

An Intercontinental Machine Learning Analysis of Factors Explaining Consumer Awareness of Food Risk

Alberto Tonda^{a,b}, Christian Reynolds^c, Rallou Thomopoulos^{d,*}

^aUMR 518 MIA-PS, INRAE, AgroParisTech, University of Paris-Saclay, France

^bUAR 3611 ISC-PIF, CNRS, France

^cCentre for Food Policy, City, University of London, United Kingdom

^dIATE, University of Montpellier, INRAE, Institut Agro, France

Abstract

Food safety is a common concern at the household level, with important variations across different countries and cultures. Nevertheless, identifying the factors that best explain similarities and differences in consumer awareness pertaining to this topic is not straightforward. Starting from a questionnaire administered in seven countries from four continents (Argentina, Brazil, Colombia, Ghana, India, Peru, and the United Kingdom), we present an analysis of the answers related to food safety concerns, aimed at identifying possible explanatory factors. As classical statistical approaches can be limited when dealing with complex datasets, we propose an analysis with machine learning techniques, that can take into account both categorical and numerical values. With the questionnaire as a base, we task a machine learning algorithm, Random Forest, with predicting consumers' answers to the target questions using information from all other answers. Once the algorithm is trained, it becomes possible to obtain a ranking of the questions considered the most important for the prediction, with the top-ranked questions likely representing explanatory factors. Top-ranked questions are then analyzed using a Random Forest regression algorithm, to test possible correlations. The results show that the most significant explanatory variables of safety concerns seem to be estimates of carbon footprints and calories

*Corresponding author

Email addresses: alberto.tonda@inrae.fr (Alberto Tonda),
christian.reynolds@city.ac.uk (Christian Reynolds),
rallou.thomopoulos@inrae.fr (Rallou Thomopoulos)

associated with food products, and primarily with beef and chicken meat. These results tend to indicate that people who are most concerned about food safety are also those who are highly aware of environmental and nutritional impacts of food, hinting at differences in food education as a possible underlying explanation for the data.

Keywords: Food Habits, Food Knowledge, Random Forest, Risk Perception, Survey, Classification, Regression

1. INTRODUCTION

To approach households' habits and beliefs at the domestic level, questionnaires and surveys have been an effective means of study. Among the fields of study, practices and perceptions about food, and more specifically food safety, have been well represented both recently (Sollid et al. 2022, Wallace et al. 2022) and in earlier decades (Cuperus et al. 1996, Medeiros et al. 2004), in various countries (Wilcock and Ball 2014, Parikh et al. 2022, Ma et al. 2019, EFSA 2019).

The collection of large quantities of data and their analysis for scientific research, social assessment, or business purposes, naturally moved to the digital world (see e.g., Jin et al. (2020) for a recent review). On the one hand, the ease of participation offered by internet and the IoT (Internet of Things) boosted this type of studies, so that datasets are increasingly being made available (Kurtz and Thomopoulos 2021, Salliou et al. 2019). On the other hand, the handling of the collected data benefited from the progress of storage technologies, while the boom of data science offers emerging efficient analysis methods (Tao et al. 2020, Vidal et al. 2015, Kurtz and Thomopoulos 2021).

Literature has addressed perceptions of food risk by consumers at different geographic scales (Haas et al. 2021, Tucker et al. 2006, Van Kleef et al. 2007). Most recent studies have focused on emerging topics related to food safety, such as the perceived safety of novel sources of proteins (Jarchlo and King 2022), or the impact of the COVID-19 pandemic on food safety perception (Sollid et al. 2022). High expectation for healthy food has been confirmed as the number-one priority for the general public (Thomopoulos et al. 2021). Within health expectations, safety comes first, closely followed by nutritional values, especially in families with young children, as shown in Kurtz and Thomopoulos (2021). The latter study also showed differential

29 levels of concern depending on the audience (families, health professionals,
30 etc.). This observation is further explored in the present work.

31 This paper is based on a survey on food-related habits and opinions,
32 carried out in 7 countries and 4 continents during March/April 2020. In
33 the analysis, we aim to explore the question: “What variables best separate
34 individuals who express worries about food safety risks, from those who do
35 not?”. In a first step, we create two classes of responders, by separating those
36 who mark at least one item over a set threshold of concern from the rest. We
37 then train a Random Forest classifier to separate the two classes, and we later
38 analyze the variables that the classifier deems more important for the classifi-
39 cation procedure, in order to find elements that might separate the two types
40 of responders. The questions identified as most relevant refer to responders’
41 guesses on greenhouse gas emissions and caloric content of meat. In a second
42 step, we use a Random Forest regressor to predict the quantitative answers
43 to these two questions, and we then analyze again the other variables ranked
44 as most important by the regressor, to find possible explanations. Section 2
45 presents the survey data used, their pre-processing, and the analysis methods
46 used for classification and regression. Section 3 provides and discusses the
47 results obtained from the two-step method applied. Section 4 summarizes
48 the outcomes and provides perspectives for future works.

49 **2. MATERIALS AND METHODS**

50 *2.1. The data collected*

51 A survey, focused on how people cook and what they know about the most
52 common food items, was carried out at the international scale (Reynolds et al.
53 2020, Armstrong et al. 2021). The results obtained from identical questions
54 asked in 7 different countries, namely Argentina, Brazil, Colombia, Ghana,
55 India, Peru, and the United Kingdom, are used as input data in this paper.

56 The Qualtrics¹ platform was used to ask everyday people to provide their
57 opinions about images of food. For each food, the questions asked cover
58 cooking and preparation of the food, food safety, food waste, how much
59 energy is in the food, and the environmental impacts of these foods. The
60 total sample size is 3,247, and goes from a minimum of 204 (Ghana) to a

¹<https://www.qualtrics.com/>

61 maximum 539 (India) answers per country. The list of questions asked in the
62 survey is provided in Appendix A².

63 *2.2. Data pre-processing*

64 Since food habits can differ greatly between countries, only common food
65 products are considered. These common food products are: Beef, Chicken,
66 Chard, Beans, Rice, Green beans, Carrot, Tomato, and Bread (roll). Most
67 commonly used green leaves vary from one country to another (chard, col-
68 lard greens, etc.) but were considered as equivalent. For India, the “Chapat-
69 ti/Roti” item was considered as equivalent to “Bread” for the other countries.

70 Among questions concerning the socio-professional status of respondents,
71 participants were asked to input their individual and household weekly in-
72 come. As these data were provided in the local currency, to avoid issues
73 related to monetary conversion, all information related to income was nor-
74 malized with respect to other participants from the same country.

75 As the analysis is focused on a specific question (Q24), all respondents
76 who did not specify an answer were filtered out, resulting in 3,198 remaining
77 samples.

78 *2.3. Input and output variables*

79 *2.3.1. Classification step*

80 The question on risk perception (Q24) is formulated as follows: “Accord-
81 ing to your best guess, please rate how safe to eat the foods listed below
82 are? i.e. how likely is it that eating them will damage your health due to
83 risks such as contamination, food poisoning, improper handling, food fraud,
84 mislabeling etc.”. The answer requires the respondent to assign weights on
85 a scale from 0 (low risk) to 10 (high risk), to 5 foods: Beef, Chicken, Chard,
86 Rice, and Beans.

87 The output variable in our study is the maximum value expressed over
88 the five foods. It is important to notice that risk perception as described by
89 the question in the survey is highly subjective, and it is thus hard to evaluate
90 the difference between a value of 1 or a value of 2, or between a value of 7
91 and a value of 8. We thus decided to arbitrarily set a threshold around the
92 mid-value (5) and just split the respondents into two classes:

²The detailed data of the study is available on request, contacting Christian Reynolds,
Christian.Reynolds@city.ac.uk

- 93 • respondents for whom the output variable has a value below 5. This
94 class (class 0) corresponds to individuals who express a low level of
95 concern about food safety risk (1,595 respondents);
- 96 • respondents for whom the output variable has a value of 5 and more.
97 This class (class 1) corresponds to individuals who express worries
98 about food safety risk (1,603 respondents).

99 For framing the task as a classification problem, all the remaining vari-
100 ables in the survey are used as input features.

101 *2.3.2. Regression step*

102 To obtain better insight into the interpretation of the classification re-
103 sults, regressions are then performed. Their objective is to further analyze
104 the top-ranked variables of the classification results. The target variables
105 of the regressions performed, are transformations of the top 4 explanatory
106 variables obtained in the classification step. The transformations computed
107 are detailed in Section 3.3 below.

108 *2.4. The analysis methods*

109 Among tens of possible candidate classification algorithms, Random For-
110 est (RF) (Breiman 2001) was selected as the reference for the experiments,
111 as: (i) it boasted one of the highest classification accuracies in a 10-fold cross-
112 validation; (ii) it is among the few classifiers that can straightforwardly deal
113 with both categorical and non-categorical features; (iii) after being trained,
114 an instance of RF can return a series of values describing the relative im-
115 portance of the features for the final result. RF creates an ensemble of
116 decision trees, training each one on a random subset of the available data,
117 thus reducing bias and delivering more robust predictions. RF determines
118 relative variable (feature) importance as the (normalized) total reduction of
119 the criterion brought by that feature, a metric also known as the Gini impor-
120 tance (Breiman et al. 1984). For all experiments reported in this work, the
121 RF classifier has parameters selected after hyperparameter optimization (Pe-
122 dregosa et al. 2011), using a total of 300 decision trees.

123 Most classifiers, alongside their predictions, are also able to return a rank-
124 ing of the relative importance of the variables in the problem, with the ones
125 that best explain the variance in the results among the top.

126 The Python code used in the experiments is publicly available on a
127 GitHub repository³.

128 3. RESULTS AND DISCUSSION

129 3.1. Major explanatory variables of food risk perception

130 The RF classifier shows a mean accuracy on test of 0.70 in a 10-fold
131 cross-validation. After training the RF classifier, it is possible to sort the
132 features it used by their Gini importance, or in other words, their relative
133 contribution to the quality of the prediction. Figure 1 shows the first 50
134 features, ranked by decreasing Gini importance.

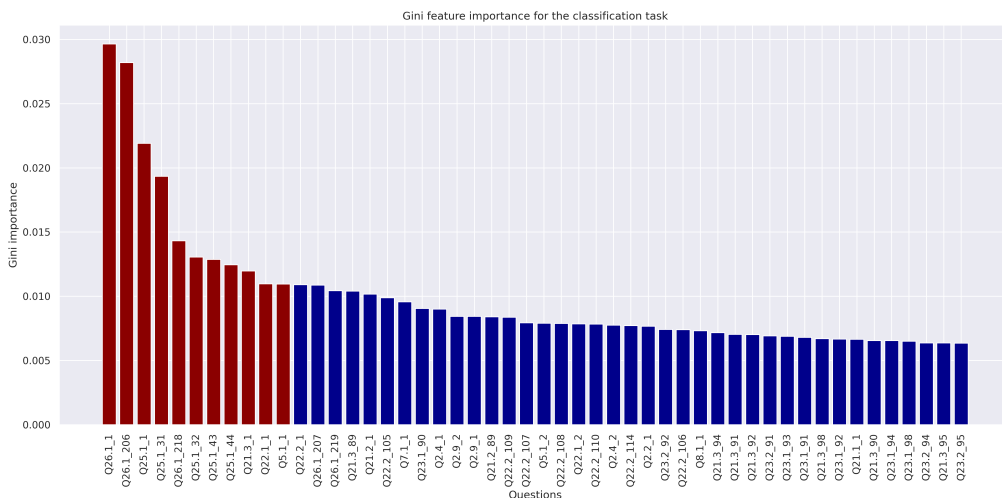


Figure 1: First 50 features, ranked by decreasing Gini importance. Features highlighted in red are later used for an in-depth analysis.

135 We then fixed an arbitrary threshold of Gini importance to perform a
136 deeper analysis, finally considering only the 11 most important variables.
137 The barplots displayed in Figure 2 show, on the X-axis, the list of variables
138 selected. The Y-axis provides the normalized mean value of each variable for
139 each of the two classes, with the corresponding standard deviation.

140 The results obtained for each country are detailed in Figure 3.

³<https://github.com/albertotonda/intercontinental-ml-analysis>

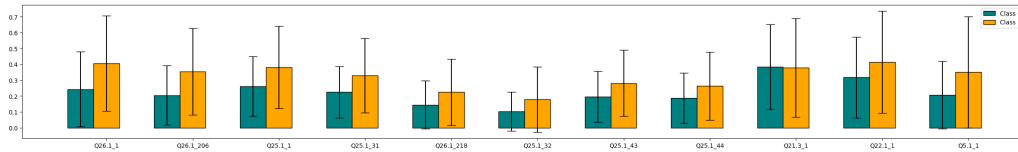


Figure 2: Mean value and standard deviation for the 11 top-ranked variables, for class 0 (teal) and class 1 (orange).

141 The survey questions corresponding to these top-ranked variables are de-
 142 tailed in Table 1, in the same order as in Figures 2 and 3.

143 Figure 4 displays the Receiver Operator Characteristics (ROC) curve for
 144 the 10-fold cross-validation, with an Area Under the Curve (AUC) of 0.76.

145 3.2. Result interpretation

146 The results reveal that the variables which best explain people’s concern
 147 about food risk are the perception of carbon footprint and the perception of
 148 calorie content, for food in general (beef or lamb, chicken, rice, green leaves,
 149 beans) and most importantly for meat products (beef or lamb, chicken) which
 150 represent the top 4 explanatory variables.

151 The results presented in the paper are quite homogeneous among the
 152 different participating countries. Some differences can be observed though,
 153 on variables Q5.1.1 (way of purchasing beef) and Q22.1.1 (method to cook
 154 beef), which are following the top-ranked 4 ones. Variable Q5.1.1 has a higher
 155 explanatory power for India and UK, the only two countries that have more
 156 than 20% of answers “I do not purchase this at all” for Q5.1.1 (21.9% for
 157 UK, 35.8% for India). This answer seems to be associated with a higher
 158 concern about food risk. The same observation applies for Q22.1.1: 18.4%
 159 of respondents answered “I do not eat this food” in UK and 23.7% in India,
 160 which seems to be associated with a higher concern about food risk.

161 Considering that meat is known to play a key part in the ecological impact
 162 of food (Godfray et al. 2018, Poore and Nemecek 2018, Vranken et al. 2014),
 163 the observation of these results raises the question of whether the “food risk
 164 concern” variable is a marker of the level of food education. Hence, we
 165 may hypothesize that the classification results obtained express a correlation
 166 between several variables representative of people’s awareness of food-related
 167 issues.



Figure 3: Results per country, top to bottom: Argentina, Brazil, Colombia, Ghana, India, Peru, and UK.

168 *3.3. Result confirmation with regression analysis*

169 In order to further explore the above hypothesis, we considered 4 new vari-
 170 ables, derived from the top 4 explanatory variables of the classification step.
 171 For each of these variables expressing respondents' estimates of greenhouse

Table 1: Most relevant explanatory questions identified

Id	Question
Q26.1_1	“According to your best guess, please estimate the carbon footprint (grams of CO ₂) embodied in the food portions that you typically eat — Beef or lamb”
Q26.1_206	“According to your best guess, please estimate the carbon footprint (grams of CO ₂) embodied in the food portions that you typically eat — Chicken”
Q25.1_1	“According to your best guess, please estimate the Calories (kcal) contained in the food portions that you typically eat — Beef or lamb”
Q25.1_31	“According to your best guess, please estimate the Calories (kcal) contained in the food portions that you typically eat — Chicken”
Q26.1_218	“According to your best guess, please estimate the carbon footprint (grams of CO ₂) embodied in the food portions that you typically eat — Rice”
Q25.1_32	“According to your best guess, please estimate the Calories (kcal) contained in the food portions that you typically eat — Green leaves”
Q25.1_43	“According to your best guess, please estimate the Calories (kcal) contained in the food portions that you typically eat — Rice”
Q25.1_44	“According to your best guess, please estimate the Calories (kcal) contained in the food portions that you typically eat — Beans”
Q21.3_1	“According to your best guess, please estimate how long (in minutes) it takes you to actively prepare the foods listed below before you to cook and eat — Beef”
Q22.1_1	“According to your best guess, please provide the typical method you used to cook the foods listed below when you eat them — Beef”
Q5.1_1	“What is the most common way you usually purchase the food items listed below? — Beef”

¹⁷² gas emissions or calories, for beef or chicken, we considered the difference,

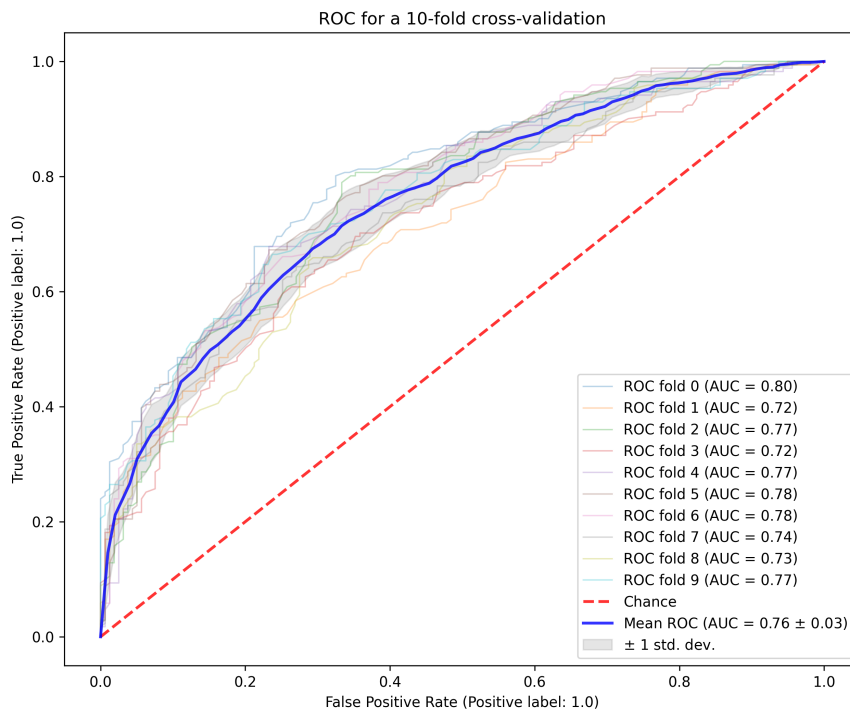


Figure 4: Receiver Operator Characteristics (ROC) curve for a 10-fold cross-validation.

173 in absolute value, from the real greenhouse gas emission / calorie value of
 174 the given food. In other words, the 4 new variables measure how much the
 175 respondents are mistaken on their assessment of greenhouse gas emissions
 176 and of caloric content, for beef and chicken, respectively.

177 The analysis was carried out resorting to a Random Forest regressor with
 178 300 estimators (regression trees), selected for the same reasons already de-
 179 tailed in Subection 2.4. The regressions were performed in two configurations:

- 180 1. Including in the explanatory variables the questions on greenhouse gas
 181 emissions (group of questions Q26) and calories (group of questions
 182 Q25), for other foods than the target one.
- 183 2. Excluding from the explanatory variables all the questions on green-
 184 house gas emissions and calories.

185 The two configurations are designed to test the assumption that a re-

186 spondent with a good knowledge of GHG emissions or caloric content would
 187 answer in a similar, satisfying way to all similar questions related to different
 188 types of food; vice-versa, a respondent not knowledgeable in this regard will
 189 likely make large mistakes for all these questions. A machine learning algo-
 190 rithm could then pick up this similarity, and start using the answers to the
 191 similar questions to predict the answer to the target. Removing the similar
 192 questions from the input would then make the predictions of the algorithm
 193 more difficult, or force it to use different variables to build its approximation
 194 of the target.

Table 2: Results of the regression experiments for the different target variables, using a 10-fold cross-validation. Mean values and standard deviation of test R2 are reported for each experiment. For reference, an R2 of 1.0 implies perfect predictions, while an R2 of 0.0 (or lower) corresponds to a poor predictive performance.

Regression target variable	R2 of a 10-fold cross-validation (including Q25 and Q26 groups of questions)	R2 of a 10-fold cross-validation (without Q25 and Q26 groups of questions)
Error on greenhouse gas emissions (beef), kg CO2	0.8007 +/- 0.0385	0.2139 +/- 0.0330
Error on greenhouse gas emissions (chicken), kg CO2	0.8281 +/- 0.0331	0.2241 +/- 0.0371
Error on caloric content (beef), kcal	0.7443 +/- 0.0395	0.2080 +/- 0.0359
Error on caloric content (chicken), kcal	0.7506 +/- 0.0266	0.2200 +/- 0.0449

195 The R2 tests obtained are reported in Table 2. From the results, we
 196 can state that it is possible to predict how much a respondent is mistaken
 197 about the greenhouse gas emissions and the calories of beef and chicken,
 198 using her/his answers about greenhouse gas emissions and calories for other
 199 foods. Interestingly, prediction performance strongly declines if we remove
 200 these explanatory variables, which might indicate that other variables in
 201 the survey do not correlate well with the regression target variable of this
 202 scenario.

203 This observation tends to confirm the existence of a correlation between
204 a group of variables representative of the level of food education.

205 4. CONCLUSIONS

206 Based on a survey on food-related habits and opinions, carried out in 7
207 countries and 4 continents, this paper investigated the factors that explain
208 people's concern about food safety. To do so, a machine learning approach
209 was proposed in two stages.

210 In the first stage, classification was used to find out the variables that best
211 separate people who worry most about food safety, from those who do not.
212 Estimates of meat carbon footprint and of meat calories revealed to be the
213 salient explanatory variables of food safety concern. In the second stage, the
214 hypothesis of a correlation between variables which are markers of people's
215 awareness of food issues, was tested and confirmed using regressions. These
216 regressions were performed on transformations of the top-ranked variables
217 obtained in the first stage.

218 From the analysis of the regression results, a correlation seems to emerge
219 between answers related to green house gas emissions and caloric content
220 of different types of food, possibly identifying groups of respondents with
221 good or scarce food education. Interestingly, no other question in the survey
222 seems to be strongly correlated, hinting that food education might be largely
223 independent from factors such as relative income. This result, if confirmed,
224 suggests that improving citizens' food education might be crucial to enhance
225 citizens' awareness about food safety related issues, and commitment to food
226 policies, whether they are linked to health, environmental, ethical or social
227 aspects. Correlation between several food-related concerns, observed in this
228 paper, has also been pointed out in previous studies (Kurtz and Thomopoulos
229 2021), independently from the general level of education which may be
230 unrelated to the food sector. Those results are thus in line with the present
231 study.

232 Nevertheless, it is also important to highlight the limitations of the anal-
233 ysis: the machine learning algorithms used in the analysis can only find
234 correlations between the variables in the survey, so there might be some
235 external, unknown factors, explaining the differences between respondents.
236 Furthermore, during the analysis, personal and family income have been
237 normalized with respect to other respondents in the same country, providing
238 just a rough approximation of reality, ignoring factors such as selection bias

239 among respondents, or country-level income distribution. It is also worth
240 mentioning that the questions related to income presented a considerable
241 number of outliers, probably because some respondents might have just re-
242 fused to provide precise data: for example, the median value of personal
243 income reported by respondents from Colombia is zero. All these factors
244 might contribute to hiding a potential correlation between wealth and food
245 education. Yet, previous studies (Yang et al. 1998, Pooler et al. 2021) do not
246 either necessarily highlight such a correlation between income and relevant
247 food choices.

248 Last but not least, the regression results did not reveal any correlation
249 between good-level food education —most likely represented by quality an-
250 swers related to green house gas emissions and caloric content of different
251 foods— and expression of responsibility for the impacts of one’s food choices
252 (represented by questions Q3.2 and Q3.3 in particular). This observation,
253 in line with Thomopoulos et al. (2021), consolidates the idea that the emer-
254 gence of changes in dietary habits is subject to resistances that should not
255 be neglected, to come to effective daily developments.

256 **References**

- 257 K. Sollid, A. Dostal, M. Paipongna, K. Smith, Food perceptions, beliefs,
258 and behaviors amid a global pandemic, *Nutrition Today* 57 (2022) 26–33.
259 doi:10.1097/NT.0000000000000524.
- 260 F. Wallace, N. Mittal, E. Lambertini, S. Nordhagen, Vendor knowledge,
261 attitudes, and practices related to food safety in low- and middle-income
262 countries: A scoping review, *J Food Prot* 85 (2022) 1069–1078. doi:10.
263 4315/JFP-21-439.PMID:35435977.
- 264 G. Cuperus, G. Owen, J. T. Criswell, S. Henneberry, Food Safety Perceptions
265 and Practices: Implications for Extension, *American Entomologist* 42
266 (1996) 201–203. URL: <https://doi.org/10.1093/ae/42.4.201>. doi:10.
267 1093/ae/42.4.201.
- 268 L. C. Medeiros, V. N. Hillers, G. Chen, V. Bergmann, P. Kendall,
269 M. Schroeder, Design and development of food safety knowl-
270 edge and attitude scales for consumer food safety education,

- 271 Journal of the American Dietetic Association 104 (2004) 1671–
272 1677. URL: [https://www.sciencedirect.com/science/article/pii/
273 S0002822304014038](https://www.sciencedirect.com/science/article/pii/S0002822304014038). doi:10.1016/j.jada.2004.08.030.
- 274 A. Wilcock, B. Ball, Food Safety: Consumer Perceptions and Practices, 2014,
275 pp. 11–29. doi:10.1002/9781118474563.ch2.
- 276 P. Parikh, N. Aparo, S. Nordhagen, H. De Steur, Food safety-related per-
277 spectives and practices of consumers and vendors in Ethiopia: A scoping
278 review, *Food Res Int* 157 (2022) 111376. doi:10.1016/j.foodres.2022.
279 111376.
- 280 L. Ma, H. Chen, H. Yan, L. Wu, W. Zhang, Food safety knowledge, attitudes,
281 and behavior of street food vendors and consumers in Handan, a third
282 tier city in China, *BMC Public Health* 19 (2019) 1128. doi:10.1186/
283 s12889-019-7475-9.
- 284 EFSA, Food safety in the EU, European Food Safety Authority, Publications
285 Office, 2019. doi:10.2805/661752.
- 286 C. Jin, Y. Bouzembrak, J. Zhou, Q. Liang, L. M. van den Bulk, A. Gavai,
287 N. Liu, L. J. van den Heuvel, W. Hoenderdaal, H. J. Marvin, Big data in
288 food safety- a review, *Current Opinion in Food Science* 36 (2020) 24–32.
289 doi:<https://doi.org/10.1016/j.cofs.2020.11.006>.
- 290 A. Kurtz, R. Thomopoulos, Consumer perceptions of infant food safety
291 in France, Technical Report, INRAE, 2021. URL: [https://doi.org/10.
292 15454/ZPP0JH](https://doi.org/10.15454/ZPP0JH). doi:10.15454/ZPP0JH.
- 293 N. Salliou, P. Taillandier, R. Thomopoulos, VITAMIN project (Vegetarian
294 Transition Argument ModellING), Technical Report, INRAE, 2019. URL:
295 <https://doi.org/10.15454/H0BUZH>. doi:10.15454/H0BUZH.
- 296 D. Tao, P. Yang, H. Feng, Utilization of text mining as a big data analysis tool
297 for food science and nutrition, *Comprehensive Reviews in Food Science
298 and Food Safety* 19 (2020) 875–894. URL: [https://ift.onlinelibrary.
299 wiley.com/doi/abs/10.1111/1541-4337.12540](https://ift.onlinelibrary.wiley.com/doi/abs/10.1111/1541-4337.12540). doi:[https://doi.org/
300 10.1111/1541-4337.12540](https://doi.org/10.1111/1541-4337.12540).
- 301 L. Vidal, G. Ares, L. Machín, S. R. Jaeger, Using twitter data for food-related
302 consumer research: A case study on “what people say when tweeting

- 303 about different eating situations”, *Food Quality and Preference* 45 (2015)
304 58–69. URL: [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0950329315001263)
305 [S0950329315001263](https://www.sciencedirect.com/science/article/pii/S0950329315001263). doi:[https://doi.org/10.1016/j.foodqual.2015.](https://doi.org/10.1016/j.foodqual.2015.05.006)
306 [05.006](https://doi.org/10.1016/j.foodqual.2015.05.006).
- 307 A. Kurtz, R. Thomopoulos, Safety vs. sustainability concerns of in-
308 fant food users: French results and european perspectives, *Sustainabil-*
309 *ity* 13 (2021). URL: <https://www.mdpi.com/2071-1050/13/18/10074>.
310 doi:10.3390/su131810074.
- 311 R. Haas, D. Imami, I. Miftari, P. Ymeri, K. Grunert, O. Meixner, Con-
312 sumer Perception of Food Quality and Safety in Western Balkan Coun-
313 tries: Evidence from Albania and Kosovo, *Foods* 10 (2021). URL: <https://www.mdpi.com/2304-8158/10/1/160>. doi:10.3390/foods10010160.
- 315 M. Tucker, S. R. Whaley, J. S. Sharp, Consumer perceptions of food-
316 related risks, *International Journal of Food Science & Technology* 41
317 (2006) 135–146. URL: [https://ifst.onlinelibrary.wiley.com/doi/](https://ifst.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2621.2005.01010.x)
318 [abs/10.1111/j.1365-2621.2005.01010.x](https://ifst.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2621.2005.01010.x). doi:10.1111/j.1365-2621.
319 [2005.01010.x](https://ifst.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2621.2005.01010.x).
- 320 E. Van Kleef, J. R. Houghton, A. Krystallis, U. Pfenning, G. Rowe,
321 H. Van Dijk, I. A. Van der Lans, L. J. Frewer, Consumer Evaluations of
322 Food Risk Management Quality in Europe, *Risk Analysis* 27 (2007) 1565–
323 1580. URL: [https://onlinelibrary.wiley.com/doi/abs/10.1111/j.](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2007.00989.x)
324 [1539-6924.2007.00989.x](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2007.00989.x). doi:10.1111/j.1539-6924.2007.00989.x.
- 325 A. I. Jarchlo, L. King, Alternative proteins: Consumer survey, *Food Stan-*
326 *dards Agency* (2022). doi:10.46756/sci.fsa.ncn554.
- 327 R. Thomopoulos, N. Salliou, P. Taillandier, A. Tonda, Consumers’ moti-
328 vations towards environment-friendly dietary changes: An assessment of
329 trends related to the consumption of animal products, in: *Handbook of*
330 *Climate Change Across the Food Supply Chain*, Springer, 2021, pp. 305–
331 319.
- 332 C. Reynolds, X. Schmidt Rivera, A. Frankowska, A. Kluczkowski, S. L. Bridle,
333 C. Martins, R. Akparibo, C. Auma, G. Bridge, M. B. Armstrong, H. Osei-
334 Kwasi, T. Bockarie, D. Mensah, Cooking as part of a global sustainable
335 food system - a 6 country pilot survey, Poster presented at the Nutrition

- 336 and Cooking Education Symposium, 12 Jun 2020, Newcastle, Australia,
337 2020. URL: <https://openaccess.city.ac.uk/id/eprint/24351>.
- 338 B. Armstrong, C. Reynolds, C. Martins, A. Frankowska, R. Levy, F. Rauber,
339 H. Osei-Kwasi, M. Vega, G. Cediél, X. Schmidt, A. Kluczkowski, R. Ak-
340 paribo, C. Auma, M. Defeyter, J. Tereza da Silva, G. Bridge, Food inse-
341 security, food waste, food behaviours and cooking confidence of UK citizens
342 at the start of the COVID-19 lockdown, *British Food Journal* 123 (2021)
343 2959–2978. doi:10.1108/BFJ-10-2020-0917.
- 344 L. Breiman, Random forests, *Machine learning* 45 (2001) 5–32.
- 345 L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification and*
346 *regression trees*, Routledge, 1984.
- 347 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel,
348 M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *Scikit-learn:*
349 *Machine learning in python*, the *Journal of machine Learning research* 12
350 (2011) 2825–2830.
- 351 H. C. J. Godfray, P. Aveyard, T. Garnett, J. W. Hall, T. J. Key, J. Lorimer,
352 R. T. Pierrehumbert, P. Scarborough, M. Springmann, S. A. Jebb, Meat
353 consumption, health, and the environment, *Science* 361 (2018) eaam5324.
- 354 J. Poore, T. Nemecek, Reducing food’s environmental impacts through pro-
355 ducers and consumers, *Science* 360 (2018) 987–992.
- 356 L. Vranken, T. Avermaete, D. Petalios, E. Mathijs, Curbing global meat
357 consumption: Emerging evidence of a second nutrition transition, *Envi-*
358 *ronmental Science & Policy* 39 (2014) 95–106.
- 359 X. Yang, B. Hsu-Hage, H. Tian, G. Hu, Q. Dong, J. Wu, M. Wahlqvist,
360 The role of income and education in food consumption and nutrient intake
361 in a Chinese population, *Asia Pac J Clin Nutr* 7 (1998) 217–26. URL:
362 <https://pubmed.ncbi.nlm.nih.gov/24393675>.
- 363 J. A. Pooler, M. Srinivasan, K. Wong, J. L. Blitstein, Food skills educa-
364 tion and low-income adults’ healthy food choices, *International Quar-*
365 *terly of Community Health Education* 23 (2021) 0272684X211004941.
366 doi:10.1177/0272684X211004941.

367 **Appendix A. Questions in the survey**

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q1.2	By ticking the button below I consent to the following	Yes/No answer
Q2.1	In which country do you currently reside?	Single choice answer
Q2.2	Please enter your age (in years)	Numerical input
Q2.3	What is your gender?	Single choice answer
Q2.4 Q2.4.1 (Height) Q2.4.2 (Weight)	Please enter your height (in cm), and weight (in kg), if you do not know, please leave blank.	Numerical inputs
Q2.5	Which of the following best describes the area you live in?	Single choice answer
Q2.6	What is your employment status	Single choice answer
Q2.7	What is the size of your household?	Single choice answer
Q2.8 Q2.8.1 (Adults aged 16 and over) Q2.8.2 (Children between 18 months and 16 years old) Q2.8.3 (Children between 0 and 18 months)	What is the size of your household?	Numerical inputs
Q2.9 Q2.9.1 (Individual) Q2.9.2 (Household)	What is your individual and total household weekly income?	Numerical inputs
Q3.1 Q3.1.6_TEXT (Comment)	How would you describe your dietary pattern?	Single choice answer with comment

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q3.2 Q3.2.1 (My health) Q3.2.2 (The environment) Q3.2.3 (Animal welfare) Q3.2.4 (The welfare of other humans)	To what extent do you agree or disagree with the following statements: “I am concerned about how the food I eat affects ...”	Array (5 point choice)
Q3.3 Q3.3.1 (I do not limit my meat intake) Q3.3.2 (Religious reasons) Q3.3.3 (Environmental concerns) Q3.3.4 (Animal welfare concerns) Q3.3.5 (I do not enjoy the taste) Q3.3.6 (Concerns for my health) Q3.3.7 (It is expensive) Q3.3.8 (Other) Q3.3.8-TEXT (Comment)	Do you limit your meat intake for any of the following reasons? (you may select more than one response)	Multiple choice with comment
Q3.4	On average, how often do you eat fast-food? (including Burgers, French fries, Potato chips)	Single choice answer

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q4.1 Q4.1_1 (Myself) Q4.1_2 (Partner) Q4.1_3 (Parents) Q4.1_4 (Children) Q4.1_5 (Other family member) Q4.1_6 (Friends) Q4.1_7 (Household staff or domestic helper) Q4.1_8 (Other) Q4.1_9 (No one cooks in my household) Q4.1_8-TEXT (Comment)	Who does the cooking in your household? (Please select all that apply)	Multiple choice with comment
Q4.2 Q4.2_1 (You) Q4.2_2 (Other household members)	How often do you and other members of your household cook or prepare food?	Array (4 point choice)
Q4.3 Q4.3_6-TEXT (Comment)	What is your main reason for cooking?	Single choice answer with comment
Q4.4 (from Q4.4_1 to Q4.4_27)	Please indicate which of these equipment you have in your kitchen. (Please select all that apply)	Multiple choice answer
Q4.5	When cooking at home, what kind of foods/ingredients do you use to prepare meals?	Single choice answer
Q4.6 (from Q4.6_1 to Q4.6_17)	When you prepare meals at home, which cooking techniques do you use? Please choose as many options as you use	Array (7 point choice)
Q4.7 (from Q4.7_1 to Q4.7_13)	Which of the following cooking techniques do you feel confident about using? Please choose as many options that apply.	Array (5 point choice)

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q4.8 (from Q4.8_1 to Q4.8_13) Q4.8_13-TEXT (Comment)	Where did you learn to cook? Who taught you to cook? Please choose as many options that apply.	Multiple choice with comment
Q4.9	What age did you start to learn to cook?	Single choice answer
Q4.10 (from Q4.10_1 to Q4.10_13)	How often do you engage in the following activities?	Array (6 point choice)
Q4.11 (from Q4.11_1 to Q4.11_4)	How often do you use these items in your cooking?	Array (5 point choice)
Q4.12	When you use a Premade Seasoning Cube (maggi, royco, onga etc.) what is the typical number of cubes that you use?	Single choice answer
Q4.13 (from Q4.13_1 to Q4.11_13)	Where do you (and your household) shop for food? Please include all shopping, including your main shopping, top-up shopping in between your main shopping trips, meat and fish, fruit and vegetables, and any other food shopping.	Multiple choice answer
Q5.1 Q5.1.1 (Beef) Q5.1.2 (Chicken) Q5.1.3 (Chard) Q5.1.4 (Beans) Q5.1.5 (Rice) Q5.1.6 (Green beans) Q5.1.7 (Carrot) Q5.1.8 (Tomato) Q5.1.11 (Bread)	What is the most common way you usually purchase the food items listed below?	Array (10 point choice)

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q7.1	<u>Beef</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q8.1	<u>Chicken</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q9.1	<u>Chard (acelga)</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q10.1	<u>Beans, in liquid</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. Do not worry about the type/colour of the beans (black, brown, yellow etc)., we are only wanting to know your typical serving size. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q11.1	<u>Rice</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q12.1	<u>Green Beans</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q13.1	<u>Carrot</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q14.1	<u>Tomato</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q16.1	<u>Bread</u> According to your best guess, please move the slider to the picture of the food that is nearest in size to your typical serving each food when you eat it. If you do not eat or cook this food, please move to the first image on the left of the slider (a big cross).	Single choice answer
Q21.1 Q21.1.1 (Beef) Q21.1.2 (Chicken) Q21.1.3 (Chard) Q21.1.4 (Beans) Q21.1.5 (Rice) Q21.1.6 (Green beans) Q21.1.7 (Carrot) Q21.1.8 (Tomato) Q21.1.11 (Bread)	How often do you eat these foods?	Array (6 point choice)
Q21.2 Q21.2.1 (Beef) Q21.2.89 (Chicken) Q21.2.90 (Chard) Q21.2.91 (Beans) Q21.2.92 (Rice) Q21.2.93 (Green beans) Q21.2.94 (Carrot) Q21.2.95 (Tomato) Q21.2.98 (Bread)	How many portions of this food do you usually cook at one time? We will ask many people the same question about these foods , so don't worry if you aren't absolutely sure. Just give us your best guess.	Numerical inputs between 0 and 25

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q21.3 Q21.3.1 (Beef) Q21.3.89 (Chicken) Q21.3.90 (Chard) Q21.3.91 (Beans) Q21.3.92 (Rice) Q21.3.93 (Green beans) Q21.3.94 (Carrot) Q21.3.95 (Tomato) Q21.3.98 (Bread)	According to your best guess, please estimate how long (in minutes) it takes you to actively prepare the foods listed below before you to cook and eat (i.e. chop, washing, mixing, weighing).	Numerical inputs between 0 and 60
Q22.1 Q22.1.1 (Beef) Q22.1.2 (Chicken) Q22.1.3 (Chard) Q22.1.4 (Beans) Q22.1.5 (Rice) Q22.1.6 (Green beans) Q22.1.7 (Carrot) Q22.1.8 (Tomato) Q22.1.11 (Bread)	According to your best guess, please provide the typical method you used to cook the foods listed below when you eat them.	Array (15 point choice)
Q22.2 Q22.2.1 (Beef) Q22.2.105 (Chicken) Q22.2.106 (Chard) Q22.2.107 (Beans) Q22.2.108 (Rice) Q22.2.109 (Green beans) Q22.2.110 (Carrot) Q22.2.111 (Tomato) Q22.2.114 (Bread)	According to your best guess, Please estimate how long (in minutes) it takes you to typically cook the foods listed below using your typical cooking method. If eaten raw please select "0".	Numerical inputs between 0 and 120

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q23.1 Q23.1_1 (Beef) Q23.1_89 (Chicken) Q23.1_90 (Chard) Q23.1_91 (Beans) Q23.1_92 (Rice) Q23.1_93 (Green beans) Q23.1_94 (Carrot) Q23.1_95 (Tomato) Q23.1_98 (Bread)	Thinking about the last time you bought the following foods, approximately what percentage of the amount you bought ended up being uneaten and thrown away (please include all food that was not eaten - e.g., put in a bin, compost bin, down the sink, given to animals etc)	Numerical inputs between 0 and 100
Q23.2 Q23.2_16 (Beef) Q23.2_90 (Chicken) Q23.2_91 (Chard) Q23.2_92 (Beans) Q23.2_93 (Rice) Q23.2_94 (Green beans) Q23.2_95 (Carrot) Q23.2_96 (Tomato) Q23.2_99 (Bread)	Thinking about the last time you cooked the following foods, approximately what percentage of the amount you cooked ended up being uneaten and thrown away (please include all food that was not eaten - e.g., put in a bin, compost bin, down the sink, given to animals etc)	Numerical inputs between 0 and 100
Q24.1 Q24.1_1 (Beef or lamb) Q24.1_31 (Chicken) Q24.1_32 (Chard) Q24.1_33 (Rice) Q24.1_34 (Beans)	According to your best guess, please rate how safe to eat the foods listed below are? i.e. how likely is it that eating them will damage your health due to risks such as contamination, food poisoning, improper handling, food fraud, mislabeling etc.	Numerical inputs between 0 and 10
Q25.1 Q25.1_1 (Beef or lamb) Q25.1_31 (Chicken) Q25.1_32 (Chard) Q25.1_43 (Rice) Q25.1_44 (Beans)	According to your best guess, please estimate the Calories (kcal) contained in the food portions that you typically eat (from your previous portion size answer).	Numerical inputs between 0 and 1000

Table A.3: Complete set of questions in the survey.

Id	Text	Type
Q26.1 Q26.1_1 (Beef or lamb) Q26.1_206 (Chicken) Q26.1_207 (Chard) Q26.1_218 (Rice) Q26.1_219 (Beans)	According to your best guess, please estimate the carbon footprint (grams of CO ₂) embodied in the food portions that you typically eat (from your previous portion size answer).	Numerical inputs between 0 and 8180
Q27.1 Q27.1.1 (We worried whether our food would run out before we got money to buy more) Q27.1.2 (The food that we bought didn't last, and we didn't have money to buy more) Q27.1.3 (We couldn't afford to eat balanced meals)	Here are several statements that people have made about their food situation. For these statements, please select the box to match if the statement was often true, sometimes true, or never true for your household in the last 12 months.	Array (4 point choice)