sample sizes (greater than 5,000) and missing data which do not exceed 40% (Jakobsen, Gluud, Wetterslev and Winkel, 2017; McNeish, 2017). Overall, 12,158 of the 46,353 individuals sampled (26% of the sample), had at least one missing value across the 13 variables. The remaining 74% of individuals sampled, had values for all 13 of the variables of interest captured in the HSE. Given the HSE's initially large sample size of 46,353 and the fact that 12 of the 13 variables of interest had missing values below 5%, complete case analysis seemed appropriate for use within the current project. However, since BMI had the most noteworthy missing percentage of 25%, additional checks (results presented in Table 4.1 and Appendix B.1) were done to ensure that persons within particular demographic groups, SEP groups and regions were not disproportionately affected by the removal of missing BMI cases (all of which were coded as "item not applicable). This detailed assessment of missing data and patterns was crucial because studies have shown that the proportion of missing data should not be the only means of deciding which method (complete case analysis or MI) to use for handling missing data (Dong and Peng, 2013; Madley-Dowd, Hughes, Tilling and Heron, 2019).

The average percentage missing across the age groups was approximately 27%, with percentages ranging from 24%-29% missing (Appendix B.1). Approximately 28% of persons aged 16-24 years had missing values for BMI, which meant that 72% of persons within the same age group had values for BMI. The percentage missing appeared to be similar across the age groups, as approximately 29%, 27%, 24% and 26% of persons aged 25-29 years, 30-39 years, 40-49 years and 50-64 years, respectively, had missing values for BMI. Approximately 25% of males and 26% of females (respectively) had missing values for BMI, which was 1% difference between the sexes. The average percentage missing across ethnic groups was approximately 29%, with percentages ranging from 25%-30%. Approximately 30% of persons belonging to the Mixed ethnic group had missing values for BMI. This was the same percentage missing for persons from the Asian, Black and Other ethnic group, respectively. Thus, there appeared to be little or no difference between the ethnic groups in terms of the percentage of missing values for BMI (Appendix B.1).

The average percentage missing across the nine GORs was approximately 25%. The percentage of persons with missing values across the nine GORs, ranged from 19%-29% (see Appendix B.1). Twenty-eight percent of persons residing in the North East of England, 27% of persons in the North West, 27% in Yorkshire and the Humber and 26% in the East of England had missing values for BMI. These percentages were marginally lower than the 29% of persons in London who had missing values for BMI. In terms of highest level of education, the percentage missing across the six education categories (No qualifications; GCSE and equivalent qualifications; GCE A level or equivalent qualifications; Higher education, below degree level; Degree or equivalent and Still in full-time education/Other foreign qualification/Other) ranged from 21%-29% (results presented in Appendix B.1). Approximately 25% of persons with GCSE and equivalent qualifications, 25% with GCE A level or equivalent qualifications and a corresponding 25% of persons with a Degree or equivalent level qualification, had missing values for BMI. These findings suggested that there was no major difference between persons with a Degree or equivalent qualification and those with below degree level qualifications, as it relates to missing values for BMI. Similarly, in terms of occupational grade (NS-SEC), approximately 26% of persons in Routine, manual and other occupations had missing values for BMI, compared 23% of persons in Higher managerial and professional occupations. Although there were slight variations between GOR and SEP categories, for the most part, differences appeared to be marginal.

Further analysis (presented in Table 4.1) showed that the percentage distribution before and after the deletion of missing BMI cases was similar for the variables of age, sex, ethnicity, GOR, NS-SEC occupation and highest level of education. The percentage of persons sampled who were 16-24 years of age was 13% before and 13% after the deletion of missing BMI cases. Likewise, 9%, 21% and 24% of the sample (before and after the deletion of missing BMI cases) were persons aged 25-

29 years, 30-39 years and 40-49 years, respectively. The percentage of persons sampled from the North East region of England was 8% before and 7% after the deletion of missing BMI cases, which represented a 1% difference (Table 4.1).

Table 4.1: Percentage (%) distribution of age, sex, ethnicity, GOR, NS-SEC/occupation and highest level of education variables before and after deletion of missing BMI cases, HSE, 2008-16

Variable	**Frequency distribution of variable before deletion of missing BMI cases Number (%) n= 46,353**	**Frequency distribution of variable after deletion of missing BMI cases Number (%) n= 34,576**
Age		
16-24	6,218 (13)	4,494 (13)
25-29	4,151 (9)	2,950 (9)
30-39	9,831 (21)	7,188 (21)
40-49	11,006 (24)	8,358 (24)
50-64	15,147 (33)	11,586 (34)
Sex		
Male	20,510 (44)	15,466 (45)
Female	25,843 (56)	19,110 (55)
Ethnicity		
White	40,386 (87)	30470 (88)
Mixed	722 (2)	503 (1)
Asian	3,280 (7)	2310 (7)
Black	1,411 (3)	994 (3)
Other	395 (1)	276 (1)
Missing	159 (0)	23 (0)
Government Office Region (GOR)		

Variable	Frequency distribution of variable before deletion of missing BMI cases Number (%) n= 46,353	Frequency distribution of variable after deletion of missing BMI cases Number (%) n= 34,576
North East	3,553 (8)	2,553 (7)
North West	6,459 (14)	4,687 (14)
Yorkshire and the Humber	4,650 (10)	3,391 (10)
East Midlands	4,278 (9)	3,326 (10)
West Midlands	4,552 (10)	3,419 (10)
East of England	5,195 (11)	3,842 (11)
London	5,849 (13)	4,125 (12)
South East	7,386 (16)	5,661 (16)
South West	4,431 (10)	3,572 (10)
Occupation/NS-SEC		
Never worked/Long-term unemployed	1,134 (2)	715 (2)
Routine, manual and other occupations	18,235 (39)	13,534 (39)
Small employers and own account workers	3,885 (8)	2,921 (8)
Intermediate occupations	17,071 (37)	12,940 (37)
Higher managerial and professional occupations	5,348 (12)	4,139 (12)
Missing	680 (1)	327 (1)
Highest level of education		
No qualifications	6,612 (14)	4,705 (14)
GCSE or equivalent qualifications	11,939 (26)	8,925 (26)

Variable	Frequency distribution of variable before deletion of missing BMI cases Number (%) n= 46,353	Frequency distribution of variable after deletion of missing BMI cases Number (%) n= 34,576
GCE A level or equivalent	6,899 (15)	5,157 (15)
Higher education, below degree level	4,797 (10)	3,766 (11)
Degree or equivalent	11,995 (26)	9,011 (26)
Still in full-time education/Other Foreign qualification/Other	3,930 (8)	2,982 (9)
Missing	181 (0)	30 (0)

Similarly, the percentage of the HSE sample from the South West of England was 10% before and 10% after the deletion of missing BMI cases. A similar pattern was observed for the SEP variables of highest level of education and NS-SEC occupation grade (Table 4.1). Overall, there appeared to be little or no difference between the distribution of age, sex, ethnicity, GOR and SEP variables, after the deletion of all missing BMI cases. This suggested that the deletion of missing cases may not have disproportionately affected persons from particular socio-demographic groups or regions. Moreover, Sperrin et al (2016) previously analysed HSE data across the 1992-2011 period (19 years) and found no discernible difference between missing and non-missing cases for BMI in relation to variables such as age, sex, educational status and income. Although highlighting the limitations associated with complete case analysis, only complete cases for BMI were analysed within the Sperrin et al (2016) study. The findings of the Sperrin et al (2016) study helped to further support this project's final decision to delete all missing cases (for the 13 HSE variables of interest) from the analysis. However, it is acknowledged that as is customary with the complete case analysis method, the removal of missing cases has the potential to introduce bias and is a possible limitation (Allison, 2001).

The removal of missing BMI cases reduced the original sample size from 46,353 to 34,576 persons. The sample size further reduced to 34,195 persons after the removal of missing cases for the remaining 12 variables of interest. However, the HSE's initially large sample size proved advantageous, as despite the removal of missing cases, the final sample size across the 2008-16 period, was still more than ten times the size of the NDNS for the corresponding period.

4.2.2.2 NDNS

Four of the 14 variables selected in the NDNS had missing values. These were equivalised household income (12% missing), BMI (6% missing), highest level of education (2% missing) and NS-SEC occupational grade (0.3% missing). Although there was no apparent pattern to the missing data (non-monotone missing pattern), equivalised income had the highest percentage missing in the NDNS, with 12% missing. The majority (11.9%) of missing values for equivalised income were coded as "no answer/refused and "don't know." Similarly, participants' refusal to have their height/weight measurements taken, accounted for approximately 3% of the 6% total missing values for BMI. Analysis of missing cases (not variables) revealed that approximately 18% (or 595 of the 3,286 individuals sampled in the NDNS), had at least one missing value across the four variables. The remaining 2,691 individuals sampled had values for all 14 variables of interest. Unlike the HSE, which had an originally large sample size, exclusion of all missing cases would have reduced the NDNS' already small sample size from 3,286 to 2,691 persons. Besides the reduced sample size and the resultant loss of information, removal of missing cases from the NDNS could have proven problematic for data modelling and could have potentially introduced bias if data were not MCAR (Allison, 2001). Based on this preliminary assessment, complete-case analysis was deemed an inappropriate means of handling missing data in the NDNS.

All four missing variables identified in the NDNS were recoded into binary categorical variables (Missing; Non-missing) to capture missing and observed values within these variables. The Pearson chi-square test of association was used to test the relationship between binary categorical missing variables and other fully observed variables of interest (such as age and sex) and to assess whether data were MCAR or MAR (Garson, 2015). There was no statistically significant association between sex and any of the binary missing variables. Similarly, no statistically significant association was found between "Achieve 5" (binary achievement of the 5-A-Day target) and any of the binary missing variables. However, age was significantly associated with equivalised income (χ^2 (4) = 21.8, p-value < 0.05, Cramer's V= 0.08) and BMI (χ^2 (4) = 10.6, p-value < 0.05, Cramer's V= 0.06). Highest level of education was also found to be significantly associated with IMD (χ^2 (4) = 12, p-value < 0.05, Cramer's V= 0.06). Despite the weak association between age, equivalised income and BMI, and between education and IMD, (based on the low Cramer's V statistic), missing data in the NDNS were assumed to be MAR. MI was selected as the best approach in this instance, because it provides valid statistical inferences under the MAR assumption, even when the level of association between variables is weak or imperfect (Dong and Peng, 2013; Garson, 2015).

MI was conducted in SPSS version 24 (IBM, 2018), using the Fully Conditional Specification (FCS) method. SPSS offers two other multiple imputation methods besides FCS (the Monotone and Auto method). However, FCS is an iterative Markov Chain Monte Carlo (MCMC) method, which is best suited for dealing with missing data that have an arbitrary, non-monotone pattern, such as observed in the NDNS (Liu and De, 2015). The FCS used in SPSS operated such that for each missing variable, univariate (single dependent variable) regression models were fitted sequentially, using all other variables in the model as predictors. SPSS used linear regression models for continuous variables entered (the initial NDNS equivalised household income and BMI continuous variables with missing values were used) and logistic regression was used as the univariate model for categorical variables (highest level of education and NS-SEC). This process was done to estimate and impute missing values for each of the four variables identified and continued until the maximum number of iterations was completed.

The default five imputations and 10 iterations offered in SPSS, was initially selected as part of the recommended model building process as well as to assess overall convergence of the data (Allison, 2001; van Buuren, 2018). Although 3-5 imputations have been found to be sufficient (Graham, Olchowski and Gilreath, 2007), 20-100 imputations and between 50 and 1,000 iterations are more ideal and widely accepted (Allison, 2001; Garson, 2015; van Buuren, 2018). Based on these statistical guidelines, the number of imputations was increased to 25 and the number of iterations to 200. Although SPSS does not formally provide diagnostics, which limits the assessment of convergence, FCS iteration history was provided (in a separate dataset) and showed the values for all four variables at each iteration. This was used as a proxy to assess data convergence at 10 and 200 iterations. Allison (2001), recommended that instead of assessing convergence diagnostics for all missing variables, emphasis should only be placed on the variable with the most missing data. Thus, SPSS FCS iteration history data were exported to Microsoft Excel and used to create trace plots of equivalised household income for the default 10 and 200 iterations. Trace plots (Figure 4.3 and Figure 4.4) were then compared to assess the speed at which the MCMC procedure converged in both instances.

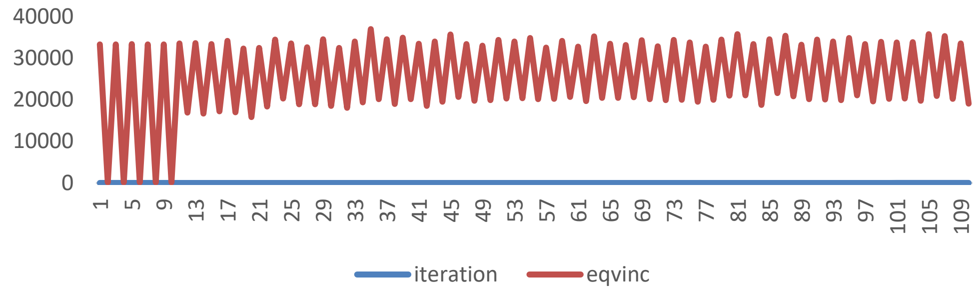


Figure 4.3: Trace plot of equivalised household income for the default 10 iterations

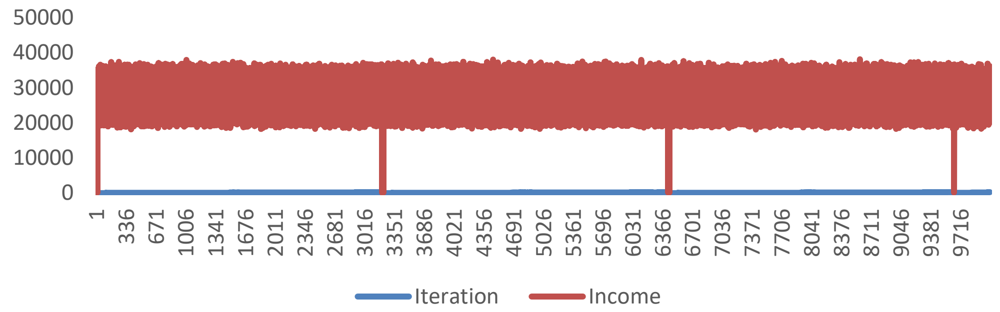


Figure 4.4: Trace plot of equivalised household income for 200 iterations

The data in Figure 4.3 did not appear to have mixed well, which implied that convergence was not reached and the number of iterations needed to be increased from the default of 10. However, the relatively horizontal pattern observed in Figure 4.4, would seem to suggest that 200 iterations was sufficient as the data appeared to converge. With diagnostics complete, imputed datasets based on 25 imputations and 200 iterations were pooled and deemed the most appropriate for use within this project.

To further assess the "success" of the MI procedure conducted, binary logistic regression models were conducted (in SPSS) with pre and post-imputed data. For both models (pre and post-imputation models), "Achieve 5" was used as the outcome variable and all remaining variables of interest (age, sex, ethnicity, marital status, highest level of education, equivalised household income, NS-SEC, BMI, GOR and IMD) entered as independent variables. In order to capture and model missing data for the pre-imputation model, all four of the initial missing variables (equivalised household income, BMI, highest level of education and NS-SEC) were recoded with "missing" as a category.

Sensitivity analyses were then conducted to assess possible differences between pre and post-imputed data (see Appendix B.2 and B.3). Models were compared and assessed by the full Supervisory Team and final sensitivity analyses revealed no significant difference between the coefficients (log-odds) for pre and post-imputation logistic regression models. For the most part, standard errors in the post-imputation model were slightly smaller, suggesting a slight improvement over the pre-imputation model. Also, the log-likelihood statistic for both the pre-imputation and post-imputation model was noted and used to assess the goodness-of-fit of the models. A reduction in the log-likelihood statistic usually indicates an improvement in the model fit. The log-likelihood statistic for both the pre and post-imputation model was approximately 3,418, when rounded to the nearest whole number. This implied that there was no significant difference between the pre and post-imputation model. This finding helped to further justify the project's use of the MI method in the NDNS and the use of imputed data (based on 25 imputations and 200 iterations) for further analysis.

4.2.3 Statistical Analysis

4.2.3.1 Survey weighting

Individual level sample weights which adjusted for sampling bias and non-response were included in the NDNS and HSE. Age (grouped), sex and GOR were the key variables used to calculate individual weights in both surveys. Prior to the analysis of data, it was considered good practice to model and assess weighted and unweighted data to determine the appropriateness of weights in the current project. Preliminary binary logistic regression models (model results not included in this project) were tested for the "Achieve 5" 5-A-Day binary outcome variable, with the remaining variables (age, sex, ethnicity, marital status, highest level of education, equivalised household income, NS-SEC occupational grade, BMI, GOR and IMD) used as independent variables. This resulted in two models; one with weights (weighted model) and the other without weights (unweighted model). Coefficients and standard errors were assessed and the minus 2-log likelihood (-2LL) value was used to assess the goodness-of-fit models tested. A reduction in the -2LL value indicated an improvement in the model. For the most part, standard errors in the weighted model were slightly smaller, suggesting a marginal improvement over the unweighted model. However, the -2LL value for the weighted and unweighted model was approximately 3,599 and 3,420 respectively. This implied that the unweighted model was the model with the better fit, which helped to justify the project's use of unweighted data.

Due to the project's extensive use of more complex methods such as multi-level modelling (particularly cross-classified multi-level modelling to be discussed in Section 4.2.3.5), the use of weights is generally not recommended (Gelman, 2007). In addition, the use of unweighted data was considered more appropriate, as this project is less focused on making population level inferences and more concentrated on the modelling of survey data. As such, a model-based approach to weighting was implemented. With this method, survey weights provided in the

NDNS and HSE were not used. Instead, all variables used to calculate survey weights (age, sex and GOR) were included as independent/explanatory variables in all models tested. In other words, models fitted were controlled by design, through the inclusion of the variables that were used to form weights. This method is widely accepted, as it accommodates the modelling of data with complex multi-level structures, produces generally unbiased results and accounts for the disproportionate sample design or non-response (Gelman, 2007; Young and Johnson, 2012). These findings further justified the project's final decision to conduct all analyses using unweighted data. Although only unweighted data are presented and analysed within the body of this thesis, weighted percentages for univariate (descriptive) statistics from the NDNS and HSE have been included in Appendix B.4 and B.5 (respectively) for general information.

4.2.3.2 Univariate and bivariate descriptive statistics for the NDNS and HSE

Univariate descriptive statistics were generated (in SPSS) for the 14 variables of interest captured in the NDNS and HSE (age, sex, ethnicity, marital status, highest level of education, equivalised household income, NS-SEC occupational grade, BMI, fruit portions, vegetable portions, fruit and vegetable portions, achievement of the 5-A-Day target, GOR and IMD). Bivariate statistics and the Pearson Chi-Square test of association were used to investigate possible associations between the achievement of the national 5-A-Day target (measured in the NDNS and HSE as the "Achieve 5" binary outcome variable) and the 13 remaining variables of interest (age, sex, GOR etc.) and to test for multi-collinearity, especially between SEP variables. The Cramer's V statistic was used to assess the strength of association between variables tested in both surveys. These statistics were primarily used to describe the characteristics of persons sampled and highlight possible similarities/differences between the two data sources. For simplicity/readability, bivariate associations between fruit and vegetable consumption and the remaining variables (in the NDNS and HSE) were restricted to the achievement of the 5-A-Day target (Yes; No, binary variable).

The analysis of HSE data as it relates to this analysis, was strictly descriptive in nature and did not involve the modelling of data. However, it should be noted that the HSE will be the primary data source used to answer the project's third research question, which investigates fruit and vegetable consumption across smaller areas of England (sub-regional variation), which will be discussed in Chapter 7. The following sections (Section 4.2.3.3- Section 4.2.3.7) describe the methods (data modelling methods) used in the NDNS, to explore regional differences in the achievement of the 5-A-Day target as well as the consumption of fruits and/or vegetables (the four outcome variables discussed in Section 4.2.1.4).

4.2.3.3 Assessment of NDNS data structure

As discussed in Section 3.3.1, multi-stage stratified random sampling was used in the NDNS. Survey participants were randomly selected from addresses, which were randomly selected from postcode sectors (or small geographical areas), randomly drawn from the Postcode Address File (PAF). Although multi-stage sampling is used in most national surveys, the various stages of the sampling process mean that data are inherently structured across multiple levels. As a result, individual observations are generally dependent and likely to be in violation of the independence assumption made in most standard statistical tests (Hox, 2010). It is therefore recommended that the structure of survey data, especially data based on a multi-stage design, be thoroughly assessed and the inherent multi-level structure be considered before data are analysed (Chromy and Abeyasekera, 2005; Hox, 2010; Leckie, 2013).

Despite these recommendations, Leckie (2013) noted the tendency for researchers to readily analyse data using traditional multiple regression or fixed effects modelling, without considering how data are collected or structured. Although frequently used and widely accepted, fixed effects (single-level) models treat individuals (the unit of analysis) as independent observations, without considering the overall structure of the data and the relationship between individuals and their contextual environment. Multi-level modelling however, is especially useful for the analysis of survey data based on a multi-stage design, as it acknowledges the structure of data. This method facilitates the exploration of relationships across and within the different levels of a multi-stage design, taking account of the variability at different levels, and the net effect on outcomes of interest. Nevertheless, the decision to use fixed or multi-level modelling is also dependent on the nature of the topic. For projects/studies strictly focused on the analysis of individual, rather than higher level factors, fixed effects models may be more appropriate if conclusions are made only at that level (Chromy and Abeyasekera, 2005). However, for projects like this, which explore more higher level geographical factors, the use of fixed effects modelling without considering the inherent structure of survey data, could lead to underestimated standard errors of regression coefficients, possibly overstated statistical significance and misleading conclusions about the relationships being studied (Hox, 2010; Leckie, 2013). Given that this analysis explores regional differences in fruit and vegetable consumption, multi-level modelling was deemed relevant to the current project.

As recommended (Chromy and Abeyasekera, 2005; Hox, 2010; Leckie, 2013), NDNS data were exported to the MLwIN multi-level (version 3.02) software package (Charlton et al, 2017) to assess the data structure. Initially, data were assumed to have a two-level hierarchical structure, in which individuals (the first level or level 1) were perfectly nested/grouped into GORs (level 2 regions). However, preliminary checks revealed a cross-classified structure (Figure 4.5).

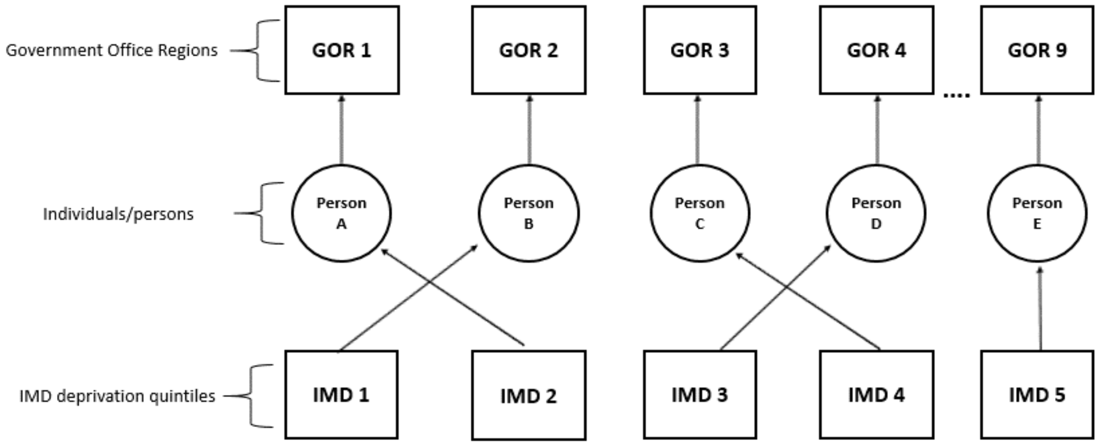


Figure 4.5: Two-way cross-classified multi-level structure of the NDNS, 2008-16

Figure 4.5 illustrates that individuals were systematically placed/grouped in regions (GORs) as initially assumed. However, when IMD quintiles were considered, a second grouping existed, as individuals were also grouped according to mutually exclusive deprivation quintiles (measured by the IMD), with cross-connecting lines depicting a two-way cross-classified structure (Figure 4.5). This finding seemed to be aligned with previously noted studies (refer to Section 4.2.1.8) which stressed the importance of acknowledging the overall structure of data as well as recognising individual and neighbourhood level effects (specifically, the IMD) on outcomes of interest (Deas, Robson, Wong and Bradford, 2003). The decision to adopt a multi-level (modelling) approach within this project was warranted for two major reasons. Firstly, IMD is a widely used neighbourhood-level measure of deprivation (not an individual level measure of deprivation or a characteristic of individuals), which like GOR, is built into the overall data structure of the NDNS and HSE and should be acknowledged. Finally, a multi-level approach recognises that individuals do not exist in insolation. Instead, they are simultaneously nested or placed within geographic regions (GOR) as well as mutually exclusive deprivation quintiles (IMD); both of which could have a potential effect on fruit and vegetable consumption and overall diet quality. Based on these findings, two-way cross-classified multi-level modelling was selected as the most suitable method for this project.

4.2.3.4 Comparison of single, two-level and cross-classified preliminary models

Although the NDNS was assumed to have a cross-classified structure, it was still considered good practice, especially during the early model building phase, to assess and compare preliminary two-way cross-classified models with corresponding fixed-effects (single-level) and two-level hierarchical models. This process was primarily carried out to assess the goodness-of-fit of preliminary models and provide further justification for the project's final decision to use two-way cross-classified multi-level modelling.

Preliminary binary logistic regression models were appropriate for modelling the "Achieve 5" (achievement of the 5-A-Day target) binary outcome variable, whereas cumulative logit or proportional odds models were suitable for the remaining ordinal outcome variables (fruit portions, vegetable portions and fruit and vegetable portions). All single-level and random intercept multi-level (two-level hierarchical and cross-classified) models were developed and tested in the MLwIN statistical package (version 3.02) (Charlton et al, 2017). Age, sex, ethnicity, marital status, equivalised household income, highest level of education, NS-SEC occupational grade and BMI were the explanatory variables entered as fixed-effects in all single, two-level hierarchical and two-way cross-classified models tested (Table 4.2). GOR and IMD were entered as fixed effects in single-level models. However, IMD was included as a fixed effect and GOR as a random effect in preliminary two-level hierarchical models and both GOR and IMD were included as random effects in two-way cross-classified random intercept models (Table 4.2).

Table 4.2: Fixed and random effects included in single-level, two-level hierarchical and two-way cross classified models tested

Model type	**Fixed effects included in model**	**Random effect(s) included in model**
Fixed-effects/single-level model	Age, sex, ethnicity, marital status, equivalised household income, highest level of education, NS-SEC occupational grade, BMI, GOR and IMD.	Not applicable
Two-level hierarchical model	Age, sex, ethnicity, marital status, equivalised household income, highest level of education, NS-SEC occupational grade, BMI and IMD.	GOR
Two-way cross-classified model	Age, sex, ethnicity, marital status, equivalised household income, highest level of education, NS-SEC occupational grade and BMI.	GOR and IMD

The Markov Chain Monte Carlo (MCMC) estimation method was selected, as it is currently the preferred method for analysing models with categorical outcomes (especially multi-level models) and models with complex data structures, such as cross-classified structures (Browne, 2019). In

MCMC estimation, a random sample of values are drawn for each parameter from its probability distribution. Thereafter, the mean and standard deviation of each random sample, produces the point estimate and standard error for that parameter. The number of draws or iterations required (in MCMC estimation) before the target distribution is reached is known as the “burn-in” period (Hox, 2010), whereas “thinning” refers to the removal of all except every k^{th} sample value/observation. The default number of iterations in MLwIN is 5,000 iterations, whereas the burn-in and thinning values are 500 and 1 respectively. Default values, specifically the default of 5,000 iterations, is not appropriate or acceptable for final models. However, this default value was only used during the preliminary model building phase and the number of iterations subsequently increased for final models selected (which will be discussed in Section 4.2.3.7).

The Deviance Information Criterion (DIC) is a measure of “model fit or the goodness-of-fit” of a model (Spiegelhalter, Best, Carlin and Van Der Linde, 2002). These scores were primarily used to compare and assess how well the tested models fit the data. A reduction in the DIC generally indicates a better or more parsimonious model. A noted benefit of the DIC (in MCMC estimation) is its ability to assess the goodness-of-fit of non-linear models (e.g. binary logit models) and models with the same outcome variable but different structures (single and multi-level structures) (Browne, 2019). Two features which were of tremendous relevance to this project.

The DIC is found by adding the “fit” and “complexity” of a particular model and is represented in Equation 4.1 as:

Equation 4.1: Formula for the calculation of the Deviance Information Criterion (DIC) statistic

$$DIC = \overline{D} + pD$$

$\overline{D}$ = the average deviance from the complete set of iterations

pD = the estimated degrees of freedom consumed in the fit

The overall results for these models were not included in the project, given that the sole purpose of this preliminary exercise was to assess the goodness-of-fit (DIC scores) for cross-classified models compared to single-level and two-level hierarchical models. The focus at this stage was on the generation and assessment of DIC scores for the 12 preliminary single, two-level hierarchical and two-way cross-classified models tested (results presented in Section 4.3.9). This process was primarily carried out to assess the goodness-of-fit of preliminary models and provide further justification for the project’s final decision to use two-way cross-classified multi-level modelling.

4.2.3.5 Two-way cross-classified model building steps

Random intercept (or variance component) models assume that the relationship between the outcome (e.g. fruit portions) and explanatory variable(s) is the same or fixed across groups (regions and IMD quintiles). This assumption can be relaxed, resulting in what is known as random coefficient/slope models. However, random intercept models were chosen as they are generally less complicated and easier to interpret (compared to random coefficient models), especially for complex cross-classified structures, such as that used in the NDNS. Binary logistic regression (using the logit link function) was used to model the achievement of the 5-A-Day target (captured by the “Achieve 5” binary outcome variable). For the three remaining ordinal outcome variables (fruit, vegetable and fruit and vegetable portions), cumulative logistic/proportional odds regression was used to model and estimate the odds of consuming at or below “5 or more portions” (the last/reference category for the three outcome variables).

A sequential approach was used to build two-way cross-classified random intercept binary and cumulative logistic models for the four outcomes. Null models were firstly tested. Thereafter, explanatory variables were entered in MLwIN (version 3.02) sequentially, beginning with individual demographic variables, followed by SEP variables and ending with BMI. The first category was selected as the reference category for each of the explanatory variables entered. Given that variables were previously recoded to be directionally consistent (ranging from the lowest to highest category), the first category (or the “lowest” category) served as a possible baseline against which the remaining categories could be compared. The MCMC estimation method was used with the default of number of iterations (5,000), resulting in the generation of 20 preliminary models (five models for each of the four outcomes). Coefficients and standard errors were assessed at each stage of this model building process and the DIC was used to assess the goodness-of-fit of models tested. The sequential process used revealed a general improvement in parameters and model fit, with the addition of each explanatory variable. For the most part, the fully adjusted model (model with all explanatory variables including BMI) appeared to be the model with the best fit and the lowest DIC compared to preceding models tested (Table 4.7). The decision was therefore taken to proceed with the fully adjusted model for each of the four outcomes. This resulted in the selection of four fully adjusted models, one for each of the four outcomes of interest. Section 4.3.10 presents the DIC results for the models tested with and without the BMI variable and presents the overall rationale for the inclusion of BMI in all final models.

4.2.3.6 Proportional odds testing for cumulative logit models

A major assumption made in cumulative logit models (also known as the ordered logit or proportional odds model) is that explanatory variables (e.g. age, sex) have the same effect on the odds for each outcome category (Hox, 2010). In other words, the effect of for example sex, on the odds of fruit and vegetable (portions) consumption, would be the same if category one (0/Under 1 portion), were to be compared with category two (1-2 portions), or category three (3-4 portions) or any other combination of outcome categories. This is known as the proportional odds assumption. Since this assumption does not always hold, it is generally considered good practice to test this prior to the selection of final models (De Silva and Sooriyarachchi, 2012).

This was done by firstly entering each explanatory variable into MLwIN individually. Thereafter, a univariate model with common coefficients (proportional odds model) was generated, followed by a model with separate coefficients (a non-proportional model). This was done for each explanatory variable included in the fully adjusted fruit, vegetable and fruit and vegetable portions model. The decision to test variables individually (univariate models) was based on the significant amount of computer memory required to conduct complex non-proportional models. MCMC estimation was used with the default of 5,000 iterations. The goodness-of-fit of univariate proportional and non-proportional odds models was assessed using the DIC. A reduction in the DIC indicated a better or more parsimonious model and helped to determine whether or not the assumption was true for the explanatory variable tested. In total, 48 univariate models were tested (results not presented).

With regards to fruit portions, the proportional odds assumption appeared to hold true for all eight explanatory variables included in the fully adjusted model, as the DIC for the proportional odds model was lower than the non-proportional model in all instances. However, for vegetable portions, the proportional odds assumption was true for all variables, except marital status, as the DIC for the proportional odds model was higher than the non-proportional odds model. Similarly, for fruit and vegetable portions, the proportional odds assumption did not hold true for sex, highest level of education, NS-SEC occupational grade and BMI. Although the non-proportional model could provide slightly more accurate results in a few instances, it would have been more difficult to interpret and much more computer-intensive, compared to the simpler proportional odds model (De Silva and Sooriyarachchi, 2012). Thus, the decision was made to maintain the proportional odds assumption for all three of the ordinal outcome variables (fruit, vegetable and fruit and vegetable portions).

4.2.3.7 Assessment of model diagnostics and selection of final models

Model diagnostics, specifically trace plots for each parameter were generated and assessed for data convergence. For all four models (achievement of 5-A-Day, fruit, vegetable and fruit and vegetable portions), data did not appear to mix well and as expected, the number of iterations needed to be increased from the default of 5,000. The Raftery-Lewis diagnostic (Nhat) was used to help determine the number of iterations needed for data convergence for each model. Based on this diagnostic, the number of iterations was increased from 5,000 to 100,000 for the achievement of the 5-A-Day target model. The number of iterations for the fruit portions, vegetable portions and fruit and vegetable portions model, was increased from 5,000 to 120,000, 150,000 and 160,000 iterations, respectively.

Following the retesting of models, trace plots were re-assessed and the data appeared to have mixed well, after the number of iterations was increased (trace plot model diagnostics for fruit portions, vegetable portions, fruit and vegetable portions and 5-A-Day achievement presented in Appendix B.13, B.17, B.18 and B.19 respectively). Kernel density graphs were assessed, and in all instances, the graph appeared to be smooth and symmetric in shape, suggesting an approximately Normal distribution. Similarly, normal probability quantile-quantile (QQ) plots were generated in MLwIN and residuals at each level (the GOR and IMD level) appeared to follow an approximately normal distribution across the four outcomes. Caterpillar plots for GOR and IMD level residuals and their ranks were also generated and compared with the overall average. Overall, based on the diagnostic statistics assessed, the increased number of iterations used in all four models appeared to be sufficient and models tested were considered final. Equation 4.2 represents the final two-way cross-classified random intercept logistic regression model for the achievement of the 5-A-Day target. Equation 4.3 represents the final two-way cross-classified cumulative logit random intercept models, for fruit, vegetable and fruit and vegetable portions. Descriptive statistics and the analysis of results for these four finalised models are presented in Section 4.3.

Equation 4.2: Two-way cross-classified random intercept logistic regression model

$$logit\ (\pi_i) = \beta_{0i} + \beta age_i + \beta sex_i + \beta ethnicity_i + \beta marital\ status_i + \beta education_i + \beta income_i + \beta NSSEC8\ Occupation_i + \beta BMI_i$$

$$\beta_{0i} = \beta_0 + u^{(2)}_{0\ imd\ (i)} + u^{(3)}_{0\ gor(i)}$$

$$\left[u_{0\,gor(i)}^{(3)}\right] \sim N\left[0, \Omega_u^{(3)}\right] :\ \Omega_u^{(3)} = \left[\Omega_{u0}^{(3)}, 0\right]$$

$$\left[u_{0\,imd\,(i)}^{(2)}\right] \sim N\left[0, \Omega_u^{(2)}\right] :\ \Omega_u^{(2)} = \left[\Omega_{u0}^{(2)}, 0\right]$$

Equation 4.3: Two-way cross-classified cumulative logit random intercept models

$$logit\ (\gamma_i) = \beta_0\ (0\ or\ Under\ 1\ portion)_i + h_i$$

$$logit\ (\gamma_i) = \beta_1\ (1 - 2\ portions)_i + h_i$$

$$logit\ (\gamma_i) = \beta_2\ (3 - 4\ portions)_i + h_i$$

$$h_i = \beta age_i + \beta sex_i + \beta ethnicity_i + \beta marital\ status_i + \beta education_i + \beta income_i + \beta NSSEC8\ Occupation_i + \beta BMI_i + u_{3\,imd\,(i)}^{(3)} + u_{3\,gor(i)}^{(4)}$$

$$\left[u_{3\,gor(i)}^{(4)}\right] \sim N\left(0, \Omega_u^{(4)}\right) :\ \Omega_u^{(4)} = \left[\Omega_{u3,3}^{(4)}\right]$$

$$\left[u_{3\,imd\,(i)}^{(3)}\right] \sim N\left(0, \Omega_u^{(3)}\right) : \Omega_u^{(3)} = \left[\Omega_{u3,3}^{(3)}\right]$$

$$\gamma_i = \pi_i;\ \gamma_i = \pi_i + \pi_i;\ \gamma_i = \pi_i + \pi_i + \pi_i;\ \ \gamma_i = 1$$

4.3 Results

4.3.1 2008-16 NDNS sample characteristics

The total (or pooled) NDNS sample for waves 1-8 (2008-16) consisted of 3,286 persons aged 16-64 years old, living in England (Table 4.3). The majority of the sample were females (57%), persons aged 50-64 years (28%), persons who were single/never married (46%) and those self-identified as white (87%) (Table 4.3).

Table 4.3: Number and percentage of persons aged 16-64 years in England sampled in the NDNS by socio-demographic and geographical characteristics and the percentage who achieved the 5-A-Day target, 2008-16

Characteristic	**Number (%) (n=3,286)**	**Achievement of 5-A-Day target (%)**
Age		
16-24	806 (25)	11
25-29	242 (7)	24
30-39	634 (19)	27
40-49	700 (21)	28

Characteristic	Number (%) (n=3,286)	Achievement of 5-A-Day target (%)
50-64	904 (28)	39
Sex		
Female	1,859 (57)	27
Male	1,427 (43)	25
Ethnicity		
White	2,854 (87)	26
All other ethnic groups	432 (13)	29
Marital Status		
Single/Never Married	1,504 (46)	19
Married/Legally recognised civil partnership	1,285 (39)	34
Separated/Divorced	443 (14)	27
Widowed	54 (2)	41
Highest level of education		
No qualifications	358 (11)	23
GCSE or equivalent qualifications	716 (22)	20
GCE A Level or equivalent	460 (14)	24
Higher education, below degree level	307 (9)	30
Degree or equivalent	830 (25)	41
Still in full-time education/Other foreign qualification/Other	615 (19)	15
Equivalised household income		
Under £20,000	987 (30)	18
£20,000-£29,999	815 (25)	23
£30,000 - £39,999	536 (16)	29
£40,000 - £49,999	357 (11)	31
£50,000 - £99,999	501 (15)	36
£100,000 or more	90 (3)	52
Occupation/NS-SEC		
Never worked/Long-term unemployed	79 (2)	19
Routine, manual and other occupations	1,094 (33)	20
Small employers and own account workers	365 (11)	26

Characteristic	Number (%) (n=3,286)	Achievement of 5-A-Day target (%)
Intermediate occupations	1,174 (36)	29
Higher managerial and professional occupations	574 (17)	34
Body Mass Index (BMI)		
Underweight	90 (3)	19
Normal weight	1,292 (39)	28
Overweight	1,184 (36)	26
Obese/morbidly obese	720 (22)	24
Government Office Region (GOR)		
North East	183 (6)	23
North West	458 (14)	21
Yorkshire and the Humber	314 (10)	24
East Midlands	312 (10)	26
West Midlands	397 (12)	23
East of England	392 (12)	29
London	395 (12)	32
South East	528 (16)	26
South West	307 (9)	33
Index of Multiple Deprivation (IMD) Quintiles		
Least deprived area	685 (21)	33
2	639 (19)	31
3	604 (18)	25
4	675 (21)	20
Most deprived areas	683 (21)	22
Survey year		
2008/09 (Wave 1)	447 (14)	32
2009/10 (Wave 2)	385 (12)	24
2010/11 (Wave 3)	390 (12)	27
2011/12 (Wave 4)	448 (14)	22
2012/13 (Wave 5)	417 (13)	29
2013/14 (Wave 6)	353 (11)	22
2014/15 (Wave 7)	427 (13)	25
2015/16 (Wave 8)	419 (13)	28

Note: Percentages may not add to 100% due to rounding

Approximately 25% of persons sampled had a degree level qualification or its equivalent, the majority of whom were aged 30 years and older. Only 2% of persons sampled never worked or were "long-term unemployed" (which included full-time students or persons with occupations which were not stated). When combined, approximately 58% of persons sampled were classified as being either overweight, obese or morbidly obese. With regards to geographical residence, 6% of the sample lived in North East England compared to 16% in South East England.

4.3.2 2008-16 HSE sample characteristics

The pooled HSE sample (for the 2008-11, 2013 and 2015-16 period) of persons aged 16-64 years living in England was 34,195. Similar to the NDNS, the majority of the HSE sample consisted of persons aged 50-64 years (33%), females (55%) and self-identified white persons (88%) (Table 4.4). Married persons (including persons in legally recognised civil partnerships) accounted for 52% of the total sample.

Table 4.4: Number and percentage of persons aged 16-64 years in England sampled in the HSE by socio-demographic and geographical characteristics and the percentage who achieved the 5-A-Day target, 2008-16

Characteristic	Number (%) (n=34,195)	Achievement of 5-A-Day target (%)
Age		
16-24	4,386 (13)	18
25-29	2,893 (9)	22
30-39	7,118 (21)	28
40-49	8,286 (24)	27
50-64	11,512 (33)	31
Sex		
Female	18,855 (55)	29
Male	15,340 (45)	25
Ethnicity		
White	30,224 (88)	26
All other ethnic groups	3,971 (12)	33
Marital Status		
Single	7,720 (23)	22
Married, including civil partnership	17,630 (52)	30
Separated/Divorced, including from a civil partnership	3,205 (9)	28
Widowed, including civil partnership	546 (2)	30
Cohabitees	5,094 (15)	22
Highest level of education		
No qualifications	4,588 (13)	19
GCSE or equivalent qualifications	8,845 (26)	20

Characteristic	Number (%) (n=34,195)	Achievement of 5-A-Day target (%)
GCE A Level or equivalent	5,136 (15)	25
Higher education, below degree level	3,753 (11)	30
Degree or equivalent	8,974 (26)	38
Still in full-time education/Other foreign qualification/Other	2,899 (9)	23
Occupation/NS-SEC		
Never worked/Long-term unemployed	712 (2)	23
Routine, manual and other occupations	13,508 (40)	20
Small employers and own account workers	2,917 (9)	30
Intermediate occupations	12,926 (38)	31
Higher managerial and professional occupations	4,132 (12)	36
Body Mass Index (BMI)		
Underweight	596 (2)	18
Normal weight	12,644 (37)	28
Overweight	12,328 (36)	28
Obese/morbidly obese	8,627 (25)	25
Government Office Region (GOR)		
North East	2,502 (7)	22
North West	4,639 (14)	24
Yorkshire and the Humber	3,347 (10)	23
East Midlands	3,295 (10)	26
West Midlands	3,374 (10)	25
East of England	3,826 (11)	27
London	4,047 (12)	34
South East	5,616 (16)	29
South West	3,549 (10)	29
Index of Multiple Deprivation (IMD) Quintiles		
Least deprived area	7,240 (21)	31
2	6,901 (20)	30
3	6,994 (21)	27
4	6,607 (19)	25
Most deprived areas	6,453 (19)	21
Survey year		
2008	10,095 (30)	27
2009	3,039 (9)	27
2010	5,317 (16)	25

Characteristic	Number (%) (n=34,195)	Achievement of 5-A-Day target (%)
2011	5,341 (16)	26
2013	3,988 (12)	27
2015	3,341 (10)	26
2016	3,074 (9)	29

Note: Percentages may not add to 100% due to rounding

Approximately 13% of persons sampled reportedly had no educational qualification, whereas 26% had attained a degree or its equivalent. Forty percent of persons had routine, manual and other occupations, compared to two percent who had never worked or were long-term unemployed (Table 4.4). Although 37% of the sample were of normal weight, when combined, 61% of persons were either overweight, obese or morbidly obese. This was only three percentage points higher than the 58% of persons who were classified as being either overweight, obese or morbidly obese in the NDNS. Seven percent of persons sampled in the HSE lived in North East region of England, whereas 16% resided in the South East.

4.3.3 Descriptive analysis of fruit and vegetable consumption: NDNS and HSE

Overall, more than a half of the persons sampled in the NDNS and HSE did not achieve the national 5-A-Day target, as they did not consume the recommended five daily portions of fruits and vegetables (Table 4.5). Only 26% and 27% of the NDNS and HSE sample (respectively) achieved the 5-A-Day target. There was only a one percent (marginal) difference between the two data sources concerning fruit and vegetable consumption and achievement of the 5-A-Day target (Table 4.5).

Table 4.5: Number and percentage of persons (aged 16-64 years) sampled in the NDNS and HSE, who achieved the 5-A-Day target, England, 2008-16

Achievement of 5-A-day target	NDNS	HSE
	Number (%)	Number (%)
Yes	864 (26)	9,200 (27)
No	2,422 (74)	24,995 (73)
Total	**3,286**	**34,195**

When disaggregated into portions, the average number of fruit and vegetable portions consumed was 3.9 in the NDNS and 3.6 in the HSE (Figure 4.6). Thirty-four percent and 33% of the NDNS and HSE sample (respectively) consumed one to two portions, whilst 34% and 30% (respectively) consumed three to four portions of fruits and vegetables. Only 6% and 10% of persons sampled in the NDNS and HSE (respectively) consumed less than one portion of fruits and vegetables (Figure 4.6).

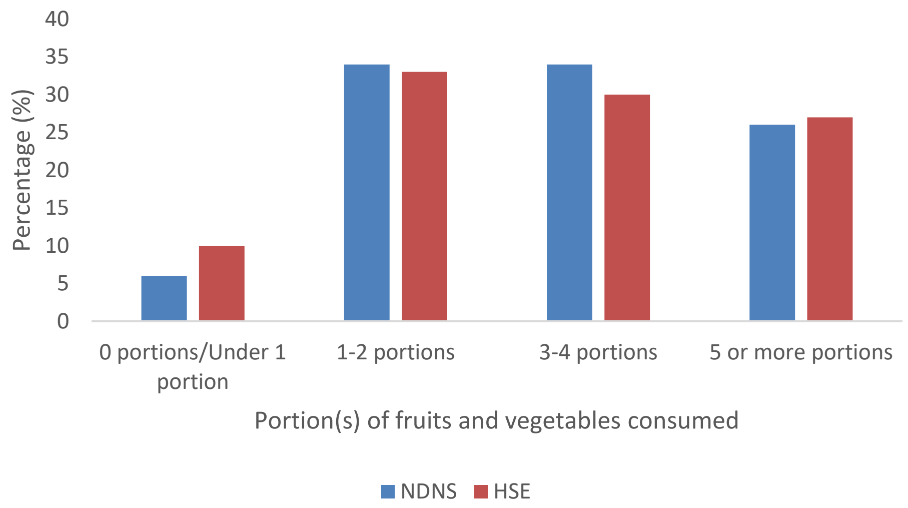


Figure 4.6: Portion(s) of fruits and vegetables consumed by persons (16-64 years old) sampled in the NDNS and HSE, England, 2008-16

Noticeable differences were observed between the two data sources in terms of fruit consumption. The majority of persons (48%) sampled in the NDNS consumed less than one portion of fruit, whereas 45% of the HSE sample consumed one to two portions of fruit (Figure 4.7). Nevertheless, only 3% and 8% of persons in the NDNS and HSE (respectively) consumed five or more portions of fruit.

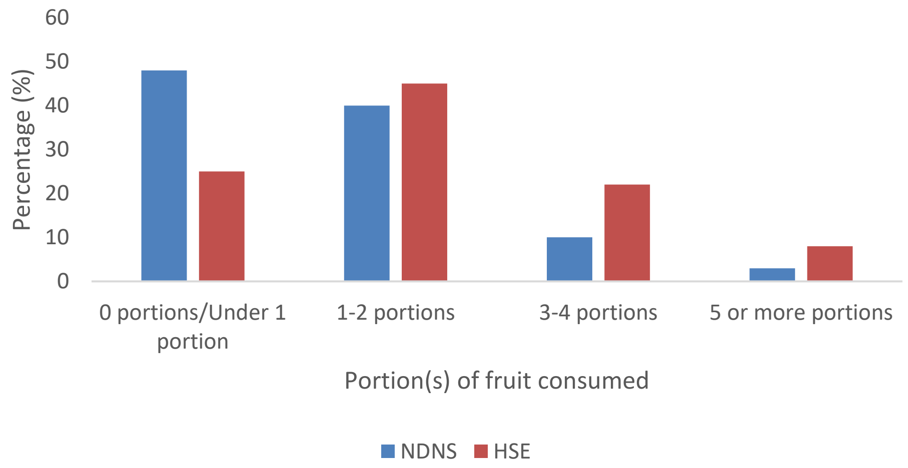


Figure 4.7: Portion(s) of fruit consumed by persons (aged 16-64 years) sampled in the NDNS and HSE, England, 2008-16

In terms of vegetable consumption, both surveys were relatively similar, in that 58% and 56% of the NDNS and HSE sample (respectively) consumed one to two portions daily (Figure 4.8).

However, 12% and 28% of the NDNS and HSE sample (respectively) consumed less than one portion, whilst 25% and 13% (respectively) consumed three to four portions of vegetables. The consumption of at least five portions of vegetables was achieved by only 6% and 3% of persons sampled in the NDNS and HSE, respectively (Figure 4.8).

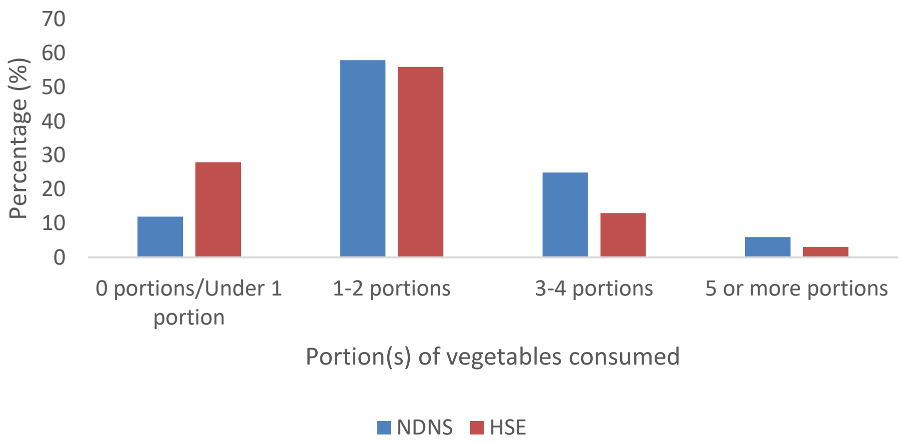


Figure 4.8: Portion(s) of vegetables consumed by persons (aged 16-64 years) sampled in the NDNS and HSE, England, 2008-16

4.3.4 Checks for multi-collinearity in the NDNS and HSE

The Pearson Chi-Square test of association was used to assess the strength of association between the SEP variables captured in the NDNS and HSE. There was a statistically significant association between education and equivalised household income ($\chi^2\ (25) = 551.15$, p-value < 0.05, Cramer's V = 0.18) and between equivalised household income and NS-SEC ($\chi^2\ (20) = 945.93$, p-value < 0.05, Cramer's V = 0.27) in the NDNS. A statistically significant association was also found between highest level of education and NS-SCEC in both surveys (NDNS: $\chi^2\ (20) = 596.04$, p-value < 0.05, Cramer's V = 0.21 and HSE: $\chi^2\ (20) = 11029.89$, p-value < 0.05, Cramer's V = 0.28). Although SEP variables were statistically significantly associated, they were not highly correlated as the Cramer's V statistic in all instances was below the typical cut-off value of 0.80 (Vatcheva, Lee, McCormick and Rahbar, 2016). Based on these findings, multi-collinearity did not appear to be an issue in the current project.

4.3.5 Association between achievement of 5-A-Day target and demographic variables: NDNS and HSE

A strong and statistically significant association was found between age and target achievement in the NDNS ($\chi^2\ (4) = 170.22$, p-value < 0.05, Cramer's V = 0.23), whereas a relatively moderate and statistically significant association was observed in the HSE ($\chi^2\ (4) = 296.27$, p-value< 0.05, Cramer's V = 0.09). The majority of persons who achieved the target in the NDNS and HSE, tended to belong to the older age groups, specifically the 50-64-year age group (Table 4.3 and 4.4 respectively). With regards to sex, 27% and 29% of female respondents in the NDNS and HSE (respectively) achieved the 5-A-Day target (Table 4.3 and Table 4.4). No statistically significant association was found between sex and target achievement in the NDNS ($\chi^2\ (1) = 1.12$, p-value = 0.291, Cramer's V = 0.02). However, a statistically significant, albeit, very weak association was found between sex and target achievement in the HSE ($\chi^2\ (1) = 66.32$, p-value < 0.05, Cramer's V = 0.04).

Approximately 26% of white respondents in the NDNS and a corresponding 26% in the HSE achieved the target (Table 4.3 and Table 4.4). Although no statistically significant association was found between ethnicity and target achievement in the NDNS ($\chi^2\ (1) = 1.79$, p-value = 0.18, Cramer's V = 0.02), a weak, but statistically significant association was found in the HSE ($\chi^2\ (1) = 92.46$, p-value < 0.05, Cramer's V = 0.05). In terms of marital status, approximately 34% and 30% of persons who were married or in a legally recognised civil partnership in the NDNS and HSE (respectively) met the 5-A-Day (Table 4.3 and Table 4.4). A strong and statistically significant association was observed between marital status and target achievement in the NDNS ($\chi^2\ (3) = 86.45$, p-value < 0.05, Cramer's V = 0.16). Although a statistically significant association was also found between marital status and target achievement in the HSE ($\chi^2\ (4) = 276.10$, p-value < 0.05, Cramer's V = 0.09), the Cramer's V statistic of 0.09 indicated a relatively moderate association between the two variables.

4.3.6 Association between achievement of 5-A-Day target and socio-economic status variables: NDNS and HSE

Statistically significant associations were found between the achievement of the 5-A-Day target and the measures of SEP (highest level of education, equivalised household income and NS-SEC occupational grade). A relatively consistent and noticeable gradient was observed for education, with more educated persons (especially those with a degree or some form of higher education) seemingly more likely to have met the target compared to those with less education (Table 4.3 and Table 4.4). For instance, 41% of persons with a degree or equivalent qualification in the NDNS achieved the 5-A-Day target compared to 20% who had up to GCSE level or equivalent qualifications. Similarly, 38% of persons who had a degree or equivalent qualification in the HSE achieved the target, compared to 20% who had GCSE level or equivalent qualifications (Table 4.4). A strong and statistically significant association was observed between highest level of education and target achievement in both surveys (NDNS: $\chi^2(5) = 156.72$, p-value < 0.05, Cramer's V = 0.22; HSE: $\chi^2(5)$ = 973.89, p-value < 0.05, Cramer's V = 0.17).

A strong and statistically significant association was found between equivalised household income and target achievement in the NDNS ($\chi^2(5) = 99.70$, p-value < 0.05, Cramer's V = 0.17). Besides household income, a moderate and statistically significant association was also found between NS-SEC and 5-A-Day achievement in both the NDNS and HSE (NDNS: $\chi^2(4) = 45.85$, p-value < 0.05, Cramer's V = 0.12; HSE: $\chi^2(4)$ = 611.11, p-value < 0.05, Cramer's V = 0.13).

4.3.7 Association between achievement of 5-A-Day target and BMI: NDNS and HSE

Twenty-eight percent of persons who were within the normal weight range in the NDNS met the 5-A-Day target (Table 4.3). Yet, no statistically significant association was observed for BMI and target achievement ($\chi^2(3) = 5.82$, p-value = 0.121, Cramer's V = 0.04). On the contrary, a statistically significant association was found between BMI and target achievement in the HSE sample ($\chi^2(3) = 52.65$, p-value < 0.05, Cramer's V = 0.04). However, the Cramer's V statistic of 0.04 (which was the same value as in the NDNS), indicated that the strength of this association was very weak.

4.3.8 Association between achievement of 5-A-Day target and geographical location: NDNS and HSE

A statistically significant association was found between deprivation quintiles (measured by IMD) and the achievement of the 5-A-Day target in both the NDNS and HSE (NDNS: $\chi^2(4) = 40.17$, p-value < 0.05, Cramer's V = 0.11; HSE: $\chi^2(4)$ = 190.83, p-value < 0.05, Cramer's V = 0.08). However, based on the Cramer's V statistic reported, the strength of this association was moderate in the NDNS and fairly weak in the HSE. In terms of GOR, approximately 23% and 22% of persons who resided in the North East in the NDNS and HSE (respectively) met the 5-A-Day compared to 26% and 29% (respectively) in the South East of England (Table 4.3 and 4.4). Although a statistically significant association was observed between GOR and target achievement in the NDNS and HSE, this was a fairly weak association given the Cramer's V value of 0.09 and 0.08 respectively (NDNS: $\chi^2(8) = 23.95$, p-value < 0.05, Cramer's V = 0.09; HSE: $\chi^2(8)$ = 223.33, p-value < 0.05, Cramer's V = 0.08).

Overall, the descriptive comparison made between the NDNS and HSE as described in Sections 4.3.1-4.3.8, showed that despite differences in sample size (3,286 in NDNS compared to 34,195 in the HSE) and dietary assessment methods used, both surveys were fairly similar in terms of fruit and vegetable data captured over the 2008-16 period. This justified the project's sole use of the NDNS to further model and explore the extent to which fruit and vegetable consumption varies by region amongst persons 16-64 years old in England.

4.3.9 Single-level, two-level and cross-classified preliminary model testing results

As mentioned in Section 4.2.3.4, preliminary two-way cross-classified models were tested along with corresponding fully adjusted fixed-effects (single-level) and two-level hierarchical models. This process was done to assess the goodness-of-fit (DIC) of preliminary models and provide further justification for the project's final decision to use two-way cross-classified multi-level modelling (Table 4.6). All models used only the NDNS.

Table 4.6: Deviance Information Criterion (DIC) for fully adjusted single-level, two-level hierarchical and two-way cross-classified models for the achievement of the 5-A-Day target, fruit portions, vegetable portions and fruit and vegetable portions outcomes, NDNS 2008-16

Model/Outcome	**Model structure**		
	Single-level	**Two-level hierarchical**	**Two-way cross-classified**
Achievement of 5-A-Day	3,496.47	3,495.45	3,493.72
Fruit portions	6,352.64	6,350.55	6,350.13
Vegetable portions	6,839.15	6,838.19	6,838.17
Fruit and vegetable portions	7,698.87	7,699.54	7,697.82

With the exception of fruit and vegetable portions, the DIC for single-level/fixed-effects models was generally higher than multi-level models tested (Table 4.6). There appeared to be little or difference between two-level hierarchical and two-way cross-classified models for fruit portions and vegetable portions, given the marginal difference in the DIC. In terms of the achievement of the 5-A-Day target and fruit and vegetable portions, two-way cross-classified models seemed to be better than single and two-level hierarchical models. Overall, multi-level models, especially two-way cross-classified models appeared to be a better fit and more appropriate than single-level and two-level hierarchical models. Based on these findings, the project was further justified in proceeding with the modelling of two-way cross-classified multi-level models for all four outcomes of interest.

4.3.10 BMI model testing results and rationale for including BMI in final NDNS models

Section 4.2.1.7 provided a brief review of BMI and obesity-related literature and the initial rationale for considering BMI for use within the current project. This section presents the results for the models tested with and without the BMI variable and helps to further justify the project's inclusion of BMI in all final models. The bivariate results presented in Section 4.3.7 showed that there was no statistically significant association between BMI and 5-A-Day target achievement in the NDNS and a weak but statistically significant association between the two variables in the HSE. The statistically insignificant association observed in the NDNS, could have been due to the survey's much smaller sample size compared to the HSE. However, the Cramer's V statistic for both surveys was 0.04, which indicated that regardless of the survey, the relationship between BMI and 5-A-Day achievement was weak, which warranted further investigation. Due to the insignificant correlation between BMI and 5-A-Day achievement, a possible approach at this stage could have been to remove the BMI variable in the NDNS from further analysis/modelling. However, the sole use of bivariate analysis for the selection of variables to be included in multivariate models, is not in keeping with good statistical practice (Heinze and Dunkler, 2017; Sun, Shook and Kay, 1996). Sun, Shook and Kay (1996) cautioned researchers against basing their decisions entirely on "p-values" obtained during bivariate analysis, as this approach could wrongfully and prematurely remove or reject potentially important variables and those which need to be controlled/adjusted for in final statistical models. For instance, although a weak association was found between BMI and 5-A-Day achievement in the NDNS, a strong and statistically significant association was found between BMI and the SEP variable of highest level of education (NDNS: $\chi^2(15) = 276.45$, p-value < 0.05, Cramer's V = 0.17). Previous UK and international studies have noted the importance of acknowledging the strong correlation between SEP and BMI and the inter-related/complex relationship with diet quality and health (Booth, Charlton and Gulliford, 2017; Hermann et al, 2011).

Heinze and Dunkler (2017) discussed common statistical myths and noted that variables of potential interest should not be pre-filtered or removed from the final analysis before they are properly assessed and tested (using a sequential model building approach) in preliminary multivariate models. Moreover, even if variables are later found to be statistically insignificant in multivariate models, they should not be immediately discarded but should be carefully considered, especially if they are found to be within the best fitting model (Heinze and Dunkler, 2017). Roberts, Cade, Dawson and Holdsworth (2018) (similar to this project) previously analysed empirically derived dietary patterns of adults in the UK, using data from the NDNS. Having included BMI in regression models, that study found that individuals with higher scores for a dietary pattern which consisted of mostly snacks, fast food and fizzy drinks were more likely to have a higher BMI than individuals with dietary patterns which consisted mostly of fruits, vegetables and oily fish. The results of the Roberts, Cade, Dawson and Holdsworth (2018) served as an additional push factor for the further testing of BMI in preliminary models, since that study examined dietary patterns in the UK using the same 60 NDNS food groups used within this project.

As mentioned in Section 4.2.3.5, a sequential approach was used to build two-way cross-classified random intercept binary and cumulative logistic models for the four outcomes. This meant that explanatory variables were entered sequentially, starting with individual demographic variables, followed by SEP variables and ending with BMI. Table 4.7 shows the DIC results for the fully adjusted model (model with all variables including BMI) and the model with all variables except BMI. The table (Table 4.7) presents the four fully adjusted models (one model for each of the four outcomes) and the four models without BMI. A reduction in the DIC statistic indicated an improvement in model fit.

Table 4.7: Deviance Information Criterion (DIC) before and after the inclusion of BMI, for two-way cross-classified achievement models tested (5-A-Day target, fruit portions, vegetable portions and fruit and vegetable portions outcomes), NDNS, 2008-16

Model/Outcome	**Model with all explanatory variables (excluding BMI)**	**Model with all explanatory variables (including BMI)**
Achievement of 5-A-Day	3,510.52	3,493.72
Fruit portions	6,361.35	6,350.13
Vegetable portions	6,836.16	6,838.17
Fruit and vegetable portions	7,705.62	7,697.82

There was a general improvement in the parameters and model fit with the addition of each variable. The fully adjusted model, which included BMI, had the lowest DIC for three of the four models (Table 4.7). It was only for the vegetable portions outcome that there was an increase in the DIC, after the addition of BMI. The DIC for the vegetable portions model increased from 6,836.16 (for the model with all variables except BMI) to 6,838.17, after BMI was added (Table 4.7). The marginal difference (a difference of 2) suggested that the model without BMI was a slight improvement over the model with BMI. On the contrary, for all three of the remaining outcomes of interest (5-A-Day achievement, Fruit portions and Fruit and Vegetable portions), the DIC decreased after the inclusion of BMI, which suggested that the fully adjusted model was the model with the best fit in these instances and that the BMI variable should be retained within the analysis. Despite the findings for vegetable portions, the fully adjusted model (inclusive of BMI) was still deemed the most appropriate and relevant for consistency. This resulted in the final selection of four fully adjusted models, one for each of the four outcomes of interest.

Overall, the findings of this section highlight the importance of conducting detailed model assessments in addition to bivariate associations and helped to further rationalise the project's final inclusion of the BMI variable in the (four) final NDNS models. Moreover, according to the Food and Agriculture Organisation (FAO et al, 2021), BMI is an internationally accepted, simple, objective and a widely used crude measure of the nutritional status of adults at the population level. Despite its noted limitations (such as its inability to account for muscle mass, bone density and body fat), BMI is closely related to diet/food consumption (the primary focus of this project) and is a recommended proxy measure for use in diet and nutrition-related studies such as the current project (FAO et al, 2021; Shetty, 2002).

In summary, obesity and the mechanisms which link BMI to diet are not within the immediate scope of this project and could signal the need for additional research in the future. Nevertheless, obesity (typically measured by BMI) continues to be acknowledged as one of the greatest diet-related public health challenges in England and the world (Theis and White, 2021; WHO, 2021). The final decision to retain BMI within this project was primarily based on: i) the complex association between diet (especially fruit and vegetable consumption which is the focus of this chapter), socio-demographics and weight status/BMI (previously highlighted in Section 4.2.1.7) and the recognised need to test/adjust for this within regression models, ii) the results of the detailed model assessment exercise conducted which highlighted BMI as being relevant and necessary for overall model fit (Table 4.7) and iii) the internationally accepted use of BMI as a proxy measure of long-term diet quality/the nutritional health status of adults. The results of the final fully adjusted models have been subsequently presented in Section 4.3.11-4.3.12.

4.3.11 Results of final two-way cross-classified binary logistic regression model

Table 4.8 presents the results of the final fully adjusted two-way cross-classified binary logistic model for the achievement of the 5-A-Day, fruit and vegetable target ("Achieve 5" outcome variable). It should be noted that statistically significant associations were tested at the 5% significance level and are denoted with a * for all models presented.

Table 4.8: Fully adjusted two-way cross-classified multi-level binary logistic regression model for the achievement of the 5-A-Day target ("Achieve 5" outcome variable) for persons 16-64 years old in England, NDNS, 2008-16

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Intercept	-2.381** (0.466)	___	___

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Age (Ref Cat: 16-24 years)			
25-29	___	2.47**	(1.60, 3.82)
30-39	___	2.65**	(1.81, 3.91)
40-49	___	3.10**	(2.10, 4.60)
50-64	___	5.69**	(3.88, 8.42)
Sex (Ref Cat: Female)			
Male	___	0.92	(0.77, 1.09)
Ethnicity (Ref Cat: White)			
All other ethnic groups	___	1.36*	(1.05, 1.76)
Marital Status (Ref Cat: Single Never Married)			
Married/In a legally recognised civil partnership	___	1.10	(0.88, 1.38)
Separated/divorced	___	0.80	(0.59, 1.07)
Widowed/other/formerly in a legally recognised civil partnership	___	1.41	(0.76, 2.56)
Highest level of education (Ref Cat: No qualifications)			
GCSE or equivalent qualifications	___	1.00	(0.72, 1.38)
GCE A level or equivalent	___	1.14	(0.81, 1.61)
Higher education, below degree level	___	1.36	(0.93, 1.97)
Degree or equivalent	___	1.99**	(1.44, 2.75)
Still in full-time education/Other foreign qualification/Other	___	1.32	(0.87, 2.00)
Equivalised household income (Ref Cat: Under £20,000)			
£20,000-£29,999	___	1.20	(0.94, 1.54)
£30,000-£39,999	___	1.37*	(1.04, 1.80)
£40,000-£49,999	___	1.35*	(0.99, 1.86)
£50,000-£99,999	___	1.54**	(1.15, 2.06)
£100,000 or more	___	2.80**	(1.70, 4.61)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
NS-SEC Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	___	0.91	(0.50, 1.70)
Small employers and own account workers	___	0.97	(0.51, 1.88)
Intermediate occupations	___	1.02	(0.55, 1.92)
Higher managerial and professional occupations	___	0.86	(0.45, 1.68)
Body Mass Index: BMI (Ref Cat: Underweight)			
Normal weight	___	1.03	(0.58, 1.88)
Overweight	___	0.69	(0.38, 1.27)
Obese	___	0.63	(0.35, 1.18)
Variance of random parameters			
Government Office Region (GOR)	0.026 (0.03)	___	___
Index of Multiple Deprivation (IMD)	0.044 (0.099)	___	___

* p-value <0.05, ** p-value<0.01, SE= standard error

The findings presented in Table 4.8 showed that age, ethnicity, highest level of education and equivalised household income were statistically significant predictors of 5-A-Day achievement whereas sex, marital status, NS-SEC and BMI were not. A consistent gradient was observed with respect to age, such that as age increased, the odds of achieving the target also increased. For instance, the odds of achieving the 5-A-Day target was 2.47 times higher for persons aged 25-29 years compared to those aged 16-24 years. However, persons aged 50-64 years had more than five times the odds of achieving the target compared to those aged 16-24 years.

Persons from other ethnic minority groups had a 36% increased odds of achieving the target compared to white persons (with all other variables held constant). Persons with a degree or equivalent qualification had almost two times (1.99) the odds of achieving the target than persons without any qualifications. The odds of achieving the target was 1.20, 1,37, 1.35, 1.54 and 2.80 for persons within the £20,000-£29,999, £30,000-£39,999, £40,000-£49,999, £50,000-£99,999 and £100,000 or more income groups, respectively. Despite a slight reduction in the odds for the £40,000-£49,999 group, for the most part, the odds of achieving the target tended to increase as equivalised household income increased. Persons in the highest income group (£100,000 or more) had the highest odds of achievement, which meant that they had 2.8 times the odds of achieving the target compared to persons with household incomes below £20,000 (lowest income group).

The estimated variance of the random GOR and IMD parameters was 0.026 and 0.044 (respectively), both of which were close to zero. Figure 4.9 and 4.10 (respectively) also showed that the confidence intervals for the nine regions and for the deprivation quintiles, all overlapped with zero (the horizontal line representative of the average). Thus, there appeared to be no statistically significant difference (at the 5% level) between the regions and deprivation quintiles, with regards to the achievement of the 5-A-Day target, once the chosen individual level covariates were taken into account.

Figure 4.9: GOR level residuals and their associated 95% confidence intervals for the achievement of the 5-A-Day target, NDNS, 2008-16

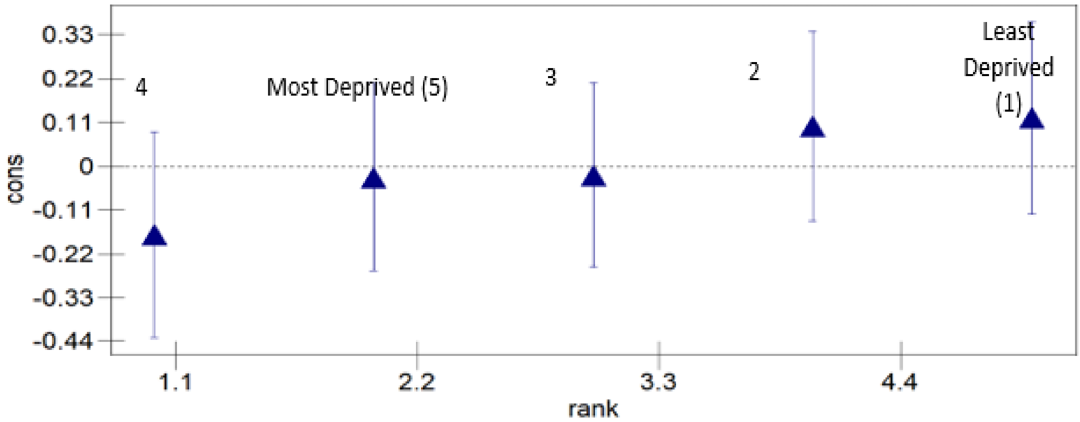


Figure 4.10: IMD level residuals and their associated 95% confidence intervals for the achievement of the 5-A-Day target, NDNS, 2008-16

4.3.12 Results of final two-way cross-classified cumulative logistic regression models

4.3.12.1 Combined Fruit and Vegetable portions consumed

Table 4.9: Fully adjusted two-way cross-classified multi-level cumulative logistic regression model for portions of fruits and vegetables (combined) consumed by persons 16-64 years old in England, NDNS, 2008-16

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Intercept (Ref Cat: 5 or more portions)			
0 portions/Under 1 portion	-1.49** (0.342)	___	___
1-2 portions	1.128** (0.339)	___	___
3-4 portions	2.767** (0.342)	___	___
Age (Ref Cat: 16-24 years)			
25-29	___	0.55**	(0.40, 0.74)
30-39	___	0.45**	(0.35, 0.59)
40-49	___	0.41**	(0.32, 0.54)
50-64	___	0.21**	(0.16, 0.28)
Sex (Ref Cat: Female)			
Male	___	1.15*	(1.01, 1.31)
Ethnicity (Ref Cat: White)			

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
All other ethnic groups	___	0.73**	(0.59, 0.90)
Marital Status (Ref Cat: Single Never Married)			
Married/In a legally recognised Civil Partnership	___	0.85*	(0.71, 1.02)
Separated/Divorced	___	1.20	(0.95, 1.51)
Widowed/Other/ Formerly in a legally recognised civil partnership	___	0.73	(0.42, 1.24)
Highest level of education (Ref Cat: No qualifications)			
GCSE or equivalent qualifications	___	0.90	(0.70, 1.16)
GCE A level or equivalent	___	0.66**	(0.51, 0.87)
Higher education, below degree level	___	0.58**	(0.43, 0.79)
Degree or equivalent	___	0.38**	(0.29, 0.49)
Still in full-time education/Other foreign qualification/Other	___	0.57**	(0.42, 0.76)
Equivalised household income (Ref Cat: Under £20,000)			
£20,000-£29,999	___	0.81*	(0.68, 0.97)
£30,000-£39,999	___	0.69**	(0.56, 0.85)
£40,000-£49,999	___	0.73**	(0.57, 0.93)
£50,000-£99,999	___	0.63**	(0.49, 0.79)
£100,000 or more	___	0.34**	(0.22, 0.53)
NS-SEC Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	___	1.01	(0.64, 1.59)
Small employers and own account workers	___	0.78	(0.48, 1.28)
Intermediate occupations	___	0.96	(0.60, 1.53)
Higher managerial and professional occupations	___	0.98	(0.60, 1.60)
Body Mass Index: BMI (Ref Cat: Underweight)			
Normal weight	___	0.87	(0.57, 1.32)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Overweight	___	1.08	(0.70, 1.64)
Obese	___	1.22	(0.79, 1.89)
Variance of random parameters			
Government Office Region (GOR)	0.029 (0.028)	___	___
Index of Multiple Deprivation (IMD)	0.033 (0.072)	___	___

* p-value <0.05, ** p-value<0.01, SE= standard error

Occupational grade (NS-SEC) and BMI were the only two variables which were not statistically significantly associated with fruit and vegetable portions (Table 4.9). Age was a highly significant predictor of fruit and vegetable (portion) consumption. As age increased, the odds of consuming fewer than the recommended five portions of fruits and vegetables decreased consistently and linearly. The odds for persons aged 25-29 years, 30-39 years, 40-49 years and 50-64 years was 0.55, 0.45, 0.41 and 0.21 respectively. Persons in the 50-64-year age group had the lowest odds (0.21) compared to persons aged 16-24 years. This meant that the odds of consuming fewer than five portions of fruits and vegetables was 79% less for persons aged 50-64 years compared to persons aged 16-24 years.

Sex was statistically significantly associated with the number of fruit and vegetable portions consumed, with the odds of consuming fewer than five portions being greater (15% greater) for males than females. In terms of ethnicity, the odds of consuming fewer than five portions was 27% less for other ethnic groups than for white persons. Married/in a legally recognised civil partnership was the only marital status category which had a relatively significant effect on the portion of fruits and vegetables consumed. The odds of consuming fewer than five portions and fruits and vegetables was 15% less for persons who were married or in a legally recognised civil partnership compared to single/never married persons (Table 4.9).

Highest level of education was a statistically significant predictor of the number of fruit and vegetable portions consumed. Although attainment of GCSE or equivalent qualifications was not associated with fruit and vegetable (portion) consumption, the odds of consuming fewer than five portions generally decreased, as the level of education increased. Persons with a degree or equivalent qualification had the lowest odds of 0.38. This meant that the odds of consuming fewer than five portions was 62% less for persons with a degree/equivalent compared to persons with no qualifications. Except for the £40,000-£49,999 equivalised household income group, the odds of consuming fewer than five portions of fruits and vegetables, generally declined as household income increased. Persons with a household income of £100,000 or more had the lowest odds of 0.34. This meant that the odds of consuming fewer than five portions of fruits and vegetables was 66% less for persons within the £100,000 or more household income group compared to those in the under £20,000 household income group.

The variance for the GOR and IMD random parameters was 0.029 and 0.033 (respectively), both of which did not appear to be much different from zero. Caterpillar residual plots helped to further illustrate this (Figure 4.11 and 4.12 respectively), by showing that the confidence intervals for the regions and IMD quintiles, all overlapped zero. Thus, there was no statistically significant difference between the regions and deprivation quintiles, with regards to the number of fruit and vegetable portions consumed by individuals.

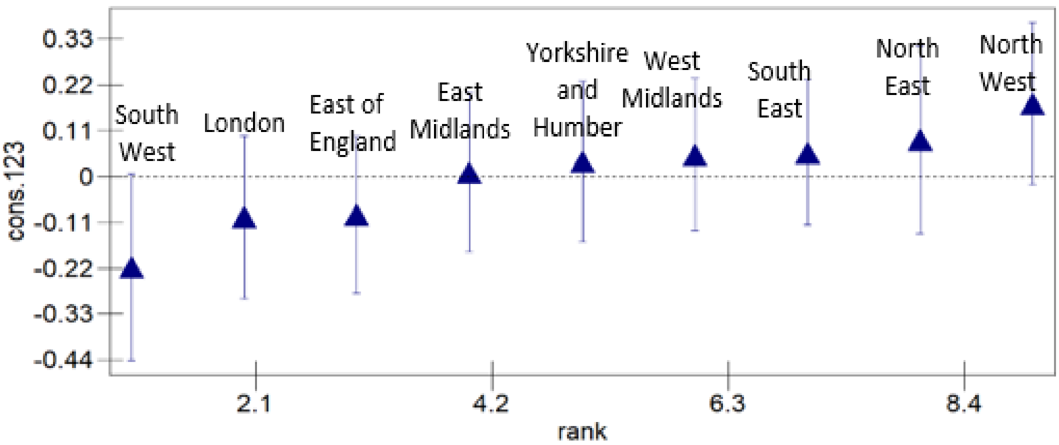


Figure 4.11: GOR level residuals and their associated 95% confidence intervals for the consumption of fruit and vegetable (combined) portions, NDNS, 2008-16

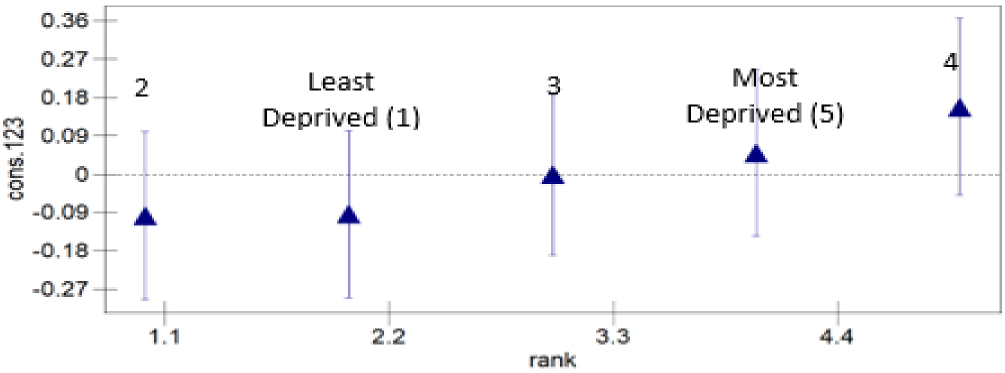


Figure 4.12: IMD level residuals and their associated 95% confidence intervals for the consumption of fruit and vegetable (combined) portions, NDNS, 2008-16

4.3.12.2 Disaggregated fruit portions and vegetable portions consumed

Similar to the findings presented in Section 4.3.11 and Section 4.3.12.1 (5-A-Day achievement and the combined fruit and vegetable portions models respectively), no statistically significant difference was found between the regions and deprivation quintiles, in terms of the number of fruit portions (only) and vegetable portions (only) consumed. Age, ethnicity, highest level of education and equivalised household income were the variables which were consistently found to be the most associated with fruit portions, vegetable portions and combined fruit and vegetable portions consumed. A relatively consistent age gradient was observed, with the odds of consuming fewer than five portions of fruits (only) and vegetables (only) generally decreasing as the age groups increased. These findings were consistent with those observed in the final combined fruit and vegetable portions model presented in Section 4.3.12.1. Similarly, the odds of consuming fewer than five portions of fruits (only) and vegetables (only) was less for persons from other ethnic groups, persons with higher levels of education (particularly those with degree level or equivalent qualifications) and higher incomes, compared to white persons and persons with lower levels of education and income. Detailed results for the final fruit portions and vegetable portions models have been presented in Appendix B.10 - B.17. This was done to avoid unnecessary repetition, given the similarity between fruit portions, vegetable portions and 5-A-Day achievement (see Section 4.3.11) and the combined fruit and vegetable portions (see Section 4.3.12.1) model results previously presented.

4.4 Discussion

The primary aim of Research Question 1 (the focus of this chapter) was to explore the extent to which fruit and vegetable consumption varied by region, amongst persons aged 16-64 years in England. Fruit and vegetable consumption data from the NDNS and HSE were compared to assess possible similarities/differences. In both surveys, over 70% of the sample did not achieve the national 5-A-Day target. Only 26% and 27% of the NDNS and HSE sample (respectively) achieved the target. Thus, there was only a one percent difference between the two data sources. These figures were relatively consistent with the 29% achievement noted in previous Governmental publications (NHS Digital, 2019). Although the surveys differed slightly when data were disaggregated into fruit portions, they were relatively on par with respect to vegetable, and fruit and vegetable portions consumed. The average number of fruit and vegetable portions consumed by the NDNS and HSE sample was 3.9 and 3.6 portions (respectively). Based on these findings, data captured in both surveys were comparable and in line with the 3.8 portions reportedly consumed by the English adult population in 2017 (NHS Digital, 2019).

During the preliminary exploratory analysis, a statistically significant (albeit weak) association was found between region of residence and 5-A-Day target achievement, with percentage achievement seemingly lower in the North and the Midlands than in the South of England (especially London and the South East). These initial findings were not surprising, as they reflected the English North-South divide in diet and health previously noted (Bambra, 2016; DEFRA, 2017; Moon et al, 2007; Morris et al, 2016; Oddy, 2003; Roberts, 2014; Scarborough and Allender, 2008; Tiffin and Salois, 2012). However, when two-way cross classified (random intercept) binary logistic and cumulative logistic regression multi-level models were tested, no statistically significant difference was found between deprivation quintiles and GORs, as it relates to fruit, vegetable and fruit and vegetable portions consumed and the overall achievement of the 5-A-Day target. These findings indicate that although the results of the descriptive/exploratory analysis hinted at a possible North-South divide in fruit and vegetable consumption, this association was no longer statistically significant once compositional factors such as age and SEP were considered.

The results of Research Question 1 can be compared with Morris et al (2016), as this is one of the few studies which attempted to explore geographic variations in diet in England/UK. In contrast to these findings for Research Question 1, Morris et al (2016) found statistically significant differences in dietary patterns at both the regional and Output Area Classification (OAC) small area level, having adjusted for socio-demographic factors such as age, sex and SEP. These conflicting findings could be because firstly, Research Question 1 used nationally representative data from the NDNS while Morris et al (2016) used data from the UK Women's Cohort Study (UKWCS), which consisted of a sample of mostly vegetarian, middle-aged, middle-class, white women. Secondly, Morris et al (2016) explored overall dietary patterns, whereas Research Question 1 only investigated fruit and vegetable consumption. The analysis of dietary patterns was not within the scope of Research Question 1 but will be done in Research Question 2. This type of analysis could reveal findings similar to those found in Research Question 1 (no regional variations) or could highlight a statistically significant association between region and dietary patterns in England.

Finally, the exploration of fruit and vegetable consumption at the sub-regional/small area level was not within the scope of Research Question 1, as the primary aim was to explore possible variations at the regional level. Smith et al (2021) did not focus on regional level variations but found statistically significant differences in adult SSB and fruit and vegetable consumption across small areas (middle layer super output areas- MSOAs) in England from small area estimation (using NDNS and 2011 census data). The statistically insignificant association between region and fruit and vegetable consumption found in Research Question 1, coupled with the findings from Morris et al (2016) and Smith et al (2021), could signal the need for further exploration across smaller areas in England (which will be addressed in Research Question 3 of this project). There could be geographic inequalities in fruit and vegetable consumption, but these may be masked at the higher regional level and more pronounced at a lower geographic scale (Cummins and Macintyre, 2002; Cummins, 2014) such as the MSOA level used in Smith et al (2021).

Although there were no statistically significant differences in fruit and/or vegetable consumption geographically, age, highest level of education, equivalised household income and ethnicity were the variables most strongly associated with fruit and/or vegetable consumption. The achievement of the 5-A-Day target and the consumption of five or more portions of fruits and/or vegetables was most apparent amongst older persons, persons in non-white ethnic groups and persons with higher levels of education and equivalised household income. Occupational grade (NS-SEC) had a slight effect on the number of vegetable portions consumed. However, highest level of education and equivalised household income were the SEP measures most strongly associated with fruit and/or vegetable consumption.

Achievement of the 5-A-Day target was highest for persons aged 50-64 years, compared to persons aged 16-24 years. These findings were consistent with previous publications which highlighted persons aged 16-24 years as the least likely to consume the recommended five portions of fruits and vegetables daily (NHS Digital, 2017). Additionally, the findings from Research Question 1 corroborated previous evidence which identified an age gradient for diet, whereby younger adults generally consumed less healthy foods compared to older adults (Imamura et al, 2015; Yau, Adams and Monsivais, 2019).

Besides age, it has been also been widely acknowledged that persons from low SEP groups generally consume fewer fruits and vegetables and have comparatively worse diets and health outcomes than those in higher SEP groups (Acheson, 1998; Roberts, Cavill, Hancock and Rutter, 2013; Roos et al, 2000; Tiffin and Salois, 2012). The findings from Research Question 1 were consistent with these studies as the consumption of the recommended five portions of fruits and/or vegetables was most apparent amongst highly educated persons (especially those with a degree or equivalent qualification) and persons with higher equivalised household incomes (especially the £100,000 or more income group).

Previous studies have suggested that persons with a higher level of education may have increased skills/competencies which enable them to better engage with dietary messages (such as 5-A-Day) and adopt healthier long-term eating habits (Maguire and Monsivais, 2015). However, the cost of food has been found to be an important mediating factor between SEP (household income and educational attainment) and diet (Darmon and Drewnowski, 2008; Monsivais, Aggarwal and Drewnowski, 2012). Besides demographic factors which are fixed, the cost of food relative to income (household), may be one of the most significant barriers which prevent individuals (especially those from low-income households) from achieving dietary recommendations such as 5-A-Day (Maguire and Monsivais, 2015). The cost of meeting the daily target of five portions of fruits and vegetables has been estimated to cost persons an additional £0.87 per day or £6 per week, compared to diets which fail to meet this target (Jones, Tong and Monsivais, 2018). Additionally, households within the lowest 20% of income continued to have the highest percentage of spend on food at 14.3% (DEFRA, 2018). With competing cost priorities, low-income households have been found to substitute healthier foods such as fruits and vegetables with relatively cheaper ultra-processed, energy-dense alternatives such as biscuits, cakes, bacon, butter and soft drinks (DEFRA, 2018). Although the cost of food could partially explain the social gradient in fruit and vegetable consumption found in Research Question 1, the association between SEP and diet is complex and requires further research beyond this project to better/fully understand the multiple pathways.

4.4.1 Strengths and limitations

Prior to this project, the extent to which fruit and vegetable consumption varied across the regions of England was a topic which was not fully explored. The few studies which previously attempted to explore regional variations in diet in England, generally noted the English North-

South divide and differences between the Scottish and English diet, focused on the relationship between diet and chronic diseases such as obesity or were not representative of the English population (Bambra, 2016; Barton, Chambers, Anderson and Wrieden, 2018; DEFRA, 2017; Moon et al, 2007; Morris et al, 2016; Oddy, 2003; Roberts, 2014; Scarborough and Allender, 2008; Scarborough, Morgan, Webster and Rayner, 2011; Shelton, 2009; Tiffin and Salois, 2012). A major strength of this project (specifically Research Question 1) is that it is the only known project which explored regional variations in fruit and vegetable consumption in England specifically.

The NDNS and HSE (both explored in Research Question 1) are two of the major nationally representative repeated cross-sectional surveys used by the English Government to assess fruit and vegetable consumption and dietary patterns in England (Bates et al, 2012). The descriptive comparison made between the NDNS and HSE was an added feature of Research Question 1, as it showed that data captured in both surveys were comparable, despite the different dietary assessment method used in each survey (four-day food diary in the NDNS and 24-hour dietary screener in the HSE) and the vast difference in their sample size (3,286 for the NDNS and 34,195 for the HSE). The analysis of HSE data as it relates to Research Question 1, was strictly descriptive in nature and did not involve the modelling of data, which is a possible limitation. However, the similarity between the surveys justified the sole use of the NDNS (for Research Question 1) to model and further explore regional/geographic variations in the achievement of the national 5-A-Day target as well as fruit, vegetable and fruit and vegetable portions consumed.

The use of multi-level modelling as opposed to single-level/fixed effects modelling in Research Question 1 was a project strength, as it acknowledged the multi-stage design and structure of NDNS data (Leckie, 2013) and aided in the analysis of regional effects on fruit and vegetable consumption. The inclusion of IMD as a random effect and not a fixed effect in multi-level models tested could be considered a limitation, as some researchers may view IMD individual/household (compositional) characteristic rather than a neighbourhood level (contextual) measure. However, IMD is the official and most widely used small area/neighbourhood level measure of relative deprivation in England (Department for Communities and Local Government, 2016). Previous studies have noted that similar to GOR, IMD is built into the overall data structure of surveys such as the NDNS and should be acknowledged as a geographic level rather than a characteristic of individuals (Deas, Robson, Wong and Bradford, 2003). Moreover, preliminary single-level, two-level hierarchical and two-way cross-classified models tested in Research Question 1 revealed that two-way cross-classified models (models with GOR and IMD as mutually exclusive levels) had the best fit. These findings helped to further justify the project's use of two-way cross-classified multi-level modelling for the four outcomes investigated in Research Question 1.

Fruit and vegetable consumption/5-A-Day achievement is a well-established indicator of a healthy diet (Agudo, 2005) and a major diet-related target in England (PHE, 2018b). The launch of the Better Health campaign in July 2020 marked the beginning of the English Government's recent shift from the sole focus on childhood obesity to improving the diet (5-A-Day consumption) of adults and children in England. Thus, the project's focus on fruit and vegetable consumption amongst working age adults (16-64 years) was further justified and well-positioned in terms of current national priorities. However, it must also be acknowledged that individuals consume a myriad of foods other than fruits and vegetables and so there is merit in assessing fruit and vegetable consumption in conjunction with dietary patterns (Roberts, Cade, Dawson and Holdsworth, 2018). The assessment of dietary patterns was not within the scope of Research Question 1 as it was the project's intention to begin by exploring fruit and vegetable consumption/5-A-Day and thereafter explore regional variations in dietary patterns in England, (which will be addressed in Research Question 2, Chapter 5). The project's use of this systematic approach was thought to provide a more comprehensive view of dietary patterns in England and possible differences across the regions. Also, besides the exploration of 5-A-Day achievement (binary yes/no outcome variable), the disaggregation of fruit and vegetable consumption data into fruit portions, vegetable portions and fruit and vegetable portions allowed for a more detailed assessment to be made, which was an added strength of Research Question 1.

Overall, the findings presented in Research Question 1 showed that although the English North-South divide has been widely acknowledged, fruit and/or vegetable consumption may be more associated with composition or who persons are (specifically their age, ethnicity, level of education and equivalised household income), than where they are geographically located (context). However, it must be acknowledged that compositional factors may also cluster geographically (for example ethnic minority areas). In addition, this chapter was an exploration of a simple summary measure (only one aspect of diet) at a coarse regional scale. The next step is to examine diet in more detail (Research Question 2) and the possibility of greater variation at a more local scale (Research Question 3).

Chapter 5 What are the dietary patterns of persons aged 16-64 years in England and to what extent do dietary patterns vary regionally?

5.1 Introduction/Context

A dietary pattern can be defined as the variety or combination of different foods and beverages regularly consumed (Sánchez-Villegas and Martínez-Lapiscina, 2018). Although there is no universally accepted definition of diet, the Eatwell Guide generally defines a healthy diet as one high in fruits, vegetables, whole-grains, high fibre starchy carbohydrates and protein rich foods (e.g. beans, eggs, pulses) and low in free sugars, salt and saturated fat (NHS Digital, 2019). The Mediterranean Diet (MD) has been recognised as one of the healthiest dietary patterns in the world and recommended as a "model" diet based on the high consumption of fruits, vegetables, nuts, legumes, whole grains, olive oil; moderate consumption of fish, poultry and wine and a low intake of dairy products (yogurt and cheese), processed and red meat and sweets (Grimaldi and Paoli, 2015; Tong et al, 2016). The traditional MD (which originated in the olive oil growing areas of South Europe and the Mediterranean region such Cyprus, Croatia, Greece, Italy, Morocco, Portugal and Spain), has been lauded for its overall health-promoting benefits and has been said to be responsible for the relatively good health of the Mediterranean people (Ogce, Ceber, Ekti and Oran, 2008; Renzella et al, 2018; Rumm-Kreuter, 2001; Sánchez-Villegas and Martínez-Lapiscina, 2018).

With the vast recognition of the MD, geographic differences in dietary patterns have been noted within Europe, with countries in North and West Europe (e.g. UK, Germany and Belgium) said to consume fewer fruits and vegetables and more ultra-processed foods than those in the South and Mediterranean regions (Croatia, Malta, Portugal, Greece and Italy) (Monteiro et al, 2018). However, the UK has been highlighted by the media as having the worst diet in Europe, with research showing that approximately 19,000 deaths due to heart disease could be averted if healthier diets such as the MD were to be adopted in the UK (Tong et al, 2016). Other studies conducted in the UK have noted stark differences between the English diet and diets in the remaining constituent countries such Wales and especially Scotland (Barton, Chambers, Anderson and Wrieden, 2018). It has been suggested that approximately 40% of deaths due to heart disease, stroke and diet-related cancers could be delayed or averted if a diet equivalent to that of the English diet were to be adopted in Scotland (Scarborough, Morgan, Webster and Rayner, 2011a).

Besides the world-renowned English fish and chips, the traditional English diet, referred to as “meat and two veg”, usually consists of cooked meat (mostly beef or mutton), potatoes, and two types of green vegetables, or a meat dish in the form of a casserole/pie (Riley, 2010). The Scottish diet on the other hand has been found to be a diet higher in saturated fat, salt, alcohol, red and processed meat, confectionery and soft drinks, and lower in fruits, vegetables and fibre, compared to the average English diet (Shelton, 2009). Despite generally negative assessments of the Scottish diet, Popkin (1993) and Newton et al (2015) have also noted the rapid pace in which dietary patterns in England and the world are changing towards a more “Western” dietary pattern (high in refined carbohydrates, ultra-processed foods, saturated fats, salt, free sugars and low in fruits and vegetables.

Statistics from DEFRA (2018) highlight changes in the English diet as evidenced by long-term reductions in the consumption of fruits, vegetables, bread and staples in the traditional English diet such as potato and the upward shift towards the consumption of unhealthier and more ultra-processed foods. Studies have found that individuals in England consume 30% fewer fruits and vegetables and almost double the amount of fat (especially saturated fat) and free sugar recommended (Global Food Security, 2016; NHS Digital, 2019). Such findings add credence to the nutrition transition and the fact that dietary patterns in England (at the national and possibly the regional level) are far from perfect and not in alignment with current dietary recommendations (NHS Digital, 2019).

Popkin (1993) noted the importance of exploring the geographic/spatial distribution of dietary patterns, as various sub-populations within a country may manifest different dietary patterns. However, the poor availability of diet-related data remains a major challenge which prevents researchers from fully exploring dietary patterns within countries such as England (Brug, 2008; Campbell et al, 2020; National Academies of Sciences, Engineering & Medicine, 2019). Fruit and vegetable consumption is a well-established and commonly used indicator of a healthy diet (Agudo, 2005). Therefore, individual level dietary consumption is oftentimes assessed based on adherence to the recommended five daily portions of fruits and vegetables or the national 5-A-Day target (Roos et al, 2000). However, it must also be recognised that individuals do not consume only fruits and vegetables or any single food/food group. Instead, they consume a variety of foods with a complex combination of nutrients and interactions (Miller et al, 2010). Dietary consumption should ideally, and as far as possible, (especially where data beyond fruit and vegetable consumption are available) be assessed by exploring these food groups in conjunction with an individual's dietary pattern (Hu, 2002; Ocké, 2013) and how patterns may vary geographically/regionally.

Robinson et al (2004) was one of the few studies which examined dietary patterns in England specifically. However, empirically derived dietary patterns identified in that study were based on a sample of young women, aged 20-34 years who resided in Southampton (a city in South East England). Morris et al (2016) attempted to explore dietary patterns in England/UK and differences at the regional and sub-regional level. However, as mentioned in Chapter 4 (Section 4.4) the study was not representative of the English population, but limited to middle-aged, middle-class women, with mostly vegetarian diets. Roberts, Cade, Dawson and Holdsworth (2018) used data from the NDNS (a nationally representative survey) to examine empirically derived dietary patterns amongst adults in the UK. Although Roberts, Cade, Dawson and Holdsworth (2018) was one of the few nationally representative studies, the exploration of regional variations in England was not within the study's scope. These findings serve as further evidence of the limited studies which have systematically explored geographic/regional differences in dietary patterns in England (Macintyre, Maciver and Sooman, 1993).

This chapter aims to fill a gap in the existing literature by: i) exploring dietary patterns (amongst traditional working age adults) in England as a whole and ii) examining whether there are regional differences in dietary patterns in England. The findings of this chapter could go towards the creation of more evidence-based policies/interventions by identifying unexpected regional differences which could go undetected if not explored. The following section (Section 5.2) presents the source of data and the method of analysis used within this chapter to address the project's second research question.

5.2 Method

5.2.1 Source of data: National Diet and Nutrition Survey (NDNS) 2008-16

The NDNS is currently the only annual, nationally representative survey which captures all foods and beverages consumed by individuals (via the food diary method) in England/UK (MRC Elsie Widdowson Laboratory and NatCen Social Research, 2018). NDNS data were analysed solely for this analysis. The survey was deemed appropriate for Research Question 2 based on the availability of data for over 7,000 foods (collected via four-day food diaries), consumed by the pooled sample (Waves 1-8, 2008-16 survey period) of 3,286 persons aged 16-64 years. Background information on the NDNS, its survey design, method of data collection, assessment of diet, initial sample size, sample characteristics and additional details regarding the accessibility and overall data cleaning procedures conducted specifically for this project (e.g. persons excluded from the analysis and the handling of missing data), were fully described in Section 3.3.1 and Section 4.2.

All foods (over 7,000, inclusive of beverages) consumed by individuals were aggregated by the NDNS survey data collectors into 61 “main food groups” and 150 “subsidiary food groups” (Roberts et al, 2018). Food groups were expressed as the average grams consumed daily by individuals sampled. Although subsidiary (sub) food groups provided greater detail than main food groups, for simplicity and ease of interpretation, main food groups were analysed within the current project. The project’s selection and use of main, as opposed to subsidiary food groups, enabled comparison with similar studies previously conducted (Roberts, Cade, Dawson and Holdsworth, 2018). Given the project’s focus on persons aged 16-64 years (traditional working age adults), the “Commercial Toddler Foods and Drinks” food group was deemed irrelevant and was excluded from the analysis. This reduced the number of food groups analysed in this project to 60 main food groups, presented in Table 5.1.

Table 5.1: NDNS main food groups (average grams consumed daily by persons aged 16-64 years in England) used as dietary input variables for the Principal Components Analysis (PCA) conducted, NDNS, 2008-16

Food Group Dietary Input Variable	
1. One percent milk	31. Other margarine, fats and oils
2. Bacon and ham	32. Other meat and meat products
3. Beef, veal and dishes	33. Other milk and cream
4. Beer, lager, cider and perry	34. Other potatoes, potato salads and dishes
5. Biscuits	35. Other white fish, shellfish and fish dishes
6. Brown, granary and wheatgerm bread	36. Pasta, rice and other cereals
7. Buns, cakes, pastries and fruit pies	37. Pork and dishes
8. Burgers and kebabs	38. Puddings
9. Butter	39. Polyunsaturated fatty acid (PUFA) margarine and oils
10. Cheese	40. Reduced fat spread not polyunsaturated

Food Group Dietary Input Variable	
11. Chicken and turkey dishes	41. Reduced fat spread polyunsaturated
12. Chips, fried and roast potatoes and potato products	42. Salad and other raw vegetables
13. Chocolate confectionery	43. Sauces, pickles and gravies
14. Coated chicken and turkey	44. Sausages
15. Crisps and savoury snacks	45. Semi skimmed milk
16. Dry weight beverages	46. Skimmed milk
17. Eggs and egg dishes	47. Soft drinks low calorie
18. Fruit	48. Soft drinks not low calorie
19. Fruit juice including smoothies	49. Soup homemade and retail
20. High fibre breakfast cereals	50. Spirits and liqueurs
21. Ice cream	51. Sugar confectionery
22. Lamb and dishes	52. Sugar, preserves and sweet spreads
23. Liver and dishes	53. Tea, coffee and water
24. Low fat spread not polyunsaturated	54. Vegetables not raw
25. Low fat spread polyunsaturated	55. White bread
26. Meat pies and pastries	56. White fish coated or fried
27. Nuts and seeds	57. Whole milk
28. Oily fish	58. Wholemeal bread
29. Other bread	59. Wine
30. Other breakfast cereals	60. Yogurt, fromage frais and dairy desserts

5.2.2 The A priori and Posteriori approach to dietary pattern analysis

There are two main approaches typically used to examine dietary patterns: i) the a priori/ hypothesis driven approach and ii) the posteriori/empirical/data driven approach. With the a priori approach, diet scores/indices are created (and defined) based on well-established scientific evidence and existing dietary recommendations/guidelines of what generally constitutes a healthy diet (Hu, 2002). A priori derived diet scores/indices are primarily used to measure how well individuals adhere to current dietary guidelines (diet quality) or to a specific dietary pattern (such as the Mediterranean Diet) of interest to researchers/policymakers. Conversely, the posteriori or data-driven approach, being more exploratory in nature, does not depend on dietary guidelines and requires no previous hypothesis or theory of what dietary patterns are to be expected in a given population. With this approach, the researcher simply lets the "data do the talking" (Moeller et al, 2007).

The a priori approach is advantageous as it is relatively easy to implement, capable of producing meaningful/interpretable results and useful for assessing adherence to existing dietary guidelines (Ocké, 2013). However, with ever increasing evidence of sub-optimal diet and non-compliance with current dietary recommendations globally (Gakidou et al, 2017; Newton et al, 2015), it was necessary at this stage to conduct more in-depth research on the actual eating habits of the population using posteriori/data driven approaches. This type of research can contribute to the evidence-base needed for public policy, as dietary guidelines and related policies/interventions which do not acknowledge or deviate greatly from current consumption patterns are less likely to succeed at the population level (De Ridder et al, 2017; Fulgoni and Drewnowski, 2019). The posteriori approach was deemed more appropriate, given that Research Question 2 seeks to explore the types of foods consumed by individuals in England and not how well individuals adhere to current dietary guidelines.

5.2.3 Statistical analysis

5.2.3.1 Dietary pattern derivation: Principal Components Analysis (PCA)

Principal components analysis (PCA) (a form of exploratory factor analysis) is a posteriori approach frequently used to extract dietary patterns within a given population (Moeller et al, 2007). PCA is a descriptive (not inferential) data reduction method which uses matrix algebra to summarise and reduce a large number of correlated dietary input variables (foods/food groups) into a smaller set of independent/uncorrelated dietary patterns referred to as components. In other words, PCA reduces the overall data structure to a more manageable size and transforms original dietary input variables into linear components (which explain the variance in the data and underlying dietary patterns), whilst retaining as much of the original information as possible (Field, 2018). PCA was the posteriori method selected to derive dietary patterns (in Research

Question 2) as it facilitated the exploration of foods actually consumed by individuals and possible variations in dietary patterns in England without any previous hypothesis or expectation.

The PCA procedure was conducted in the SPSS version 24 statistical software package (IBM, 2018). Figure 5.1 provides an overview of the PCA process and the steps (discussed below) used in Research Question 2 of this project.

STEP 1: PCA generation and preliminary assessment	**STEP 2: Component extraction and labelling of dietary patterns**	**STEP 3: Multi-level regression modelling**
•PCA conducted in SPSS for the national NDNS sample and for each of the nine GORs in England (10 PCAs conducted) using the 60 pre-defined NDNS main food groups •Orthogonal varimax rotation used •Component scores or dietary pattern scores for each individual computed and saved in SPSS during PCA process •KMO test and the Bartlett's test of sphericity used to assess the suitability of PCA method for NDNS data	•Initial PCA components extracted •Percentage variance assessed to determine the number of components to be retained •Component one retained for all 10 PCAs (PCA for the national sample and nine region-specific PCAs) •Component loadings with an absolute value of 0.25 assessed as being moderate/significant contributors to dietary patterns identified	•Dietary pattern scores saved (from Step 1) for the national sample selected as the sole continuous outcome variable in final two-way cross-classified multi-level model •Model results assessed to examine the effect of region on dietary patterns identified in the national sample

Figure 5.1: Summary of the PCA process and the steps used in Research Question 2

All 60 of the NDNS pre-defined food groups (listed in Table 5.1) were included in the PCA, as they represented the total dietary intake of individuals sampled and were useful to identify underlying dietary patterns in the data. PCA was firstly conducted for the national sample. Thereafter, the national sample dataset was split by region and the procedure repeated for each of the nine regions (GORs) in England (North East; North West; Yorkshire and the Humber; East Midlands; West Midlands; East of England; London; South East; and South West). The separate analysis by region was purposefully done to identify dietary patterns of each sub-group and explore (descriptively) possible regional differences (the aim of Research Question 2). The Kaiser-Meyer-Olkin (KMO) test and the Bartlett's test of sphericity were the preliminary tests carried out to measure the sampling adequacy and suitability of the data for the PCA method. These tests were conducted for the national sample as well as the nine region-specific PCAs. In all instances, the KMO value reached the acceptable value of 0.50 and the Bartlett's test of sphericity was less than the 0.05 significance level (Field, 2018). These results indicated/verified that PCA was indeed an appropriate method.

The number of components produced following the PCA procedure corresponds with the number of input variables used (Field, 2018). Thus, the 60 diet input variables/food groups used in this project (Table 5.1), resulted in 60 initial components. Orthogonal rotation (specifically varimax rotation) was applied because it is an accepted means of simplifying the data structure and the interpretation of results (Field, 2018; Smith, Emmett, Newby and Northstone, 2013). The first component in the PCA generally accounts for the highest possible variance in the data (Marchioni et al, 2005). The second component has the second highest variance and thereafter the variance continuously decreases up to the final component (which was 60). In studies of this nature, it is customary for all components with "eigenvalues" (or variances) greater than one to be selected and analysed since they explain the largest proportion of the variance in the data (Field, 2018). However, following the assessment of scree plots (not presented) and the overall characteristics (inclusive of the proportion of variance explained) for the first four components, only the first component was retained for the national sample and the nine region-specific PCAs. The rationale was that the first component explained most of the variance in the data and provided the most interpretable and meaningful pattern of foods consumed by the sample. This approach was consistent with previous research on dietary patterns (DelaCruz, 2010; Robinson et al, 2004).

The PCA method operates such that each component describes a dietary pattern and the linear combination allows for a dietary pattern score to be computed for each individual sampled (Hu et al, 1999; Smith, Emmett, Newby and Northstone, 2013). The extracted dietary pattern score for component one for each individual was calculated (see Equation 5.1) as the sum of the food intake variables (the 60 food groups) weighted by the component loadings generated by the method (Sauvageot et al, 2016). These scores (measured as a continuous variable) indicated the extent to which the individual's diet conformed to the pattern identified for component one.

Equation 5.1: Calculation of individual dietary pattern scores

$$Component\ score\ for\ pattern\ i = \sum_j \left\lfloor \left(b_{ij}/\lambda_i \right) X_j \right\rfloor$$

b_{ij} = the (component) loading for the j^{th} food group on the i^{th} pattern

λ_i = the associated eigenvalue (the eigenvalue represents the total amount of variance explained by a particular component)

X_j = the standardised value of the j^{th} food group

Component loadings describe how much each input variable/food group contributes to a particular component. Thus, the higher the component loading (positive or negative), the stronger the association between the input variable/food group and the component. Component loadings with a positive sign indicate a positive association with the component, whereas a negative sign indicates an inverse association. In keeping with previous statistical guidelines, food groups which had a component loading greater than the absolute value of 0.30 (that is, ≥ 0.30 or ≤− 0.30) were considered high/significant contributors (Field, 2018; Marchioni et al, 2005; Miller et al, 2010) to component one in the national sample and for each of the nine region-specific PCAs. Component loadings with an absolute value of 0.25-0.29 were considered moderate contributors, in keeping with previous research (Roberts, Cade, Dawson and Holdsworth, 2018). Food groups with component loadings below the absolute value of 0.25 (cut-off point) were considered weak contributors and were therefore not included in the interpretation of components/dietary patterns. For ease of interpretation, components were generally described and labelled based on the food groups which had moderate/high component loadings.

5.2.3.2 Two-way cross classified multi-level modelling

Previous studies have recommended the use of regression analysis for projects interested in exploring the relationship between dietary patterns and other variables of interest (Hu, 2002). Research Question 2 seeks to explore regional variations in dietary patterns in England. As such, multi-level modelling was considered an appropriate method to further examine the effect of region on dietary patterns in the national sample. Standardised (with a mean of zero and a standard deviation of one) dietary pattern scores for component one of the national sample (generated by SPSS during the PCA process and previously described in in Section 5.2.3.1) were used as the sole continuous outcome variable for Research Question 2. Although dietary pattern scores for component one were also available for the nine region-specific PCAs, these scores were unnecessary/irrelevant, as the focus of the analysis was on modelling the effect of region on dietary patterns within the national sample.

As good practice, preliminary two-way cross-classified multi-level models and corresponding two-level hierarchical multi-level models and fixed-effects (single-level) models were tested in the MLwIN multi-level (version 3.02) software package (Charlton et al, 2017). This approach was taken to verify the structure of the data, assess goodness-of-fit of preliminary models and provide further justification for the project’s final decision to use two-way cross-classified multi-level modelling. Age, sex, ethnicity, marital status, equivalised household income, highest level of education, NS-SEC occupational grade and BMI were entered as fixed-effects in the preliminary single, two-level hierarchical and two-way cross-classified models tested. However, IMD was included as a fixed effect and GOR as a random effect in the preliminary two-level hierarchical model. GOR and IMD were both included as random effects in the preliminary two-way cross-classified random intercept model tested. Additional details regarding these independent variables, such as the description of categories, recodes carried out and the rationale for their inclusion, have been fully described in Section 4.2.1.

The Markov Chain Monte Carlo (MCMC) estimation method was applied and the MLwiN default settings (5,000 iterations and burn-in and thinning values of 500 and 1 respectively) were used to build preliminary models. The Deviance Information Criterion (DIC) (used to compare and assess the goodness-of fit) for the preliminary single, two-level hierarchical and two-way cross-classified model was 8,898.03, 8,894.55 and 8,893.94 respectively. Based on the DIC, multi-level models, especially two-way cross-classified models appeared to be a better fit and more appropriate than single-level models. This meant that the project was justified in proceeding with two-way (random intercept) cross-classified multi-level modelling for Research Question 2.

Two-way (random intercept) cross-classified models were built (in MLwIN) by entering variables in a sequential manner, beginning with the null model and ending with the full adjusted model (age, sex, ethnicity, marital status, equivalised household income, NS-SEC occupational grade and highest level of education and BMI). The first category was selected as the reference category for each of the explanatory variables entered. The fully adjusted model (model with all explanatory variables) appeared to be the model with the best fit and the lowest DIC compared to preceding models tested.

Following the decision to proceed with the fully adjusted model, trace plots (based on the default 5,000 iterations) for each parameter were generated and assessed for data convergence. As expected, the data did not appear to mix well, and the number of iterations needed to be increased from the default of 5,000 iterations. The Raftery-Lewis diagnostic (Nhat) was assessed and based on the results obtained, the number of iterations subsequently increased to 25,000. Trace plots were reassessed, and the data appeared to have mixed well following this increase (trace plot model diagnostics presented in Appendix C.2). Kernel density graphs were assessed and appeared to be smooth and symmetric in shape, suggesting an approximately Normal distribution. Similarly, normal probability quantile-quantile (QQ) plots were generated in MLwIN and residuals at each level (the GOR and IMD level) appeared to follow an approximately normal distribution. Caterpillar plots (presented in Section 5.3.2) for GOR and IMD level residuals and their ranks were also generated and compared with the overall average. Overall, based on the diagnostic statistics assessed, the increased number of iterations (25,000) appeared to be sufficient and was considered the final model. Equation 5.2 represents the final two-way cross-classified random intercept model for the PCA-derived dietary pattern (outcome) scores for component one of the national sample.

Equation 5.2: Two-way cross-classified random intercept model for PCA derived outcome scores

$$PCA - derived\ dietary\ pattern\ score_{\ i} = \beta_{0i} + \beta age_i + \beta sex_i + \beta ethnicity_i + \beta marital\ status_i + \beta education_i + \beta income_i + \beta NSSEC8\ Occupation_i + \beta BMI_i$$

$$\beta_{0i} = \beta_0 + u^{(2)}_{0\ imd\ (i)} + u^{(3)}_{0\ gor(i)} + e_{0i}$$

$$\left[u^{(3)}_{0\ gor(i)}\right] \sim N\left[0, \Omega^{(3)}_u\right] : \Omega^{(3)}_u = \left[\Omega^{(3)}_{u0}, 0\right]$$

$$\left[u^{(2)}_{0\ imd\ (i)}\right] \sim N\left[0, \Omega^{(2)}_u\right] : \Omega^{(2)}_u = \left[\Omega^{(2)}_{u0}, 0\right]$$

$$[e_{0i}] \sim N\ (0, \Omega_e) : \Omega_e = [\Omega_{eo}, 0]$$

5.3 Results

Univariate statistics, particularly the characteristics of the NDNS waves 1-8 sample, will not be presented in this chapter/section, as they were previously discussed in Chapter 4 (Section 4.3). The results presented within this section (Section 5.3) relate specifically to the PCA dietary pattern analysis and multi-level modelling procedures conducted for the project's second research question.

5.3.1 PCA derived dietary pattern results

Having completed the PCA procedure (using varimax rotation) and visually assessed scree plots, the first four components of the national sample explained approximately 11.1% of the variance in the data, with each component accounting for 3.5%, 2.6%, 2.5% and 2.5% of the variance, respectively (Table 5.2).

Table 5.2: Percentage of variance explained for the first four components of the national sample and the nine region-specific PCAs conducted, NDNS, 2008-16

Component	**Percentage (%) of variance explained following varimax rotation**									
	National sample	**North East**	**North West**	**Yorkshire and the Humber**	**East Midlands**	**West Midlands**	**East of England**	**London**	**South East**	**South West**
1	3.5	3.9	3.2	4.3	3.6	3.2	3.4	3.6	3.2	3.3
2	2.6	3.6	3.2	3.2	3.2	3.0	3.1	3.3	2.9	3.0
3	2.5	3.5	3.1	2.8	3.0	2.9	2.9	3.1	2.9	3.0
4	2.5	3.4	3.0	2.8	2.9	2.8	2.7	2.8	2.7	2.9
Total*	**11.1**	**14.4**	**12.5**	**13.0**	**12.7**	**12.0**	**12.0**	**12.8**	**11.8**	**12.1**

*Percentage totals may differ due to rounding

At the regional level, the percentage of variance explained by the first four components ranged from 11.8% in the South East to a maximum of 14.4% in the North East (Table 5.2). In general, there appeared to be little or no difference between the components for the national sample and the nine regions, in terms of the percentage of variance explained (Table 5.2). For instance, in the North West region, the first four components explained approximately 12.5% of the variance in the data, with each component accounting for 3.2%, 3.2%, 3.1% and 3.0%, respectively. Similarly, in the South West, the first four components explained 12.1% of the variance in the data, with each accounting for 3.3%, 3.0%, 3.0% and 2.9%, respectively (Table 5.2). As expected, the first component explained most of the variance in the data and provided the most interpretable and meaningful pattern of foods consumed by the sample. Based on these findings, the project was thought to be justified in its decision to retain and analyse only the first component for the national sample and the nine region-specific PCAs carried out.

Table 5.3 presents the PCA results (rotated factor loadings) for component one of the national sample and the nine regions. Food groups with moderate/high component loadings were interpreted as the major contributors to the dietary patterns identified (Table 5.3). The full/detailed results of the PCA for the national sample and the nine regions (with moderate/high and weak component loadings) are presented in Appendix C.1. The first component of the national sample was labelled “Health conscious” because it was characterised by moderate/strong (positive) component loadings for fruit (0.66), yogurt, fromage frais and dairy desserts (0.58), salad and other raw vegetables (0.52), skimmed milk (0.44), tea, coffee and water (0.42), high fibre breakfast cereals (0.27) and other margarine, fats and oils (0.25). It should be noted that the "Health conscious" label given to the national sample refers only to the first component and not overall diet. Other aspects of diet were captured in the remaining components (i.e. components two, three, four etc). However, as stated in Section 5.2.3.1, component one was retained for the national sample and the nine region-specific PCAs because it explained most of the variance in the data and provided the most interpretable and meaningful pattern of foods consumed by the sample.

When disaggregated by region, six of the nine regions (North East, Yorkshire and the Humber, East Midlands, London, South East and South West) had moderate/strong positive loadings for food groups such as fruit; salad and other raw vegetables; tea, coffee and water; yogurt, fromage frais and dairy desserts; skimmed milk; wholemeal bread; high fibre breakfast cereals; oily fish; brown, granary and wheatgerm bread and moderate/strong negative loadings for chips, fried and roast potatoes and potato products and meat pies and pastries (Table 5.3). Given the similarity with the national sample, the dietary patterns (for component one) identified in the North East, Yorkshire and the Humber, East Midlands, London, South East and South West were also labelled "Health conscious", as they primarily consisted of food groups generally considered to be beneficial to human health.

Table 5.3: Rotated PCA component loadings for the first component of the national sample and the nine regions of England, NDNS, 2008-16

Food groups with moderate/strong positive component loading (≥0.25 or ≤−0.25)	**Component loadings**	**Dietary pattern label**
National sample		
Fruit	0.66	Health conscious
Yogurt, fromage frais and dairy desserts	0.58	
Salad and other raw vegetables	0.52	
Skimmed milk	0.44	
Tea, coffee and water	0.42	
High fibre breakfast cereals	0.27	
Other margarine, fats and oils	0.25	
North East		
Yogurt, fromage frais and dairy desserts	0.85	Health conscious
Fruit	0.77	
Salad and other raw vegetables	0.35	
Chicken and turkey dishes	0.31	
Meat pies and pastries	-0.29	

Food groups with moderate/strong positive component loading (≥0.25 or ≤−0.25)	Component loadings	Dietary pattern label
Vegetables not raw	0.27	
Wholemeal bread	0.25	
Chips, fried and roast potatoes and potato products	-0.25	
North West		
Burgers and kebabs	0.74	Western/Takeaway
Chips, fried and roast potatoes and potato products	0.63	
Soup homemade and retail	-0.36	
Coated chicken and turkey	0.34	
Other milk and cream	0.33	
Soft drinks not low calorie	0.30	
Salad and other raw vegetables	-0.25	
Yorkshire and the Humber		
Fruit	0.72	Health conscious
Yogurt, fromage frais and dairy desserts	0.61	
Salad and other raw vegetables	0.60	
Tea, coffee and water	0.51	
Other potatoes, potato salads and dishes	0.33	
Vegetables not raw	0.29	
Chips, fried and roast potatoes and potato products	-0.29	
Skimmed milk	0.28	
Biscuits	0.27	
High fibre breakfast cereals	0.25	
East Midlands		
Salad and other raw vegetables	0.69	Health conscious
Fruit	0.68	
High fibre breakfast cereals	0.62	
Tea, coffee and water	0.35	
Yogurt, fromage frais and dairy desserts	0.33	
Other margarine, fats and oils	0.32	
Chips, fried and roast potatoes and potato products	-0.25	
West Midlands		
Low fat spread poly-unsaturated	0.73	Plain/Traditional English
White fish coated or fried	0.59	
Sugar, preserves and sweet spreads	0.51	
Whole milk	0.50	

Food groups with moderate/strong positive component loading (≥0.25 or ≤−0.25)	Component loadings	Dietary pattern label
Chips, fried and roast potatoes and potato products	0.29	
Sausages	0.26	
East of England		
Other milk and cream	0.68	Sugar, Fats and Oils
Other margarine, fats and oils	0.65	
Fruit	0.63	
PUFA margarine and oils	0.37	
Tea, coffee and water	0.32	
Yogurt, fromage frais and dairy desserts	0.32	
Puddings	0.28	
London		
Fruit	0.67	Health conscious
Skimmed milk	0.62	
Yogurt, fromage frais and dairy desserts	0.58	
Salad and other raw vegetables	0.43	
Brown, granary and wheatgerm bread	0.31	
Oily fish	0.26	
South East		
Salad and other raw vegetables	0.72	Health conscious
Fruit	0.59	
Yogurt, fromage frais and dairy desserts	0.56	
Tea, coffee and water	0.36	
Other margarine, fats and oils	0.31	
South West		
Salad and other raw vegetables	0.71	Health conscious
Fruit	0.54	
Other margarine, fats and oils	0.39	
Butter	0.37	
Tea, coffee and water	0.32	
Chocolate confectionery	0.28	
Other white fish, shellfish and fish dishes	0.26	
High fibre breakfast cereals	0.26	

The North West, West Midlands and East of England were the three regions with dietary patterns (for component one) which deviated from the national sample and the six remaining regions. The component one dietary pattern for the East of England was labelled “Sugar, Fats and Oils” as it was characterised by moderate/strong positive loadings for other milk and cream (0.68), other margarine, fats and oils (0.65), fruit (0.63), PUFA margarine and oils (0.37), tea, coffee and water (0.32), yogurt, fromage frais and dairy desserts (0.32) and puddings (0.28). The North West had a component one dietary pattern which was labelled “Western/Takeaway,” as it had moderate/strong positive loadings burgers and kebabs (0.74), chips, fried and roast potatoes and potato products (0.63), coated chicken and turkey (0.34), other milk and cream (0.33), soft drinks not low calorie (0.30) and moderate/strong negative loadings for soup homemade and retail (-0.36) and salad and other raw vegetables (-0.25). The West Midlands had a component one dietary pattern which was labelled “Plain/Traditional English”, with moderate/strong positive loadings for low fat spread poly-unsaturated (0.73), white fish coated or fried (0.59), sugar, preserves and sweet spreads (0.51), whole milk (0.50), chips, fried and roast potatoes and potato products (0.29) and sausages (0.26).

5.3.2 Results of final two-way cross-classified multi-level model of PCA-derived dietary pattern scores for the national sample

As mentioned previously in Section 5.2.3.2, dietary pattern scores for component one of the national sample were used to model the effect of region on dietary patterns within the national sample. Age, sex, ethnicity, highest level of education and equivalised household income were the variables most strongly associated with the “Health conscious” dietary pattern (for component one) identified for the national sample (Table 5.4).

Table 5.4: Fully adjusted two-way cross-classified multi-level regression model of the standardised PCA-derived dietary pattern scores (for component one/”Health conscious” dietary pattern for the national sample) for persons aged 16-64 years in England, NDNS, 2008-16

Variables in the model	Beta coefficient (SE)	95% Confidence Interval
Intercept	-0.711** (0.161)	___
Age (Ref Cat: 16-24 years)		
25-29	0.102 (0.077)	(-0.047, 0.255)
30-39	0.232** (0.066)	(0.103, 0.359)
40-49	0.471** (0.068)	(0.338, 0.603)
50-64	0.699** (0.069)	(0.564, 0.835)
Sex (Ref Cat: Female)		
Male	-0.159** (0.033)	(-0.225, -0.094)

Variables in the model	Beta coefficient (SE)	95% Confidence Interval
Ethnicity (Ref Cat: White)		
All other ethnic groups	-0.218** (0.053)	(-0.323, -0.115)
Marital Status (Ref Cat: Single Never Married)		
Married/In a legally recognised civil partnership	0.006 (0.046)	(-0.085, 0.096)
Separated/divorced	-0.066 (0.059)	(-0.183, 0.049)
Widowed/other/formerly in a legally recognised civil partnership	0.13 (0.135)	(-0.135, 0.394)
Highest level of education (Ref Cat: No qualifications)		
GCSE or equivalent qualifications	0.116* (0.062)	(-0.005, 0.238)
GCE A level or equivalent	0.224** (0.068)	(0.091, 0.358)
Higher education, below degree level	0.373** (0.075)	(0.223, 0.519)
Degree or equivalent	0.395** (0.066)	(0.265, 0.524)
Still in full-time education/Other foreign qualification/Other	0.243** (0.074)	(0.096, 0.388)
Equivalised household income (Ref Cat: Under £20,000)		
£20,000-£29,999	0.061 (0.045)	(-0.027, 0.151)
£30,000-£39,999	0.144** (0.053)	(0.039, 0.249)
£40,000-£49,999	0.258** (0.063)	(0.134, 0.383)
£50,000-£99,999	0.274** (0.06)	(0.157, 0.391)
£100,000 or more	0.233* (0.111)	(0.015, 0.449)
NS-SEC Occupation Grade (Ref Cat: Never worked/Long term unemployed)		
Routine, manual and other occupations	0.007 (0.109)	(-0.204, 0.224)
Small employers and own account workers	-0.044 (0.118)	(-0.276, 0.186)
Intermediate occupations	-0.001 (0.112)	(-0.219, 0.219)
Higher managerial and professional occupations	-0.046 (0.119)	(-0.28, 0.187)
Body Mass Index: BMI (Ref Cat: Underweight)		
Normal weight	0.157 (0.102)	(-0.043, 0.357)

Variables in the model	Beta coefficient (SE)	95% Confidence Interval
Overweight	0.124 (0.104)	(-0.08, 0.329)
Obese	0.104 (0.107)	(-0.105, 0.313)
Variance of random parameters		
Government Office Region (GOR)	0.004 (0.004)	___
Index of Multiple Deprivation (IMD)	0.007 (0.035)	___
Individual level	0.868 (0.021)	___

* p-value <0.05, ** p-value<0.01, SE= standard error

Although dietary pattern scores for persons aged 25-29 years were not statistically significantly different from persons aged 16-24 years, persons aged 30-39 years had scores which were approximately 0.23 of a standard deviation higher than those aged 16-24 years (Table 5.4). Similarly, older persons aged 40-49 years and 50-64 years had scores which were approximately 0.47 and 0.70 (respectively) of a standard deviation higher than those in the age group 16-24 years. It appeared that older persons, especially persons aged 50-64 years, had better adherence to the first component health conscious dietary pattern compared to persons aged 16-24 years. In terms of sex, men tended to have a generally less health conscious dietary pattern than women, with scores being 0.16 of a standard deviation lower for men than women. Ethnicity was also statistically significantly associated with the health conscious dietary pattern identified for component one of the national sample. Persons from “other ethnic groups” had dietary pattern scores which were approximately 0.22 of a standard deviation lower than for white persons.

Persons with GCSE or equivalent qualifications had scores which were approximately 0.12 of a standard deviation higher than persons with no qualifications. Persons with GCE A Level or equivalent and Higher education, below degree level, had dietary pattern scores which were 0.22 and 0.37 (respectively) of a standard deviation higher than those with no qualifications. However, persons with a degree or equivalent level qualification, had scores which were approximately 0.40 of a standard deviation higher than persons with no qualifications. Based on these findings, it appeared that persons with higher levels of education had a dietary pattern which closely resembled the health conscious pattern observed for component one of the national sample. Similarly, persons with higher levels of equivalised household income seemed to have more elements of a health conscious dietary pattern compared to those in the under £20,000 equivalised household income group. For instance, persons in the £30,000-£39,999 and £40,000-£49,999 income groups, had scores which were 0.14 and 0.26 (respectively) of a standard deviation higher than those in the under £20,000 reference group. Likewise, persons in the higher £50,000-£99,999 group had scores which were 0.27 of a standard deviation higher than persons in the under £20,000 income group. With the exception of the £100,000 or more income group, a slight gradient in equivalised household income was observed.

The variance for the GOR and IMD random parameters was 0.004 and 0.007 (respectively) (Table 5.4), both of which did not appear to be much different from zero. Caterpillar residual plots presented in Figure 5.2 and 5.3 further illustrated this, as the confidence intervals for the nine regions and IMD quintiles, all overlapped zero. This meant that there was no statistically significant difference between the regions and deprivation quintiles, with regard to dietary patterns observed for component one of the national sample after controlling for relevant covariates. However, at the individual level, the plot of individual residuals presented in Figure 5.4 showed that not all the confidence intervals overlapped the horizontal zero line (representative of the average dietary pattern score). Figure 5.4 highlighted differences in dietary patterns at the individual level, as some individuals sampled had scores which deviated significantly from the average score at the 5% level. This finding was illustrated by confidence intervals which were either above or below the average line (Figure 5.4). Overall, this result indicated that differences in dietary patterns were between the individuals sampled, rather than between regions and IMD quintiles.

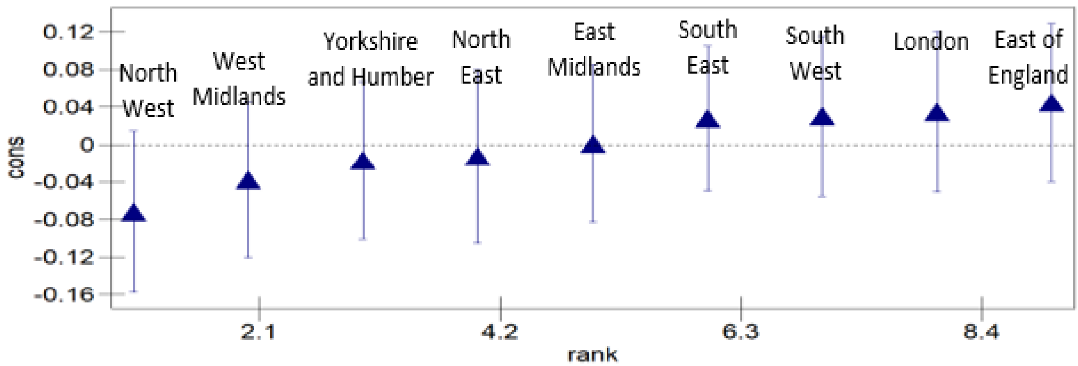


Figure 5.2: GOR level residuals and their associated 95% confidence intervals for diet quality, NDNS, 2008-16

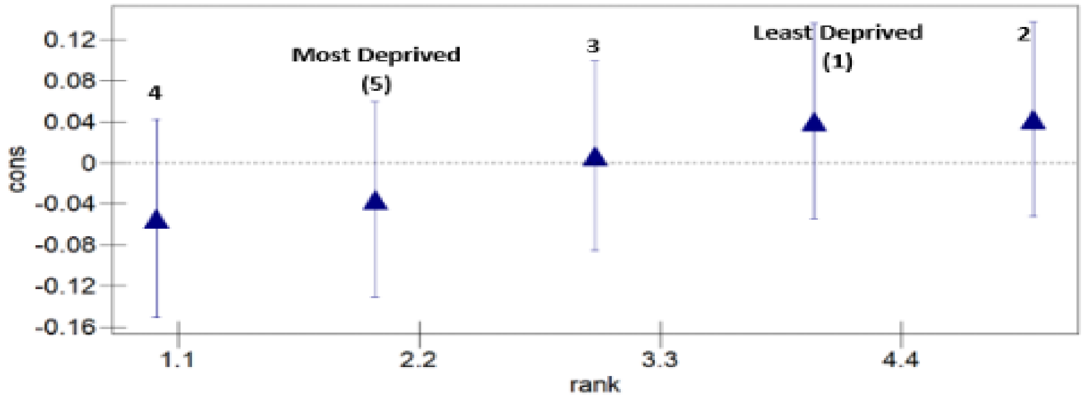


Figure 5.3: IMD level residuals and their associated 95% confidence intervals for diet quality, NDNS, 2008-16

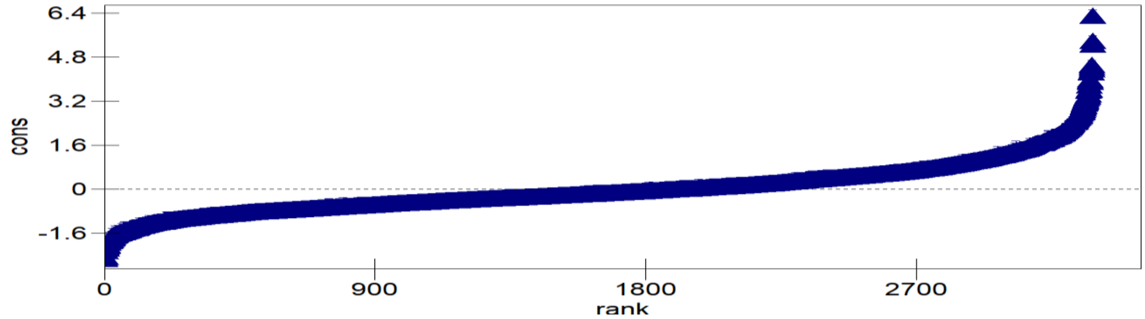


Figure 5.4: Individual level residuals and their associated 95% confidence intervals for diet quality, NDNS, 2008-16

5.4 Discussion

The primary aim of Research Question 2 was to explore the dietary pattern of individuals aged 16-64 years in England as a whole (not the UK) and to examine whether there were regional differences in dietary patterns in England. Dietary patterns (at the national level and disaggregated by region) were derived from the PCA method and two-way cross-classified multi-level modelling was used to examine the effect of region on dietary patterns identified within the national sample. The total diet of individuals, expressed as the average grams consumed daily, was captured in 60 NDNS (2008-16) pre-defined food groups, which were used in the PCA for the national sample and for each of the nine regions in England. The percentage variance explained (following varimax rotation) for the first four components of the national and region-specific samples, ranged from 11.1% to 14.4%. The first component of the national sample explained approximately 3.5% of the variance in the data, whereas the first component for the region-specific samples, ranged from 3.2% in the North West, West Midlands and South East to 4.3% in Yorkshire and the Humber. Overall, the first component explained most of the variance in the data and provided the most interpretable and meaningful pattern of foods consumed in the national and region-specific samples. This was not surprising as previous studies have found the first component to the best and most appropriate indicator of the variability in the data (DelaCruz, 2010; Robinson et al, 2004). These findings, when coupled with the marginal difference observed between the four components, helped to further justify the project's sole use and analysis of the first component for the national sample and the nine region-specific PCAs.

The first component dietary pattern identified for the national sample had moderate/strong positive loadings for fruits, vegetables, water, tea, coffee, skimmed milk, yogurt, fromage frais and dairy desserts, high fibre breakfast cereals and other margarine, fats and oils. At the regional level, first component dietary patterns for the North East, Yorkshire and the Humber, East Midlands, London, South East and South West regions closely resembled the national sample with moderate/strong positive loadings for fruit; salad and other raw vegetables; tea, coffee and water; yogurt, fromage frais and dairy desserts; skimmed milk; wholemeal bread; high fibre breakfast cereals; oily fish; brown, granary and wheatgerm bread and moderate/strong negative loadings for chips, fried and roast potatoes and potato products and meat pies and pastries. Thus, the leading dimension to diet was health conscious in most regions, given that the majority of food groups included (in component one) were generally recommended and considered beneficial to overall health (PHE, 2018b).

PCA-derived dietary patterns will inevitably vary across studies given that PCA is an approach which is data-specific and based on the study population used (Ocké, 2013). Nevertheless, the health conscious dietary patterns identified in this project were consistent with previous research. Several studies have previously noted and defined a health conscious, "Prudent" or "Rational" dietary pattern as one with high component loadings on fruits, vegetables, whole grains, legumes and low-fat dairy products, and in some instances, fish and seafood (Judd et al, 2015; Miller et al, 2010; Ocké, 2013). Similar to this project, the first component of the Ericson et al (2019) study was labelled a "Health conscious" dietary pattern, as it was characterised by high positive loadings for fruits and berries, nuts and seeds, legumes, other vegetables (non-legumes), plain yogurt, fresh cheese, tea, animal replacement foods, breakfast cereals, cooked grains such as bulgur, oil-based dressings, fish and fibre-rich bread, and high negative loadings for SSBs, red and processed meat, white bread and fried/deep-fried potatoes. Robinson et al (2004) explored the dietary pattern of a sample of young women in Southampton England and labelled the first component "Prudent", as it was characterised by high positive loadings for fruits and vegetables, wholemeal bread, rice and pasta, yogurt and breakfast cereals and high negative loadings for chips, processed meats and sugar. Similarly, Maksimov, Karamnova, Shalnova and Dapkina (2020) explored socio-demographic and regional determinants of PCA-derived dietary patterns in Russia and identified a fairly similar "Rational" dietary pattern, characterised by high loadings for fruits, vegetables, cereals, pasta, sweets and pastries and dairy products. Roberts, Cade, Dawson and Holdsworth (2018) used data from the NDNS to explore PCA-derived dietary patterns in the UK. The study used the same 60 food groups used in this project and identified what appeared to be a health conscious “Fruit, vegetables, oily fish” dietary pattern which was characterised by moderate/strong positive loadings (≥ +0.25) for fruit, salad and other raw vegetables, yogurt, fromage frais, and dairy desserts, oily fish, skimmed milk, nuts and seeds and moderate/strong negative loadings (≤−0.25) for chips (fried), roast potatoes, and potato products, white bread, sugar, preserves, and sweet spreads and white fish, coated or fried. Despite slight and inevitable differences across these studies, such as numerical differences in component loadings, the different combination of food groups included in dietary patterns and the use of different labels (e.g. prudent and rational instead of health conscious), for the most part, the findings from these empirical studies corroborated this project’s findings and provided further justification for the final decision to label component one (for the national sample and the North East, Yorkshire and the Humber, East Midlands, London, South East and South West) as “Health conscious.”

The East of England, North West and West Midlands were the three regions with a first component dietary pattern which deviated from the health conscious pattern observed for the national sample and the six remaining regions discussed. The East of England had a first component dietary pattern which was a mixture of margarine, fats, oils, dairy, dairy alternatives, fruit and pudding. Hence, it was labelled “Sugar, Fats and Oils”. On the other hand, the North West had a first component dietary pattern which was characterised by moderate/strong negative loadings for salad and other raw vegetables and moderate/strong positive loadings for burgers and kebabs, other milk and cream (inclusive of sour cream, purchased hot chocolate, condensed milk and all alternative milk products) and soft drinks (not low calorie). In contrast to the national sample, burgers and kebabs, which included any type of purchased/retail or takeaway burger or kebab products, contributed the most to the dietary pattern identified for the North West, followed by chips, fried and roast potatoes and potato products (which included any type of purchased retail/takeaway chips, French fries or potatoes roasted in fat). The North West seemed to have a more Western style dietary pattern for component one, as it consisted of mostly ultra-processed takeaway food, high fat dairy products, fried/coated foods and soft/fizzy drinks. The Western dietary pattern has been previously described as a diet low in fruits and vegetables and high in refined carbohydrates, ultra-processed foods, saturated fats, salt, free sugars, alcohol, red and processed meat, potatoes, refined grains and high fat dairy (Bere and Brug, 2010; Judd et al, 2015; Miller et al, 2010; Ocké, 2013; Popkin, Adair and Ng, 2012).

The West Midlands had a first component dietary pattern which was comprised of mostly spreads (low fat poly-unsaturated and sweet spreads), fried/coated fish, chips, fried and roast potatoes and potato products and sausages. Although low fat poly-unsaturated spreads (spreads which contained 40% or less fat and those with cholesterol lowering properties) contributed the most to the dietary pattern observed in the West Midlands, the remaining food groups were fried/coated, ultra-processed and presumably high in free sugar. Interestingly, the West Midlands and the North West were the only two regions in which fruits and vegetables (that is the "fruit" and "salad and other raw vegetables" food groups) were negatively associated and chips, fried and roast potatoes and potato products were positively associated with the fist component dietary pattern identified for these regions. The West Midlands also had moderate/strong positive loadings for both "white fish coated or fried" (which included purchased/retail/takeaway or homemade white fish (e.g. cod or haddock) coated, battered and/or fried) and "chips, fried and roast potatoes and potato products". The inclusion of both of these food groups was unique to the West Midlands and was reminiscent of the famous English fish and chips. Moreover, the West Midlands was the only region in which sausage was positive and moderately associated with the region's first component. Sausage included beef, pork, chicken or turkey sausages as well as sausage-based English dishes such as Toad in the Hole and Sausage and Mash. Thus, the first component dietary pattern observed for the West Midlands was labelled "Plain/Traditional English" because it had varying elements of the traditional English diet.

Janssen, Davies, Richardson and Stevenson (2018) conducted a pilot study which investigated PCA-derived dietary patterns for a small sample (26 persons) of individuals within the North West of England. Despite the very low sample size, that study identified a "Convenience" dietary pattern (comprised of refined grains, cakes, pastries and fast/takeaway food such as kebabs), which closely resembled the "Western/Takeaway" first component pattern observed in this project for the North West region. Although Morris et al (2016) used K-means cluster analysis instead of PCA, their results identified an unhealthy dietary pattern for the North West. The North West reportedly had the highest proportion of individuals with a "Monotonous Low Quantity Omnivore" dietary pattern, characteristic of a Western style diet with moderate/high consumption of white bread, milk, sugar, meat and potatoes. Conversely, regions towards the South, especially Greater London reportedly had the highest proportion of individuals with a "Health conscious" or vegetarian style dietary pattern (Morris et al, 2016).

When compared to the national sample, the unique (first component) dietary patterns observed in this project for the North West and West Midlands and the findings from previously discussed studies, initially reflected the North-South divide in diet/health in England (Bambra, 2016). However, after using two-way cross-classified multi-level modelling to examine the effect of region on dietary pattern scores for component one (the health-conscious dietary pattern) of the national sample, region was not statistically significant once compositional factors were considered. Age, sex, ethnicity and SEP, specifically highest level of education and equivalised household income, were the variables most associated with the first component dietary pattern for the national sample. This suggested that older persons, women and individuals within higher SEP groups were more inclined to have elements of a health conscious dietary pattern than younger persons, men and those from lower SEP groups. These findings were consistent with Roberts, Cade, Dawson and Holdsworth (2018) which found age, sex, ethnicity and SEP (specially NS-SEC occupational grade) to be positively associated with the health conscious, “Fruit, Vegetables, Oily Fish” dietary pattern identified in that study. Similarly, Robinson et al (2004) found age and especially highest level of education to be the variables most associated with the “Prudent” (health conscious) dietary pattern identified in that study. Although the association between ethnicity and diet (and related health outcomes such as obesity) remains inconclusive in the literature (Higgins, Nazroo and Brown, 2019; National Obesity Observatory, 2011), the statistically significant association between diet and age, sex, education and equivalised household income found in Research Question 2 has been previously identified and widely accepted, which further strengthened the findings of this project (Acheson, 1998; Imamura et al, 2015; NHS Digital, 2017; Roos et al, 2000; Roberts, Cavill, Hancock and Rutter, 2013; Tiffin and Salois, 2012).

Differences in dietary patterns at the sub-regional/small area level were not explored in Research Question 2. However, Morris et al (2016) observed statistically significant differences in dietary patterns at the small area level, with individuals living in more affluent areas said to have healthier, more diverse dietary patterns than persons in less affluent areas. Smith et al (2021) found statistically significant differences in estimated adult SSB and fruit and vegetable consumption across small areas in England, using NDNS and 2011 census data. Similarly, Newton et al (2015) although noting general differences between the North and South of England, found that inequalities in England were greater within regions than between them. The findings of these studies could help to explain the statistically insignificant association between region and dietary patterns found in this project. It is possible that geographic differences in dietary patterns in England may be more evident at a lower level of geography than the coarse regional level analysed in Research Question 2.

Previous studies have noted the relationship between the local food retail environment and dietary patterns (Saunders, Saunders and Middleton, 2015). Takeaway or “fast food” outlets typically offer energy dense food which is readily available and includes (but is not limited to) burger bars, kebab and chicken shops, chip shops and pizza outlets (PHE, 2018c). More than a quarter of all eateries in England are takeaway outlets (PHE, 2018d), with consumer spending on takeaways estimated to be £12.5 billion in 2018 and expected to increase to £15 billion by 2023

(British Takeaway Campaign, 2019). The increased demand for takeaway food and the drastic increase in the number of takeaway outlets (especially in the North of England) in recent years has been associated with the increased access to and the excess consumption of takeaway foods, especially in low-income and ethnically diverse areas (Maguire, Burgoine and Monsivais, 2015; PHE, 2018c). Areas with higher levels of deprivation and those which are more ethnically diverse tend to have greater access to takeaway (fast food) outlets than more affluent white areas (Black, Moon and Baird, 2014). Although it may be possible to purchase healthy takeaway options, frequent consumption of takeaway/fast food has been associated with increased excess weight over time, given that these foods are usually high in saturated fat (Maguire, Burgoine and Monsivais, 2015).

Statistics from PHE (2018c) have alluded to a North-South divide, with the number of takeaway outlets in the North West estimated to be 118 per 100,000 persons, compared to 92 and 75 per 100,000 in London and the South East respectively. At the sub-regional level, urban centres in the North West such as Blackpool, Manchester and Liverpool, which have the highest proportion of neighbourhoods among the most deprived in England, have been found to be potential hotspots for takeaway outlets (PHE, 2018c). Other studies (Blow, Gregg, Davies and Patel, 2019; Saunders, Saunders and Middleton, 2015) conducted in deprived areas in Manchester (North West England) and Sandwell (West Midlands) have also highlighted the high number of takeaway outlets in these areas, most of which offer a mixture of cheap, largely portioned, energy dense foods such as fried chicken, burgers, pizzas and kebabs. The lower operating cost to businesses (e.g. lower price to purchase land or access commercial spaces in low-income areas) has been associated with the high concentration of takeaway outlets on commercial streets within low-income areas (Blow, Gregg, Davies and Patel, 2019).

There was no clear rationale for the “Sugar, Fats and Oils” first component dietary pattern observed for the East of England and so this could possibly be flagged as an area for future studies. The increased number of takeaway outlets in deprived areas of the North West and the neighbouring West Midlands region (the second most ethnically diverse region outside of London), could help to explain the high component loadings for takeaway foods such as burgers, kebabs, chips and fried/coated fish and chicken observed for these regions in Research Question 2 of this project. Moreover, with the recent COVID-19 pandemic, the increased usage of convenient online takeaway delivery applications such as Just Eats and Deliveroo, especially amongst younger and less educated persons (groups identified in this project as less inclined to have elements of a health conscious dietary pattern), could also have future implications for the dietary patterns and overall health of individuals in England (Keeble et al, 2020). Although the association between the food environment and dietary patterns has been well-established in the USA, the situation within

England/the UK is more complex (Cummins and Macintyre, 2002; Cummins, 2014). Thus, persons in poorer or ethnic minority areas in England may not always be disproportionately worse off than persons in more affluent white areas (Black, Moon and Baird, 2014). Nevertheless, the results of Research Question 2 described some of the dietary patterns in England at the national and regional level, identified food groups (e.g. foods such as kebabs, burgers and chips associated with the North West) which could be flagged for future research at the sub-regional/small area level and highlighted the need for additional studies which investigate the influence of the overall food environment on dietary patterns in England.

5.4.1 Strengths and limitations

PCA, a well-established posteriori method, was used in this project to reduce 60 NDNS pre-defined food groups into smaller and more manageable components, which explained the variance and underlying dietary patterns in the data. PCA was advantageous given that the project's primary aim was to explore the types of foods actually consumed by individuals and possible differences across the regions of England, with no previous hypothesis or expectation. Although adherence to current dietary guidelines (as measured by the hypothesis/a priori method) is necessary and could have been assessed, this was not within the project's scope. PCA was beneficial as the use of this method made it possible to identify healthy as well as unhealthy food groups within dietary patterns identified. However, a major limitation of PCA is the method is data and population specific. PCA derived dietary patterns will inevitably differ across studies due to the method's dependence on the study population, data used and the many subjective decisions which have to be made by researchers such as the number of factors to extract, cut-off values and the labelling of components (Hearty and Gibney, 2008). Regardless, the dietary patterns identified in Research Question 2 of this project, specifically the "Health conscious" and "Western" style patterns, have been found in approximately 30% and 19% (respectively) of at least 84 previously published studies (Moeller et al, 2007). The reproducibility of these findings across studies meant that fewer subjective decisions had to be made and the analysis and interpretation of project findings were backed by current and well-established literature. Although there are other posteriori methods such as cluster analysis which could have been used in this project, previous studies have found dietary patterns identified with PCA to be similar or comparable to those identified with cluster analysis (Hearty and Gibney, 2008). The use of PCA in Research Question 2 was further justified as previous studies have found this method to be more practical and informative in characterising the dietary pattern of individuals, compared to other posteriori methods such as cluster analysis (Crozier et al, 2006).

The current project only analysed the first component for the national sample and the nine region-specific PCAs. Although the other components could have been considered, the first component was prioritised because it explained most of the variance in the data. The variance for component one for the national sample was only 3.5% and could be considered low and a possible limitation. Interestingly, Roberts, Cade, Dawson and Holdsworth (2018) used the same 60 pre-defined NDNS food groups (captured in NDNS waves 1-4/the 2008-12 period) and reported a variance of 3.9% for the first component, which was only 0.4 percentage points higher than that observed in this project. Although a slightly higher variance of 8% was observed in the Robinson et al (2004) study, this was based on 49 diet input variables. Therefore, the seemingly low percentage variance observed in this project, could be due to the number of input variables used (Field, 2018; Roberts, Cade, Dawson and Holdsworth, 2018). Another limitation was that some of the 60 NDNS pre-defined main food groups were broad and merged what appeared to be healthy and less healthy foods with different nutritional compositions. Although not within the scope of this project, it is possible that further disaggregation of food groups could have yielded slightly different results. Geographic data captured in the NDNS were constrained to the regional level or the GOR in which survey participants resided. As a result, the relationship between dietary patterns and the overall food environment (including activity spaces/areas where individuals typically visit in their daily lives such as work and school) could not be explored in Research Question 2, which is another possible limitation.

Despite these limitations, the project's use of the PCA method to produce empirically-derived dietary patterns for the national NDNS sample as well as for each of the nine regions in England was considered a major strength, as very few studies have done separate PCAs, especially based on geography/region (Moeller et al, 2007). Judd et al (2015) and Maksimov, Karamnova, Shalnova and Dapkina (2020) previously examined PCA-derived dietary patterns in the national sample as well as by region. However, these studies were conducted in the USA and Russia, respectively. As far as known, this is the first project to explore PCA-derived dietary patterns in England as a whole as well as by region, which is a major project strength. Previous studies conducted in England tended to focus on health outcomes such as obesity, the consumption of fruits and vegetables as opposed to overall diet or differences across the UK and not across the regions of England specifically.

This project found that dietary patterns especially the health conscious pattern, may have more to do with who persons are (compositional factors), than where they reside (geographic/contextual factors). This is not to say that the unique differences observed in this and previous studies for specific regions (e.g. the North West and West Midlands) are inconsequential. Instead, the findings from Research Question 2 helped to identify regions which may need to be flagged for future studies. The poor availability of diet consumption data remains a major challenge which prevents researchers from fully exploring this topic (Brug, 2008; National Academies of Sciences, Engineering & Medicine, 2019). Nevertheless, further exploration of dietary patterns and possible differences at the sub-regional or small area level is warranted and will be investigated in this project's third research question. Such information could help to inform social policy as well as more localised interventions aimed at improving the diets of individuals, especially those in lower SEP groups.

Chapter 6 To what extent does estimated achievement of 5-A-Day vary across smaller areas of England?

Sections of this chapter have been published as Smith et al (2021).
Smith, D. M., Vogel, C., Campbell, M., Alwan, N. and Moon, G. (2021) 'Adult diet in England: Where is more support needed to achieve dietary recommendations?', *PLOS ONE,* 16(6), pp. e0252877. doi:10.1371/journal.pone.0252877

6.1 Introduction/Context

It has been widely acknowledged that fruit and vegetable consumption/diet is not uniformly distributed across population sub-groups and that there are differences in diet according to age, sex, ethnicity and SEP (Imamura et al, 2015; Roberts, Cavill, Hancock and Rutter, 2013; Yau, Adams and Monsivais, 2019). Inequalities in diet and health have also been observed across small and larger geographic areas of the UK (Barton, Chambers, Anderson and Wrieden, 2018; Shelton, 2009). Since the 19^{th} century geographic differences in diet have been found between the regions of England, with persons in the North of England said to have comparatively worse diets than those in the South (Bambra, 2016). Although the English North-South divide in diet/health has been previously noted, less is known about how diets vary within countries such as England (Imamura et al 2015), especially at the local level. Macintyre, Maciver and Sooman (1993) noted the tendency for researchers to rely on crude stereotypes (e.g. North-South divide in England) to make general assumptions about the diet and general health of persons living in particular areas, without conducting the research necessary to substantiate these claims. It is recommended that researchers systematically compare the diet/health of persons in different areas instead of relying on past assumptions (Macintyre, Maciver and Sooman, 1993 p. 230). These types of investigations could reveal findings similar or contrary to what was initially assumed or findings which could go undetected if not explored. However, a major issue/gap identified is that very few studies have explored possible differences in fruit and vegetable consumption/diet at the sub-regional level in England (Smith et al, 2021).

The introduction of the Localism Act (2011) gave local authorities more freedom and power to liaise with constituents to address the issues identified within local areas (Gadsby et al, 2017). Subsequently, the passing of the Health and Social Care Act (2012) saw local authorities gaining more responsibility for improving the overall health of local populations. This included increased oversight for the monitoring and evaluation of national diet-related targets such as the 5-A-Day target. With this added responsibility, local level diet-related data became even more crucial to monitor and assess the diet of individuals in local areas, identify potential inequalities and evaluate the success of current diet/health-related policies and interventions. However, a major challenge is that diet-related data are primarily obtained from national Government funded surveys such as the HSE, which are not powered to capture more local variation (Pryce et al, 2020). Thus, inequalities in diet and health could go undetected due to the poor availability of data at the sub-regional level (Twigg, Moon and Jones, 2000).

Small sample sizes, the different area boundaries used by data collectors and the different ways in which data are captured or formatted are noted challenges which hinder researchers from fully exploring the diet of individuals at the sub-regional level (Brug, 2008; Campbell et al, 2020; Macintyre, Maciver and Sooman, 1993; Moretti and Whitworth, 2021; Whitworth, 2013). The Active Lives Survey (ALS) is the only survey which captures consistent reliable fruit and vegetable consumption data at the local authority level in England (Campbell et al, 2020). However, a major limitation is the survey's overestimation of fruit and vegetable consumption/5-A-Day compared to the NDNS and HSE (PHE, 2020a). Although some local level surveys have been conducted by Public Health England (PHE) over the years, these surveys are limited to a few select local authorities, typically small in terms of sample size, and to date, have been primarily focused on smoking, binge drinking and diet-related outcomes such as obesity (Pryce et al, 2020). The issue of small sample sizes at the sub-regional level, could likely be due to the tremendously high physical and financial resources associated with survey data collection (Moretti and Whitworth, 2021; MRC, 2014). The focus on smoking and binge drinking (instead of diet) could be because these factors are generally easier to identify and have more short-term direct impacts on health (Bambra et al, 2010).

In the absence of representative local level survey data, PHE and public health researchers have come to rely on proxy measures and small area estimation (SAE) methods to estimate health outcomes at the local level (Moon et al, 2007; Pryce et al, 2020; Twigg, Moon and Jones, 2000). Some researchers have collaborated with supermarket chains (such as Sainsburys) to use loyalty card food purchasing data to estimate the diet of individuals at the regional and sub-regional level (Jenneson, Morris, Greenwood and Clarke, 2021). Although a novel approach, loyalty card data captures household food purchasing and expenditure and does not capture foods actually consumed by individuals (Appelhans et al, 2017). In light of General Data Protection Regulation (GDPR) and other disclosure guidelines, there may also be challenges in accessing loyalty card data at the small area level, as this type of information is sensitive and could possibly be used to reveal the identity of survey participants (Stourm et al, 2020). These and the data-related challenges previously mentioned, restrict the assessment of diet/5-A-Day achievement to the national and/or regional level and potentially mask possible geographic variations which are likely to be more pronounced at the local level (Cummins and Macintyre, 2002; Cummins, 2014).

SAE (regardless of the method used) is an important means of overcoming the many data-related challenges mentioned, as it makes use of existing survey data to produce robust and precise estimates of health-related outcomes, at small geographic scales, which would not have been otherwise possible (Rahman, 2017; Whitworth, Carter, Ballas and Moon, 2017). However, a lacuna still exists as very few studies have used SAE to explore possible variations in diet (not obesity) across small areas of England. With the exception of Smith et al (2021), this project is the only known study to use SAE to estimate 5-A-Day achievement across small areas in England. Research Question 3 therefore aims to use SAE to estimate and visualise 5-A-Day achievement across small local areas in England. The findings of this project are valuable and could help to fill a current gap in literature, given the dearth of diet-related SAE studies and the general unavailability of small area survey data. The following section (Section 6.2) will highlight the source of data and the method of analysis used within this chapter.

6.2 Method

There are several methods which can been used to conduct SAE for health-related outcomes. SAE methods typically fall within two broad categories: direct and indirect model-based estimation methods (Rahman, 2017). The direct estimation method uses only survey data collected at the small area level to explore an outcome of interest. Indirect estimation methods utilise survey data and data from additional sources such as the census to estimate outcomes of interest (such as 5-A-Day achievement) for persons living within small areas. Direct estimation is not a suitable method where survey data at the small area level are unavailable or where sample sizes are low/not representative of small areas (Whitworth, 2013). As customary with most health-related studies, the poor availability of (diet-related) survey data at the small area level (Whitworth, Carter, Ballas and Moon, 2017), made indirect estimation the more appropriate method for Research Question 3.

Indirect SAE can be further divided into statistical and geographic modelling methods, with the latter based on spatial microsimulation models. A major benefit of indirect SAE (regardless of the method used), is the use of detailed survey data to produce robust and precise estimates of health-related outcomes, at small geographic scales, which would not have been otherwise possible (Rahman, 2017; Whitworth, Carter, Ballas and Moon, 2017). The statistical approach is based on regression modelling and explores the relationship between a target outcome of interest (for which small area estimates will be made) and a set of explanatory variables such as age, sex and SEP. There are three main types of models used in the statistical approach: i) individual level, ii) area level and iii) individual and area level (multi-level) modelling. Multi-level SAE statistical modelling is preferred to individual and areal level modelling, as it acknowledges that the outcome of interest is dependent on compositional factors (individual factors related to who persons are) and contextual factors (where persons reside) (Twigg, Moon and Jones, 2000). The spatial microsimulation (geographic) modelling method works by combining individual level survey microdata with spatially disaggregate small area census data, to create synthetic microdata estimates for small areas (Lovelace, Birkin, Ballas and van Leeuwen, 2015).

A major strength of the multi-level modelling SAE approach is its acknowledgement of both compositional and contextual factors and their effect on the outcome of interest. This method has been successfully used to estimate smoking and high alcohol consumption and obesity across small areas in England (Moon et al, 2007; Twigg, Moon and Jones, 2000). The multi-level SAE method requires that all individual level explanatory variables exist as a single census cross-tabulation (Whitworth, 2013). However, there are strict limits placed on the number of covariates that can be included in any census cross-tabulation. Besides variables such as age and sex, detailed cross-tabulated data at the small area level (e.g. the MSOA level) were not available in the census for other socio-demographic variables of interest such as housing tenure. These observations were consistent with those previously noted by Twigg, Moon and Jones (2000). Based on these challenges, the use of the multi-level SAE method in Research Question 3 would have reduced the number of individual level explanatory variables that could have been included in the model. Other relevant factors could only have been included as higher-level aggregate variables.

Unlike multi-level and other statistical SAE approaches, spatial microsimulation approaches such as IPF are more flexible and do not require that explanatory variables be represented as a single census cross-tabulation. Instead, a list-based approach is used in spatial microsimulation, where an individual’s attributes are stored in models as lists rather than matrices (Rahman, Harding, Tanton and Liu, 2010). Rahman, Harding, Tanton and Liu (2010) also noted the benefit and flexibility of spatial microsimulation methods, especially in instances where census and/or survey data are limited or unavailable. Moreover, Rahman (2017, p.3) (although noting that each SAE approach has its own strengths and weaknesses), found spatial microsimulation-based models to be robust and a “precise means of estimating health-related population characteristics over small areas.” Due to the flexibility and the limited data constraints, spatial microsimulation was deemed the more appropriate approach for Research Question 3.

The IPF microsimulation method is referred to as a deterministic approach, because it does not involve the use of random numbers and give the same results each time the model is run (Tomintz, Kosar and Clarke, 2016). The IPF method operates by replicating the population in small areas from survey respondents, based on matched demographic characteristics which predict the health outcome of interest. IPF was selected for Research Question 3 because it is a relatively simple method, which is widely used in microsimulation models for healthcare research (Harland, Heppenstall, Smith and Birkin, 2012), transparent and computationally efficient (Lomax and Norman, 2016) and capable of producing reliable estimates which are comparable with other techniques (Whitworth, 2013; Whitworth, Carter, Ballas and Moon, 2017). IPF has previously been used to estimate smoking in New Zealand (Smith, Pearce and Harland, 2011); smoking in Austria (Tomintz, Kosar and Clarke, 2016); as well as obesity (Edwards and Clarke, 2009) and mental health and alcohol consumption in England (Riva and Smith, 2012) Although successfully used in these previous studies, with the exception of Smith et al (2021), there are very few studies which have used IPF to estimate diet (rather than diet-related outcomes such as obesity) at the small area level. This finding further justified the use of IPF (in Research Question 3), as this project would be the first known study to use IPF to estimate 5-A-Day achievement across small areas in England.

Figure 6.1 provides an overview of the IPF spatial microsimulation methodology and the steps (discussed in subsequent sections) used within the current project.

Figure 6.1: Steps used to conduct the IPF spatial microsimulation method

6.2.1 Step 1: Selection of data

At least two sources of data are needed to conduct IPF. The first data source, known as the population dataset, is a nationally representative survey with information on the outcome of interest, socio-demographic details, but limited (if any) geographic information, especially at the small area level. The second data source, referred to as the geographic dataset, contains information on the socio-demographic characteristics of a population aggregated at the small area level, with generally limited information on health or related behaviours.

The population dataset used for Research Question 3 was the 2008-11, 2013 and 2015-16 waves of the Health Survey for England (HSE) (NatCen Social Research and University College London Department of Epidemiology and Public Health, 2008-2016). The HSE was selected as it is currently the primary survey used by the PHE to monitor achievement of the national “5-A-Day” target, which is the focus and binary outcome variable of interest for Research Question 3. The decision to use the HSE was also based on the survey’s large annual sample size, which when pooled across the seven survey waves, resulted in a total sample size of 111,158 persons. Missing cases were assessed, and complete-case analysis was deemed an appropriate method for handling missing data in the HSE. The removal of all missing cases is a possible limitation, as bias could be introduced (Allison, 2001). However, the total percentage of missing cases fell below five percent (statistical rule of thumb), which meant that the potential impact of the missing data was negligible and ignorable (Jakobsen, Gluud, Wetterslev and Winkel, 2017). The project’s focus on the working age 16-64 age group meant that persons under the age of 16 years and those 65 years and older were excluded. The removal of missing cases and the age groups not relevant to the project, reduced the HSE sample size to 34,195 persons. This sample was more than ten times the NDNS’ sample size (3,286 persons) for the corresponding 2008-16 period. Background information on the HSE, its survey design, method of data collection, sample size, sample characteristics and additional details regarding data accessibility and cleaning procedures conducted, inclusive of the handling of missing data were fully described in Chapter 3 (Section 3.3.2) and Chapter 4 (Section 4.2). Additionally, details regarding the binary outcome variable (“Achieve5”, fruit and vegetable consumption) and other explanatory variables captured in the HSE were previously described in Chapter 4 (Section 4.2.1.4 - 4.2.1.8).

The 2011 census was the geographic data set selected. As customary with most geographic datasets, data pertaining to diet were not captured within the 2011 census. Rather, it provided the socio-demographic characteristics of the English population (34,329,091 persons aged 16-64 years) at the middle layer super output area (MSOA). MSOAs are statistical units built from groups of contiguous lower layer super output areas (LSOAs) and designed to facilitate the reporting of small area statistics in England. There are currently 6,791 MSOAs in England, each with a minimum population of 5,000 and a mean of 7,200 (Sun, Hu, Huang and On Chan, 2020). Small area estimates have been found to have better precision at the MSOA level compared to other levels of geography, such as the local authority district (LAD) level (ONS, 2017b). This precision relates to MSOAs being more comparable in size to the primary sampling unit (PSU) used in surveys such as the HSE (ONS, 2017b). Previous studies have also noted the instability of models where individuals were nested within the smaller LSOAs (Moon et al, 2019). Thus, small area estimates of fruit and vegetable consumption conducted for Research Question 3 were based at the MSOA level, which was in keeping with national statistical guidelines, and an approach adopted by similar studies (Moon et al, 2019; Taylor, Moon and Twigg, 2016).

6.2.2 Step 2: Data preparation

The 2011 census and HSE data were assessed simultaneously to identify the set of explanatory variables (commonly referred to as constraint variables in the IPF method) captured in both datasets, which best predicted the outcome of achievement of the 5-A-Day target. Constraint variables were the set of variables used to predict the outcome of 5-A-Day achievement (Whitworth, 2013). Age, sex, ethnicity, marital status, highest level of education, NS-SEC (occupational grade) and housing tenure were initially identified as possible constraint variables, as they were captured in both the HSE and the 2011 census and previously found to be significantly associated with 5-A-Day achievement (results of the Chi-square correlation analysis discussed in Chapter 4, Section 4.3). Previous studies have recognised equivalised household income, education and NS-SEC occupational grade to be measures of SEP most associated with fruit and vegetable consumption and overall diet (Acheson, 1998; Berzofsky et al, 2014; Darmon and Drewnowski, 2008; Roos et al, 2000; Tiffin and Salois, 2012). Equivalised household income was not captured in the 2011 census and the (continuous) equivalised household income variable was removed from the HSE in 2015 (onwards) following a disclosure review by NHS Digital. As such, HSE data related to this variable could not be included or analysed in Research Question 3, based on the inconsistency across the 2008-16 period. Housing tenure, also considered a possible indicator of SEP, was not analysed/included in Research Question 1 and 2 of this project. However, since housing tenure was available in both the HSE and 2011 census, it was still considered good practice (specifically for the IPF method used in Research Question 3) to test it as a possible constraint variable during the preliminary assessment.

Data tables for the 2011 census were downloaded from the “Local characteristics” section of the ONS’ NOMIS data repository website (ONS, 2011). Data were filtered to capture members of the English population aged 16-64 years at the MSOA level. The NOMIS census table labelled LC 6206 EW was a cross-tabulation of age by NS-SEC by ethnicity. Table LC 1108 EW captured living arrangements by sex by age; LC 5102 EW captured highest level of qualification gained by age; and LC 3409 EW captured general health by tenure by age. Multi-level modelling is not required for the selection of constraint variables used in IPF but was eventually used in this project for the construction of credible intervals, which will be discussed in Step 5 (Section 6.2.5). The use of multi-level modelling was also beneficial because the method (unlike single-level/fixed effects modelling typically used) acknowledged that HSE data were hierarchically structured, with individuals, nested within small areas (captured by the primary sampling unit (PSU) variable in the HSE), nested within GORs.

Checks for collinearity were conducted (in SPSS version 24 (IBM, 2018)) and as customary with the IPF methodology, univariate and multivariate binary logistic regression models were tested using the seven constraint variables initially identified: age, sex, ethnicity, marital status, highest level of education, NS-SEC and housing tenure. Three-level hierarchical binary logistic multi-level regression models were conducted in MLwIN version 3.02 (Charlton et al, 2017), with level one (first level) being the individual level, level two representative of the area/PSU level and level three, the region/GOR-level. Although small area/MSOA level data were not available in the HSE, previous studies have found PSUs (level two) to be similar in size and the most comparable to MSOAs in England (ONS, 2017b). Details regarding the PSU variable captured in the HSE were discussed in Chapter 3, Section 3.3.2.

Previous studies have strongly recommended that the final number of constraint variables used in IPF should not exceed four, as this number provides the most reliable estimates without over constraining the IPF model (Smith, Clarke and Harland, 2009). Based on this recommendation, variables were entered sequentially, resulting in 7 univariate and 20 multivariate model combinations, used to identify the four best predictors/constraint variables for 5-A-Day achievement. The MCMC estimation method was used with the default 5,000 iterations and burn-in and thinning equal to five and one, respectively. The Raftery-Lewis diagnostic (Nhat) was used to assess the number of iterations needed for data convergence. Following this assessment, the number of iterations was increased from 5,000 to 1,000,000 (data convergence). The Deviance Information Criterion (DIC) was used to assess the goodness-of-fit for the 27 univariate and multivariate models tested (results not included in project). A reduction in the DIC value indicated an improvement in the model. The model with highest level of education, age, ethnicity and NS-SEC was selected as the final, because it had the lowest DIC value of all models tested. In addition, as good statistical practice, the corresponding single-level model with the four constraint variables was also tested and the DIC found to be higher than the multi-level model (model results presented in Appendix D.1). These findings helped to confirm the project's final selection of the three-level hierarchical multi-level model and the selection of the variables education, age, ethnicity and NS-SEC as the four final constraint variables to be used in the IPF estimation process.

Prior to IPF model execution, variables in the HSE and 2011 census were recoded so that the categories used in both datasets corresponded, as far as possible. This categorisation was required to ensure accuracy during the IPF estimation process (Smith, Pearce and Harland, 2011). The final recoded education variable consisted of five categories: i) No qualifications; ii) NVQ1/CSE other grade equivalent/NVQ2/GCE O Level; iii) NVQ3/GCE A Level equivalent/Apprenticeship; iv) NVQ4/NVQ5 /Degree or equivalent and v) Still in full-time education/Other foreign qualification/Other. Age was recoded into a five-category variable: i) 16-24 years; ii) 25-29 years; iii) 30-39 years; iv) 40-49 years and v) 50-64 years. The NS-SEC occupational grade variable was collapsed from eight into five categories based on ONS guidelines previously outlined in Chapter 4, Section 4.2.1.6.3 (ONS, 2019): i) Never worked and long-term unemployed; ii) Routine, manual and other occupations; iii) Small employers and own account workers; iv) Intermediate occupations and v) Higher managerial, administrative and professional occupations. Previous studies have noted the importance of acknowledging possible differences between ethnic groups, especially when estimating health-related lifestyle behaviours such as fruit and vegetable consumption at the small area level (Smith, Clarke and Harland, 2009). As such, ethnicity was recoded into a five-category variable: i) White; ii) Mixed/multiple ethnic groups; iii) Asian/Asian British; iv) Black African/Caribbean/Black British and v) Other (inclusive of Arabs and other ethnic groups).

Before executing the IPF model, a noted challenge to be considered is the model's attempt to smooth the population distributions of each constraint variable towards a global mean. Previous geo-demographic studies (McLachlan and Norman, 2020; Smith, Clarke and Harland, 2009) have recommended conducting a K-means cluster analysis in order to group output areas (e.g. MSOAs) to create aggregate groups with shared characteristics in the population that predict the outcome of interest. Based on these recommendations, a K-means cluster analysis (the details of which can be found in the aforementioned studies) was conducted in SPSS version 24 (IBM, 2018). Clusters were based on the percentage of persons in the 2011 census population who were 50-64 years of age, had a degree level or equivalent qualification and were of the Asian/Asian British ethnic group. These three categories were selected because they were the most significant predictors of 5-A-Day achievement observed during the model constraint selection process (see final model results in Section 6.3).

The "elbow" method (the details of which have been described in Yuan and Yang (2019) was used to select the optimal number of (K) clusters to be retained within this project. This process involved firstly conducting a cluster analysis (in SPSS) for different values of K, ranging from 2 to 20 (that is, K= 2, 3, 4….20). Twenty (that is, K=20) was selected as the maximum number to be tested, because beyond this value was thought to be unwieldly. A scree plot (presented in

Appendix D.2) was then generated by plotting the within-group sum of squares distance (also known as the squared Euclidean distance) on the y-axis and the number of clusters (K=2, 3, 4…20) on the x-axis. The elbow or point of inflection in the curve indicates the optimal or most appropriate number of clusters to be retained (Yuan and Yang, 2019). The scree plot was visually assessed and five (K=5) was selected as the optimal number of clusters to be used within the current project. Thereafter, MSOAs were sorted based on their newly assigned cluster, with 2,315, 409, 1,831, 908 and 1,328 MSOAs assigned to cluster 1, 2, 3, 4 and 5, respectively.

6.2.3 Step 3: IPF model execution in R open source software package

MSOA census population estimates for each of the four constraint variables (education, age, ethnicity and NS-SEC) were adjusted to reflect the 2018 mid-year age-sex population estimates (the most up to date estimates available at the time data were analysed), produced by the ONS (2018b). This step was necessary because data from the 2011 census were outdated and not reflective of more recent mid-year population estimates. Constraint variables included in the IPF model were entered based on the order of statistical significance (or association from the regression) with the (binary Yes/No) 5-A-Day achievement outcome variable. Highest level of education (the variable most statistically significantly associated with 5-A-Day achievement) was entered first, followed by age, ethnicity and finally, NS-SEC occupational grade. The IPF model, executed in the R open source software package (version 3.4.4) (R Development Core Team, 2018), calculated the probability that each of the 34,195 respondents in the HSE lived in one of the 6,791 MSOAs across England. The model was run one cluster at a time to address the issue of smoothing previously mentioned in Section 6.2.2. Any smoothing would be addressed with this step. The IPF model was executed using 10 iterations which was sufficient for complete convergence and was in keeping with the number of iterations recommended in previous studies (Lovelace, Birkin, Ballas and van Leeuwen, 2015).

Probabilities were calculated by matching HSE survey respondents to known residents in the English population, based on the four socio-demographic constraint variables. The probabilities for each individual were summed to total the local population based on the ONS 2018 mid-year population data and resulted in the estimated prevalence of fruit and vegetable consumption in each MSOA. Figure 6.2 presents a scenario/worked example which was used to demonstrate the overall IPF re-weighting process and the subsequent re-scaling of the new probabilities to fit ONS mid-year population data. Details of the IPF re-weighting process (including a visual summary of the IPF process) are discussed elsewhere (Campbell and Ballas, 2013; Lovelace, Birkin, Ballas and van Leeuwen, 2015; Smith, Clarke and Harland, 2009).

HSE respondents

ID	Ethnicity
001	White
002	Mixed
003	Asian
004	Black

MSOA 001 (s_{ij} values)

White	Mixed	Asian	Black	Other
800	125	325	200	50

HSE totals (m_{ij} values)

White	Mixed	Asian	Black	Other
3,000	500	300	1,120	80

Survey respondents: new weight calculation

ID	Ethnicity	Weight (w_i)	w_i s_{ij} / m_{ij} = n_i
1	White	1	1 x 800/3,000 = 0.27
2	Mixed	1	1 x 125/500 = 0.25
3	Asian	1	1 x 325/300 = 1.08
4	Black	1	1 x 200/1,120 = 0.18

$n_{i=} w_i s_{ij}/m_{ij}$

$n_{i=}$ the new weight of individual i

w_i = the old weight of individual i

s_{ij} = the element of the MSOA ONS table for individual i and variable j

m_{ij}= the element of the NDNS data table for individual i and variable j

New weights are then scaled to fit known ONS populations:

$n_0^c = w^c \; X \; (t_o^c \; / \; t_s^c \;)$

$N_o^c = \sum n_o^c$

with scaling factor

$t_o^c \; / N_o^c$

In this example, where n_o is new weight for individual in $MSOA_o$, N_o the total weight for all individuals in $MSOA_o$, c the variable subcategory (e.g. White, Mixed, Asian, Black or Other for the ethnicity category), t^c_o the total population in variable c in MSOA o (ONS totals used here), t^c_s the total population in variable c in survey data (HSE used here) and w_c is new weight for an individual. The sum of all the new weights after the ethnicity variable is reweighted is calculated for $MSOA_o$ (N_o) and will sum to the total population of $MOSA_{o.}$ Then, the sum of the new weights for all White persons is calculated (N^c_o). The ratio of the number of White persons reweighted by the model (N^c_o) to the reweighted population for that area (N_o) is used as a scaling factor on all the new weights generated by this variable reweighting process.

Figure 6.2: Worked example of the IPF reweighting process

6.2.4 Step 4: Data validation

Following the execution of the IPF model, data should be validated to check the robustness of model estimates. External validation seeks to establish the similarity between simulated model estimates and the actual spatial distribution of an outcome(s) (5-A-Day achievement in the case of this project). Internal validation examines how well the simulated data represent the input or source data. Although both equally important, the poor availability of detailed survey data at the small area level (e.g. MSOA level), remains a challenge to the external validation of IPF models. The poor availability of diet-related survey data at the small area level meant that there was no external data source with which IPF model estimates could have been compared or validated (externally) (Tomintz, Kosar and Clarke, 2016). This lack of lower level geographic data justifies the overall need for IPF and could explain the tendency for studies (similar to this project) to focus primarily/solely on internal validation (James, Lomax and Birkin, 2019).

The total absolute error (TAE), standardised absolute error (StAE) and percent error, are typically used for the internal validation of IPF model estimates. The TAE, which measures the error between the simulated population and the census/mid-year population, was calculated for each constraint variable (across the 6,791 MSOAs in England) using the formula presented in Equation 6.1.

Equation 6.1: Formula for the calculation of the Total Absolute Error (TAE)

$$TAE = \sum_{ij} |U_{ij} - T_{ij}|$$

U_{ij} is the observed count for MSOA i in category j

T_{ij} is the expected count for MSOA i in category j

The StAE was found by dividing the TAE by the total census-defined population for each MSOA and the percent error by multiplying the StAE by 100%. The percent error (to be discussed in Section 6.3) was thereafter assessed to ensure that errors did not exceed the 20% error in 80% of areas recommended in previous studies (Lovelace, Birkin, Ballas and van Leeuwen, 2015).

6.2.5 Step 5: Calculation of credible intervals and data visualisation

Measures of uncertainty around point estimates have not been traditionally a part of IPF or other spatial microsimulation methods. This could explain why most studies which use IPF tend not go beyond Step 4 or the (internal) validation of data. Small area point estimates and 95% credible intervals were calculated within this project, as they usually infer the likely precision of small area estimates (Moretti and Whitworth, 2021). This additional step (Step 5) specifically required the use of multi-level logistic regression modelling (steps previously described in Step 2, Section 6.2.2) because unlike single-level modelling, it provided the separate area level residual error variance necessary for the calculation of credible intervals (Whitworth, Carter, Ballas and Moon, 2017). It should be noted that the use of the Bayesian terminology "credible" instead of the usual "confidence interval" reflects the simulation basis to interval identification.

The variance of the residual level two (PSU) error was saved, exported to SPSS version 24 (IBM, 2018) and converted to log odds. The residual level two error was replicated stochastically to represent a range of 1,000 possible values for each MSOA. This was done by taking a random sample of 1,000 values (for each MSOA) from the known distribution of the residual level two error, with the mean equal to zero and the standard deviation estimated by the multi-level model with the constraints used in the IPF, and normally distributed. Each of the randomly selected residual level two error terms were added separately to the microsimulated central point estimate for that MSOA. These estimates were converted from predicted log odds to predicted probabilities. The 2.5^{th} and 97.5^{th} percentiles of the resulting distribution of 1,000 values formed the lower and upper credible intervals, which were plotted in Microsoft Excel. Further details on the calculation of credible intervals have been described elsewhere (Whitworth, Carter, Ballas and Moon, 2017). Finally, IPF model estimates for the achievement of the 5-A-Day target across MSOAs in England, were mapped using ArcGIS version 10.7.

6.3 Results

6.3.1 Results of the final model with constraint variables

As mentioned in Section 6.2.2, a combination of univariate and multivariate binary logistic three-level hierarchical multi-level regression models were tested to identify the four best predictors/constraint variables for the IPF. Highest level of education, age, ethnicity and NS-SEC was the best fitting model, with a DIC value of 36,985. Table 6.1 presents the results of the final multi-level binary logistic regression model with the four constraint variables selected as the best predictors of 5-A-Day achievement in the IPF.

Table 6.1: Three-level hierarchical multi-level binary logistic regression model with the four constraint variables selected and used in the IPF small area estimation method, HSE, 2008-11, 2013 and 2015-16

Variables in the model	Log-odds coefficient (SE)	Exp (β)/Odds ratio	Exp (Confidence Interval)
Intercept**	-2.133** (0.122)	___	___
Age (Ref Cat: 16-24 years)			
25-29	___	1.08	(0.94, 1.23)
30-39	___	1.51**	(1.35, 1.69)
40-49	___	1.65**	(1.47, 1.84)
50-64	___	2.19**	(1.96, 2.44)
Ethnicity (Ref Cat: White)			
Mixed	___	1.15	(0.93, 1.42)
Asian	___	1.62**	(1.46, 1.79)
Black	___	1.00	(0.85, 1.18)
Other	___	1.75**	(1.34, 2.30)
Highest level of education (Ref Cat: No qualifications)			
NVQ1/CSE other grade equivalent/NVQ2/GCE O Level equivalent	___	1.15**	(1.05, 1.27)
NVQ3/GCE A Level equivalent/Higher education below degree level	___	1.65**	(1.50, 1.82)
NVQ4/NVQ5/Degree or equivalent	___	2.48**	(2.24, 2.74)
Still in full-time education/Other foreign qualification/Other	___	1.91**	(1.67, 2.20)

Variables in the model	Log-odds coefficient (SE)	Exp (β)/Odds ratio	Exp (Confidence Interval)
NS-SEC8 Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	___	0.89	(0.73, 1.08)
Small employers and own account workers	___	1.23	(1.00, 1.52)
Intermediate occupations	___	1.17	(0.96, 1.42)
Higher managerial and professional occupations	___	1.19	(0.97, 1.46)
Variance of random parameters			
Government Office Region (GOR)	0.022 (0.017)	___	___
Primary Sampling Unit (PSU)	0.210 (0.022)	___	___

* p-value <0.05, ** p-value<0.01, SE= standard error

NS-SEC occupational grade was not statistically significantly associated with 5-A-Day achievement but was retained as a constraint variable in the model with the best “fit” (Table 6.1). Persons belonging to the Asian and Other ethnic group had increased odds of achieving of the 5-A-Day target compared to White persons. However, highest level of education and age appeared to be the strongest predictors of 5-A-Day achievement, as the odds of achievement increased as the level of education and age increased (Table 6.1).

6.3.2 IPF model results

On average, 26% of persons aged 16-64 years, living across MSOAs in England, achieved the 5-A-Day target, with the percentage achievement ranging from 18% to 36%. This meant that approximately 74% of persons did not consume the requisite daily portion of fruits and vegetables. IPF model estimates for the achievement of the 5-A-Day target were mapped for the 6,791 MSOAs, across the nine regions of England and presented in Figure 6.3. Figures 6.4, 6.5 and 6.6 show the results (disaggregated) for regions in the North of England (North West, North East and Yorkshire and Humber); Midlands and East of England (West Midlands, East Midlands and East of England) and the South of England and London (South West, South East and London) respectively.

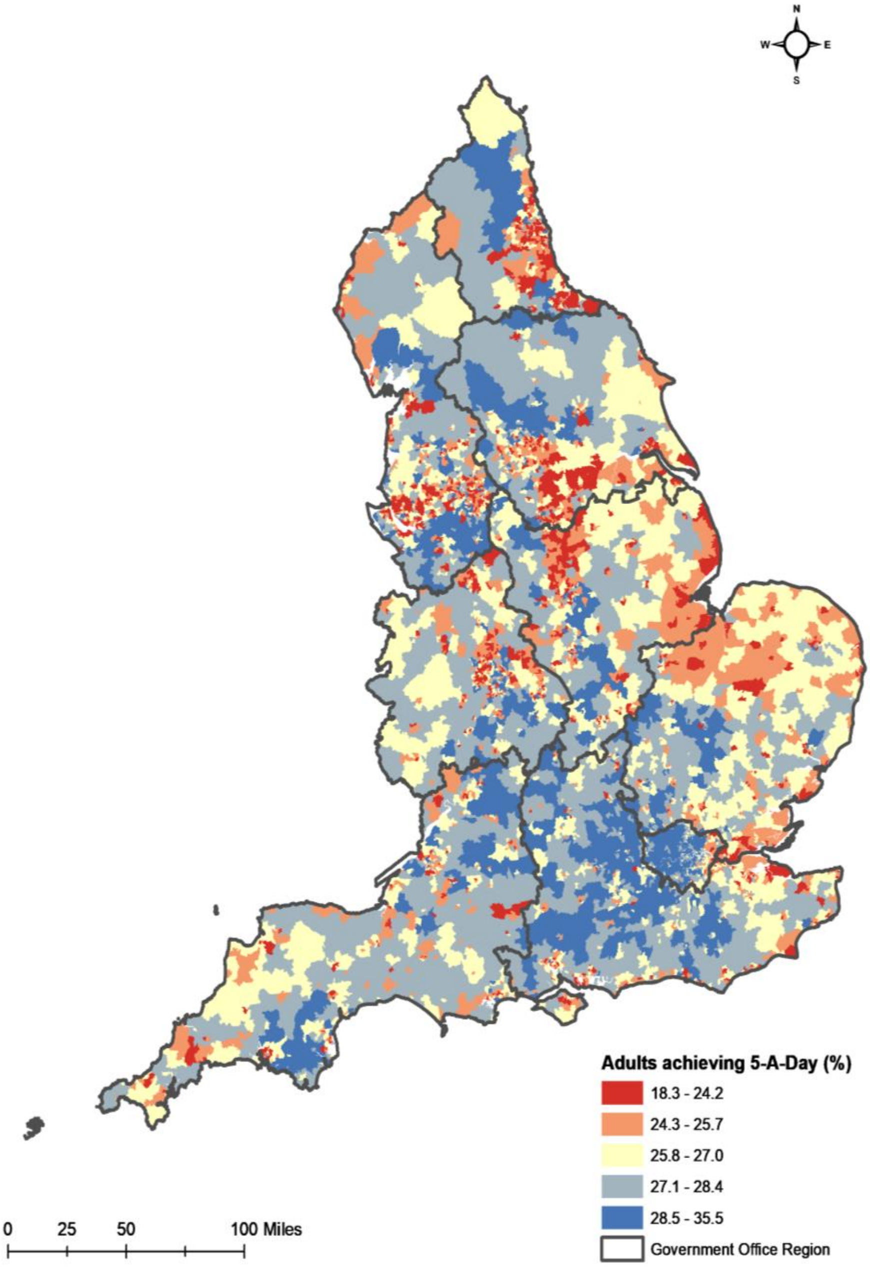


Figure 6.3: IPF model estimates for the achievement of the 5-A-Day target (national level)

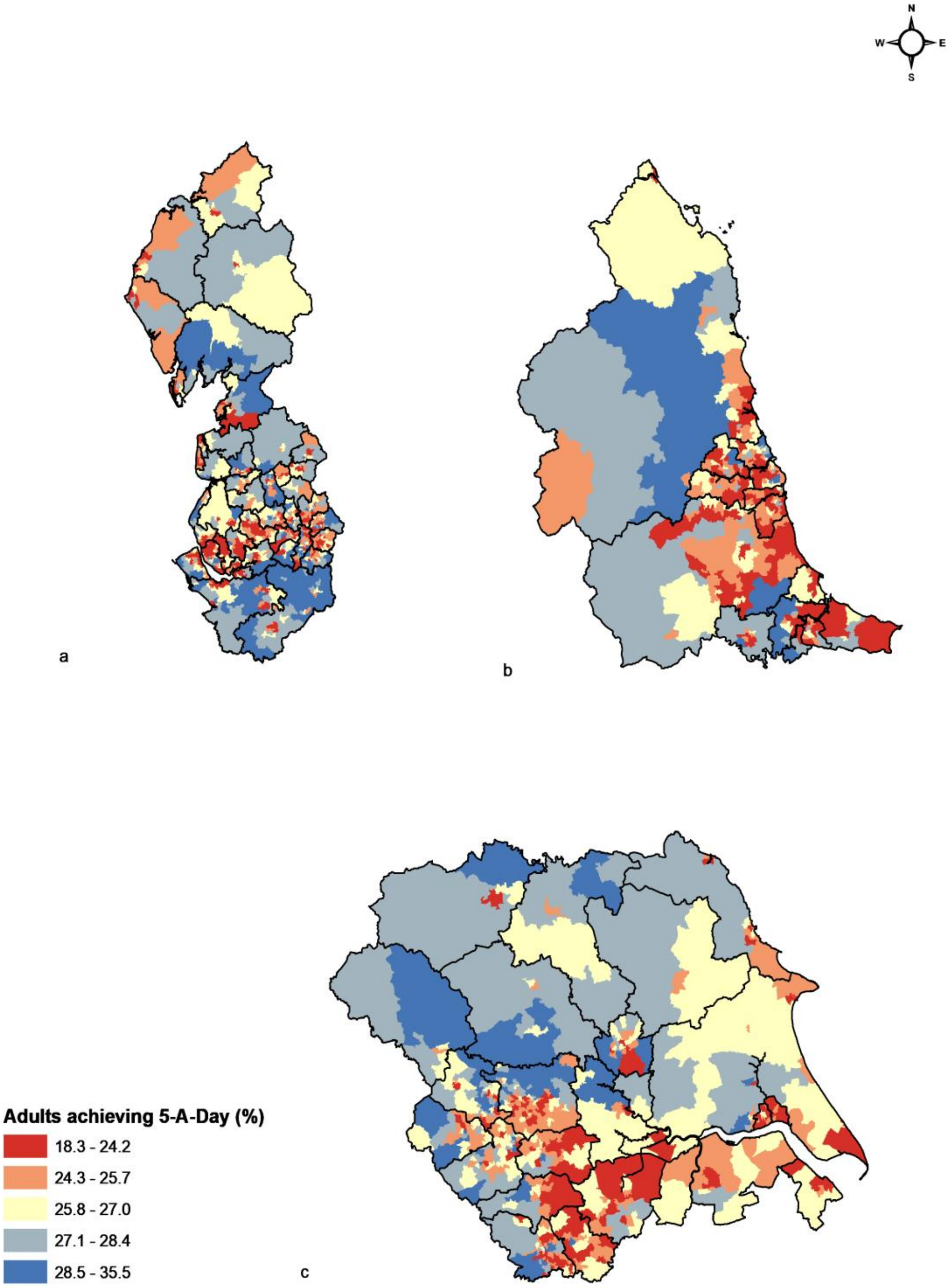


Figure 6.4: IPF model estimates for the achievement of the 5-A-Day target for the a) North West b) North East and c) Yorkshire and Humber region

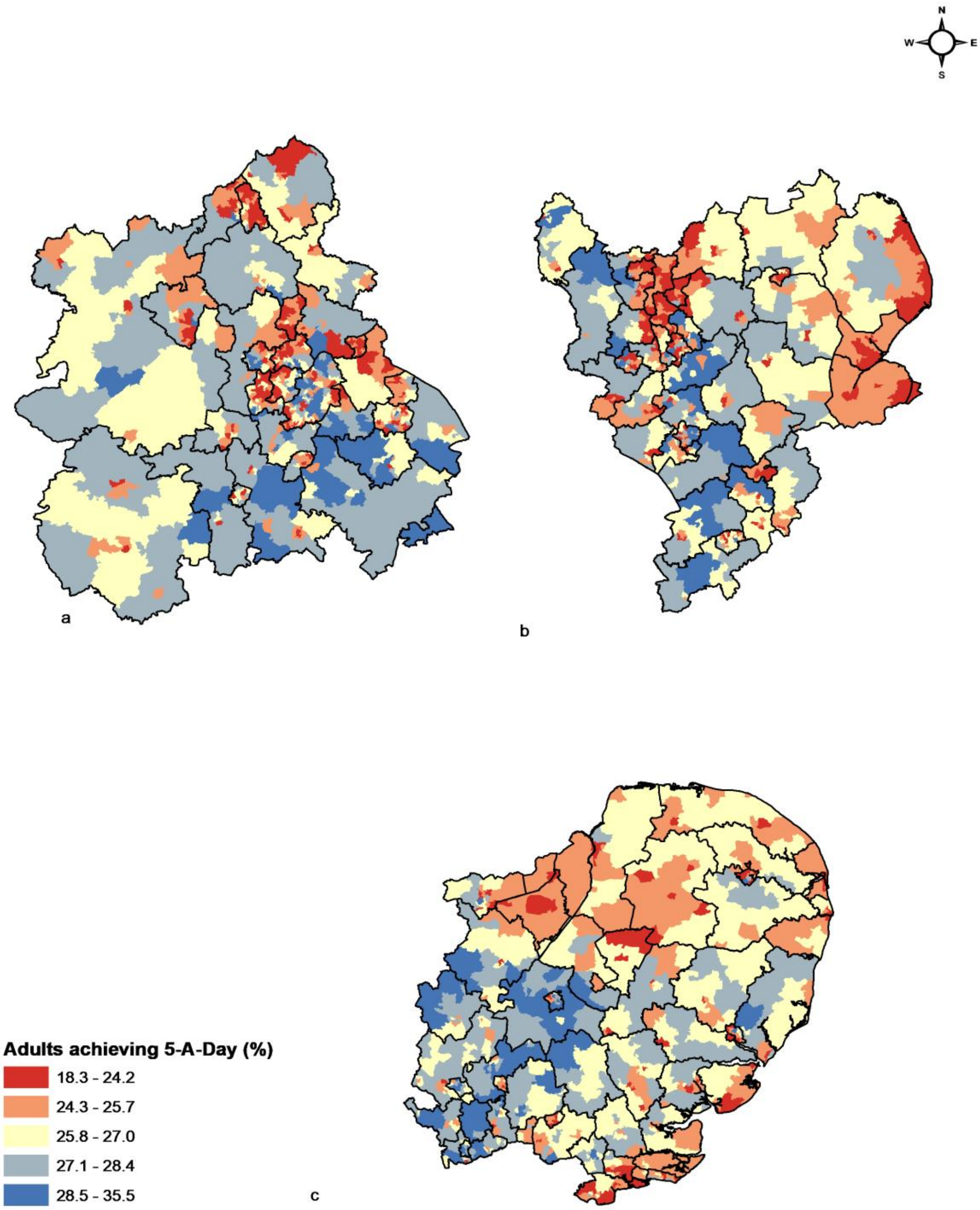


Figure 6.5: IPF model estimates for the achievement of the 5-A-Day target for the a) West Midlands, b) East Midlands and c) East of England region

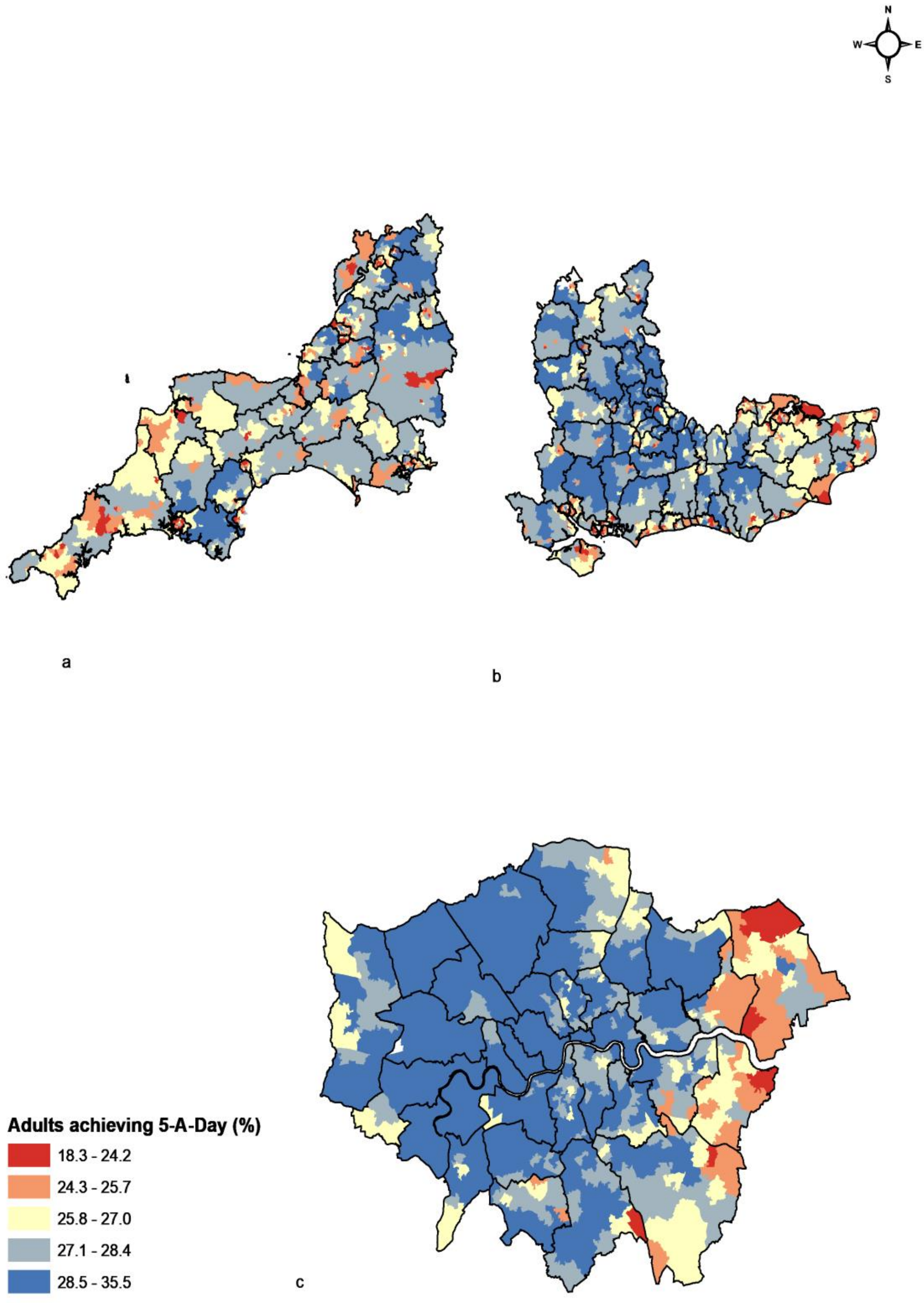


Figure 6.6: IPF model estimates for the achievement of the 5-A-Day target for the a) South West, b) South East and c) London region

Figures 6.3-6.6 showed that all quintiles were represented within each region, with the lowest quintile (shaded in dark red) representing 18.3%-24.2% of the MSOA population achieving the 5-A-Day target and the highest quintile (shaded in dark blue) representing 28.5%-35.5% of the MSOA population achieving this target. Potential "hotspots" (MSOAs with the lowest quintile, adherence) appeared to be concentrated towards the more urban sections of the North and Midlands, particularly areas within Liverpool, Newcastle, Kingston-Upon-Hull, Stoke-on-Trent and Nottingham. On the contrary, MSOAs within the highest quintile tended to be concentrated in the South of England and London. A distinct pattern was observed for London, with the majority of MSOAs in this region falling either within the fourth or fifth quintile (Figure 6.6, inset C). Harrow, Westminster and Richmond-Upon-Thames were some of the areas in London with the highest quintile of adherence to 5-A-Day.

6.3.3 Data validation and credible intervals results

As described in Section 6.2.5, three-level hierarchical multi-level logistic regression modelling was used primarily to obtain the area level (PSU level two) residual error variance needed to calculate credible intervals. The GOR and PSU level variance which accompanied the final three-level hierarchical multi-level model (model with the constraint variables education, age, ethnicity and NS-SEC) was 0.022 and 0.210 respectively. The closeness of the GOR level variance to zero meant that there was no statistically significant difference between regions/GORs. However, the PSU level variance showed that there was more variation between MSOAs (PSUs) than GORs in terms of 5-A-Day achievement. Although credible intervals varied across the 6,791 MSOAs, the average credible interval ranged from 12.7%-45.5% (Figure 6.7).

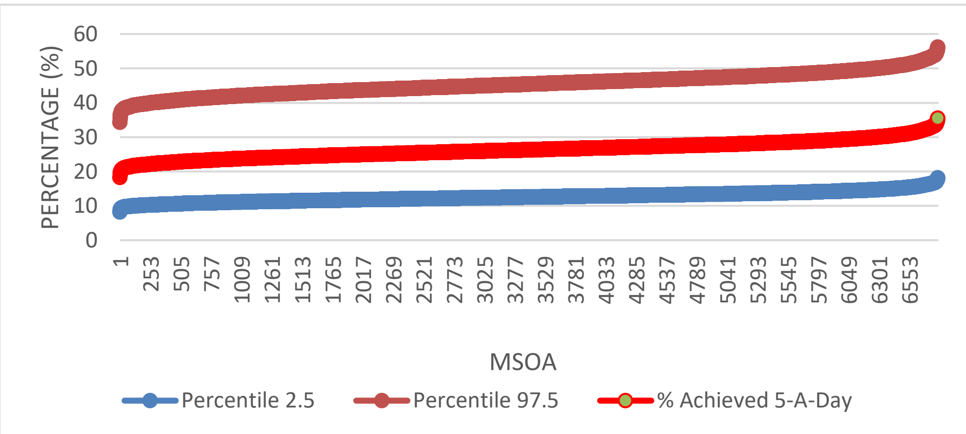


Figure 6.7: Credible intervals for 5-A-Day achievement across MSOAs in England,2008-16

In terms of the internal validation of data, the average percent error for each of the four constraint variables was low and in keeping with the recommended internal goodness-of-fit error of 20% (Lovelace, Birkin, Ballas and van Leeuwen, 2015). Education, ethnicity and age had a percent error of 0.014%, 0.014% and 0.086% (respectively), whilst NS-SEC had the highest percent error of approximately 0.464% compared to the known census population (updated with ONS (2018b) data).

6.4 Discussion

The primary aim of this chapter was to estimate and visualise 5-A-Day achievement/adherence across MSOAs (small areas) in England. This aim was achieved by using the IPF deterministic spatial microsimulation method. Overall, the results presented showed that each region had MSOAs which fell within each of the five quintiles. However, MSOAs which had a percentage achievement below 24.2% (the quintile with the lowest 5-A-Day achievement) tended to be concentrated in and around urban centres in the North of England and the Midlands. Noteworthy areas (i.e. areas in the lowest quintile of achievement) identified within the North of England included, Liverpool within the North West; Newcastle in the North East and Kingston-Upon-Hull in Yorkshire and the Humber. Approximately 54% of the 61 MSOAs/small areas in Liverpool were in the lowest quintile. Similarly, 48% of the 29 MSOAs in Newcastle and 72% of the 32 MSOAs in Kingston-Upon-Hull fell within the lowest quintile and can thus be flagged as potential hotspots for intervention. Although some sections of Birmingham, Wolverhampton and Coventry were within the lowest quintile, Stoke-on-Trent was the most noteworthy within the West Midlands region. None of the 34 MSOAs in Stoke-on-Trent fell within the highest quintile, but approximately 59% fell within the lowest quintile. Nottingham was the most notable in the East Midlands, given that approximately 47% of the 38 MSOAs in Nottingham fell within the lowest quintile.

MSOAs within the highest quintile tended to be concentrated more towards the South of England, especially in and around London. With the exception of some sections of Havering (East London) which fell within the lowest quintile, the majority of MSOAs in London fell within the 4^{th} or 5^{th} quintile. For instance, all of the 30 MSOAs in Harrow (Central London) fell within the highest quintile, with the percentage of achievement ranging from approximately 31%-34%. Similarly, all of the 21, 24 and 23 MSOAs in Kensington & Chelsea, Westminster and Richmond-Upon-Thames, respectively, fell within the highest quintile of achievement. Previous studies have noted geographic variations in diet/fruit and vegetable consumption and related health behaviours in England. Roberts (2014) found that adults in London and the South of England were more likely to meet the national 5-A-Day target compared to those in the North of England. Oddy (2003) estimated that "Londoners" were more likely to consume foods such as fruits, more than any other region in England. The findings of this project and those from the aforementioned studies, highlight inequalities in fruit and vegetable consumption at the small area/MSOA level and could also point to the traditional English North-South divide, in which persons in the North of England are said to have relatively poorer diets and worse health outcomes than those in the South of England. However, there were some MSOAs (primarily cities) in the South of England which fell within the lowest quintile. For instance, 44% of the 32 MSOAs in Southampton (South East of England) and 50% of the 32 MSOAs in Plymouth (South West of England) fell within the lowest quintile of 5-A-Day achievement. These findings were consistent with Smith et al (2021) which found low fruit and vegetable consumption to be most prevalent in urban centres in the North of England such as Liverpool and Manchester, eastern boroughs in London and a few coastal cities in the South of England such as Southampton and Portsmouth. The IPF method used in the Smith et al (2021) study was replicated in the current project using data from the HSE instead of NDNS data used in Smith et al (2021). Thus, the findings of Smith et al (2021) corroborated the results obtained in Research Question 3 and helped to identify small areas across England (not only in the North of England or the Midlands) which could be prioritised for dietary improvements.

Previous studies have noted the association between the local food environment and diet (Maguire, Burgoine and Monsivais, 2015). Areas with higher levels of deprivation and those which are more ethnically diverse tend to have greater access to takeaway (fast food) outlets than more affluent white areas (Black, Moon and Baird, 2014). The marked increase in the number of takeaway outlets (especially in the North of England) in recent years has been associated with the excess consumption of takeaway foods, especially in low-income areas (Blow, Gregg, Davies and Patel, 2019; Maguire, Burgoine and Monsivais, 2015; PHE, 2018c; Saunders, Saunders and Middleton, 2015). These findings could be linked to the low prevalence of 5-A-Day achievement observed (in this project) for areas in the North of England such as Liverpool (the local authority with the second largest proportion of highly deprived neighbourhoods in England (Ministry of Housing, Communities & Local Government, 2019). However, due to its complexity, additional studies are needed to further explore the relationship between diet and the local food environment in England (Black, Moon and Baird, 2014).

Age, ethnicity and SEP factors such as education which were found in this chapter to be key predictors of dietary quality, have been noted in previous studies (Bambra, 2016) and was consistent with the results obtained in Research Question 1 (Chapter 4), even with the modelling of NDNS (rather than HSE) data within that chapter. Overall, these findings could explain some of the inequalities in 5-A-Day adherence observed across England. With regards to age, it has been generally accepted that older persons tend to reside in coastal and rural areas, while greater numbers of younger persons live in the more urban areas. It is possible that the greater prevalence of low fruit and vegetable consumption observed in the urban North and the

Midlands, could be due to the higher student and young adult population in areas such as Liverpool and Nottingham, in which the median age is below 30 years (McCurdy, 2019).

Ethnicity, also found in this project to be a strong predictor of 5-A-Day achievement, was one of the constraint variables used in the IPF model. The final model with constraint variables showed that persons within the Asian ethnic group had a 68% increased odds of target achievement, compared to White persons. Although the overall association between ethnicity and diet/related outcomes remains inconclusive (Higgins, Nazroo and Brown, 2019; Moon et al, 2007; National Obesity Observatory, 2011), this project's findings appeared to be in line with previous studies which reported 5-A-Day target achievement to be highest amongst Asians, especially Chinese and Indians (ONS, 2006).

Highest level of education, found previously (Acheson, 1998; Roberts, Cavill, Hancock and Rutter, 2013; Tiffin and Salois, 2012) and within the current project to be a strong predictor of 5-A-Day achievement/diet, could help to explain the higher percentage achievement observed in the South of England (especially in and around London) compared to the remaining regions. Statistics from ONS (2020b) showed that in 2019, approximately 54% of working aged adults (16-64 years) in England who had a Degree or equivalent qualification resided in London. Westminster, Kensington & Chelsea and Richmond-Upon-Thames, identified in the current project as being in the highest quintile of 5-A-Day achievement, were also found to be some of the areas in London with highest proportion of working age adults with a degree level or equivalent qualification. In contrast, Havering (identified as a potential hotspot for low fruit and vegetable consumption in this project), has been previously highlighted as the area in London with the second lowest proportion (34%) of working age adults with a degree level or equivalent qualification (ONS, 2020b).

Equivalised household income was not available in the 2011 census and was not consistently captured in the HSE. This variable was therefore not analysed for Research Question 3; hence, a limitation of this project. Nevertheless, small area (MSOA level) model-based income estimates from ONS (2020a) showed that in 2018, approximately 82% of the 50 MSOAs with the highest (equivalised household) incomes within England and Wales were located in London; including the previously mentioned local authorities of Richmond-Upon-Thames, Westminster and Kensington & Chelsea. These statistics, coupled with the findings of this project, appear to corroborate previous studies which noted that persons within high SEP groups (e.g. high-income persons or those with higher levels of education) tend to eat more fruits and vegetables than those in lower SEP groups (Acheson, 1998; Roos et al, 2000; Tiffin and Salois, 2012). It has also been postulated that the poor diet/health of persons from low SEP groups could be further amplified by living in disadvantaged neighbourhoods, which tend to have increased takeaway/less healthy food options (Macintyre, Maciver and Sooman, 1993). However, more research is needed to explore the association between the local food environment and diet, because persons living in disadvantaged areas in England may not always be disproportionately worse off than persons in more affluent areas (Black, Moon and Baird, 2014).

6.4.1 Strengths and limitations

In the absence of diet-related data at the small area level in the census and national surveys such as the HSE, IPF has been successfully used to estimate smoking in New Zealand and Austria (Smith, Pearce and Harland, 2011; Tomintz, Kosar and Clarke, 2016), in addition to obesity (Edwards and Clarke, 2009) and mental health and alcohol consumption in England (Riva and Smith, 2012). Despite advancements in its use globally, there are still very few IPF studies which have modelled diet/fruit and vegetable consumption. James, Lomax and Birkin (2019) although noting the dearth of IPF studies on diet, focused on food and drink expenditure and not diet consumption. Smith et al (2021) is the only known study to date (besides this project) which used IPF to derive small area estimates for diet, specifically fruit and/or vegetable portions and the consumption of SSBs. Unlike Smith et al (2021), Research Question 3 of this project focused on overall 5-A-Day (binary) achievement across MSOAs as opposed to fruit and/or vegetable portions. The NDNS was the population dataset used by Smith et al (2021), whereas the HSE was used in the current project. This meant that the exploration of additional foods groups such as SSBs (captured in the NDNS) was not only outside this project's scope but not possible, as the HSE only captured data on fruit and vegetable consumption. Despite these differences, the findings from Smith et al (2021) helped to strengthen this project's findings, as both identified geographic inequalities in diet across MSOAs and highlighted potential hotspots within urban centres in the North of England and the Midlands. Fruit and vegetable consumption data captured in the Active Lives Survey (ALS) could have been used to estimate diet at the local authority level in England. However, a major limitation of the ALS was the survey's overestimation of fruit and vegetable consumption/5-A-Day compared to the NDNS and HSE (PHE, 2020a). Thus, the use of the HSE in Research Question 3 of this project can be considered an added strength, given that it is the primary survey used to monitor progress towards the achievement of national diet/health targets in England such as the 5-A-Day target (NatCen Social Research and University College of London, 2017) and its sample size over the 2008-16 survey period was more than six times the size of the corresponding 2008-16 NDNS pooled dataset used in Smith et al (2021).

The use of multi-level modelling for Research Question 3 (though not a requirement of IPF) was a particular added strength of this project because it acknowledged the structure of HSE data and provided the separate PSU level variance needed to calculate credible intervals. The findings from the final multi-level model showed that there was more variation between MSOAs (PSUs) than regions, which highlighted the importance of systematically exploring regional and sub-regional differences in diet (Cummins and Macintyre, 2002; Cummins, 2014). The width of credible intervals showed that there was still a lot of uncertainty as to the extent to which observed variation equated to true differences between MSOAs. However, the wide credible intervals observed in this project is commonplace in SAE at the MSOA level or more local levels, particularly with health behaviours (Taylor, Moon and Twigg, 2016).

The number of constraint variables used in the IPF model was restricted to four, which could be considered a limitation. However, this number has been proven to provide the most reliable estimates without over constraining the IPF model (Smith, Clarke and Harland, 2009). The IPF model also assumed that working age adults aged 16-64 years, living in England, replicate the behaviours of respondents in the HSE, which is an assumption which may not always hold true. Another limitation was the inability to compare IPF model estimates with actual 5-A-Day achievement at the MSOA level, due to the poor availability of detailed survey data at the small area level. However, it can be argued that the lack of data at the small area level, substantiates the overall need for IPF. Although external data validation continues to be a limitation of IPF, model estimates obtained in the current project appeared to be robust, as the average percent error for all four constraint variables did not exceed 1%, which was well below the recommended internal goodness-of-fit error of 20%.

Besides the low percent error, another major project strength is that this is the only known project which used IPF to estimate 5-A-Day achievement across MSOAs in England using data from the HSE. As mentioned, the project identified potential hotspots primarily in urban centres in the North of England and Midlands, reflecting the traditional North-South divide (Bambra, 2016) and areas possibly in need of more targeted dietary interventions. However, this does not negate the fact that 74% of working age adults in England failed to achieve the national 5-A-Day target (figures which were in alignment with NDNS data). This meant that although the prevalence of 5-A-Day achievement appeared to be greater in the South of England and London, compared to some urban areas in the North and Midlands, most of the population did not consume the recommended daily portion of fruits and vegetables. Due to the dearth of diet-related IPF studies, the findings of this project could be considered a baseline or starting point from which future studies can easily replicate, expand and investigate 5-A-Day achievement as well as non-achievement across MSOAs/small areas.

PHE and local authorities have recognised the importance of monitoring health outcomes at the small area level and have commissioned local level surveys for a few select local authorities. The findings of Research Question 3 highlighted the need for survey data at the small area level and additional studies which investigate the influence of the local food environment on dietary patterns in England. The increased availability of local level survey data and studies could improve the monitoring and evaluation of key diet-related targets such as the 5-A-Day target and could pinpoint opportunities for more targeted interventions. However, these types of surveys require a significant commitment in terms of time and financial resources, especially if they were to be rolled out across all local authorities or small areas. For this reason, the PHE regularly commissions small area estimates primarily for smoking and alcohol consumption (Pryce et al, 2020). However, to the best of our knowledge, this has not been done for diet/fruit and vegetable consumption. The project's identification of urban centres in the North of England and the Midlands, eastern boroughs in London and some coastal cities in the South of England, highlighted possible areas for prioritisation and more targeted interventions such as fruit and vegetable voucher schemes and subsidies, especially for persons in lower SEP groups. The findings of the projects also highlight the usefulness of SAE/IPF in providing policy makers and local authorities (who are ultimately responsible for localised public health) with reliable estimates of diet within respective areas. With ever tightening research budgets and the recent association made between diet (particularly fruit and vegetable consumption) and COVID-19 (Butler and Barrientos, 2020) now more than ever, local authorities and public health researchers need to make the best use of the data currently available to conduct the research necessary for the creation of more evidence-based and targeted dietary policies and interventions in England.

Chapter 7 Conclusion

The primary aim of this project was to explore regional and small area differences in dietary patterns in England amongst the traditional working age adult population (persons aged 16-64 years) in England. This aim was achieved by addressing three primary research questions.

7.1 To what extent does fruit and vegetable consumption (fruit and vegetable portions and 5-A-Day achievement) vary by region amongst persons aged 16-64 years in England?

The first research question (Research Question 1) investigated the extent to which fruit and vegetable consumption varied across the regions of England. Overall, no statistically significant differences were observed regionally. However, the achievement of the 5-A-Day target and the consumption of five or more portions of fruits and/or vegetables was most apparent amongst older persons, persons in non-white ethnic groups and those with higher levels of education and equivalised household income. These findings suggested that inequalities in fruit and vegetable consumption may be more associated with the characteristics of individuals (compositional factors), than geographic region of residence. In contrast to this project, Morris et al (2016) previously noted statistically significant differences in dietary patterns at both the regional and small area level, having adjusted for socio-demographic factors. The inconsistent findings observed between this project and Morris et al (2016) was likely due to this project's analysis of nationally representative data from NDNS, unlike Morris et al (2016) which utilised data from the UK Women's Cohort Study, which was constrained to middle-aged white women with predominantly vegetarian diets. Another major difference between the two studies was that Morris et al (2016) examined the overall dietary pattern of individuals, whereas the assessment of diet in Research Question 1 was based solely on fruit and vegetable consumption, which is a relatively simple summary measure and captures only one aspect of diet. The results of this project were based on nationally representative data, which is an advantage over the Morris et al (2016) study. However, it was also recognised that individuals consume a variety of foods other than fruit and vegetables. Thus, the second research question aimed to provide a more detailed snapshot of aspects of the English diet and the extent to which dietary patterns in England varied across the regions.

7.2 What are the dietary patterns of persons aged 16-64 years in England and to what extent do dietary patterns vary regionally?

Principal Components Analysis (PCA) was used in Research Question 2 to gain insight into the dietary pattern of working age adults at the national level and disaggregated by region. With the exception of the North West of England, East of England and West Midlands regions, the leading dimension (first component) to diet was health conscious (consisting of fruit; yogurt, fromage frais and dairy desserts; salad and other raw vegetables; skimmed milk; tea, coffee and water; high fibre breakfast cereals and other margarine, fats and oils) in most regions. Sweet puddings, fats and oils were the main aspects of the East of England dietary pattern, whereas the West Midlands had distinct elements of the traditional English diet with favourites such as fish and chips. The leading dimension to diet in the North West was characterised as “Western/Takeaway” as it mainly featured takeaway foods, chips, processed meat, fried foods and fizzy drinks. The Western style dietary pattern (for the first component) identified for the North West suggested a North-South divide in diet/health as noted in previous studies (Billson, Pryer and Nichols, 1999; Whichelow, Erzinclioglu and Cox, 1991). However, similar to the findings of Research Question 1, no statistically significant difference was found between the regions after adjusting for compositional factors.

The effect of geographic region on diet was not within the scope of the Roberts, Cade, Dawson and Holdsworth (2018) study. Nevertheless, Roberts, Cade, Dawson and Holdsworth (2018) used PCA, applied the same 60 pre-defined NDNS food groups used in this project and found compositional factors, primarily age, sex, ethnicity and SEP (specially NS-SEC occupational grade) to be positively associated with the health conscious, “Fruit, Vegetables, Oily Fish” dietary pattern identified in that study. Likewise, Robinson et al (2004) having explored PCA-derived dietary patterns of a sample of young women in Southampton England, found age and especially highest level of education to be the variables most associated with the “Prudent” (health conscious) dietary pattern identified in that study. These findings corroborated those obtained within the current project and suggested that some of the differences in diet could be associated with compositional factors (predominantly age, SEP and ethnicity) or factors associated with who persons are, rather than where persons reside as measured regionally (context). Moreover, the statistically insignificant difference between region and diet could have been partially due to the geographic scale, particularly the higher/coarse regional level investigated in both research questions, which was the only scale available in the data. The final research question therefore aimed to examine whether there was greater variation in diet at a more local scale than the regional scale examined in Research Questions 1 and 2.

7.3 To what extent does estimated achievement of 5-A-Day vary across smaller areas of England?

IPF (a deterministic reweighting SAE) method was used in Research Question 3 to estimate and visualise 5-A-Day achievement across MSOAs in England using data from the HSE 2008-16. The results of this exercise showed that differences in 5-A-Day achievement were greater between MSOAs (PSUs) than GORs in England. Urban centres in the North of England and West Midlands, eastern boroughs in London and some coastal cities in the South of England were identified as potential hotspots or areas for prioritisation and more targeted interventions. The IPF method used in the recent Smith et al (2021) study was replicated in the current project using data from the HSE instead of NDNS data used in Smith et al (2021). Despite the different datasets used, Smith et al (2021) also found low fruit and vegetable consumption (estimated) to be most prevalent in urban centres in the North of England such as Liverpool and Manchester, eastern boroughs in London and a few coastal cities in the South of England such as Southampton and Portsmouth. The results of that study corroborated those obtained in Research Question 3 and suggested that the statistically insignificant geographic differences found in Research Questions 1 and 2 could have been due to the coarse geographic scale (GOR/region) examined within those questions. Although the current project identified some small/local areas possibly in need of dietary change, more than 70% of working age adults in England failed to achieve the national 5-A-Day target. This meant that even though the prevalence of 5-A-Day achievement was seemingly lower in some urban areas in the North and Midlands and sections of London and the South of England, the majority of the population did not consume the recommended daily portion of fruits and vegetables. It is acknowledged that 5-A-Day is a simple measure and only captures one aspect of diet. However, because 5-A-Day continues to be a major indicator of diet quality, its use and the findings presented within the current project were relevant and further highlighted the issue of sub-optimal diets in England and the need for more evidence-based policies which tackle this issue (Newton et al, 2015).

7.4 Overall project strengths and limitations

Prior to this project, the majority of diet-related studies conducted in England tended to focus on childhood obesity or the general association between diet and related outcomes such as obesity or dietary differences across the UK rather than England specifically. The few studies which attempted to explore the diet of adults in England were either not representative of the English population or generally alluded to the stereotypical North-South divide without comparing diets at the regional and/or sub-regional level. The diet of persons under the age of 16 years and those

65 years and over was not assessed within the current project, which could be a limitation. However, as far as is known, this is the only project which carried out such a detailed assessment of fruit and vegetable consumption and dietary patterns of working age adults in England and explored geographic differences in diet at both the regional and sub-regional level, using data from two nationally representative surveys (the NDNS and HSE). The findings of this project demonstrated the importance of systematically exploring regional and sub-regional differences in diet, highlighted current data-related challenges which hinder researchers from properly assessing diets in England, helped to identify gaps in the literature and topics in need of further research and revealed possible areas for intervention which would have otherwise gone undetected.

The availability of data for over 7,000 foods (collected via four-day food diaries) was one of its key features of NDNS which proved beneficial to the current project. Although the HSE only captured data on fruit and vegetable consumption, its target annual sample size of approximately 10,000 persons annually was advantageous. However, the cross-sectional nature of the NDNS and HSE meant that causal inferences could not have been made since both surveys were designed to provide a snapshot of the diet of individuals sampled at a particular point in time. Regardless, the NDNS and HSE are two of the major repeated cross-sectional diet-related surveys used by the Government to monitor the diet of individuals in England and inform policy decisions (Bates et al, 2012). However, it must be acknowledged that the removal of missing cases (complete case analysis) from the HSE is a possible limitation. Nevertheless, the inclusion of both surveys in this project made it possible for a descriptive comparison to be made, the results of which showed that diet-related data captured in both surveys were comparable, despite the vast difference in their sample size (over the 2008-16 period) and the different dietary assessment method used in each survey.

The data landscape process used to assess and eventually select the NDNS and HSE for analysis in the current project was another project strength, given that it was detailed and highlighted the strengths and limitations of some of the major diet-related repeated cross-sectional and longitudinal surveys conducted in England over the last 48 years (1970–2018). This detailed assessment not only justified the project's final decision to analyse data from the NDNS and HSE but was deemed (via peer reviewed journal publication; Campbell et al, 2020) a valuable source of information for other/future researchers interested in conducting secondary (quantitative) research on diet. Despite the many data challenges highlighted generally and in the data landscape, the project was still able to make the best use of the (secondary) diet-related survey data currently available. The pooling of data over the 2008-16 period helped to increase the sample size of the HSE and particularly the NDNS, which further strengthened project findings. In addition, the project's use of several advanced quantitative statistical methods (multi-level modelling, PCA and IPF spatial microsimulation) demonstrated how secondary survey data can be used to identify inequalities in diet (not obesity) in England despite the many data-related challenges.

The use of fixed effects/single-level modelling without considering the inherent structure of survey data, could have resulted in underestimated standard errors of regression coefficients, possibly overstated statistical significance and misleading conclusions. Thus, the project's utilisation of the multi-level modelling method for all three research questions was a major strength because it acknowledged the multi-stage design/structure of the NDNS and HSE, accounted for the variability at the different levels and facilitated the exploration of the effect of region/geographic location and individual level characteristics on diet. The PCA method used was advantageous given that the project's primary aim was to explore the types of foods actually consumed by individuals and possible differences across the regions of England, with no previous hypothesis or expectation. The project's ability to produce PCA (empirically) derived dietary patterns for the overall NDNS sample as well as for each of the nine regions in England was considered another major strength, as very few studies have done separate PCAs, especially based on geography/region. However, future studies could gain from exploring the remaining components/patterns (e.g. the first four components and not just the first component as done in this project), especially for regions such as the North West, West Midlands and the East of England, which had dietary patterns which deviated from that observed at the national level.

The IPF microsimulation method used in this project was similar to that used in the recent Smith et al (2021) study, with the major difference being this project's use of data from the HSE rather than NDNS data used in Smith et al (2021). The IPF method produced reliable estimates which were consistent with the results from Smith et al (2021) and helped to identify inequalities in 5-A-Day achievement/adherence across MSOAs (small areas) in England, which were novel findings in the absence of representative local level survey data. IPF model estimates obtained appeared to be robust and in keeping with recommended internal goodness-of-fit measures, which served to further strengthen the overall findings of this project. Additionally, measures of uncertainty around point estimates have not been traditionally a part of IPF or other spatial microsimulation methods. Therefore, the addition of credible intervals to the point estimates within this project was a further improvement to the previous use of spatial microsimulation for health outcomes (Riva and Smith 2012; Edwards and Clarke, 2009; Tomintz, Kosar and Clarke, 2016). Although the credible intervals produced in this project were relatively wide, this is customary in SAE at the MSOA level or more local levels, particularly with health behaviours (Taylor, Moon and Twigg, 2016).

The poor availability of diet-related survey data at the small area level meant that there was no external data source with which IPF model estimates could have been compared or validated (externally), which was another limitation. It is also acknowledged that the exposures or activity spaces people experience daily should be considered when assessing the potential impact of context on diet (Perchoux, Chaix, Cummins and Kestens, 2013). However, because geographic data captured in the NDNS and HSE were constrained to the regional level (GOR in which participants resided), the relationship between dietary patterns and the overall food environment (including activity spaces/areas where individuals typically visit in their daily lives such as work and school) could not be explored, which is another possible limitation. The project's sole use of GOR and MSOA (which are pre-defined and conventional geographic measures) could be viewed as a limitation because additional area attributes/features such as the family and home food environment, social capital, the built environment (among others) and the causal processes which link place and diet/health were not explored within the current project. Key area attributes/features such as familial influence, the home food environment, local and cultural norms and social capital are acknowledged as contributors to diet (Cummins, Curtis, Diez-Roux and Macintyre, 2007). However, data on the features/attributes of areas and home food environment data are typically not available in nationally representative cross-sectional surveys such as the NDNS and HSE used in this project (Cummins, Macintyre, Davidson and Ellaway, 2005). In addition, the project's primary aim was to fill a current research gap by assessing the dietary patterns of individuals (not households) and possible geographic variations in England, using pre-defined geographic units, GOR and MSOA. Nevertheless, the findings of the current project highlighted the need for more detailed studies which examine the relationship between diet and the local food environment and more specifically, the causal mechanisms which connect diet to area, in order to pinpoint opportunities for intervention. Overall, given the dearth of regional and sub-regional diet-related studies in England, the current project can be considered a baseline or starting point from which future studies can easily replicate, expand and investigate 5-A-Day achievement as well as non-achievement across MSOAs/small areas, which is a major strength.

7.5 Project implications

England has gone through several political changes and has seen the publication of numerous policy reports, papers, strategies and the establishment of several health-related interventions and initiatives over the past several decades. However, similar to diet-related research, associated policies and interventions in England are generally more focused on individual responsibility and choice through educational policy measures and dietary guidelines and the prevalence of childhood obesity, than actual dietary consumption at the population level (Asaria, 2017; Caroline Walker Trust, 2009; Mozaffarian, Angell, Lang and Rivera, 2018; Ulijaszek and McLennan, 2016).

With the launch of the Health and Social Care Act (2012), public health became the responsibility of local authorities and not the NHS (Gadsby et al, 2017). Since then, several local authorities have attempted to address some of the wider environmental and structural determinants of diet and obesity through the development of initiatives which encourage food outlets to use healthier ingredients, menus and cooking practices and the placement of restrictions on the opening of new hot food takeaway outlets close to schools, leisure centres and other locations frequented by children (PHE, 2017). Although laudable, these local level interventions are still centred primarily within schools and their surrounding areas and focused mainly on children. The recent launch of the national Better Health campaign in July 2020 signalled the Government's transition from the sole focus on childhood obesity to the improvement of diet/health amongst adults in England. Thus, this project's focus on fruit and vegetable consumption/5-A-Day and dietary patterns within the working age adult population was timely and in keeping with current national priorities. The emerging evidence of the indirect influence of diet/fruit and vegetable consumption on COVID-19 (Butler and Barrientos, 2020) also brings to the fore the urgent need for more policies and interventions aimed at tackling dietary inequalities, which if not addressed in the short-term could be further exacerbated or widened as a result of the pandemic.

With increasingly limited research budgets, local authorities, funding agencies and research institutions are forced to consider new, cost-effective and creative methods (e.g. big data and digital technology) of maintaining existing repeated cross-sectional and cohort studies and overcoming geographical constraints (MRC, 2014). Loyalty card supermarket data and web-based food dietary assessment surveys such as MyFood24 are some of the recently proposed big data sources/emerging technologies capable of providing more real time food and nutrient data at the regional and sub-regional/small area level with sufficient sample sizes (Jenneson, Morris, Greenwood and Clarke, 2021; Wark et al, 2018). Although more cost-effective and less burdensome for researchers and survey participants, these technology-based dietary assessment methods are still not clearly defined and are not widely available for re-use at present. As such, SAE methods such as the IPF method used in this project are a cost-effective tool which can be further utilised to assist local authorities (ultimately responsible for the diet and health of constituents) in identifying the areas of greatest need, prioritisation and intervention (Smith, Heppenstall and Campbell, 2021).

Current interventions such as the Change4Life social marketing campaign (launched in 2009) and the more recent Better Health campaign are primarily focused on the distribution of information (via TV, print and digital media) and tools to help people make better choices related to diet and health (Ulijaszek and McLennan, 2016). Although the provision of diet-related information is necessary, interventions such as Change4Life are downstream or high-agency interventions which require individuals to be motivated to engage with the information received and be able to use their personal resources to make the recommended changes (Adams, Mytton, White and Monsivais, 2016). The establishment of a tax on sugary drinks in 2018 was an attempt made by central Government to acknowledge the wider structural influences on food choice and consumption. The sugar tax was an indirect tax levied at beverage manufacturers, with the major aim of driving the reformulation of sugary drinks (The Food Foundation, 2017). Advocates of this "low-agency" response to food choice argue that interventions such as the sugar tax are more equitable and effective as individuals may not directly incur any added costs, do not have to actively engage with information and are not required to actively change their behaviours (Adams, Mytton, White and Monsivais, 2016). Instead, the onus is on food manufacturers to reformulate and reduce the amount of added sugars in products. However, the taxing of unhealthy foods tends to have a more negative impact on poorer segments of society (Tiffin and Salois, 2012), as producers still have the option of absorbing the costs of the levy or transferring reformulation costs to consumers through increased prices (The Food Foundation, 2017). In addition, the taxing of SSBs only without the inclusion of other high fat, sugar and salt (HFSS) foods, could cause individuals to simply switch from sugary drinks to other unhealthy foods and beverages not included in the levy (The Food Foundation, 2017). Therefore, other interventions such as price reductions and food subsidies on healthier foods such as fruit and vegetables could also help to improve the diet of individuals in England, especially those from low SEP groups (Beech, Holmes, McKenna and Cooper, 2020; McFadden et al, 2014).

In light of COVID-19 and the likely short and long-term impact on the diet, particularly for low SEP groups, the use of existing Governmental programmes and/or the increased collaboration/partnership between local authorities and private sector organisations/charities could be a cost-effective and practical means of assisting those most in need of dietary improvement. The Healthy Start programme is a means tested food subsidy programme which offers food vouchers for fruit, vegetables, milk and vitamin supplements to pregnant women and breastfeeding children from low-income households (NHS, 2021). Although currently focused on pregnant women and children, the Healthy Start programme is an existing Government funded programme which could be further expanded to provide vouchers (e.g. fruit and vegetable vouchers) for adults from low SEP groups (Beech, Holmes, McKenna and Cooper, 2020). There is also increased opportunities for local authorities to partner with charitable organisations such as FareShare which currently seek to reduce food waste and redistribute food to persons and communities identified as most in need. It is acknowledged that there is no single strategy or intervention which will reduce diet inequalities in England (Beech, Holmes, McKenna and Cooper, 2020). However, a combination of interventions (regulatory and fiscal policies), backed by the supporting empirical data/evidence, such as that presented in the current project, could help to create a more supportive environment, possibly improve the diet of individuals and reduce identified diet/health inequalities in England.

7.6 Project Summary and Closing Remarks

Prior to this project, the dietary pattern of individuals and how patterns vary across (administrative) regions and small areas in England was a relatively unexplored topic. Previous diet-related studies conducted in the UK tended to focus on health outcomes such as obesity instead of diet or placed more emphasis on differences between constituent countries (particularly the Scottish and English diet) than diets in England specifically. Morris et al (2016) was one of the few studies which attempted to examine the geography of diets in England/UK. However, that study was not representative of the English population. In order to help improve upon this limitation, the current project sought to gain insight into the types of foods actually consumed by individuals across all nine administrative regions of England (nationally representative), provide a more recent empirical snapshot of dimensions of the "English diet" and examine whether traditional region/area-specific dietary patterns, (beyond traditionally held stereotypes) had still existed or could have possibly been replaced by a more Western style dietary pattern. A multi-scale approach was used within the current project to assess diets in England at the higher national, regional and finally the local level using pre-defined geographic measures (GOR, MSOAs). This three-tiered geographic assessment of diets was carried out in recognition that the relationship between compositional and contextual factors is complex and operates at various geographic scales. Moreover, the analysis of more than one geographic scale was in acknowledgement of the MAUP and the possibility that statistical results obtained were likely to vary according to the geographic scale at which data were measured and analysed (Openshaw, 1984; Wong, 2009).

The English diet (derived in the current project) seemed to have some traditional and Western elements, notably in the West Midlands and the North West administrative regions. Traditional dishes such as the famous English Fish and Chips, Toad in the Hole and Sausage and Mash were observed for the West Midlands region and indicated the possible retention of some aspects of the traditional English diet. However, future research on the remaining components or dimensions of diet (not explored in this project) would need to be carried out before any further assertions can made. The identification of mostly takeaway foods as the leading dimension in the North West of England, was not only synonymous with the Western diet, but also initially pointed to the stereotypical North-South divide in diet and health in England. Despite the unique patterns identified particularly in the North West and West Midlands, overall, variations in fruit and vegetable consumption and overall diet were not statistically significant at the GOR level. Estimated inequalities in diet were found to be greater between small areas (MSOAs) than the regions of England. Urban centres in the North of England and Midlands, sections of London and some coastal cities in the South of England were identified as areas which could benefit from dietary improvements and more targeted interventions such as food subsidies or vouchers. However, regardless of geographic region/area, more than a half of working age adults in England had diets (fruit and vegetable consumption) which were not in alignment with current dietary recommendations. This suggested that although there were potential "hotspots" or small areas in need of prioritisation, some of the differences in diet (inclusive of fruit and vegetable consumption) may be more associated with composition or the characteristics of individuals such as their age, ethnicity and SEP, than region or area of residence. Although compositional factors appeared to be more associated with diet than geographic region, it must be acknowledged that there is a complex, mutually reinforcing (rather than mutually exclusive) and reciprocal relationship between people and place (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Frohlich, 2000).

Whilst it is acknowledged that socio-demographic factors such as an individual's age, ethnicity, sex, their level of education and income can contribute to their diet and overall health, individuals are also simultaneously exposed to multiple "contexts" beyond their residential area (e.g. household/family environment, social, school, peers, workplace, places of leisure etc), which vary in time and space and operate at various geographic scales, from the global, national, regional to local scale/level (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Dearden, Lloyd and Green, 2020; Frohlich, Corin and Potvin, 2001). Moreover, just as individuals are a part of and/or interact with multiple social groups/sub-groups within a population; areas within England also have different population compositions, physical and social contextual area characteristics and varying opportunity structures which set them apart from each other and which collectively contribute to diet and health inequalities (Bambra, 2016; Dearden, Lloyd and Catney, 2019; Diez-Roux, 2001). Individuals, their attitudes and their overall diet/health have the potential to shape and be shaped by different types of people and the various "contextual" environments in which they live and operate regularly (Twigg, 2014). This is partially due to the fact that persons with similar socio-demographic characteristics tend to cluster in space and because individuals in the same area typically share common contextual influences (Dearden, Lloyd and Catney, 2019). The complex, multi-dimensional and reciprocal relationship between diet/health, people and place, simply means that people cannot and should not be separated from place (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Dearden, Lloyd and Green, 2020; Diez-Roux, 2002).

Diet-related research should ideally aim to understand and conceptualise/hypothesise the reason(s) for and the causal processes which connect area to diet (Chen and Antonelli, 2020). However, given the limited research conducted previously on regional variations in diet in England, this project was focused on filling this particular research gap by gaining empirical and more current insights into the types of food working age adults in England were eating and how dietary patterns varied geographically. The project was therefore more concerned with assessing *where* there was variance in diet (in relation to pre-defined geographic levels) rather than *why* there was variance at different geographic levels/scales or why place influences diet. Thus, additional features of the physical and social contextual environment likely to influence diet (such as social capital, the built environment, crime and community safety, the family/home food environment, local norms and values, the price and availability of healthy and unhealthy foods, the location of takeaway outlets etc) and the complex (causal) pathways which link area of residence to diet were not within the project's immediate scope, which was a limitation. Besides project scope, the generally poor availability of "true" area level data within secondary cross-sectional surveys and the census also limited the project's definition of context to conventional/pre-defined (aggregate) measures of residential region/area (GOR and MSOAs). This limitation hindered the possible examination of additional area effects on diet. Previous scholars have noted that due to the lack of "true" area level data, quantitative studies tend to or in many instances are forced to use more conventional/pre-defined, static definitions of space and place, which could explain why many of these studies find only a small proportion of variation attributable to "context", when compared to individual-level risk factors (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Pickett and Pearl, 2001) It is possible that the statistically insignificant association found between GOR and diet within this project, could be because additional characteristics/features of the physical and social contextual environment were not explored.

It would be impossible for this or any study to examine every attribute of the physical and social contextual environment likely to influence diet. Nevertheless, examination of additional features of the physical and social local contextual environment in future studies could reveal results different from those found within the current project and could provide greater insight into which aspects of "context" contribute to the dietary pattern of individuals in England (Cummins, Curtis, Diez-Roux and Macintyre, 2007; Frohlich, 2000). The adoption of a more "collective" or "relational" approach in future studies would be an improvement on the more conventional approach used within this project. However, future researchers would need to consider all possible data-related challenges (some of which were noted in this project) and devise the most appropriate work-around strategy, which may include some primary quantitative and qualitative data collection. The incorporation of qualitative data or a mixed methods approach to health and place research studies could help to fill in some of the missing relational qualities/mechanisms not easily detected in strictly quantitative studies (Diez-Roux, 2001; Gong and Hassink, 2020; Potvin and Hayes, 2007). A key first step would be for researchers to carefully hypothesise the pathways or mechanisms which they theorise to link diet to the wider contextual environment. Structural equation modelling is an approach which could aid in the development and testing of theoretical models that link area attributes to diet and health (Stafford et al, 2007). Though, researchers would have to determine whether this method best suits their research needs.

"True" area level contextual data is difficult to collect at the national level, partly due to the lack of consensus on which features of the contextual environment most influence diet and health (Frohlich, 2000; Mollborn, Lawrence and Saint Onge, 2021). For instance, the family and the overall home environment has been generally recognised for its positive influence on diet quality and health outcomes. Yet there is still no formal consensus and very limited data/research regarding the features of the home food environment most associated with adult diet quality (Glanz et al, 2021; Kegler, Hermstad and Haardorfer, 2021). Another related issue is the significant allotment of time as well as financial, physical and human resources required to collect, process and analyse area level contextual data at the national level (Cummins, Macintyre, Davidson and Ellaway, 2005, Frohlich, 2000; Mollborn, Lawrence and Saint Onge, 2021). This could explain why true area data are not usually available in routine nationally representative surveys such as the NDNS and HSE analysed in this project. If sufficient funding is available, future studies could consider the collection of diet and area level data from a sample of cities/small areas across the regions of England. Some of the potential "hotspots" identified within this project could be considered for more detailed examination. These include areas such as: Liverpool, Newcastle and Kingston-Upon-Hull in the North of England, Birmingham, Wolverhampton, Stoke-on-Trent and Nottingham in the Midlands and Southampton and Portsmouth in the South of England. Although not generalisable, these small-scale case studies could provide researchers, policymakers and local authorities (ultimately responsible for the diet and health of individuals) with more insight into local dietary patterns and the features of the social and physical environment likely to have a positive or negative impact on diets in England and the areas or sub-groups within the population in need of more targeted interventions (Diez-Roux, 2001). This more detailed assessment would be especially useful for regions/local authorities such as in the North West of England, where mostly takeaway foods were identified within this project.

Future studies could (as far as possible) also benefit from acknowledging the activity spaces people experience (daily and over the lifecourse) and how these vary over time and space and at different geographic scales. The limitations of cross-sectional survey designs (such as the NDNS and HSE surveys analysed in this project) have been well established. However, there is an opportunity for future studies to make use of more longitudinal data/research designs and increasingly affordable global positioning system (GPS) technology (Mollborn, Lawrence and Saint Onge, 2021). Unlike cross-sectional studies which capture the diet of individuals at one point in time, longitudinal study designs have the potential to track eating habits, possible changes over the lifecourse and where possible, the influence of the social and physical environment on diet over time. The capturing of GPS coordinates could also allow future researchers to track the movements of individuals over the course of their day, week, or lifetime (based on project funding and scope) through multiple “contexts” (beyond the residential area, such as work, school, shopping centres, food outlets, social support network, the home location of family and friends) which have potential health-promoting or health-damaging features (Cummins, Curtis, Diez-Roux and Macintyre, 2007). The incorporation and use of more longitudinal and activity-space/GPS data could help to further enhance research on health and place (Mollborn, Lawrence and Saint Onge, 2021).

Overall, the findings of this project came at a time when adult sub-optimal diet and COVID-19 were thrust into the forefront as major public health challenges in England, following the Government’s “reset moment” in July 2020. Furthermore, the discontinuation of the Universal Credit uplift could affect more than five million families which benefited from an additional £20/week, towards the purchase of food and other basic amenities during the pandemic (Spoor, 2021). In the absence of a suitable replacement financial support scheme, the removal of the Universal Credit uplift benefit could significantly affect the long-term diet and health of persons in England (Mackley, Hobson and McInnes, 2021; Winchester, 2021). The implications are that diet-related data, especially at the local level, will become even more crucial to monitor and assess the diet of individuals and evaluate the success of current diet/health-related policies and interventions. With increasingly limited research budgets, researchers and policymakers have been encouraged to make “proper, effective and good” use of the secondary survey data currently available (Campbell et al, 2020; WHO, 2015b) and where possible/financially feasible, conduct the primary data collection necessary to fill in at least some of the noted data and research gaps (Gong and Hassink, 2020).

Despite the many noted data challenges, this project still demonstrated how secondary survey data along with a combination of quantitative statistical techniques can be utilised to provide an empirical snapshot of aspects of the English diet, examine geographic variations of aggregate data and identify possible food items and small areas in need of more targeted nutritional interventions. In the absence of representative local level secondary survey data, the use of the IPF microsimulation method produced robust and reliable estimates of diet quality across small areas in England. The identification and visualisation of potential "hotspots" helped to highlight areas/regions possibly in need of more detailed investigation in future studies. The findings of this project could be used as a guide for future studies to expand upon, given that methods such as IPF microsimulation can be easily replicated (with the use of existing secondary survey data). When coupled, IPF and additional research on the causal pathways which connect diet to the contextual environment could help local authorities to better understand the key drivers of dietary inequalities and the policy actions which could provide the most impact, especially amongst the more disadvantaged groups of the society.

From as far back as 1927, English novelist and critic Virginia Woolf noted that "what passes for cookery in England is an abomination. It is putting cabbages in water. It is roasting meat till it is like leather. It is cutting off delicious skins of vegetables" (Woolf, 1927). However, based on this project's detailed assessment, the issue now is not so much about the mundane elements of the traditional English diet but more about sub-optimal diets in England and the fact that despite decades of interventions and strategies, much work is still required to tackle this major public health issue.

Appendix A Detailed review of 17 major repeated cross-sectional and longitudinal surveys conducted in England over the January 1970-December 2018 period

	Repeated cross-sectional surveys	
	Living Costs and Food Survey (LCFS)	**Active Lives Survey (ALS)**
Survey Background	The LCFS (formerly known as the Expenditure and Food Survey (EFS) prior to 2008) is the UK's premier household expenditure survey, which captures information on the spending patterns and cost of living across the UK.	The ALS, which replaced the Active People Survey in November 2015, is a sport and recreation survey which measures physical activity levels of over 198,000 persons living across England.
Survey Design and Methodology	Annual repeated cross-sectional survey. Sample selected using multi-stage stratified random sampling with clustering. Household addresses with small user postcodes are randomly selected from the Royal Mail's postcode address file (PAF). Face-to face interviews (individual and household questionnaires administered) and 2-week self-reported expenditure diaries completed by all members of the household, aged 16 years and over. Simplified expenditure diaries are completed by children 7 to 15 years old.	Annual repeated cross-sectional survey. Multi-stage stratified random sample. Each year, approximately 198,250 persons are targeted for inclusion in the survey. Household addresses are randomly selected from the Royal Mail's postcode address file (PAF) and letters sent inviting up to two adults (16 years and over) per household to complete a questionnaire online or via post (for persons without internet access). Participants are rewarded with a £5 voucher from a range of retailers. During the adult survey, persons are asked if there are any 14-15-year olds in their household. Children aged 14-15 who are interested and receive parental consent to participate in the study are contacted and asked to complete a young person questionnaire.
Target population and level of geography covered	Families/households within the UK (England, Scotland, Northern Ireland and Wales). Data for England are available at the national and Government Office Region (GOR) level. Local authority level data can be made available upon request and approval by the UK Data Service.	Individuals 14 years and older living in England during the 2015-2016 and 2016-2017 periods. Data are available for the Government Office Region (GOR), County Sport Partnerships, Counties and Local Authority District level. The survey was designed to achieve a minimum annual sample size of 500 for each local authority, with the exception of the City of London and Isles of Scilly, in which the target sample size was 250.
Type of dietary assessment used	Household food expenditure data captured in the Family Food Module of the survey are used as a proxy measure for food consumption.	Single 24-hour screener/brief/shortened instrument (fruit and vegetable only) completed online or via post.
Primary users of diet-related data	Academics/Researchers and several Governmental Departments. The Family Food Module of the LCFS is primarily used by the Department for Environment Food and Rural Affairs (Defra) to monitor food consumption and to produce the annual Family Food Report (a report which provides estimates of nutrient content and statistics on household food purchases by food type).	Academics/Researchers, Local Authorities, Public Health England (PHE)
Data Accessibility/ Availability	Data accessible through the UK Data Service. Data currently available for the 2008- 2017/18 period	Data accessible through the UK Data Service. Data currently available for the 2015-2016 and 2016-2017 survey periods.
Types of variables captured	Socio-demographic information (age, sex, occupation, education), GOR, local authority level geography, data garnered from 2-week expenditure diary (expenditure on energy, bills, utilities and food).	Socio-demographic information (age, sex, employment) and health measures such as obesity and fruit and vegetable consumption over a 24-hour period.
Cost to access	Not applicable	Not applicable
Key features/ potential benefits	1. Nationally representative annual survey with relatively large sample size (approximately 5,000 households each year) 2. Two (2) week expenditure diaries (completed by each member of the household 16 years and over) detailing purchased quantities of food and drink are used to estimate food consumption in England. 3. Possible to make comparisons between low and high-income households.	1. Large sample size and a nationally representative sample of the English population. 2. Although focused on sport and recreation, the survey also includes data on fruit and vegetable consumption. 3. The availability of local authority data makes it possible to analyse dietary consumption below the regional (GOR) level.

Key considerations	1. Difficult to compare data prior to 2008 as a different survey methodology was used for the previous EFS. 2. Survey designed to capture household expenditure on food and quantities of food and drink purchased. The survey does not capture foods actually consumed by individuals.	1. Survey overestimates 5-A-Day consumption, (55% achievement) and does not compare well with NDNS and HSE, 2. The survey only captures self-reported fruit and vegetable consumption over a single 24-hour period. 3. Difficult to compare data prior to 2015 as a different survey methodology was used for the previous Active People Survey 4. Survey focused on physical activity and sport in England.
	National Diet and Nutrition Survey (NDNS)	**Health Survey for England (HSE)**
Survey Background	The NDNS was originally established in 1992 as a series of four separate cross-sectional surveys, capturing information on: children ages 1 ½ -4 ½ years (1992-1993), young people 4-18 years old (1997), adults 19-64 years old (2000-2001) and persons 65 years and over (1994-1995). In 2008, the new NDNS Rolling Programme (RP) was introduced as a nationally representative repeated cross-sectional survey which captures information on the type and quantity of foods and beverages consumed by 1,000 persons (500 adults and 500 children) annually in the UK.	The HSE is an annual survey used to monitor and assess changes in the overall health and lifestyle of persons living within England.
Survey Design and Methodology	Annual repeated cross-sectional survey. Multi-stage stratified random sample. Face-to face interviews conducted with respondents to capture food preparation, smoking and drinking habits. Self-completed 4-day food diaries are completed by persons 12 years and older and parents and/or carers are asked to complete food diaries for children 11 years and younger. Anthropometric measurements and blood and urine samples collected via nurse interview.	Annual repeated cross-sectional survey. Multi-stage stratified random sample. Face-to face interviews, self-completed questionnaires and a follow-up nurse visit carried out to collect anthropometric measurements and blood samples.
Target population and level of geography covered	Individuals 1 ½ years and older, residing in private households in the UK. Data for England are available at the national and Government Office Region (GOR) level.	Adults (defined as persons 16 years and older) and children (0-15 years old) living in private households in England. Data available at the national, Government Office Region (GOR) and Strategic Health Authorities level. Local authority level data only available upon request and approval by NatCen Social Research at a cost.
Type of dietary assessment used	Four (4) day food diary	Food frequency questionnaire (FFQ) used prior to 2009. Single 24-hour screener/brief/shortened instrument (fruit and vegetable only) used since 2009.
Primary users of diet-related data	Academics/Researchers, policymakers, UK Health Departments, Scientific Advisory Committee on Nutrition's (SACN), Food Standards Agency (FSA) and several Governmental Departments.	Academics/Researchers, policymakers, the Department of Health & Social Care, Public Health England (PHE), NHS England, other NHS bodies, Local Authorities, charities and voluntary organisations. Data used to track the national achievement of the 5-A-Day, fruit and vegetable target.
Data Accessibility/ Availability	Data for the NDNS RP are accessible through the UK Data Service. Data currently available for the 2008-2016/17 period (survey wave 1-9).	Data are accessible through the UK Data Service. Data currently available for the 1991-2017 period.
Types of variables captured	Socio-demographic information (age, sex, occupation, education), GOR and all foods and beverages consumed over a 4-day period.	Socio-demographic information (age, sex, occupation, education), GOR, general health, height and weight measurements and fruit and vegetable consumption.
Cost to access	Not applicable	No cost to access GOR level data but lower level geography (e.g. local authority level) can be accessed at a minimum cost of £1,000.
Key features/ potential benefits	1. Availability of annual food consumption data at the national level and 2. Detailed information available on all foods and beverages actually consumed by individuals over a 4-day period using the food diary method.	1. Nationally representative annual survey with large sample size of approximately 10,000 individuals (8,000 adults and 2,000 children). 2. Data captured could be used to explore relationships between diet (specifically fruit and vegetable consumption), obesity and associated chronic diseases.
Key considerations	1. Relatively small annual sample size compared to larger cohort studies which employ methods which are less tedious than the food diary method. 2. Difficult to compare data prior to 2008 with NDNS RP data, as a different survey methodology was used previously. This makes it difficult for comparisons to be made across the survey waves and for changes in diet to be assessed over time.	1. Significant changes (e.g. the complete omission of the fruit and vegetable module in the 2012 survey wave) have been made to the type of diet questions asked, which makes it difficult for comparisons to be made across the survey waves and for changes in diet to be assessed over time.

	Food and You Survey
Survey Background	Food and You is a random probability survey commissioned by the Food Standards Agency (FSA) every two years. The survey captures information on public attitudes and self-reported knowledge as it relates to food safety, production and other food-related issues.
Survey Design and Methodology	Bi-annual repeated cross-sectional survey. Multi-stage stratified random sample. Face-to face interviews conducted with adults, defined as persons aged 16 years and over.
Target population and level of geography covered	Adults (16 and over) residing in private households the UK. Data for England are available at the national and Government Office Region (GOR) level.
Type of dietary assessment used	Food frequency questionnaire (FFQ) conducted at each wave of the survey
Primary users of diet-related data	Academics/Researchers, policymakers and several Governmental Departments, particularly the Food Standards Agency (FSA)
Data Accessibility/ Availability	Data are accessible through the UK Data Service. Data currently available for the five survey waves completed to date: 2010, 2012, 2014, 2016 and 2018.
Types of variables captured	Socio-demographic information (age, sex, occupation, education, household income), GOR, frequency of consumption of foods such as beef, poultry, burgers, ready meals, diary, fruits and vegetables.
Cost to access	Not applicable
Key features/potential benefits	1. Nationally representative survey with sample size of about 3,000-3,500 individuals every 2 years, 2. Besides data collected via FFQs, the survey also captures respondents' knowledge of current dietary recommendations and perceptions of what constitutes a healthy and balanced diet.
Key considerations	1. Changes made to diet-related questions asked over the years, makes it difficult for comparisons to be made across the survey waves and for changes in diet to be assessed over time.

Appendix A

	Longitudinal surveys	
	Southampton Women's Survey (SWS)	**Born in Bradford (BIB)**
Survey Background	The SWS was established between 1998 and 2002 with the primary aim of measuring non-pregnant women aged 20-34 years living in Southampton (England) and to follow-up members of the cohort who subsequently became pregnant. The study's major aim was to examine the effect of diet and lifestyle factors on the health of mothers and their children throughout the lifecourse.	BIB is a study which tracks the health of over 13,500 children (and their parents) born at the Bradford Royal Infirmary between March 2007 and December 2008. The study tracks the health of these children from pregnancy throughout childhood and into adulthood.
Survey Design and Methodology	Longitudinal Birth Cohort study. Pre-pregnancy home visits were made to 12,583 non-pregnant women (who were 20-34 years old during the 1998-2002 period) who resided in Southampton, England and surrounding areas. Pre-pregnancy food diaries were completed by participants and face-to-face interviews and blood samples were taken by a research nurse. Follow-up nurse visits were made to 3,158 women who became pregnant and delivered a live born child; blood samples taken, and follow-up interviews conducted. Participants were asked to keep a food diary during early and late pregnancy. Follow-up surveys were conducted when children were 6 and 12 months and 3, 6-7, 8-9 and 11-13 years old.	Longitudinal Birth Cohort study. Women who planned to be give birth during the 2007-2011 period were recruited and baseline data on socio-economic position, ethnicity and family trees, diet, physical and mental health were collected from 12,453 women at 26-28 weeks of pregnancy. Baseline data were also collected from 3,448 partners of recruited mothers. Follow-up self-administered questionnaires were completed by partners at 6 and 12 months. Follow-up home visits were made with 2 sub-groups within the cohort when children were 6, 12, 18 months and 2, 3 and 4 years old to collect information on growth trajectories, risk factors for childhood obesity and exposures to asthma and atopy. Follow-up waves are heavily dependent on the level of funding available.
Target population and level of geography covered	12,583 non-pregnant women aged 20-34 years during the 1998-2002 period, living in Southampton (South East of England) and surrounding areas; 3,158 women who became pregnant and delivered a live born child subsequent to recruitment and their children.	Pregnant women (26-28 weeks) who delivered babies at the Bradford (North England) Royal Infirmary, fathers of the children and the children born to recruited mothers. Geographical area captured: Bradford (North of England)
Type of dietary assessment used	Interviewer administered FFQ and 24-hour recall conducted at each survey wave, food diaries completed by mothers at pre-pregnancy, early pregnancy and when children were 3 years old and 24-hour diet recalls administered when children were 6 months old.	Food frequency questionnaire (FFQ) at each wave of the survey.
Primary users of diet-related data	Academics/Researchers, Local Authority.	Academics/Researchers, Local Authority, National Health Service (NHS).
Data Accessibility/ Availability	Data accessible through the MRC Lifecourse Epidemiology Unit, University of Southampton. Data available for women before pregnancy (1998-2002) and during early and late pregnancy. Data for children are available for 6 and 12 months, 3, 6-7, 8-9 and 11-13 years old.	Data (and details regarding survey data currently available) accessible through the Bradford Institute for Health Research.
Types of variables captured	Socio-demographic information (age, sex, occupation, employment, education), general diet, dietary changes and a 100-point FFQ asking the frequency of consumption in the last 3 months of fruits, vegetables, potatoes, rice, soft drinks, dairy, bread and a host of other foods across the various food groups.	Socio-demographic information (age, sex, occupation, employment, education) and a more than 100-point FFQ asking the frequency of consumption in the last 2-3 months of fruits, vegetables and a host of other foods across the various food groups.
Cost to access	Not applicable	Not applicable
Key features/ potential benefits	1. Food consumption data available for a wide variety of foods. 2. Cohort study data can be used to track changes over time. 3. Availability of pre- and post-pregnancy data.	1. Food consumption data available for a wide variety of foods 2. Cohort study data can be used to track changes over time. 3. Bradford has a large ethnic community and so the study captures ethnic minority groups which are usually underrepresented
Key considerations	1. Study not representative of English population; only focuses on Southampton (South of England). 2.The study only focuses on women and their children over time. 3. Complete data on children are not available for the entire cohort at each age of follow-up.	1.Study not representative of English population; only focuses on Bradford (North of England) 2. Changes made to diet-related questions across the survey waves may make it difficult to make comparisons over time. 3. Follow-up waves are heavily dependent on the level of funding available.

	Understanding Society	British Cohort Study 1970 (BCS70)
Survey Background	Understanding Society is an annual large-scale, multi-topic longitudinal cohort study established to understand social and economic changes in the UK at the individual and household level.	The BCS70 is a large national longitudinal birth cohort study which tracks over 17,000 persons born in England, Scotland and Wales in a single week in 1970. The study has gathered information related to the health, social, economic and educational development of participants.
Survey Design and Methodology	Annual Longitudinal/panel/cohort study. Multi-stage stratified random sample. The first wave was conducted in 2009 when over 40,000 households were selected. Since then, follow-up interviews have been conducted with the same individuals every 12 months. At each survey wave, one member of the household is asked to complete a household questionnaire and each person 16 years and older is interviewed and asked to complete a separate (self-completed) questionnaire. Members of the household aged 10-15 years (young people) are also asked to complete a separate (self-completed) paper or web-based/online questionnaire. Web-based surveys were introduced in wave 7 (2016) of the survey.	Longitudinal Birth Cohort study. All children born in England, Scotland and Wales in 1970 were recruited and eight follow-up surveys have been conducted to date. Follow-up interviews were done when children were 5, 10, 16, 26, 30, 34, and 42 years of age (in 2012). Although data are not currently available, a follow-up survey was conducted at age 46 (in 2016) and information is currently being processed. In the 2004 study (age 34) cohort members were given a basic skills (numeracy and literacy) assessment test and a self-completion questionnaire to complete.
Target population and level of geography covered	Individuals living within over 40,000 households in the UK. Data for England are available at the national and Government Office Region (GOR) level. Local authority level, Westminster Parliamentary Constituencies, Local Education Authorities and Travel to Work Areas are available upon request and approval by the UK Data Service under its Special License Agreement.	Children born in England, Scotland and Wales in a single week in 1970.
Type of dietary assessment used	Short food frequency screener/brief instrument which primarily captured fruit and vegetable consumption.	4-day food diary and a 24-hour diet recall included in 1986 wave of survey. Online diet diary also included in the 2016 wave, when respondents were 46 years old.
Primary users of diet-related data	Academics/Researchers.	Academics/Researchers.
Data Accessibility/ Availability	Data accessible through the UK Data Service. Data currently available for the 2009-2018 period (survey wave 1-9)	Data accessible through the UK Data Service. Data currently available for the 1975-2016-18 survey period.
Types of variables captured	Socio-demographic information (age, sex, education, family, social life), self-reported health, type of milk, bread usually consumed, daily and weekly consumption of fruits and vegetables.	Socio-demographic information (age, sex, occupation, education). Consumption of fruits, vegetables, meat, dairy, soup, potatoes, biscuits, crisps, fizzy drinks, sweets and ice-cream consumed over a 24-hour period. All foods consumed over a 4-day period in 1986 (paper-based food diary) and in 2016 (online food diary) when respondents were 46 years old.
Cost to access	Not applicable	Not applicable
Key features/ potential benefits	1. Large sample size, nationally representative and conducted annually. 2. Cohort study data can be used to track changes over time.	1. Large sample size and nationally representative 2. Cohort study data can be used to track changes over time. 3. Detailed information on all foods consumed by participants over several days were captured in food diaries conducted in the 1986 and 2016 wave of the survey.
Key considerations	1. Very few diet-related questions included in the study (fruit and vegetable consumption, dairy, bread). Questions posed in the main questionnaire primarily focused on the type of bread and milk consumed and portions of fruits and vegetables consumed in a typical week. 2. Differences in the number and types of diet-related questions asked across survey waves could make it difficult for comparisons to be made over time.	1. Food diary data for the 1986 and 2016 wave are being cleaned and the expected date of release is undetermined 2. Changes made to diet-related questions across survey waves could make it difficult for comparisons to be made over time.

Appendix A

	Avon Longitudinal Study of Parents and Children (ALSPAC)	UK Women's Cohort Study (UKWCS)
Survey Background	ALSPAC also known as the Children of the 90s Study, is a study which tracks the health and well-being of 14,400 families living within the Bristol area.	The UKWCS is a large-scale cohort study which explores the relationship between diet (including foods, nutrients and supplements) and health outcomes such as cancer, cardiovascular disease and obesity amongst over 35,000 middle aged women in the UK.
Survey Design and Methodology	Longitudinal Birth Cohort study. Study posters were disseminated, and local community midwives discussed the study with women with expected deliveries between April 1991 and December 1992. Persons who contacted the study team were included in the study. Baseline data were captured during pregnancy and follow-up assessments carried out when children were 4 weeks to 24 years of age. Self-completed postal questionnaires were completed by mothers, children and teachers (of children) and clinical assessment visits were carried out at different stages of the study.	Longitudinal Cohort study. Direct mail questionnaires were sent by the World Cancer Research Fund to persons, particularly women, living in England, Scotland and Wales, listed on direct mailing lists. Female survey responders aged 35-69, who self-identified as vegetarian or non-red meat eaters were included in the study. Baseline data were collected during the 1995-1998 period and follow up (known as phase 2 of the study) was done during the 1999-2002 period. Several sub-studies have been carried out over the years. For instance, an iron status sub-study in 2000-2002, a snacking study in 2006 and a pilot study to test a web-based 24-hour dietary assessment tool in 2014.
Target population and level of geography covered	All women pregnant during 1990-1992, who resided in Bristol/Avon Health Authority and surrounding areas, their partners and all children born out of these pregnancies. Geographical area captured: Bristol and surrounding areas (South West of England)	Middle aged women (aged 35-69 at recruitment) living in England, Scotland and Wales, who self-reported as being vegetarian or non-red meat eaters. Geographical area captured: England, Scotland and Wales and English regions. Regions included in the study's data set can be easily converted to Government Office Region (GOR) categories
Type of dietary assessment used	Food frequency questionnaires (FFQs). Food diaries were completed by parents when children were 7, 10 and 13 years of age.	Food frequency questionnaire (FFQ); a 4-day food diary (completed during the follow up study in 1999-2002) and a 24-hour web-based diet recall assessment pilot in 2014.
Primary users of diet-related data	Academics/Researchers.	Academics/Researchers.
Data Accessibility/ Availability	Data (and details regarding survey data currently available) accessible through the University of Bristol	Data (and details regarding survey data currently available) accessible through the Consumer Data Research Centre
Types of variables captured	Socio-demographic information (age, sex, occupation, employment, education), consumption of fruits, vegetables and a host of other foods which vary across the survey waves.	Socio-demographic information (age, sex, occupation, education), food consumption data captured from FFQs and food diaries conducted at different survey waves.
Cost to access	Minimum cost of £2,715 to access	Not applicable
Key features/ potential benefits	1. Large sample size. 2. Cohort study data can be used to track changes over time.	1. Large sample size. 2. Cohort study data could be used to track changes over time. 3. Availability of food diary data provides detailed information on all foods consumed by participants.

Key considerations	1. Costly to access. 2. Study not representative of English population; only focuses on Bristol and surrounding areas (South West of England). 3. Changes made to diet-related questions across survey waves could make it difficult for comparisons to be made over time	1. Food diaries completed in phase 2 of the study (1999-2002) and diaries completed during the 2014 online pilot study were still being processed at the time of this assessment. As such, these data are not available, and the date of release is undetermined 2. Study not representative of the English population. Participants were mostly vegetarian, middle aged, middle class, white women who volunteered to be a part of the study during the late 1990s 3. Changes made to diet-related questions across the survey waves may make it difficult to make comparisons over time.
	Whitehall II Study	**Millennium Cohort Study (MCS)**
Survey Background	The Whitehall II study is a cohort study conducted to assess the causes of social inequalities in health in England.	The MCS is a large national longitudinal birth cohort study which tracks 19,000 children born in the UK during 2000-2001, from childhood into adulthood.
Survey Design and Methodology	Longitudinal Cohort study. A cohort of 10,308 middle-aged persons (3,413 females and 6,895 males, aged 35-55 years old) who worked in the London offices of 20 Whitehall departments in 1985–1988 were included in the study. During the 2015-2016 period, research clinics were established in London, Bristol, Birmingham and Liverpool to allow persons (especially retired persons) now living within these and surrounding areas to be a part of the study and reduce the level of attrition. Members of the cohort were invited to attend a clinic research screening every 5 years and a postal survey sent to participants between clinic phases. Overall, data has been collected over 12 waves, from 1985-1988 to 2015-2016	Longitudinal Birth Cohort study. Multi-stage stratified random sample. The sample consisted of all children born (live births) over 12 months (from 1 September 2000 in England and Wales and for 59 weeks from 22 November 2000 in Scotland and Northern Ireland). Six surveys have been conducted to date, capturing information when children were 9 months and 3, 5, 7, 11 and 14 years of age (in 2015). Although data are currently unavailable, the 7th wave was conducted in 2018 captures children at age 18. A combination of data collection methods has been used. These include face-to-face interviews, self-completed questionnaires; psychological measurements, observation; time use diaries and physical measurements.
Target population and level of geography covered	Middle-aged persons who worked in the London offices of 20 Whitehall departments in 1985–1988.	Children born in the UK (England, Scotland, Northern Ireland and Wales) during 2000-2001. Data for England are available at the national and Government Office Region (GOR).
Type of dietary assessment used	Food frequency questionnaire (FFQ)	Food frequency questionnaire (FFQ)
Primary users of diet-related data	Academics/Researchers.	Academics/Researchers.
Data Accessibility/ Availability	Data accessible through the University College London. Data available for waves 1-12 (1986-2016)	Data accessible through the UK Data Service. Data currently available for the 2001-2015 survey period.
Types of variables captured	Socio-demographic information (age, sex, occupation, employment, retirement, education, income), self-reported health and frequency of consumption in the last 12 months of fruits, vegetables, meat, fish, soups, sauces, spreads, eggs, dairy products, fats, bread, pasta, potato, rice, sweets and snacks were consumed.	Socio-demographic information (age, sex, occupation, employment, education of parents), consumption of fruits and vegetables and other foods such as bread, milk, sugary drinks and fast foods.
Cost to access	Not applicable	Not applicable

Key features/ potential benefits	1. Food consumption data available for a wide variety of foods. 2. Fairly large sample size across the 12 waves (10,308 in 1985-1988 to 5,632 in 2015-2016). 3. Cohort study data can be used to track changes over time.	1. Large sample size and nationally representative. 2. Cohort study data can be used to track changes over time. 3. Children were asked to state their consumption of fruits and vegetables and other foods such as bread, sugary drinks and fast food at age 14.
Key considerations	1. Study not representative of English population. Study focused on middle-aged civil servants. 2. Changes made to diet-related questions across the survey waves may make it difficult to make comparisons over time. 3. Based on the current age-group of participants, the study is now primarily focused on issues surrounding population ageing.	1. Cohort members are still very young, which currently limits the assessment of diet by age/over lifecourse. 2. Changes made to diet-related questions across survey waves could make it difficult for comparisons to be made over time.
	European Prospective Investigation into Cancer and Nutrition (EPIC Norfolk/Oxford)	**UK Biobank**
Survey Background	EPIC is a large cohort study which aims to examine diet as a risk factor for cancer and other chronic diseases amongst over 80,000 middle aged persons in the UK.	The UK Biobank is a large-scale longitudinal study which follows 500,000 middle-aged persons across the UK to investigate the association between diet and a range of diseases such as cancer, heart disease, stroke, diabetes and dementia.
Survey Design and Methodology	Longitudinal Cohort study. EPIC Oxford: 65,000 persons from the general population were recruited between 1993 and 1999 via EPIC nurses in GP practices in Greater Manchester, Oxfordshire and Buckinghamshire, England. Postal questionnaires were also sent to members of the Vegetarian Society of the UK and Vegan Society, and study information distributed through health magazines and shops, to capture persons located across the entire UK. Follow-up surveys were conducted 5, 10 and 15 years later. EPIC Norfolk: Invitations were sent to all 40-79-year olds on collaborating GP listings. Over 30,000 persons within Norwich and surrounding areas (East of England) were recruited over the 1993-1997 period. Participants were followed up at 18 months, 3, 13 and 20 years after recruitment. A combination of data collection methods was used for both studies (nurse interview to collect anthropometric measurements and blood samples, self-completed questionnaires (on physical activity) and record linkages via hospital diagnoses, death certification and cancer registration	Longitudinal Cohort study. Population-based registers such as those held by the National Health Service (NHS) were used as a sampling frame to identify persons living within proximity to study assessment centres. Each assessment centre aimed to recruit as many persons within the target population. Baseline data (for the 2006-2010 period) were collected at assessment centres, where self-reported baseline questionnaires were used to collect health and lifestyle-related data and interviews conducted to collect physical measurements and biological samples. A follow up survey was conducted in 2011-2012.
Target population and level of geography covered	EPIC Oxford: Men and women 35 years and over (at recruitment) who lived in Greater Manchester, Oxfordshire and Buckinghamshire in England and vegetarians /vegans located across the UK. EPIC Norfolk: Men and women aged 40-79 (at recruitment) who lived in Norwich and surrounding towns and rural areas.	Middle-aged males and females (persons aged 40-69 during the 2006-2010 period) who lived within a 10-mile radius of 35 study centres strategically located across England, Wales and Scotland.
Type of dietary assessment used	Food frequency questionnaire (FFQ) and a 7-day food diary (completed at recruitment and at the 2nd wave of the study)	Food frequency questionnaire (FFQ) with foods related to increased cancer risk conducted at baseline. Web-based 24-hour recall repeated on four occasions over a 16-month period.
Primary users of diet-related data	Academics/Researchers.	Academics/Researchers.
Data Accessibility/ Availability	The EPIC Oxford study is accessible through the University of Oxford and EPIC Norfolk through the University of Cambridge. Details on current data availability accessible from both institutions.	Data accessible through the UK Biobank. Data available (at the time of assessment) for the 2006-2010 (baseline) and 2011-2012 period.

Types of variables captured	Socio-demographic information (age, sex, occupation, education), food consumption data captured from FFQs and food diaries conducted at different survey waves.	Socio-demographic information (age, sex, employment) and fruits, vegetables, meat, dairy and a host of other foods consumed (total of over 200 foods) over a 24-hour period.
Cost to access	Not applicable	Minimum £2,000 to cover application and data access cost. Possibly reduced cost of £500 for research students (subject to review and approval).
Key features/ potential benefits	1. Large sample size. 2. Cohort study data can be used to track changes over time. 3. Availability of food diary data (at recruitment and wave 2) which provides detailed information on foods consumed by participants.	1. Large sample size. 2. Cohort study data can be used to track changes over time. 3. Detailed information on foods consumed by participants over repeated days (repeated 24-hour diet recalls).
Key considerations	1. Study not representative of English population. Focused on middle-aged persons living in Norwich, Greater Manchester, Oxfordshire and Buckinghamshire who were in some instances selected via purposive sampling. 2. 50% of participants were vegetarians/vegans. 3. Changes made to diet-related questions across the survey waves may make it difficult to make comparisons over time	1. Study not representative of English population. Focused on middle-aged persons from less-deprived areas (based on the target population). 2. The baseline survey captured some aspects of diet consumption but was not as comprehensive as the 2011-2012 survey wave. 3. Differences in the number and types of diet-related questions asked across survey waves could make it difficult for comparisons to be made over time. 4. Relatively high cost to access data. 5. Lengthy application process and possible lag time for approval.
	British Regional Heart Study (BRHS)	**British Women's Heart and Health Study (BWHHS)**
Survey Background	The BRHS is a cohort study, established in 1978-1980, which explores the factors associated with heart disease, hypertension and stoke amongst 7,735 middle-aged men (40-59 years at recruitment) recruited from General Practices (GPs) in 24 towns in England, Scotland and Wales.	The BWHHS is a cohort study, established in 1999 as a complement to the BRHS. The study follows 4,286 women, aged 60 years and over (at recruitment) from 24 General Practices (GPs), in 23 towns in England, Scotland and Wales
Survey Design and Methodology	Longitudinal Cohort study. Almost 8,000 middle-aged men who were selected at random from one GP in each of the 24 towns, were examined over the 1978-1980 period. Self-completed health and lifestyle questionnaires and clinical assessments/examinations (inclusive of anthropometric measurements) completed at baseline (1978-80). Follow-up self-completed questionnaires were completed in 1985,1992,1996,1998-2000, 2003,2005,2007,2010-12, 2014,2015, 2016,2017 and 2018. A review of GP records (including all hospital and clinic correspondence) was also conducted bi-annually. A clinical re-examination was done in the 1998-2000 wave. Participants were also given a self-completed activity survey questionnaire and asked to wear an activity monitor and keep a 3-day activity diary in 2010, 2011, 2012, 2013, 2015 and 2017.	Longitudinal Cohort study. Almost 8,000 middle-aged women were randomly selected from 24 GPs, in 23 towns from 1999-2000. Self-completed health and lifestyle questionnaires, and nurse administered interviews and medical examinations were completed at baseline (1999-2000). A review of GP records (including all hospital and clinic correspondence) was completed at baseline and in 2002, 2004, 2007, 2011-12 and 2016-17. Self-completed health and lifestyle questionnaires were completed in 2003, 2007 and 2010-2011. Participants were also given a self-completed activity survey questionnaire and asked to wear an activity monitor/belt and keep a 3-day activity diary in 2010-2011.
Target population and level of geography covered	Middle-aged men aged 40-59 years (at recruitment) who resided in 24 towns across England, Scotland and Wales.	Middle-aged women aged 60 years and over (at recruitment) from 23 towns across England, Scotland and Wales.
Type of dietary assessment used	Food frequency questionnaire (FFQ)	Food frequency questionnaire (FFQ)
Primary users of diet-related data	Academics/Researchers.	Academics/Researchers.
Data Accessibility/ Availability	Data accessible through University College London	Data accessible through University College London

Types of variables captured	Socio-demographic information (age, sex), health status, consumption of fruits and vegetables, fish, meat, bread and a host of other foods which vary across the survey waves.	Socio-demographic information (age, sex), consumption of fruits, vegetables, cheese, milk, red meat and other foods which vary across the survey waves.
Cost to access	Unknown (Information inaccessible at time of assessment).	Unknown (Information inaccessible at time of assessment).
Key features/ potential benefits	1. Cohort study data can be used to track changes over time. 2. Data captured could be used to explore relationships between diet, cardiovascular disease and associated chronic diseases.	1. Cohort study data can be used to track changes over time. 2. Data captured could be used to explore relationships between diet, cardiovascular disease and associated chronic diseases.
Key considerations	1. Study not representative English population. Study only captures middle-aged men from 24 towns across sections of Scotland, England and Wales. 2. Differences in the number and types of diet-related questions asked across survey waves could make it difficult for comparisons to be made over time. 3. Based on the current age-group of participants, the study is now primarily focused on issues surrounding population ageing.	1. Study not representative English population. Study only captures middle-aged women from 23 towns across sections of Scotland, England and Wales. 2. Differences in the number and types of diet-related questions asked across survey waves could make it difficult for comparisons to be made over time.

Appendix B Research Question 1 supplemental results

B.1 Number and percentage of observed and missing cases for the variables of age, sex, ethnicity, GOR, NS-SEC and highest level of education, HSE, 2008-16

Variable	Observed cases Number (%)	Missing cases Number (%)	Total
Age			
16-24	4,494 (72)	1,724 (28)	6,218
25-29	2,950 (71)	1,201 (29)	4,151
30-39	7,188 (73)	2,643 (27)	9,831
40-49	8,358 (76)	2,648 (24)	11,006
50-64	22,990 (74)	8,216 (26)	31,206
Sex			
Male	15,466 (75)	5,044 (25)	20,510
Female	19,110 (73)	6,733 (26)	25,843
Ethnicity			
White	30,470 (75)	9,916 (25)	40,386
Mixed	503 (70)	219 (30)	722
Asian	2,310 (70)	970 (30)	3,280
Black	994 (70)	417 (30)	1,411
Other	276 (70)	119 (30)	395
Missing	23 (14)	136 (86)	159
Government Office Region (GOR)			
North East	2,553 (72)	1,000 (28)	3,553
North West	4,687 (73)	1,772 (27)	6,459

Variable	Observed cases Number (%)	Missing cases Number (%)	Total
Yorkshire and the Humber	3,391 (73)	1,259 (27)	4,650
East Midlands	3,326 (78)	952 (22)	4,278
West Midlands	3,419 (75)	1,133 (25)	4,552
East of England	3,842 (74)	1,353 (26)	5,195
London	4,125 (71)	1,724 (29)	5,849
South East	5,661 (77)	1,725 (23)	7,386
South West	3,572 (81)	859 (19)	4,431
Occupation/NS-SEC			
Never worked/Long-term unemployed	715 (63)	419 (37)	1,134
Routine, manual and other occupations	13,534 (74)	4,701 (26)	18,235
Small employers and own account workers	2,921 (75)	964 (25)	3,885
Intermediate occupations	12,940 (76)	4,131 (24)	17,071
Higher managerial and professional occupations	4,139 (77)	1,209 (23)	5,348
Missing	327 (48)	353 (52)	680
Highest level of education			
No qualifications	4,705 (71)	1,907 (29)	6,612
GCSE or equivalent qualifications	8,925 (75)	3,014 (25)	11,939
GCE A level or equivalent	5,157 (75)	1,742 (25)	6,899
Higher education, below degree level	3,766 (79)	1,031 (21)	4,797
Degree or equivalent	9,011 (75)	2,984 (25)	11,995
Still in full-time education/Other foreign qualification/Other	2,982 (76)	948 (24)	3,930

Variable	Observed cases Number (%)	Missing cases Number (%)	Total
Missing	30 (17)	151 (83)	181

B.2 Pre-imputation fully adjusted single-level regression model for 5-A-Day achievement for persons aged 16-64 years old in England, NDNS, 2008-16 (sensitivity analysis)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Intercept	-2.082** (0.506)	——	——
Age (Ref Cat: 16-24 years)			
25-29	——	2.52**	(1.63, 3.90)
30-39	——	2.67**	(1.82, 3.93)
40-49	——	3.14**	(2.12, 4.65)
50-64	——	5.70**	(3.86, 8.43)
Sex (Ref Cat: Female)			
Male	——	0.92	(0.77, 1.09)
Ethnicity (Ref Cat: White)			
All other ethnic groups	——	1.36*	(1.04, 1.77)
Marital Status (Ref Cat: Single Never Married)			
Married/In a legally recognised civil partnership	——	1.10	(0.88, 1.38)
Separated/divorced	——	0.80	(0.59, 1.07)
Widowed/other/formerly in a legally recognised civil partnership	——	1.39	(0.76, 2.53)
Highest level of education (Ref Cat: No qualifications)			
GCSE or equivalent qualifications	——	1.00	(0.71, 1.37)
GCE, A level or equivalent	——	1.13	(0.80, 1.61)
Higher education, below degree level	——	1.31	(0.90, 1.92)
Degree or equivalent	——	1.97**	(1.42, 2.72)
Still in full-time education/Other foreign qualification/Other	——	1.34	(0.88, 2.02)
Missing	——	1.03	(0.50, 2.10)
Equivalised household income (Ref Cat: Under £20,000)			
£20,000-£29,999	——	1.22	(0.94, 1.58)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
£30,000-£39,999	—	1.39*	(1.04, 1.86)
£40,000-£49,999	—	1.26	(0.90, 1.76)
£50,000-£99,999	—	1.44*	(1.07, 1.94)
£100,000 or more	—	2.68**	(1.63, 4.41)
Missing	—	1.19	(0.88, 1.61)
NS-SEC8 Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	—	0.87	(0.46, 1.67)
Small employers and own account workers	—	0.93	(0.47, 1.83)
Intermediate occupations	—	0.99	(0.51, 1.90)
Higher managerial and professional occupations	—	0.83	(0.42, 1.64)
Missing	—	1.29	(0.30, 5.46)
Body Mass Index: BMI (Ref Cat: Underweight)			
Normal weight	—	1.03	(0.57, 1.84)
Overweight	—	0.70	(0.39, 1.27)
Obese	—	0.64	(0.35, 1.17)
Missing	—	0.63	(0.32, 1.25)
Government Office Region (GOR) (Ref Cat: North East)			
North West	—	0.73	(0.47, 1.13)
Yorkshire and the Humber	—	0.83	(0.53, 1.32)
East Midlands	—	0.99	(0.63, 1.56)
West Midlands	—	0.83	0.53, 1.30)
East of England	—	1.04	(0.67, 1.60)
London	—	1.12	(0.72, 1.75)
South East	—	0.88	(0.58, 1.34)
South West	—	1.36	(0.87, 2.12)
Index of Multiple Deprivation (IMD) quintiles (Ref Cat: Least deprived)			
2	—	0.98	(0.76, 1.25)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
3	——	0.80	(0.62, 1.04)
4	——	0.64**	(0.49, 0.83)
Most deprived areas	——	0.80	(0.60, 1.05)

* p-value <0.05, ** p-value<0.01, SE= standard error

B.3 Post-imputation (25 imputations) fully adjusted single-level regression model for 5-A-Day achievement for persons aged 16-64 years old in England, NDNS, 2008-16 (sensitivity analysis)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Intercept	-2.085**(0.495)	—	—
Age (Ref Cat: 16-24 years)			
25-29	—	2.49**	(1.61, 3.84)
30-39	—	2.64**	(1.80, 3.87)
40-49	—	3.09**	(2.09, 4.57)
50-64	—	5.63**	(3.81, 8.30)
Sex (Ref Cat: Female)			
Male	—	0.92	(0.78, 1.09)
Ethnicity (Ref Cat: White)			
All other ethnic groups	—	1.36*	(1.05, 1.78)
Marital Status (Ref Cat: Single Never Married)			
Married/In a legally recognised civil partnership	—	1.10	(0.88, 1.37)
Separated/divorced	—	0.80	(0.60, 1.07)
Widowed/other/formerly in a legally recognised civil partnership	—	1.39	(0.76, 2.53)
Highest level of education (Ref Cat: No qualifications)			
GCSE or equivalent qualifications	—	0.99	(0.72, 1.37)
GCE, A level or equivalent	—	1.12	(0.79, 1.58)
Higher education, below degree level	—	1.33	(0.92, 1.94)
Degree or equivalent	—	1.94**	(1.40, 2.68)
Still in full-time education/Other foreign qualification/Other	—	1.29	(0.85, 1.95)
Equivalised household income (Ref Cat: Under £20,000)			
£20,000-£29,999	—	1.20	(0.94, 1.53)

Appendix B

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
£30,000-£39,999	——	1.36*	(1.03, 1.78)
£40,000-£49,999	——	1.33	(0.97, 1.76)
£50,000-£99,999	——	1.51**	(1.07, 1.94)
£100,000 or more	——	2.76**	(1.63, 4.41)
NS-SEC8 Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	——	0.89	(0.48, 1.63)
Small employers and own account workers	——	0.92	(0.48, 1.76)
Intermediate occupations	——	0.98	(0.53, 1.82)
Higher managerial and professional occupations	——	0.82	(0.43, 1.57)
Body Mass Index: BMI (Ref Cat: Underweight)			
Normal weight	——	1.02	(0.57, 1.82)
Overweight	——	0.68	(0.38, 1.24)
Obese	——	0.64	(0.35, 1.16)
Government Office Region (GOR) (Ref cat: North East)			
North West	——	0.74	(0.48, 1.14)
Yorkshire and the Humber	——	0.85	(0.53, 1.34)
East Midlands	——	1.01	(0.64, 1.58)
West Midlands	——	0.84	(0.54, 1.31)
East of England	——	1.05	(0.68, 1.62)
London	——	1.13	(0.73, 1.76)
South East	——	0.89	(0.58, 1.35)
South West	——	1.38	(0.88, 2.15)
Index of Multiple Deprivation (IMD) quintiles (Ref cat: Least deprived)			
2	——	0.98	(0.76, 1.25)
3	——	0.80	(0.62, 1.05)
4	——	0.64**	(0.49, 0.84)
Most deprived areas	——	0.80	(0.61, 1.05)

* p-value <0.05, ** p-value<0.01, SE= standard error

B.4 Number and percentage (unweighted and weighted) of persons aged 16-64 years old in England sampled in the NDNS by socio-demographic and geographical characteristics, 2008-16

Characteristic	Number	Unweighted (%)	Weighted (%)
Age			
16-24	806	25	18
25-29	242	7	11
30-39	634	19	21
40-49	700	21	23
50-64	904	28	28
Sex			
Female	1,859	57	50
Male	1,427	43	50
Marital status			
Single/Never Married	1,504	46	39
Married/Legally recognised civil partnership	1,285	39	49
Separated/Divorced	443	14	10
Widowed	54	2	1
Ethnicity			
White	2,854	87	85
All other ethnic groups	432	13	15
Equivalised household income			
Under £20,000	987	30	29
£20,000-£29,999	815	25	24
£30,000 - £39,999	536	16	17
£40,000 - £49,999	357	11	12
£50,000 - £99,999	501	15	16
£100,000 or more	90	3	3
Highest level of education			
No qualifications	358	11	11
GCSE or equivalent qualifications	716	22	22
GCE, A level or equivalent	460	14	16
Higher education, below degree level	307	9	11
Degree or equivalent	830	25	29

Characteristic	Number	Unweighted (%)	Weighted (%)
Still in full-time education/Other foreign qualification/Other	615	19	12
Occupation/NS-SEC			
Never worked/Long-term unemployed	79	2	2
Routine, manual and other occupations	1,094	33	33
Small employers and own account workers	365	11	11
Intermediate occupations	1,174	36	35
Higher managerial and professional occupations	574	17	18
Body Mass Index (BMI)			
Underweight	90	3	2
Normal weight	1,292	39	37
Overweight	1,184	36	38
Obese/morbidly obese	720	22	23
Government Office Region (GOR)			
North East	183	6	4
North West	458	14	13
Yorkshire and the Humber	314	10	10
East Midlands	312	10	9
West Midlands	397	12	11
East of England	392	12	11
London	395	12	17
South East	528	16	16
South West	307	9	10
Index of Multiple Deprivation (IMD) quintiles			
Least deprived area	685	21	20
2	639	19	20
3	604	18	18
4	675	21	21
Most deprived areas	683	21	21
Survey year (Wave)			
2008/09 (Wave 1)	447	14	15
2009/10 (Wave 2)	385	12	14
2010/11 (Wave 3)	390	12	14
2011/12 (Wave 4)	448	14	14
2012/13 (Wave 5)	417	13	12
2013/14 (Wave 6)	353	11	10
2014/15 (Wave 7)	427	13	11
2015/16 (Wave 8)	419	13	11

Characteristic	Number	Unweighted (%)	Weighted (%)
TOTAL	**3,286**	**100**	**100**

B.5 Number and percentage (unweighted and weighted) of persons aged 16-64 years old in England sampled in the HSE by socio-demographic and geographical characteristics, 2008-16

Characteristic	Number (n)	Unweighted (%)	Weighted (%)
Age			
16-24	4,386	13	17
25-29	2,893	9	10
30-39	7,118	21	21
40-49	8,286	24	23
50-64	11,512	33	30
Sex			
Female	18,855	55	50
Male	15,340	45	50
Ethnicity			
White	30,224	88	87
All other ethnic groups	3,971	12	13
Marital Status			
Single	7,720	23	26
Married, including civil partnership	17,630	52	49
Separated/Divorced, including from a civil partnership	3,205	9	8
Widowed, including civil partnership	546	2	1
Cohabitees	5,094	15	16
Highest level of education			
No qualifications	4,588	13	13
NVQ1/CSE other grade equivalent /NVQ2/GCE O Level equivalent	8,845	26	25
NVQ3/GCE A Level equivalent	5,136	15	15
Higher education, below degree level	3,753	11	11
NVQ4/NVQ5/Degree or equivalent	8,974	26	26
Still in full-time education/Other foreign qualification/Other	2,899	9	11
Occupation/NS-SEC 8			
Never worked/Long-term unemployed	712	2	2
Routine, manual and other occupations	13,508	40	41
Small employers and own account workers	2,917	9	9

Characteristic	Number (n)	Unweighted (%)	Weighted (%)
Intermediate occupations	12,926	38	36
Higher managerial and professional occupations	4,132	12	12
Body Mass Index (BMI)			
Underweight	596	2	2
Normal weight	12,644	37	38
Overweight	12,328	36	36
Obese/morbidly obese	8,627	25	24
Government Office Region (GOR)			
North East	2,502	7	5
North West	4,639	14	13
Yorkshire and the Humber	3,347	10	10
East Midlands	3,295	10	9
West Midlands	3,374	10	10
East of England	3,826	11	11
London	4,047	12	15
South East	5,616	16	17
South West	3,549	10	11
Index of Multiple Deprivation (IMD) Quintiles			
Least deprived area	7,240	21	21
2	6,901	20	20
3	6,994	21	21
4	6,607	19	20
Most deprived areas	6,453	19	19
Survey year			
2008	10,095	30	29
2009	3,039	9	9
2010	5,317	16	16
2011	5,341	16	15
2013	3,988	12	12
2015	3,341	10	10
2016	3,074	9	9
TOTAL	**34,195**	**100**	**100**

B.6 Unweighted and weighted percentage of 5-A-Day achievement captured in the NDNS 2008-16 and HSE 2008-16

Achievement of 5-A-Day target	NDNS			HSE		
	Number	Unweighted (%)	Weighted (%)	Number	Unweighted (%)	Weighted (%)
Yes	864	26	28	9,200	27	26
No	2,422	74	72	24,995	73	74
Total	**3,286**	**100**	**100**	**34,195**	**100**	**100**

B.7 Portions of fruit consumed (unweighted and weighted percentage) captured in the NDNS 2008-16 and HSE 2008-16

Portion(s) of fruit consumed	NDNS			HSE		
	Number	Unweighted (%)	Weighted (%)	Number	Unweighted (%)	Weighted (%)
0/Under 1 portion	1,582	48	46	8,466	25	25
1-2 portions	1,298	40	42	15,339	45	45
3-4 portions	323	10	10	7,568	22	22
5 or more portions	83	3	3	2,822	8	8
Total	**3,286**	**100**	**100**	**34,195**	**100**	**100**

B.8 Portions of vegetables consumed (unweighted and weighted percentage) captured in the NDNS 2008-16 and HSE 2008-16

Portion(s) of vegetables consumed	NDNS			HSE		
	Number	Unweighted (%)	Weighted (%)	Number	Unweighted (%)	Weighted (%)
0/Under 1 portion	392	12	10	9,721	28	29
1-2 portions	1,904	58	58	19,174	56	56
3-4 portions	810	25	25	4,307	13	13
5 or more portions	180	6	6	993	3	3
Total	**3,286**	**100**	**100**	**34,195**	**100**	**100**

B.9 Portions of fruits and vegetables consumed (unweighted and weighted percentage) captured in the NDNS 2008-16 and HSE 2008-16

Portion(s) of fruits and vegetables consumed	NDNS			HSE		
	Number	Unweighted (%)	Weighted (%)	Number	Unweighted (%)	Weighted (%)
0/Under 1 portion	190	6	4	3,470	10	10
1-2 portions	1,132	34	34	11,214	33	33
3-4 portions	1,100	34	34	10,311	30	30
5 or more portions	864	26	28	9,200	27	26
Total	**3,286**	**100**	**100**	**34,195**	**100**	**100**

B.10 Fully adjusted two-way cross-classified multi-level cumulative logistic regression model for portions of fruits (only) consumed by persons 16-64 years old in England, NDNS, 2008-16

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Intercept (Ref Cat: 5 or more portions)			
0 portions/Under 1 portion	1.543** (0.364)	___	___
1-2 portions	3.857** (0.37)	___	___
3-4 portions	5.635** (0.384)	___	___
Age (Ref Cat: 16-24 years)			
25-29	___	0.66**	(0.47, 0.91)
30-39	___	0.59**	(0.44, 0.79)
40-49	___	0.41**	(0.30, 0.55)
50-64	___	0.21**	(0.16, 0.29)
Sex (Ref Cat: Female)			
Male	___	1.43**	(1.24, 1.64)
Ethnicity (Ref Cat: White)			
All other ethnic groups	___	0.72**	(0.58, 0.90)
Marital Status (Ref Cat: Single Never Married)			
Married/In a legally recognised Civil Partnership	___	0.77**	(0.64, 0.94)
Separated/Divorced	___	1.19	(0.93, 1.53)
Widowed/Other/ Formerly in a legally recognised civil partnership	___	0.56*	(0.32, 0.99)
Highest level of education (Ref Cat: No qualifications)			
GCSE or equivalent qualifications	___	0.71**	(0.54, 0.93)
GCE A level or equivalent	___	0.57**	(0.43, 0.77)
Higher education, below degree level	___	0.40**	(0.29, 0.55)
Degree or equivalent	___	0.30**	(0.23, 0.40)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval
Still in full-time education/Other foreign qualification/Other	___	0.38**	(0.28, 0.53)
Equivalised household income (Ref Cat: Under £20,000)			
£20,000-£29,999	___	0.80*	(0.66, 0.97)
£30,000-£39,999	___	0.72**	(0.58, 0.90)
£40,000-£49,999	___	0.78*	(0.60, 1.02)
£50,000-£99,999	___	0.71**	(0.55, 0.90)
£100,000 or more	___	0.56**	(0.36, 0.88)
NS-SEC Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	___	0.97	(0.58, 1.59)
Small employers and own account workers	___	0.91	(0.53, 1.54)
Intermediate occupations	___	0.90	(0.54, 1.49)
Higher managerial and professional occupations	___	0.81	(0.47, 1.37)
Body Mass Index: BMI (Ref Cat: Underweight)			
Normal weight	___	0.94	(0.60, 1.44)
Overweight	___	1.17	(0.74, 1.82)
Obese	___	1.44	(0.91, 2.28)
Variance of random parameters			
Government Office Region (GOR)	0.039 (0.037)	___	___
Index of Multiple Deprivation (IMD)	0.017 (0.045)	___	___

* p-value <0.05, ** p-value<0.01, SE= standard error

Similar to fruit and vegetable portions, occupational grade (NS-SEC) and BMI were not statistically significantly associated with the number of fruit portions consumed by individuals (Appendix B.10). Age had a highly significant effect on the number of fruit portions consumed, with the odds of consuming less than five portions of fruit consistently decreasing, as the age groups increased (holding all other variables constant). Persons aged 25-29 years had a 34% reduced odds of consuming below five portions of fruit compared to those aged 16-24 years. Persons in the older

40-49 age group had a 59% reduced odds of consuming less than five portions compared to persons aged 16-24 years. However, the odds of consuming less than five portions was smallest for 50-64-year olds (0.21). This meant that persons in the 50-64-year age group had a 79% reduced odds of consuming less than five portions of fruit compared to persons aged 16-24 years, with the true population effect (measured by exponent of confidence intervals) between 71% and 84% (Appendix B.10).

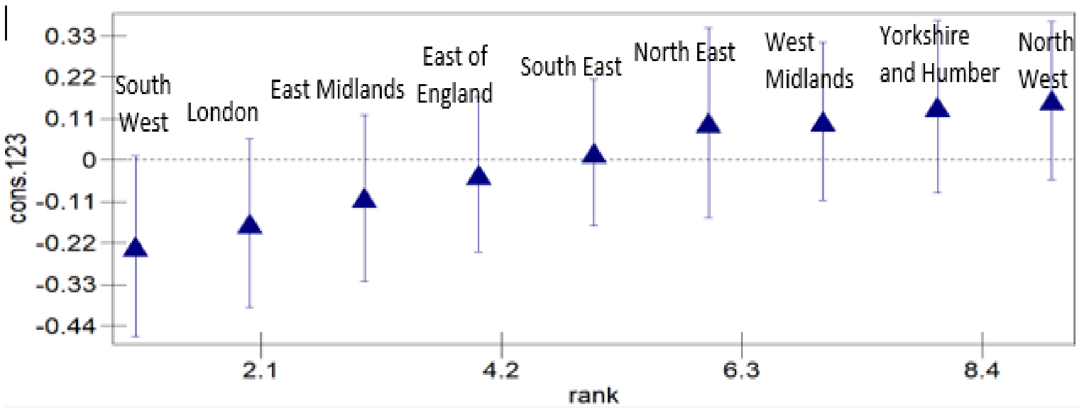
Sex also had a highly significant effect on fruit consumption. The odds that males consumed less than five portions was 1.43 times that of females (Appendix B.10). Persons from other ethnic groups had a 28% reduced odds of consuming less than five portions compared to white persons. The odds of consuming less than five portions of fruit was 23% less for married persons or those in legally recognised civil partnerships compared to single/never married persons. Being separated/divorced had no significant effect on fruit portions consumed. However, persons who were widowed or in formerly recognised civil partnerships had a 44% reduced odds of consuming less than five portions compared to those who were single/never married.

Highest level of education was statistically significantly associated with the number of fruit portions consumed. It appeared that the higher the level of education, the lower the odds of consuming below five portions of fruit. The odds of consuming less than five portions of fruit was 60% less for persons with below degree level qualifications compared to those with no qualifications. However, persons with a degree or equivalent qualification had the lowest odds of 0.30. This implied that the odds of consuming less than five portions of fruit was 70% less for persons with a degree, than persons without any qualifications.

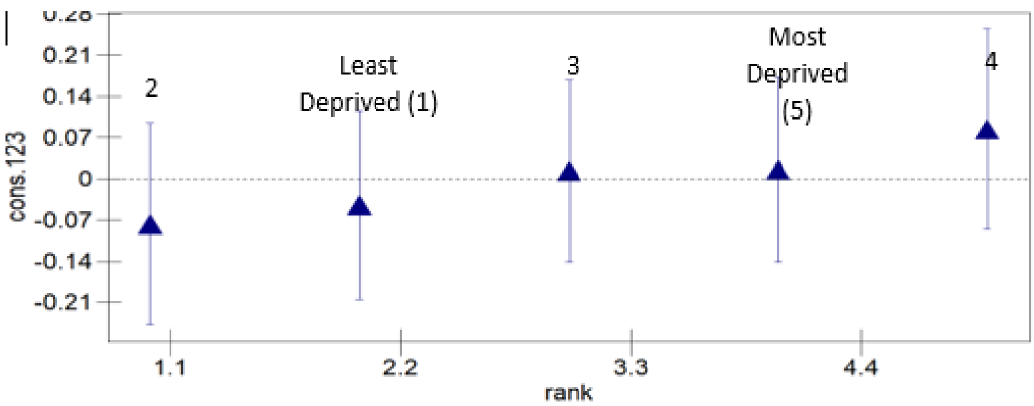
Apart from the £40,000-£49,999 income group, the odds of consuming less than five portions of fruit generally decreased, as household income increased (Appendix B.10). Persons within the £20,000-£29,999 income group had a 20% reduced odds of consuming less than five portions compared to those with household incomes below £20,000. The higher £50,000-£99,999 group had a 29% reduced odds compared to those within the less than £20,000 group. However, the odds of consuming less than five portions of fruit was 44% less for persons with a household income of £100,000 or more compared to those with a household income less than £20,000.

The variance for the GOR and IMD random parameters was 0.039 and 0.017 (respectively), both of which were fairly close to zero. Caterpillar residual plots helped to further illustrate this (Appendix B.11 and B.12 respectively), by showing that the confidence intervals for the regions and IMD quintiles, all overlapped zero. This meant that there was no statistically significant difference between the regions and deprivation quintiles, in terms of the number of fruit portions consumed.

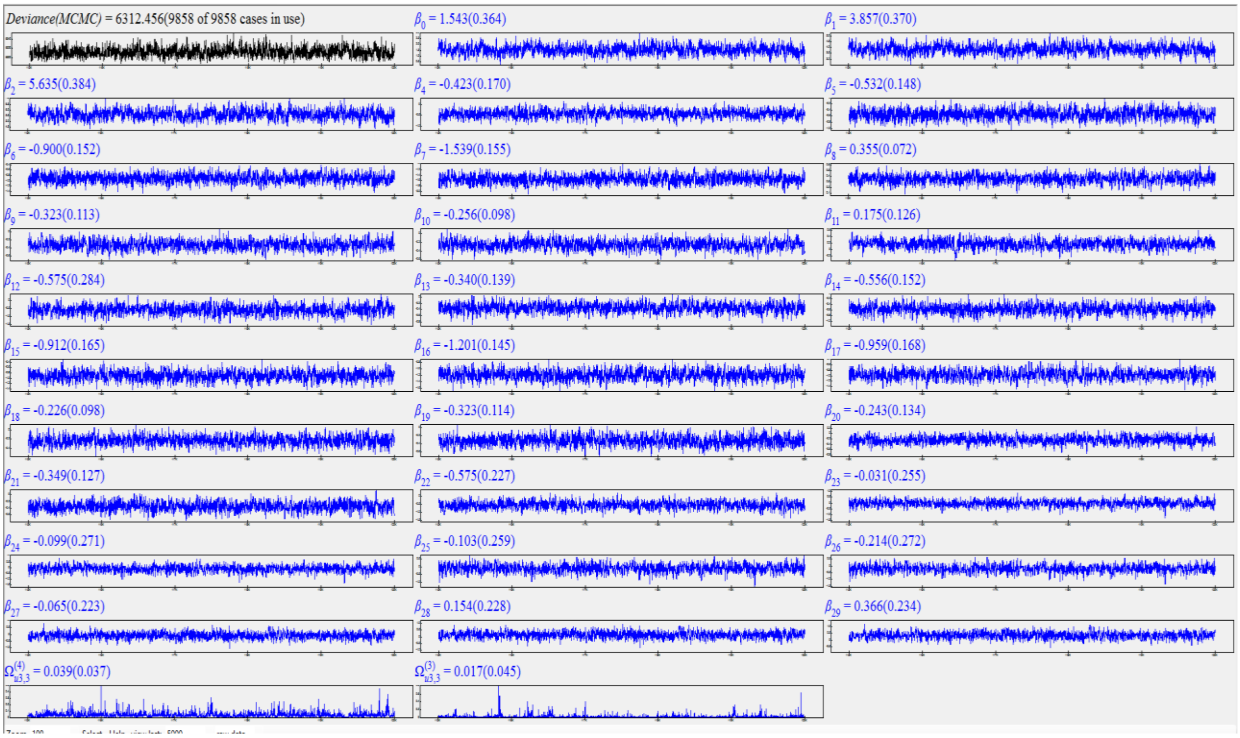
B.11 GOR level residuals and their associated 95% confidence intervals for the consumption of fruit portions, NDNS, 2008-16



B.12 IMD level residuals and their associated 95% confidence intervals for the consumption of fruit portions, NDNS, 2008-16



B.13 Trace plots showing convergence of parameters (Betas) for the final fruit portions model using 120,000 iterations, NDNS 2008-16



B.14 Fully adjusted two-way cross-classified multi-level cumulative logistic regression model for portions of vegetables (only) consumed by persons 16-64 years old in England, NDNS, 2008-16

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval)
Intercept (Ref Cat: 5 or more portions)			
0 portions/Under 1 portion	-0.375 (0.339)	___	___
1-2 portions	2.678** (0.344)	___	___
3-4 portions	4.772** (0.351)	___	___
Age (Ref Cat: 16-24 years)			
25-29	___	0.57**	(0.42, 0.79)
30-39	___	0.49**	(0.37, 0.65)
40-49	___	0.55**	(0.41, 0.73)
50-64	___	0.35**	(0.26, 0.47)
Sex (Ref Cat: Female)			
Male	___	0.97	(0.84, 1.11)
Ethnicity (Ref Cat: White)			
All other ethnic groups	___	0.80*	(0.64, 0.99)
Marital Status (Ref Cat: Single Never Married)			
Married/In a legally recognised Civil Partnership	___	0.88	(0.72, 1.06)
Separated/Divorced	___	1.10	(0.86, 1.40)
Widowed/Other/ Formerly in a legally recognised civil partnership	___	0.89	(0.52, 1.54)
Highest level of education (Ref Cat: No qualifications)			
GCSE or equivalent qualifications	___	0.83	(0.64, 1.08)
GCE A level or equivalent	___	0.70**	(0.53, 0.94)
Higher education, below degree level	___	0.81	(0.59, 1.12)
Degree or equivalent	___	0.51**	(0.39, 0.67)

Variables in the model	Log-odds coefficient (SE)	Odds ratio	95% Confidence Interval)
Still in full-time education/Other foreign qualification/Other	___	0.70*	(0.51, 0.95)
Equivalised household income (Ref Cat: Under £20,000)			
£20,000-£29,999	___	0.81*	(0.67, 0.98)
£30,000-£39,999	___	0.81*	(0.65, 1.02)
£40,000-£49,999	___	0.74*	(0.57, 0.96)
£50,000-£99,999	___	0.65**	(0.51, 0.84)
£100,000 or more	___	0.42**	(0.27, 0.66)
NS-SEC Occupation Grade (Ref Cat: Never worked/Long term unemployed)			
Routine, manual and other occupations	___	0.75	(0.48, 1.19)
Small employers and own account workers	___	0.62*	(0.38, 1.02)
Intermediate occupations	___	0.69	(0.43, 1.10)
Higher managerial and professional occupations	___	0.68	(0.41, 1.11)
Body Mass Index: BMI (Ref Cat: Underweight)			
Normal weight	___	0.76	(0.49, 1.17)
Overweight	___	0.90	(0.58, 1.40)
Obese	___	0.89	(0.56, 1.41)
Variance of random parameters			
Government Office Region (GOR)	0.022 (0.024)	___	___
Index of Multiple Deprivation (IMD)	0.017 (0.044)	___	___

* p-value <0.05, ** p-value<0.01, SE= standard error

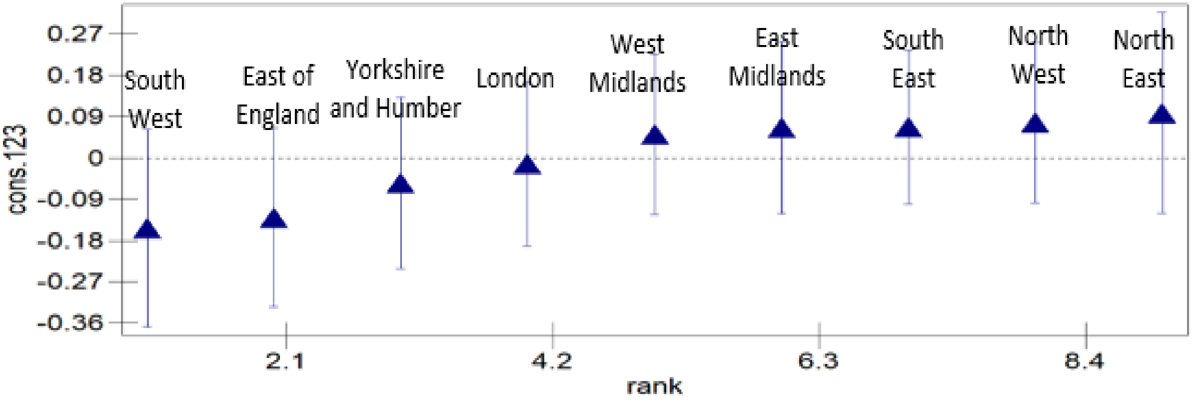
Sex, marital status and BMI were not statistically significantly associated with the number of vegetable portions consumed by individuals (Appendix B.14). Unlike combined fruit and vegetable and fruit portions, it was only for vegetable portions that NS-SEC occupational grade had a statistically significant association. The odds of consuming below five portions of vegetables was 38% less for small employers and own account workers compared to those who never worked or were long-term unemployed.

Compared to white persons, persons from other ethnic groups had a 20% reduced odds of consuming less than five portions of vegetables (Appendix B.14). Age was a highly significant predictor of vegetable consumption. Despite an increase in the odds for the 40-49 age group, the odds of consuming less than five portions of vegetables tended to decrease, as the age groups increased (Appendix B.14). The odds of consuming below five portions of vegetables was 43% less for persons aged 25-29 years compared to those aged 16-24 years. However, persons aged 50-64 years had the lowest odds (0.35). Thus, the odds of consuming below five portions of vegetables was 65% less for persons aged 50-64 years compared to persons aged 16-24 years.

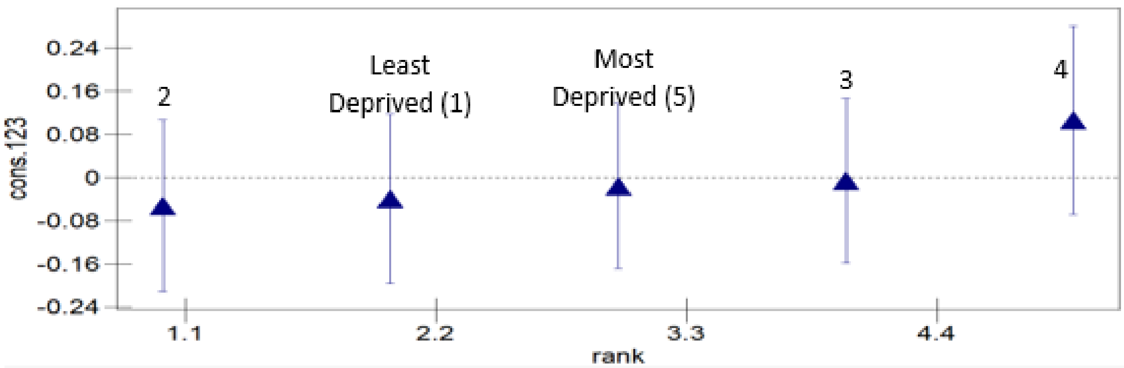
In terms of SEP, attainment of GCE A level or equivalent qualifications was statistically significantly associated with the number of vegetable portions consumed. However, persons with a degree or equivalent level qualification had the lowest odds (0.51). The odds of consuming less than five portions was 49% less for persons with a degree/equivalent, than for persons with no qualifications (holding all other variables in the model constant). In general, the odds of consuming less than five portions of vegetables tended to decrease as household income increased. The £100,000 or more household income group had the lowest odds of 0.42. This meant that the odds of consuming less than five portions of vegetables was 58% less for persons with a household income of £100,000 or more compared to persons within the under £20,000 household income group.

Similar to the fruit and fruit and vegetable portions models, the variance for the GOR and IMD random parameters (0.022 and 0.017 respectively) were both fairly close to zero. In addition, the caterpillar residual plots presented in Appendix B.15 and B.16 (respectively), both showed that the confidence intervals for the regions and IMD quintiles, all overlapped zero. This suggested that there was no statistically significant difference between the regions and deprivation quintiles, in terms of the number of vegetable portions consumed.

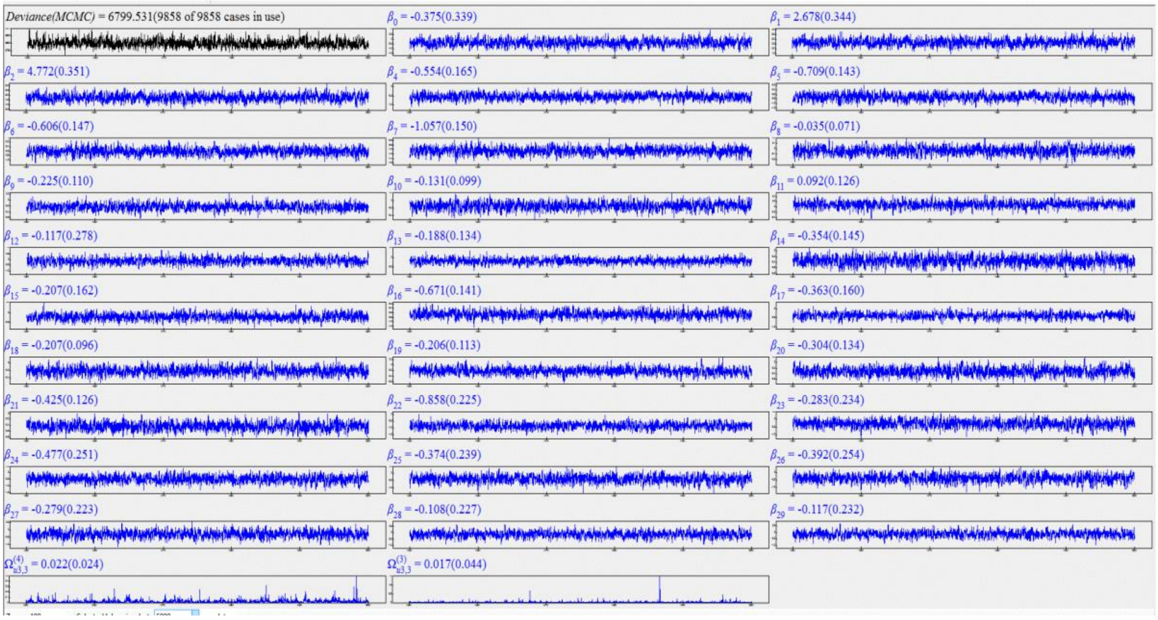
B.15 GOR level residuals and their associated 95% confidence intervals for the consumption of vegetable portions, NDNS, 2008-16



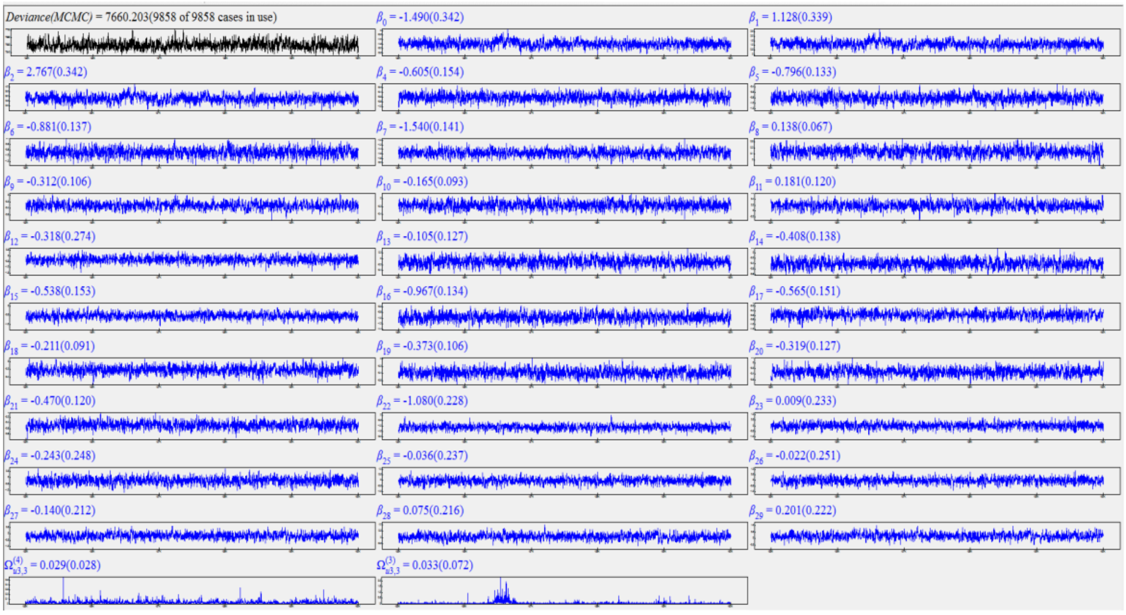
B.16 IMD level residuals and their associated 95% confidence intervals for the consumption of vegetable portions, NDNS, 2008-16



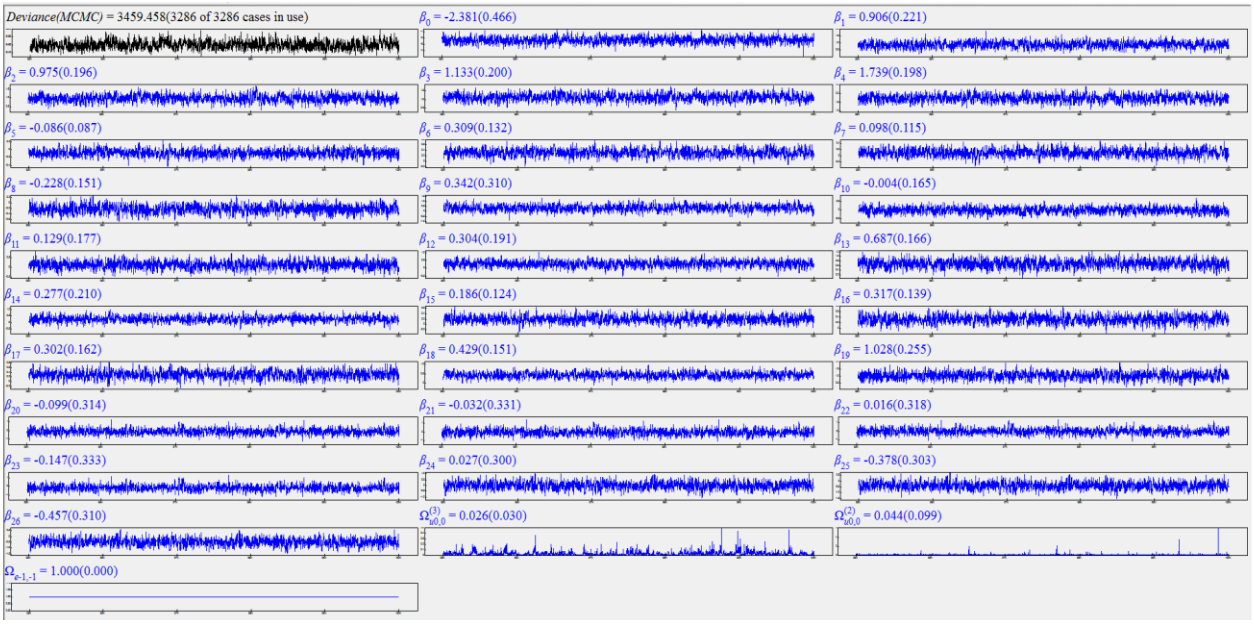
B.17 Trace plots showing convergence of parameters (Betas) for the final vegetable portions model using 150,000 iterations, NDNS 2008-16



B.18 Trace plots showing convergence of parameters (Betas) for the final fruit and vegetable (combined) portions model using 160,000 iterations, NDNS 2008-16



B.19 Trace plots showing convergence of parameters (Betas) for the final 5-A-Day achievement model using 100,000 iterations, NDNS 2008-16



Appendix C Research Question 2 supplemental results

C.1 Rotated PCA component loadings for the first component of the national sample and the nine regions of England, NDNS, 2008-16

Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
Fruit	**0.66**	**0.77**	-0.15	**0.72**	**0.68**	-0.21	**0.63**	**0.67**	**0.59**	**0.54**
Yogurt, fromage frais and dairy desserts	**0.58**	**0.85**	0.09	**0.61**	**0.33**	-0.08	**0.32**	**0.58**	**0.56**	0.13
Salad and other raw vegetables	**0.52**	**0.35**	**-0.25**	**0.60**	**0.69**	-0.07	0.22	**0.43**	**0.72**	**0.71**
Skimmed milk	**0.44**	0.12	-0.13	**0.28**	0.16	-0.02	-0.01	**0.62**	0.06	-0.11
Tea, coffee and water	**0.42**	0.17	-0.24	**0.51**	**0.35**	0.15	**0.32**	0.21	**0.36**	**0.32**
High fibre breakfast cereals	**0.27**	0.01	-0.09	**0.25**	**0.62**	0.00	0.11	0.08	0.11	**0.26**
Other margarine, fats and oils	**0.25**	0.06	-0.01	0.06	**0.32**	-0.05	**0.65**	0.07	**0.31**	**0.39**

Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
Nuts and seeds	0.22	-0.10	-0.02	0.24	0.09	-0.05	0.16	0.02	0.11	0.24
Wholemeal bread	0.22	**0.25**	-0.03	0.13	0.15	0.00	-0.04	0.10	0.13	0.13
Other white fish, shellfish and fish dishes	0.18	0.05	-0.04	0.24	0.23	0.03	-0.05	-0.01	0.10	**0.26**
Oily fish	0.15	0.06	-0.02	0.24	0.20	-0.01	0.12	**0.26**	0.16	0.01
Biscuits	0.15	0.04	-0.08	**0.27**	0.04	0.08	0.14	0.11	0.05	-0.04
Vegetables not raw	0.12	**0.27**	0.00	**0.29**	0.08	0.03	0.11	0.15	-0.03	0.17
Other milk and cream	0.12	0.05	**0.33**	0.06	0.02	-0.10	**0.68**	0.00	0.00	0.03
Low fat spread not polyunsaturated	0.11	0.12	-0.04	-0.12	0.03	0.03	0.03	0.10	0.04	-0.07
Bacon and ham	0.10	0.13	0.04	0.09	-0.03	-0.06	-0.05	0.10	0.04	0.17
Cheese	0.10	0.13	-0.05	0.19	0.11	-0.03	0.09	0.14	0.15	0.00
Other potatoes, potato salads and dishes	0.10	-0.01	-0.02	**0.33**	0.04	0.05	0.11	0.18	0.04	0.08

Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
Fruit juice including smoothies	0.08	0.13	-0.04	0.01	0.11	-0.10	-0.09	-0.04	0.07	-0.01
Other breakfast cereals	0.07	0.11	-0.08	-0.03	-0.02	-0.08	0.05	0.02	-0.01	-0.03
Puddings	0.06	0.01	-0.01	0.06	-0.05	0.02	0.28	-0.04	-0.01	0.00
Low fat spread polyunsaturated	0.05	0.09	0.00	0.10	-0.06	**0.73**	-0.01	0.09	-0.02	-0.14
Buns, cakes, pastries and fruit pies	0.05	0.10	-0.10	0.15	-0.02	-0.13	-0.04	0.17	0.04	-0.03
Brown, granary and wheatgerm bread	0.05	-0.15	-0.11	-0.04	0.06	-0.04	-0.03	**0.31**	0.05	0.23
Semi skimmed milk	0.05	0.00	-0.08	-0.01	0.06	0.01	-0.08	0.08	0.10	-0.03
Beef, veal and dishes	0.05	-0.04	0.01	-0.02	-0.02	-0.17	-0.03	-0.01	-0.08	0.08
Wine	0.05	0.02	-0.19	0.06	0.00	-0.19	0.03	-0.03	0.08	0.00
Chicken and turkey dishes	0.05	**0.31**	0.03	0.14	-0.01	-0.11	0.00	-0.05	0.01	0.01

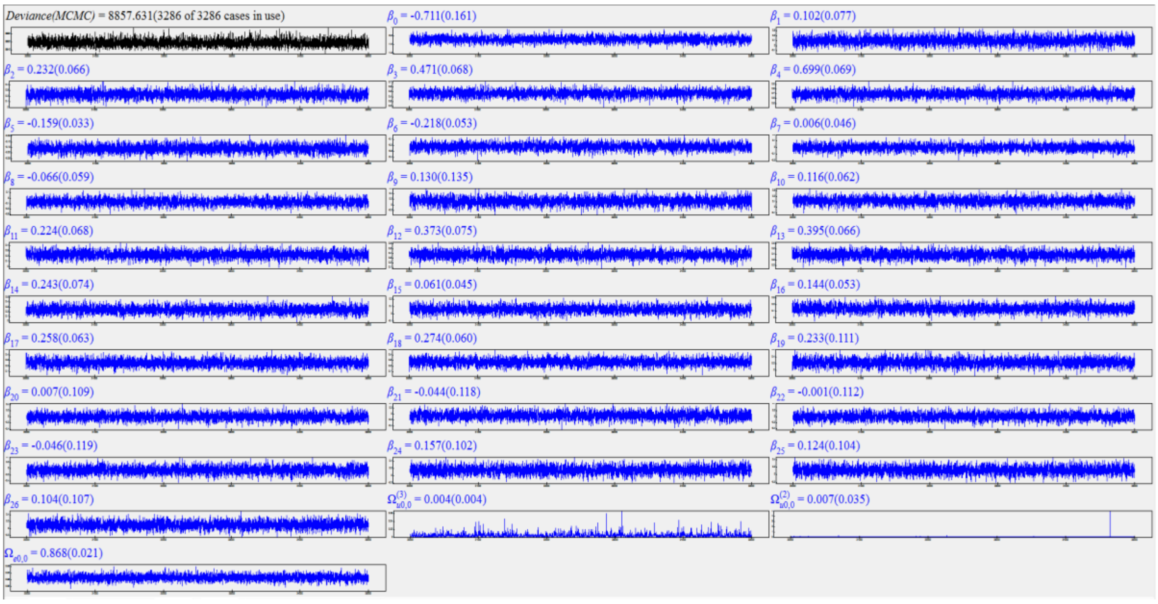
Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
Eggs and egg dishes	0.05	0.21	0.03	0.17	0.05	-0.07	-0.04	0.02	-0.11	0.07
Chocolate confectionery	0.04	0.01	0.03	0.02	-0.07	-0.04	-0.06	0.18	0.12	**0.28**
Dry weight beverages	0.03	0.06	0.06	-0.02	-0.05	0.03	-0.01	0.02	0.10	-0.06
Sugar confectionery	0.03	-0.03	0.01	-0.14	-0.07	0.08	0.04	0.05	0.04	0.01
Reduced fat spread polyunsaturated	0.02	0.03	0.05	-0.12	-0.03	0.02	-0.06	0.14	0.04	-0.10
Soup homemade and retail	0.02	0.03	**-0.36**	0.01	0.05	0.05	0.05	0.22	-0.05	0.04
Other bread	0.01	0.06	-0.01	0.07	-0.01	-0.04	-0.02	0.23	-0.06	0.05
One percent milk	0.00	0.13	-0.01	-0.04	-0.06	0.02	-0.04	0.02	0.09	0.08
Sauces, pickles and gravies	0.00	-0.03	0.04	0.10	0.09	0.10	-0.07	-0.04	0.00	0.09
Lamb and dishes	-0.01	-0.09	0.24	0.02	0.02	-0.03	-0.03	-0.02	-0.03	0.01

Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
Spirits and liqueurs	-0.01	0.00	0.00	0.02	0.09	-0.03	0.00	-0.15	-0.09	0.02
Liver and dishes	-0.03	0.00	0.17	0.06	0.00	-0.11	0.16	0.00	-0.03	-0.03
Burgers and kebabs	-0.03	-0.10	**0.74**	-0.21	-0.23	-0.06	-0.03	-0.01	-0.10	-0.09
Ice cream	-0.04	-0.06	-0.01	0.01	-0.01	0.04	-0.04	-0.01	0.12	-0.01
Sausages	-0.05	-0.16	0.21	-0.08	-0.02	**0.26**	-0.12	0.00	-0.10	-0.18
Pork and dishes	-0.05	0.00	-0.08	0.01	0.04	0.01	0.06	-0.12	0.15	-0.14
Soft drinks low calorie	-0.06	-0.07	0.07	-0.01	0.02	0.00	0.05	0.14	-0.21	-0.04
PUFA margarine and oils	-0.06	0.00	-0.04	-0.03	-0.22	-0.04	**0.37**	-0.05	-0.01	-0.01
Meat pies and pastries	-0.06	**-0.29**	0.18	-0.10	-0.01	0.00	-0.06	-0.21	0.01	-0.14
Coated chicken and turkey	-0.07	-0.09	**0.34**	0.00	-0.17	0.19	-0.05	-0.14	-0.07	-0.01
Butter	-0.08	-0.24	0.00	-0.05	0.03	-0.01	-0.04	0.03	-0.06	**0.37**
Sugar, preserves and sweet spreads	-0.09	-0.12	0.05	0.08	-0.14	**0.51**	0.06	-0.02	-0.08	-0.14

Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
White fish coated or fried	-0.09	-0.01	-0.02	-0.04	-0.01	**0.59**	-0.06	-0.02	-0.07	-0.17
Beer, lager, cider and perry	-0.10	-0.18	0.16	0.03	-0.05	0.01	-0.17	-0.11	-0.04	-0.06
Crisps and savoury snacks	-0.13	-0.08	0.14	-0.04	-0.14	0.13	0.01	-0.06	-0.14	-0.03
Reduced fat spread not polyunsaturated	-0.14	0.01	-0.05	-0.06	0.01	-0.07	-0.06	-0.01	-0.21	-0.18
Soft drinks not low calorie	-0.15	-0.10	**0.30**	-0.22	-0.22	-0.03	-0.03	-0.18	-0.05	-0.04
Other meat and meat products	-0.16	-0.07	0.00	-0.04	0.01	0.15	-0.06	-0.14	-0.02	-0.07
Pasta, rice and other cereals	-0.17	0.19	-0.02	0.03	-0.09	-0.08	0.06	-0.09	-0.12	-0.17
Whole milk	-0.17	-0.13	0.02	-0.04	-0.07	**0.50**	-0.09	-0.13	-0.08	-0.17
Chips, fried and roast potatoes and potato products	-0.18	**-0.25**	**0.63**	**-0.29**	**-0.25**	**0.29**	-0.08	-0.16	-0.23	-0.21

Food group	National sample	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
White bread	-0.22	-0.13	0.12	-0.16	0.00	0.13	-0.05	-0.09	-0.22	-0.09

C.2 Trace plots showing convergence of parameters (Betas) for the final PCA derived dietary pattern scores model using 25,000 iterations, NDNS 2008-16



Appendix D Research Question 3 supplemental results

D.1 Comparison of single-level and three-level hierarchical multi-level binary logistic regression model with the four constraint variables selected and used in the IPF small area estimation method, HSE, 2008-11, 2013 and 2015-16

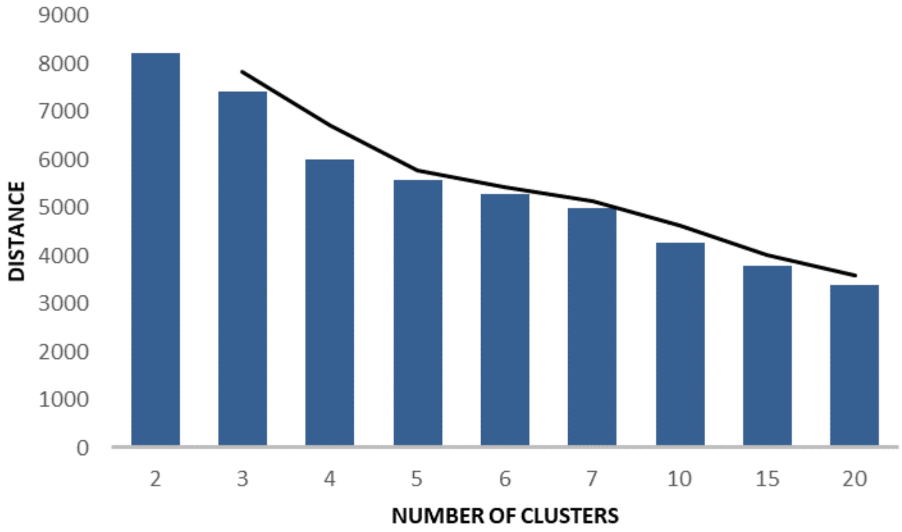
	Single-level model			Three-level hierarchical multi-level model		
Variables in the model	**Log-odds coefficient (SE)**	**Exp (β)/Odds ratio**	**Exp (Confidence Interval)**	**Log-odds coefficient (SE)**	**Exp (β)/Odds ratio**	**Exp (Confidence Interval)**
Intercept	-2.046** (0.106)	___	___	-2.133**(0.122)	___	___
Age (Ref Cat: 16-24 years)						
25-29	___	1.08	(0.95, 1.24)	___	1.08	(0.94, 1.23)
30-39**	___	1.50	(1.34, 1.67)	___	1.51	(1.35, 1.69)
40-49**	___	1.63	(1.46, 1.81)	___	1.65	(1.47, 1.84)
50-64**	___	2.13	(1.92, 2.37)	___	2.19	(1.96, 2.44)
Ethnicity (Ref Cat: White)						
Mixed	___	1.21	(0.99, 1.48)	___	1.15	(0.93, 1.42)

Variables in the model	Single-level model Log-odds coefficient (SE)	Exp (β)/Odds ratio	Exp (Confidence Interval)	Three-level hierarchical multi-level model Log-odds coefficient (SE)	Exp (β)/Odds ratio	Exp (Confidence Interval)
Asian**	___	1.68	(1.53, 1.85)	___	1.62	(1.46, 1.79)
Black	___	1.11	(0.96, 1.29)	___	1.00	(0.85, 1.18)
Other**	___	1.93	(1.49, 2.50)	___	1.75	(1.34, 2.30)
Highest level of education (Ref Cat: No qualifications)						
NVQ1/CSE other grade equivalent/NVQ2/GCE O Level equivalent**	___	1.14	(1.04, 1.25)	___	1.15	(1.05, 1.27)
NVQ3/GCE A Level equivalent/Higher education below degree level**	___	1.63	(1.48, 1.79)	___	1.65	(1.50, 1.82)
NVQ4/NVQ5/Degree or equivalent**	___	2.5	(2.27, 2.75)	___	2.48	(2.24, 2.74)
Still in full-time education/Other foreign qualification/Other**	___	1.92	(1.68, 2.20)	___	1.91	(1.67, 2.20)
NS-SEC8 Occupation Grade (Ref Cat: Never worked/Long term unemployed)						
Routine, manual and other occupations	___	0.86	(0.72, 1.04)	___	0.89	(0.73, 1.08)

	Single-level model			Three-level hierarchical multi-level model		
Variables in the model	**Log-odds coefficient (SE)**	**Exp (β)/Odds ratio**	**Exp (Confidence Interval)**	**Log-odds coefficient (SE)**	**Exp (β)/Odds ratio**	**Exp (Confidence Interval)**
Small employers and own account workers	___	1.22	(1.00, 1.49)	___	1.23	(1.00, 1.52)
Intermediate occupations	___	1.14	(0.94, 1.38)	___	1.17	(0.96, 1.42)
Higher managerial and professional occupations	___	1.17	(0.96, 1.42)	___	1.19	(0.97, 1.46)
Variance of random parameters						
Government Office Region (GOR)	___	___	___	0.022 (0.017)	___	___
Primary Sampling Unit (PSU)	___	___	___	0.210 (0.022)	___	___

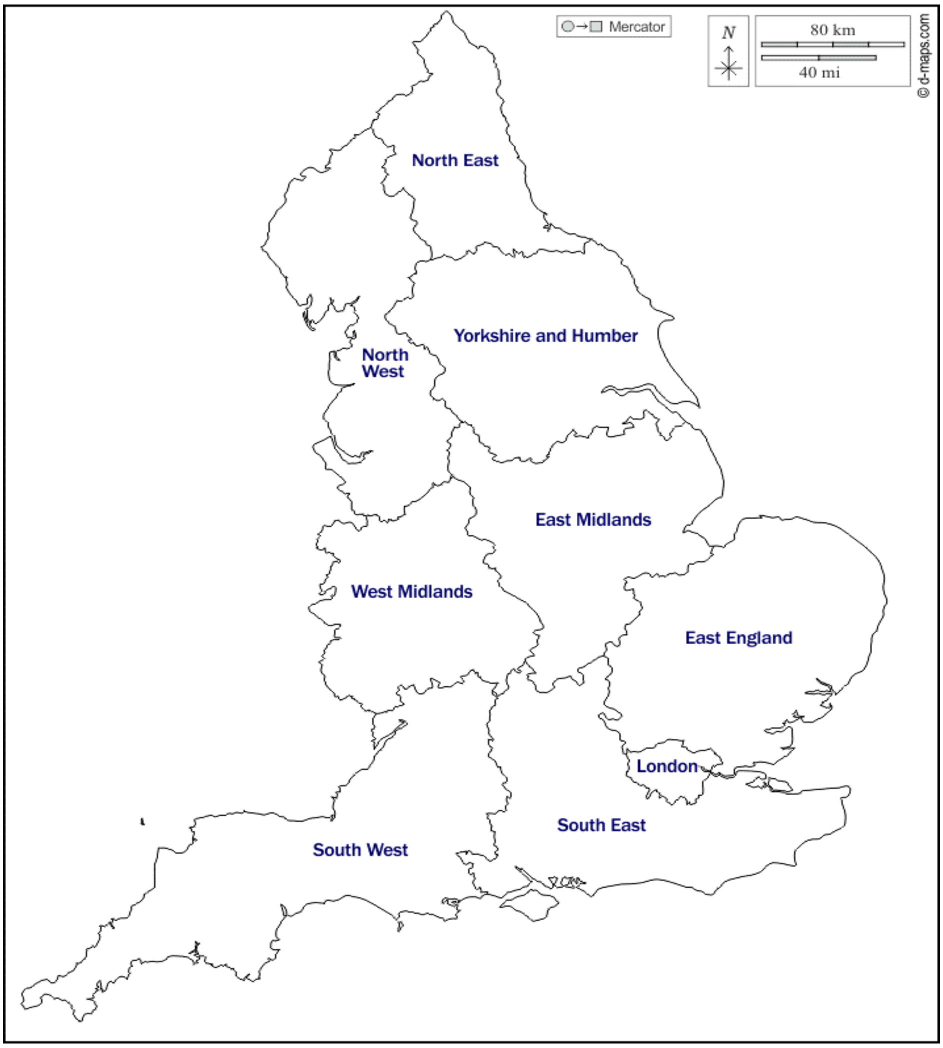
* p-value <0.05, ** p-value<0.01, SE= standard error

D.2 Elbow plot used to select the optimal number of (K) clusters to be retained HSE, 2008-11, 2013 and 2015-16



Appendix E General information and project outputs

E.1 Map of the GORs of England



E.2 Publications

Publication	Date of publication
Smith, D. M., Vogel, C., Campbell, M., Alwan, N. and Moon, G. (2021) 'Adult diet in England: Where is more support needed to achieve dietary recommendations?', PLOS ONE, 16(6), pp. e0252877. doi:10.1371/journal.pone.0252877	June 23, 2021
Smith, D. M., Heppenstall, A. and Campbell, M. (2021) 'Estimating Health over Space and Time: A Review of Spatial Microsimulation Applied to Public Health', J, 4(2), pp. 182-192. doi: https://doi.org/10.3390/j4020015	June 9, 2021
Campbell, M., Smith, D., Baird, J., Vogel, C. and Moon, E. G. (2020) 'A critical review of diet-related surveys in England, 1970-2018', Archives of Public Health, 78(1), pp. 66. doi:10.1186/s13690-020-00447-6	July 7, 2020

E.3 Conference presentations: 2018-2021

Conference	Location	Date
The 2019 UK Data Service Health Users' Conference	University College of London	July 10, 2019
The 2019 Emerging and New Researchers in the Geographies of Health & Impairment (ENRGHI) Conference	University of Exeter	June 18, 2019
The 2019 Annual Geography PGR Conference	University of Southampton	June 14, 2019

E.4 Training and related activities completed: 2018-2021

Name of course/Activity	Affiliated Institution/Location	Date
Introduction to Multilevel Modelling Using MLwiN	University of Bristol, Centre for Multi-level Modelling (virtual)	January 26-28, 2021
Practicing Human Geographical Research (GEOG2035) SPSS Course Demonstration	University of Southampton	March 12, 2020
ESRC Impact: Healthy Ageing Sector Scoping Event	University of Southampton	January 29, 2020
Personal Leadership training workshop	University of Southampton, Leadership and Management Development Unit	December 2, 2019
Optimising Your Time training workshop	University of Southampton, Leadership and Management Development Unit	November 25, 2019
Successful Communication training workshop	University of Southampton, Leadership and Management Development Unit	November 18, 2019
Invigilation training	University of Southampton	June 4, 2019
Word: creating a thesis - save time and tears part 2 (PC users) thesis development course	University of Southampton	May 15, 2019
SCDTP Boosting business engagement of your research workshop Part 2	University of Southampton	March 26, 2019
Introduction to R course	SAGE Campus (virtual)	March 25, 2019
SCDTP Research Impact workshop	University of Southampton	February 18, 2019
LEMMA Online Multi-level Modelling self-taught course	University of Bristol, Centre for Multi-level Modelling (virtual)	October 23, 2018
SCDTP Boosting business engagement in research Part 1	University of Southampton	October 15, 2018

Name of course/Activity	Affiliated Institution/Location	Date
Introduction to STATA self-taught module	University of Southampton, MRC Lifecourse and Epidemiology Unit	October 11, 2018
Word: creating a thesis - Save time and tears part 3 (PC users) thesis development course	University of Southampton	May 23, 2018
Word: creating a thesis - save time and tears part 1 (PC users) thesis development course	University of Southampton	April 25, 2018
Introduction to Mandarin	University of Southampton, Department of Humanities, Avenue Campus	March 19-23, 2018
EndNote for Beginners (Bibliographic Software) thesis development course	University of Southampton. Hartley Library	March 22, 2018
Quantitative methods/Text analysis webinar	University of Essex (virtual)	February 21, 2018
Orientation to teaching & demonstrating	University of Southampton	February 6, 2018

Glossary of Terms

ALS..Active Lives Survey

ALSPAC ....Avon Longitudinal Study of Parents and Children

BCS70 ....British Cohort Study 1970

BiB ....Born in Bradford

BMI.......................................Body Mass Index

BRHS.....................................British Regional Heart Study

BWHHS.................................British Women's Heart and Health Study

CAPI ....Computer Assisted Personal Interviewing

CDRC.....................................Consumer Data Research Centre

CLOSER ....Cohort and Longitudinal Studies Enhancement Resources

CLS..Centre for Longitudinal Studies

DEFRA...................................Department for Environment Food & Rural Affairs

DIC..Deviance Information Criterion

EPIC ....European Prospective Investigation into Cancer and Nutrition

EU ....European Union

FCS..Fully Conditional Specification

FDI ....Foreign Direct Investment

FFQ ....Food Frequency Questionnaire

GBD ....Global Burden of Disease

GCE.......................................General Certificate of Education

GCSE ....General Certificate of Secondary

GI ....Geographical Indications

GOR ....Government Office Region

GPS ....Global Positioning System

HCES ....Household Consumption and Expenditure Survey

HFSS ....High fat sugar and salt

HRP .... Household Reference Person

HSE...................................... Health Survey for England

IMD .... Index of Multiple Deprivation

IPF .... Iterative Proportional Fitting

KMO .... Kaiser-Meyer-Olkin

LAD...................................... Local Authority District

LEMMA .... Learning Environment for Multilevel Methodology and Applications

LCFS..................................... Living Costs and Food Survey

LSOA.................................... Lower Layer Super Output Area

MAR .... Missing at random

MAUP.................................. Modifiable Area Unit Problem

MCAR .... Missing completely at random

MCMC................................. Markov Chain Monte Carlo

MCS..................................... Millennium Cohort Study

MD .... Mediterranean Diet

MI... Multiple Imputation

MRC .... Medical Research Council

MRC EWL .... Medical Research Council Elsie Widdowson Laboratory

MSOA.................................. Middle Layer Super Output Area

NDNS................................... National Diet and Nutrition Survey

NDNS RP.............................. National Diet and Nutrition Survey Rolling Programme

NGO .... Non-Governmental Organisation

NHS .... National Health Service

NS-SEC................................. National Statistics Socio-economic Classification

NVQ..................................... National Vocational Qualification

OAC .... Output Area Classification

OECD................................... Organisation for Economic Co-operation and Development

ONS .... Office of National Statistics

PAF ....Postcode Address File

PCA..Principal Components Analysis

PHE..Public Health England

PSU ....Primary Sampling Unit

PUFA.......................................Polyunsaturated fatty acid

QQ ....Quantile-Quantile

SAE ....Small Area Estimation

SCDTP ....South Coast Doctoral Training Partnership

SE..Standard Error

SEP...Socio-economic position

SSB...Sugar sweetened beverage

StAE ....Standardised Absolute Error

SWS ....Southampton Women's Survey

TAE ....Total Absolute Error

UCL ....University College London

USA..United States of America

UK...United Kingdom

UKWCS ....UK Women's Cohort Study

WHO......................................World Health Organisation

List of References

Acheson, D. (1998) 'Independent Inquiry into Inequalities in Health Report'. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/265503/ih.pdf (Accessed March 19, 2018).

Adams, J., Mytton, O., White, M. and Monsivais, P. (2016) 'Why Are Some Population Interventions for Diet and Obesity More Equitable and Effective Than Others? The Role of Individual Agency', *PLoS Med,* 13(4), pp. e1001990. doi:10.1371/journal.pmed.1001990

Afshin, A. and Sur, P. J. and Fay, K. A. and Cornaby, L. and Ferrara, G. and Salama, J. S., . . . Murray, C. J. L. (2019) 'Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017', *The Lancet,* 393(10184), pp. 1958-1972. doi:10.1016/S0140-6736(19)30041-8

Agudo, A. (2005) *Measuring intake of fruits and vegetables*: World Health Organisation. Available at: https://www.who.int/dietphysicalactivity/publications/f&v_intake_measurement.pdf (Accessed July 3, 2019).

Allison, P. D. (2001) *Missing Data. Quantitative Applications in the Social Sciences.* SAGE Publications.

Anyaegbu, G. (2010) 'Using the OECD equivalence scale in taxes and benefits analysis', *Economic & Labour Market Review,* 4(1), pp. 49-54. doi:10.1057/elmr.2010.9

Appelhans, B. M., French, S. A., Tangney, C. C., Powell, L. M. and Wang, Y. (2017) 'To what extent do food purchases reflect shoppers' diet quality and nutrient intake?', *Int J Behav Nutr Phys Act,* 14(1), pp. 46. doi:10.1186/s12966-017-0502-2

Arsenault, J., Michel, P., Berke, O., Ravel, A. and Gosselin, P. (2013) 'How to choose geographical units in ecological studies: Proposal and application to campylobacteriosis', *Spatial and Spatio-temporal Epidemiology*, 7, pp. 11-24. doi: https://doi.org/10.1016/j.sste.2013.04.004.

Asaria, M. (2017) *The economics of health inequality in the English NHS: the long view*: Centre for Health Economics, University of York. Available at: https://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP142_economics_health_inequality_NHS.pdf (Accessed June 10, 2018).

Baker, C. (2021) 'Obesity statistics'. Available at: https://researchbriefings.files.parliament.uk/documents/SN03336/SN03336.pdf (Accessed: October 14, 2021).

Bambra, C. (2016) *Health Divides: Where You Live Can Kill You.* Great Britain: Policy Press.

Bambra, C., Gibson, M., Sowden, A., Wright, K., Whitehead, M. and Petticrew, M. (2010) 'Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews', *J Epidemiol Community Health,* 64(4), pp. 284-91. doi:10.1136/jech.2008.082743

Barao, K. and Forones, N. (2012) 'Body mass index: different nutritional status according to WHO, OPAS and Lipschitz classifications in gastrointestinal cancer patients', *Arquivos de Gastroenterologia*, 49(2), pp. 169-171. doi:https://doi.org/10.1590/S0004-28032012000200013

Barton, K. L., Chambers, S., Anderson, A. S. and Wrieden, W. L. (2018) 'Time to address the double inequality of differences in dietary intake between Scotland and England', *Br J Nutr,* 120(2), pp. 220-226. doi:10.1017/S0007114518001435

Bates, B., Lennox, A., Prentice, A., Bates, C. and Swan, G. (2012) 'National Diet and Nutrition Survey Headline results from Years 1, 2 and 3 (combined) of the Rolling Programme (2008/2009 – 2010/11) '. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/207708/NDNS-Y3-report_All-TEXT-docs-combined.pdf (Accessed June 14, 2021).

Beech, J., Holmes, J., McKenna, H. and Cooper, E. (2020) 'What role do taxes and regulation play in promoting better health?'. Available at: https://www.kingsfund.org.uk/sites/default/files/2020-03/tax-regulation-briefing-2020.pdf (Accessed February 10, 2021).

Bere, E. and Brug, J. (2010) 'Is the term 'Mediterranean diet' a misnomer?', *Public Health Nutr,* 13(12), pp. 2127-9. doi:10.1017/S1368980010000480

Bernard, P., Charafeddine, R., Frohlich, K.L., Daniel, M., Kestens, Y. and Potvin, L. (2007) 'Health inequalities and place: a theoretical conception of neighbourhood', *Soc Sci Med*, 65(9), pp. 1839-52. doi: 10.1016/j.socscimed.2007.05.037.

Berzofsky, M., Smiley-McDonald, H., Moore, A., Krebs, C. and Creel, D. (2014) 'Measuring Socioeconomic Status (SES) in the NCVS: Background, Options, and Recommendations'. Available at: https://www.bjs.gov/content/pub/pdf/Measuring_SES-Paper_authorship_corrected.pdf (Accessed May 6, 2021).

Billson, H., Pryer, J. A. and Nichols, R. (1999) 'Variation in fruit and vegetable consumption among adults in Britain. An analysis from the dietary and nutritional survey of British adults', *Eur J Clin Nutr,* 53(12), pp. 946-52. doi:10.1038/sj.ejcn.1600877

Black, C., Moon, G. and Baird, J. (2014) 'Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment?', *Health Place,* 27, pp. 229-42. doi:10.1016/j.healthplace.2013.09.015

Black, C., Ntani, G., Kenny, R., Tinati, T., Jarman, M., Lawrence, W., . . . Baird, J. (2012) 'Variety and quality of healthy foods differ according to neighbourhood deprivation', *Health Place,* 18(6), pp. 1292-9. doi:10.1016/j.healthplace.2012.09.003

Blow, J., Gregg, R., Davies, I. G. and Patel, S. (2019) 'Type and density of independent takeaway outlets: a geographical mapping study in a low socioeconomic ward, Manchester', *BMJ open,* 9(7), pp. e023554-e023554. doi:10.1136/bmjopen-2018-023554

Bonaccio, M., Di Castelnuovo, A., Pounis, G., Costanzo, S., Persichillo, M., Cerletti, C., . . . Investigators, o. b. o. t. M.-s. S. (2017) 'High adherence to the Mediterranean diet is associated with cardiovascular protection in higher but not in lower socioeconomic groups: prospective findings from the Moli-sani study', *International Journal of Epidemiology,* 46(5), pp. 1478-1487. doi:10.1093/ije/dyx145

Boo, S. and Froelicher, E.S. (2013) 'Secondary analysis of national survey datasets', *Jpn J Nurs Sci*, 10(1), pp. 130-5. doi: 10.1111/j.1742-7924.2012.00213.x.

Booth, H.P., Charlton, J. and Gulliford, M.C. (2017) 'Socioeconomic inequality in morbid obesity with body mass index more than 40kg/m2 in the United States and England', *SSM - Population Health*, 3, pp. 172-178. doi: https://doi.org/10.1016/j.ssmph.2016.12.012.

British Medical Association (BMA) (2016) 'Obesity and diet-related illness in the UK'. Available at: https://archive.bma.org.uk/-/media/files/pdfs/news%20views%20analysis/press%20briefings/obesity%20and%20dietary%20related%20illness%20in%20the%20uk.pdf?la=en (Accessed May 6, 2019).

British Takeaway Campaign (2019) 'Cooking up growth, serving up talent in the takeaway sector'. Available at: https://www.britishtakeawaycampaign.co.uk/wp-content/uploads/2019/08/BTC-report_FINAL.pdf (Accessed January 20, 2021).

Browne, W.J. (2019) 'MCMC Estimation in MLwiN v3.03'. Available at: http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-03/mcmc-print.pdf (Accessed September 19, 2018).

Brug, J. (2008) 'Determinants of healthy eating: motivation, abilities and environmental opportunities', *Fam Pract,* 25 Suppl 1, pp. i50-5. doi:10.1093/fampra/cmn063

Buettner, E. (2008) 'Going for an Indian: South Asian Restaurants and the Limits of Multiculturalism in Britain', *The Journal of Modern History,* 80(4), pp. 865-901. doi:10.1086/591113

Burnett, J. (1989) *Plenty and Want: A Social History of Food in England from 1815 to the Present Day.* 3rd edn. London: Routledge.

Butler, M. J. and Barrientos, R. M. (2020) 'The impact of nutrition on COVID-19 susceptibility and long-term consequences', *Brain Behav Immun,* 87, pp. 53-54. doi:10.1016/j.bbi.2020.04.040

Campbell, M. and Ballas, D. (2013) 'A spatial microsimulation approach to economic policy analysis in Scotland', *Regional Science Policy & Practice,* 5(3), pp. 263-288. doi:10.1111/rsp3.12009

Campbell, M., Smith, D., Baird, J., Vogel, C. and Moon, E. G. (2020) 'A critical review of diet-related surveys in England, 1970-2018', *Archives of Public Health,* 78(1), pp. 66. doi:10.1186/s13690-020-00447-6

Caroline Walker Trust (2009) 'Nutrition policy across the UK: Briefing paper'. Available at: https://www.cwt.org.uk/wp-content/uploads/2014/07/Publichealthpolicyreportfinal.pdf (Accessed February 8, 2018).

Charlton, K., Kowal, P., Soriano, M.M., Williams, S., Banks, E., Vo, K. and Byles, J. (2014) 'Fruit and vegetable intake and body mass index in a large sample of middle-aged Australian men and women', *Nutrients*, 6(6), pp. 2305-2319. doi: 10.3390/nu6062305.

Charlton, C., Rasbash, J., Browne, W.J., Healy, M. and Cameron, B. (2017) *MLwiN* (Version 3.02) [Computer program]. Centre for Multilevel Modelling, University of Bristol. Available at: http://www.bristol.ac.uk/cmm/software/mlwin/download/ (Downloaded September 19, 2018).

Chen, P.-J. and Antonelli, M. (2020) 'Conceptual Models of Food Choice: Influential Factors Related to Foods, Individual Differences, and Society', *Foods (Basel, Switzerland)*, 9(12), p. 1898. doi: 10.3390/foods9121898.

Chromy, J. and Abeyasekera, S. (2005) 'Statistical analysis of survey data ', in *Household sample surveys in developing and transition countries* pp. 389-418 [Online]. Version. Available at: https://unstats.un.org/unsd/hhsurveys/pdf/Household_surveys.pdf (Accessed February 20, 2018).

Clary, C., Matthews, S.A. and Kestens, Y. (2017) 'Between exposure, access and use: Reconsidering foodscape influences on dietary behaviours', *Health & Place*, 44, pp. 1-7. doi: https://doi.org/10.1016/j.healthplace.2016.12.005.

Cornish Pasty Association (2021) *The Mighty Cornish Pasty*. Available at: https://cornishpastyassociation.co.uk/ (Accessed: October 15, 2021).

Corris, V., Dormer, E., Brown, A., Whitty, P., Collingwood, P., Bambra, C. and Newton, J. L. (2020) 'Health inequalities are worsening in the North East of England', *British Medical Bulletin*, 134(1), pp. 63-72. doi:10.1093/bmb/ldaa008

Crozier, S. R., Robinson, S. M., Borland, S. E., Inskip, H. M. and the, S. W. S. S. G. (2006) 'Dietary patterns in the Southampton Women's Survey', *European Journal of Clinical Nutrition,* 60(12), pp. 1391-1399. doi:10.1038/sj.ejcn.1602469

Cruz, J. A. (2000) 'Dietary habits and nutritional status in adolescents over Europe--Southern Europe', *Eur J Clin Nutr,* 54 Suppl 1, pp. S29-35. doi:10.1038/sj.ejcn.1600981

Cummins, S. (2014) 'Food Deserts', *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society*, pp. 562-564.

Cummins, S., Curtis, S., Diez-Roux, A. V. and Macintyre, S. (2007) 'Understanding and representing 'place' in health research: a relational approach', *Soc Sci Med,* 65(9), pp. 1825-38. doi:10.1016/j.socscimed.2007.05.036

Cummins, S. and Macintyre, S. (2002) '"Food deserts"-evidence and assumption in health policy making', *BMJ (Clinical research ed.),* 325(7361), pp. 436-438. doi:10.1136/bmj.325.7361.436

Cummins, S., Macintyre, S., Davidson, S. and Ellaway, A. (2005) 'Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources', *Health & Place*, 11(3), pp. 249-260. doi: https://doi.org/10.1016/j.healthplace.2004.05.003

Cummins, S., Petticrew, M., Higgins, C., Findlay, A. and Sparks, L. (2005) 'Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment', *J Epidemiol Community Health,* 59(12), pp. 1035-40. doi:10.1136/jech.2004.029843

Curtis, S. and Rees Jones, I. (1998) 'Is There a Place for Geography in the Analysis of Health Inequality?', *Sociology of Health & Illness,* 20(5), pp. 645-672. doi: https://doi.org/10.1111/1467-9566.00123

D'Angelo, C., Gloinson, E., Draper, A. and Guthrie, S. (2020) 'Food consumption in the UK: Trends, attitudes and drivers'. Available at: https://www.rand.org/content/dam/rand/pubs/research_reports/RR4300/RR4379/RAND_RR4379.pdf (Accessed: October 4, 2021).

Dahlgren, G. and Whitehead, M. (1991) *Policies and strategies to promote social equity in health*. Available at: https://core.ac.uk/download/pdf/6472456.pdf (Accessed April 10, 2019).

Darmon, N. and Drewnowski, A. (2008) 'Does social class predict diet quality?', *The American Journal of Clinical Nutrition,* 87(5), pp. 1107-1117. doi:10.1093/ajcn/87.5.1107

De Ridder, D., Kroese, F., Evers, C., Adriaanse, M. and Gillebaart, M. (2017) 'Healthy diet: Health impact, prevalence, correlates, and interventions', *Psychology & Health,* 32(8), pp. 907-941. doi:10.1080/08870446.2017.1316849

De Silva, D. and Sooriyarachchi, M. (2012) 'Generalized Linear Multilevel Models for Ordinal Categorical Responses: Methods and Application to Medical Data.', *Sri Lankan Journal of Applied Statistics,* 12, pp. 83–99. doi:http://doi.org/10.4038/sljastats.v12i0.4969

Dearden, E.K., Lloyd, C.D. and Catney, G. (2019) 'A spatial analysis of health status in Britain, 1991–2011', *Social Science & Medicine*, 220, pp. 340-352. doi: https://doi.org/10.1016/j.socscimed.2018.11.014.

Dearden, E.K., Lloyd, C.D. and Green, M. (2020) 'Exploring the histories of health and deprivation in Britain, 1971–2011', *Health & Place*, 61, p. 102255. doi: https://doi.org/10.1016/j.healthplace.2019.102255.

Deas, I., Robson, B., Wong, C. and Bradford, M. (2003) 'Measuring Neighbourhood Deprivation: A Critique of the Index of Multiple Deprivation', *Environment and Planning C: Government and Policy,* 21(6), pp. 883-903. doi:10.1068/c0240

DelaCruz, N. (2010) *Process Monitoring and Quality by Design for Biotechnology Products.* Morgan & Claypool Publishers.

Denoth, F., Scalese, M., Siciliano, V., Di Renzo, L., De Lorenzo, A. and Molinaro, S. (2016) 'Clustering eating habits: frequent consumption of different dietary patterns among the Italian general population in the association with obesity, physical activity, sociocultural characteristics and psychological factors', *Eat Weight Disord,* 21(2), pp. 257-68. doi:10.1007/s40519-015-0225-9

Department for Communities and Local Government (2016) 'English indices of deprivation 2015: Frequently asked questions'. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/579151/English_Indices_of_Deprivation_2015_-_Frequently_Asked_Questions_Dec_2016.pdf (Accessed March 4, 2020).

Department for Environment Food & Rural Affairs (DEFRA) (2017) 'Food Statistics Pocketbook 2016'. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/608426/foodpocketbook-2016report-rev-12apr17.pdf (Accessed July 17, 2019).

Department for Environment Food & Rural Affairs (DEFRA) (2018) 'National Statistics: Family Food 2016/17: Purchases'. Available at: https://www.gov.uk/government/publications/family-food-201617/purchases (Accessed August 29, 2018).

Department for Environment Food & Rural Affairs (DEFRA) (2020) 'Protected geographical food and drink names: UK GI scheme'. Available at: https://www.gov.uk/guidance/protected-geographical-food-and-drink-names-uk-gi-schemes (Accessed: October 12, 2021).

Department for Environment Food & Rural Affairs (DEFRA) (2021) 'Cornish Pasty: Protected food name with Protected Geographical Indication (PGI)'. Available at: https://www.gov.uk/protected-food-drink-names/cornish-pasty (Accessed: October 15. 2021).

D'Innocenzo, S., Biagi, C. and Lanari, M. (2019) 'Obesity and the Mediterranean Diet: A Review of Evidence of the Role and Sustainability of the Mediterranean Diet', *Nutrients,* 11(6), pp. 1306. doi:10.3390/nu11061306

Diez Roux, A.V. (2001) 'Investigating neighborhood and area effects on health', *Am J Public Health*, 91(11), pp. 1783-9. doi: 10.2105/ajph.91.11.1783.

Diez Roux, A.V. (2002) 'Invited commentary: places, people, and health', *Am J Epidemiol*, 155(6), pp. 516-9. doi: 10.1093/aje/155.6.516.

Diez Roux, A.V. and Mair, C. (2010) 'Neighborhoods and health', *Ann N Y Acad Sci*, 1186, pp. 125-45. doi: 10.1111/j.1749-6632.2009.05333.x.

Dong, Y. and Peng, C.-Y. J. (2013) 'Principled missing data methods for researchers', *SpringerPlus,* 2(1), pp. 222-222. doi:10.1186/2193-1801-2-222

Dowler, E. (2008) 'Food and health inequalities: the challenge for sustaining just consumption†', *Local Environment,* 13(8), pp. 759-772. doi:10.1080/13549830802478736

Dye, C. (2014) 'After 2015: infectious diseases in a new era of health and development', *Philos Trans R Soc Lond B Biol Sci,* 369(1645), pp. 20130426. doi:10.1098/rstb.2013.0426

Edwards, K. L. and Clarke, G. P. (2009) 'The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity', *Social Science & Medicine,* 69(7), pp. 1127-1134. doi:https://doi.org/10.1016/j.socscimed.2009.07.037

Emmett, P.M., Tweney, E., Golding, J. and Taylor, C.M. (2017) 'A regional comparison of dietary intakes of 7-year-old children enrolled in observational birth cohort studies on the Isle of Man and in south-west England', *Proceedings of the Nutrition Society*, 76(OCE4), p. E212. doi: 10.1017/S0029665117003743.

Ericson, U., Brunkwall, L., Hellstrand, S., Nilsson, P. M. and Orho-Melander, M. (2019) 'A Health-Conscious Food Pattern Is Associated with Prediabetes and Gut Microbiota in the Malmö Offspring Study', *The Journal of Nutrition,* 150(4), pp. 861-872. doi:10.1093/jn/nxz293

European Food Information Council (2012) 'Parental influence on children's food preferences and energy intake'. Available at: https://www.eufic.org/en/healthy-living/article/parental-influence-on-childrens-food-preferences-and-energy-intake (Accessed August 28, 2018).

Field, A. (2018) *Discovering Statistics Using IBM SPSS Statistics.* SAGE Publications.

Food and Agriculture Organisation (FAO), International Fund for Agricultural Development (IFAD), UNICEF., World Food Programme (WFP) and World Health Organization (WHO) (2021) 'The State of Food Security and Nutrition in the World 2021. Transforming food systems for food security, improved nutrition and affordable healthy diets for all'. doi: 10.4060/cb4474en.

Forouzanfar, M., Alexander, L., Anderson, HR., Bachman, VF., Biryukov, S., Brauer, M….GBD 2013 Risk Factors Collaborators (2015) 'Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013', *The Lancet,* 386, pp. 2287–323. doi:10.1016/S0140-6736(15)00128-2

Freund, F. and Springmann, M. (2021) 'Policy analysis indicates health-sensitive trade and subsidy reforms are needed in the UK to avoid adverse dietary health impacts post-Brexit', *Nature Food*, 2(7), pp. 502-508. doi: 10.1038/s43016-021-00306-9.

Friel, S., Hattersley, L., Snowdon, W., Thow, A.-M., Lobstein, T., Sanders, D., . . . INFORMAS (2013) 'Monitoring the impacts of trade agreements on food environments', *Obesity Reviews*, 14(S1), pp. 120-134. doi: https://doi.org/10.1111/obr.12081.

Frohlich, K.L. (2000) 'The collective lifestyles framework : a contextual analysis of social practices, social structure and disease'. Available at: http://www.santecom.qc.ca/Bibliothequevirtuelle/santecom/35567000034279.pdf (Accessed: October 15, 2021).

Frohlich, K.L., Corin, E. and Potvin, L. (2001) 'A theoretical proposal for the relationship between context and disease', *Sociology of Health & Illness*, 23(6), pp. 776-797. doi: https://doi.org/10.1111/1467-9566.00275.

Frohlich, K.L. and Potvin, L. (1999) 'Collective lifestyles as the target for health promotion', *Can J Public Health*, 90 Suppl 1(Suppl 1), pp. S11-4. doi: 10.1007/bf03403571.

Fulgoni, V. and Drewnowski, A. (2019) 'An Economic Gap Between the Recommended Healthy Food Patterns and Existing Diets of Minority Groups in the US National Health and Nutrition Examination Survey 2013–14', *Frontiers in Nutrition,* 6(37). doi:10.3389/fnut.2019.00037

Fulkerson, J.A., Larson, N., Horning, M. and Neumark-Sztainer, D. (2014) 'A review of associations between family or shared meal frequency and dietary and weight status outcomes across the lifespan', *J Nutr Educ Behav*, 46(1), pp. 2-19. doi: 10.1016/j.jneb.2013.07.012.

Fuller, E., Bankiewicz, U., Davies, B., Mandalia, D. and Stocker, B. (2019) 'The Food and You Survey Wave 5: Combined report for England, Wales and Northern Irelan'. Available at: https://www.food.gov.uk/sites/default/files/media/document/food-and-you-wave-5-combined-report.pdf (Accessed: October 12, 2021).

Gadsby, E. W., Peckham, S., Coleman, A., Bramwell, D., Perkins, N. and Jenkins, L. M. (2017) 'Commissioning for health improvement following the 2012 health and social care reforms in England: what has changed?', *BMC Public Health,* 17(1), pp. 211. doi:10.1186/s12889-017-4122-1

Gakidou, E., Afshin, A., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., ... GBD 2016 Risk Factors Collaborators. (2017) 'Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016', *The Lancet,* 390(10100), pp. 1345-1422. doi:10.1016/S0140-6736(17)32366-8

Galobardes, B., Shaw, M., Lawlor, D., Lynch, J. and Smith, G. D. (2006) 'Indicators of socioeconomic position (part 1)', *Journal of Epidemiology and Community Health,* 60, pp. 12 - 7. doi:10.1136/jech.2004.023531

Garson, G. (2015) 'Missing values analysis & data imputation'. Available at: http://www.statisticalassociates.com/missingvaluesanalysis_p.pdf (Accessed March 4, 2020).

Gelman, A. (2007) 'Struggles with survey weighting and regression modeling', *Statistical Science* 22(2), pp. 184-188. doi:10.1214/088342306000000691

Glanz, K., Metcalfe, J.J., Folta, S.C., Brown, A. and Fiese, B. (2021) 'Diet and Health Benefits Associated with In-Home Eating and Sharing Meals at Home: A Systematic Review', *International journal of environmental research and public health*, 18(4), p. 1577. doi: 10.3390/ijerph18041577.

Global Food Security (2016) 'Insight: Overconsumption and influences on diet'. Available at: https://www.foodsecurity.ac.uk/publications/ (Accessed July 16, 2019).

Global Panel on Agriculture and Food Systems for Nutrition (2016) 'Food systems and diets: Facing the challenges of the 21st century'. Available at: http://glopan.org/sites/default/files/ForesightReport.pdf (Accessed June 1, 2018).

Gong, H. and Hassink, R. (2020) 'Context sensitivity and economic-geographic (re)theorising', *Cambridge Journal of Regions, Economy and Society*, 13(3), pp. 475-490. doi: 10.1093/cjres/rsaa021.

Graham, J. W., Olchowski, A. E. and Gilreath, T. D. (2007) 'How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory', *Prevention Science,* 8(3), pp. 206-213. doi:10.1007/s11121-007-0070-9

Grimaldi, K. and Paoli, A. (2015) 'Chapter 10 - Genomic Determinants of Mediterranean Diet Success', in Preedy, V.R. and Watson, R.R. (eds.) *The Mediterranean Diet*. San Diego: Academic Press, pp. 105-113.

Harland, K., Heppenstall, A., Smith, D. and Birkin, M. (2012) 'Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques', *Journal of Artificial Societies and Social Simulation,* 15(1), pp. 1. doi:10.18564/jasss.1909

Hawkes, C. (2006) 'Uneven dietary development: linking the policies and processes of globalization with the nutrition transition, obesity and diet-related chronic diseases', *Globalization and Health*, 2(1), p. 4. doi: 10.1186/1744-8603-2-4.

Hawkes, C., Harris, J. and Gillespie, S. (2017) 'Urbanization and the Nutrition Transition', *Global Food Policy Report,* (4), pp. 34-41. doi:10.2499/9780896292529

Hawkesworth, S., Dangour, A. D., Johnston, D., Lock, K., Poole, N., Rushton, J., . . . Waage, J. (2010) 'Feeding the world healthily: the challenge of measuring the effects of agriculture on health', *Philos Trans R Soc Lond B Biol Sci,* 365(1554), pp. 3083-97. doi:10.1098/rstb.2010.0122

Hearty, Á. P. and Gibney, M. J. (2008) 'Comparison of cluster and principal component analysis techniques to derive dietary patterns in Irish adults', *British Journal of Nutrition,* 101(4), pp. 598-608. doi:10.1017/S0007114508014128

Heinze, G. and Dunkler, D. (2017) 'Five myths about variable selection', *Transplant International*, 30(1), pp. 6-10. doi: https://doi.org/10.1111/tri.12895.

Heo, M., Kim, R.S., Wylie-Rosett, J., Allison, D.B., Heymsfield, S.B. and Faith, M.S. (2011) 'Inverse association between fruit and vegetable intake and BMI even after controlling for demographic, socioeconomic and lifestyle factors', *Obesity facts*, 4(6), pp. 449-455. doi: 10.1159/000335279.

Hermann, S., Rohrmann, S., Linseisen, J., May, A.M., Kunst, A., Besson, H., . . . Peeters, P.H.M. (2011) 'The association of education with body mass index and waist circumference in the EPIC-PANACEA study', *BMC Public Health*, 11(1), p. 169. doi: 10.1186/1471-2458-11-169.

Higgins, V., Nazroo, J. and Brown, M. (2019) 'Pathways to ethnic differences in obesity: The role of migration, culture and socio-economic position in the UK', *SSM - population health,* 7, pp. 100394-100394. doi:10.1016/j.ssmph.2019.100394

Holt, G. and Amilien, V. (2007) 'Introduction: From local food to localised food '. Available at: https://journals.openedition.org/aof/405 (Accessed: October 7, 2021).

Howarth, N.C., Saltzman, E. and Roberts, S.B. (2001) 'Dietary fiber and weight regulation', *Nutr Rev*, 59(5), pp. 129-39. doi: 10.1111/j.1753-4887.2001.tb07001.x.

Hox, J. (2010) *Multilevel analysis: Techniques and applications.* 2nd edn.: Routledge/Taylor & Francis Group.

Hu, F. B. (2002) 'Dietary pattern analysis: a new direction in nutritional epidemiology', *Curr Opin Lipidol,* 13(1), pp. 3-9. doi:10.1097/00041433-200202000-00002

Hu, F. B., Rimm, E., Smith-Warner, S. A., Feskanich, D., Stampfer, M. J., Ascherio, A., . . . Willett, W. C. (1999) 'Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire', *Am J Clin Nutr,* 69(2), pp. 243-9. doi:10.1093/ajcn/69.2.243

IBM (2018) *IBM SPSS Statistics for Windows* (Version 24) [Computer program]. IBM Corp.

Imamura, F., Micha, R., Khatibzadeh, S., Fahimi, S., Shi, P., Powles, J. and Mozaffarian, D. (2015) 'Dietary quality among men and women in 187 countries in 1990 and 2010: a systematic assessment', *The Lancet Global Health,* 3(3), pp. e132-e142. doi:10.1016/s2214-109x(14)70381-x

Jakobsen, J. C., Gluud, C., Wetterslev, J. and Winkel, P. (2017) 'When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts', *BMC Medical Research Methodology,* 17(1), pp. 162. doi:10.1186/s12874-017-0442-1

James, W. H. M., Lomax, N. and Birkin, M. (2019) 'Local level estimates of food, drink and tobacco expenditure for Great Britain', *Scientific Data,* 6(1), pp. 56. doi:10.1038/s41597-019-0064-z

Janssen, H. G., Davies, I. G., Richardson, L. and Stevenson, L. (2018) 'Dietary patterns of takeaway (fast) food consumers in the North West of England: A Pilot study', *Proceedings of the Nutrition Society,* 77(OCE1), pp. E18. doi:10.1017/S0029665117004372

Jebb, S. A. (2018) 'Interventions to accelerate change towards a healthier diet', *Proc Nutr Soc,* 77(2), pp. 106-111. doi:10.1017/S0029665117004086

Jenneson, V., Morris, M., Greenwood, D. and Clarke, G. (2021) 'STRIDE Study: Supermarket Transaction Records In Dietary Evaluation'. doi: 10.17605/OSF.IO/VUKTQ

Jia, L., Lu, H., Wu, J., Wang, X., Wang, W., Du, M., . . . Zhang, N. (2020) 'Association between diet quality and obesity indicators among the working-age adults in Inner Mongolia, Northern China: a cross-sectional study', *BMC Public Health*, 20(1), p. 1165. doi: 10.1186/s12889-020-09281-5.

Jones, N. R., Tong, T. Y. and Monsivais, P. (2018) 'Meeting UK dietary recommendations is associated with higher estimated consumer food costs: an analysis using the National Diet and Nutrition Survey and consumer expenditure data, 2008-2012', *Public Health Nutr,* 21(5), pp. 948-956. doi:10.1017/S1368980017003275

Judd, S. E., Letter, A. J., Shikany, J. M., Roth, D. L. and Newby, P. K. (2015) 'Dietary Patterns Derived Using Exploratory and Confirmatory Factor Analysis are Stable and Generalizable Across Race, Region, and Gender Subgroups in the REGARDS Study', *Frontiers in nutrition,* 1, pp. 29-29. doi:10.3389/fnut.2014.00029

Keeble, M., Adams, J., Sacks, G., Vanderlee, L., White, C. M., Hammond, D. and Burgoine, T. (2020) 'Use of Online Food Delivery Services to Order Food Prepared Away-From-Home and Associated Sociodemographic Characteristics: A Cross-Sectional, Multi-Country Analysis', *International journal of environmental research and public health,* 17(14), pp. 5190. doi:10.3390/ijerph17145190

Kegler, M.C., Hermstad, A. and Haardörfer, R. (2021) 'Home food environment and associations with weight and diet among U.S. adults: a cross-sectional study', *BMC Public Health*, 21(1), p. 1032. doi: 10.1186/s12889-021-11102-2.

Kestens, Y., Lebel, A., Daniel, M., Thériault, M. and Pampalon, R. (2010) 'Using experienced activity spaces to measure foodscape exposure', *Health & Place*, 16(6), pp. 1094-1103. doi: https://doi.org/10.1016/j.healthplace.2010.06.016.

Kyriacou, A., Evans, J. M., Economides, N. and Kyriacou, A. (2015) 'Adherence to the Mediterranean diet by the Greek and Cypriot population: a systematic review', *Eur J Public Health,* 25(6), pp. 1012-8. doi:10.1093/eurpub/ckv124

Leckie, G. (2013) *Cross-classified Multilevel Models-Concepts. LEMMA VLE*. Available at: http://www.bristol.ac.uk/cmm/learning/course.html (Accessed July 31, 2019).

Lee, J., Alnasrallah, M., Wong, D., Beaird, H. and Logue, E. (2014) 'Impacts of Scale on Geographic Analysis of Health Data: An Example of Obesity Prevalence', *ISPRS International Journal of Geo-Information*, 3(4), pp. 1198-1210. doi: https://doi.org/10.3390/ijgi3041198.

Liu, Y. and De, A. (2015) 'Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study', *International journal of statistics in medical research,* 4(3), pp. 287-295. doi:10.6000/1929-6029.2015.04.03.7

Lomax, N. and Norman, P. (2016) 'Estimating Population Attribute Values in a Table: “Get Me Started in” Iterative Proportional Fitting', *The Professional Geographer,* 68(3), pp. 451-461. doi:10.1080/00330124.2015.1099449

Lovelace, R., Birkin, M., Ballas, D. and van Leeuwen, E. (2015) 'Evaluating the Performance of Iterative Proportional Fitting for Spatial Microsimulation: New Tests for an Established Technique', *Journal of Artificial Societies and Social Simulation,* 18(2), pp. 21. doi:10.18564/jasss.2768

Macintyre, S., Ellaway, A. and Cummins, S. (2002) 'Place effects on health: how can we conceptualise, operationalise and measure them?', *Soc Sci Med*, 55(1), pp. 125-39. doi: 10.1016/s0277-9536(01)00214-3.

Macintyre, S., Maciver, S. and Sooman, A. (1993) 'Area, Class and Health: Should we be Focusing on Places or People?', *Journal of Social Policy,* 22(02), pp. 213-234. doi:10.1017/s0047279400019310

Mackley, A., Hobson, F. and McInnes, R. (2021) *Coronavirus: Legacy benefits and the Universal Credit 'uplift'* House of Commons Library. Available at: https://researchbriefings.files.parliament.uk/documents/CBP-9246/CBP-9246.pdf (Accessed: October 12, 2021).

Madley-Dowd, P., Hughes, R., Tilling, K. and Heron, J. (2019) 'The proportion of missing data should not be used to guide decisions on multiple imputation', *Journal of Clinical Epidemiology*, 110, pp. 63-73. doi: https://doi.org/10.1016/j.jclinepi.2019.02.016.

Maguire, E. R., Burgoine, T. and Monsivais, P. (2015) 'Area deprivation and the food environment over time: A repeated cross-sectional study on takeaway outlet density and supermarket presence in Norfolk, UK, 1990–2008', *Health & Place,* 33, pp. 142-147. doi:https://doi.org/10.1016/j.healthplace.2015.02.012

Maguire, E. R. and Monsivais, P. (2015) 'Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data', *Br J Nutr,* 113(1), pp. 181-9. doi:10.1017/S0007114514002621

Maksimov, S., Karamnova, N., Shalnova, S. and Drapkina, O. (2020) 'Sociodemographic and Regional Determinants of Dietary Patterns in Russia', *International Journal of Environmental Research and Public Health,* 17(1), pp. 328. doi:10.3390/ijerph17010328

Malhotra, A., Noakes, T. and Phinney, S. (2015) 'It is time to bust the myth of physical inactivity and obesity: you cannot outrun a bad diet', *Br J Sports Med,* 49(15), pp. 967-8. doi:10.1136/bjsports-2015-094911

Marchioni, D. M. L., Latorre, M. d. R. D. d. O., Eluf-Neto, J., Wünsch-Filho, V. and Fisberg, R. M. (2005) 'Identification of dietary patterns using factor analysis in an epidemiological study in São Paulo', *Sao Paulo Medical Journal,* 123, pp. 124-127. doi:10.1590/S1516-31802005000300007

McCurdy, C. (2019) 'Ageing, fast and slow: When place and demography collide'. Available at: https://www.resolutionfoundation.org/publications/ageing-fast-and-slow/ (Accessed February 23, 2021).

McFadden, A., Green, J. M., Williams, V., McLeish, J., McCormick, F., Fox-Rushby, J. and Renfrew, M. J. (2014) 'Can food vouchers improve nutrition and reduce health inequalities in low-income mothers and young children: a multi-method evaluation of the experiences of beneficiaries and practitioners of the Healthy Start programme in England', *BMC Public Health,* 14, pp. 148. doi:10.1186/1471-2458-14-148

McGuinness, F. and Harari, D. (2019) 'Income inequality in UK', 7484. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-7484/ (Accessed May 11, 2021).

McLachlan, G. and Norman, P. (2020) 'Analysing Socio-Economic Change Using a Time Comparable Geodemographic Classification: England and Wales, 1991–2011', *Applied Spatial Analysis and Policy*. doi:10.1007/s12061-020-09346-3

McNeish, D. (2017) 'Missing data methods for arbitrary missingness with small samples', *Journal of Applied Statistics*, 44(1), pp. 24-39. doi: 10.1080/02664763.2016.1158246.

Medical Research Council (MRC) (2014) *Maximising the value of UK population cohorts MRC: Strategic Review of the Largest UK Population Cohort Studies*. Available at: https://mrc.ukri.org/publications/browse/maximising-the-value-of-uk-population-cohorts/ (Accessed June 8, 2020).

Mela, D.J. (2001) 'Determinants of Food Choice: Relationships with Obesity and Weight Control', *Obesity Research*, 9(S11), pp. 249S-255S. doi: https://doi.org/10.1038/oby.2001.127.

Micha, R., Coates, J., Leclercq, C., Charrondiere, U. R. and Mozaffarian, D. (2018) 'Global Dietary Surveillance: Data Gaps and Challenges', *Food Nutr Bull,* 39(2), pp. 175-205. doi:10.1177/0379572117752986

Miller, P. E., Lazarus, P., Lesko, S. M., Muscat, J. E., Harper, G., Cross, A. J., . . . Hartman, T. J. (2010) 'Diet index-based and empirically derived dietary patterns are associated with colorectal cancer risk', *The Journal of nutrition,* 140(7), pp. 1267-1273. doi:10.3945/jn.110.121780

Ministry of Housing, Communities & Local Government (2019) 'The English Indices of Deprivation 2019 (IoD2019)'. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/835115/IoD2019_Statistical_Release.pdf (Accessed April 14, 2021).

Moeller, S. M., Reedy, J., Millen, A. E., Dixon, L. B., Newby, P. K., Tucker, K. L., . . . Guenther, P. M. (2007) 'Dietary Patterns: Challenges and Opportunities in Dietary Patterns Research: An Experimental Biology Workshop, April 1, 2006', *Journal of the American Dietetic Association,* 107(7), pp. 1233-1239. doi:https://doi.org/10.1016/j.jada.2007.03.014

Mollborn, S., Lawrence, E.M. and Saint Onge, J.M. (2021) 'Contributions and Challenges in Health Lifestyles Research', *Journal of Health and Social Behavior*, 62(3), pp. 388-403. doi: 10.1177/0022146521997813.

Monsivais, P., Aggarwal, A. and Drewnowski, A. (2012) 'Are socio-economic disparities in diet quality explained by diet cost?', *J Epidemiol Community Health,* 66(6), pp. 530-5. doi:10.1136/jech.2010.122333

Monteiro, C. A. (2009) 'Nutrition and health. The issue is not food, nor nutrients, so much as processing', *Public Health Nutr,* 12(5), pp. 729-31. doi:10.1017/S1368980009005291

Monteiro, C. A., Cannon, G., Levy, R., Moubarac, J.-C., Jaime, P., Martins, A. P., . . . Sattamini, I. (2016) 'NOVA. The star shines bright', *World Nutrition*, 7(1-3), pp. 28-38. Available at: https://worldnutritionjournal.org/index.php/wn/article/view/5/4 (Accessed August 7, 2018).

Monteiro, C. A., Moubarac, J. C., Levy, R. B., Canella, D. S., Louzada, M. and Cannon, G. (2018) 'Household availability of ultra-processed foods and obesity in nineteen European countries', *Public Health Nutr,* 21(1), pp. 18-26. doi:10.1017/S1368980017001379

Monterrosa, E.C., Frongillo, E.A., Drewnowski, A., de Pee, S. and Vandevijvere, S. (2020) 'Sociocultural Influences on Food Choices and Implications for Sustainable Healthy Diets', *Food and Nutrition Bulletin*, 41(2_suppl), pp. 59S-73S. doi: 10.1177/0379572120975874.

Moon, G., Quarendon, G., Barnard, S., Twigg, L. and Blyth, B. (2007) 'Fat nation: deciphering the distinctive geographies of obesity in England', *Soc Sci Med,* 65(1), pp. 20-31. doi:10.1016/j.socscimed.2007.02.046

Moon, G., Twigg, L., Jones, K., Aitken, G. and Taylor, J. (2019) 'The utility of geodemographic indicators in small area estimates of limiting long-term illness', *Social Science & Medicine,* 227, pp. 47-55. doi:https://doi.org/10.1016/j.socscimed.2018.06.029

Moretti, A. and Whitworth, A. (2021) 'Estimating the Uncertainty of a Small Area Estimator Based on a Microsimulation Approach', *Sociological Methods & Research*. doi:10.1177/0049124120986199

Morris, M. A., Clarke, G. P., Edwards, K. L., Hulme, C. and Cade, J. E. (2016) 'Geography of Diet in the UK Women's Cohort Study: A Cross-Sectional Analysis', *Epidemiology - Open Journal,* 1(1), pp. 20-32. doi:10.17140/epoj-1-104

Mozaffarian, D., Angell, S. Y., Lang, T. and Rivera, J. A. (2018) 'Role of government policy in nutrition-barriers to and opportunities for healthier eating', *BMJ,* 361, pp. k2426. doi:10.1136/bmj.k2426

MRC Elsie Widdowson Laboratory and NatCen Social Research (2018) 'National Diet and Nutrition Survey Years 1-8, 2008/09-2015/16. [data collection].' Available at: https://www.ukdataservice.ac.uk/ (Accessed March 24, 2021).

Musingarimi, P. (2009) 'Obesity in the UK: a review and comparative analysis of policies within the devolved administrations', *Health Policy,* 91(1), pp. 10-6. doi:10.1016/j.healthpol.2008.11.004

Narzisi, K. and Simons, J. (2021) 'Interventions that prevent or reduce obesity in children from birth to five years of age: A systematic review', *J Child Health Care*, 25(2), pp. 320-334. doi:10.1177/1367493520917863

NatCen Social Research and University College London Department of Epidemiology and Public Health (2008-2016) 'Health Survey for England, 2008-2016 [data collection]'. Available at: https://www.ukdataservice.ac.uk/ (Accessed April 11, 2021).

NatCen Social Research and University College of London (UCL) (2017) 'Health Survey for England 2016 Methods'. Available at: http://healthsurvey.hscic.gov.uk/media/63778/HSE2016-Methods-text.pdf (Accessed May 11, 2021).

National Academies of Sciences, Engineering & Medicine (2019) *Sustainable Diets, Food, and Nutrition: Proceedings of a Workshop* Washington, DC: The National Academies Press. Available at: https://www.nap.edu/catalog/25192/sustainable-diets-food-and-nutrition-proceedings-of-a-workshop (Accessed July 19, 2018).

National Health Service (NHS) (2021) 'Get help to buy food and milk: The Healthy Start scheme'. Available at: https://www.healthystart.nhs.uk/ (Accessed July 8, 2021).

National Institutes of Health (NIH) (2019) 'Eating highly processed foods linked to weight gain'. Available at: https://www.nih.gov/news-events/nih-research-matters/eating-highly-processed-foods-linked-weight-gain (Accessed: October 14, 2021).

National Obesity Observatory (NOO) (2011) 'Obesity and ethnicity'. Available at: http://webarchive.nationalarchives.gov.uk/20170110170207/http://www.noo.org.uk/NOO_about_obesity/inequalities (Accessed August 29, 2018).

Newton, J. N., Briggs, A. D. M., Murray, C. J. L., Dicker, D., Foreman, K. J., Wang, H., . . . Davis, A. C. J. (2015) 'Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013', *The Lancet,* 386(10010), pp. 2257-2274. doi:10.1016/s0140-6736(15)00195-6

NHS Digital (2017) 'Statistics on Obesity, Physical Activity and Diet, England: 2017'. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/613532/obes-phys-acti-diet-eng-2017-rep.pdf (Accessed May 31, 2018).

NHS Digital (2019) 'Statistics on Obesity, Physical Activity and Diet, England, 2019'. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet/statistics-on-obesity-physical-activity-and-diet-england-2019 (Accessed May 5, 2021).

NHS Digital (2020) 'Statistics on Obesity, Physical Activity and Diet, England, 2020'. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet/england-2020 (Accessed: October 14, 2021).

Nour, M., Lutze, S.A., Grech, A. and Allman-Farinelli, M. (2018) 'The Relationship between Vegetable Intake and Weight Outcomes: A Systematic Review of Cohort Studies', *Nutrients*, 10(11), p. 1626. doi: 10.3390/nu10111626

Ocké, M. C. (2013) 'Evaluation of methodologies for assessing the overall diet: dietary quality scores and dietary pattern analysis', *Proceedings of the Nutrition Society,* 72(2), pp. 191-199. doi:10.1017/S0029665113000013

O'Connell, R., Owen, C., Padley, M., Simon, A. and Brannen, J. (2018) 'Which Types of Family are at

Risk of Food Poverty in the UK? A Relative Deprivation Approach', *Social Policy and Society*, pp. 1-18. doi:10.1017/s1474746418000015

Oddy, D. (2003) *From Plain Fare to Fusion Food: British Diet from the 1890s to the 1990s* Great Britain: Boydell Press.

Office of National Statistics (ONS) (2006) 'Health Survey for England - 2004, Health of Ethnic Minorities, Main report'. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2004-health-of-ethnic-minorities-main-report (Accessed May 5, 2021).

Office of National Statistics (ONS) (2011) '2011 Census: Local characteristics'. Available at: https://www.nomisweb.co.uk/sources/census_2011_lc (Accessed February 23, 2021).

Office of National Statistics (ONS) (2017a) 'Overview of the UK population: July 2017'. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/july2017 (Accessed March 4, 2020).

Office of National Statistics (ONS) (2017b) 'Model-Based Estimates of households in poverty for Middle Layer Super Output Areas, 2011/12 Technical Report'. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/modelbasedestimatesofhouseholdsinpovertyformiddlelayersuperoutputareas201112technicalreport (Accessed March 14. 2019).

Office of National Statistics (ONS) (2018a) 'Regional ethnic diversity'. Available at: https://www.ethnicity-facts-figures.service.gov.uk/british-population/national-and-regional-populations/regional-ethnic-diversity/latest (Accessed June 6, 2019).

Office of National Statistics (ONS) (2018b) 'Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland'. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland (Accessed July 26, 2019).

Office of National Statistics (ONS) (2019) 'The National Statistics Socio-economic classification (NS-SEC)'. Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 (Accessed May 11, 2021).

Office of National Statistics (ONS) (2020a) 'Income estimates for small areas, England and Wales: financial year ending 2018'. Available at: https://www.ons.gov.uk/releases/smallareamodelbasedincomeestimatesenglandandwalesfinancialyearending2018 (Accessed May 11, 2021).

Office of National Statistics (ONS) (2020b) 'Qualifications of Working Age Population (NVQ)'. Available at: https://data.london.gov.uk/dataset/qualifications-working-age-population-nvq-borough (Accessed February 23, 2021).

Ogce, F., Ceber, E., Ekti, R. and Oran, N. T. (2008) 'Comparison of Mediterranean, Western and Japanese diets and some recommendations', *Asian Pac J Cancer Prev,* 9(2), pp. 351-6.

Openshaw, S. (1984) *The modifiable areal unit problem.* Norfolk: Geo Books.

Organisation for Economic Co-operation and Development (OECD) (2021) 'Obesity and the economics of prevention: Fit not fat-United Kingdom (England) Key Facts'. Available at: https://www.oecd.org/els/health-systems/obesityandtheeconomicsofpreventionfitnotfatunitedkingdomenglandkeyfacts.htm (Accessed: October 14, 2021).

Otter, C. (2012) 'The British Nutrition Transition and its Histories', *History Compass,* 10, pp. 812-825. doi:10.1111/hic3.12001

Oyebode, O., Gordon-Dseagu, V., Walker, A. and Mindell, J. S. (2014) 'Fruit and vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health Survey for England data', *J Epidemiol Community Health,* 68(9), pp. 856-62. doi:10.1136/jech-2013-203500

Pem, D. and Jeewon, R. (2015) 'Fruit and Vegetable Intake: Benefits and Progress of Nutrition Education Interventions- Narrative Review Article', *Iranian journal of public health,* 44(10), pp. 1309-1321.

Perchoux, C., Chaix, B., Cummins, S. and Kestens, Y. (2013) 'Conceptualization and measurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility', *Health Place,* 21, pp. 86-93. doi:10.1016/j.healthplace.2013.01.005

Pickett, K.E. and Pearl, M. (2001) 'Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review', *J Epidemiol Community Health*, 55(2), pp. 111-22. doi: 10.1136/jech.55.2.111.

Pinho-Gomes, A.-C., Knight, A., Critchley, J. and Pennington, M. (2021) 'Addressing the low consumption of fruit and vegetables in England: a cost-effectiveness analysis of public policies', *Journal of Epidemiology and Community Health,* 75(3), pp. 282-288. doi:10.1136/jech-2020-214081

Popkin, B. (1993) 'Nutritional Patterns and Transitions', *Population and Development Review,* 19(1), pp. 138-157. doi:10.2307/2938388

Popkin, B. (2002) 'An overview on the nutrition transition and its health implications: the Bellagio meeting', *Public Health Nutrition,* 5(1a). doi:10.1079/phn2001280

Popkin, B. (2006) 'Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases', *The American Journal of Clinical Nutrition,* 84(2), pp. 289-298. doi:10.1093/ajcn/84.2.289

Popkin, B., Adair, L. and Ng, S. W. (2012) 'Global nutrition transition and the pandemic of obesity in developing countries', *Nutr Rev,* 70(1), pp. 3-21. doi:10.1111/j.1753-4887.2011.00456.x

Popkin, B. and Ng, S. W. (2007) 'The nutrition transition in high- and low-income countries: what are the policy lessons?', *Agricultural Economics,* 37(s1), pp. 199-211. doi:https://doi.org/10.1111/j.1574-0862.2007.00245.x

Potvin, L. and Hayes, M.V. (2007) 'Place and health research in Canada', *Canadian journal of public health=Revue canadienne de sante publique*, 98 Suppl 1(Suppl 1), pp. S6-S9. doi: 10.1007/BF03403720.

Prouse, V., Ramos, H., Grant, J. L. and Radice, M. (2014) 'How and when Scale Matters: The Modifiable Areal Unit Problem and income inequality in Halifax', *Canadian Journal of Urban Research*, 23(1), pp. 61-82. Available at: http://www.jstor.org/stable/26195268 (Accessed July 8, 2021).

Pryce, R., Angus, C., Holmes, J., Gillespie, D., Buykx, P., Meier, P., . . . Brennan, A. (2020) 'Reweighting national survey data for small area behaviour estimates: modelling alcohol consumption in Local Authorities in England', *Population Health Metrics,* 18(1), pp. 1.

doi:10.1186/s12963-019-0201-0

Public Health England (PHE) (2017) 'Health matters: obesity and the food environment'. Available at: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2 (Accessed July 10, 2019).

Public Health England (PHE) (2018a) 'PHE publishes latest data on nation's diet', *Public Health England has today published new data detailing the state of the nation's diet.* Available at: https://www.gov.uk/government/news/phe-publishes-latest-data-on-nations-diet (Accessed July 10, 2019).

Public Health England (PHE) (2018b) 'The Eatwell Guide: Helping you eat a healthy, balanced diet'. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/742750/Eatwell_Guide_booklet_2018v4.pdf (Accessed July 10, 2019).

Public Health England (PHE) (2018c) 'Density of fast food outlets in England by local authority and ward: data tables and metadata'. Available at: https://www.gov.uk/government/publications/fast-food-outlets-density-by-local-authority-in-england (Accessed July 10, 2019).

Public Health England (PHE) (2018d) 'England's poorest areas are fast food hotspots: New figures from PHE show higher concentrations of fast food outlets in England's most deprived communities'. Available at: https://www.gov.uk/government/news/englands-poorest-areas-are-fast-food-hotspots (Accessed July 10, 2019).

Public Health England (PHE) (2020a) 'Healthy eating among adults'. Available at: https://www.ethnicity-facts-figures.service.gov.uk/health/diet-and-exercise/healthy-eating-of-5-a-day-among-adults/latest (Accessed May 11, 2021).

Public Health England (PHE) (2020b) 'Major new campaign encourages millions to lose weight and cut COVID-19 risk', *Public Health England has today published new data detailing the state of the nation's diet.* Available at: https://www.gov.uk/government/news/major-new-campaign-encourages-millions-to-lose-weight-and-cut-covid-19-risk (Accessed May 11, 2021).

R Development Core Team (2018) *R: A language and environment for statistical computing* (Version 3.4.4) [Computer program]. R Foundation for statistical computing. Available at: https://www.r-project.org/ (Downloaded August 24, 2020).

Rahman, A. (2017) 'Estimating small area health-related characteristics of populations: a methodological review', *Geospat Health,* 12(1), pp. 495. doi:10.4081/gh.2017.495

Rahman, A., Harding, A., Tanton, R. and Liu, S. (2010) 'Methodological Issues in Spatial Microsimulation Modelling for Small Area Estimation', *The International Journal of Microsimulation,* 3(2), pp. 3-22. doi:10.34196/IJM.00035

Rauber, F., Chang, K., Vamos, E.P., da Costa Louzada, M.L., Monteiro, C.A., Millett, C. and Levy, R.B. (2021) ' Ultra-processed food consumption and risk of obesity: a prospective cohort study of UK Biobank', *Eur J Nutr*, 60(4), pp. 2169-2180. doi: 10.1007/s00394-020-02367-1.

Rayner, G., Hawkes, C., Lang, T. and Bello, W. (2006) 'Trade liberalization and the diet transition: a public health response', *Health Promot Int*, 21 Suppl 1, pp. 67-74. doi: 10.1093/heapro/dal053.

Renzella, J., Townsend, N., Jewell, J., Breda, J., Roberts, N., Rayner, M. and Wickramasinghe, K. (2018) 'What national and subnational interventions and policies based on Mediterranean and Nordic diets are recommended or implemented in the WHO European Region, and is there evidence of effectiveness in reducing noncommunicable diseases?', *Health Evidence Network synthesis report 58*. Available at: http://www.euro.who.int/__data/assets/pdf_file/0011/365285/hen-58-eng.pdf (Accessed July 11, 2019).

Riley, H. (2010) 'Potato consumption in the UK – why is 'meat and two veg' no longer the traditional British meal?', *Nutrition Bulletin,* 35, pp. 320–331. doi:https://doi.org/10.1111/j.1467-3010.2010.01864.x

Riva, M., Gauvin, L. and Barnett, T. A. (2007) 'Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998', *J Epidemiol Community Health,* 61(10), pp. 853-61. doi:10.1136/jech.2006.050740

Riva, M. and Smith, D. M. (2012) 'Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method', *Soc Psychiatry Psychiatr Epidemiol,* 47(5), pp. 745-55. doi:10.1007/s00127-011-0376-6

Roberts, C. (2014) 'Health Survey for England 2013: Fruit and vegetable consumption', 1(7). Available at: https://files.digital.nhs.uk/publicationimport/pub16xxx/pub16076/hse2013-ch7-fru-veg-com.pdf (Accessed July 17, 2019).

Roberts, C., Steer, T., Maplethorpe, N., Cox, L., Meadows, S., Nicholson, S., . . . Swan, G. (2018) 'Appendix R: Main and subsidiary food groups and disaggregation categories. ', in *National Diet and Nutrition Survey. Headline results from Years 7-8 (combined) of the Rolling Programme (2014/2015 –2015/16)* [Online]. Version. Available at: https://www.gov.uk/government/statistics/ndns-results-from-years-7-and-8-combined (Accessed May 21, 2020).

Roberts, K., Cade, J., Dawson, J. and Holdsworth, M. (2018) 'Empirically Derived Dietary Patterns in UK Adults Are Associated with Sociodemographic Characteristics, Lifestyle, and Diet Quality', *Nutrients,* 10(2). doi:10.3390/nu10020177

Roberts, K., Cavill, N., Hancock, C. and Rutter, H. (2013) 'Social and economic inequalities in diet and physical activity'. Available at: http://www.oneeastmidlands.org.uk/sites/default/files/library/vid_19253_Social_and_economic_inequalities_in_diet_and_physical_activity_04.11.13.pdf (Accessed June 6, 2019).

Robinson, S. M., Crozier, S. R., Borland, S. E., Hammond, J., Barker, D. J. and Inskip, H. M. (2004) 'Impact of educational attainment on the quality of young women's diets', *Eur J Clin Nutr,* 58(8), pp. 1174-80. doi:10.1038/sj.ejcn.1601946

Roos, G., Johansson, L., Kasmel, A., Klumbiené, J. and Prättälä, R. (2000) 'Disparities in vegetable and fruit consumption: European cases from the north to the south', *Public Health Nutrition,* 4(1), pp. 35-43. doi:10.1079/phn200048

Rumm-Kreuter, D. (2001) 'Comparison of the eating and cooking habits of northern Europe and the Mediterranean countries in the past, present and future', *Int J Vitam Nutr Res,* 71(3), pp. 141-8. doi:10.1024/0300-9831.71.3.141

Sánchez-Villegas, A. and Martínez-Lapiscina, E. H. (2018) 'Chapter 11 - A Healthy Diet for Your Heart and Your Brain', in Sánchez-Villegas, A. and Sánchez-Tainta, A. (eds.) *The Prevention of Cardiovascular Disease Through the Mediterranean Diet*: Academic Press, pp. 169-197.

Saunders, P., Saunders, A. and Middleton, J. (2015) 'Living in a 'fat swamp': exposure to multiple sources of accessible, cheap, energy-dense fast foods in a deprived community', *British Journal of Nutrition,* 113(11), pp. 1828-1834. doi:10.1017/S0007114515001063

Sauvageot, N., Leite, S., Alkerwi, A. a., Sisanni, L., Zannad, F., Saverio, S., . . . Guillaume, M. (2016) 'Association of Empirically Derived Dietary Patterns with Cardiovascular Risk Factors: A Comparison of PCA and RRR Methods', *PLOS ONE,* 11(8), pp. e0161298. doi:10.1371/journal.pone.0161298

Savage, J. S., Fisher, J. O. and Birch, L. L. (2007) 'Parental influence on eating behavior: conception to adolescence', *The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics,* 35(1), pp. 22-34. doi:10.1111/j.1748-720X.2007.00111.x

Scaglioni, S., De Cosmi, V., Ciappolino, V., Parazzini, F., Brambilla, P. and Agostoni, C. (2018) 'Factors Influencing Children's Eating Behaviours', *Nutrients,* 10(6), pp. 706. doi:10.3390/nu10060706

Scarborough, P. and Allender, S. (2008) 'The North-South gap in overweight and obesity in England', *Br J Nutr,* 100(3), pp. 677-84. doi:10.1017/S0007114508911582

Scarborough, P., Bhatnagar, P., Wickramasinghe, K. K., Allender, S., Foster, C. and Rayner, M. (2011) 'The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs', *J Public Health (Oxf),* 33(4), pp. 527-35. doi:10.1093/pubmed/fdr033

Scarborough, P., Morgan, R. D., Webster, P. and Rayner, M. (2011) 'Differences in coronary heart disease, stroke and cancer mortality rates between England, Wales, Scotland and Northern Ireland: the role of diet and nutrition', *BMJ Open,* 1(1), pp. e000263. doi:10.1136/bmjopen-2011-000263

Shelton, N. J. (2009) 'Regional risk factors for health inequalities in Scotland and England and the "Scottish effect"', *Soc Sci Med,* 69(5), pp. 761-7. doi:10.1016/j.socscimed.2009.06.044

Shepherd, R. (1999) 'Social determinants of food choice', *Proceedings of the Nutrition Society,* 58(04), pp. 807-812. doi:10.1017/s0029665199001093

Shetty, P. (2002) 'Keynote Paper: Measures of nutritional status from anthropometric survey data'. Available at: https://www.fao.org/3/y4249e/y4249e0b.htm (Accessed: October 15, 2021).

Shetty, P. (2013) 'Nutrition transition and its health outcomes', *Indian J Pediatr,* 80 Suppl 1, pp. S21-7. doi:10.1007/s12098-013-0971-5

Shim, J. S., Oh, K. and Kim, H. C. (2014) 'Dietary assessment methods in epidemiologic studies', *Epidemiol Health,* 36, pp. e2014009. doi:10.4178/epih/e2014009

Slimani, N., Fahey, M., Welch, A. A., Wirfalt, E., Stripp, C., Bergstrom, E., . . . Riboli, E. (2002) 'Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project', *Public Health Nutr,* 5(6B), pp. 1311-28. doi:10.1079/PHN2002407

Smith, A. D., Emmett, P. M., Newby, P. K. and Northstone, K. (2013) 'Dietary patterns obtained through principal components analysis: the effect of input variable quantification', *Br J Nutr,* 109(10), pp. 1881-91. doi:10.1017/s0007114512003868

Smith, D., Clarke, G. and Harland, K. (2009) 'Improving the Synthetic Data Generation Process in Spatial Microsimulation Models', *Environment and Planning A: Economy and Space,* 41(5), pp. 1251-1268. doi:10.1068/a4147

Smith, D. M., Heppenstall, A. and Campbell, M. (2021) 'Estimating Health over Space and Time: A Review of Spatial Microsimulation Applied to Public Health', *J,* 4(2), pp. 182-192. doi: https://doi.org/10.3390/j4020015

Smith, D. M., Pearce, J. R. and Harland, K. (2011) 'Can a deterministic spatial microsimulation model provide reliable small-area estimates of health behaviours? An example of smoking prevalence in New Zealand', *Health Place,* 17(2), pp. 618-24. doi:10.1016/j.healthplace.2011.01.001

Smith, D. M., Vogel, C., Campbell, M., Alwan, N. and Moon, G. (2021) 'Adult diet in England: Where is more support needed to achieve dietary recommendations?', *PLOS ONE,* 16(6), pp. e0252877. doi:10.1371/journal.pone.0252877

Sperrin, M., Marshall, A. D., Higgins, V., Renehan, A. G. and Buchan, I. E. (2016) 'Body mass index relates weight to height differently in women and older adults: serial cross-sectional surveys in England (1992–2011)', *Journal of Public Health,* 38(3), pp. 607-613. doi:10.1093/pubmed/fdv067

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002) 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 64(4), pp. 583-639. doi:https://doi.org/10.1111/1467-9868.00353

Spoor, E. (2021) 'The real impact of removing the Universal Credit uplift'. Available at: https://www.trusselltrust.org/2021/02/08/the-real-impact-of-removing-the-universal-credit-uplift/ (Accessed: October 12, 2021).

Sproesser, G., Ruby, M.B., Arbit, N., Akotia, C.S., Alvarenga, M.d.S., Bhangaokar, R., . . . Renner, B. (2019) 'Understanding traditional and modern eating: the TEP10 framework', *BMC Public Health*, 19(1), p. 1606. doi: 10.1186/s12889-019-7844-4.

Stafford, M., Cummins, S., Ellaway, A., Sacker, A., Wiggins, R.D. and Macintyre, S. (2007) 'Pathways to obesity: identifying local, modifiable determinants of physical activity and diet', *Soc Sci Med*, 65(9), pp. 1882-97. doi: 10.1016/j.socscimed.2007.05.042.

Stephenson, B.J.K., Herring, A.H. and Olshan, A. (2020) 'Robust Clustering with Subpopulation-specific Deviations', *J Am Stat Assoc*, 115(530), pp. 521-537. doi: 10.1080/01621459.2019.1611583.

Stourm, V., Neslin, S. A., Bradlow, E. T., Breugelmans, E., Chun, S. Y., Gardete, P., . . . Venkatesan, R. (2020) 'Refocusing loyalty programs in the era of big data: a societal lens paradigm', *Marketing Letters,* 31(4), pp. 405-418. doi:10.1007/s11002-020-09523-x

Sun, G.W., Shook, T.L. and Kay, G.L. (1996) 'Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis', *J Clin Epidemiol*, 49(8), pp. 907-16. doi: 10.1016/0895-4356(96)00025-x.

Sun, Y., Hu, X., Huang, Y. and On Chan, T. (2020) 'Spatial Patterns of Childhood Obesity Prevalence in Relation to Socioeconomic Factors across England', *ISPRS International Journal of Geo-Information,* 9(10), pp. 599. doi:10.3390/ijgi9100599

Sung, H., Park, J.M., Oh, S.U., Ha, K. and Joung, H. (2021) 'Consumption of Ultra-Processed Foods Increases the Likelihood of Having Obesity in Korean Women', *Nutrients*, 13(2), p. 698. doi: 10.3390/nu13020698

Swinburn, B., Egger, G. and Raza, F. (1999) 'Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity', *Prev Med,* 29(6 Pt 1), pp. 563-70. doi:10.1006/pmed.1999.0585

Taylor, J., Moon, G. and Twigg, L. (2016) 'Using geocoded survey data to improve the accuracy of multilevel small area synthetic estimates', *Social Science Research,* 56, pp. 108-116. doi:https://doi.org/10.1016/j.ssresearch.2015.12.006

The Food Foundation (2017) 'The UK's sugar levy', *International Learning Series* 2. Available at: https://foodfoundation.org.uk/wp-content/uploads/2017/07/2-Briefing-Sugar-Levy_v3.pdf (Accessed August 29, 2018).

Theis, D. and White, M. (2021) 'Is Obesity Policy in England Fit for Purpose? Analysis of Government Strategies and Policies, 1992–2020', *The Milbank Quarterly*, 99(1), pp. 126-170. doi: https://doi.org/10.1111/1468-0009.12498.

Tiffin, R. and Salois, M. (2012) 'Inequalities in diet and nutrition', *Proc Nutr Soc,* 71(1), pp. 105-11. doi:10.1017/S0029665111003284

Tinkler, J. and Ritchie, J. (2016) 'Barriers to Healthy Food', (522). Available at: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0522 (Accessed June 6, 2019).

Tomintz, M., Kosar, B. and Clarke, G. (2016) 'smokeSALUD: exploring the effect of demographic change on the smoking prevalence at municipality level in Austria', *International Journal of Health Geographics,* 15(1), pp. 36. doi:10.1186/s12942-016-0066-4

Tong, T. Y. N., Wareham, N. J., Khaw, K.-T., Imamura, F. and Forouhi, N. G. (2016) 'Prospective association of the Mediterranean diet with cardiovascular disease incidence and mortality and its population impact in a non-Mediterranean population: the EPIC-Norfolk study', *BMC Medicine,* 14(1), pp. 135. doi:10.1186/s12916-016-0677-4

Trichopoulou, A., Naska, A. and Costacou, T. (2002) 'Disparities in food habits across Europe', *Proc Nutr Soc,* 61(4), pp. 553-8. doi:10.1079/pns2002188

Trichopoulos, D. and Lagiou, P. (2004) 'Mediterranean diet and overall mortality differences in the European Union', *Public Health Nutrition*, 7(7), pp. 949-951. doi: 10.1079/PHN2004559.

Trichopoulou, A., Soukara, S. and Vasilopoulou, E. (2007) 'Traditional foods: a science and society perspective', *Trends in Food Science & Technology*, 18(8), pp. 420-427. doi: https://doi.org/10.1016/j.tifs.2007.03.007.

Turrell, G., Hewitt, B., Patterson, C. and Oldenburg, B. (2003) 'Measuring socio-economic position in dietary research: is choice of socio-economic indicator important?', *Public Health Nutr,* 6(2), pp. 191-200. doi:10.1079/PHN2002416

Tuson, M., Yap, M., Kok, M.R., Murray, K., Turlach, B. and Whyatt, D. (2019) 'Incorporating geography into a new generalized theoretical and statistical framework addressing the modifiable areal unit problem', *International Journal of Health Geographics*, 18(1), p. 6. doi: 10.1186/s12942-019-0170-3.

Tweney, E. M., Emmett, P. M., Golding, J., Goodfellow, S. and Taylor, C. M. (2017) 'Comparison of Dietary Intakes of 7-Year-Old Children Enrolled in Observational Birth Cohort Studies on the Isle of Man and in South-west England', *Nutrients,* 9(7), pp. 724. doi:10.3390/nu9070724

Twigg, L. (2014) 'Geographies of space, place, and population health '. Available at: https://pure.port.ac.uk/ws/portalfiles/portal/1824070/GEOGRAPHIES_OF_SPACE_text_revised_jan_2013.pdf (Accessed: October 4, 2021).

Twigg, L., Moon, G. and Jones, K. (2000) 'Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators', *Social science & medicine (1982),* 50(7-8), pp. 1109-1120. doi:10.1016/s0277-9536(99)00359-7

Tyrrell, R.L., Greenhalgh, F., Hodgson, S., Wills, W.J., Mathers, J.C., Adamson, A.J. and Lake, A.A. (2017) 'Food environments of young people: linking individual behaviour to environmental context', *J Public Health (Oxf)*, 39(1), pp. 95-104. doi: 10.1093/pubmed/fdw019.

Ulijaszek, S. J. and McLennan, A. K. (2016) 'Framing obesity in UK policy from the Blair years, 1997-2015: the persistence of individualistic approaches despite overwhelming evidence of societal and economic factors, and the need for collective responsibility', *Obes Rev,* 17(5), pp. 397-411. doi:10.1111/obr.12386

van Buuren, S. (2018) *Flexible Imputation of Missing Data.* 2nd edn. New York Chapman and Hall/CRC

Vatcheva, K. P., Lee, M., McCormick, J. B. and Rahbar, M. H. (2016) 'Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies', *Epidemiology (Sunnyvale, Calif.),* 6(2), pp. 227. doi:10.4172/2161-1165.1000227

Vogler, P. (2020) *Scoff: A History of Food and Class in Britain.* London: Atlantic Books Ltd.

Walls, H. L., Johnston, D., Mazalale, J. and Chirwa, E. W. (2018) 'Why we are still failing to measure the nutrition transition', *BMJ Glob Health,* 3(1), pp. e000657. doi:10.1136/bmjgh-2017-000657

Walsh, D., McCartney, G, Collins, C, Taulbut, M, Batty, G.D. (2016) 'History, politics and vulnerability: explaining excess mortality in Scotland and Glasgow'. Available at: http://www.gcph.co.uk/assets/0000/5988/Excess_mortality_final_report_with_appendices.pdf (Accessed March 12, 2019).

Wark, P. A., Hardie, L. J., Frost, G. S., Alwan, N. A., Carter, M., Elliott, P., . . . Cade, J. E. (2018) 'Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews', *BMC Med,* 16(1), pp. 136. doi:10.1186/s12916-018-1113-8

Weigand, M., Wurm, M., Dech, S. and Taubenböck, H. (2019) 'Remote Sensing in Environmental Justice Research—A Review', *ISPRS International Journal of Geo-Information*, 8(1), p. 20. doi: https://doi.org/10.3390/ijgi8010020

Whichelow, M. J., Erzinclioglu, S. W. and Cox, B. D. (1991) 'Some regional variations in dietary patterns in a random sample of British adults', *Eur J Clin Nutr,* 45(5), pp. 253-62.

Whitworth, A. (2013) 'Evaluations and improvements in small area estimation methodologies', *National Centre for Research Methods Methodological Review paper*. Available at: http://eprints.ncrm.ac.uk/3210/1/sme_whitworth.pdf (Accessed April 10, 2021).

Whitworth, A., Carter, E., Ballas, D. and Moon, G. (2017) 'Estimating uncertainty in spatial microsimulation approaches to small area estimation: A new approach to solving an old problem', *Computers, Environment and Urban Systems,* 63, pp. 50-57. doi:https://doi.org/10.1016/j.compenvurbsys.2016.06.004

Widener, M. J. (2018) 'Spatial access to food: Retiring the food desert metaphor', *Physiol Behav,* 193(Pt B), pp. 257-260. doi:10.1016/j.physbeh.2018.02.032

Williamson, S., McGregor-Shenton, M., Brumble, B., Wright, B. and Pettinger, C. (2017) 'Deprivation and healthy food access, cost and availability: a cross-sectional study', *J Hum Nutr Diet,* 30(6), pp. 791-799. doi:10.1111/jhn.12489

Winchester, N. (2021) 'Universal credit: an end to the uplift'. Available at: https://lordslibrary.parliament.uk/universal-credit-an-end-to-the-uplift/ (Accessed: October 12, 2021).

Wong, D.W. (2009) 'Modifiable Areal Unit Problem', in Kitchin, R. and Thrift, N. (eds.) *International Encyclopedia of Human Geography*. Oxford: Elsevier, pp. 169-174.

Woolf, V. (1927) *To the lighthouse.* Florida: Harcourt.

World Health Organization (WHO) (2015a) 'Healthy diet: Key facts'. Available at: http://www.who.int/news-room/fact-sheets/detail/healthy-diet (Accessed July 11, 2018).

World Health Organization (2015b) *European Food and Nutrition Action Plan2015–2020*. Available at: http://www.euro.who.int/__data/assets/pdf_file/0003/294474/European-Food-Nutrition-Action-Plan-20152020-en.pdf?ua=1 (Accessed June 8, 2020).

World Health Organization (WHO) (2021) 'Overweight and Obesity'. Available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (Accessed: October 15, 2021).

Wrieden, W., Chambers, S., Barton, K., Albani, V. and Anderson, A. (2015) 'Diet and excess mortality in Glasgow and Scotland: Exploring differences in diet and nutrition'. Available at: http://www.healthscotland.com/uploads/documents/27303-Diet%20and%20excess%20mortality%20in%20Glasgow%20and%20Scotland.pdf (Accessed April 10, 2018).

Wrieden, W. L., Armstrong, J., Sherriff, A., Anderson, A. S. and Barton, K. L. (2013) 'Slow pace of dietary change in Scotland: 2001-9', *Br J Nutr,* 109(10), pp. 1892-902. doi:10.1017/S0007114512003789

Wrigley, N. (2002) ''Food Deserts' in British Cities: Policy Context and Research Priorities', *Urban Studies,* 39(11), pp. 2029–2040. doi:10.1080/0042098022000011344

Wrigley, N., Warm, D. and Margetts, B. (2003) 'Deprivation, Diet, and Food-Retail Access: Findings from the Leeds 'Food Deserts' Study', *Environment and Planning A,* 35(1), pp. 151-188. doi:10.1068/a35150

Yau, A., Adams, J. and Monsivais, P. (2019) 'Time trends in adherence to UK dietary recommendations and associated sociodemographic inequalities, 1986-2012: a repeated cross-sectional analysis', *Eur J Clin Nutr,* 73(7), pp. 997-1005. doi:10.1038/s41430-018-0347-z

Young, R. and Johnson, D. (2012) 'To Weight or Not to Weight, That is the Question: Survey weights and multivariate analysis'. Available at: http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2012/03_-Young-Johnson_A2_Weighting-paper_aapor-2012-ry.pdf (Accessed March 3, 2020).

Yuan, C. and Yang, H. (2019) 'Research on K-Value Selection Method of K-Means Clustering Algorithm', *J,* 2(2), pp. 226-235. doi:10.3390/j2020016