A typology of food safety activities

Paper 1 – Technical Report

Food and You Waves 1-3 secondary analysis

NatCen Social Research

Social Science Research Unit

Food Standards Agency

FS409014

October 2016

Unit Report 2016/04.1b

GOVERNMENT SOCIAL RESEARCH

Social Science in Government

NatCen
Social Research that works for society

A typology of food safety activities

Technical report

Authors: Klaudia Lubian[1], Caireen Roberts[1], Sally McManus[1], Edward Eaton[2]

[1] NatCen Social Research

[2] Food Standards Agency

Date: October 2016

Prepared for: Food Standards Agency

1.Introduction

This technical report accompanies the briefing paper, entitled ‘A typology of food safety activities', alongside which it should be read. The briefing paper contains the background to the study, along with the findings from the analysis, the interpretation and conclusions. This report presents a detailed methodological description of the analysis.

The aim of the analysis was to use data from Waves 1-3 of the FSA's Food and You survey (conducted in 2010, 2012 and 2014) to explore linkages between different food-safety-related activities and so develop a typology of consumers. The Food and You survey is a biennial, random probability, cross-sectional survey of adults (16 years and over) living in private households. The analysis is based on 9,056 participants.

2.Latent Class Analysis

The first stage of the analysis was to explore whether people could be classified into distinct clusters on the basis of their reported food-safety-related activities. A typology was derived using latent class analysis (LCA), a statistical method for finding subtypes of related cases (latent classes) from multivariate data. The technique can be used to assign individual respondents to discrete, non-overlapping clusters on the basis of patterns of responses to certain survey questions. There are however, certain limitations to LCA, notably in that the selection of the model (that is, the decision about which variables to include, and the number of groups that best fits the data). The labels given to groups, can also inevitably be somewhat subjective.

The data for this analysis were modelled using the package *Latent GOLD*. A typical analysis involves fitting several models to the data with different numbers of classes. The models were compared using several criteria – including statistical properties and interpretability. We also bore in mind the FSA's objective of identifying 'at risk' groups, which helped inform the selection of the most useful model in terms of the number of classes and maximise potential insight into any patterning that was found in food-safety-related activity.

2.1 Selection of variables and number of classes

The source of data for the LCA was responses to food-safety-related questions from the FSA's Food and You survey. The survey provides a rich dataset, with a range of questions around behaviour, attitudes and knowledge relating to food safety in the home; other topics including eating habits, shopping, cooking and eating out; and a range of demographic and socio-economic information.

As the Agency commissioned this analysis to investigate whether different food-safety-related activities cluster together in the population, it was decided that the basis for the primary LCA would be variables that represented a range of different food-safety-related activities, which could help to reduce the risk of contracting foodborne illnesses both inside and outside the home:[1]

- Cleaning
- Chilling
- Leftovers
- Use of information
- Shopping
- Eating out

Using the combined dataset from Food and You Waves 1-3, 14 variables were initially selected (see Appendix A, Table A1), with the additional criteria that they had been asked at all three waves of the survey and in all four UK countries, thus providing the largest possible sample size. In selecting variables for an LCA, there are also technical considerations which have an impact on the choice of variables that are included in the final model. Formulating an LCA requires selecting a suitable number of clusters in the data, and choosing the variables which are most informative for the clusters. The modelling framework itself cannot provide for the selection of variables to be used in any one study. Therefore, selection requires understanding, both of the topic at hand and

[1] The final list of variables was selected in consultation with Alan Warde, Professor of Sociology, University of Manchester.

the variables available from which the selection is to be made. Typically, all variables that are selected are used in the model to test their ability to discriminate between clusters. Using a large number of variables can cause a problem of sparseness of data. In this case, while it may not be possible to statistically evaluate a single model, one may obtain some insight by comparing the fit of alternative models, either with a difference chi-squared test, or with parsimony indices. Often, removing unnecessary variables and parameters (those that do not discriminate between the clusters) can also improve classification performance and the precision of parameter estimates.

Another technical consideration is the fact that the usual LCA model assumes that variables are independent within latent classes. This is usually an untenable assumption as it is almost impossible to explain complex relationships between different behaviours with one unobserved latent variable. For example, two items may be alternative measures of the same basic construct, or may measure closely-related traits. In such cases, variables would be assumed to be associated within latent classes, a situation termed *local dependence*. As this analysis required the inclusion of food-safety-related variables that were highly correlated with each other, the standard LCA model had to be modified to account for this, by allowing for associations between variables, which meant increasing the number of parameters to be estimated in the model.

Estimation of the model including all 14 variables confirmed that most of them were conditionally dependent, and it was not possible to estimate fully-specified models (accounting for *local dependence*) which included more than eight variables.[2] Reducing the number of variables meant that some dimensions (such as eating out) were now only represented by one question. This reduced our confidence in being able to describe clusters in terms of broader dimensions of food-safety-related activity. A partially-specified model suggested that the population could be divided into five clusters. This assumption was used in the process of choosing the final list of variables to be included, which was based on a criterion of the amount of the variance of each variable being explained by the model. R^2 values were used to indicate how much of the variance of each variable was explained by the model and as the variable with the smallest R^2 value was removed, the LCA model was re-estimated and the new variable that was 'least' explained by the model was identified.[3] From eight variables it was possible to fully specify the model so we then re-considered the number of clusters. The statistical parameters suggested choosing between a model with three, five or six clusters (because, as shown in Table 1, the 4-class model did not provide a significant improvement in terms of fit to data over the 3-class model). The 3-class model would have provided the best fit to data due to a very small error of classification (1.7%, Table 2). However, the 3-class model would not have helped identify 'at risk' groups in terms of those reporting activities that were not in line with FSA recommendations, and so would have been of less use to the FSA. In order to identify groups of interest to the FSA, and to avoid creating clusters that were too small (as in the case of the 6-class model), the 5-cluster model was selected.

[2] Trying to estimate models accounting for local dependence with more than 8 variables was not possible due to technical limitations.

[3] In practice this means that questions were removed with response patterns that were not significantly contributing to the clustering of respondents.

Table 1: Statistical comparison of models with different number of clusters (-2LL Diff bootstrap test for the difference between the models)			
Models compared	**-2LL Diff**	**p-value**	**interpretation**
3-cluster vs 2-cluster	2566	0.00	Since p < 0.05, the 3-Class Model does provide a significant improvement over the 2-class model
4 vs 3	0.7	1.00	Since p > 0.05, the 4-Class Model does not provide a significant improvement over the 3-class model
5 vs 4	226	0.00	Since p < 0.05, the 5-Class Model does provide a significant improvement over the 4-class model
6 vs 5	211	0.00	Since p < 0.05, the 6-Class Model does provide a significant improvement over the 5-class model

Table 2: Model parameters: 3 to 5-class models with 8 variables included					
Models	**L^2**	**BIC(LL)**	**AIC(LL)**	**Npar**	**Class.Err.**
3-Cluster	4583.817	95805	95528	39	1.7%
4-Cluster	4396.678	95709	95361	49	2.3%
5-Cluster	3973.95	95350	94952	56	20.7%
6-Cluster BVRs	3784.441	95183	94762	66	21.2%

The models considered are accounting for local dependencies

The final list of the eight variables included in the model is presented in Table 3 with 9,056 respondents providing responses to all eight.[4] Table 3 shows that all variables contribute significantly to the model,[5] with R^2 values indicating the extent to which the variance of each indicator is explained by the model. The variable '*How do you know that the food has been reheated properly?'* proved to differentiate among the groups most strongly (the LCA model explains 99% of the variance of the responses to this question). In other words, people's responses concerning re-heating food help, to the greatest extent (compared to the other seven variables) to derive five distinguishable clusters.

Table 3: Variables included in the 5-class model			
Variable	**Wald**	**p-value**	**R^2**
Food re-heated properly (how do you know that…)	60414.096	<0.000	0.999
Washing raw meat and poultry	85250.674	<0.000	0.664
How many times re-heating food	120.7014	<0.000	0.632
Washing hands after handling raw meat/fish	158.161	<0.000	0.617
Check use-by dates when cooking	688.9162	<0.000	0.308
Check use-by dates when buying	22.0573	<0.000	0.135
Cooking food to steaming hot	35.648	<0.000	0.073
Storing raw meat and poultry in the fridge	76.7363	<0.000	0.067

Final model, accounting for local dependencies

[4] The question about checking use-by-dates had only a few non-responses (which were then re-classified as 'Never' checking). For all other questions, non-responses were entered into the model as a separate 'Not Applicable' category.

[5] A significant p-value (p<0.05) associated with Wald statistic means that the indicator discriminates between the clusters in a statistically significant way.

The briefing paper accompanying this report provides a summary description of the main characteristics of each of the five clusters, outlining some key distinctive features. The next section in this report presents the statistical analyses that were used to derive the descriptions of the clusters.

2.2 Classifying individuals and describing classes

The 5-class model revealed that more than half of the sample was clustered in the first cluster (54%). In general, the clusters were not greatly differentiated, which was reflected in the limitation of the model's ability to predict the classification of respondents into clusters.[6] The classification of cases was done by modal assignment, based on data from responses to the eight food safety-related activities, which means that the cases are classified to a cluster for which they have the highest membership probability. Our model misclassifies around 20% of the cases. Prediction could possibly be improved by inclusion of other variables, e.g. demographic and socio-economic factors. However, as the aim of our analysis was to find out whether people could be clustered based on their food-safety-related activities, these variables were not used as predictors at this stage, but they were included as covariates in an 'inactive' mode. It was then possible to profile the identified clusters, without influencing the parameters of the model, by using observed frequencies of socio-economic and demographic variables, as well as other food-related variables not included in the original model.

Thus, we have used conditional probabilities for the eight food-safety-related activities to describe the clusters (Appendix A, Table A2). Observed frequencies of socio-economic, demographic and other food-related variables have been used to provide additional description for the profile of each cluster (Appendix A, Table A3).

[6] *Reduction of Errors*, which is a pseudo R-squared statistic indicating how well one can predict class memberships based on the observed variables, has a value of 0.5 (the closer it is to 1 the better the predictions).

3.Key differentiating factors - Chi-Squared Automatic Interaction Detection (CHAID) analysis

To identify the key factors that explain the differences between the clusters, a statistical technique called*Chi-Squared Automatic Interaction Detection* (CHAID) analysis was used. CHAID analysis is a tree-based segmentation technique and an effective approach for obtaining clusters that are predictive of a nominal dependent variable. Each of the resulting clusters, depicted as a node in a tree diagram, is defined as a combination of categorical predictors (e.g. age, income) of the dependent variable. The general rules for interpreting the output from the CHAID analysis are as follows: descriptive entries in each tree node consist of the sample size and the distribution of the dependent variable, so that it is possible to see how groups defined by the significant predictors differ with regards to distribution of the dependent variable. Chi-squared goodness-of-fit tests are used to identify significant predictors, and to merge predictor categories that do not differ in their prediction of the dependent variable - a significant advantage over other statistical techniques.

We ran a CHAID analysis using various potential predictors of a classification to five clusters (modal assignment from the LCA model). A number of models were tested, and one was selected with the best predictive ability, i.e. the highest percentage of correctly-classified cases. Using the information gathered in variables entered into the model, only 56% of cases can be correctly classified to the clusters. The reasons may be twofold: 1) small differences between the clusters with regards to socio-economic/demographic variables, 2) classification error in the modal assignment from the LCA model.

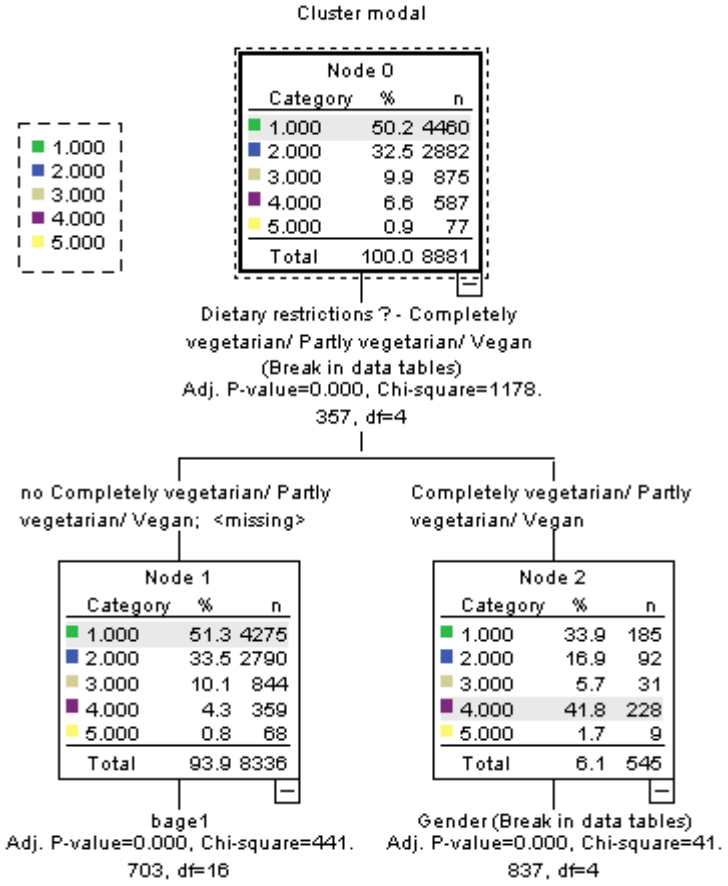
The following variables were included as potential predictors: country, age, region, gender, marital/relationship status, ethnicity, being vegetarian/vegan, reported allergy to certain food, children under 16 in the household, religion, household size, working status, household annual income, housing tenure, disability/long lasting illness, highest educational qualification, and socio-economic status.

The most significant predictor of classification, according to chi-squared goodness-of-fit tests, was **whether the respondent is completely/partially vegetarian or vegan**. It means that this variable, compared with the other covariates included as potential predictors, can offer the best explanation for the differences between the identified clusters (different patterns of food-related practices). Node 0 in Figure 1 presents the distribution of all the cases with membership to one of the five clusters allocated using modal assignment from the LCA model. Splitting the sample by vegetarian/vegan and non-vegetarian/vegan produces two new nodes: 1 and 2, with a significantly different distribution of cases to clusters. For example, cluster four makes up 42% of the vegetarian/vegan sample, compared to 4% of the non-vegetarian/vegan sample.

The next differentiating factors differ between vegetarian/vegan and non-vegetarian/vegan respondents. For the vegetarian/vegan group it is gender, while for the non-vegetarian/vegan group it is age.[7]

[7] The full tree can be viewed at the following link: https://www.food.gov.uk/sites/default/files/fs409014-paper1-chaid-analysis-tree.png

Figure 1: CHAID classification tree (nodes 0-3)



4.Predicting membership to classes - logistic regression

Two clusters were chosen for comparison: Cluster 1 'The majority' (53.7% of the sample) with a mean IRP score of 61 and Cluster 2 'Most in line with recommendations' (28.9%) with a mean IRP score of 72. This was to better understand the activities and nature of the people in the 'Most in line with recommendations' category, so that we could gain insights that might support the aim of improving behaviours for people in the 'The majority' category.

In total, 7,232 cases were included in the model (96.6% of respondents classified to Clusters 1 and 2).[8] The dependent variable comes from the modal assignment of the cases from the LCA into Custers 1 and 2. Possible predictors (socio-economic/demographic and other food-related questions not used in the creation of the clusters) were first cross-tabulated with the two clusters for an initial review of relationships. A list of the factors (socio-economic/demographic and food-related) that were considered as potential predictors of class membership is presented in tables B1 and B2 in Appendix B.

We entered the variables into the model in blocks at a time: first socio-economic/demographic variables (block 1) and then food-related variables (block 2), which allowed for assessing the difference between the two models with regards to their predictive abilities.[9] The first step, called Step 0, includes no predictors and just the intercept. Estimation of a model with socio-economic/demographic characteristics yielded a highly significant model chi-square statistic (indicating that the model including the predictors is significantly better than one without those predictors). However, a significant Hosmer and Lemeshow test value[10] indicated that the model predicted the real-world data quite poorly. Information about the respondents' socio-economic characteristics allows for correct prediction of classification of around 82% of respondents classified to Cluster 1 but only 35% to Cluster 2.[11] Therefore, we may say that the derived clusters are not distinguishable using only information about respondents' socio-economic/ demographic characteristics.

Inclusion of the second block of variables allowed us to test whether we can improve the prediction of the classification using the data about respondents' food-related activities (other than those used to derive the clusters – Table B2 in Appendix B). Reduction in the -2 Log likelihood value[12] tells us that the model is better at predicting the classification than it was before the new variables were added. The ability to predict classification to Cluster 2 improved significantly to 42% of the cases. A non-significant value of the Hosmer-Lemeshow test confirmed that predictions made by the new model fitted better with observed group memberships.[13] Few variables proved to be significant predictors of the classification to either Cluster 1 or 2. The results of the regression for the significant variables (when controlling for any

[8] Not all of respondents could be included in the model due to missing values for some of the predictors.
[9] For some of the variables we allowed for the missing values to be valid categories included in the model.
[10] Chi-square=16.204, p-value=0.04. This statistic tests the hypothesis that the observed data are significantly different from the predicted values from the model. So, in effect, a non-significant value for this test would be desirable (i.e. > 0.05, because this would indicate that the model does not differ significantly from the observed data).
[11] In logistic regression, the classification of a case is based on the predicted probability that the case will be an event (the higher value on the dependent variable), as calculated with the current model equation. By default, a case is predicted to be in the event class (say, the 1 in dependent variable coded as 0 and 1) if its predicted probability is equal to at least 0.5. For this analysis the cut-off value for prediction was set to 0.5.
[12] Reduction from 9170 to 8902 in -2Log likelihood value.
[13] Chi-square= 10.134, p-value=0.256.

other variables) are shown in Appendix B, Table B3. They can be interpreted in combination with Table A3 in Appendix A, which shows the percentage distribution, as well as an average for the whole sample.

Appendix A. Latent class analysis

Table A1 List of 14 variables and their dimensions

Dimension	Question	Number of valid responses (missing values)	Included in the final model
Chilling	Do you ever check your fridge temperature?	9789 (58)	
Chilling	Where/how you store raw meat and poultry in the fridge?	8917 (930)	Yes
Cleaning	Do you do the following things at all when you are in the kitchen and if so how frequently; Wash raw meat and poultry	8755 (1092)	Yes
Cleaning	Do you do the following things at all when you are in the kitchen and if so how frequently; Wash hands after handling raw meat/fish	9284 (563)	Yes
Cooking	Do you do the following things at all when you are in the kitchen and if so how frequently; cook food to steaming hot	9517 (330)	Yes
Cooking	Do you do the following things at all when you are in the kitchen and if so how frequently; Eat chicken or turkey if the meat is pink or has pink or red juices	9368 (479)	
Cooking	How do you usually tell that food has been re-heated properly?	8015 (1832)	Yes
Cooking	How often do you cook for others?	9746 (101)	
Eating out	How often have you eaten out in the last 7 days?	7740 (2107)	
Eating out	Generally, when you're deciding where to eat out, which of the following are important to you? - A good hygiene rating/score	7740 (2107)	
Information	Do you check use by dates when you are about to cook or prepare food?	9847	Yes
Information	Do you check use-by dates when you are buying food?	9847	Yes
Leftovers	If you made a meal on Sunday, What is the last day that you would consider eating the leftovers?	9847	
Leftovers	How many times would you consider re-heating food after it was cooked for the first times?	9056 (791)	Yes

Table A2 Conditional probabilities of variables included in model					
	The majority	*Most in line with recommend-ations*	*Cook but never re-heat*	*Least likely to handle raw meat*	*Least likely to be involved in cooking*
Cluster Size (n=9056)	53.7%	28.9%	9.9%	6.7%	0.9%
How many times re-heating food					
Not at all	0	0	1	0	1
Once	0.86	1	0	0.86	0
Twice or more	0.14	0	0	0.14	0
Check use-by dates when buying					
Never	0.07	0.01	0.03	0.12	0.13
Sometimes/depending on the food type	0.27	0.1	0.19	0.33	0.34
Yes, always	0.66	0.89	0.78	0.55	0.53
Washing hands after handling raw meat/fish					
N/A	0	0	0.01	0.73	0.63
Not following recommended practice	0.13	0	0.13	0.04	0.05
Following recommended practice	0.87	1	0.86	0.23	0.32
Washing raw meat and poultry					
N/A	0.05	0.03	0.04	0.98	0.99
Not following recommended practice	0.64	0.65	0.64	0.02	0.01
Following recommended practice	0.31	0.32	0.32	0.01	0
Cooking food to steaming hot					
N/A	0.01	0	0.01	0.36	0.49
Not following recommended practice	0.22	0.05	0.22	0.14	0.11
Following recommended practice	0.78	0.95	0.77	0.49	0.39
Check use-by dates when cooking					
Never	0.11	0.01	0.05	0.19	0.27
Sometimes/depending on the food type	0.29	0.15	0.23	0.33	0.36
Yes, always	0.6	0.84	0.72	0.48	0.38
Storing raw meat and poultry in the fridge					
N/A	0.09	0	0.09	0.44	0.35
Not following recommended practice	0.62	0.51	0.63	0.39	0.45
Following recommended practice	0.28	0.49	0.29	0.17	0.2
Food re-heated properly (how do you know that…)					
N/A	0	0	1	0	1
Not following recommended practice	0.34	0.18	0	0.35	0
Following recommended practice	0.66	0.82	0	0.65	0

Table A3 Socio-economic and demographic factors and food-related activities: comparison of 5 clusters

Factor	Categories		The majority	Most in line with recommend-ations	Cook but never re-heat	Least likely to handle raw meat	Least likely to be involved in cooking	Total
Mean of IRP score			61	72	67	60	63	66
Country	England	%	86	81	82	85	83	84
	Wales	%	4	6	6	5	6	5
	Scotland	%	7	10	9	8	7	8
	Northern Ireland	%	3	3	2	2	4	3
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Age (8 categories)	16-24	%	17	11	11	23	6	15
	25-34	%	18	20	11	13	8	17
	35-44	%	16	20	13	14	5	17
	45-54	%	18	20	17	14	10	18
	55-64	%	13	14	15	13	18	14
	65-74	%	11	11	20	12	19	12
	75-84	%	6	4	11	9	26	6
	85+	%	2	1	3	3	9	2
	Unweighted bases		4266	3216	954	525	88	9049
	Weighted bases		4499	2996	892	583	80	9049
Age (3 categories)	16-44	%	51	50	34	49	19	49
	45-74	%	42	45	52	39	47	44
	75+	%	8	5	14	12	35	8
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Region	North East	%	3	5	4	3	8	4
	North West	%	10	12	12	12	11	11
	Yorkshire and The Humber	%	8	9	10	7	7	8
	East Midlands	%	7	8	5	7	9	7
	West Midlands	%	8	9	10	8	8	9
	East of England	%	10	9	10	8	6	9
	London	%	15	10	8	17	7	13
	South East	%	15	11	17	12	14	14
	South West	%	9	7	7	11	13	9
	Wales	%	4	6	6	5	6	5
	Scotland	%	7	10	9	8	7	8
	Northern Ireland	%	3	3	2	2	4	3
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Gender	Male	%	52	38	51	66	77	49
	Female	%	48	62	49	34	23	51
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Marital/ relationship status	Married/living as married	%	50	58	54	48	55	53
	Single/widowed/divorced/separated	%	50	42	46	52	45	47
	Unweighted bases		4267	3216	953	525	88	9049
	Weighted bases		4502	2996	889	584	80	9050

Ethnicity	White	%	85	91	96	83	86	88
	Other	%	13	8	3	16	6	11
	missing	%	1	1	1	1	8	1
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Completely vegetarian/ Partly vegetarian/ Vegan	No	%	95	96	96	61	89	93
	Yes	%	4	3	3	38	11	6
	missing	%	-	1	-	1	1	1
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Allergic to certain food	No	%	94	94	97	97	98	95
	Yes	%	5	5	3	3	1	5
	missing	%	-	1	-	1	1	1
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Child aged under 16 in the household	Yes	%	26	31	21	23	13	27
	No	%	74	69	79	77	87	73
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Child aged under 5 in the household[14]	No	%	87	85	90	92	97	87
	Yes	%	13	15	10	8	3	13
	Unweighted bases		2923	2158	626	348	64	6119
	Weighted bases		3102	2013	588	380	55	6139
Religion	Christian	%	60	66	69	48	71	62
	Non-christian	%	9	5	2	17	6	8
	No religion	%	30	28	28	35	22	30
	missing	%	1	1	1	1	2	1
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Household size	One	%	17	13	21	14	25	16
	Two	%	35	38	43	33	52	37
	Three	%	18	21	15	24	6	19
	Four	%	18	19	15	19	11	18
	Five or more	%	12	9	6	9	6	10
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Work status	In work	%	55	62	43	51	29	56
	Retired	%	21	18	36	23	49	22
	Unemployed	%	5	5	6	5	7	5
	Other[15]	%	19	15	15	20	15	17
	Unweighted bases		4268	3217	954	527	88	9054
	Weighted bases		4501	2996	892	586	80	9055
Household annual income	Up to 10,399	%	10	11	15	9	14	10
	10,400 to 25,999	%	22	24	29	19	25	23
	26,000 to 51,999	%	24	26	18	22	18	24
	52k+	%	20	20	14	17	12	19
	missing	%	23	19	23	33	32	23
	Unweighted bases		4270	3217	954	527	88	9056

[14] Question asked only in Waves 2 and 3
[15] Includes those in full-time education and caring for family and home

	Weighted bases		4503	2996	892	586	80	9056
Housing tenure	Owner occupier	%	64	65	64	64	71	64
	Private tenant	%	18	16	12	15	7	16
	Social tenant	%	14	17	22	13	20	15
	Other	%	-	-	1	2	-	1
	missing	%	4	2	2	6	2	3
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Disability/ long-lasting illness	Yes	%	16	16	18	18	40	17
	No	%	84	84	82	82	60	83
	Unweighted bases		4270	3215	954	527	88	9054
	Weighted bases		4503	2994	892	586	80	9054
Highest educational qualification	Degree or higher	%	29	24	12	28	17	26
	A level/ Diploma/ Apprenticeship	%	34	35	29	31	28	34
	GCSE	%	19	25	27	21	15	22
	Other/ None	%	18	16	32	20	40	19
	Unweighted bases		4261	3209	951	525	87	9033
	Weighted bases		4494	2983	888	582	79	9025
Socio-economic status (NS-SEC)	Large employers and higher managerial and administrative occupations	%	4	5	4	4	3	4
	Higher professional occupations	%	11	8	4	9	5	9
	Lower managerial, administrative and professional occupations	%	25	24	18	23	24	24
	Intermediate occupations	%	8	10	9	8	6	9
	Small employers and own account workers	%	10	9	11	15	16	10
	Lower supervisory and technical occupations	%	9	12	12	8	8	10
	Semi-routine occupations	%	12	14	18	8	7	13
	Routine occupations	%	10	11	13	15	15	11
	Never worked and long-term unemployed	%	1	1	2	1	6	1
	Not classified	%	8	6	8	7	7	7
	missing	%	1	1		2	2	1
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Eating chicken or turkey when pink/red juices	No	%	92	96	95	48	78	90
	Sometimes	%	4	2	3	5	-	3
	Yes/ most of the time	%	2	1	2	1	-	2
	missing	%	2	1	1	47	21	5
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Storing raw meat and poultry in the fridge	NRP	%	68	46	55	35	38	57
	RP	%	22	54	36	15	20	34
	missing	%	10	-	8	50	42	9
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
A good hygiene rating/score	No	%	76	66	72	76	78	72
	Yes	%	24	34	28	24	22	28

important when eating out[16]	Unweighted bases		3363	2516	741	409	76	7105
	Weighted bases		3555	2365	701	449	68	7138
Checking fridge temperature	Not following recommended practice	%	52	43	48	49	47	48
	Following recommended practice	%	48	57	52	51	53	51
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Cooking for others	At least once a day	%	34	47	39	15	8	38
	3-6 times a week	%	18	18	16	10	3	17
	2-8 times a month	%	22	18	20	15	7	20
	Once a month or less	%	26	16	24	59	83	25
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
How many times re-heating food	Not at all	%	-	-	100	-	100	11
	Once	%	85	100	-	86	-	81
	Twice or more/ DK	%	15	-	-	14	-	9
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Check use-by dates when buying	Never	%	7	-	5	15	23	5
	Sometimes/depending on the food type	%	35	3	16	26	13	22
	Yes, always	%	58	97	80	59	64	73
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Washing raw meat and poultry	Not following recommended practice	%	61	67	69	1	-	60
	Following recommended practice	%	33	31	27	-	-	29
	missing	%	6	1	4	98	100	11
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Cooking food to steaming hot	Not following recommended practice	%	28	-	12	16	12	16
	Following recommended practice	%	72	100	87	47	38	81
	missing	%	1	-	1	38	50	3
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Check use-by dates when cooking	Never	%	12	-	8	20	38	8
	Sometimes/depending on the food type	%	39	5	17	30	11	25
	Yes, always	%	50	95	75	50	51	67
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
The last day would consider eating leftovers	Not following recommended practice	%	24	12	8	20	9	18
	Following recommended practice	%	76	88	92	80	91	82
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
How do you usually tell that food has been re-heated properly?	Not following recommended practice	%	40	9	-	40	-	26
	Following recommended practice	%	60	91	-	60	-	64
	missing	%	-	-	100	-	100	11
	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
How many times have you eaten out in the last 7 days?	Never	%	18	21	25	20	39	20
	Once	%	25	27	25	26	30	26
	More than twice	%	36	32	29	31	17	33
	missing	%	21	21	21	23	15	21

[16] Question asked only of a third of respondents in Wave 1

	Unweighted bases		4270	3217	954	527	88	9056
	Weighted bases		4503	2996	892	586	80	9056
Responsible for food shop[17]	All/ Most	%	49	55	56	26	21	50
	Half	%	20	23	21	18	11	21
	Little/ None	%	31	22	24	56	68	29
	Unweighted bases		3353	2499	738	414	69	7073
	Weighted bases		3532	2313	691	451	59	7045
Ever had food poisoning?	Yes	%	39	40	36	31	28	39
	No	%	54	55	60	64	63	55
	missing	%	7	5	4	4	9	6
	Unweighted bases		4005	4270	3217	954	527	88
	Weighted bases		4197	4503	2996	892	586	80

[17] Question asked only of a third of respondents in Wave 1

Appendix B Logistic regression

Table B1 Socio-economic and demographic factor category distribution for Clusters 1 and 2

Factor	Category	n
Socio-economic status (NS-SEC)	Large employers and higher managerial and administrative occupations	248
	Higher professional occupations	640
	Lower managerial, administrative and professional occupations	1715
	Intermediate occupations	744
	Small employers and own account workers	653
	Lower supervisory and technical occupations	712
	Semi-routine occupations	995
	Routine occupations	873
	Never worked and long-term unemployed	119
	Not classified	533
Household annual income	Up to 10,399	1050
	10,400 to 25,999	1903
	26,000 to 51,999	1691
	52k+	1149
	missing	1439
Household size	One	2028
	Two	2637
	Three	1120
	Four	981
	Five or more	466
Work status	In work	3804
	Retired	1837
	Unemployed	418
	Other[19]	1173
Housing tenure	Owner occupier	4699
	Private tenant	1107
	Social tenant	1275
	Other	30
	missing	121
Highest educational qualification	Degree or higher	1816
	A level/ Diploma/ Apprenticeship	2357
	GCSE	1527
	Other/ None	1532
Country	England	4481

[19] Includes those in full-time education and caring for family and home

Table B1 Socio-economic and demographic factor category distribution for Clusters 1 and 2 (cont.)		
	Wales	538
	Scotland	1075
	Northern Ireland	1138
Age	16-24	633
	25-34	1210
	35-44	1301
	45-54	1287
	55-64	1163
	65-74	947
	75-84	537
	85+	154
Ethnicity	White	6690
	Other	542
Religion	Christian	5074
	Non-christian	339
	No religion	1819
Disability/long-lasting illness	Yes	1442
	No	5790
Gender	Male	2820
	Female	4412
Marital/relationship status	Married/living as married	3468
	Single/widowed/divorced/separated	3764
Allergic to certain food	No	6870
	Yes	362
Completely vegetarian/ Partly vegetarian/ Vegan	No	6967
	Yes	265
Child aged under 16 in the household	Yes	1901
	No	5331

Table B2 Food-related practices category distribution for Clusters 1 and 2		
Factor	**Category**	**n**
How often do you cook or prepare food for others?	At least once a day	2998
	3-6 times a week	1149
	2-8 times a month	1416
	Once a month or less	1669
Responsible for food shop	All/ Most	3578
	Half	986
	Little/ None	1056
	missing	1612
How many times have you eaten out in the last 7 days?	Never	1556
	Once	1945
	More than twice	2152
	missing	1579
Eating chicken or turkey when pink/red juices	No	6805
	Sometimes	204
	Yes/ most of the time	113
	missing	110
Had food poisoning?	Yes	2673
	No	4155
	missing	404
What is the last day that you would consider eating the leftovers?	NRP	1195
	RP	6037
A good hygiene rating important when eating out	No	4020
	Yes	1633
	missing	1579
Do you ever check your fridge temperature?	Not following recommended practice	3675
	Following recommended practice	3557

In Table B3 we have presented the odds ratios of the significant factors from the final model (p-value of less than 0.05) after controlling for other factors included in the model (presented in the Tables B1 and B2). A statistically significant factor means the outcome of the model varies according to that factor. If the factor is significant we can then look at the p-values for each of the categories within each factor, if the p-value for a category is less than 0.05 then the category is significantly different from the reference category.

The odds ratio (Exp(B)) is the change in odds; if the value is greater than 1 then it indicates that as the predictor increases, the odds of the outcome occurring increase. Conversely, a value less than 1 indicates that as the predictor increases, the odds of the outcome occurring decrease. Where there is a trend in the odds ratio, for example, with age, we have discussed this in the main briefing paper. If the odds ratio for a factor is statistically significant but there is no trend, this suggests that the statistical finding could be a chance occurrence, rather than an informative result.

Table B3 Significant predictors of the classification to Cluster 2 compared to Cluster 1 when controlling for socio-economic and demographic factors and food related-questions (other than those used in the LCA model)

		OR	95% C.I.		p-value
			Lower	Upper	
Gender (p<0.001)	Male (Ref)				
	Female	1.8	1.6	2.0	0.000
Age (p<0.001)	16-24 (Ref)				
	25-34	1.5	1.2	1.8	0.000
	35-44	1.4	1.1	1.8	0.002
	45-54	1.3	1.0	1.6	0.028
	55-64	1.2	0.9	1.6	0.135
	65-74	1.0	0.8	1.4	0.801
	75-84	0.7	0.5	1.1	0.099
	85+	0.5	0.3	0.9	0.030
Country (p<0.001)	England (Ref)				
	Wales	1.5	1.2	1.9	0.000
	Scotland	1.3	1.1	1.6	0.001
	Northern Ireland	1.1	0.8	1.5	0.523
Ethnicity (p=0.033)	White (Ref)				
	Other	0.8	0.6	1.0	0.033
Marital status (p=0.003)	Married/living as married (Ref)				
	Single/widowed/divorced/separated	0.8	0.7	0.9	0.003
Completely vegetarian/ Partly vegetarian/ Vegan (p=0.004)	No (Ref)				
	Yes	0.6	0.5	0.9	0.004
Child aged under 16 in the household (p=0.025)	No (Ref)				
	Yes	1.2	1.0	1.4	0.025
Religion (p=0.007)	Christian (Ref)				
	Non-christian	0.7	0.5	0.9	0.002
	No religion	0.9	0.8	1.1	0.255

Table B3 Significant predictors of the classification to Cluster 2 compared to Cluster 1 when controlling for socio-economic and demographic factors and food related-questions (other than those used in the LCA model) (cont.)

Work status (p=0.001)	In work (Ref)				
	Retired	0.9	0.7	1.1	0.303
	Unemployed	0.9	0.7	1.1	0.388
	Other	0.7	0.6	0.8	0.000
Household size (p=0.001)	One (Ref)				
	Two	1.0	0.8	1.2	0.932
	Three	0.9	0.7	1.2	0.480
	Four	0.8	0.6	1.1	0.161
	Five or more	0.6	0.5	0.8	0.001
Highest educational qualification (p<0.001)	Degree or higher (Ref)				
	A-level/ Diploma/ Apprenticeship	1.2	1.0	1.3	0.034
	GCSE	1.4	1.2	1.6	0.000
	Other/ None	1.0	0.8	1.2	0.811
Socio-economic status (NS-SEC) (p<0.001)	Large employers and higher managerial and administrative occupations (Ref)				
	Higher professional occupations	0.6	0.5	0.9	0.003
	Lower managerial, administrative and professional occupations	0.8	0.6	1.0	0.054
	Intermediate occupations	1.0	0.7	1.3	0.847
	Small employers and own account workers	0.7	0.5	1.0	0.025
	Lower supervisory and technical occupations	1.0	0.8	1.4	0.759
	Semi-routine occupations	0.9	0.7	1.2	0.389
	Routine occupations	0.8	0.6	1.1	0.213
	Never worked and long-term unemployed	1.2	0.7	2.1	0.550
	Not classified	0.7	0.5	1.0	0.051
Responsible for food shop (p=0.033)	All/ Most (Ref)				
	Half	1.2	1.1	1.5	0.007
	Little/ None	1.0	0.9	1.2	0.866
	missing	1.0	0.8	1.2	0.916
Eating chicken or turkey when pink/red juices (p<0.001)	No (Ref)				
	Sometimes	0.5	0.4	0.7	0.000
	Yes/ most of the time	0.6	0.4	0.9	0.023
	missing	0.6	0.3	0.9	0.014
Hygiene score important when deciding where to eat out (p<0.001)	No (Ref)				
	Yes	1.5	1.3	1.7	0.000
Do you ever check your fridge temperature? (p<0.001)	Not recommended practice				
	Recommended practice	1.4	1.3	1.6	0.000
How many times have	Never (Ref)				

Table B3 Significant predictors of the classification to Cluster 2 compared to Cluster 1 when controlling for socio-economic and demographic factors and food related-questions (other than those used in the LCA model) (cont.)

you eaten out in the last 7 days? (p=0.029)	Once	0.9	0.8	1.0	0.103
	More than twice	0.8	0.7	1.0	0.010
	missing	0.8	0.6	1.0	0.019
What is the last day that you would consider eating the leftovers? (p<0.001)	Not recommended practice				
	Recommended practice	2.0	1.7	2.3	0.000
Cooking for others (p=0.006)	At least once a day (Ref)				
	3-6 times a week	0.9	0.7	1.0	0.050
	2-8 times a month	0.9	0.8	1.1	0.313
	Once a month or less	0.8	0.6	0.9	0.001

*Model also included age & gender combined but this was not significant