Future Foods 7 (2023) 100233

Contents lists available at ScienceDirect

Future Foods

journal homepage: www.elsevier.com/locate/fufo

An intercontinental machine learning analysis of factors explaining consumer awareness of food risk

Alberto Tonda [a,b], Christian Reynolds [c], Rallou Thomopoulos [d,*]

[a] UMR 518 MIA-PS, INRAE, AgroParisTech, University of Paris-Saclay, France
[b] UAR 3611 ISC-PIF, CNRS, France
[c] Centre for Food Policy, City, University of London, United Kingdom
[d] IATE, University of Montpellier, INRAE, Institut Agro, France

ARTICLE INFO

Keywords:
Food habits
Food knowledge
Random forest
Risk perception
Survey
Classification
Regression

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 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.

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

This paper is based on a survey on food-related habits and opinions, carried out in 7 countries and 4 continents during March/April 2020. In the analysis, we aim to explore the question: "What variables best

* Corresponding author.
E-mail address: rallou.thomopoulos@inrae.fr (R. Thomopoulos).

https://doi.org/10.1016/j.fufo.2023.100233
Received 12 December 2022; Received in revised form 4 May 2023; Accepted 6 May 2023

separate individuals who express worries about food safety risks, from those who do not?". In a first step, we create two classes of responders, by separating those who mark at least one item over a set threshold of concern from the rest. We then train a Random Forest classifier to separate the two classes, and we later analyze the variables that the classifier deems more important for the classification procedure, in order to find elements that might separate the two types of responders. The questions identified as most relevant refer to responders' guesses on greenhouse gas emissions and caloric content of meat. In a second step, we use a Random Forest regressor to predict the quantitative answers to these two questions, and we then analyze again the other variables ranked as most important by the regressor, to find possible explanations. Section 2 presents the survey data used, their pre-processing, and the analysis methods used for classification and regression. Section 3 provides and discusses the results obtained from the two-step method applied. Section 4 summarizes the outcomes and provides perspectives for future works.

2. Materials and methods

2.1. The data collected

A survey, focused on how people cook and what they know about the most common food items, was carried out at the international scale (Reynolds et al., 2020; Armstrong et al., 2021). The results obtained from identical questions asked in 7 different countries, namely Argentina, Brazil, Colombia, Ghana, India, Peru, and the United Kingdom, are used as input data in this paper. The Qualtrics[1] platform was used to ask everyday people to provide their opinions about images of food. For each food, the questions asked cover cooking and preparation of the food, food safety, food waste, how much energy is in the food, and the environmental impacts of these foods. The total sample size is 3247, and goes from a minimum of 204 (Ghana) to a maximum 539 (India) answers per country. The list of questions asked in the survey is provided in Appendix A[2]

2.2. Data pre-processing

Since food habits can differ greatly between countries, only common food products are considered. These common food products are: Beef, Chicken, Chard, Beans, Rice, Green beans, Carrot, Tomato, and Bread (roll). Most commonly used green leaves vary from one country to another (chard, collard greens, etc.) but were considered as equivalent. For India, the "Chapatti/Roti" item was considered as equivalent to "Bread" for the other countries. Among questions concerning the socio-professional status of respondents, participants were asked to input their individual and household weekly income. As these data were provided in the local currency, to avoid issues related to monetary conversion, all information related to income was normalized with respect to other participants from the same country.

As the analysis is focused on a specific question (Q24), all respondents who did not specify an answer were filtered out, resulting in 3198 remaining samples.

2.3. Input and output variables

2.3.1. Classification step

The question on risk perception (Q24) is formulated as follows: "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.". The answer requires the respondent to assign weights on a scale from 0 (low risk) to 10 (high risk), to 5 foods: Beef, Chicken, Chard, Rice, and Beans.

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

- respondents for whom the output variable has a value below 5. This class (class 0) corresponds to individuals who express a low level of concern about food safety risk (1595 respondents);
- respondents for whom the output variable has a value of 5 and more. This class (class 1) corresponds to individuals who express worries about food safety risk (1603 respondents).

For framing the task as a classification problem, all the remaining variables in the survey are used as input features.

2.3.2. Regression step

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

2.4. The analysis methods

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

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

The Python code used in the experiments is publicly available on a GitHub repository.[3]

3. Results and discussion

3.1. Major explanatory variables of food risk perception

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

We then fixed an arbitrary threshold of Gini importance to perform a deeper analysis, finally considering only the 11 most important variables. The barplots displayed in Fig. 2 show, on the X-axis, the list of

[1] https://www.qualtrics.com/

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

[3] https://github.com/albertotonda/intercontinental-ml-analysis

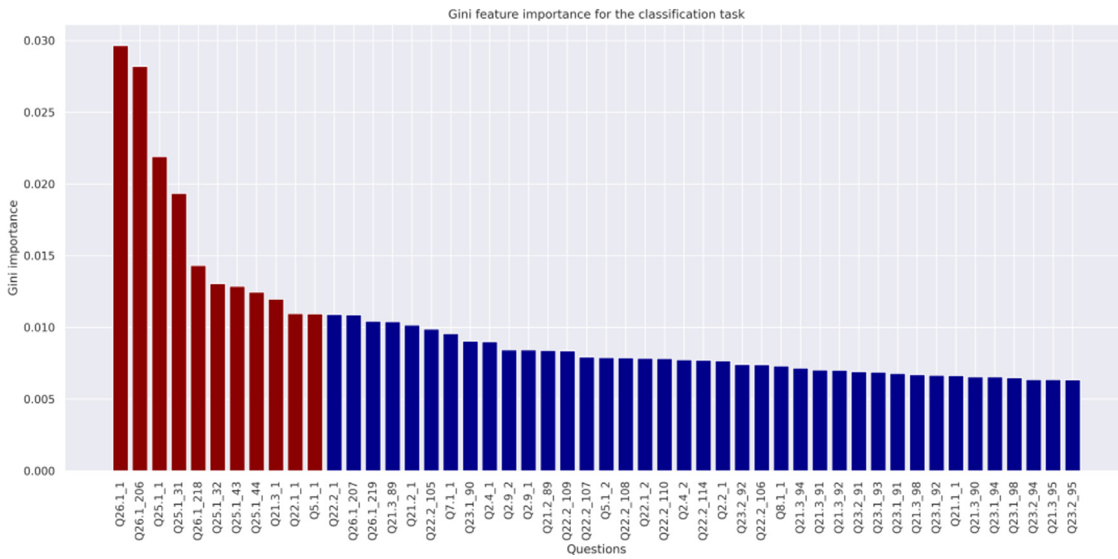

Fig. 1. First 50 features, ranked by decreasing Gini importance. Features highlighted in red are later used for an in-depth analysis.

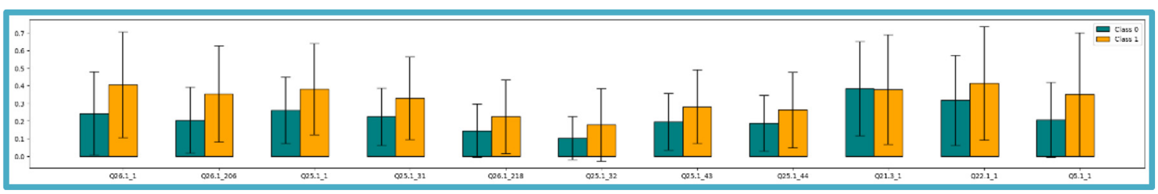

Fig. 2. Mean value and standard deviation for the 11 top-ranked variables, for class 0 (teal) and class 1 (orange).

Table 1
Most relevant explanatory questions identified.

Id	Question
Q26.1_1	"According to your best guess, please estimate the carbon footprint (grams of CO2) 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 CO2) 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 CO2) 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"

variables selected. The Y-axis provides the normalized mean value of each variable for each of the two classes, with the corresponding standard deviation.

The results obtained for each country are detailed in Fig. 3.

The survey questions corresponding to these top-ranked variables are detailed in Table 1, in the same order as in Figs. 2 and 3.

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

3.2. Result interpretation

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

The results presented in the paper are quite homogeneous among the different participating countries. Some differences can be observed though, on variables Q5.1_1 (way of purchasing beef) and Q22.1_1 (method to cook beef), which are following the top-ranked 4 ones. Variable Q5.1_1 has a higher explanatory power for India and UK, the only two countries that have more than 20% of answers "I do not purchase this at all" for Q5.1_1 (21.9% for UK, 35.8% for India). This answer seems to be associated with a higher concern about food risk. The same observation applies for Q22.1_1: 18.4% of respondents answered "I do not eat this food" in UK and 23.7% in India, which seems to be associated with a higher concern about food risk.

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

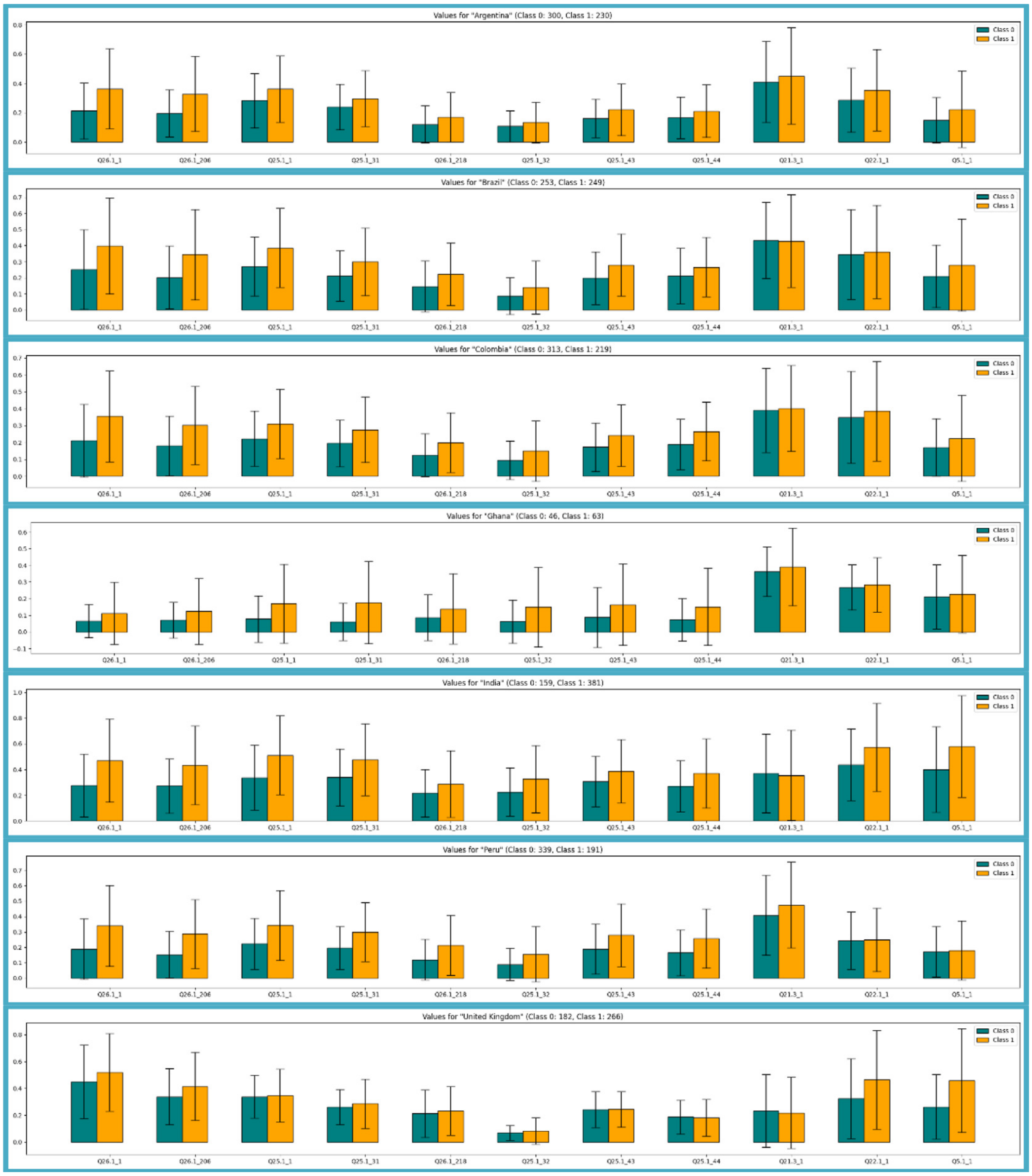

Fig. 3. Results per country, top to bottom: Argentina, Brazil, Colombia, Ghana, India, Peru, and UK.

3.3. Result confirmation with regression analysis

In order to further explore the above hypothesis, we considered 4 new variables, derived from the top 4 explanatory variables of the classification step. For each of these variables expressing respondents' estimates of greenhouse gas emissions or calories, for beef or chicken, we considered the difference, in absolute value, from the real greenhouse gas emission / calorie value of the given food. In other words, the 4 new variables measure how much the respondents are mistaken on their assessment of greenhouse gas emissions and of caloric content, for beef and chicken, respectively.

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

1 Including in the explanatory variables the questions on greenhouse gas emissions (group of questions Q26) and calories (group of questions Q25), for other foods than the target one.
2 Excluding from the explanatory variables all the questions on greenhouse gas emissions and calories.

The two configurations are designed to test the assumption that a respondent with a good knowledge of GHG emissions or caloric content would answer in a similar, satisfying way to all similar questions related to different types of food; vice-versa, a respondent not knowledgeable in this regard will likely make large mistakes for all these questions. A machine learning algorithm could then pick up this similarity, and start using the answers to the similar questions to predict the answer to the target. Removing the similar questions from the input would then make the predictions of the algorithm more difficult, or force it to use different variables to build its approximation of the target.

The R2 tests obtained are reported in Table 2. From the results, we can state that it is possible to predict how much a respondent is mistaken about the greenhouse gas emissions and the calories of beef and chicken, using her/his answers about greenhouse gas emissions and calories for other foods. Interestingly, prediction performance strongly declines if we remove these explanatory variables, which might indicate that other

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 ques- tions)	R2 of a 10-fold cross-validation (without Q25 and Q26 groups of ques- tions)
Error on greenhouse gas emissions (beef), kg CO2	0.8007 +/− 0.0385	0.2139 +/− 0.0330
Error on green-house 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

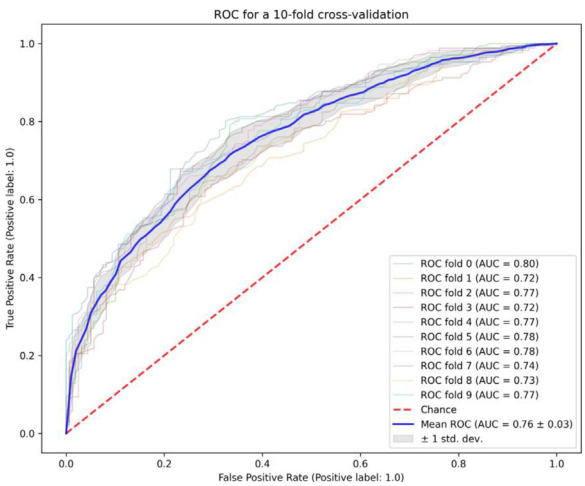

Fig. 4. Receiver operator characteristics (ROC) curve for a 10-fold cross-validation.

variables in the survey do not correlate well with the regression target variable of this scenario.

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

4. Conclusions

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

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

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

Nevertheless, it is also important to highlight the limitations of the analysis: the machine learning algorithms used in the analysis can only find correlations between the variables in the survey, so there might be some external, unknown factors, explaining the differences between respondents. Furthermore, during the analysis, personal and family income have been normalized with respect to other respondents in the same country, providing just a rough approximation of reality, ignoring factors such as selection bias among respondents, or country-level income distribution. It is also worth mentioning that the questions related to income presented a considerable number of outliers, probably because some respondents might have just refused to provide precise data: for example, the median value of personal income reported by respondents from Colombia is zero. All these factors might contribute to hiding a potential correlation between wealth and food education. Yet, previous studies (Yang et al., 1998; Pooler et al., 2021) do not either necessarily highlight such a correlation between income and relevant food choices.

Last but not least, the regression results did not reveal any correlation between good-level food education —most likely represented by quality answers related to greenhouse gas emissions and caloric content of different foods— and expression of responsibility for the impacts of one's food choices (represented by questions Q3.2 and Q3.3 in particular). This observation, in line with Thomopoulos et al. (2021), consolidates the idea that the emergence of changes in dietary habits is subject to resistances that should not be neglected, to come to effective daily developments.

Ethics statement

The study is based on a non-interventional study (a survey). All participants were fully informed why the research was being conducted, how their data would be used and if there were any risks associated. Participants gave informed consent via the statement:

"By ticking the button below I consent to the following:

I confirm that I have read and understood the information for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

I understand that my participation is voluntary and that I am free to leave the study at any time, without giving any reason. I understand that I can withdraw my data within 6 weeks of taking part in the study.

I understand that any information given by me may be used in future reports, academic articles, publications or presentations by the researcher/s, but I will not be identifiable.

I understand that my name/my organisation's name will not appear in any reports, articles or presentation without my consent.

I understand that data will be kept according to University guidelines for a minimum of 10 years after the end of the study.

I agree to take part in the above study." where an affirmative reply was required to enter the survey. They were able to withdraw from the survey at any time without giving a reason.

Declaration of Competing Interest

None

CRediT authorship contribution statement

Alberto Tonda: Methodology, Software, Formal analysis, Writing – original draft. **Christian Reynolds:** Resources, Data curation, Writing – review & editing. **Rallou Thomopoulos:** Conceptualization, Formal analysis, Visualization, Writing – original draft.

Data availability

Data will be made available on request.

Appendix A. Questions in the survey

Table A.3.

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 re-side?	Single choice answer
Q2.2	Please enter your age (in years)	Numerical in-put
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 in-puts
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 in-puts
Q2.9 Q2.9 1 (Individual) Q2.9 2 (House- hold)	What is your individual and total household weekly income?	Numerical in-puts
Q3.1 Q3.1 6 TEXT (Com-ment)	How would you describe your dietary pattern?	Single choice Answer with comment
Q3.2 Q3.2 1 (My health) Q3.2 2 (The environment) Q3.2 3 (Animal welfare) Q3.2 4 (The welfare of other hu-mans)	To what extent do you agree or dis-agree 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 con- cerns) Q3.3_4 (An- imal welfare con- cerns) 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
Q4.1 Q4.1 1 (My-self) 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 house-hold? (Please select all that apply)	Multiple choice with comment
Q4.2 Q4.2 1 (You) Q4.2 2 (Other house- hold 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 (Com-ment)	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 pre- pare 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)
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)

(*continued on next page*)

Table A.3 (*continued*)

Id	Text	Type
Q4.13 (from Q4.13 1 to Q4.11 13)	Where do you (and your household) shop for food? Please include all shop- ping, 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)
Q7.1	Beef 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	Chicken 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	Chard (acelga) 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	Beans,inliquid According to your best guess, please move the slider to the picture of the food that is near- est in size to your typical serving each food when you eat it.Do not worry about the type/color 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
Q11.1	Rice 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	GreenBeans 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	Carrot 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	Tomato 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
Q16.1	Bread 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 in-puts between 0 and 25
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 in-puts between 0 and 60

(*continued on next page*)

Table A.3 (*continued*)

Id	Text	Type
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 cook- ing method. If eaten raw please select “0″.	Numerical in-puts between 0 and 120
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, approxi- mately 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 in-puts 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, approxi- mately 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 in-puts 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 srate how safe to eat the foods listed be- low are? i.e. how likely is it that eat- ing them will damage your health due to risks such as contamination, food poisoning, improper handling, food fraud, mislabeling etc.	Numerical in-puts 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 in-puts between 0 and 1000
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 CO2) embodied in the food portions that you typically eat (from your pre- vious portion size answer).	Numerical in-puts 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 peo- ple have made about their food sit- uation. For these statements, please select the box to match if the state- ment was often true, sometimes true, or never true for your household in the last 12 months.	Array (4 point choice)

References

Armstrong, B., Reynolds, C., Martins, C., Frankowska, A., Levy, R., Rauber, F., Osei-Kwasi, H., Vega, M., Cediel, G., Schmidt, X., Kluczkovski, A., Akparibo, R., Auma, C., Defeyter, M., Tereza da Silva, J., Bridge, G., 2021. Food inse- curity, food waste, food behaviours and cooking confidence of uk citizens at the start of the covid-19 lock-down. Br. Food J. 123, 2959–2978. doi:10.1108/BFJ-10-2020-0917.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Routledge.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

G. Cuperus, G. Owen, J.T. Criswell, S. Henneberry, Food safety perceptions and prac- tices: implications for extension, American Entomologist 42 (1996) 201–203. URL: 10.1093/ae/42.4.201.

EFSA, Food safety in the EU, European Food Safety Authority, Publications Office, 2019. doi:10.2805/661752.

Godfray, H.C.J., Aveyard, P., Garnett, T., Hall, J.W., Key, T.J., Lorimer, J., Pierrehum- bert, R.T., Scarborough, P., Springmann, M., Jebb, S.A., 2018. Meat consumption, health, and the environment. Science 361, eaam5324.

Haas, R., Imami, D., Miftari, I., Ymeri, P., Grunert, K., Meixner, O., 2021. Con- sumer perception of food quality and safety in Western Balkan Countries: ev- idence from Albania and Kosovo. Foods 10. doi:10.3390/foods10010160. URL: https://www.mdpi.com/2304-8158/10/1/160.

A.I. Jarchlo, L. King, Alternative proteins: consumer survey, Food Stan- dards Agency (2022). doi:10.46756/sci.fsa.ncn554.

Jin, C., Bouzembrak, Y., Zhou, J., Liang, Q., van den Bulk, L.M., Gavai, A., Liu, N., van den Heuvel, L.J., Hoenderdaal, W., Marvin, H.J., 2020. Big data in food safety- a review. Curr. Opin. Food Sci. 36, 24–32. doi:10.1016/j.cofs.2020.11.006.

A. Kurtz, R. Thomopoulos, Consumer perceptions of infant food safety in France, Technical Report, INRAE, 2021. URL: 10.15454/ZPPOJH.

Kurtz, A., Thomopoulos, R., 2021b. Safety vs. sustainability concerns of infant food users: french results and european perspectives. Sustainability 13. doi:10.3390/su131810074. URL: https://www.mdpi.com/2071-1050/13/18/10074.

Ma, L., Chen, H., Yan, H., Wu, L., Zhang, W., 2019. Food safety knowledge, attitudes, and behavior of street food vendors and consumers in Handan, a third tier city in China. BMC Public Health 19, 1128. doi:10.1186/s12889-019-7475-9.

Medeiros, L.C., Hillers, V.N., Chen, G., Bergmann, V., Kendall, P., Schroeder, M., 2004. Design and development of food safety knowledge and attitude scales for consumer food safety education. J .Am. Diet. Assoc. 104, 1671–1677. doi:10.1016/j.jada.2004.08.030. URL: https://www.sciencedirect.com/science/article/pii/S0002822304014038.

Parikh, P., Aparo, N., Nordhagen, S., De Steur, H., 2022. Food safety-related per- spectives and practices of consumers and vendors in Ethiopia: a scoping review. Food Res. Int. 157, 111376. doi:10.1016/j.foodres.2022.111376.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: machine learning in python, the. J. Mach. Learn. Res. 12, 2825–2830.

J.A. Pooler, M. Srinivasan, K. Wong, J.L. Blitstein, Food skills educa- tion and low-income adults' healthy food choices, International Quar- terly of Community Health Education 23 (2021). doi:10.1177/0272684X211004941.

Poore, J., Nemecek, T., 2018. Reducing food's environmental impacts through pro- ducers and consumers. Science 360, 987–992.

Reynolds, C., Schmidt Rivera, X., Frankowska, A., Kluczkovski, A., Bridle, S.L., Martins, C., Akparibo, R., Auma, C., Bridge, G., Armstrong, M.B., Osei- Kwasi, H., Bockarie, T., Mensah, D., 2020. Cooking as part of a global sustainable food system - a 6 country pilot surveyProceedings of the Poster presented at the Nutrition and Cooking Education Symposium. Newcastle, Australia. 12 Jun 2020URL:.

N. Salliou, P. Taillandier, R. Thomopoulos, VITAMIN project (VegetarIan Transition Argument ModellINg), Technical Report, INRAE, 2019. URL: 10.15454/HOBUZH.

Sollid, K., Dostal, A., Paipongna, M., Smith, K., 2022. Food perceptions, beliefs, and behaviors amid a global pandemic. Nutr Today 57, 26–33. doi:10.1097/NT.0000000000000524.

Tao, D., Yang, P., Feng, H., 2020. Utilization of text mining as a big data analysis tool for food science and nutrition. Compr. Rev. Food Sci. Food Saf. 19, 875–894. doi:10.1111/1541-4337.12540. URL: https://ift.onlinelibrary.wiley.com/doi/abs/10.1111/1541-4337.12540.

Thomopoulos, R., Salliou, N., Taillandier, P., Tonda, A., 2021. Consumers' motivations towards environment-friendly dietary changes: an assessment of trends related to the consumption of animal products. In: Handbook of Climate Change Across the Food Supply Chain. Springer, pp. 305–319.

Tucker, M., Whaley, S.R., Sharp, J.S., 2006. Consumer perceptions of food- related risks. Int. J. Food Sci. Technol. 41, 135–146. doi:10.1111/j.1365-2621.2005.01010.x. URL: https://ifst.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2621.2005.01010.x.

Van Kleef, E., Houghton, J.R., Krystallis, A., Pfenning, U., Rowe, G., Van Dijk, H., Van der Lans, I.A., Frewer, L.J., 2007. Consumer evaluations of food risk management quality in Europe. Risk Anal. 27, 1565–1580. doi:10.1111/j.1539-6924.2007.00989.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2007.00989.x.

Vidal, L., Ares, G., Mach´ın, L., Jaeger, S.R., 2015. Using twitter data for food-related consumer research: a case study on "what people say when tweeting about different eating situations. Food Qual. Prefer. 45, 58–69. doi:10.1016/j.foodqual.2015.05.006. URL: https://www.sciencedirect.com/science/article/pii/S0950329315001263.

Vranken, L., Avermaete, T., Petalios, D., Mathijs, E., 2014. Curbing global meat consumption: emerging evidence of a second nutrition transition. Envi ron. Sci. Policy 39, 95–106.

Wallace, F., Mittal, N., Lambertini, E., Nordhagen, S., 2022. Vendor knowledge, attitudes, and practices related to food safety in low- and middle-income countries: a scoping review. J. Food Prot. 85, 1069–1078 doi:10.4315/JFP-21-439.PMID:35435977.

Wilcock, A., Ball, B., Food safety: consumer perceptions and practices, 2014, pp. 11–29. doi:10.1002/9781118474563.ch2.

Yang, X., Hsu-Hage, B., Tian, H., Hu, G., Dong, Q., Wu, J., Wahlqvist, M., 1998. The role of income and education in food consumption and nutrient intake in a Chinese population. Asia Pac. J. Clin. Nutr. 7, 217–226. URL: https://pubmed.ncbi.nlm.nih.gov/24393675.