

Annex A

Methodological protocol for the quality evaluation of the food consumption and related data collected under the EU Menu project

Catarina Carvalho^{1,2}, Milton Severo¹, Daniela Correia^{1,3}, Andreia Oliveira^{1,3},
Caroline van Rossum⁴, Marga Ocké⁴, Duarte Torres^{1,2}, Carla Lopes^{1,3}

1. EPIUnit/ITR, Institute of Public Health, University of Porto, Portugal,
2. Faculty of Food and Nutrition Sciences, University of Porto, Portugal,
3. Faculty of Medicine, University of Porto, Portugal,
4. RIVM, The Netherlands

Abstract

The nature, direction, and magnitude of assessment errors vary across surveys depending on the methods and procedures conducted, which highlights the importance of establishing guidelines for data quality assessment within the EU-Menu framework. The use of objective and standardized indicators of data quality across countries may help to evaluate and track the validity and reliability of harmonized food consumption and related data within and between countries. Within the ERA EU Menu project ('Evaluation, Review and Advice on methods and tools for EU Menu phase 2'), work package 2 aims to provide robust and scientific-based evidence to consent on an update of the EU menu guidance, namely by the evaluation of the current data, collected under the EU menu framework, and the assessment of their quality and harmonisation. This protocol will propose a list of quality indicators identified, as the most relevant, to be assessed in National Dietary Surveys, as well as the procedures to assess their quality, and the statistical approach to evaluate the main factors associated with the overall quality of the surveys.

© RIVM and University of Porto, 2022

Keywords: quality indicators, measurement errors, national dietary surveys, dietary assessment

Correspondence: iDATA@efsa.europa.eu



Summary

The project 'Evaluation, Review and Advice on methods and tools for EU Menu phase 2', with acronym ERA EU Menu, has been conducted by RIVM and the University of Porto (U.Porto) in a joint effort (EFSA grant: GP/EFSA/DATA/2021/03).

The objective of the ERA EU-Menu project is to give advice to EFSA for an update of the EU menu guidance, based on an extensive literature review on methodologies and tools that are currently used in National Dietary Surveys, combined with an evaluation of the data quality collected under the EU menu framework. Based on these different purposes, the ERA EU Menu project is divided into three work packages (WP): WP1: review on methods and tools; WP2: Evaluation of current data; WP3: Advise for EU-menu phase 3.

This protocol is designed specifically to answer WP2 'Evaluation of current data', in which the project team aims to:

1. identify which are the most relevant data quality indicators related to EU Menu dietary surveys;
2. define what should be the statistical approach for evaluating the overall quality of the EU Menu surveys and for finding factors associated with the quality indicators;
3. evaluate the quality and harmonisation of the surveys in the EU Menu framework and the main factors associated with overall data quality.

In this protocol, a list of identified quality indicators will be proposed, as the most relevant to be assessed in National Dietary Surveys, as well as the procedures to quantify their quality. It will also describe the statistical approach for evaluating the overall quality of the surveys and for finding factors associated to the quality indicators. Then, the quality indicators defined will be assessed along the EU Menu surveys (using both data from the literature, namely reports and survey coordinator's consultation, and from ad hoc statistical analysis) to evaluate the overall quality of indicators and to identify the major gaps.

The comparisons of these quality indicators across the EU Menu surveys will play an essential role for the updating of the EU menu guidance, and support an even more strict harmonisation and methodological robustness in future National Dietary Surveys.



Table of contents

Abstract	1
Summary	2
1 Introduction	4
1.1 ERA EU Menu	4
2 Guidelines for measuring data quality	6
2.1 Quality indicators.....	6
2.1.1 Sources of information	9
2.1.2 Description and assessment of quality indicators.....	10
2.2 Factors associated with data quality	25
3 Deliverables work package 2 Evaluation of current data	26
3.1 Consultation with EFSA.....	26
4 Planning	27
References	28
Appendix A – Surveys to be included in the data quality evaluation	30



1 Introduction

1.1 ERA EU Menu

The ERA EU Menu project ('Evaluation, Review and Advice on methods and tools for EU Menu phase 2') aims to provide robust and scientific-based evidence to consent an update of the EU menu guidance, namely by the evaluation of the current data, collected under the EU menu framework, and the assessment of their quality (objectives included in work package 2 of the refereed project).

An accurate measurement of data, particularly dietary data, across populations from different countries, as the ones included in the EU-menu framework, is a challenging task. Surveys are easily prone to random and systematic errors that might affect the accuracy and precision of the final estimates. Random errors will decrease the precision of the measurement estimates, resulting in a loss in statistical power. These random errors can result, for example, from the natural day-to-day variation in food intake that arises from differences in food intake both between persons (between- or inter-person variation) and within one person (within- or intra-person variation) (Rutishauser 2005; de Boer et al. 2011). At the same time, surveys are also prone to systematic errors that can reduce study accuracy, and that can be introduced at any stage of the survey, from the study sampling to the publication of results. Potential sources of systematic errors can be related with the use of non-probabilistic samples, the procedures used in data collection (day of the week or season reported, the methods used to quantify dietary intake, etc.), the magnitude of the energy misreporting, among others (Gibson et al. 2017). Ultimately, systematic errors will bias dietary intake measurements, yielding potentially erroneous conclusions with regard to the absolute amount of foods and nutrients consumed. Previous studies have identified procedures to overcome these errors, namely by incorporating standardized quality-control procedures and collecting more than one 24-h recall per person, as advised by EFSA guidance (EFSA 2014). Moreover, standardization of methodologies in the Pan-European context, such as the EU Menu framework, enables consistency and harmonisation of data collection for risk assessment and other purposes.

The nature, direction, and magnitude of these errors will vary across surveys depending on the methods and procedures conducted, which highlights the importance of establishing guidelines for data quality assessment within the EU-Menu framework. Quality indicators are objective, standardized, evidence-based measures that may help to collect and analyse better quality data and track the performance of accurate and harmonized food consumption outcomes within and among countries.

The specific objectives within work package 2 are:

1. To identify the most relevant data quality indicators related to EU Menu dietary surveys;
2. To describe the statistical approach for evaluating the overall quality of the surveys and for finding factors associated to the quality indicators;
3. To summarise the quality of the surveys in the EU Menu framework and its main associated factors.

The methodology used and the results on data quality will be described in a report. See also figure 1 for an overview of the activities in this work package.

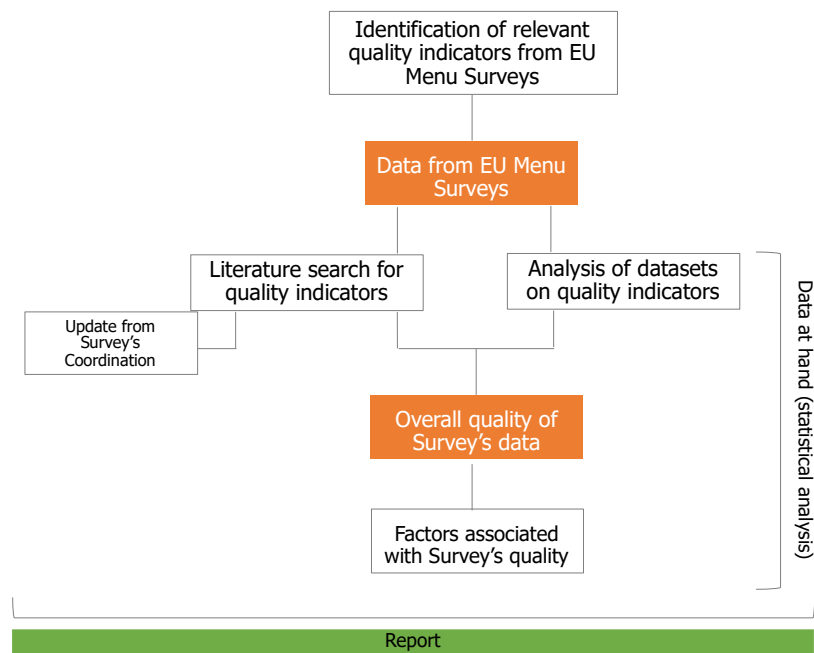


Figure 1. Graphical representation of work package 'Evaluation of current data'

The protocol describes the guidelines to assess the most relevant quality indicators, the statistical approach to assess overall quality of national dietary surveys based on the quality indicators appraised, and the statistical approach to evaluate the factors associated with data quality in the EU Menu surveys. The deliverables and the planning are described in chapters 3 and 4.

2 Guidelines for measuring data quality

2.1 Quality indicators

In the context of surveys data, 'quality' refers to how well the outputs meet the user needs, or whether they are 'fit for purpose'. As data from National Dietary Surveys may greatly differ in terms of their quality, it is important to predefine a list of quality indicators that, ideally, would be followed by all the EU-Menu surveys for harmonization and high-standard data.

Quality indicators are objective, standardized, evidence-based measures that may help to collect and analyse better quality data and track the performance of accurate and harmonized food consumption outcomes within and among surveys.


The quality indicators proposed are grouped in 9 dimensions related with the usual dietary Survey's phases.

identifies those dimensions, sub-dimensions and all indicators proposed to assess surveys quality, with hyperlinks to their full description and assessment approach. The proposed list includes all the quality indicators that we aim to assess throughout the project. However, it remains to be seen in practice which aspects can really be evaluated in the current available data (both the EU menu database and the characteristics that were extracted from the survey report). The aspects not possible to evaluate will be identified as data gaps in the project final report. A total of 29 surveys from the EU Menu framework will be assessed for the proposed quality indicators.

Table 1: Organization of quality indicators into survey dimensions and sub-dimensions

Survey Dimension	Sub-dimension	Indicators
<u>A. Sampling plan: methods and coverage</u>	<u>A1. Target population</u>	A1.1. Sampling frame A1.2 Coverage of the target population
	<u>A2. Sampling design and procedures</u>	A2.1 Probabilistic sampling design A2.2 Sampling by waves
	<u>A3. Sample size</u>	A3.1 Target sample size estimated by statistical assumptions
	<u>A4. Sampling error</u>	A4.1 Estimation of standard error for key estimates
	<u>A5. Representativeness</u>	A5.1 Study sample with similar distribution as target population A5.2 Weighting procedures
<u>B. Recruitment</u>	<u>B1. Participation rate</u>	B1.1 Participation rate, contact rate, cooperation rate
	<u>B2. Comparison between participants and non-participants</u>	B2.1 Absence of selection bias
<u>C. Training and supervision of interviewers</u>	<u>C1. Interviewers' selection criteria</u>	C1.1 Background in Nutrition/Dietetics and/or interviewing experience
	<u>C2. Training of interviewers</u>	C2.1 Standardized training procedures C2.2 Training duration and monitoring

Survey Dimension	Sub-dimension	Indicators
<u>D. Dietary data collection: Interview administration procedures</u>	<u>C3. Supervision of interviewers</u>	<u>C3.1 Interviewer monitoring</u> <u>C3.2 Observer Bias</u>
	<u>C4. Pilot survey</u>	<u>C4.1 Pilot methods and tools</u>
	<u>D1. Mode of administration of the interview</u>	<u>D1.1 At least one face-to-face interview</u> <u>D1.2 Single mode of administration to all participants at each interview</u>
	<u>D2. Interview setting and timing</u>	<u>D2.1 Interview setting</u> <u>D2.2 Average time interval between interviews</u> <u>D2.3 Distribution of dietary data by weekdays and season</u> <u>D2.4 Duration of interview</u>
<u>E. Data processing: software tools and validation procedures</u>	<u>E1. Dietary assessment software: validation and automatic quality control procedures</u>	<u>E1.1 Use of a validated dietary assessment tool</u> <u>E1.2 Use of the multi-pass method (or similar)</u> <u>E1.3 Monitoring of the interview time</u> <u>E1.4 Processing errors identification and description of quality control procedures</u>
	<u>E2. Quantification of dietary consumption</u>	<u>E2.1 Use of a validated food picture book</u> <u>E2.2 Minimum number of pictures per photo series in picture book</u> <u>E2.3 Total number of picture series in the picture book</u> <u>E2.4 Availability of food standard units as a quantification method</u> <u>E2.5 Availability of default quantities as a quantification method</u>
	<u>E3. Food propensity questionnaire (FPQ) or Food frequency questionnaire (FFQ)</u>	<u>E3.1 Application of a FPQ or FFQ</u> <u>E3.2 Adequacy of the FPQ/FFQ</u>
<u>F. Non-dietary data collection</u>	<u>F1. Anthropometric data</u>	<u>F1.1 Measured (vs. self-reported) weight and height</u> <u>F1.2 Standardized measurements of weight and height</u> <u>F1.3 Two standardized measurements available for each parameter (weight, height)</u> <u>F1.4 Digit preference in anthropometric measurements</u>
	<u>F2. Physical activity</u>	<u>F2.1 Physical activity assessment</u> <u>F2.2 Collection of accurate physical activity measurements</u>
<u>G. Data handling and cleaning</u>	<u>G1. Completeness of food composition database</u>	<u>G1.1 Number of food items/recipes in the food composition database</u> <u>G1.2 Availability of a food supplements database</u>



Survey Dimension	Sub-dimension	Indicators
		G1.3 Availability of a recipe composition database
		G1.4 Food items with only FoodEx2 base term
		G1.5 Availability of energy and nutrients (#6) per 100 grams of food
	G2. Completeness of food consumption database	G2.1 Disaggregation of recipes into ingredients and identification of mixed dishes G2.2 Missing values in non-mandatory variables G2.3 Total number of facets G2.4 Minimum recommended facets G2.5 Prevalence of foods classified as level 4 (or above) in the FoodEx2 hierarchy
	G3. Completeness of subjects' database	G3.1 Missing values in non-mandatory variables
H. Data analyses	H1. Dietary intake validity	H1.1 Energy outliers H1.2 Food groups outliers H1.3 Number of main meals per day H1.4 Minimum acceptable number of food items H1.5 Digit preference in food amounts
	H2. Data distribution and variability	H2.1 Proportion of total variance explained by differences between individuals
	H3. Indicators to validate dietary data	H3.1 Availability of data, such as biomarkers, to validate dietary intake
	H4. Energy misreporting (under and over-reporting)	H4.1 Calculation of misreporting of energy intake through recommended methods
	I1. External validity of results	I1.1 Weighted results to ensure the representativeness
I. Results reporting	I2. Usual nutritional intake	I2.1 Adjustment of nutritional intake for the intra-individual variability I2.2 Usual intake estimated using the food propensity/frequency questionnaire
	I3. Energy misreporting	I3.1 Sensitivity analysis excluding misreporters of energy intake

From a statistical point of view, the quality indicators proposed can be dichotomous or continuous variables. The assessment of continuous quality indicators can be done through a categorization into specific literature-based or expert-based cutoffs, or through a benchmark approach. This approach ascertains the benchmark, i.e., the survey with the best performance

for a certain quality indicator and describes the relative performance of the remaining surveys when compared to the point of reference.

Whenever the unit of observation to assess a quality indicator at the survey level is the participant or the interview, summary statistics for each survey, and by indicator, will be estimated using a pooled analysis, if all datasets present similar standard deviation (SD). Otherwise, a random or fixed-effects meta-analysis will be employed, depending on the heterogeneity of the quality indicator.

2.1.1 Sources of information

There are two main information sources for quality indicators assessment which are the surveys' methodological reports and the surveys' databases provided by EFSA. The data extracted from surveys' methodological reports under WP1 tasks will be used to assess the respective quality indicators in WP2. The classification of the indicators according to the respective information source is described in **Error! Reference source not found..**

Table 2: Sources of information used for assessment of quality indicators, by survey's dimensions.

	Information for quality indicators assessment	
	Methodological reports	EFSA databases
A. Sampling plan: methods and coverage	A1.1, A1.2, A2.1, A2.2, A3.1, A5.1, A5.2	A4.1
B. Recruitment	B1.1, B2.1	-
C. Training and supervision of interviewers	C1.1, C2.1, C2.2, C3.1, C3.2, C4.1	-
D. Dietary data collection: Interview administration procedures	D1.1, D1.2, D2.1, D2.4	D2.2, D2.3
E. Data processing: software tools and validation procedures	E1.1, E1.2, E1.3, E1.4, E2.1, E2.2, E2.3, E2.4, E2.5, E3.1, E3.2	-
F. Non-dietary data collection	F1.1, F1.2, F2.1, F2.2	F1.3, F1.4
G. Data handling and cleaning	G1.2, G1.3	G1.1, G1.4, G1.5, G2.1, G2.2, G2.3, G2.4, G3.1
H. Data analyses	H3.1, H4.1.1	H1.1, H1.2, H1.3, H1.4, H1.5, H2.1, H2.2, H4.1.2
I. Results' reporting	I2.1, I2.2, I3.1	-

2.1.2 Description and assessment of quality indicators

The complete list of quality indicators, divided by survey's dimensions, including a brief description of their main purpose as well as a set of questions or procedures, which have been developed to simplify their assessment, are presented in detail in **Error! Reference source not found..**

Table 3: Complete description of quality indicators according to their dimensions and sub-dimensions.

A. Sampling plan: methods and coverage

A1. Target population

A1.1 Sampling frame

The description of the sampling frame provides information if it derives from a national population register and whether it is current or out of date. If available, a national population register is recommended as sampling frame. However, if this is not available, not current (i.e. older than 10 years), or do not cover some age groups, such as small infants, other non-national sampling frames can be used as alternatives. Examples of alternative sampling frames are data from the most recent census, local population registers, lists of schools and kindergartens and lists of patients from medical centres, general practitioners (GPs) or paediatricians. Nonetheless, potential selection bias may occur from these sampling frames.

1. [Has a national population register \(updated and accessible\) been used for sampling?](#)

A1.2 Coverage of the target population

Coverage error arises when it is not possible to sample from the whole target population. Usually, it is virtually impossible to have complete and accurate lists of target populations against which to check frame coverage.

It is important to acknowledge coverage error as it may reduce the representativeness of the study sample, and impact on the accuracy of key estimates.

1. [Does the sampling frame cover the entire target population? \(e.g. if the target population are adults 18-74y, the sampling frame should be a registry in which all adults are included; if they are children, schools could be a good sampling frame to children with compulsory education at that age\).](#)

2. [Does the Survey's report estimates of under coverage, duplication, ineligibility and/or misclassification to give an indication of coverage error?](#)

A2. Sampling design and procedures

A2.1 Probabilistic sampling design

The sample design affects the accuracy of population estimates. Random probability sample designs, such as sample random sampling, stratified random sampling, systematic sampling, multistage sampling or cluster sampling, should be used to derive population estimates. According to EFSA guidance (EFSA 2014), the sample, where possible, should at least be stratified by age group and sex.

1. [Has the sampling been probabilistic, at individual level?](#)

2. Has the sampling been stratified by age classes, sex, and/or other important characteristics, such as region of residence, urban vs rural areas, seaside vs countryside areas, ethnic groups or household size, in case of household-based sampling?

A2.2 Sampling by waves

Populations are not constant, and it is important that sampling frames are updated with information on births, deaths, and any changes in classification to units in the population. This is particularly relevant for infants and toddlers, for whom it is recommended by EFSA (EFSA 2014) that a new sample should be drawn at the beginning of every season using an up-to-date sampling frame.

- Has the sampling been done by waves (in children)?

A3. Sample size

A3.1 Target sample size estimated by statistical assumptions

The sample size estimation should follow an adequate statistical procedure that considers the balance between statistical precision of results and feasibility to conduct the survey. EFSA guidance (EFSA 2009; (EFSA) 2014) recommends that each country should include, at least, 260 subjects with two complete interviews, 130 males and 130 females, in each of the six age classes to account for the statistics of interest (i.e., mean, median or variance), the level of precision needed, and the accuracy of high consumption level estimates, relevant from a risk assessment perspective. However, each country should estimate the sample size needed for the survey based on their goals and population characteristics.

1. Has an adequate statistical procedure for the sample size estimation been followed?

2. Has the minimum target sample size defined by EFSA been achieved for all age groups?

A4. Sampling error

A4.1 Estimation of standard error for key estimates

The standard error/confidence interval gives indication of how close the sample estimator (mean, proportion, etc.) is to the population value: the larger the standard error, the less precise the estimator. The coefficient of variation is a measurement of the relative variability of an estimate, sometimes called the relative standard error (RSE).

- To estimate the coefficient of variation (RSE) for key variables under study (e.g., key FoodEx2 food groups (level 2), energy intake, BMI...) as the:

$$RSE = \frac{\text{Standard error estimate}}{\text{estimate}} \times 100$$

The estimated RSE values for each variable will be compared against the benchmark.

A5. Representativeness

A5.1 Study sample with similar distribution as target population

The target population and the study sample should be compared in, at least, some characteristics. This comparison gives an indication of coverage error as the study sample is derived from a frame that may not perfectly enumerate the population.

- Has the study sample the same distribution of core characteristics with the target population?

A5.2 Weighting procedures

Weighting factors should be used to correct for the complex sampling design and non-participation. Sample weights reported to EFSA should be adapted according to the sample reported.

- If a sampling bias from the sampling frame is identified, have weighting factors been used to ensure the representativeness of the study population?

B. Recruitment

B1. Participation rate

B1.1 Participation rate, contact rate, cooperation rate

The response rate measures the proportion of individuals who responded to a survey. There are different types of rates, and these are indirect indicators on how significant the non-response bias is likely to be. When response is high, non-response bias is likely to be lower than when there are high rates of non-response.

The participation should be reported considering different rates:

- $Contact\ Rate\ (\%) = \frac{eligible}{eligible + unknown\ eligible\ individuals}$
- $Cooperation\ Rate\ (\%) = \frac{participants}{eligible\ individuals}$
- $Participation\ Rate\ (\%) = \frac{participants}{eligible + unknown\ eligible\ individuals}$

The cooperation rate is the requested response rate for EFSA, which is the rate mainly available in survey reports.

In the case where the sampling frame is a market research panel or similar type of list the calculation of the response rate should be adjusted for the original panel response rate.

- To compare survey response rates as relative differences to the pooled average/median rate of all surveys.

$$Relative\ difference_i = \frac{Survey\ response\ rate_i}{Average\ or\ median\ rate} - 1$$

B2. Comparison between participants and non-participants

B2.1 Absence of selection bias

Characteristics of non-participants may differ from participants, which leads to selection bias. Describing the differences in core variables between participants and non-participants indicates how relevant this bias is likely to be.

- Is there a comparison of core study variables (e.g. age, sex, diet-related variables, self-reported BMI) between participants and non-participants?

C. Training and supervision of interviewers

C1. Interviewers' selection criteria



C1.1 Background in Nutrition/Dietetics and/or interviewing experience

The use of selection criteria for interviewers is considered a process to reduce measurement error. Adequate training and knowledge on the dietary assessment methods to be used are mandatory criteria for interviews in dietary surveys. Ideally, interviewers should have a background on Nutrition or Dietetics, but other interviewers with experience in health data collection and/or training on the dietary assessment methods are acceptable, as long as they were properly trained for dietary assessment.

-To classify surveys according to the interviewers' background or previous experience:

- Dietetics/Nutrition background,
- Other interviewers experienced in health assessment
- Other situations

C2. Training of interviewers

C2.1 Standardized training procedures

Training of interviewers is a crucial procedure to improve the accuracy of the survey, to collect comprehensive data, and to reduce measurement error. The training should ensure that interviewers become familiar with survey protocols and tools (FAO 2018; Turrini et al. 2021).

-
1. Are there standard operation procedures (SOP) in place concerning training of interviewers?
 2. Has the training of interviewers been conducted according to these SOP?
 3. Has the training phase been conducted during the pilot survey?
 4. Has the training been covered in all aspects of the survey?

C2.2 Training duration and monitoring

The training duration should be sufficient for the interviewers to become familiar with the protocols and tools, to improve the communication and general interviewing techniques and to ensure the collection of comprehensive data.


Moreover, during data collection, the training should be conducted at a regular basis to tackle possible doubts and deviations detected, as well as to train new interviewers hired due to dropouts.

-
1. Has the training of interviewers been conducted at regular basis in order to assess possible deviations or problems raised during data collection?
 2. To compare the training duration of interviewers per country against the benchmark, i.e., the country with highest training duration.

C3. Supervision of interviewers

C3.1 Interviewer monitoring

During fieldwork, it is important to regularly monitor the performance of interviewers. The identification of interviewers with lower performance and the identification of main faults allows the survey coordination team to give fieldwork staff feedback and to adjust training programs. The interview coordinator should conduct regular checks of data (e.g. double data entry of the 24h-recall/food checks), should observe the frequency of specific answers to



look for potential errors, and should solve problems or doubts that may arise during the interviews communicating the results to all the fieldwork staff. These procedures help to improve interviewers' performance, impacting on data quality by reducing observer bias in data collection.

-
1. Has an interviewer coordinator/ coordination team been assigned in the survey?
 2. What was the percentage of interviews re-contacted via telephone to double-check whether interviews were actually undertaken?
 3. Has the coordination team checked the frequency of specific answers to identify possible errors, missing values etc.?
 4. Has the coordination team established a dynamic strategy to address issues emerging in the field (e.g. communication issues between the participant and the interviewer, issues regarding the meaning of questions, etc) in timely manner?
 5. Has the coordination team reported solutions for the problems risen in the fieldwork to all the staff?
-

C3.2 Observer bias

Observer bias is the systematic difference between a true value and the value actually observed due to observer (interviewer) variation (Porta 2014). Interviewer variance is the variability between the results obtained by different interviewers. Monitoring the estimated interviewer variance and some indicators such as the proportion of non-response by interviewer should be performed continuously throughout the fieldwork to minimize observer bias, improving the validity of the collected data.

-
1. Has the explained variance (%) for key variables (e.g., average time of interview, energy intake, BMI) been compared between interviewers regularly during the fieldwork?
 2. Has the proportion of item non-response, by interviewer, been continuously monitored?

C4. Pilot survey

C4.1 Pilot methods and tools

The protocols, questionnaires, and tools to be used in data collection should be tested before the actual survey to prevent and solve any shortcomings that may arise. Furthermore, some pilot study characteristics, such as sample size or setting may impact the quality of the actual survey data; thus, it is important to assure that the pilot study properly reflects the conditions of the survey.


-
1. Have methods been tested in practice (either pilot study or previous studies)?
 2. Has the setting and protocols used for pilot study been the same as the ones used in the actual survey?
-

D. Dietary data collection: Interview administration procedures

D1. Mode of administration of the interview

D1.1 At least one face-to-face interview

Face-to-face interviews are the standard and most used dietary assessment method in large surveys as it allows for better communication that may increase response rate due to



personal contact with the interviewer and more detailed probing of participant responses may be obtained. It is recommended by EFSA (EFSA 2014) that at least one face-to-face interview is conducted to ensure comprehensive reporting by the respondent.

- Has at least one interview been conducted face-to-face?

D1.2 Single mode of administration to all participants at each interview

Face-to-face interviews are the preferred mode of administration of the interview to ensure as much as possible a comprehensive reporting by the respondent. However, the 2014 EFSA (EFSA 2014) stated that telephone (or Internet-based e.g. Skype) interviews may be applied if a face-to-face encounter is not feasible. Nonetheless, different modes of administration of the interviews may affect the participants responses creating a bias in the consumption data, and the same methods should be applied to all participants in each interview moment. The decision or need of applying different modes of administration (e.g. due to Covid-19 pandemic) should imply an analysis to assess possible bias.

1. Has the same mode of administration been applied to all participants at each interview moment?

2. If different methods were applied, has a possible bias been identified and reported in the survey final reports?

D2. Interview setting and timing

D2.1 Interview setting

Participant's home is the recommended interview setting because of easy access to consumed foods and its packaging as well as specific measurements. However, other settings such as health centres, school canteens, or work may be acceptable. To provide flexibility for the participant to choose the interview location that best suits his/her needs reduces the participants' burden and is expected to increase the participation rates.

1. Has the setting of interview been decided according to participant's preference?

2. Have deviations from the pre-defined interview setting (including changes in the order through which questionnaires are administered) been monitored?

D2.2 Average time interval between interviews

Food consumption data should be collected on non-consecutive days. The time interval should be 8 to 15 days apart to ensure independence, which allows for a better estimate of intra-individual variability than data collected on consecutive days.

- To classify the surveys according to the average number of days between interviews: 8-15 days; >15 days.

D2.3 Distribution of dietary data by weekdays and season

Dietary data collection should be equally distributed through the seven weekdays and the four seasons, to capture the natural variability in food consumption according to these factors.

1. Is there a similar distribution (%) of 24-h recalls/diet records according to week and weekend days?

2. Is there a similar distribution (%) of 24-h recalls/diet records according to seasons?

D2.4 Duration of interview

It is important to keep the dietary interview time as short as possible, without decreasing the survey output quality, while keeping the subject burden reasonable. 30 minutes is considered the ideal duration of a 24-hour recall (EFSA 2014), and its application should not exceed 45 minutes.

Usually, surveys do not consider only the specific duration of a 24-hour recall but refer to the total interview time, including not only CAPI/CATI for dietary data collection, but also anthropometric evaluation, application of general questionnaires, etc. This refers to the first interview, where 60 minutes is considered as a reasonable amount of time for the total interview, and its application should not exceed 90 minutes. The second is expected to have the same duration as the 24h recall.

To compare the interview duration of each country with the benchmark: 30 minutes for 24-hour recall (as recommended by EFSA) and/or the average of total interview duration (24hR and other questionnaires).

E. Data processing: software tools and validation procedures

E1. Dietary assessment software: validation and automatic quality control procedures

E1.1 Use of a validated dietary assessment tool

The software used to collect dietary data along with the embedded databases must ensure the highest possible level of standardisation of the interviews within and between countries, and increase the likelihood that, if measurement errors exist, they will occur equally in all centres. Considering that 24-hour recalls and food diaries are prone to measurement errors, the software tools used across countries to collect data should be validated. The type and quality of databases embedded in the tool are addressed in other dimensions and subdimensions of this protocol.

- Has a previously validated tool for dietary assessment been used?

E1.2 Use of the multi-pass method (or similar)

Software should allow to enter the information on all foods and beverages consumed directly into the computer in electronic format and coded in accordance with common rules, e.g., according to the Multi-Pass Method (Moshfegh et al. 2008). This aims at reducing measurement error and increase comparability across countries.


- Does the software follow the multi-pass method or similar?

E1.3 Monitoring of the interview time

It is important to keep the interview time as short as possible, without decreasing the survey output quality, while keeping the subject burden reasonable. The software used must keep track of the interview time per participant. If the interviews are performed first on paper and only after inserted in the platform, the interviewer should monitor and register the duration of the interview for quality control.

- Does the software allow the tracking of the interview time per subject? Or, if the interview is first conducted on paper, has the interviewer monitored and registered its duration?

E1.4 Processing errors identification and description of quality control procedures



Within data processing, errors may occur. These processing errors could be present in data capture, coding, editing, and weighting. They usually cannot be calculated, but the main sources of errors can be identified and clearly described. It is important to describe the mechanisms in place to minimise processing errors by ensuring accurate data capture, processing, and computer-assisted procedures.

It is relevant to identify possible systematic errors within the standardised operating procedures, such as unacceptable values, empty food consumption occasions, missing quantities, easily forgotten foods, etc. and describe them as possible threats to your data.

The software for dietary data collection should include procedures for automatic checking of these issues.

-
1. Is there automatic checking for empty food consumption occasions, minimum/maximum accepted quantities per food type, and/or missing quantities?
 2. Is there automatic checking for easily forgotten foods (use of probing questions)?
 3. Is there a calculation of energy and macronutrient intake outliers at the end of interviews?

E2. Quantification of dietary consumption

E2.1 Use of a validated food picture book

The process of quantifying foods, usually using portion-size measurement aids is subject to inaccuracies that ultimately may lead to misclassification of subjects according to the amount of food consumed. Thus, it is essential to use validated tools for quantification of foods and recipes.

The picture books should be validated at country-level and should include foods commonly available on the market at the time of the survey, and portion sizes representative of the dietary habits of the survey country. Furthermore, it should contain pictures of household measurement tools and standard units.

-
1. Has a validated food picture book (or portion size measurement aids) been used?
 2. If the picture book has been updated, has it been validated again (with the new added pictures)?

E2.2 Minimum number of pictures per photo series in picture book

To avoid regression towards the mean, an even number of pictures should be included, ideally 6 (or a minimum of 4) colour pictures should be used. Each picture in the same photo series should represent only one portion size and should have the same shape and be taken at the same viewing angle as others.

-
- If available, does the picture book include a minimum of 4 colour pictures per series?

E2.3 Total number of picture series in the picture book

The 2014 EFSA guidance (EFSA 2014) recommended 45 as a reasonable number of picture series to be included in the picture book; however, more picture series could be included at the country level, if necessary.

-
- To classify the surveys according to the total number of picture series:
 - low: < 40
 - reasonable: 40-50
 - high: >50

E2.4 Availability of food standard units as a quantification method

For specific foods or occasions, it is recommended that other quantification tools are made available, such as standard units: units for fruits, rolls, cups of yoghurts, cans/packages of beverages, etc.

- Are there food standard units available?

E2.5 Availability of default quantities as a quantification method

For specific foods or occasions, it is recommended that other quantification tools are made available, such as default quantities, when the participant cannot provide any accurate information regarding the quantity consumed.

- Are there default quantities available?

E3. Food propensity questionnaire (FPQ) or Food frequency questionnaire (FFQ)

E3.1 Application of an FPQ or FFQ

An FPQ or FFQ should be included in the survey to complement the repeated 24-hour recalls or food diaries with important covariate information for estimating usual intake of food groups (Subar et al. 2006; EFSA 2009). The inclusion of portion sizes is not recommended, as they can be obtained through the multiple 24-hour recalls/food diaries.

4. Has a FPQ or a FFQ been applied?

5. If the FPQ/FFQ was self-administered, has it been subsequently checked for completeness?

E3.2 Adequacy of the FPQ/FFQ

The FPQ should include a short-selected list of age-appropriate foods that are consumed episodically or in specific seasons and that are relevant from a risk-assessment perspective (i.e., likely to contain higher concentrations of relevant food-borne hazards). It should also include questions about dietary supplements.

1. Has the FPQ/FFQ included less frequently eaten foods?

2. Has the FPQ/FFQ included food that are likely to contain higher concentrations of relevant food-borne hazards?

3. Has information on food/dietary supplements been recorded in the FPQ/FFQ?

4. Has the FPQ/FFQ been designed to capture the seasonal variation of certain foods?

F. Non-dietary data collection

F1. Anthropometric data

F1.1 Measured (vs. self-reported) weight and height

Anthropometric data collection is important in the context of a National Dietary Survey. For example, recording body weight of participants enables the calculation of dietary exposure per kg/bodyweight at individual level for direct comparison with toxicological reference units.

In adults, the collection of body weight and height data, can be performed through self-report or measured during the interview.

For children, it is recommended that body weight and height are measured during the interview rather than (or in addition to) self-reported.

1. Have objective measurements of weight and height been collected in adults?
2. Have objective measurements of weight and height been collected in children?
3. In the case of pregnant women, has the self-reported weight before pregnancy been collected?

F1.2 Standardized measurements of weight and height

According to the 2009 EFSA guidance (EFSA 2009), actual measurements of body weight and height, through standardized procedures enable more accurate estimates, reducing the bias related to differences between countries in self-reporting. The standardization ensures that anthropometric measurements are taken under the same conditions for all subjects, considering the interview setting, and possible deviations from the protocol should be registered and reported. Moreover, the protocol should reflect the frequency of equipment verification and calibration and interviewers must have adequate and regular training.

1. Have measures of weight and height been collected by standardized procedures in adults?
2. Have measures of weight and height been collected by standardized procedures in children?
3. Have possible deviations from the protocol for measuring weight and height been reported?
4. Have the interviewers been trained regularly through repeated anthropometric measurements?
5. Has the equipment used for anthropometric measurements been regularly checked (e.g. proper calibration of stadiometer or body scale on a regular basis)?

F1.3 Repeated standardized measurements available for each parameter (weight, height)

To account for measurement error, more than one standardized measurement of anthropometric variables should be available.

- If measured, has a second measurement of weight and height been obtained?

F1.4 Digit preference in anthropometric measurements

Digit preference is the observation that the final number in a measurement occurs with a greater frequency that is expected by chance. This can occur because of rounding, the practice of increasing or decreasing the value in a measurement to the nearest whole or half unit, or because data are made up.

- To estimate by which level by which digit preference exists in each survey, for weight and height according to the equation (WHO & UNICEF 2019):

$$\sum_{i=0}^9 \frac{|\text{observed } prev_i - 0.1|}{2}$$

The estimated preference of digits will be compared against the benchmark.

F2. Physical activity

F2.1 Physical activity assessment

The collection of physical activity level data influences the survey data quality as it is used for the estimation of misreporting to rank individuals into low, medium, and high physical activity level categories. EFSA recommends the use of a validated PA questionnaire as a feasible method for assessing habitual physical activity in large populations, despite its limitations (i.e., questionnaires tend to over-report activity levels). Ideally, the physical activity questionnaire should be validated by comparison with an objective method, such as an accelerometer (Wareham et al. 2002; Ferrari et al. 2007).

- Has physical activity been measured by a validated questionnaire (e.g. IPAQ) in adults?
- Has physical activity been collected in children using a reporting method (e.g., diaries)?

F2.2 Collection of accurate physical activity measurements

Heart rate monitoring or motion sensors (accelerometers) are more accurate methods to assess physical activity levels. Despite being too demanding in terms of logistical and financial resources to be included in a pan-European dietary survey, these can be used in a smaller sub-sample to validate the self-reporting physical activity levels (diaries and questionnaires) that tend to be overestimated.

- Have more accurate physical activity measurements been collected (e.g. heart rate monitoring, motion sensors (accelerometers))?

G. Data handling and cleaning

G1. Completeness of food composition database

G1.1 Number of food items/recipes in the food composition database

The food list used for data collection should be comprehensive, i.e., it should include most of the foods consumed in the country and be open-ended, allowing for regular updates for including new foods and dishes, reported during the data entry phase.


- To compare the number of food items and recipes reported in the country-specific food lists against the benchmark. Because of differences in the reporting strategy between countries (some countries had very extensive food lists but used few facets in the consumption eating occasions, others had a more knit food list but were very descriptive), the benchmark should be set after extracting the distinct food items based on the complex FoodEx2 term from the consumption file.

G1.2 Availability of a food supplements database

Detailed information on the consumption of dietary supplements should be collected through the 24h-recall/ food diary, as it is useful for nutrient and chemical exposure assessments. Thus, if the food list available should include dietary supplements.

1. Are there food supplements available in the food list to be reported during the 24-hour recall/food diaries.
2. If available, are food supplements quantified in grams?

G1.3 Availability of a recipe composition database



It is recommended to disaggregate the composite dishes into ingredients whenever possible. For that purpose, an open-ended recipe database, subject to regular updates, should be available in the software used for data collection.

1. Is there a recipe composition database available?

2. Has the recipe database been updated during the survey fieldwork?

G1.4 Food items with only FoodEx2 base term

Foodex2 codes with only the base term may indicate low specificity in the food items description.

- To identify the proportion of food items with only the base term at country level and compare against the benchmark.

G1.5 Availability of energy and nutrients (#6) per 100 grams of food

The energy and other nutrients per 100g of food is a non-mandatory variable in the Foods dataset, but relevant for nutritional intake analyses, at country and European level.

- To compare against the benchmark the percentage of missing values in nutrients reported in the foods dataset per country.

G2. Completeness of food consumption database

G2.1 Disaggregation of recipes into ingredients and identification of mixed dishes

Composite dishes should be disaggregated into their ingredients, according to the EU Menu methodology. It is also important to be able to identify foods that were eaten together as part of a recipe

1. To estimate the percentage of composite dishes in FoodEx2 codes and compare it against the benchmark.

2. Have a recipe code been reported for composite dishes, allowing to identify foods that were consumed together as part of a recipe?

G2.2 Missing values in non-mandatory variables

Variables such as brand, raw and cooked quantities, recipe codes and amounts and eating location are relevant descriptors of food items. When reported, it may have a low or high percentage of missing values.

- To estimate the percentage of missing values in the identified variables from the consumption dataset and compare it against the benchmark.


G2.3 Total number of facets

Facets are relevant descriptors to a better understanding of the food items reported. According to the FoodEx2 system, there are 29 facets available for reporting.

- Out of the possible 29 facets, compare the total of facets used per country against the benchmark.

G2.4 Minimum recommended facets

The following facets, from FoodEx2 classification system, are identified as the minimum recommended facets available at survey level for a proper foods description, when



applicable: source, part-consumed-analysed, processing, qualitative info, fortification, sweetening-agent, and packaging-material facets.

- Has the recommended minimum facets using the FoodEx2 classification system been reported?

G2.5 Prevalence of foods classified as level 4 in the FoodEx2 hierarchy

The classification with FoodEx2 should be as specific as possible. Thus, a high prevalence of food items classified as level 4 in the FoodEx2 hierarchy is desirable, while lower hierarchy levels may represent low specificity which may lead to inconclusive estimates.

- To estimate the proportion of foods classified as level 4 or above and level 3, if accompanied by a facet that specifies the food better, in the FoodEx2 hierarchy per country and compare it against the benchmark.

G3. Completeness of subjects' database

G3.1 Missing values in non-mandatory variables

This indicator will evaluate the occurrence of missing values in variables such as sample weights or sociodemographic and anthropometric variables (education level, geographical region, household size, labour status, professional category, special conditions/diet).

- To estimate proportion of missing values in non-mandatory variables (e.g. sample weights, sociodemographic and anthropometric) and compare it against the benchmark.

H. Data analyses

H1. Dietary intake validity

H1.1 Energy outliers

24-hour recalls or food diaries with a very low daily energy intake are likely to be incomplete or are not representative of an usual food consumption day.


- For countries that reported Energy in the Foods dataset, identify the percentage of interviews with total energy intake lower or higher than 3 SD/interquartile range (excluding special days) or lower than 500 and higher than 3500 (Willett 2012). We will consider as benchmarking the country with lower percentage and we will compare the countries against this value.

H1.2 Food groups outliers

The amount consumed of different foods can present outlier values representing inconsistencies and errors during data collection that may have not been captured by the automatic control checks in the software.

- To identify the percentage of interviews with food groups consumed amounts lower or higher than 3 SD/interquartile range (excluding special days). The percentages will be compared to the benchmark. For standardization, we will consider the FoodEx2 hierarchy level 2 food groups.

H1.3 Number of main meals per day



24-hours recalls or food diaries without at least one of the three main meals (empty values) are likely to be incomplete or are not representative of an usual food consumption day.

- To identify the percentage of incomplete 24-hour recalls or food diaries, i.e., 24-hour recalls or food diaries with less than two main meals (breakfast, lunch, dinner), excluding special days.

H1.4 Minimum acceptable number of food items

24-hour recalls or food diaries with severe negative outliers for the total number of food items reported are likely to be incomplete or are not representative of an usual food consumption day. Moreover, severe positive outliers may indicate a report bias.

- To identify interviews with total number of food items reported lower or higher than 3 SD/interquartile range, excluding special days, within and between surveys.

H1.5 Digit preference in food amounts

Digit preference is the observation that the final number in a measurement occurs with a greater frequency that is expected by chance. This can occur because of rounding, the practice of rounding (increasing or decreasing) the value in a measurement to the nearest whole or half unit, or because data are made up.

- To estimate by which level by which digit preference exists in each survey, for food amounts, according to the equation (WHO & UNICEF 2019):

$$\sum_{i=0}^9 \frac{|\text{observed prev}_i - 0.1|}{2}$$

The estimated preference of digits will be compared against the benchmark.

H2. Data distribution and variability

H2.1 Proportion of total variance explained by differences between individuals

The intraclass correlation coefficient (ICC) measures the proportion of total variance explained by differences between individuals. Thus, ICC may inform on the adequacy of the number of repeated 24-hour recalls or food diaries to accurately estimate the intake of food groups and nutrients at individual level.

- To estimate the between and within-variability in core dietary variables (e.g. food groups, nutrients) using the average ICC and to compare against the benchmark.

H3. Indicators to validate dietary data


H3.1 Availability of data, such as biomarkers, to validate dietary intake

Traditional dietary assessment methods (food diaries, 24-hour recall, FFQ) may be prone to systematic bias affected by factors such as sex, age, socioeconomic status, and social desirability. Biomarkers are not prone to these biases, thus representing a possible accurate way to validate the dietary intake.

- Are there data available to validate dietary intake even in sub-samples (e.g., biomarkers)?

H4. Energy misreporting (under and over-reporting)

H4.1 Calculation of misreporting of energy intake through recommended methods



Misreporting, including both under and over-reporting, is one of the main sources of error in dietary assessment. The assessment of the prevalence of misreporting (i.e., under- and over-reporting of dietary energy intake) should be performed using the Goldberg cut-off method (Goldberg et al. 1991) updated by Black in 2000 (Black 2000), considering the physical activity levels (low, moderate, or vigorous) of the subjects.

1. Has misreporting of energy intake been calculated using the Goldberg cut-off method, taking into account physical activity?

2. To compare the estimated prevalence of energy misreporting (under and over-reporting) per country against the benchmark. (if physical activity levels and energy intake are available)

I. Results reporting

I1. External validity of results

I1.1 Weighted results to ensure the representativeness

If a sampling bias from the sampling frame is identified, weighting factors could be used to ensure the results are representative of the study population.

- Have weighting factors been used to ensure the representativeness of the study population (if identified this need in E21)?

I2. Usual nutritional intake

I2.1 Adjustment of nutritional intake for the intra-individual variability

Usual intake can be estimated through independent (non-consecutive) repeated measures of 24-hour recalls or food diaries. The statistical models aim to reduce the intra-individual variability, adjusting the observed distribution of food/nutrient intake.

- Have nutritional intake estimates been adjusted for the intra-individual variability of intake?

I2.2 Usual intake estimated using the food propensity/frequency questionnaire

The FPQ may provide important covariate information to supplement multiple recalls for estimating usual intake of food groups (Subar et al. 2006; EFSA 2009). The FPQ is especially relevant for foods that were consumed by less than 50 % of the population, that contained high levels of constituents (e.g. nutrients) for which the lower percentiles of intake were relevant (van Klaveren et al. 2012).

- If usual intake was estimated, has the food propensity/frequency questionnaire been used?

I3. Energy misreporting

I3.1 Sensitivity analysis excluding misreporters of energy intake

Sensitivity analysis excluding misreporters allows to obtain insight on the bias in estimating dietary exposure due to misreporting.

- Has a sensitivity analysis been conducted excluding misreporters of energy intake (but not excluding them from dataset)?

2.2 Factors associated with data quality

An exploratory analysis will be conducted to find the factors associated with data quality. This analysis will be performed at the survey level, comprising about 30 surveys. The identification of these associations might be an important step for improving and updating the recommendations for the next round of European national surveys under the EU Menu methodology.

Among all proposed quality indicators, those with large or moderate effect for distinguishing surveys will be identified, discarding those with small effect. For quality indicators at the survey level, the coefficient of variation will be estimated. In case the indicator is available at participant or interview level, these effects will be quantified using median odds ratio (OR), through generalized linear logistic regression model with random effects by survey, or ICC through mixed-linear regression model. An example would be a mixed linear regression model using as outcome the number of food items per interview and with random effects by survey. From this model, the variance of the random effect will be extracted to calculate the ICC, to assess the effect of the survey in the number of food items (outcome used as example). A second model can be done adjusted for other variables (e.g. age or sex) as fixed effects.

To diminish the multiple comparisons, the quality indicators with a moderate or large effect will be summarized into 4 composite indicators: *Sampling and recruitment*, *Dietary data collection and processing*, *Non-dietary data collection*, and *Data analyses and results reporting*, as showed in **Error! Reference source not found..** For this, an expert-judgment approach will be preferred over a data-driven approach, due to the small number of observations for statistical purposes. To create each composite indicator, an expert panel will weight each individual quality indicator within the dimensions included.

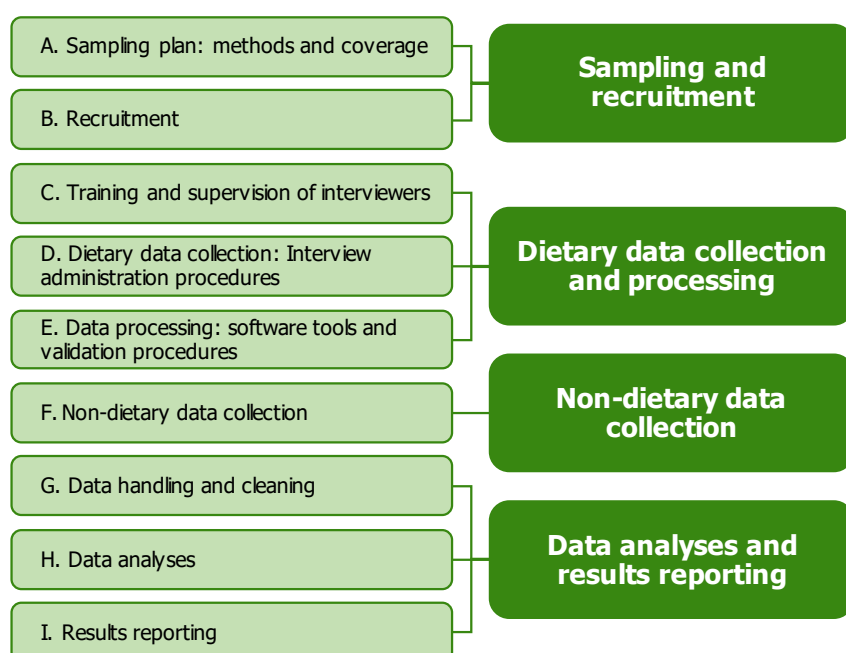



Figure 2. Organization of survey quality indicators dimensions into composite quality indicators



To identify associations between composite indicators, Pearson or Spearman correlation coefficients will be estimated between each pair. If a moderate correlation is found, the associations between the individual quality indicators within each composite quality indicator will be analyzed.

Lastly, an ecological study will be carried out to assess if country characteristics, such as average educational level, gross domestic product (GDP), and unemployment rate, attenuate the associations between quality indicators. These analyses will be conducted using linear or binary/multinomial logistic regression models. The information on the country-specific variables mentioned will be retrieved from OECD data (<https://data.oecd.org>), Eurostat (<https://ec.europa.eu/eurostat/data/database>), and from country-specific national statistics.

3 Deliverables work package 2 Evaluation of current data

This protocol is the deliverable assigned to WP2 (D3). The results obtained in WP2 analyses will be presented in a table with the overall summary of each quality indicator, comprising the benchmark and the mean/median value of the relative performance for the remaining surveys. If the quality indicator assessment result in a proportion, that value will be described in the table. The results of WP2 will also include a description of the associated factors analysis.

These outputs will be used to prepare the draft final report of this project (D4), including advice for updates in the EU Menu guidance.

3.1 Consultation with EFSA

This protocol was discussed with EFSA during the second meeting in June 2022 and feedback was considered to prepare the final protocol that includes reformulations and updates.

4 Planning

The planning of the activities, milestones (M) and deliverables (D) are presented in Table 2. The results of milestone 4 will be discussed with EFSA to focus the work of M5. The draft report was discussed with EFSA in the second meeting that was held in 14th June 2022.

	Month (number and date)																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Jan 17 2022	Feb 16 2022	Mar 16 2022	Apr 16 2022	May 16 2022	Jun 16 2022	Jul 16 2022	Aug 16 2022	Sep 16 2022	Oct 16 2022	Nov 16 2022	Dec 16 2022	Jan 16 2023	Feb 16 2023	Mar 16 2023	Apr 16 2023	May 16 2023	Jun 16 2023
M2																		
Define the draft list of quality indicators		M2																
D3																		
Prepare protocol							D3											
M7																		
Map surveys according to quality indicators											M7							
M8																		
Conduct associated factors analyses												M8						
M9																		
Prepare draft report																		M9

All the tasks from WP2 are leaded by the U. Porto team.



References

- (EFSA) EFSA. 2014. Guidance on the EU Menu methodology. EFSA J [Internet]. 12(12):3944. <https://doi.org/10.2903/j.efsa.2014.3944>
- Black A. 2000. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. Int J Obes [Internet]. 24(9):1119–1130. <http://www.nature.com/articles/0801376>
- de Boer EJ, Slimani N, van 'T Veer P, Boeing H, Feinberg M, Leclercq C, Trolle E, Amiano P, Andersen LF, Freisling H, et al. 2011. Rationale and methods of the European Food Consumption Validation (EFCHOVAL) Project. Eur J Clin Nutr [Internet]. 65(S1):S1–S4. <https://www.nature.com/articles/ejcn201182>
- EFSA. 2009. General principles for the collection of national food consumption data in the view of a pan-European dietary survey. EFSA J [Internet]. 7(12):1435. <http://doi.wiley.com/10.2903/j.efsa.2009.1435>
- EFSA. 2014. Guidance on the EU Menu methodology. EFSA J. 12(12):1–80.
- FAO. 2018. Dietary Assessment: A resource guide to method selection and application in low resource settings. Rome.
- Ferrari P, Friedenreich C, Matthews CE. 2007. The Role of Measurement Error in Estimating Levels of Physical Activity. Am J Epidemiol [Internet]. 166(7):832–840. <https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwm148>
- Gibson RS, Charrondiere UR, Bell W. 2017. Measurement Errors in Dietary Assessment Using Self-Reported 24-Hour Recalls in Low-Income Countries and Strategies for Their Prevention. Adv Nutr An Int Rev J [Internet]. 8(6):980–991. <http://advances.nutrition.org/lookup/doi/10.3945/an.117.016980>
- Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, Prentice AM. 1991. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. Eur J Clin Nutr. 45(12):569–581.
- van Klaveren JD, Goedhart PW, Wapperom D, van der Voet H. 2012. A European tool for usual intake distribution estimation in relation to data collection by EFSA. EFSA Support Publ [Internet]. 9(6). <http://doi.wiley.com/10.2903/sp.efsa.2012.EN-300>
- Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler W V, Paul DR, Sebastian RS, Kuczyński KJ, Ingwersen LA, et al. 2008. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am J Clin Nutr [Internet]. 88(2):324–332. <https://academic.oup.com/ajcn/article/88/2/324/4650197>
- Porta M. 2014. A dictionary of epidemiology. [place unknown]: Oxford university press.
- Rutishauser IH. 2005. Dietary intake measurements. Public Health Nutr [Internet]. 8(7a):1100–1107. https://www.cambridge.org/core/product/identifier/S1368980005001369/type/journal_article
- Subar AF, Dodd KW, Guenther PM, Kipnis V, Midthune D, McDowell M, Tooze JA, Freedman LS, Krebs-Smith SM. 2006. The Food Propensity Questionnaire: Concept, Development, and Validation for Use as a Covariate in a Model to Estimate Usual Food Intake. J Am Diet Assoc [Internet]. 106(10):1556–1563. <https://linkinghub.elsevier.com/retrieve/pii/S0002822306016890>
- Turrini A, Catasta G, Censi L, Comendador Azcarraga FJ, D'Addezio L, Ferrari M, Le Donne C, Martone D, Mistura L, Pettinelli A, et al. 2021. A Dietary Assessment Training Course Path: The Italian IV SCAI Study



on Children Food Consumption. *Front Public Heal.* 9(March):1–10.

Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. 2002. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. *Int J Epidemiol* [Internet]. 31(1):168–174. <https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/31.1.168>

WHO, UNICEF. 2019. Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old. Geneva, Switzerland.

Willett W. 2012. *Nutritional epidemiology*. [place unknown]: Oxford university press.

Appendix A – Surveys to be included in the data quality evaluation

Country	Survey code	Survey name
Austria	AT3	EU Menu Austria: Food consumption data for Austrian adults
	AT5	EU Menu Austria: Food consumption data for Austrian adolescents
Belgium	BE1	Belgian national food consumption survey in children, adolescents and adults
Bosnia and Herzegovina	BA1	The study of Bosnia-Herzegovinian Dietary Survey of Adolescents, Adults and Pregnant Women (B&H MENU)
Croatia	HR1	Croatian food consumption survey on adults
Cyprus	CY1	National dietary survey of the children of Cyprus
	CY2	National dietary survey of the adult population of Cyprus
Estonia	EE1	National Dietary Survey among 11-74 years old individuals in Estonia
	EE2	National Dietary Survey among children up to ten years old and breastfeeding mothers in Estonia
Finland	FI1	The Finnish National Dietary Survey in Adults and
France	FR3	The French national dietary survey (INCA3, 2014-2015)
Greece	GR1	The EFSA-funded collection of dietary and related data in the general population aged 10-74 years in Greece
Hungary	HU1	Hungarian national food consumption survey
Italy	IT2	Italian national dietary survey on children population from three months up to nine years old – IV SCAI CHILD
Latvia	LV3	Latvian National Dietary Survey on the general population
Montenegro	ME1	National Dietary Survey on Adolescents, Adults, Elderly and Pregnant Women in Montenegro
Netherlands	NL4	Dutch National Food Consumption survey 2012-2016
North Macedonia	MK1	National dietary survey on the children population in the Republic of North Macedonia
Portugal	PT1	National Food, Nutrition and Physical Activity Survey of the Portuguese general population (IAN-AF) 2015-2016
	PT2	National Food, Nutrition and Physical Activity Survey of the Portuguese general population (IAN-AF) 2015-2016
Romania	RO3	Romanian national food consumption survey for adolescents, adults and elderly
	RO4	Ad-hoc consumption survey for Romanian pregnant women
	RO5	Ad-hoc consumption survey for Romanian vegetarian adults
Serbia	RS1	National Food Consumption Survey among children from 1 to 9 years old in Serbia
	RS2	National Food Consumption Survey among adults from 10 to 74 years old in Serbia



Slovenia	SI2	Slovenian national food consumption survey on children (infants and toddlers)
	SI3	Slovenian national food consumption survey in adolescents, adults and elderly
Spain	ES2	Spanish National dietary survey in adults, elderly and pregnant women
	ES3	Spanish National dietary survey on children and adolescents