

## Article

# Dietary Patterns Derived from UK Supermarket Transaction Data with Nutrient and Socioeconomic Profiles

Stephen D. Clark <sup>1</sup>, Becky Shute <sup>2</sup>, Victoria Jenneson <sup>1</sup>, Tim Rains <sup>2</sup>, Mark Birkin <sup>1</sup> and Michelle A. Morris <sup>3,\*</sup>

<sup>1</sup> Leeds Institute for Data Analytics and School of Geography, University of Leeds, Leeds LS2 9JT, UK; S.D.Clark@leeds.ac.uk (S.D.C.); fs10vl@leeds.ac.uk (V.J.); m.h.birkin@leeds.ac.uk (M.B.)

<sup>2</sup> Sainsbury's Supermarkets Ltd., Holborn, London EC1N 2HT, UK; Becky.Shute@sainsburys.co.uk (B.S.); Tim.Rains@sainsburys.co.uk (T.R.)

<sup>3</sup> Leeds Institute for Data Analytics and School of Medicine, University of Leeds, Leeds LS2 9JT, UK

\* Correspondence: m.morris@leeds.ac.uk

**Abstract:** Poor diet is a leading cause of death in the United Kingdom (UK) and around the world. Methods to collect quality dietary information at scale for population research are time consuming, expensive and biased. Novel data sources offer potential to overcome these challenges and better understand population dietary patterns. In this research we will use 12 months of supermarket sales transaction data, from 2016, for primary shoppers residing in the Yorkshire and Humber region of the UK ( $n = 299,260$ ), to identify dietary patterns and profile these according to their nutrient composition and the sociodemographic characteristics of the consumer purchasing with these patterns. Results identified seven dietary purchase patterns that we named: Fruity; Meat alternatives; Carnivores; Hydrators; Afternoon tea; Beer and wine lovers; and Sweet tooth. On average the daily energy intake of loyalty card holders -who may buy as an individual or for a household- is less than the adult reference intake, but this varies according to dietary purchase pattern. In general loyalty card holders meet the recommended salt intake, do not purchase enough carbohydrates, and purchase too much fat and protein, but not enough fibre. The dietary purchase pattern containing the highest amount of fibre (as an indicator of healthiness) is bought by the least deprived customers and the pattern with lowest fibre by the most deprived. In conclusion, supermarket sales data offer significant potential for understanding population dietary patterns.

**Keywords:** dietary patterns; transaction data; nutrients; socioeconomic; big data; nutrition analytics; dietary assessment



**Citation:** Clark, S.D.; Shute, B.; Jenneson, V.; Rains, T.; Birkin, M.; Morris, M.A. Dietary Patterns Derived from UK Supermarket Transaction Data with Nutrient and Socioeconomic Profiles. *Nutrients* **2021**, *13*, 1481. <https://doi.org/10.3390/nu13051481>

Academic Editor: Marika Massaro

Received: 23 March 2021

Accepted: 20 April 2021

Published: 27 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Poor diet is responsible for 90,000 deaths per year in the United Kingdom (UK) alone, equating to 1 in 7 of all deaths [1]. It is a major modifiable risk factor for a range of non-communicable disease [2] and poor diet can additionally leave people more susceptible to infectious diseases [3]. There are numerous drivers of diet related behaviours including biological, economic, environmental and societal [4]. These factors can contribute to inequity with respect to access to a nutritious balanced diet. Access to food and drinks will systematically vary by country due to availability of products within local food systems, and to a different extent vary locally within countries [2,5].

Diet is measured based on consumption of food and drinks, typically self-reported through a food diary or food frequency questionnaire which capture typical diet behaviours. However, these methods present limitations around error and bias in reporting [6]. When time and money permit, these dietary assessments may be administered by a trained nutritionist or dietitian, in the form of an interview. Traditionally such food records are collected in paper form and then foods and drinks are coded by a researcher in specialist dietary assessment software. Increasingly, online tools are available for individuals to record their food and drink consumption, with the benefit of features such as typical

pack sizes and portion size images, ‘commonly consumed together’ prompts and bar code scanning of products [7,8]. Rich databases lie behind these tools containing nutrient composition data and automating the coding of the diaries. Dietary surveys in the UK, such as the National Diet and Nutrition Survey (NDNS) [9], or cohort studies [10] designed to investigate diet–disease relationships are often used to inform policy decisions. These are excellent resources, even though they are likely to contain biases such as reporting bias and selection bias, and relatively modest sample sizes.

An alternative to recording consumption is to capture food and drink purchases. Here there is an assumption that food and drinks purchased are later consumed. Purchase records or sales data could provide a better reflection of household dietary patterns; however, they are unlikely to be sensitive enough to reflect individual behaviour, except where people live alone. The Family Food Unit of the Living Costs and Food Survey (LCFS) captures household expenditure from purchase records, so the principle of using purchase rather than consumption records is established [11]. Additionally, market research companies such as Kantar and Nielsen use panel surveys where participants scan their foods and drinks purchased, again offering a novel method for assessing diet using purchase records, but at a significant cost to the research team [12]. These methods reduce mis-reporting caused by forgetting what has been consumed, or inaccurate recollection of portion size; however, they do not remove social desirability bias, as participants can “choose” not to scan certain items, for example snacks and soft drinks that may be bought away from home [13]. Research that utilised purchase records from collection of till receipts to estimate nutrient composition showed promise, although did not reduce the data collection burden [14].

There is increasing appetite around the world to use sales data direct from supermarket loyalty card schemes or transaction records [15]. These data have been used in the evaluation of randomized control trials, with a modest number of participants and with their consent [16]. There are examples of sales data at scale from loyalty card transactions in location planning research, rather than in a nutritional context, using data to predict demand in tourist destinations [17]. We have additionally seen transaction data in sustainability research investigating purchase patterns of organic and free range foods [18]. More recently research is emerging using transaction data to better understand nutrition and related health outcomes [19,20]. Research from Finland has investigated the representativeness of data from supermarket loyalty card schemes, showing that women in their forties are overrepresented and that food expenditure recorded on the cards was equivalent to two thirds of the nationally estimated averages [21]. However, they report that these data present great potential to collect data dynamically and at scale in order to enable timely societal dietary insight to be generated.

Access to supermarket transaction data for academic research is challenging due to the commercial sensitivity of such data. In most cases this means that research is prohibited, but where partnerships and research have emerged, transaction data are aggregated in order to reduce disclosure of information which may provide a competitive advantage to other supermarkets. Alternatively, data are only provided for a particular sample of customers, a specific region or specific age group. In some cases, the retailer chooses to keep their identity secret. In spite of these challenges, the use of sales data in nutrition research offers substantial opportunity to transform our understanding of population dietary behaviour.

The aim of this paper is to use sales data from a loyalty card scheme to identify dietary purchase patterns and their associated nutrient profiles and how these vary according to sociodemographic characteristics. We hypothesise that distinct dietary purchasing patterns will be identified from the typical purchasing behaviours of different demographic groups.

## 2. Materials and Methods

### 2.1. Subjects

Participants in this research were 311,972 unique Nectar loyalty card holders that reside in the Yorkshire and Humber region of the United Kingdom. All transactions for the whole of 2016 were included in this research. Participants were identified as “primary” shoppers, where we have estimated that they complete the majority of their shop with Sainsbury’s supermarket. Two conditions were set to identify these shoppers. Firstly, they needed to have shopped with the retailer on a least 10 occasions during 2016. This rule removes those loyalty card holders who just made occasional purchases, e.g., when on vacation or at Christmas. The second condition was that they must purchase from a variety of food groups. Each item purchased is categorised into one of 15 groups. Eleven were derived from the Living Cost and Food Survey (LCFS) categories: Carbohydrate Products; Meat and Fish; Dairy; Fats; Fruit; Salad/Vegetables; Potatoes; Sweets; Other; Non-alcoholic drinks; Alcoholic drinks and a further four generated by the research team: Ready Meals; Baby Food; Cakes and Biscuits; Crisps and Nuts. Only loyalty card holders that purchased from at least seven of these 15 categories, or purchased ready meals and from three other categories, are included in the cohort. This excludes more casual shoppers, e.g., those who habitually only purchase a meal per deal (typically composed of a sandwich, a drink and a snack) or top-up a main shop made elsewhere with occasional purchases of bread, milk or snacks from the retailer. The sample of loyalty card customers is therefore designed to include only those who regularly do their main shop with the retailer.

In addition to transaction records, we received information on the gender of the loyalty card holder and their age band. We also received a neighbourhood geographic identifier, an output area (typically composed of 125 households), to facilitate linkage to area characteristics including 2015 Index of Multiple Deprivation (IMD) ranked deciles [22] and the Output Area Classification (OAC) geodemographic classification [23]. This OAC classification uses 2011 Census data from a range of domains to group similar types of characteristic within an area.

### 2.2. Transaction Records

The transaction data provided included records that linked sales to the loyalty card via a hashed ID to maintain anonymity of customers. Information on product sales provided include: the stock keeping unit (SKU) code that identifies the product; the number of these items that were purchased; the price paid for the items; and, where the item was sold by weight, its weight in g.

During our sampling time frame from the 1 January to 31 December 2016 the retailer carried just over 25,000 products, as identified by their SKU, and in total they sold over 334 million items to our sample of loyalty card holders. Following initial inspection, we identified unusual purchasing patterns at the extremes of loyalty card spending, therefore we removed these outliers, based on information from the 2016 edition of the Family Food Survey (FFS), for the Yorkshire and the Humber region [22] in order to estimate a reasonable upper bound on what a family may spend on non-alcoholic food items at a supermarket through the year. This upper bound is defined as  $1.5 \times$  the inter-quartile range beyond the upper quartile (a common criteria to identify large outliers in box plots). Thus, any loyalty card holder that spent more than this amount on non-alcoholic food items with the retailer was excluded from our sample. This removes 2.04% of loyalty card holders, leaving  $n = 305,616$ . For consistency we additionally removed customers spending in the bottom 2.04% of spends leaving a final sample for analysis of 299,260 loyalty card holders.

### 2.3. Nutritional Information

Nutrient composition data were linked to our sales data at product level. These were “back of packet” values per 100 g or per 100 mL and included: Energy (kcal); Carbohydrates; Protein; Fats; Saturated fats; Salt; Sugar and Fibre. These nutrient composition data were obtained from the supermarket own brand databases and from the commercial brandbank

database for branded products. Where products still had missing nutritional information, we sourced these from the McCance and Widdowson food tables [24], or by searching the internet, recognising that nutrient information may not always match temporally. It was not possible to link nutrient composition information for 2.1% of the 2016 sales, which were associated with the least popular products.

Nutrient composition for the sales were calculated using values per 100 g (or mL) and product weight. Product weight data was sourced from the retailer or mined from the product description in our transaction files. We used the nutrient information for foods as sold, rather than as eaten, as we could not estimate cooking methods or waste. We did not account also for edible portions of food. It was not possible to generate weight data for 0.05% of total items sold in 2016.

Each product was assigned one of the 82 detailed categories that include the LCFS sub-categories, in addition to researcher generated categories: Savoury snacks; Ready Meals; Baby Food; Meat Alternatives; Dairy Alternatives. We were unable to categorise 0.1% of items sold.

## 2.4. Analysis

### 2.4.1. Identifying Dietary Patterns

k-means clustering was used to identify dietary patterns in the sales data. Loyalty card holders are represented by the distribution of their purchase volume (weight in grams) in the 82 categories that include food, non-alcoholic and alcoholic beverages. Provided with a number of classes to find, k-means attempts to group together loyalty card holders with similar patterns of purchasing. For k groups, the method begins with k starting points defined by a random composition amongst the 82 categories. Each loyalty card holder is then allocated to the starting point that they are most similar to, and the mean composition across the 82 categories in each of the k groups is calculated. These calculated means then replace the initial k starting points, and each loyalty card holder is re-allocated to which of these updated k points they are most similar to. This iterative process of re-allocation and calculation is repeated until no loyalty card holder changes their group membership between iterations, or a maximum number of iterations are reached. When this process has finished the k-means solution provides the typical composition for each of the k groups and the group that each loyalty card holder belongs to.

Whilst this process is largely automatic, an important requirement is that the number of groups, k, is known. Typically, this is not the case and some judgement is required to decide on a suitable value for k. The “quality” of a solution given a value of k can be accessed via a within group sum of squares (WSS). This quantity is a measure of how similar loyalty card holders within a group are, a lower value signifying that the loyalty card holders are more similar to each other. By design, as the number of groups increases this value will not decrease, meaning that, as k increases, the quality of the solution will not deteriorate, so that looking for a measure of maximum quality (i.e., a minimum WSS) will not be possible. Instead, what is adopted in practice is to identify when this improvement in quality (reduction in WSS) as k increases becomes negligible or constant. This is best done using a scree plot, with the value of k along the x axis and the value of WSS for k on the y axis. The value of k where this plot develops an “elbow” indicates negligible improvement and is chosen as the value for k.

Another consideration with k-means is that the method works best when the ranges of the values within the categories (e.g., the weights of different foods) are similar (to prevent unequal weight been given to some categories relative to others) and not skewed (k-means looks to form “circular” groupings rather than elongated ones). This is achieved here by range standardising each category, so that the values lie within a range from 0.0 to 1.0, and applying the inverse arc-sin square-root transformation to reduce skewness.

Once a value of k is determined then each loyalty card holder is assigned to a group and, to help identify the nature of the group, the (untransformed) mean volume purchased in each category for members of that grouping can be calculated. For example, in one

group the average purchase of meat amongst its members may be much lower than the overall sample, or in another group purchase of ready meals may be higher.

#### 2.4.2. Profiling Dietary Patterns

The dietary patterns identified by the k-means will be further described, according to their nutrient composition, by the demographic characteristics of the loyalty card holders associated with each pattern and by the typical area characteristics of the loyalty card holders. The nutrient composition will provide an estimated daily consumption of the eight back-of-pack nutrients, calculated by dividing the average total nutrient for each dietary pattern by 366 (as 2016 was a leap year). Adult reference intake will be provided in the results for comparison [25]. Tests for statistical significance in difference were not carried out. Due to the large sample size, all differences are likely to be highly statistically significant. We interpret difference in terms of a clinically meaningful difference.

#### 2.5. Data Sharing

Due to the commercial nature of the data used in this research, it is not possible for data to be published alongside the manuscript.

#### 2.6. Ethical Approval

This research was reviewed by the University of Leeds ethics committee reference: AREA 18-050.

### 3. Results

#### 3.1. The Sample

Our sample of 299,260 loyalty card holders includes large numbers in each demographic group. However, women (Figure 1A), those aged 45–64 (Figure 1B) and those living in areas within the six least deprived deciles of deprivation (Figure 1C) are over-represented. Customers typically living in areas characterized by the 2011 OAC classification as Rural residents, Urbanites and Suburbanites are also over-represented (Figure 1D). The median frequency of shopping occasions for our classification sample was 53 with interquartile range 33 to 82.

#### 3.2. Dietary Patterns

Using the scree plot (Figure 2A) and change in WSS (Figure 2B) we identified seven distinct dietary patterns. The scree plot looks inconclusive, but for the first differences, the reduction in WSS when moving from 6 to 7 groups is  $-1543$ , but thereafter there are much smaller step change reductions of around  $-1000$ . This provides support for having seven groupings within the data.

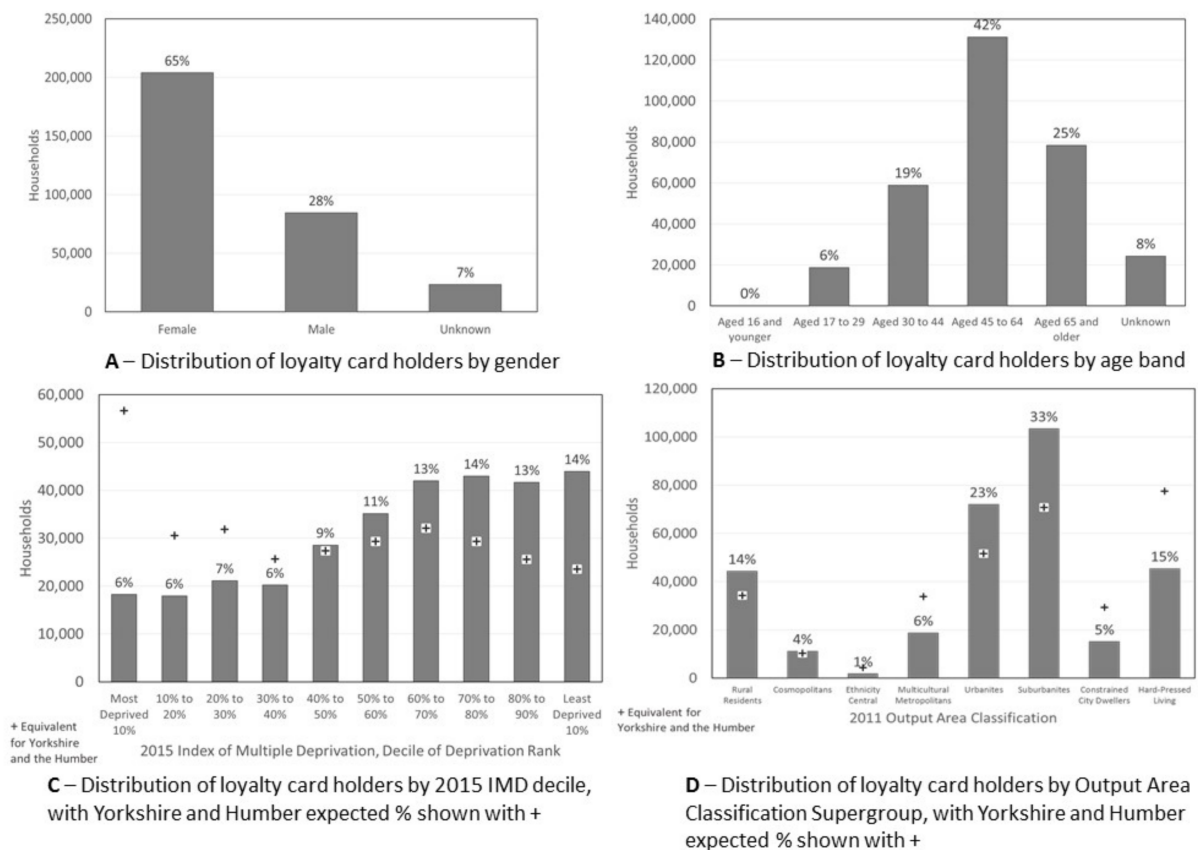


Figure 1. Characteristics of our loyalty card holders.

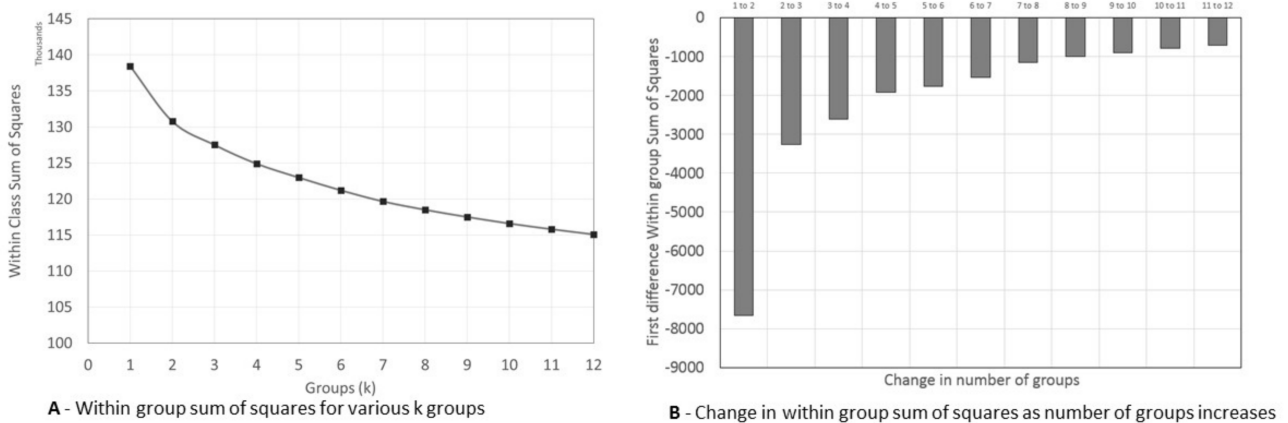


Figure 2. (A) Scree plot and (B) change in within group sum of square for identifying number of dietary patterns.

Features of the seven dietary patterns are presented in Figure 3. Further information and radar plots for each pattern are provided in Appendix A.

### 3.3. Demographic Characteristics Associated with Each Dietary Pattern

Dietary patterns are ordered by their approximate healthfulness, using amount of fibre as a proxy for healthfulness. Figure 4 shows how each dietary pattern is purchased within deciles of deprivation, with variation across deciles of deprivation. We see that the “Sweet tooth”, “Afternoon tea” and “Carnivores” patterns are over-represented in the most deprived decile and the “Fruity” cluster is over-represented in the 3 least deprived deciles. Additionally, the bottom row of Figure 4 provides the ratio of the percentage of loyalty card

holders in the top (least deprived) two deciles relative to the bottom (most deprived) two, clearly highlighting where patterns are more likely to be purchased by the most deprived (lowest number) or least deprived (highest number). The “Meat alternatives” pattern is most likely to be purchased by customers living in the least deprived areas and the “Sweet tooth” in the most deprived.

**Fruity** ( $n = 57,232$ ): Fresh fruit is purchased in a greater proportion by these loyalty card holders relative to the average, with seven of the top ten purchases being fruit. The remaining three in the top ten are: fresh vegetables, dairy alternatives and sugar and sugar products. However, sugary drinks, both soft and alcoholic, are least popular with purchases below average.

**Meat alternatives** ( $n = 61,065$ ): These loyalty card holders are more likely to purchase pasta and meat alternative products than the average and they purchase large amounts of baby food. Least popular purchases are unsurprisingly meat products but also alcoholic drinks such as wines, spirits and liqueurs.

**Carnivores** ( $n = 49,923$ ): A comprehensive variety of different meat products dominate purchases by these loyalty card holders. Other than meats, there is some purchasing of staple vegetables such as potatoes and cabbage. Alcohol, meat and dairy alternatives and baby food are not commonly purchased by these customers.

**Hydrators** ( $n = 22,804$ ): The dominant purchase by these customers is mineral and spring waters, which they purchase nearly seven times more frequently than other customers. Other popular purchases include: dairy alternatives, preserved and whole milks, meat alternatives, olive oil, seafood and baby food. Buns, cakes, biscuits, pastry and sausages are the least commonly purchased in this group.

**Afternoon tea** ( $n = 41,309$ ): The pattern of purchasing for these loyalty card holders resembles the type of food stuffs found in a tea room, with high purchases of: milk, cakes and puddings, margarine, preserves (jams) and pastries. Fresh vegetables are least popular in this group.

**Beer and wine lovers** ( $n = 35,486$ ): These customers are characterised by the purchase of alcoholic drinks, with eight of the ten most purchased items being alcoholic. Fresh fruit is not often purchased by these customers.

**Sweet tooth** ( $n = 31,441$ ): These loyalty card holders purchase a variety of products that are sweet in nature. This includes drinks (alcoholic and non-alcoholic) and chocolates/confectionery. These customers also make regular purchases of ready meals. Ingredients for the preparation of traditional meals, vegetables and some meats are purchased less than average.

**Figure 3.** Summary of the dietary pattern contents. Please see Appendix A for more details.

The same type of social patterning is observed in Figure 5 where the dietary patterns are profiled against the OAC classification. The “Sweet tooth” customers are under-represented in Rural residents neighbourhoods, but over-represented in those that are categorised as Hard-pressed living. The “Fruity” customers are over-represented in Hard-pressed neighbourhoods, which is coupled with over representations among the Rural residents and Suburbanites.

IMD decile	Whole sample	Dietary pattern							
	All	Fruity	Meat alternatives	Carnivores	Hydrators	Afternoon tea	Beer and wine lovers	Sweet tooth	
Lowest 10% (most deprived)	6%	4%	4%	7%	6%	8%	4%	10%	
20%	6%	4%	4%	7%	5%	8%	5%	8%	
30%	7%	5%	6%	8%	7%	8%	6%	9%	
40%	6%	5%	6%	7%	6%	7%	7%	8%	
50%	9%	8%	9%	10%	9%	9%	10%	10%	
60%	11%	11%	11%	11%	12%	11%	12%	11%	
70%	13%	14%	14%	13%	14%	13%	14%	12%	
80%	14%	15%	15%	13%	14%	12%	14%	12%	
90%	13%	16%	15%	12%	13%	12%	14%	11%	
Highest 10% (least deprived)	14%	18%	16%	12%	14%	12%	14%	9%	
ratio top 20%/bottom 20%	2.38	4.6	3.61	1.79	2.41	1.47	2.89	1.08	

Figure 4. Distribution of groups of loyalty card holders amongst Index of Multiple Deprivation (IMD) deciles.

OAC Super Group	Whole sample	Dietary pattern							
	All	Fruity	Meat alternatives	Carnivores	Hydrators	Afternoon tea	Beer and wine lovers	Sweet tooth	
Rural Residents	14%	18%	14%	14%	15%	12%	15%	9%	
Cosmopolitans	4%	4%	5%	1%	4%	2%	3%	5%	
Ethnicity Central	1%	0%	0%	0%	1%	1%	0%	1%	
Multicultural Metropolitan	6%	5%	5%	6%	7%	8%	5%	8%	
Urbanites	23%	24%	25%	20%	24%	21%	24%	22%	
Suburbanites	33%	37%	34%	35%	32%	31%	34%	26%	
Constrained City Dwellers	5%	3%	4%	6%	5%	7%	4%	7%	
Hard-Pressed Living	15%	9%	12%	18%	13%	18%	14%	22%	

Figure 5. Distribution of groups of loyalty card holders amongst Output Area Classification (OAC) Super Groups.

Table 1 provides a breakdown of the daily nutrient purchasing of each dietary pattern. Those within the “Carnivores” pattern purchase the highest daily amount of calories, just over the UK adult reference intake. By contrast, “Sweet tooth” customers gain the lowest amount of their daily nutritional requirement from this supermarket. None of the patterns contain the recommended daily fibre intake, but the “Fruity” pattern contains the most at 21 g/day, just over two thirds of the daily recommended amount. In contrast, the “Sweet tooth” pattern only contains 10 g/day fibre, one third of the daily recommended amount. None of the patterns contain enough carbohydrate, with the highest carbohydrate purchases in the “Meat alternatives” group, buying 80% of the recommended amount. All of the patterns, except “Sweet tooth” contain enough protein. “Afternoon tea”, “Sweet tooth” and “Beer and wine lovers” meet the target of <70 g/day fat; however, all patterns exceed the daily saturated fat recommendations, with the “Carnivores” pattern most exceeding both fat and saturated fat. While the “Sweet tooth”, pattern is the least healthy, with the lowest fibre content, and typified by sugary food and drinks, this cluster purchase on average, the least amount of calories, which appears to be the reason they do not exceed the daily fat target and have lower sugar purchases than other patterns. Conversely, “Fruity”—the healthiest pattern—contains the highest amount of sugar, likely explained by natural sugars from fruit. That said, despite being dominated by fruit purchases, it also contains above average purchases of sugar.



**Table 1.** Daily nutritional values for loyalty card holder by group. \* Fibre not included in adult reference intake values, but UK recommended intake included below.

Nutrient	Adult Reference Intake	All	Fruity	Meat Alternatives	Carnivores	Hydrators	Afternoon Tea	Beer and Wine Lovers	Sweet Tooth
Energy (kcal)	2000	1757	1818	1869	2026	1729	1632	1585	1384
Fat (g)	<70	72	75	76	89	72	66	61	56
Saturates (g)	<20	28	29	29	35	27	27	23	22
Carbohydrates (g)	At least 260	186	197	207	203	183	189	141	155
Sugars (g)	90	87	97	93	90	86	91	61	72
Salt (g)	<6	5	6	6	7	5	5	5	4
Protein (g)	50	66	69	70	83	65	58	57	47
Fibre (g)	30 *	16	21	18	18	16	14	12	10

## 4. Discussion

We present a novel analysis of large and complex supermarket transaction data that identifies dietary patterns purchased in the UK. We believe this to be the first work of its kind using such detailed food transactions with associated nutritional information. We were able to generate data driven dietary patterns for ~300,000 loyalty card holders across one region of the UK and profile these by average daily nutrient intake. While the sample was biased towards certain demographic characteristics, data were available for large numbers of loyalty card holders in each sub-group.

### 4.1. Dietary Pattern Insight

Our sample only included shoppers that we believe do the majority of their food shopping at Sainsbury's supermarket. We were interested in capturing an indicator of habitual intake, akin to a diet diary or food frequency questionnaire, rather than looking at occasional shoppers, or those who only buy in limited food groups. Our method for identifying these customers used both a frequency metric and an indicator of variety, based on national survey categories. Whilst we set a threshold of at least 10 shops with the retailer, in reality our cohort of loyalty carders shopped more frequently, with a median of 53 shopping occasions throughout the year. Subgroup analysis (not reported here) showed that loyalty card holders shopping at least biweekly had slightly higher daily energy purchases, at 1887 kcal/day compared to the whole cohort at 1757 kcal/day.

Our data include food and beverage purchases, including alcohol. Two of the dietary patterns: "Beer and wine lovers" and "Sweet tooth" were driven by alcohol sales, of different types. These two patterns were associated with the two lowest daily calorie intakes, suggesting perhaps that while these shoppers met the criteria for purchasing regularly from a range of different food groups, their main priority during their shop with this retailer was for alcohol purchases. Characteristics of the loyalty card holders purchasing each of these patterns were quite different, with the "Sweet tooth" pattern more commonly containing customers living in a neighbourhood characterised by the OAC geodemographic classification as Hard-pressed living. The "Beer and wine lovers" dietary pattern is distributed across OAC groups in a similar way to the total cohort of loyalty card holders, with no noticeable over or under-representation.

All but the "Carnivores" pattern contain less than adult reference intake for energy (2000 kcal), but all except the "Sweet tooth" and "Beer and wine lovers" contain more calories than the mean adult woman's intake in 2016 reported by the NDNS of 1632 kcal [9]. It is known that the NDNS under-reports dietary intake, and results here are broadly supportive of that, with our cohort average intake of 1762 kcal/day.

The "Fruity" pattern is interesting, with seven of the top ten purchased items being types of fruit. This pattern is most popular among Rural residents and in Suburbanite neighbourhoods, which are typically quite affluent. It contains the highest daily fibre intake, but this is still way below the daily recommended intake. Fibre does not form part of the adult reference intake guidance, nor is it mandated to appear on the back-of-pack of food products [26]. In 2015 the UK Scientific Advisory Committee on Nutrition recommended

an increase in daily fibre intake to 30 g/day in the UK following a comprehensive review of the literature [27]. These sales data are from 2016, the year following the report. In the case of fibre, it would be interesting to follow up with more recent data to see how consumers have responded to these recommendations.

Conversely, in the case of salt, where recommendations to reduce salt intake have been around for much longer, only the “Carnivores” pattern exceeds the daily recommended salt intake, and this is by 1 g at an average of 7 g/day. This is suggestive that the highly publicised guidance and product reformulations—with voluntary targets introduced in 2005 and the most recent legislation brought in in 2015—are largely successful at helping consumers achieve targets [28]. Considering that loyalty card holders may be buying for more than one person, this implies that salt targets are likely to be met by all, and that fibre targets are even further from being achieved.

#### 4.2. Study Strengths and Limitations

The most striking strengths of this research are the large sample size and the objectively measured sales records for dietary intake. The work presented here highlights the potential scope of sales data for identifying dietary purchase patterns. The transaction data are comprehensively matched to detailed nutrient composition data at the product level, providing a powerful dataset to investigate nutritional patterns and trends. These types of data would enable comprehensive modelling of proposed nutritional policies, such as the planned legislation expected to be announced in summer 2021 to restrict price and location promotions of high fat, sugar and salt products [29]. It has been recognised that there are limitations in the data that retailers hold which would enable them to successfully implement these new rules across their whole product portfolio [30], but this new dataset, which combines sales and nutrient composition data, would make that easier and would also enable, for the first time, quantification of the impact of legislation in objective sales weighted data, not subject to the limitations of self-reporting, and therefore unlike survey and panel data.

These linked sales and nutrient data, with demographic characteristics, additionally make a bold step towards digitising the food system. While they may only represent part of the consumer facing retail environment, this could be the starting point for combination with other sources of retail data from the food supply chain [5].

However, this is matched by some notable limitations. These data are only from one supermarket, which as the results demonstrate is used by people residing in areas across all deciles of deprivation, but over-represented in the least deprived areas. The loyalty card holders are predominantly women, which is the same observation found by Nevalainen, Erkkola [21] in Finland, suggesting that women still take on the main responsibility of providing food for their family, or that women register for cards that men also use, and they are middle aged, again consistent with the Finnish study.

In our loyalty card data, we do not know whether the loyalty card holder is shopping for themselves, or for a family. This significantly limits the conclusions that can be drawn from our results and the reason why we only present descriptive statistics. Average daily nutrient intakes were calculated as if the loyalty card holder were living alone, but we can see from the contents of the patterns that some customers are likely to be buying for a family. With this in mind, the nutrient profiles presented in Table 1 are likely to under-represent the daily intake of these consumers. Compared with the NDNS, the average daily energy intake is still >100 kcal greater per day. That said, it is widely recognised that the NDNS under-reports diet in the UK. Supermarket sales data do not capture out of home food purchases that dietary surveys have the potential to capture—yet still our study demonstrates a higher daily calorie intake than NDNS. LCFS purchase data from 2016/17 suggests that 88% of purchases, by volume (g or mL) are for home consumption, compared to 12% eating out [31].

Another limitation is that we do not know whether the loyalty card holder does all of their shopping at Sainsbury's, nor whether they use their loyalty card each time. They may forget their card, or do "top-up" shops from another retailer.

Data driven dietary patterns are beneficial as they identify patterns relative to the data subjects' behaviour, rather than trying to score them against a pre-existing dietary pattern. Being data driven, using detailed food categories, they are more difficult to compare to patterns observed in other studies, generated from subjects with different consumption or purchasing behaviours. That said, there are common themes in our patterns seen in many other studies, for example: presence of a carnivore pattern, a sweet tooth/snacking pattern and a meat alternative, or vegetarian pattern [32,33]. Our patterns differ from many previous studies, due to the inclusion of alcohol in the cluster generating process, which we believe to be a strength as alcohol contributes to energy intake. The naming of dietary patterns can be subjective and as such we have included detailed information on the content of the patterns in Appendix A for transparency.

#### 4.3. Future Work

We hope that this study is the first of many that make use of supermarket sales data to generate insight into food purchasing behaviours around the world. In order to translate food purchase behaviours to dietary consumption, evaluation of transaction records as a source of dietary assessment is required [34].

It is important to better understand the bias in these data, with respect to demographic characteristics of the shoppers, household composition and how much shopping is carried out at this retailer using a loyalty card. Incorporating an indication of out of home purchases to sales data would facilitate understanding of total dietary intake.

Comparison with established dietary patterns or frameworks, such as the Eatwell Guide [35], nutrient density measure [36] or healthy eating index [37], would facilitate comparison to patterns generated in other research, for example cohort studies.

Linkage of dietary purchase patterns to health records, with explicit consent, to better understand diet disease relationships could be paradigm shift in prevention and management of non-communicable disease.

## 5. Conclusions

Sales data recorded by supermarket loyalty cards, linked with nutrient composition records and demographic characteristics, offer potential to understand food purchases at a scale not seen before, which unlocks the potential for more robust analysis to inform diet related policy. Further work is required to better understand the biases within these data and to extend data providers to include representation from other retailers and the out of home sector.

**Author Contributions:** Conceptualization, M.A.M. and M.B.; methodology, M.A.M., S.D.C. and V.J.; formal analysis, M.A.M., S.D.C. and V.J.; resources, B.S. and T.R.; data curation, T.R., B.S., S.D.C., V.J. and M.A.M.; writing—original draft preparation, M.A.M. and S.D.C.; writing—review and editing, M.A.M., S.D.C., V.J., M.B., B.S. and T.R.; funding acquisition, M.B. and M.A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The researcher time was funded by the Economic and Social Research Council funded Consumer Data Research Centre REF: ES/L011891/1.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the University of Leeds Ethics committee reference: AREA 18-050.

**Informed Consent Statement:** Informed consent was not required for this secondary data analysis, and not possible to obtain as all data were anonymized.

**Data Availability Statement:** Due to the commercial nature of the data used in this research, it is not possible for data to be published alongside the manuscript.

**Acknowledgments:** Thank you to Sainsbury's for providing data in kind for this research; Nilani Sritharan and Bethan Tempest for nutrition expertise; and to Adam Keeley in the Data Analytics team at the Leeds Institute for Data Analytics for support with data import to the secure environment. Also thank you to the reviewers who provided timely and supportive comments on an earlier draft of this article.

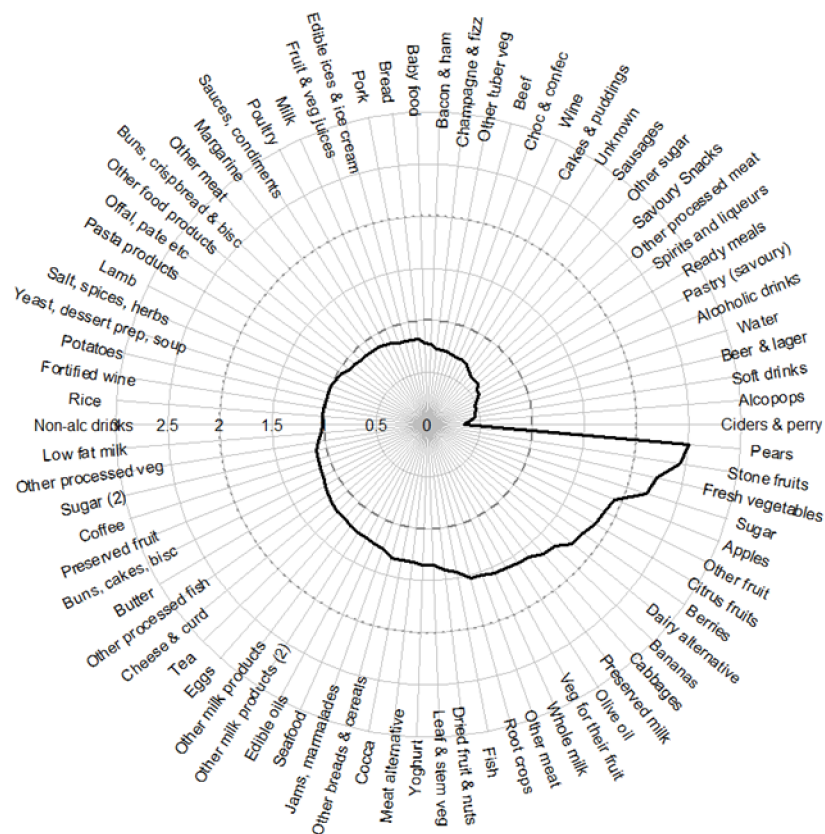
**Conflicts of Interest:** M.A.M. is an inventor and shareholder at Dietary Assessment Ltd., but receives no financial income, and is not involved in the running of the company. B.S. and T.R. are employees at Sainsbury's.

## **Appendix A Detailed Information on the Dietary Patterns**

### *Interpretation of Groupings*

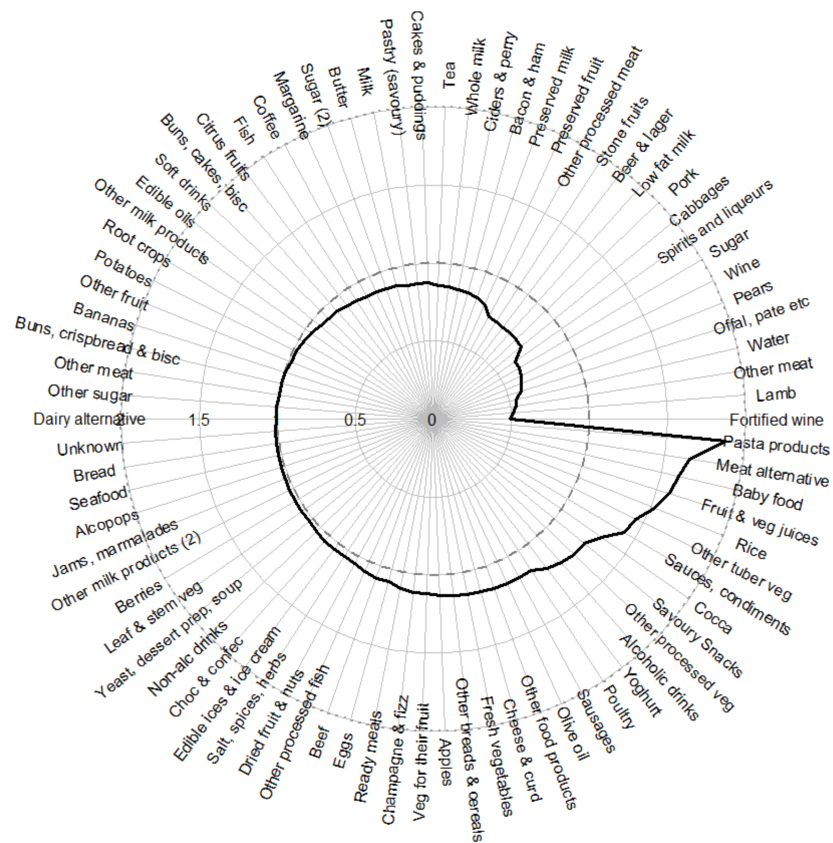
Each subsequent page contains the information necessary to profile each loyalty card holder grouping. A loyalty card holder may be purchasing for just themselves, or for a household which may be more than one person. The top of the page contains an ordered radial plot showing how many times MORE or LESS than a typical (average) loyalty card holder in this grouping is purchased in each category of food or drink. For example, a value of 2.0 indicates that they purchase twice as much of this category as a typical customer. A value of 1.0 indicates that they spend exactly the same as a typical household, and a value of 0.5 indicates that they spend half as much. The noteworthy categories are shown around the 3 o'clock position, either categories that are purchased more or items that are purchased less than a typical household.

Below the radial plot are the top and bottom ten categories within this group (i.e., categories in the region of 3 o'clock in the radial plot).



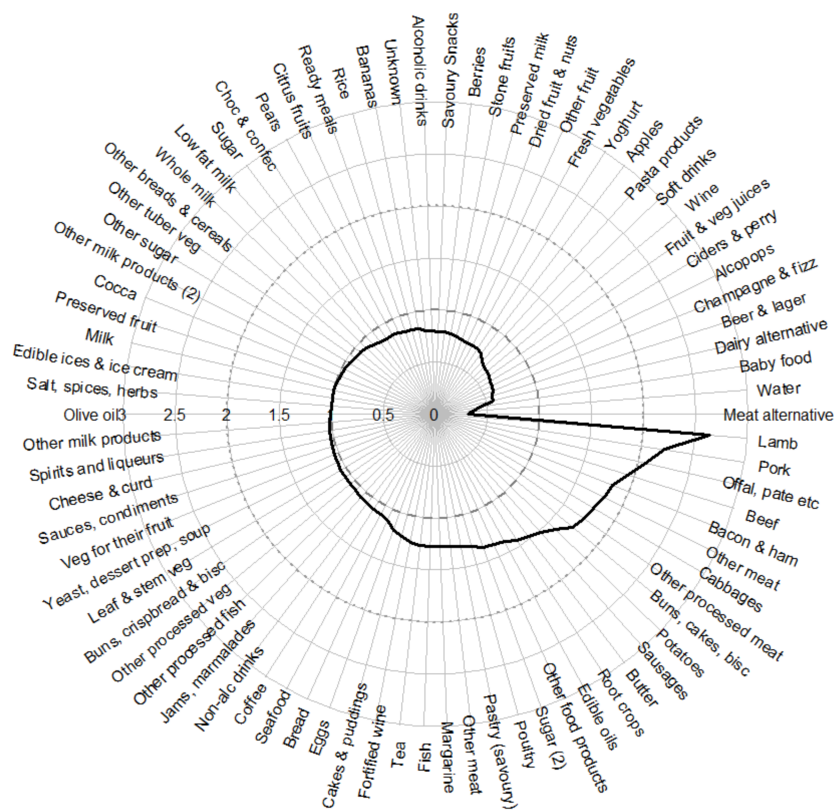
1	<b>Pears (fresh)</b>
2	<b>Stone fruits (fresh)</b>
3	<b>Fresh vegetables</b>
4	<b>Sugar and sugar products</b>
5	<b>Apples (fresh)</b>
6	<b>Other fresh, chilled or frozen fruits</b>
7	<b>Citrus fruits (fresh)</b>
8	<b>Berries (fresh)</b>
9	<b>Dairy alternative</b>
10	<b>Bananas (fresh)</b>
<hr/>	
73	<i>Other preserved or processed meat and meat preparations</i>
74	<i>Spirits and liqueurs (brought home)</i>
75	<i>Ready meals</i>
76	<i>Pastry (savoury)</i>
77	<i>Alcoholic drinks</i>
78	<i>Mineral or spring waters</i>
79	<i>Beer and lager (brought home)</i>
80	<i>Soft drinks (inc. fizzy and ready to drink fruit drinks)</i>
81	<i>Alcopops (brought home)</i>
82	<i>Ciders and perry (brought home)</i>

Figure A1. Fruity (n = 57,232).



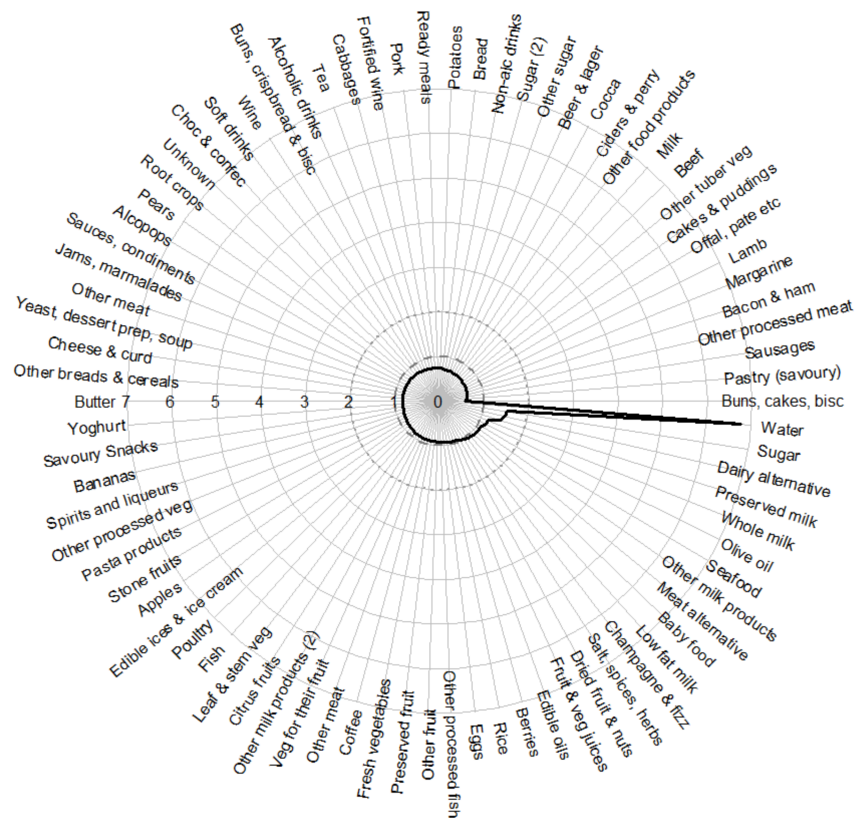
1	<b>Pasta products</b>
2	<b>Meat alternatives</b>
3	<b>Baby food</b>
4	<b>Fruit and vegetable juices</b>
5	<b>Rice</b>
6	<b>Other tubers and products of tuber vegetables</b>
7	<b>Sauces, condiments</b>
8	<b>Cocoa and powdered chocolate</b>
9	<b>Savoury snacks</b>
10	<b>Other dried, preserved or processed vegetables</b>
<hr/>	
73	<i>Cabbages (fresh or chilled)</i>
74	<i>Spirits and liqueurs (brought home)</i>
75	<i>Sugar and sugar products</i>
76	<i>Wine from grape or other fruit (brought home)</i>
77	<i>Pears (fresh)</i>
78	<i>Offal, pate etc</i>
79	<i>Mineral or spring waters</i>
80	<i>Other meats and meat preparations</i>
81	<i>Lamb (fresh, chilled or frozen)</i>
82	<i>Fortified wine (brought home)</i>

Figure A2. Meat Alternatives (n = 61,065).



1	<b>Lamb (fresh, chilled or frozen)</b>
2	<b>Pork (fresh, chilled or frozen)</b>
3	<b>Offal, pate etc</b>
4	<b>Beef (fresh, chilled or frozen)</b>
5	<b>Bacon and ham</b>
6	<b>Other meats and meat preparations</b>
7	<b>Cabbages (fresh or chilled)</b>
8	<b>Other preserved or processed meat and meat preparations</b>
9	<b>Buns, cakes, biscuits etc</b>
10	<b>Potatoes</b>
73	<i>Wine from grape or other fruit (brought home)</i>
74	<i>Fruit and vegetable juices</i>
75	<i>Ciders and perry (brought home)</i>
76	<i>Alcopops (brought home)</i>
77	<i>Champagne and sparkling wines (brought home)</i>
78	<i>Beer and lager (brought home)</i>
79	<i>Dairy alternative</i>
80	<i>Baby food</i>
81	<i>Mineral or spring waters</i>
82	<i>Meat alternatives</i>

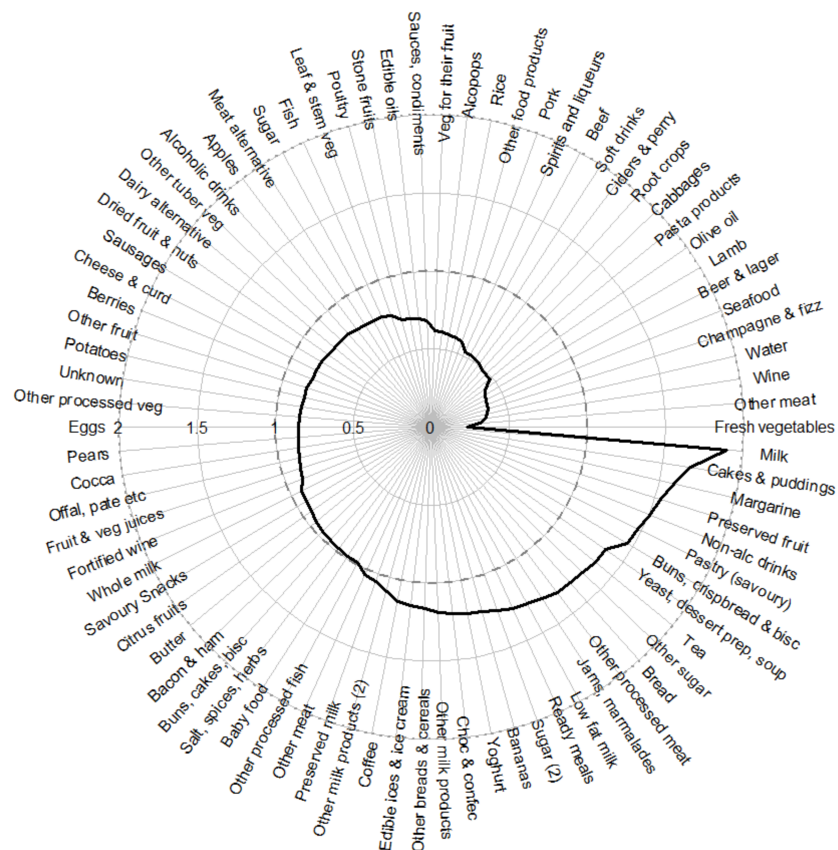
Figure A3. Carnivores (n = 49,923).



1	<b>Mineral or spring waters</b>
2	<b>Sugar and sugar products</b>
3	<b>Dairy alternative</b>
4	<b>Preserved milk</b>
5	<b>Whole milk</b>
6	<b>Olive oil</b>
7	<b>Seafood, dried, smoked or salted fish</b>
8	<b>Other milk products</b>
9	<b>Meat alternatives</b>
10	<b>Baby food</b>
-----	
73	<i>Other tubers and products of tuber vegetables</i>
74	<i>Cakes and puddings</i>
75	<i>Offal, pate etc</i>
76	<i>Lamb (fresh, chilled or frozen)</i>
77	<i>Margarine, other vegetable fats and peanut butter</i>
78	<i>Bacon and ham</i>
79	<i>Other preserved or processed meat and meat preparations</i>
80	<i>Sausages</i>
81	<i>Pastry (savoury)</i>
82	<i>Buns, cakes, biscuits etc</i>

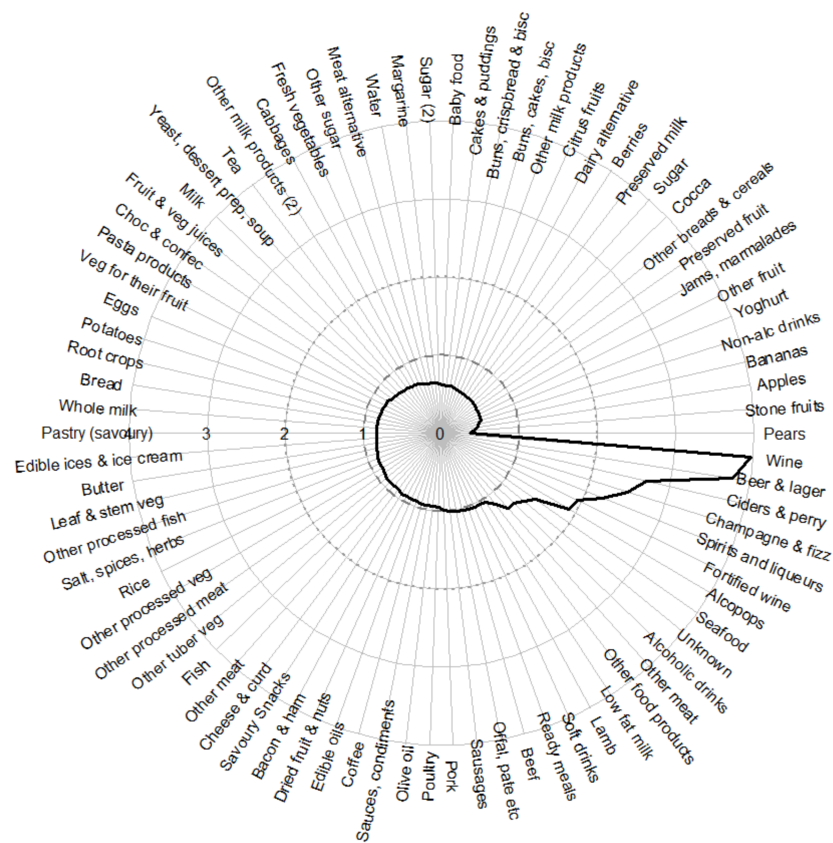
Figure A4. Hydrators ( $n = 22,804$ ).





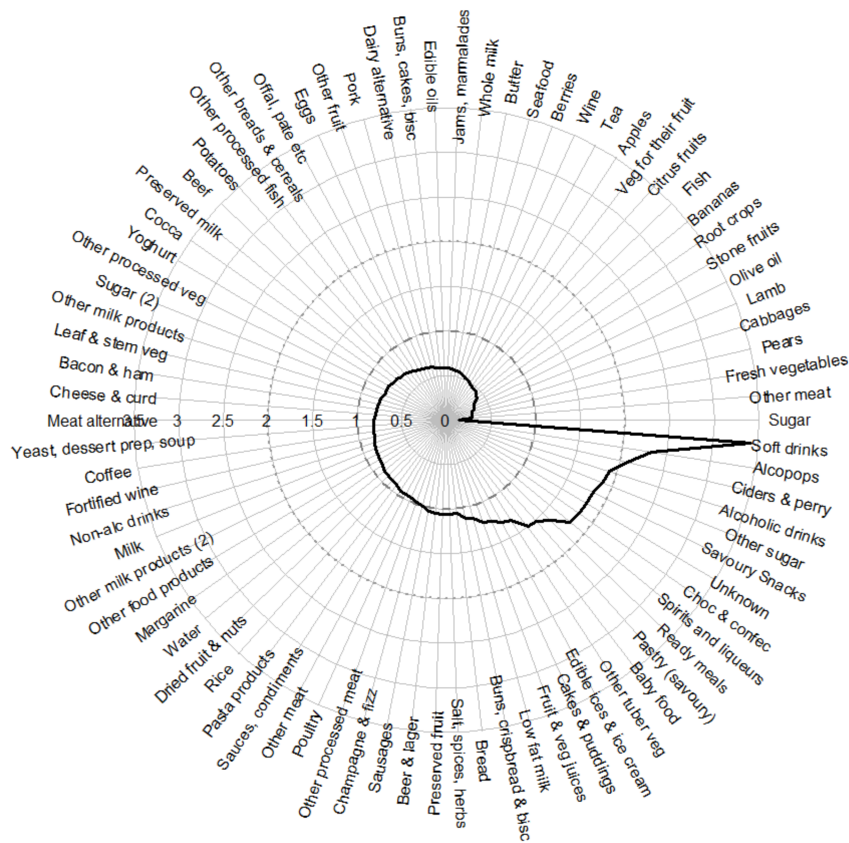
1	<b>Milk</b>
2	<b>Cakes and puddings</b>
3	<b>Margarine, other vegetable fats and peanut butter</b>
4	<b>Preserved fruit and fruit based products</b>
5	<b>Non-alcoholic drinks</b>
6	<b>Pastry (savoury)</b>
7	<b>Buns, crispbread and biscuits</b>
8	<b>Baker's yeast, dessert preparations, soups</b>
9	<b>Tea</b>
10	<b>Other sugar products</b>
73	<i>Pasta products</i>
74	<i>Olive oil</i>
75	<i>Lamb (fresh, chilled or frozen)</i>
76	<i>Beer and lager (brought home)</i>
77	<i>Seafood, dried, smoked or salted fish</i>
78	<i>Champagne and sparkling wines (brought home)</i>
79	<i>Mineral or spring waters</i>
80	<i>Wine from grape or other fruit (brought home)</i>
81	<i>Other meats and meat preparations</i>
82	<i>Fresh vegetables</i>

Figure A5. Afternoon tea (n = 41,309).



1	<b>Wine from grape or other fruit (brought home)</b>
2	<b>Beer and lager (brought home)</b>
3	<b>Ciders and perry (brought home)</b>
4	<b>Champagne and sparkling wines (brought home)</b>
5	<b>Spirits and liqueurs (brought home)</b>
6	<b>Fortified wine (brought home)</b>
7	<b>Alcopops (brought home)</b>
8	<b>Seafood, dried, smoked or salted fish</b>
9	<b>Unknown</b>
10	<b>Alcoholic drinks</b>
-----	
73	<i>Other breads and cereals</i>
74	<i>Preserved fruit and fruit based products</i>
75	<i>Jams, marmalades</i>
76	<i>Other fresh, chilled or frozen fruits</i>
77	<i>Yoghurt</i>
78	<i>Non-alcoholic drinks</i>
79	<i>Bananas (fresh)</i>
80	<i>Apples (fresh)</i>
81	<i>Stone fruits (fresh)</i>
82	<i>Pears (fresh)</i>

Figure A6. Beer and wine lovers ( $n = 35,486$ ).



1	<b>Soft drinks (inc. fizzy and ready to drink fruit drinks)</b>
2	<b>Alcopops (brought home)</b>
3	<b>Ciders and perry (brought home)</b>
4	<b>Alcoholic drinks</b>
5	<b>Other sugar products</b>
6	<b>Savoury snacks</b>
7	<b>Unknown</b>
8	<b>Chocolate &amp; confectionary</b>
9	<b>Spirits and liqueurs (brought home)</b>
10	<b>Ready meals</b>
73	<b>Bananas (fresh)</b>
74	<b>Root crops, non-starchy bulbs and mushrooms (fresh, chilled or frozen)</b>
75	<b>Stone fruits (fresh)</b>
76	<b>Olive oil</b>
77	<b>Lamb (fresh, chilled or frozen)</b>
78	<b>Cabbages (fresh or chilled)</b>
79	<b>Pears (fresh)</b>
80	<b>Fresh vegetables</b>
81	<b>Other meats and meat preparations</b>
82	<b>Sugar and sugar products</b>

Figure A7. Sweet tooth (n = 31,441).

## References

1. National Food Strategy. *National Food Strategy—Part One*; Department for Environment, Food & Rural Affairs: London, UK, 2020.
2. World Health Organisation. Healthy Diet Factsheet. 2020. Available online: <https://www.who.int/news-room/fact-sheets/detail/healthy-diet> (accessed on 1 February 2021).
3. Iddir, M.; Brito, A.; Dinger, G.; Del Campo, S.S.F.; Samouda, H.; La Frano, M.R.; Bohn, T. Strengthening the Immune System and Reducing Inflammation and Oxidative Stress through Diet and Nutrition: Considerations during the COVID-19 Crisis. *Nutrients* **2020**, *12*, 1562. [[CrossRef](#)] [[PubMed](#)]
4. Butland, B.; Jebb, S.; Kopelman, P.; McPherson, K.; Thomas, S.; Mardell, J.; Parry, V. *Tackling Obesities: Future Choices-Project Report*; Department of Innovation, Universities and Skills: London, UK, 2007; Volume 10.
5. Parsons, K.; Hawkes, C.; Wells, R. Brief 2. What is the Food System? In *A Food Policy Perspective, in Rethinking Food Policy: A Fresh Approach to Policy and Practice*; Centre for Food Policy: London, UK, 2019.
6. Schatzkin, A.; Subar, A.F.; Moore, S.; Park, Y.; Potischman, N.; Thompson, F.E.; Leitzmann, M.; Hollenbeck, A.; Morrissey, K.G.; Kipnis, V. Observational Epidemiologic Studies of Nutrition and Cancer: The Next Generation (with Better Observation). *Cancer Epidemiol. Biomark. Prev.* **2009**, *18*, 1026–1032. [[CrossRef](#)] [[PubMed](#)]
7. Carter, M.C.; Albar, S.A.; Morris, M.A.; Mulla, U.Z.; Hancock, N.; Evans, C.E.; Alwan, N.A.; Greenwood, D.C.; Hardie, L.J.; Frost, G.S.; et al. Development of a UK Online 24-h Dietary Assessment Tool: myfood24. *Nutrients* **2015**, *7*, 4016–4032. [[CrossRef](#)] [[PubMed](#)]
8. Subar, A.F.; Kirkpatrick, S.I.; Mittl, B.; Zimmerman, T.P.; Thompson, E.F.; Bingley, C.; Willis, G.; Islam, N.G.; Baranowski, T.; McNutt, S.; et al. The automated self-administered 24-hour dietary recall (ASA24): A resource for researchers, clinicians, and educators from the national cancer institute. *J. Acad. Nutr. Diet.* **2012**, *112*, 1134–1137. [[CrossRef](#)] [[PubMed](#)]
9. National Diet and Nutrition Survey. *NDNS: Results from Years 7 and 8 (Combined)*; Public Health England and the Food Standards Agency: London, UK, 2018.
10. Cade, J.E.; Burley, V.J.; Alwan, N.A.; Hutchinson, J.; Hancock, N.; Morris, M.A.; Threapleton, D.E.; Greenwood, D.C. Cohort Profile: The UK Women’s Cohort Study (UKWCS). *Int. J. Epidemiol.* **2017**, *46*, e11. [[CrossRef](#)] [[PubMed](#)]
11. Office for National Statistics. Living Costs and Food Survey. 2017. Available online: <https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey> (accessed on 1 February 2021).
12. Bandy, L.; Adhikari, V.; Jebb, S.; Rayner, M. The use of commercial food purchase data for public health nutrition research: A systematic review. *PLoS ONE* **2019**, *14*, e0210192. [[CrossRef](#)] [[PubMed](#)]
13. Einav, L.; Leibtag, E.; Nevo, A. *On the Accuracy of Nielsen Homescan Data*; United States Department of Agriculture, Ed.; Economic Research Service: Washington, DC, USA, 2008.
14. Ransley, J.K.; Donnelly, J.K.; Khara, T.N.; Botham, H.; Arnot, H.; Greenwood, D.C.; E Cade, J. The use of supermarket till receipts to determine the fat and energy intake in a UK population. *Public Health Nutr.* **2001**, *4*, 1279–1286. [[CrossRef](#)] [[PubMed](#)]
15. Jenneson, V.; Greenwood, D.C.; Clarke, G.; Morris, M.A. The use of electronic supermarket sales data for population dietary surveillance: A systematic review. **2021**. forthcoming.
16. Piernas, C.; Aveyard, P.; Lee, C.; Tsiountsioura, M.; Noreik, M.; Astbury, N.M.; Oke, J.; Madigan, C.; Jebb, S.A. Evaluation of an intervention to provide brief support and personalized feedback on food shopping to reduce saturated fat intake (PC-SHOP): A randomized controlled trial. *PLoS Med.* **2020**, *17*, e1003385. [[CrossRef](#)] [[PubMed](#)]
17. Newing, A.; Clarke, G.; Clarke, M. Identifying seasonal variations in store-level visitor grocery demand. *Int. J. Retail. Distrib. Manag.* **2013**, *41*, 477–492. [[CrossRef](#)]
18. Chintakayala, P.K.; Young, W.; Barkemeyer, R.; Morris, M.A. Breaking niche sustainable products into the mainstream: Organic milk and free-range eggs. *Bus. Strat. Environ.* **2018**, *27*, 1039–1051. [[CrossRef](#)]
19. Green, M.A.; Watson, A.W.; Brunstrom, J.M.; Corfe, B.M.; Johnstone, A.M.; Williams, E.A.; Stevenson, E. Comparing supermarket loyalty card data with traditional diet survey data for understanding how protein is purchased and consumed in older adults for the UK, 2014–16. *Nutr. J.* **2020**, *19*, 1–10. [[CrossRef](#)] [[PubMed](#)]
20. Aiello, L.M.; Quercia, D.; Schifanella, R.; Del Prete, L. Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London. *Sci. Data* **2020**, *7*, 1–11. [[CrossRef](#)] [[PubMed](#)]
21. Nevalainen, J.; Erkkola, M.; Saarijärvi, H.; Näppilä, T.; Fogelholm, M. Large-scale loyalty card data in health research. *Digit. Health* **2018**, *4*, 2055207618816898. [[CrossRef](#)] [[PubMed](#)]
22. Ministry of Housing Communities & Local Government. *English Indices of Deprivation*; Ministry of Housing, Communities and Local Government: London, UK, 2015. Available online: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015> (accessed on 1 February 2021).
23. Gale, C.G.; Singleton, A.D.; Bates, A.G.; A Longley, P. Creating the 2011 area classification for output areas (2011 OAC). *J. Spat. Inf. Sci.* **2016**, *2016*, 1–27. [[CrossRef](#)]
24. McCance and Widdowson. Composition of Foods Integrated Dataset (CoFID). 2015. Available online: <https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid> (accessed on 1 February 2021).
25. NHS. Reference Intakes Explained. 2020. Available online: <https://www.nhs.uk/live-well/eat-well/what-are-reference-intakes-on-food-labels/> (accessed on 1 February 2021).
26. Department of Health London. *Technical Guidance on Nutrition Labelling*; Department of Health London: London, UK, 2016.

27. The Scientific Advisory Committee on Nutrition. *SACN Carbohydrates and Health Report*; SACN: London, UK, 2015.
28. Food Standard Agency. Salt Reduction Targets. 2018. Available online: <https://www.food.gov.uk/business-guidance/salt#salt-reduction-targets> (accessed on 1 February 2021).
29. Department of Health and Social Care. *Restricting Promotions of Products High in Fat, Sugar and Salt by Location and by Price: Government Response to Public Consultation*; DHSC: London, UK, 2020.
30. Jenneson, V.; Greenwood, D.C.; Clarke, G.P.; Hancock, N.; Cade, J.E.; Morris, M.A. Restricting promotions of ‘less healthy’ foods and beverages by price and location: A big data application of UK Nutrient Profiling Models to a retail product dataset. *Nutr. Bull.* **2020**, *45*, 389–402. [[CrossRef](#)]
31. Department for Environment Food and Rural Affairs. *Family Food Datasets*; DEFRA: London, UK, 2020.
32. Greenwood, D.C.; Cade, J.E.; Draper, A.; Barrett, J.H.; Calvert, C.; Greenhalgh, A. Seven unique food consumption patterns identified among women in the UK Women’s Cohort Study. *Eur. J. Clin. Nutr.* **2000**, *54*, 314–320. [[CrossRef](#)] [[PubMed](#)]
33. Morris, M.A.; Wilkins, E.L.; Galazoula, M.; Clark, S.D.; Birkin, M. Assessing diet in a university student population: A longitudinal food card transaction data approach. *Br. J. Nutr.* **2020**, *123*, 1406–1414. [[CrossRef](#)] [[PubMed](#)]
34. Jenneson, V. *STRIDE (Supermarket Transaction Records in Dietary Evaluation) Study Protocol: Validation of Supermarket Transactions against Self-Reported Dietary Intake*; Open Science Framework: Charlottesville, VA, USA, 2020.
35. Public Health England. *The Eatwell Guide*; PHE: London, UK, 2018.
36. Drewnowski, A.; Fulgoni, V.L. Nutrient density: Principles and evaluation tools. *Am. J. Clin. Nutr.* **2014**, *99* (Suppl. 5), 1223S–1228S. [[CrossRef](#)] [[PubMed](#)]
37. Krebs-Smith, S.M.; Pannucci, T.E.; Subar, A.F.; Kirkpatrick, S.I.; Lerman, J.L.; Toozé, J.A.; Wilson, M.M.; Reedy, J. Update of the Healthy Eating Index: HEI-2015. *J. Acad. Nutr. Diet.* **2018**, *118*, 1591–1602. [[CrossRef](#)] [[PubMed](#)]