

‘Ultra-processed’ foods, household expenditure and trends in obesity.



Image by [Kevin Phillips](#) from [Pixabay](#)

Elsa Ekevall

24 September 2022

Project Overview

Motivation

Poor diet within the UK is a public health issue and the UK is ranked as one of the worst in Europe for levels of obesity, particularly among children. A [study in the journal Public Health Nutrition](#)¹ showed that families buy more 'ultra-processed' food than other European countries; with these products accounting for more than half of all UK family food purchases. 'Ultra-processed' food - food made in a factory with industrial ingredients and additives you wouldn't find in your kitchen - bears little resemblance to the fruit, vegetables, meat or fish used to cook a fresh meal at home. Could these products be driving the rise in obesity?

Objective

This project will investigate if there are any relationships between 'ultra-processed' foods, household expenditure and obesity.

Scope

This project will focus on the UK and look at data from 2015 up to 2020.

Summary Of Data Sets

- Living Costs and Food Survey
- Health Survey for England
- Scottish Health Survey
- National Survey for Wales
- Health Survey Northern Ireland

Living Costs and Food Survey

The [Living Costs and Food Survey \(LCF\)](#)², collects information on spending patterns and the cost of living reflecting household budgets across the UK. The LCF provides an invaluable supply of economic and social data "primarily used to provide information for the Retail Prices Index, National Accounts estimates of household expenditure, analysis of the effect of taxes and benefits, and trends in nutrition".

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

² Anthony Rafferty and Pierre Walthery (uploaded by Rosalyn Southern, 2014). *Introductory Guide to the Living Costs and Food Survey*. UK Data Service, University of Essex and University of Manchester;

Data Sourcing

An external data source provided by the [Office for National Statistics](#) in the UK, “the UK’s largest independent producer of official statistics and its recognised national statistical institute”. This is government data and can be verified as a trustworthy data source.

Data Collection

The data is administrative data collected annually from face-to-face interviews, spread evenly over the year to ensure that seasonal effects are covered, from a sample of around 6,000 households a year within the UK. It is a repeated cross-sectional study that uses multi-stage stratified random sampling with clustering. Households are selected at random from the Royal Mail's Postcode Address File in Great Britain or the Land and Property Services Agency's (LPSA) property database in Northern Ireland. The survey is voluntary and households are not obliged to take part. The data is collected through two questionnaires and diary records of daily expenditure. Adults and children, aged 7 to 15, who also complete their daily expenditure diary receive a monetary gift voucher.

Further information about the LCF, including links to published reports based on the survey, may be found on the ONS [Living Costs and Food Survey](#) webpage.

Data Contents

The household questionnaire collects information at a household-level on a range of subjects including family relationships, ethnicity, employment details, family food, and the ownership of household durables. It also contains expenditure not recorded in the daily diary; such as regular payments and large, infrequently purchased items. The individual income questionnaire collects the key person-level variables and covers income from employment, benefits and assets; forming an overview of the total household income, as well as each individual household member. In addition to the two questionnaires, individuals are asked to keep diary records of daily expenditure for two weeks.

Data Relevance

The LCF data is available annually from 2008 and the five most recent years (2015 to 2020) downloaded from the [UK Data Service](#)³ are used in this analysis. As it was provided by the Office of National Statistics it can be assumed to be the most trustworthy and complete version of the data available. The data sets provide a comprehensive picture of household income and spending, including food purchases, and is critical to addressing the project objective.

Data Sets

- Office for National Statistics, Department for Environment, Food and Rural Affairs. (2022). Living Costs and Food Survey, 2019-2020. [data collection]. 2nd Edition. UK Data Service. SN: 8803, [DOI: 10.5255/UKDA-SN-8803-2](#)

³ Office for National Statistics, Department for Environment, Food and Rural Affairs., Living Costs and Food Survey. UK Data Service. Accessed on 24 September 2022.

- Office for National Statistics, Department for Environment, Food and Rural Affairs. (2021). Living Costs and Food Survey, 2018-2019. [data collection]. 4th Edition. UK Data Service. SN: 8686, [DOI: 10.5255/UKDA-SN-8686-4](https://doi.org/10.5255/UKDA-SN-8686-4)
- Office for National Statistics, Department for Environment, Food and Rural Affairs. (2020). Living Costs and Food Survey, 2017-2018. [data collection]. 3rd Edition. UK Data Service. SN: 8459, [DOI: 10.5255/UKDA-SN-8459-3](https://doi.org/10.5255/UKDA-SN-8459-3)
- Office for National Statistics, Department for Environment, Food and Rural Affairs. (2020). Living Costs and Food Survey, 2016-2017. [data collection]. 3rd Edition. UK Data Service. SN: 8351, [DOI: 10.5255/UKDA-SN-8351-3](https://doi.org/10.5255/UKDA-SN-8351-3)
- Department for Environment, Food and Rural Affairs, Office for National Statistics. (2020). Living Costs and Food Survey, 2015-2016. [data collection]. 3rd Edition. UK Data Service. SN: 8210, [DOI: 10.5255/UKDA-SN-8210-5](https://doi.org/10.5255/UKDA-SN-8210-5)

The LCF Derived Urban and Rural Household Characteristics Data for each of the years above (dvhh_urbanrural_ukanon.dta) were used in this analysis. These files contain information on household composition, household characteristics, aggregate household spending and items that are contained in a household.

Health Survey for England

The [Health Survey for England](#) (HSE) provides information about adults aged 16 and over, and children aged 0 to 15, living in private households in England. “The survey is used to estimate the proportion of people in England who have health conditions, and the prevalence of risk factors and behaviours associated with certain health conditions.”

Data Sourcing

An external data source provided by [NHS Digital](#) on behalf of the [Office for National Statistics](#) in the UK, “the UK’s largest independent producer of official statistics and its recognised national statistical institute”. This is government data and can be verified as a trustworthy data source.

Data Collection

“The survey combines questionnaire-based answers with physical measurements and the analysis of blood samples. Blood pressure, height and weight, smoking, drinking and general health are covered every year.” The data is collected through self-administered questionnaires, clinical measurements (nurse visits the household) and face-to-face interviews with each eligible person living in the private household in England (around 9,600 addresses selected at random). It is an annually repeated cross-sectional study of adults (aged 16 and over) and children (aged 0-15 years) in the general population, which uses multi-stage stratified random sampling. The survey is voluntary and households are not obliged to take part.

Further information about the survey, including links to published reports based on the HSE, may be found on the NHS Digital [Health Survey for England](#) webpage.

Data Contents

The questionnaires cover general health, longstanding illness, alcohol consumption, smoking, socio-economic information and prescribed medications. The nurse visit collects measurements such as height and weight, blood pressure, waist and hip circumference, as well as taking of blood samples and saliva samples.

Data Relevance

The HSE data is available annually from 1993 and the years (2015 to 2019) downloaded from the UK Data Service⁴ are used in this analysis. As it was provided by NHS Digital for the Office of National Statistics it can be assumed to be the most trustworthy and complete version of the data available. The data sets monitor the population's health including BMI/Obesity.

Data Sets

- NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2021). Health Survey for England, 2019. [data collection]. UK Data Service. SN: 8860, [DOI: 10.5255/UKDA-SN-8860-1](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=8860)
- NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2022). Health Survey for England, 2018. [data collection]. 2nd Edition. UK Data Service. SN: 8649, [DOI: 10.5255/UKDA-SN-8649-2](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=8649)
- University College London, Department of Epidemiology and Public Health, National Centre for Social Research (NatCen). (2021). Health Survey for England, 2017. [data collection]. 2nd Edition. UK Data Service. SN: 8488, [DOI:10.5255/UKDA-SN-8488-2](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=8488)
- NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2019). Health Survey for England, 2016. [data collection]. 3rd Edition. UK Data Service. SN: 8334, [DOI: 10.5255/UKDA-SN-8334-3](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=8334)
- NatCen Social Research, University College London, Department of Epidemiology and Public Health. (2019). Health Survey for England, 2015. [data collection]. 2nd Edition. UK Data Service. SN: 8280, [DOI: 10.5255/UKDA-SN-8280-2](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=8280)

The HSE Adult Child BMI/Obesity tables⁵ were used in this analysis.

Scottish Health Survey

The [Scottish Health Survey](#) (SHeS) provides information about adults aged 16 and over, and children aged 0 to 15, living in private households in Scotland. The survey “provides regular information on aspects of the public's health and factors related to health .”

⁴ Office for National Statistics, Department for Environment, Food and Rural Affairs., Living Costs and Food Survey. UK Data Service. Accessed on 24 September 2022.

⁵ Downloaded on 28 September.

Data Sourcing

An external data source commissioned by the [Scottish Government Health Directorates](#), the series provides regular information on aspects of the public's health and factors related to health. Conducted with the [Office for National Statistics](#) in the UK, “the UK’s largest independent producer of official statistics and its recognised national statistical institute”. This is government data and can be verified as a trustworthy data source.

Data Collection

“Each survey in the series includes a set of core questions and measurements (height and weight and, if applicable, blood pressure, waist circumference, urine and saliva samples), plus modules of questions on specific health conditions that vary from year to year.”

The data is collected through face-to-face interviews with people living in private households in Scotland (around 6,000 addresses selected at random), and where eligible biological measurements and samples are taken by a trained interviewer. It is an annually repeated cross-sectional study of adults (aged 16 and over) and children (aged 0-15 years) in the general population, which uses multi-stage stratified random sampling. The survey is voluntary and households are not obliged to take part.

Further information about the survey, including links to published reports based on the SHeS, may be found on the Scottish Government [Scottish Health Survey](#) webpage.

Data Contents

“Topics covered included household composition, demographics (including ethnicity, religion, educational background and economic activity), general health including caring, mental health and wellbeing, cardiovascular disease, respiratory disease and asthma, physical activity, sedentary activity, barriers and motivations to exercise, eating habits, fruit and veg consumption, smoking and drinking, dental health, social capital.”

Data Relevance

The SHeS data is available annually from 1995 and the years (2015 to 2019) downloaded from the [UK Data Service](#) are used in this analysis. As it was provided by Scottish Government with the Office of National Statistics it can be assumed to be the most trustworthy and complete version of the data available. The data sets monitor the population’s health including BMI/Obesity.

Data Sets

- ScotCen Social Research. (2021). Scottish Health Survey, 2019. [data collection]. UK Data Service. SN: 8737, [DOI: 10.5255/UKDA-SN-8737-1](#)
- ScotCen Social Research. (2021). Scottish Health Survey, 2018. [data collection]. 4th Edition. UK Data Service. SN: 8569, [DOI: 10.5255/UKDA-SN-8569-4](#)
- ScotCen Social Research. (2021). Scottish Health Survey, 2017. [data collection]. UK Data Service. SN: 8398, [DOI: 10.5255/UKDA-SN-8398-1](#)
- ScotCen Social Research. (2017). Scottish Health Survey, 2016. [data collection]. UK Data Service. SN: 8290, [DOI: 10.5255/UKDA-SN-8290-1](#)

- ScotCen Social Research. (2016). Scottish Health Survey, 2015. [data collection]. UK Data Service. SN: 8100, [DOI: 10.5255/UKDA-SN-8100-1](https://doi.org/10.5255/UKDA-SN-8100-1)

The SHes Adult Child Obesity/BMI data accessed online⁶ through the UK Data Service using Nesstar were used in this analysis.

National Survey for Wales

The [National Survey for Wales](#) provides information about adults aged 16 and over living in private households in Wales. The survey provides “representative, reliable and up-to-date information about the people of Wales down to local authority level.”

Data Sourcing

An external data source carried out by Office for National Statistics in the UK, “the UK’s largest independent producer of official statistics and its recognised national statistical institute,” on behalf of the [Welsh Government](#). This is government data and can be verified as a trustworthy data source.

Data Collection

The survey covers a random sample of adults (around 12,000) a year living in private households across Wales. Interviewers visit each address and carry out a 45-minute face-to-face interview with one randomly selected adult in the household, covering a range of views, behaviours, and characteristics. It is an annually repeated cross-sectional study of adults (aged 16 and over) in the general population, which uses one-stage stratified or systematic random sampling. The survey is voluntary and households are not obliged to take part.

Further information about the survey, including links to published reports based on the survey, may be found on the Welsh Government [National Survey for Wales](#) webpage.

Data Contents

The surveys cover “a range of issues including health (e.g. diet, alcohol consumption, smoking, and physical activity), childcare, and child education, climate change and environmental action, visits to the outdoors, participation in arts events, and sports activities, use of and satisfaction with public services, material deprivation and income, wellbeing and loneliness, use of and attitudes towards the use of the Welsh language, internet access and use and tax devolution.”

Data Relevance

Data is available annually from 2012 and the years (2017 to 2019) downloaded from the [UK Data Service](#) are used in this analysis. As it was provided by the Welsh Government with the Office of National Statistics it can be assumed to be the most trustworthy and complete version of the data available. These data sets monitor the population’s health including BMI/Obesity.

⁶ Accessed on the 28 September 2022 <http://nesstar.ukdataservice.ac.uk/webview/index.jsp?v=2&mode=documentation&submode=abstract&study=http://nesstar.ukdataservice.ac.uk/to80/obj/fStudy/8569&top=yes>

Data Sets

- Welsh Government, Office for National Statistics. (2020). National Survey for Wales, 2018-2019. [data collection]. UK Data Service. SN: 8591, [DOI: 10.5255/UKDA-SN-8591-1](https://doi.org/10.5255/UKDA-SN-8591-1)
- Welsh Government, Office for National Statistics. (2020). National Survey for Wales, 2017-2018. [data collection]. 2nd Edition. UK Data Service. SN: 8390, [DOI: 10.5255/UKDA-SN-8390-2](https://doi.org/10.5255/UKDA-SN-8390-2)
- Office for National Statistics, Welsh Government. (2020). National Survey for Wales, 2016-2017. [data collection]. UK Data Service. SN: 8301, [DOI: 10.5255/UKDA-SN-8301-1](https://doi.org/10.5255/UKDA-SN-8301-1)

The National Survey for Wales Adult Child Obesity/BMI data accessed online⁷ through the UK Data Service using Nesstar were used in this analysis.

Health Survey Northern Ireland

The [Health Survey Northern Ireland](#) provides information about adults aged 16 and over living in private households in Northern Ireland. “The survey covers a range of health topics that are important to the lives of people in Northern Ireland today.”

Data Sourcing

An external data source carried out by the [Central Survey Unit \(CSU\) of the Northern Ireland Statistics and Research Agency \(NISRA\)](#), “the principal source of official statistics and social research on Northern Ireland,” on behalf of the [Department of Health in Northern Ireland](#). This is government data and can be verified as a trustworthy data source.

Data Collection

The survey covers a random sample of adults living in private households across Northern Ireland. Data is collected through face-to-face interviews, self-completed questionnaires, and physical measurements. It is an annually repeated cross-sectional study of adults (aged 16 and over) in the general population, which uses simple random sampling (a systematic random sample of 6,240 addresses selected from the NISRA Address Register). The survey is voluntary and households are not obliged to take part.

Further information about the survey, including links to published reports based on the survey, may be found on the Northern Ireland [Department of Health](#) webpage.

Data Contents

The survey covers a range of topics including basic household information, general health, dental health, specific medical conditions, flu vaccinations, health and lifestyle attitudes, home care and carers, dietary information and food security, breast feeding, smoking and drinking, and religion .

⁷ Accessed on the 28 September 2022 <http://nesstar.ukdataservice.ac.uk/webview/index.jsp?v=2&mode=documentation&submode=abstract&study=http://nesstar.ukdataservice.ac.uk/to80/objectStudy/8390&top=yes>

Data Relevance

The [Health Survey NI Trend Tables](#)⁸ containing BMI/Obesity data covering 2010/11 to 2020/21 from the Health Survey Northern Ireland was downloaded and used in this analysis. As it was provided by the Northern Ireland government it can be assumed to be the most trustworthy and complete version of the data available.

Data Set

- Health Survey NI Trend Tables: httpsto://www.health-ni.gov.uk/sites/default/files/publications/health/hsni-trend-tables-20-21_1.xlsx

Data Limitations And Ethics

Living Costs and Food Survey

Data Limitations

The survey is voluntary and although efforts are made to ensure that the results represent the whole population, it may not accurately represent the population of the UK.

The survey does not provide a balance sheet of income and expenditure. The majority of expenditure information collected relates to the two week period immediately following the interview, whereas income components refer to a much longer period.

Respondents' diaries are checked after a few days to ensure they have understood how to record their daily expenditure and again when they are returned. Prior to publication LCF data are subjected to a rigorous process of quality assurance. Missing data are imputed by a team of editors within ONS.

Data Bias

Exclusion bias: There may be bias if non-respondents to the survey are different to those responding.

Measurement bias: Differences in interviewer techniques and the respondent's interpretation of the question may introduce bias. The interviewers are trained to try to reduce this bias.

Sample bias: The addresses are chosen at random and even with stratification may not provide an accurate representation of the variation between households. Low response rates, with fewer households overall represented, could have an affect on the variety of households and increase the possibility of bias.

⁸ Accessed on 28 September 2022. <httpsto://www.health-ni.gov.uk/publications/tables-health-survey-northern-ireland>

Data Ethics

The data was collected in a transparent manner in accordance with the laws and regulations. It has been anonymised to stop individuals and households being identified. In order to access the data you must agree to comply with the regulations that govern the use of the data.

Timeliness

The data obtained is the most recent complete set with the same methodology. At the time of this analysis the Living Costs and Food Survey, 2020-2021 was not available.

UK BMI Surveys

Data Limitations

The surveys are voluntary and although efforts are made to ensure that the results represent the whole population, it may not accurately represent the population of that part of the UK. There may be differences in the way the measurements to calculate BMI have been taken.

Data Bias

Exclusion bias: There may be bias if non-respondents to the survey are different to those responding.

Measurement bias: Although the people measuring are trained to reduce bias there may be differences in techniques, especially between the different countries within the United Kingdom, who run separate surveys.

Sample bias: The addresses are chosen at random and even with stratification may not provide an accurate representation of the variation between households. Low response rates, with fewer households overall represented, could have an affect on the variety of households and increase the possibility of bias.

Data Ethics

The data was collected in a transparent manner in accordance with the laws and regulations and has been anonymised to stop individuals being identified. In order to access some of the data sets you must agree to comply with the regulations that govern the use of the data.

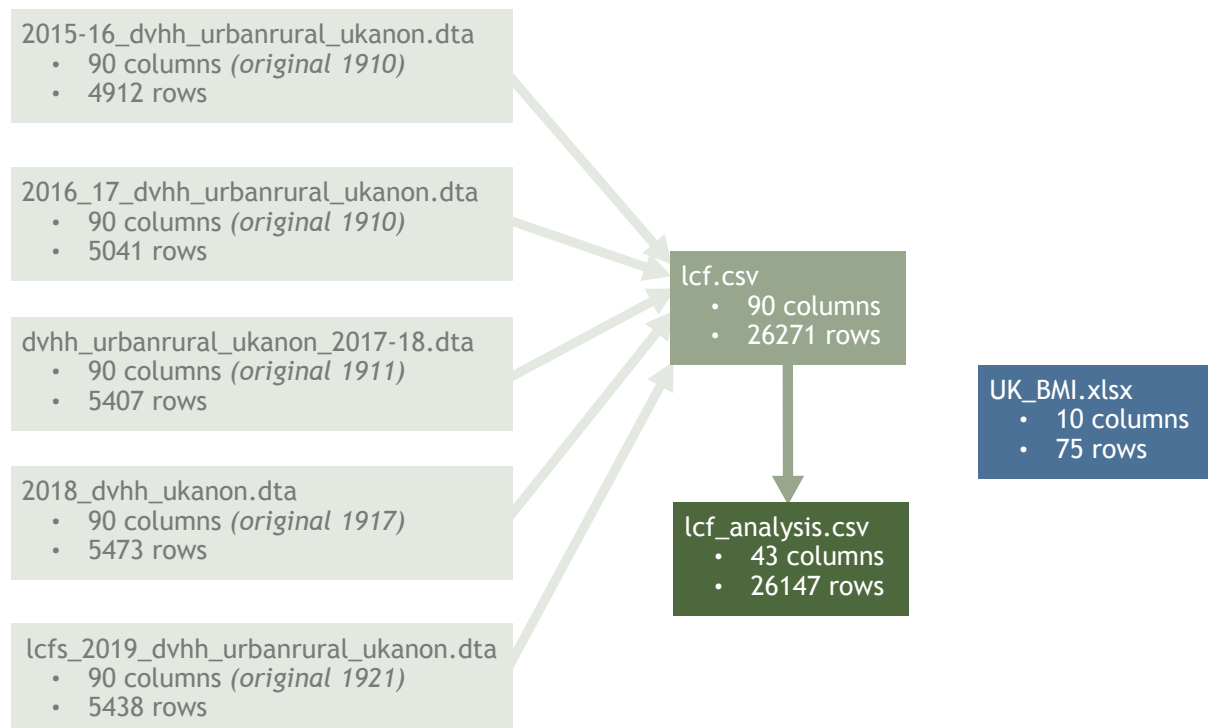
Timeliness

The data obtained are the most recent complete sets with the same methodology. Data from 2020 was not included due to the pandemic impacting the way the data was collected.

Merging The Data Sets

From each of the [Living Costs and Food Survey](#) data sets 90 variables (see [Data Dictionary](#)) were imported out of the original variables and a unique identifier row added. The data sets were then merged into one.

The relevant BMI/obesity data was directly copied from the Health Survey for England, Scottish Health Survey, National Survey for Wales and Health Survey Northern Ireland data sets into an Microsoft Excel file named UK_BMI containing 10 variables (see [Data Dictionary](#)).



Data Profiles

Living Costs and Food Survey Combined Data Sets

The four data set files were imported with only the relevant columns and subsequently joined to create one data set. A unique_id column was added to the data set during the cleaning process to make checking rows and columns easier.

Data Dictionary

None of the columns from the joined data sets contained unstructured data. All income and expenditure is in pounds sterling (£).

| Variable | Description | Data Type | | |
|---------------------------------|---|----------------|--------------|------------|
| unique_id | Unique number assigned to that record during processing | time-invariant | qualitative | ordinal |
| no_people | Number of persons in household | time-variant | qualitative | ordinal |
| household_type | Type of household – index, wealthy or pensioner | time-variant | qualitative | nominal |
| quarter | Quarter (Jan-Mar, Apr-Jun, Jul-Sep or Oct-Dec) | time-invariant | qualitative | ordinal |
| OECD_disposable_income | Equivalised disposable income (OECD Scale) | time-variant | quantitative | continuous |
| region | Government Office Region | time-invariant | qualitative | nominal |
| total_income | Anonymised household income and allowances | time-variant | quantitative | continuous |
| OECD_scale | OECD Scale factor | time-invariant | quantitative | scalar |
| gross_income | Gross current income of household | time-variant | quantitative | continuous |
| weekly_disposable_income | Normal weekly disposable household income - anonymised | time-variant | quantitative | continuous |
| income_source | Main source of household income | time-variant | qualitative | nominal |
| adult_food_cost | Cost of food for adults | time-variant | quantitative | continuous |
| child_food_cost | Cost of food for children | time-variant | quantitative | continuous |
| total_food_cost | Cost of food for the household | time-variant | quantitative | continuous |
| adult_total_expenditure | Total expenditure by adults | time-variant | quantitative | continuous |
| child_total_expenditure | Total expenditure by children | time-variant | quantitative | continuous |
| total_expenditure | Total expenditure by the household | time-variant | quantitative | continuous |
| eng_rural_urb | Rural and Urban Classification 2004 for England & Wales | time-invariant | qualitative | nominal |
| scot_rural_urb | Scottish Executive Urban Rural Classification | time-invariant | qualitative | nominal |
| quarterly_weight | Quarterly weight | time-invariant | quantitative | scalar |
| year | Survey year | time-invariant | qualitative | ordinal |
| no_children | No of children in the household | time-variant | qualitative | ordinal |
| no_adult | No of adults in the household | time-variant | qualitative | ordinal |
| unprocessed_food | Cost of unprocessed food for the household | time-variant | quantitative | continuous |
| processed_food | Cost of processed food for the household | time-variant | quantitative | continuous |
| ultra_processed_food | Cost of ultra-processed food for the household | time-variant | quantitative | continuous |
| no_households | Weighted number of households (ie UK total) | time-variant | quantitative | discrete |

Data Formatting

- 'eng_rural_urb' and 'scot_rural_urb' columns changed to string due to mixed data

Data Integrity Issues

- No empty rows.
- No duplicates found.
- North West & Merseyside 568 entries and North West and Merseyside 2189 entries
- NaN values in 'eng_rural_urb' (4819) and 'scot_rural_urb' (23173)
- 114 records have no food costs

Data Changed/Fixed Records

- North West & Merseyside changed to North West and Merseyside
- 114 entries with no food costs i.e. they didn't keep a food diary.

Outliers

The food columns contained outliers which were left in the data set. The diaries are checked with the respondents and there could be a number of reasons for a larger than normal spend on certain items during the survey e.g. celebrations, buying in bulk, special offers, etc. When looked at in further detail different households accounted for the outliers in the food categories that were checked. In some cases the household had spent less in other related categories, for example more on beef and nothing on poultry and lamb. Or more on processed foods such as meat and seafood and nothing on fresh or frozen meat and seafood. Removing these outliers from each variable would severely impact on the number of food entries within the data set.

Outliers that were removed.

- Oils unique_id 1469 spent £169.44 compared to the next closest entry £40.73.
- Total food costs unique_id's 1037 (£1574.87), 17487 (£1468.80), 20420 (£1572.48) compared to the next closest entry £977.45.
- Adult total expenditure unique_id 4383 spent £10404.18 compared to the next closest entry £8109.47.
- Child total expenditure unique_id 165 spent £722.94 compared to the next closest entry £467.74.

Cleaned Data Accuracy

Adult food cost

- number of outliers: 721
- max outlier value: 977.445
- min outlier value: 230.74

Adult total expenditure

- number of outliers: 1163
- max outlier value: 8109.473
- min outlier value: 1324.891

Child food cost

- number of outliers: 3306
- max outlier value: 86.02
- min outlier value: 0.05

Total food cost

- number of outliers: 711
- max outlier value: 977.445
- min outlier value: 234.34

Child total expenditure

- number of outliers: 3689
- max outlier value: 467.735
- min outlier value: 0.05

Total expenditure

- number of outliers: 1152
- max outlier value: 8109.473
- min outlier value: 1332.758

Descriptive Statistics (Cleaned Data Set)

| | no_people | adult_food_cost | child_food_cost | total_food_cost | adult_total_expenditure | child_total_expenditure | total_expenditure | year |
|-------|-----------|-----------------|-----------------|-----------------|-------------------------|-------------------------|-------------------|-------|
| count | 26151 | 26151 | 26151 | 26151 | 26151 | 26151 | 26151 | 26151 |
| mean | 2 | 90.47 | 0.82 | 91.52 | 549.03 | 2.30 | 551.04 | 2017 |
| std | 1 | 59.94 | 3.42 | 60.86 | 419.03 | 10.33 | 420.13 | 1 |
| min | 1 | 0.30 | 0.00 | 0.30 | -157.82 | 0.00 | -157.82 | 2015 |
| 0.25 | 1 | 47.18 | 0.00 | 47.46 | 279.55 | 0.00 | 280.46 | 2016 |
| 0.50 | 2 | 78.25 | 0.00 | 79.13 | 454.36 | 0.00 | 456.58 | 2017 |
| 0.75 | 3 | 120.60 | 0.00 | 122.17 | 697.62 | 0.00 | 701.24 | 2019 |
| max | 9 | 977.45 | 86.02 | 977.45 | 8109.47 | 467.74 | 8109.47 | 2020 |

- A typical household in the UK contains up to 3 people. The minimum is one and the maximum is nine people.

Data Completeness

The data set is complete. The missing values in the 'eng_rural_urb' and 'scot_rural_urb' columns are due to these variables only covering certain regions.

Living Costs and Food Survey Analysis Data Set

Data Wrangling

- Combined all the children columns to get the total number of children ('no_children').
- Combined all the adult columns to get the total number of adults (no_adult).
- Created three food group columns - unprocessed, processed⁹ and ultra-processed.¹⁰
- Used quarterly weight to calculate the number of households represented (no_households). [Weights are assigned using ONS population estimates so that each case represents a larger number of households, with the sum of weights being equal

⁹ Manufactured with the addition of unprocessed or minimally processed foods.

¹⁰ Industrial formulations manufactured mostly or entirely from sugar, salt, oils and fats, starches and many substances derived from foods not normally used in kitchens.

to the total number of UK households. The number of households represented is equal to 1,000 x the quarterly weight.]

- Removed all the columns that are no longer required, i.e. children and adults by age and food categories.
- Added weighted columns for all the relevant income and expenditure columns.
- Added columns for processed and unprocessed food as a percentage of the total food cost.
- Removed two households with an expenditure less than zero. Total expenditure below £0, could cause problems when weighted.
- Removed one household where the percentage of ultra_processed_food was greater than 100%.
- Removed one household where the percentage food categories when added together were greater than 100%

Descriptive Statistics (Wrangled Data Set)

| | no_households | W_adult_food_cost | W_child_food_cost | W_total_food_cost | W_adult_total_expenditure | W_child_total_expenditure | W_total_expenditure | W_OECD_disposable_income |
|-------|---------------|-------------------|-------------------|-------------------|---------------------------|---------------------------|---------------------|--------------------------|
| count | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 |
| mean | 20824 | 1892.47 | 16.26 | 1913.34 | 11531.06 | 45.86 | 11571.58 | 9056.96 |
| std | 8748 | 1678.02 | 73.27 | 1699.29 | 10772.83 | 222.39 | 10801.58 | 7755.28 |
| min | 4 | 0.42 | 0.00 | 0.43 | -2632.83 | 0.00 | -2632.83 | -20493.30 |
| 25% | 15587 | 822.42 | 0.00 | 829.40 | 4866.45 | 0.00 | 4880.59 | 4285.14 |
| 50% | 19851 | 1450.74 | 0.00 | 1464.15 | 8557.23 | 0.00 | 8603.24 | 7062.15 |
| 75% | 24780 | 2421.98 | 0.00 | 2450.87 | 14584.74 | 0.00 | 14650.20 | 11391.57 |
| max | 98000 | 30982.09 | 1891.94 | 31102.98 | 179804.15 | 10790.42 | 179804.15 | 133649.03 |

| | W_total_income | W_gross_income | W_disposable_income | W_unprocessed_food | W_processed_food | W_ultra_processed_food | %processed_food | %unprocessed_food | %ultra_processed_food |
|-------|----------------|----------------|---------------------|--------------------|------------------|------------------------|-----------------|-------------------|-----------------------|
| count | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 | 26149 |
| mean | 17165.48 | 16059.73 | 14211.77 | 445.34 | 132.40 | 577.70 | 24.89 | 7.25 | 33.29 |
| std | 15388.59 | 15143.57 | 12199.31 | 435.97 | 163.60 | 504.39 | 12.86 | 5.90 | 14.72 |
| min | 0.00 | 0.00 | -20493.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 25% | 6985.45 | 6073.83 | 6061.12 | 172.24 | 37.60 | 253.28 | 15.69 | 3.27 | 22.94 |
| 50% | 12498.33 | 11435.84 | 10765.16 | 334.04 | 87.08 | 447.70 | 23.30 | 6.01 | 32.03 |
| 75% | 22107.18 | 20953.45 | 18429.05 | 577.96 | 170.97 | 749.27 | 32.35 | 9.77 | 42.42 |
| max | 201567.66 | 193298.76 | 163239.94 | 8650.52 | 7572.26 | 10198.48 | 100.00 | 100.00 | 100.00 |

- There are seventeen households with an expenditure less than zero, who are all on social security benefits and across the regions. With the exception of one the period is July to September and the year 2018. These households have been left in the data set.
- The total food cost also includes items consumed outside the home and drinks. Only itemised foods are included in the unprocessed, processed and ultra-processed food categories.

UK BMI Data Set

The relevant obesity data was obtained from each health survey and compiled in excel.

Data Dictionary

None of the columns contain unstructured data.

| Variable | Description | | Data Type | |
|--------------------|--|----------------|--------------|------------|
| Region | Region/country within the UK | time-invariant | qualitative | nominal |
| % Underweight | Percentage with BMI <18.5 | time-variant | quantitative | discrete |
| % Normal | Percentage with BMI 18.5-25 | time-variant | quantitative | discrete |
| % Overweight | Percentage with BMI 25-30 | time-variant | quantitative | discrete |
| % Obese | Percentage with BMI 30-40 | time-variant | quantitative | discrete |
| % Morbidly obese | Percentage with BMI 40+ | time-variant | quantitative | discrete |
| % Overweight (25+) | Percentage with BMI 25+ | time-variant | quantitative | discrete |
| % Obese (30+) | Percentage with BMI 30+ | time-variant | quantitative | discrete |
| Year | Year the survey was carried out (if survey spans two years the latter year is used e.g. 2015/16 is 2016) | time-invariant | quantitative | continuous |
| Age Group | Age range: either 2to15 (child England and Scotland) or 16+ (adult) | time-variant | quantitative | discrete |

Data Formatting

- Reformatted: '% underweight' to 'float32', '% normal' to 'float32', '% overweight' to 'float32', '% obese' to 'float32', '% morbidly obese' to 'float32', '% overweight (25+)' to 'float32', '% obese (30+)' to 'float32', and 'year' to 'int32'.

Data Integrity Issues

- No empty rows.
- No duplicate values found.
- No mixed column types.
- % underweight 16 NaN entries
- % normal 16 NaN entries
- % overweight 11 NaN entries
- % obese 16 NaN entries
- % morbidly obese 16 NaN entries
- % overweight (25+) 16 NaN entries

Data Changed/Fixed Records

No changes made.

Descriptive Statistics

Adults (16+)

| | % underweight | % normal | % overweight | % obese | % morbidly obese | % overweight (25+) | % obese (30+) | year |
|-------|---------------|----------|--------------|---------|------------------|--------------------|---------------|------|
| count | 54 | 54 | 54 | 59 | 59 | 63 | 63 | 65 |
| mean | 1.64 | 34.07 | 35.97 | 25.28 | 3.33 | 64.44 | 28.43 | 2017 |
| std | 0.56 | 3.25 | 1.46 | 2.70 | 0.82 | 3.50 | 3.39 | 1 |
| min | 0.56 | 28.04 | 32.24 | 20.50 | 1.93 | 56.94 | 21.75 | 2015 |
| 25% | 1.26 | 31.20 | 35.21 | 23.76 | 2.82 | 61.92 | 25.76 | 2016 |
| 50% | 1.69 | 34.58 | 35.67 | 25.47 | 3.31 | 64.22 | 29.06 | 2017 |
| 75% | 2.04 | 36.51 | 36.86 | 27.26 | 3.84 | 67.58 | 30.65 | 2018 |
| max | 2.86 | 39.56 | 38.61 | 29.72 | 5.66 | 70.11 | 35.21 | 2019 |

Children (2to15)

| | % underweight | % normal | % overweight | % obese | % morbidly obese | % overweight (25+) | % obese (30+) | year |
|-------|---------------|-----------|--------------|----------|------------------|--------------------|---------------|-----------|
| count | 5.000000 | 5.000000 | 10.000000 | 5.000000 | 5.000000 | 10.000000 | 10.000000 | 10.000000 |
| mean | 1.28 | 69.860001 | 13.42069 | 6.36 | 8.879999 | 28.86521 | 15.411917 | 2017 |
| std | 0.370135 | 1.731474 | 0.989208 | 0.572713 | 1.171324 | 1.191465 | 1.07369 | 1.490712 |
| min | 0.8 | 67.699997 | 12 | 5.7 | 7.1 | 26.9 | 14.019164 | 2015 |
| 25% | 1.1 | 69 | 12.825 | 5.9 | 8.3 | 28.056515 | 14.4 | 2016 |
| 50% | 1.2 | 69.400002 | 13 | 6.4 | 9.4 | 28.763029 | 15.55 | 2017 |
| 75% | 1.6 | 71.199997 | 14.155171 | 6.7 | 9.7 | 29.85 | 16.224999 | 2018 |
| max | 1.7 | 72 | 15.3 | 7.1 | 9.9 | 30.700001 | 17 | 2019 |

Data Completeness

- % underweight, % normal, % obese, % morbidly obese, % overweight (25+) all had 16 NaN entries (around 21% of the data missing).
- % overweight had 11 NaN entries (around 15% of the data missing).

The missing data makes up more than 5% of the data set, it is mainly due to the lack of comparable data for children and only overweight and obese data available in 2015 for England. The missing data mainly relates to the child obesity statistics and a full analysis from 2016-2019 can be done on the adults.

Data Uniqueness

Data Grain: Region-Year-Age-Group

Results in a count of one record for each unique row combination and there are no duplicates.

Questions To Explore

Assumptions and Constraints

The food categories - unprocessed, processed and ultra-processed is only based on food bought by households and does not include foods consumed while eating out or drinks.

Relies on households completing the food diaries accurately.

Questions

How does consumption of ultra-processed foods, etc vary throughout the year and over the years?

Is there a rise in the proportion of ultra-processed foods being consumed?

Is there a correlation between ultra-processed food consumption and BMI?

Is there a link between income / expenditure and the consumption of ultra-processed foods?

Are there common features among households consuming different types of foods?

Are different regions consuming different types of foods?

Appendix

Unprocessed or minimally processed foods: fresh, dry or frozen fruits and vegetables; packaged grains and pulses; grits, flakes or flours made from corn, wheat, rye; pasta, fresh or dry, made from flours and water; eggs; fresh or frozen meat and fish; fresh or pasteurised milk.

Processed culinary ingredients and processed foods: sugar, oils, fats, preserved vegetables, preserved fruits, preserved meat and fish, cheese, bread and other similar ready-to-consume products.

Ultra-processed foods: sweet, fatty or salty packaged snack products; ice cream, chocolate, candies; mass-produced packaged breads, cookies, pastries, cakes; breakfast cereals; preserves; margarines; milk drinks, including 'fruit' yoghurts; cocoa drinks; reconstituted meat product.

- Index Households are non-pensioner households where the income is less than a specified amount.
- Wealthy households are non-pensioner households where the income is above a specified amount.

The specified amount is adjusted each year and throughout the year.