nesta

A review of how eating out contributes to our diets: technical appendix

Background

This technical appendix accompanies our report, *How eating out contributes to our diets*. It provides details about the data that underpins the report and the analytical methods used to build the data and analyse it.

Summary of the project methodology

This research consisted of two core components.

- Data building: we built a novel dataset combining sales data in the out-of-home (OOH) sector and nutritional information from several sources.
- Data analysis: we analysed the linked sales and nutrition data to provide individual- and sector-level statistics describing the contribution of the OOH sector to diets.

Building the data

One of the main aims of the project was to generate a dataset with complete linked sales and calorie information. Secondly, a key variable for our analysis was a reliable measure of excess weight prevalence, which we constructed by adjusting and imputing respondents' self-reported height and weight measurements.

Matching calorie and product weight information to sales

We used the OOH subset of Kantar's Take-Home Panel to capture purchases of food and non-alcoholic beverages for OOH consumption in 2021. The Kantar Worldpanel FMCG panel records food and drink household purchases for approximately 30,000 British households. The OOH subset of the FMCG panel for 2021 contains approximately 7,500 individuals aged 13+ who record their purchases on a mobile phone app. This particular analysis focused on a subset of these panellists, approximately 5,800, who were recording purchasing on a continuous basis for the lifetime of the project analysis period and met project eligibility criteria. The data collection happens at individual, rather than household, level.

The OOH panel is a comprehensive record of sales but does not record nutritional data. We used additional data sources to backfill this information. We separately followed a four-step process to backfill calories and a five-step process for product weights, outlined below. When talking about calories, we refer to a standardised measure of calories per 100g of a product. When talking about product weights, we refer to the mass of the product in kilograms. For liquid products, we convert litres to kilograms by using specific gravities for each drink type (for example, the specific gravity for fruit juices is 1.04 meaning that 100 ml of fruit juice corresponds to 104g).

All analysis and interpretation was conducted independently of Kantar. Kantar cannot independently verify the findings, nor can it endorse the views or findings of this report. Where data is sourced as Kantar, some elements have been augmented by data from additional sources

Step-by-step calorie imputation process

1. Exact matches

We used additional data provided by Kantar to find exact matches via a product identifier. The two additional data sources are the nutritional file provided to us as part of the Kantar Worldpanel FMCG take-home panel, which contains data on products sold in the retail environment, and a file with web-scraped menu and nutritional information from 92 OOH chain establishments collected by Kantar in 2021. In this file, Kantar has manually generated unique product identifiers that allow us to match products to the sales data. Using the first source we matched a large number of products that are pre-packaged: 49.9% of products in our final data are matched via this method, corresponding to 23.5% of sales. Using the web-scraped menu file we matched 2.5% of products, corresponding to 13.1% of sales.

2. Products with similar names

The second step involved matching calorie information from products that have similar names. For this step we used three data sources:

- the web-scraped menu information obtained from Kantar and already used in step 1,
- MenuTracker data from 2021 collected by the University of Cambridge and,
- product data provided by NIQ Brandbank in 2023.

MenuTracker is a quarterly collection of web-scraped menu data from 89 chain restaurants, fast food, cafe and coffee shop outlets. There is an overlap with the data collected from Kantar but MenuTracker provides the additional benefit of being collected four times over the course of the year, so increases the likelihood of including seasonal items that are only available on a limited time basis. On-pack product data is sourced from NIQ Brandbank 2023, which is a database of branded products sold on the market and their nutritional information. To match products between the sales data and each of these sources we used a data science method called fuzzy matching. This step allowed us to match 33.3% of products, corresponding to 50.1% of sales.

3. Manual matches with keywords

Using all the data sources above we filled in additional missing values by manually looking for keywords that would allow us to find appropriate calorie information for a product. For example, to fill in missing calorie information for ice cream cones we looked for all ice cream products that had the word ‘cone’ in their description and took the median value of their calories as the value for imputation. This step allowed us to match 9.7% of products corresponding to 6.1% of sales.

4. Food and drink category median

The last step involved taking the median calorie value for a food and drink category and assigning it to the remaining products that had missing calorie data in that category. This was used to fill in calorie information for the remaining 4.6% of products, corresponding to 7.2% of sales. The outcome was a database with 100% matched calorie information.

Step-by-step product weight imputation process

1. Extraction from product description and other attributes

For the majority of products (54.6%, corresponding to 33.5% of sales) the weight information is contained in several parts of the data so we used various data manipulation techniques to extract it and make it usable. For example, a product's weight might be reported within the product description as a multiplication (such as 6 x 100g), so we extracted this information and calculated the total weight (600g).

2. Exact matches

This step mirrors step 1 for calories and allowed us to match 2.6% of products, corresponding to 1.7% of sales.

3. Products with similar names

This step mirrors step 2 for calories and it allowed us to match 8.5% of products, corresponding to 16.4% of sales.

4. Manual matches with keywords

This step mirrors step 3 for calories and it allowed us to match 27.6% of products, corresponding to 39.4% of sales.

5. Food and drink category median

This step mirrors step 4 for calories and it allowed us to match 6.5% of products, corresponding to 9.0% of sales. The outcome was a dataset with 100% complete product weights.

Limitations

To our knowledge this is the first attempt of its kind to systematically and comprehensively link sales to nutritional data. However, it is worth being mindful that a large proportion of the nutrition information is based on estimates rather than actual values and as such might differ from data internal to OOH businesses, which is not in the public domain. The source data available to us covers large chain restaurants, so one of the key assumptions is that meals with similar names would have the same calorie content regardless of whether they are sold in a large chain or in a small independent venue. Sales data is self-reported and Kantar puts in place incentives and extensive panel quality measures.

To test the robustness of our findings we ran a simulation exercise using bootstrapping where we simulated the imputation exercise several times. Each time we assumed that, for a random 20% of products with imputed values, a different value was imputed within a range between +- 50% of the value that was eventually chosen. This exercise allowed us to obtain a distribution of some key statistics to understand the extent to which our final figures are robust to methodological choices. We calculate that 50% of the estimated values lie within +-3% of the average, which is the statistic we report.

Body Mass Index (BMI)

Kantar collects self-reported data on the height and weight of adults in their panels. We found that the values reported suffered from self-reporting bias, resulting in a lower prevalence of excess weight than reported by objectively collected data (where excess weight is defined as BMI values higher than 25): the prevalence of adult excess weight based on self-reported figures was 60.9%, whereas the Health Survey for England survey in 2019 and the Scottish Health Survey in 2019 found that adult excess weight prevalence was approximately 64.2% in England and 66% in Scotland.

To reliably use the self-reported data in our analysis we adjusted the BMI values using a formula developed by OHID based on the Health Survey for England data. This formula takes the gender and age of respondents into account and adjusts the values of height and weight, based on known patterns of mis-reporting of BMI in the population.

Having adjusted the height and weight values we calculated new BMI values that uplift the prevalence of excess weight. However, it should be noted that the Kantar figures are not directly comparable to the health survey statistics due to differences in reporting years and the age of the sample respondents (health surveys report figures for all individuals aged 16 and above).

Table 3. BMI distribution for the sample

	Unadjusted (Age: 18+)	Adjusted (Age: 18+)	HSE 2019 (Age: 16 +)	SHEs 2019 (Age: 16 +)
Underweight	1.9%	1.4%	1.8%	2.0%
Healthy weight	37.2%	29.3%	34.0%	33.0%
Living with overweight	33.5%	36.3%	36.2%	37.0%
Living with obesity	27.4%	33.0%	28.0%	29%
Sample	4,025	4,025	6,681	3,889

Analysing the data

Individual-level analysis

By individual-level analysis we refer to all figures that refer to a user of the OOH sector and their behaviours (key findings 1,2,8; key finding 3 uses both individual and sector level analysis).

Sample selection

We made some restrictions to the sample to ensure the robustness of our analysis. We subsetted to the months of April to December 2021 to ensure our findings are not directly impacted by COVID-19 restrictions. We also removed respondents who were under 18 years of age. The final sample contained 5,196 individuals aged 18 and above. We compared the age, gender, geographical and BMI profile of our sample to relevant population figures.

Compared to mid-year population statistics we find our sample is broadly representative of the split observed across nations and the age profile but it has a large proportion of females to which Kantar applies weights to be representative of Great Britain. However, direct comparisons with population statistics are not possible due to our sample being built to describe the users of the OOH sector, rather than the population as a whole.

Table 1. Basic sample characteristics

Adults aged 18 +	Sample	Comparable figure	Source for comparable figure
Median age	50	49	Mid year population estimates table MYE2
% Female	60.7%	51.6%	Mid year population estimates table MYE1 and MYE2 - Females

% England	87.8%	86.7%	Mid year population estimates table MYE1
% Wales	4.5%	4.8%	Mid year population estimates table MYE1
% Scotland	7.7%	8.6%	Mid year population estimates table MYE1

Calculating average daily calories purchased

To calculate the average daily calories purchased per person we took advantage of the statistical weights that are provided by Kantar. By multiplying the weight by the calorie content of the product purchased and the quantity of product purchased we obtained a figure that represents the total calories purchased in the population by similar individuals on the same day. We aggregated these across the whole reporting period and obtained a figure that represents the total number of calories purchased in the population for the months April to December 2021 for all individuals similar to the representative individual. To obtain the daily average we divided this figure by the number of days between 1 April and 31 December (275) and the number of people in the population that each respondent represents (calculated from the 2021 mid-year population and assuming each respondent represents an equivalent share of the population).

Adjustment for the COVID effect

The COVID-19 pandemic had a significant impact on OOH businesses. The size of the sector shrank during the pandemic, but customers adapted to restrictions by changing the way they engaged with the sector, with more takeaway food being purchased. By COVID effect we refer to all possible changes to the OOH food environment and customer behaviours that were a result of the pandemic restrictions.

We addressed the COVID effect by restricting the reporting period to April to December 2021. We observed that per capita calorie estimates were much lower in the first quarter of the year and they gradually increased. Please note that these are raw figures and do not account for the 'purchasing for others' effect explained in the next section.

Table 2. Daily/per capita calorie purchased for each quarter of 2021

	Jan-Mar	Apr-Jun	Jul-Sept	Oct-Dec
GB	268	325	308	398

Adjustment for 'purchasing for others' effect

The sample is designed to measure sales of purchases regardless of who is consuming the food and drinks. It is possible that people may purchase items for consumption by others and not just themselves. This would introduce bias when considering per capita figures as for some people the figure may predominantly reflect their own consumption, while for others it would reflect consumption of a larger group of people. It might also be possible that people consume OOH food that is not purchased by them. If this was the case, then our figures would be underestimating the food and drinks consumed by respondents.

To adjust for possible bias in the per capita figures introduced by purchasing for other household members, we standardise our figures using the sub-group of panellists that live in households where all household members are panellists. People may be purchasing for others outside the household, so this method does not completely adjust for the purchasing for others effect, but it goes some way towards it.

To standardise our figures we split the sample between panellists that belong to a household where every member older than 18 years old is a member of the OOH panel (1,609 individuals) and those where some adult household members do not participate in the OOH panel (3,587 individuals). We observe some significant differences between the sub-samples in terms of age, BMI and household size. The assumption is that the behaviour observed in the complete sub-sample is a more reliable estimate of per capita calorie purchased than using the full sample. However, the different population structures mean that we cannot directly infer the whole sample figures from the sub-sample alone. So to calculate a full sample figure that corrects for purchasing for others bias we calculate what average we would observe in the full sample if they had the same behaviours as the sub-sample.

To do so, we calculate the average calories purchased in the sub-sample for each group defined by household size, age and BMI group. For example, in the sub-sample we calculate that, on average, respondents who are under 35, have a BMI lower than 25 and are in a two-person household purchase 194 calories/pp/day. Considering the full sample we calculate that respondents who have the same characteristics make up 1.2% of the sample. We use these proportions for each combination as weights to calculate the overall average, which is the sum of the multiplication of each group-specific average daily calories and the respective proportion.

The raw figure for average daily calories purchased is 308 calories/day/pp. Once we adjust this to control for the purchasing for others effect the figure is revised downwards to 296 calories/day/pp.

Estimating the potential impact of food waste

Food waste is the main source of discrepancy between calories purchased and consumed. To estimate the potential magnitude of food waste in terms of calories consumed in the OOH sector we rely on a survey conducted by the Waste and Resource Action Programme (WRAP) in 2022. In this survey WRAP estimated that, in 2022, on average 18% of the food purchased in the OOH sector was thrown away, with some variation between different types of food: from 14.8% for main courses to 19.6% for sides. We estimate the size of food waste by calculating the total amount of kilograms sold for each (column kg purchased in Table 4) computing the average content of calories per kilogram of product purchased (column calories/kg in Table 4) and calculating the calories that end up in waste according to the rate of wastage for that group, normalising the figure to person per day (column calories wasted/pp/day in

Table 4). We are only able to make assumptions of food waste for the categories that WRAP reported on. For all other types of food we estimate the sector average level of food waste at 18%. We estimate that on average 58 calories/day/pp end up in waste.

Table 4. Food waste calculations

	% wasted	kg purchased	calories/kg	calories wasted/pp/day
Starters	19.6%	221 billion	4,303	3
Main course	14.8%	1,799 billion	2,432	19
Side	19.5%	222 billion	2,114	3
Desserts	18.1%	341 billion	3,349	4
Other	18.0%	2,287 billion	749	29

Identifying patterns of engagement in the OOH sector

We used a clustering algorithm called k-means to segment the population according to how they engage with the OOH sector. Clustering is an unsupervised machine learning technique that allows grouping of observations that are similar based on several features. We clustered the sample of users of the OOH sector based on: the total amount of money they spent, the calories they purchased, the number of trips they took overall as well as for each of the OOH channels (such as fast food, cafe etc). We identified 11 distinct groups of individuals that have different patterns of engagement with the sector. To describe the behaviour of each group we used pen portraits defined by the median or modal value of several background characteristics for the individuals in each cluster.

Sector-level analysis

By sector-level analysis we refer to all figures that relate to products or channels of the OOH sector (key findings 4, 5, 6, 7, 9, 10; key finding 3 uses a combination of individual and sector level analysis; figure 1 uses sector-level analysis). The only exclusion that we apply when reporting sector level statistics is to restrict to the months between April and December 2021 to ensure that findings are not affected by COVID-19 restrictions. Because we do not make the restriction that only adults be included, to take full advantage of the statistical weights provided, this data covers the full OOH Kantar panel that covers the British population that is 13 years old or older.

Measuring meals as product bundles

In addition to describing products that are sold by OOH businesses we wanted to identify products that are commonly purchased together as a way of capturing what is actually eaten by customers of the OOH sector. These feed into key findings 5 and 10.

We used association rule mining, also known as market basket analysis. More specifically, we used the apriori algorithm, which relies on computing probabilities that products are purchased alone or in conjunction with others to measure the strength of an association rule. An association rule is strong when there is a high probability that purchasing one product will lead to purchasing another one together, which means that this combination is being observed several times in the data. We apply this separately for all channels in our data. We define these as meals and they are the combination of products that are most likely to be purchased together.