External Scientific Report

APPROVED: December 2023
doi: 10.5281/zenodo.10100898

Quality evaluation of the food consumption and related data collected under the EU Menu Project

Catarina Carvalho[1,2], Milton Severo[1], Daniela Correia[1,3], Andreia Oliveira[1,3], Caroline van Rossum[4], Marga Ocké [4], 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 ERA EU Menu project aims to give evidence to update the EU menu guidance by evaluating methods and tools used in National Dietary Surveys and assessing data quality collected under the EU menu framework. This report focuses on the activities of workpakage 2: 'Evaluation of current data'.

Data from thirty-one surveys conducted under EU Menu guidelines was used, considering the surveys' datasets and their methodological reports. Surveys were mapped based on a list of 96 quality indicators, which cover nine dimensions Sampling, Recruitment, Interviewer Training, Data Collection, Software Tools, Non-dietary Data Collection, Data Completeness, Data Analysis, and Reporting. An exploratory analysis was also conducted to examine how the survey dimensions impacted data quality in terms of missing data, food and recipe description, data reliability, the presence of outliers, and energy misreporting.

Overall, most surveys adhered well to EFSA 2014 guidelines, achieving a commendable level of harmonization and compliance. However, some challenges persist, namely, the inaccurate use of definitions, missing information in the final reports, differences in the seasons coverage during data collection, and discrepancies in database sharing with EFSA. The dimensions that mostly impacted data quality were the sampling, training & supervision of interviewers and interview administration procedures (seasonality, interview duration).

This report describes in detail the analysis and results of quality indicators to inform recommendations for future survey rounds in the ERA EU Menu project's final deliverable.

Key words: quality indicators, measurement errors, national dietary surveys, dietary assessment

Correspondence: IDATA@efsa.europa.eu

Acknowledgements: We highly appreciate all input received from EFSA colleagues (Anastasia Livaniou, Ashraf Khosravi and Sofia Ioannidou) and collaborators under the tasking grant GP/EFSA/DATA/2021/01 Lot 3 (Androniki Naska, Krystalia Niforou, Nikolaos Koffas and Elissavet Maria Valanou).

Suggested citation: Carvalho C., Severo M., Correia D., Oliveira A., van Rossum C., Ocké M., Torres D., Lopes C., 2023. Quality evaluation of the food consumption and related data collected under the EU Menu Project. Zenodo https://doi.org/10.5281/zenodo.10100898

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 quality of data 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 report refers to WP2 'Evaluation of current data' activities, in which the project team aimed 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 and associated factors of the EU Menu surveys; and finally (3) evaluate the quality and harmonization of the surveys in the EU Menu framework and the main associated factors (survey dimensions and countries' sociodemographic factors).

Thirty-one surveys conducted under the EU Menu guidelines were assessed through the datasets shared with EFSA and the methodological reports to map the surveys according to the 96 quality indicators proposed in the protocol. The quality indicators focus on nine survey dimensions: Sampling, Recruitment, Training and Supervision of Interviewers, Data collection procedures, Software tools and validation, Non-dietary data collection, Data completeness, Data analyses and Reporting. Moreover, exploratory analyses were done to investigate inter-correlations and dependencies among quality indicators within different dimensions.

Our findings indicate that, in general, the surveys adhered well to the EFSA 2014 guidelines, resulting in a commendable level of harmonization and compliance with the recommendations. Nevertheless, several inconsistencies within and between surveys were identified throughout this study that deserve to be highlighted for future improvement. For instance, challenges were noted in the dimension related to recruitment, where definitions of participation rate, contact rate, and cooperation rate were often inaccurately reported. Similarly, the dimension assessing food and recipe description faced challenges due to the lack of consistency and harmonization on reporting FoodEx2 facets. Other issues included the omission of crucial information in reports, differences in seasonality coverage in the dietary data collection, and variations in data reporting to EFSA.

Generally, an enhanced sampling plan was linked to greater data reliability, a reduced occurrence of outliers, and improved collection of non-dietary data. Thus, ad-hoc surveys exhibited lower quality in these aspects. Furthermore, improved interviewer training and the application of appropriate interview procedures were associated with more accurate food and recipe descriptions and a reduced rate of missing data. The country's level of education and demographic factors also played a significant role in data quality. Highly educated populations are likely to better report and describe the foods consumed despite of investing less in the training and supervision of interviewers. Countries with higher proportion of the population living in rural areas presented lower data reliability.

This report highlights the relevance of defining and measuring data quality indicators and emphasizes their significance as valuable lessons that will inform recommendations for future survey rounds in the ERA EU Menu final deliverable.

Table of contents

1 Introduction

1.1 Background – The ERA EU Menu Project

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) (de Boer et al., 2011; Rutishauser, 2005). 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, 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 of the ERA EU Menu project 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 identifying its associated factors;

3. To summarise the quality of the surveys in the EU Menu framework and its main associated factors.

See also Figure 1 for an overview of the activities in this work package.

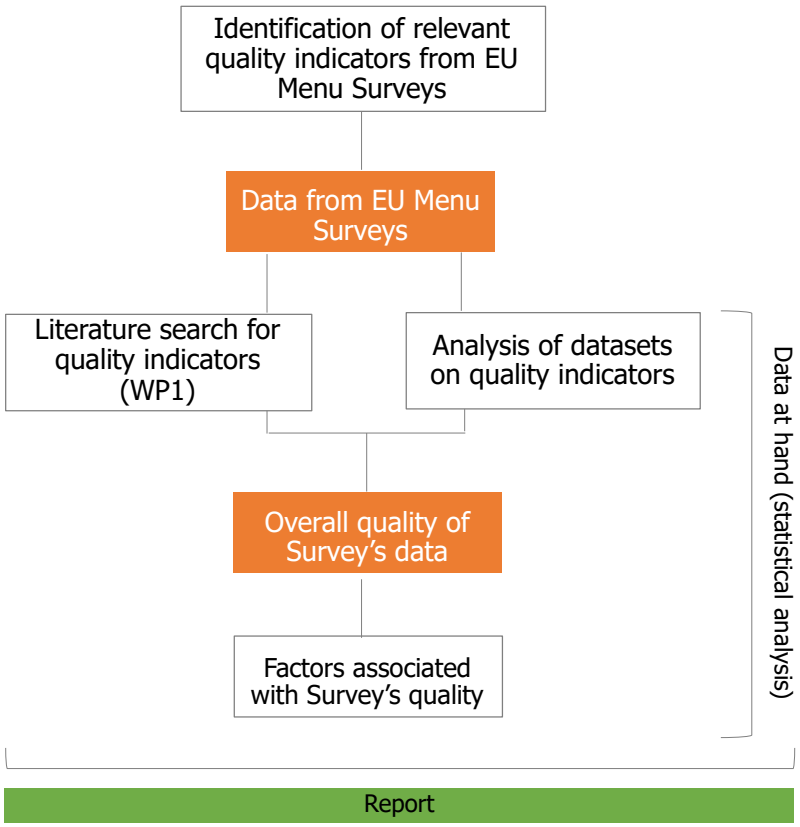

Figure 1. Graphical representation of work package 2 'Evaluation of current data'

A detailed protocol (Annex A) was developed supporting the quality indicators identification, the procedure for its assessment and important methodological decisions for guiding the activities of the WP2, answering to the first and second objectives indicated. The next chapters of this report describe the methodology used, the results on the data quality indicators of the EU Menu surveys along with its interpretation and a discussion focusing on the strengths and limitations of the methods used, referring to the third objective of WP2 . The results from this report will feed the final project deliverable, a report with advice for recommendations for the next round of national dietary surveys in Europe.

2 Data and Methodologies

2.1 EU Menu Surveys

Thirty-one **datasets of EU Menu surveys provided by EFSA** were considered for the analyses. **Seven** included the **total population** (children and adults together), **seven** included only **children**, **one** included only **adolescents**, **eleven** only **adults and adolescents,** and **five** were **ad-hoc** surveys (three among pregnant women and two among vegetarians). Furthermore, one of the children's surveys also included an oversampled ad-hoc group of breastfeeding mothers, and five of the adult's surveys included oversampled ad-hoc groups of pregnant women.

Guidance on EU-MENU methodology for conducting national dietary surveys is different according to age groups, children <10 years old, or adolescents and adults ≥10 years old. Thus, the assessment of the quality indicators was stratified into these two age groups plus an additional category, the ad-hoc surveys. Accordingly, the datasets that included both age groups and ad-hoc samples were split resulting in a total of 43 survey target groups, as presented in Table 1. In Table 1, and the remaining figures from this report, each survey target group is represented by a three-digit code that identifies the country and the age group. Thus, this identifier comprised two components: the ISO-alpha-2 code of the country and a number representing the age group of the participants. For participants below 10 years old, the number 1 was used, while for those aged 10 and above, the number 2 was used. Furthermore, ad-hoc surveys received the numbers 3 or 4, depending on the number of ad-hoc surveys conducted in the same country.

Table 1. EU Menu surveys considered and respective organization according to survey target groups: < 10 years old, ≥10 years old and ad-hoc.

EU Menu Country	Survey Name – EU Menu datasets received	Survey Target Group	Code
Austria	EU Menu Austria: Food consumption data for Austrian adolescents	≥ 10 years	AT2
	EU Menu Austria: Food consumption data for Austrian adults		
	EU Menu Austria: Food consumption data for Austrian pregnant women	Ad-hoc (Pregnant)	AT3
Belgium	Belgian national food consumption survey in children, adolescents and adults	<10 years	BE1
		≥10 years	BE2
Bosnia & Herzegovina	Bosnia-Herzegovinian Dietary Survey of adolescents, adults and pregnant women	≥10 years	BA2
		Ad-hoc (Pregnant)	BA3
Croatia	Croatian national food consumption survey on children from 3 months to 9 years of age	<10 years	HR1
Cyprus	National dietary survey of the children of Cyprus	<10 years	CY1
	National dietary survey of the adult population of Cyprus	≥10 years	CY2
		Ad-hoc (Pregnant)	CY3
Estonia	National Dietary Survey among children up to ten years old and breastfeeding mothers in Estonia	<10 years	EE1
		Ad-hoc (Breastfeeding)	EE3

EU Menu Country	Survey Name – EU Menu datasets received	Survey Target Group	Code
	National Dietary Survey among 11-74 years old individuals in Estonia	≥10 years	EE2
Finland	The Finnish National Dietary Survey in Adults and Elderly (FinDiet 2017)	≥10 years	FI2
France	The French national dietary survey (INCA3, 2014-2015)	<10 years	FR1
		≥10 years	FR2
Greece	The EFSA-funded collection of dietary and related data in the general population aged 10-74 years in Greece	≥10 years	GR2
Hungary	Hungarian national food consumption survey	<10 years	HU1
		≥10 years	HU2
Italy	Italian national dietary survey on children population from three months up to nine years old	<10 years	IT1
	Italian national dietary survey on adult population from 10 up to 74 years old	≥10 years	IT2
Latvia	Latvian National Dietary survey	<10 years	LV1
		≥10 years	LV2
Montenegro	Montenegrin National Dietary Survey on the general population	≥10 years	ME2
		Ad-hoc (Pregnant)	ME3
Republic of North Macedonia	National dietary survey on the children population in the Republic of North Macedonia	<10 years	MK1
Netherlands	Dutch National Food Consumption Survey 2012-2016 (DNFCS)	<10 years	NL1
		≥10 years	NL2
Portugal	National Food, Nutrition and Physical Activity Survey of the Portuguese general population (IAN-AF 2015-2016)	<10 years	PT1
		≥10 years	PT2
	National Food, Nutrition and Physical Activity Survey of the Portuguese pregnant women	Ad-hoc (Pregnant)	PT3
Romania	Romanian national food consumption survey for adolescents, adults and elderly	≥10 years	RO2
	Ad-hoc consumption survey for Romanian pregnant women	Ad-hoc (Pregnant)	RO3
	Ad-hoc consumption survey for Romanian vegetarian adults	Ad-hoc (Vegetarian)	RO4
Serbia	Serbian Food Consumption survey on children	<10 years	RS1
	Serbian Food Consumption Survey on adults	≥10 years	RS2
		Ad-hoc (Pregnant)	RS4
	Serbian Food Consumption Survey on vegetarians	Ad-hoc (Vegetarian)	RS3
Slovenia		<10 years[1]	SI1

EU Menu Country	Survey Name – EU Menu datasets received	Survey Target Group	Code
	Slovenian national food consumption survey	≥10 years	SI2
Spain[2]	Spanish National dietary survey on children and adolescents[2]	<10 years	ES1
	Spanish National dietary survey in adults, elderly and pregnant women[2]	≥10 years	ES2
		Ad-hoc (Pregnant)	ES3

1. Slovenia survey <10 years old only comprises infants and toddlers (<3 years)
2. The dataset from the Spanish National dietary survey on children and adolescents was split, and the data from adolescents combined with the dataset from the Spanish National dietary survey in adults, elderly and pregnant women to comply with the age groups defined (<10 and ≥10 years).

2.2 Quality indicators

Quality indicators are measurable, standardized, evidence-based measures that improve collection, analysis and monitoring of quality data and track the performance of accurate and harmonized food consumption outcomes within and among surveys.

The quality indicators proposed are grouped into nine dimensions and 27 sub-dimensions related to the usual phases of Dietary Surveys. Table 2 identifies those dimensions and sub-dimensions and the 96 indicators proposed to assess the surveys' data quality. The complete description of each indicator and the procedure for its assessment are described in the annexed protocol.

Table 2. Organization of quality indicators into survey dimensions and sub-dimensions

Survey Dimension	Sub-dimension	Indicators
A. Sampling plan: methods and coverage	A1. Target population	A1.1. Sampling frame
		A1.2.1 Coverage of the target population
		A1.2.2 Estimates of undercoverage, duplication, ineligibility and/or misclassification
	A2. Sampling design and procedures	A2.1.1 Probabilistic sampling design
		A2.1.2 Stratification by relevant variables (age, sex, etc.)
		A2.2 Sampling by waves
	A3. Sample size	A3.1 Target sample size estimated by statistical assumptions
	A4. Sampling error	A4.1.1 Estimation of standard error for BMI
		A4.1.2 Estimation of standard error for Energy Intake
		A4.1.3 Estimation of standard error for Food Groups
	A5. Representativeness	A5.1 Study sample with similar distribution as the target population

		A5.2 Weighting procedures
B. Recruitment	B1. Participation rate	B1.1 Response rate, participation rate, contact rate
	B2. Comparison between participants and non-participants	B2.1 Absence of selection bias
C. Training and supervision of interviewers	C1. Interviewers' selection criteria	C1.1 Background in Nutrition/Dietetics and/or interviewing experience
	C2. Training of interviewers	C2.1.1 Standardized training procedures – SOP in place?
		C2.1.2 Standardized training procedures – training according to SOP?
		C2.1.3 Standardized training procedures – training during pilot?
		C2.1.4 Standardized training procedures – all survey aspects covered?
		C2.2.1 Training duration and monitoring – regular training?
		C2.2.2 Training duration and monitoring – compare duration against benchmark
	C3. Supervision of interviewers	C3.1.1 Interviewer monitoring – coordination team?
		C3.1.2 Interviewer monitoring - % of re-contacts?
		C3.1.3 Interviewer monitoring – checking of answers, errors, missings?
		C3.1.4 Interviewer monitoring – dynamic strategy to address issues?
		C3.1.5 Interviewer monitoring – report solutions to all staff?
		C3.2.1 Observer Bias – variance for key-variables compared between interviewers?
		C3.2.2 Observer Bias - % of item-non response per interviewer monitored?
	C4. Pilot survey	C4.1.1 Pilot methods and tools – previous methods testing?
		C4.1.2 Pilot methods and tools – setting and protocols were the same?
D. Dietary data collection: Interview administration procedures	D1. Mode of administration of the interview	D1.1 At least one face-to-face interview
		D1.2.1 Single mode of administration to all participants at each interview
		D1.2.2 If different modes of administration – possible bias identified and reported?
	D2. Interview setting and timing	D2.1.1 Interview setting – participant's preference?
		D2.1.2 Interview setting – deviations monitored?
		D2.2 Participants with ≥2 interviews and average time interval between interviews
		D2.3.1 Distribution of dietary data by weekdays

		D2.3.2 Distribution of dietary data by season
		D2.4 Duration of interview
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
		E1.2 Use of the multi-pass method (or similar)
		E1.3 Monitoring of the interview time
		E1.4.1 Automatic checking for empty food consumption occasions, minimum/maximum accepted quantities per food type, and/or missing quantities
		E1.4.2 Automatic checking for easily forgotten foods (use of probing questions)?
		E1.4.3 Calculation of energy and macronutrient intake outliers at the end of interview
	E2. Quantification of dietary consumption	E2.1.1 Use of a validated food picture book
		E2.1.1 Use of a validated food picture book – updates validated?
		E2.2 Minimum number of pictures per photo series in picture book
		E2.3 Total number of picture series in the picture book
		E2.4 Availability of food standard units as a quantification method
		E2.5 Availability of default quantities as a quantification method
	E3. Food propensity questionnaire (FPQ) or Food frequency questionnaire (FFQ)	E3.1.1 Application of a FPQ or FFQ
		E3.1.2 Check for completeness if self-administered FPQ
		E3.2.1 Adequacy of the FPQ/FFQ – less frequently eaten foods
		E3.2.2 Adequacy of the FPQ/FFQ – probable hazardous
		E3.2.3 Adequacy of the FPQ/FFQ – dietary supplements
		E3.2.4 Adequacy of the FPQ/FFQ – seasonal variations
F. Non-dietary data collection	F1. Anthropometric data	F1.1.1 & F1.1.2 Measured (vs. self-reported) weight and height, in adults and children
		F1.1.3 Pre-pregnancy weight collected
		F1.2.1 Standardized measurements of weight and height – adults
		F1.2.2 Standardized measurements of weight and height – children
		F1.2.3 Standardized measurements of weight and height – protocol deviations
		F1.2.4 Standardized measurements of weight and height – trained interviewers
		F1.2.5 Standardized measurements of weight and height – equipment verification

		F1.3 Two standardized measurements available for each parameter (weight, height)
		F1.4 Digit preference in anthropometric measurements
	F2. Physical activity	F2.1.1 Physical activity assessment – validated questionnaire adults
		F2.1.2 Physical activity assessment – reporting method children
		F2.2 Collection of accurate physical activity measurements
G. Data completeness	G1. Completeness of food composition database	G1.1 Number of food items/recipes in the food composition database
		G1.2.1 Availability of a food supplements database
		G1.2.2 Availability of a food supplements database – quantification in grams
		G1.3.1 Availability of a recipe composition database
		G1.3.2 Availability of a recipe composition database – updated during fieldwork
		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	G2.1.1 Percentage of composite dishes in Foodex2 codes
		G2.1.2 Recipe code
		G2.2 Unclassified values in 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 Complete interviews (≥2 main meals per day)
		H1.4 Number of food items per interview
		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	H4.1 Calculation of misreporting of energy intake through recommended methods H4.2 Prevalence of plausible reporters

	and over-reporting)	
I. Results' reporting	I1. External validity of results	I1.1 Weighted results to ensure the representativeness
	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

Some indicators were evaluated using the data extracted from the surveys' methodological reports under WP1 tasks (literature-driven quality indicators) (van Rossum et al., 2022), while others were assessed from the surveys' datasets provided by EFSA (data-driven quality indicators).

The classification of the indicators according to the respective information source is described in Table 3.

Table 3. Sources of information used to assess data quality indicators by survey dimensions.

	Information for quality indicators assessment	
	Methodological reports	EFSA datasets
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	F1.2, F2.1, F2.2	F1.1, F1.3, F1.4
G. Data completeness	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	H1.1, H1.2, H1.3, H1.4, H1.5, H2.1, H2.2
I. Results' reporting	I2.1, I2.2, I3.1	-

2.3 Mapping of surveys according to quality indicators

Literature quality indicators, extracted from the EFSA survey reports, are at survey level. Most literature quality indicators were categorical variables that were summarized using

absolute and relative frequencies. Heatmaps per survey dimension were created to cluster the indicators and surveys in order to help identify patterns and trends. The heatmap was built using a hierarchical cluster technique using the complete method and the Gower distance, which applies to different variable types. Both indicators and surveys were clustered, and a transposed data matrix was used to group the indicators.

Data-driven quality indicators are estimated from the variables available in the EU Menu datasets and can be computed following different approaches. To summarize the information of indicators with available estimates and their corresponding standard errors, random-effects meta-analysis was employed to combine the results from all surveys. The meta-average obtained, and the respective 95%CI, was considered to summarize the information. For the remaining indicators, the median and interquartile range of the indicator were computed, and the survey with the best performance was defined as the benchmark. Furthermore, indicators representing several food groups (e.g., H1.2, H2.1), nutrients (e.g. H2.1), or other dataset variables (e.g., G2.2, G3.1), were visually organized in survey and variables clusters, using heatmaps with the Euclidian distance.

For a better interpretation of results, a benchmark approach was used by setting a reference point to which all the surveys were compared, because the gold standard result may not be achievable. Thus, the survey with the best performance was defined as the benchmark.

The visual representation of the data quality analysis of indicators is presented in the results section of this report and in the Appendix A. Whenever there was missing information for the indicator in any survey, the respective entries were omitted from the graphs. These analyses were conducted using the datasets stratified by age and ad-hoc groups, i.e., the overall 43 survey target groups.

2.4 Summary quality scores and associated factors

Considering the extensive array of individual quality indicators evaluated, a more concise approach to establish correlations among various survey quality aspects was proposed using summary quality scores. The quality indicators were combined according to the dimensions originally defined, and Principal Component Analysis (PCA) was performed per dimension, to obtain summary quality scores. The number of summary scores used to represent each dimension was defined according to the elbow rule. The elbow rule is a heuristic method used to determine the number of principal components (i.e., in this context "scores") to retain in a principal component analysis (PCA). It is based on the scree plot, which is a graph of the eigenvalues of the principal components, in decreasing order. The eigenvalues represent the amount of variance explained by each principal component. To determine the number of principal components to retain, you look for an "elbow" or a point in the scree plot where the eigenvalues start to level off. This point is interpreted as the point where the remaining principal components explain only a small amount of variance and can be discarded (Cattell, 1966). Each dimension score was obtained by aggregating multiple indicators within the dimension, following standardization (z-score), and applying the respective weights derived from the PCA analysis. For the interpretation of summary scores, it was considered that an indicator was represented in that score if the factor loading (correlation between the indicator and score) was higher than 0.4 (Hair et al., 2014).

Before conducting the PCA, indicators that did not distinguish surveys (e.g., same response to all surveys, 100% missing information) were removed for each dimension, and negative scoring indicators (i.e., indicators in which higher values correspond to lower quality) were reversed. If the survey had missing information for one continuous indicator, the average value of the indicator from the remaining surveys were attributed to create the score. For categorical indicators, missing information was considered equivalent to the response "no" (i.e., 0). Considering that the number of surveys is limited, first the occurrence of multidimensional outliers was verified, and if that was the case, the outliers were removed using the cut-off points given by the Mahalanobis distance technique.

PCA was not performed on Dimension B because indicator *B1.1 Response rate* did not adequately represented the dimension, as it was the only one that provided relevant and comprehensive information for most surveys. Moreover, indicator H4.1.2 Proportion of plausible reporters was also not included in the PCA for dimension H, because it was not calculated for children nor for surveys without energy intake. Its inclusion would expressively bias the conclusions for dimension H.

The Pearson correlation was used to test correlations between the different dimension scores. A search directed acyclic graph (DAG) was created to explore the associations observed, assessing which factors affect the survey quality outputs and the association's direction. To search the DAG, the rules from the Spirtes, Glymour, and Scheines (SGS) algorithm (Spirtes et al., 2000) were used. This algorithm allows to identify if there is an association between dimensions and the respective direction. In the end, we used confirmatory analysis to assess the goodness of fit of the DAG identified. The Robust Comparative Fit Index (CFI), Robust Tucker-Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA) were used to assess the fit. Values higher than 0.9 for CFI and TLI and lower than 0.08 for RSMEA are considered as acceptable fit.

To improve the interpretation of these results, and further explore the effect of several survey aspects on quality, for the scores that were associated with each other, indicator-indicator correlations were estimated, using the Pearson correlation (for continuous variables) or the point biserial correlation (for categorical variables). Furthermore, for some indicators that were not meaningful, were negatively associated with the PCA component used to obtain the score (i.e., factor loading <0.4 or negative factor loading, respectively), or were not included in the PCA, further exploratory analyses were conducted to observe the association between these individual indicators and the other summary scores. For this purpose, boxplots (for categorical indicators) and scatterplots (for continuous indicators) were created. These plots depict the distribution of the various dimension scores (Dimension scores A to I) based on the indicator.

Moreover, in an ecological analysis, the association between the summary quality dimensions and socioeconomic characteristics of the countries, collected from EUROSTAT and OECD matching the year of the survey, was assessed through Pearson correlation. These characteristics reflected different domains, namely income (GDP per capita, in product purchase standards), education level (% of the population with at least upper secondary educational attainment) and geographical area (% of the population living in rural areas).

The associations between summary quality scores, socioeconomic variables at the country-level, and age target group (<10 years, ≥10 years or both) were graphically represented for better visualization. Because most of the indicators were mainly dependent on the country and not on age groups, the analysis of the factors associated with data quality was not stratified by survey target groups (<10 years, ≥10 years and ad-hoc), and the original EU menu datasets were used (n=31).

3 Results

3.1 Mapping of surveys according to quality indicators

The following sub-sections briefly depict the main descriptive findings concerning the quality indicators across surveys and its interpretation. Additional graphic results for some indicators can be consulted in the Appendix A – Map surveys according to quality indicators: Extra figures.

3.1.1 **Dimension A** – Sampling plan: methods and coverage

Table 4 summarizes the information for categorical indicators from dimension A. Except from ad-hoc surveys, all surveys use a probabilistic sampling design, mostly multistage sampling (64.3% <10 years and 66.7% ≥10 years). Moreover, 92.9% of survey units <10 years and 88.9% ≥10 years stratify the sample according to relevant characteristics (sex, age, region). According to the survey's methodological reports, despite no survey presenting estimates of coverage error (A1.2.2), 21 survey target groups report that their sampling frames cover the entire target population (71% of <10 years and 61% of ≥10 years surveys). Except from ad-hoc surveys, one survey among <10 years and one ≥10 years, the surveys report that their samples present similar distribution to the target population. Nonetheless, nearly 60% of surveys report the application of weighting procedures to improve representativeness. The absence of further detailed statistics on the sample representativeness limits the objective analysis on the survey's sampling quality.

Table 4. Summary statistics for categorical indicators of Dimension A – Sampling plan: methods and coverage.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
A1.1.1	National population register (updated and accessible)	no	8	57.1	12	66.7	11	100.0
		yes	6	42.9	6	33.3	0	0.0
A1.2.1	Sampling frame covering the entire target population	no	4	28.6	7	38.9	11	100.0
		yes	10	71.4	11	61.1	0	0.0
A1.2.2	Estimates of under coverage, duplication, ineligibility and/or misclassification	no	14	100.0	18	100.0	11	100.0
		yes	0	0.0	0	0.0	0	0.0
A2.1.1	Probabilistic sampling design	convenience	0	0.0	0	0.0	11	100.0
		multistage	9	64.3	12	66.7	0	0.0
		stratified	2	14.3	1	5.6	0	0.0

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
		stratified random	3	21.4	5	27.8	0	0.0
A2.1.2	Sampling stratified by age classes. sex. and/or other important characteristics	no	1	7.1	2	11.1	6	54.5
		yes	13	92.9	16	88.9	5	45.5
A2.2	Sampling by waves in infants	no	7	50.0	0	0.0	0	0.0
		not applicable	0	0.0	18	100	11	100.0
		yes	7	50.0	0	0.0	0	0.0
A3.1.1	Adequate statistical procedure implemented for the sample size estimation	no	1	7.1	0	0.0	0	0.0
		yes	6	42.9	6	33.3	2	18.2
		missing	*7*	*50.0*	*12*	*66.7*	9	81.8
A5.1	Study sample with similar distribution as target population	no	1	7.1	1	5.6	11	100.0
		yes	13	92.9	17	94.4	0	0.0
A5.2	Weighting procedures applied	no	3	21.4	3	16.7	11	100.0
		yes	8	57.1	11	61.1	0	0.0
		missing	*3*	*21.4*	*4*	*22.2*	0	0.0

For literature indicators, as expected due to their convenience sampling design, ad-hoc surveys performed worse in this dimension, as shown in the heatmap (Figure A 1, from Appendix A) where these are grouped together. Moreover, the heatmap did not show a cluster pattern based on the target age group, but instead it created clusters based on countries.

Concerning sample size, most surveys do not report whether there was a proper statistical procedure for determining the target sample size (Indicator A3.1.1, Table 4). According to EFSA, the target sample size was, at minimum, 130 participants per sex in each age group (<1y, 1-2y, 3-9y, 10-17y, 18-64y, >64y). The Indicator A3.1.2 (Figure 2) depicts the percentage of the minimum target sample size achieved by the surveys. Most surveys reach the target, particularly among adults, with the exception of one, in elderly, for both sexes. Among children, particularly for younger ages, four surveys did not reach the cut-off of 130 participants for at least one gender. Among infants, four surveys did not reach the minimum of 130 participants per sex, and one does not even reach half of the target sample size. Among toddlers, two surveys do not reach the minimum (130 participants per sex).

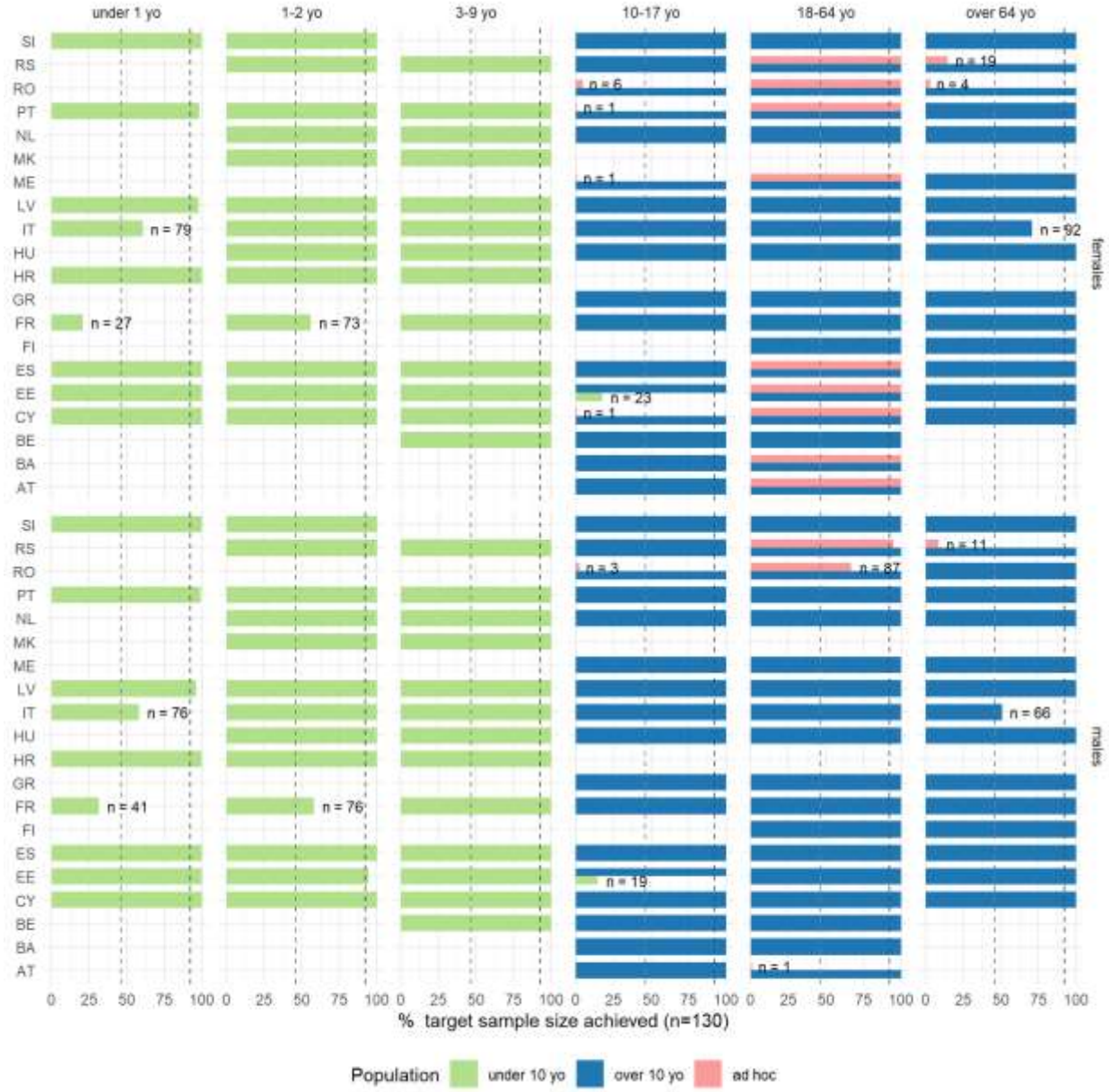

Figure 2. Indicator A3.1.2 – Minimum target sample size defined by EFSA per sex and age group, by EU Menu country. Note: In the figure, it is possible to see some participants classified as children (<10 years old, green bar) but falling within the column of 10-17 years. These are borderline participants that were sampled and evaluated as children but at the moment of the interview were already 10 years old.

The relative standard error (RSE), indicator A4.1, measures the precision of the estimates for some key variables, per unit of the variable. There may be two main reasons for higher RSE results and less precision in the estimates. First, a small sample size and second, a high heterogeneity in the groups evaluated, which can be seen in surveys including individuals from 10 to >80 years old or surveys from infants to 9 years old. In general, for the core variables tested (Energy intake, BMI, and Food Groups), the ad-hoc surveys performed worse regarding precision, with higher values of RSE (Figure A 2 to Figure A 4). Ad-hoc surveys are

characterized by relatively homogeneous participants (e.g., all pregnant women or all vegetarian adults) but smaller sample sizes. Thus, it is likely that the sample size is the factor that mostly affects RSE in these variables. Figure 3 supports this hypothesis by illustrating that, within each survey target group (comprising individuals aged <10 years, those aged ≥10 years, and ad-hoc groups), increased average sample size for specific age and sex subgroups is associated with lower values of Relative Standard Error (RSE) for both energy intake and BMI. Consequently, having a sample size greater than the minimum requirement serves to reduce the level of uncertainty surrounding the estimates. Nonetheless, even for ad-hoc surveys, which have relatively higher RSE, overall, the absolute RSE values for most variables tested are low. High values of RSE were observed only for some food groups that are not frequently consumed in specific surveys (Figure A 4).

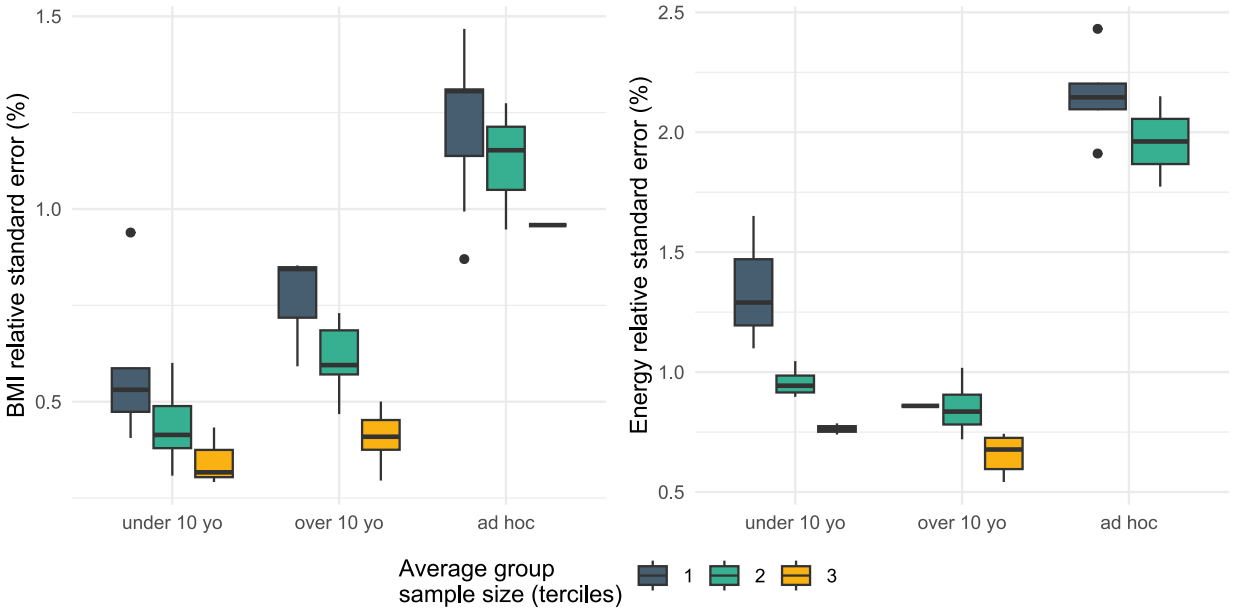

Figure 3. Comparison of Indicator A4.1 Relative Standard Error (RSE) for energy and BMI by terciles of average group sample size stratified by survey target group (<10 years, ≥10 years and ad-hoc).

3.1.2 **Dimension B** – Recruitment

The EU Menu Guidance proposes several rates to address subjects' participation in national dietary surveys, namely the contact rate (*eligible/(eligible + unknown eligible individuals)*), the cooperation rate (*participants / eligible individuals*) and the participation rate (*participants / (eligible + unknown eligible individuals)*). Nonetheless, in EFSA survey methodological reports, survey authors mostly report a so-called "Response-rate" somewhat equivalent to the cooperation rate. Others present a participation rate that does not compare to the one proposed in the EFSA EU Menu guidance, creating heterogeneity in the survey reports. Moreover, most surveys report only the response rate, disregarding the participation and contact rates, as shown in Table 5.

Table 5. Summary statistics for categorical indicators of Dimension B – Recruitment.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
B1.1	Response rate	with information	13	92.9	17	94.4	8	72.7
		missing	1	7.1	1	5.6	3	27.3
	Contact rate	with information	2	14.3	3	16.7	0	0.0
		missing	12	85.7	15	83.3	11	100.0
	Participation rate	with information	3	21.4	5	27.8	0	0.0
		missing	11	78.6	13	72.2	11	100.0
B2.1	Absence of selection bias: comparison of core study variables between participants and non-participants	no	3	21.4	2	11.1	0	0.0
		yes	4	28.6	4	22.2	2	18.2
		missing	*7*	*50.0*	*12*	*66.7*	*9*	*81.8*

The median (and min-max range) response rates (or cooperation rate) are 62.0% (22.8-85.0%), 59.0% (26.7-84.1%), and 77.6% (48.0-90.0%) for surveys <10 years, ≥10 years and ad-hoc, respectively (Figure 4 and Figure A 5, ad-hoc surveys are not presented in the figure due to their methodological differences and the corresponding plot is in appendix). The high variability ranges may also reflect that surveys are not considering the same denominator when defining participation, which hampers straight comparisons between surveys.

Most surveys (more than 50%) did not report on having performed a comparison between participants and non-participants, hampering the evaluation of a potential selection bias. A comparison between participants and non-participants for some variables (Indicator B2.1) was only reported in a minority of surveys: eight national (four in <10 years and four in ≥10 years and two ad-hoc surveys (Table 5). This comparison is relevant for evaluating a potential participation bias, and thus the analysis of survey results in the view of sample representativeness would benefit from the comparison between participants and non-participants (data from registers or refusal questionnaires, when possible) for at least some core variables.

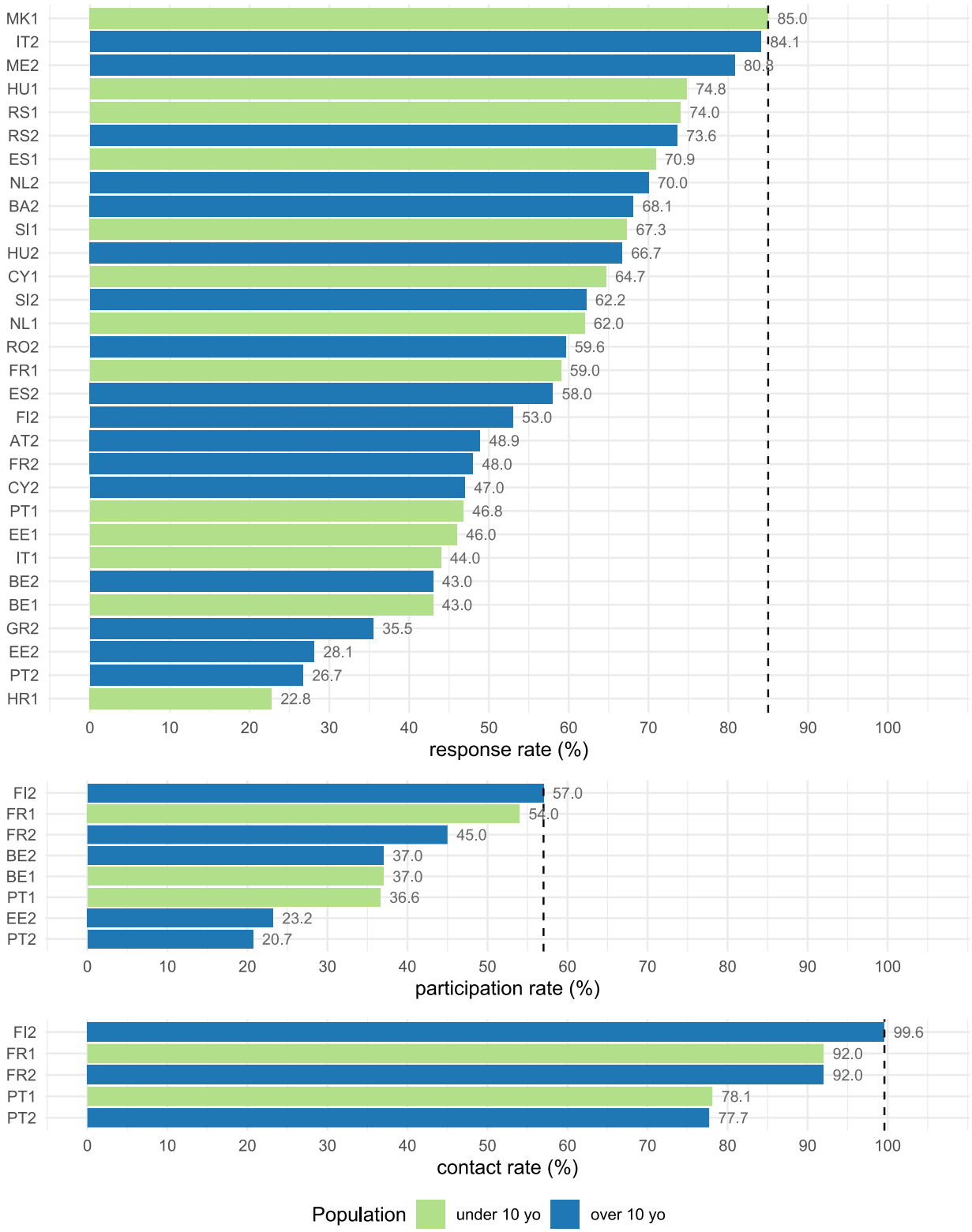

Figure 4. Indicator B1.1 Response rate, participation rate and contact rate.

3.1.3 **Dimension C –** Training and supervision of interviewers

Table 6 presents the information for most indicators from dimension C. Most surveys complied with the recommendation of having interviewers with a background in nutrition or dietetics or at least experienced in health assessment. Although all surveys report training of interviewers, there is a reasonable extent of missing information in the survey methodological reports for some proposed indicators. More than 70% of surveys (71.4% <10 years; 72.2% ≥10 years; 81.8% ad-hoc) do not report whether there were standard operating procedures in place for training, around 60% (57.1% <10 years; 61.1% ≥10 years; 36.4% ad-hoc) do not report if there was a training phase during the pilot or if the training was conducted regularly during the fieldwork, and no survey reports if the training of interviewers covered all survey aspects.

Table 6. Summary statistics for categorical indicators of Dimension C – Training and supervision of interviewers.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
C1.1	Background in Nutrition/Dietetics and/or interviewing experience[1]	a)	8	57.1	11	61.1	7	63.6
		a); b)	3	21.4	3	16.7	0	0.0
		a); c)	1	7.1	1	5.6	1	9.1
		b)	1	7.1	1	5.6	1	9.1
		c)	1	7.1	2	11.1	2	18.2
C2.1.1	Training: standard operating procedures (SOP) in place	no	0	0.0	0	0.0	0	0.0
		yes	4	28.6	5	27.8	2	18.2
		missing	*10*	*71.4*	*13*	*72.2*	9	81.8
C2.1.2	Training: conducted according to SOP	no	0	0.0	0	0.0	0	0.0
		yes	4	28.6	5	27.8	2	18.2
		missing	*10*	*71.4*	*13*	*72.2*	9	81.8
C2.1.3	Training: conducted during the pilot	no	0	0.0	0	0.0	0	0.0
		yes	6	42.9	7	38.9	7	63.6
		missing	*8*	*57.1*	*11*	*61.1*	4	36.4
C2.1.4	Training: all survey aspects covered	no	0	0.0	0	0.0	0	0.0
		yes	1	7.1	0	0.0	0	0.0
		missing	13	*92.9*	*18*	*100*	11	100.0
C2.2.1	Training: conducted at regular basis	no	6	42.9	8	44.4	5	45.5
		yes	8	57.1	10	55.6	6	54.5
C3.1.1	Interviewer monitoring: coordinator or coordination team defined	no	0	0.0	0	0.0	0	0.0
		yes	14	100.0	18	100.0	11	100.0
C3.1.3	Interviewer monitoring: Identification of possible errors. missing values by the coordination team	no	0	0.0	0	0.0	0	0.0
		yes	14	100.0	18	100.0	11	100.0
C3.1.4		no	0	0.0	0	0.0	0	0.0
		yes	6	42.9	8	44.4	6	54.5

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
	Interviewer monitoring: dynamic strategy to address issues emerging in the field	*missing*	*8*	*57.1*	*10*	*55.6*	5	45.5
C3.1.5	Interviewer monitoring: coordinator timely report solutions to all staff	no	0	0.0	0	0.0	0	0.0
		yes	3	21.4	5	27.8	7	63.6
		missing	*11*	*78.6*	*13*	*72.2*	4	36.4
C3.2.1	Observer bias: explained variance (%) for key variables compared between interviewers during the fieldwork	no	2	14.3	2	11.1	2	18.2
		yes	5	35.7	6	33.3	5	45.5
		missing	*7*	*50.0*	*10*	*55.6*	4	36.4
C3.2.2	Observer bias: proportion of item non-response, by interviewer, continuously monitored?	no	0	0.0	0	0.0	0	0.0
		yes	8	57.1	8	44.4	5	45.5
		missing	*6*	*42.9*	*10*	*55.6*	6	54.5
C4.1.1	Pilot study conducted	no	0	0.0	0	0.0	0	0.0
		yes	13	92.9	17	94.4	11	100.0
		missing	*1*	*7.1*	*1*	*5.6*	0	0.0
C4.1.2	Pilot study with similar setting and methods	no	0	0.0	0	0.0	0	0.0
		yes	0	0.0	0	0.0	0	0.0
		missing	14	100	18	100	11	100.0

[1] a) Dietetics/nutrition background; b) other interviewers experienced in health assessment; c) other situations

A similar trend is observed for subdimension C3 - Supervision of interviewers, where all surveys report having a coordination team responsible for monitoring the fieldwork and managing possible errors, but a large proportion of surveys fail to report details regarding the interviewer monitoring, strategies to solve fieldwork constraints or to deal with observer bias.

More than 90% of all surveys report conducting a pilot study before the survey, independently of the target group (<10, ≥10 years, and ad-hoc). However, the indicator from this dimension related to the pilot study's conduction in a similar setting and with similar methods (C4.1.2) could not be assessed due to missing information in all survey reports.

Overall, the performance of quality indicators of this dimension is dependent on the country and not on the age target group, because surveys from the same country have the same characteristics regarding the background and training of interviewers. This was confirmed and can be checked in the corresponding heatmap (Figure A 6, from Appendix A) and continuous indicator figures (Figure 5 and Figure A 7.), where survey target groups from the same country are grouped due to identical results.

Figure 5 shows the results for indicator C2.2.2 – Training duration of interviewers per survey, in hours. It ranges from 7 to 80 hours (considered as the benchmark) and is dependent on the country.

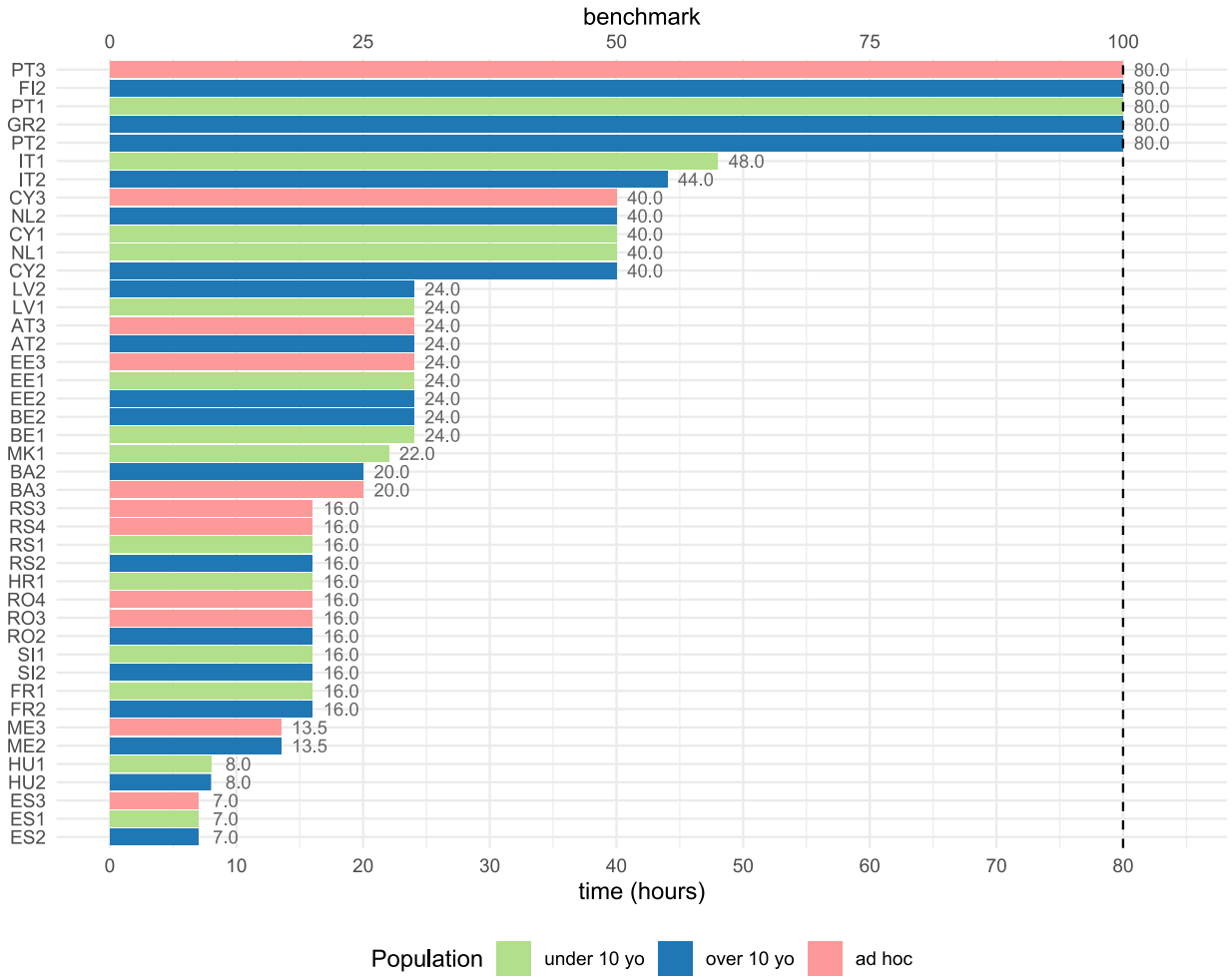

Figure 5. Indicator C2.2.2 – Training duration of interviewers per survey, in hours. The number of hours was estimated by multiplying the number of days reported in the methodological reports by 8 working hours.

3.1.4 **Dimension D –** Data collection: Interview administration procedures

Table 7 presents the summary of categorical indicators from dimension D. Most surveys conducted at least one face-to-face interview (>70% in all target groups). Around 60% of surveys have the same mode of administration to all participants at each interview versus 40% that do not comply with this. When applicable, only one survey checked the possible bias due to different setting. The participants preference for defining the interview setting was considered only in around 40% of surveys. The option of 'no choice' was higher in children surveys (35.7% for <10y vs. 27.8% for > 10y). No information was available to assess the indicator D2.1.2 Monitoring of deviations from pre-defined interview setting: 100% of missings.

Table 7. Summary statistics for categorical indicators of Dimension D – Data collection: Interview administration procedures.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
D1.1	At least one face-to-face interview	no	4	28.6	3	16.7	0	0.0
		yes	10	71.4	15	83.3	11	100.0
D1.2.1	Same mode of administration to all participants at each interview moment	no	6	42.9	6	33.3	4	36.4
		yes	8	57.1	12	66.7	7	63.6
D1.2.2	If different methods were applied, a possible bias identified and reported?	no	5	35.7	5	27.8	4	36.4
		not applicable	8	57.1	12	66.7	7	63.6
		yes	1	7.1	1	5.6	0	0.0
D2.1.1	Interview setting according to participant's preference	no	5	35.7	5	27.8	3	27.3
		yes	6	42.9	8	44.4	6	54.5
		missing	*3*	*21.4*	*5*	*27.8*	*2*	*18.2*
D2.1.2	Monitoring of deviations from pre-defined interview setting	no	0	0.0	0	0.0	0	0.0
		yes	0	0.0	0	0.0	0	0.0
		missing	*14*	*100.0*	*18*	*100.0*	*11*	*100.0*

In general, also the scoring of the indicators from Dimension D depend more on the country than on the survey target group (<10 years, ≥10 years and ad-hoc), as shown in the heatmap (Figure A 8, Appendix A). This was expected as interview administration procedures are defined mainly by the survey coordination team and are less dependent on the target group assessed, which is especially evident whenever the different population groups were assessed together under the same survey.

Dietary interviews were expected to be distributed uniformly by weekdays (14.3% each) and seasons (25% each). Despite most surveys tried to comply with this recommendation, some discrepancies were observed. Regarding weekdays, when surveys drift from the uniform distribution, it is mainly due to a lack of interviews on Fridays and Saturdays. Sundays also present an irregular proportion of dietary interviews, ranging from 3 to 30% depending on the survey. In general, <10 years surveys better cover all weekdays than the remaining groups. Concerning seasons, three surveys (two ad-hoc and one ≥10 years) only have interviews in two different seasons, which was the minimum requirement from EFSA. The remaining surveys have at least one interview each season, but the proportions vary significantly. There are, however, no clear differences between survey target groups. Figure 6 presents the indicator results, and the figures representing the benchmark can be found in the Appendix (Figure A 9 and Figure A 10).

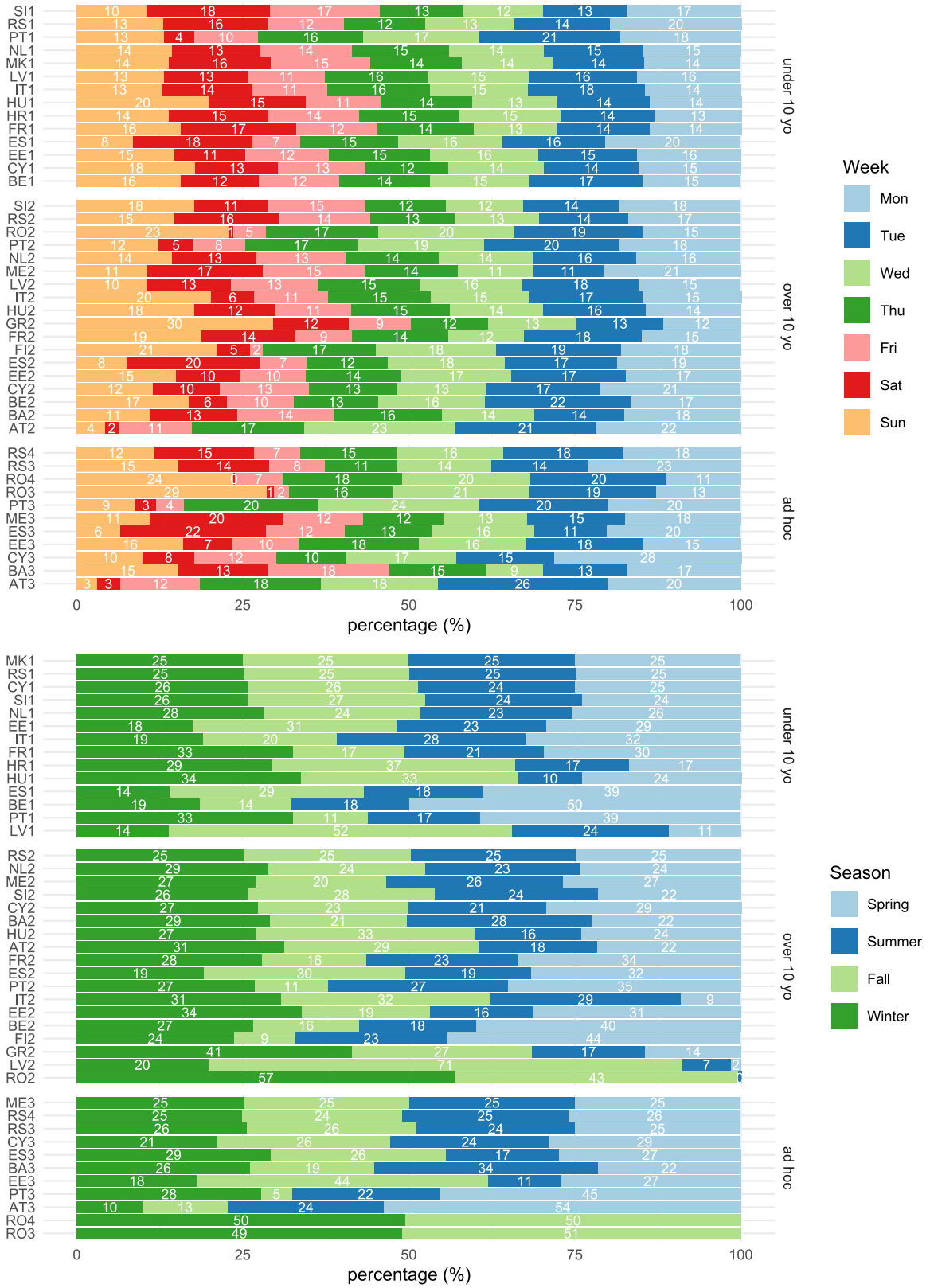

Figure 6. Indicator D2.3 – Distribution of dietary data by weekdays and season.

According to the EU Menu guidance, participants should have undergone two non-consecutive interviews. The percentage of complete participants (Figure 7, Indicator D2.2.0) varies between 81.9% to 100%. However, it is unclear whether data providers choose to share with EFSA data from all subjects or only complete cases. For complete participants, Indicator D2.2.1 shows the meta-average of the percentage of participants with an interview gap of 8-15 days was around 52%, ranging from 0.5% to 90.3% (Figure 8).

Regarding this indicator (D2.2.1 Percentage of participants with an interview gap of 8-15 days), an important remark has to be made. The indicators assessed in this work were mainly based in the recommendations from the EFSA guidance. However, in some cases we found some difficulties in interpreting the information from the guidance, which may have impacted the quality assessment. For example, in the EFSA Guidance of 2014 it is mentioned that the interviews should be separated by a period of at least 8 days. However, also in the guidance it is mentioned that the interviews should be "one-to-two weeks apart". The interpretation of this last sentence is ambiguous and led our team to define one indicator as the "proportion of participants with interview gap between 8-15 days".

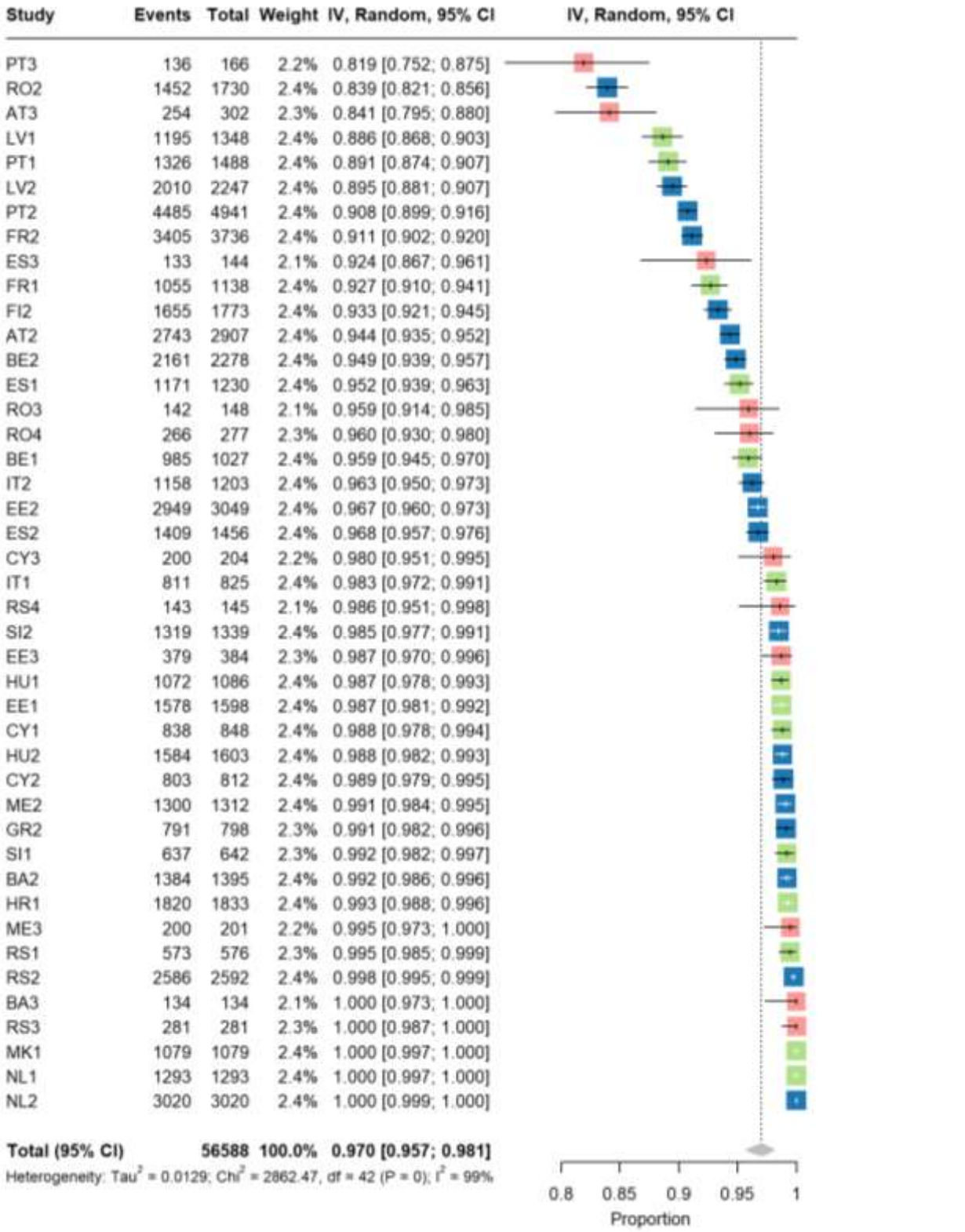

Figure 7. Indicator D2.2.0 – Proportion of complete participants (≥2 interviews) in the databases

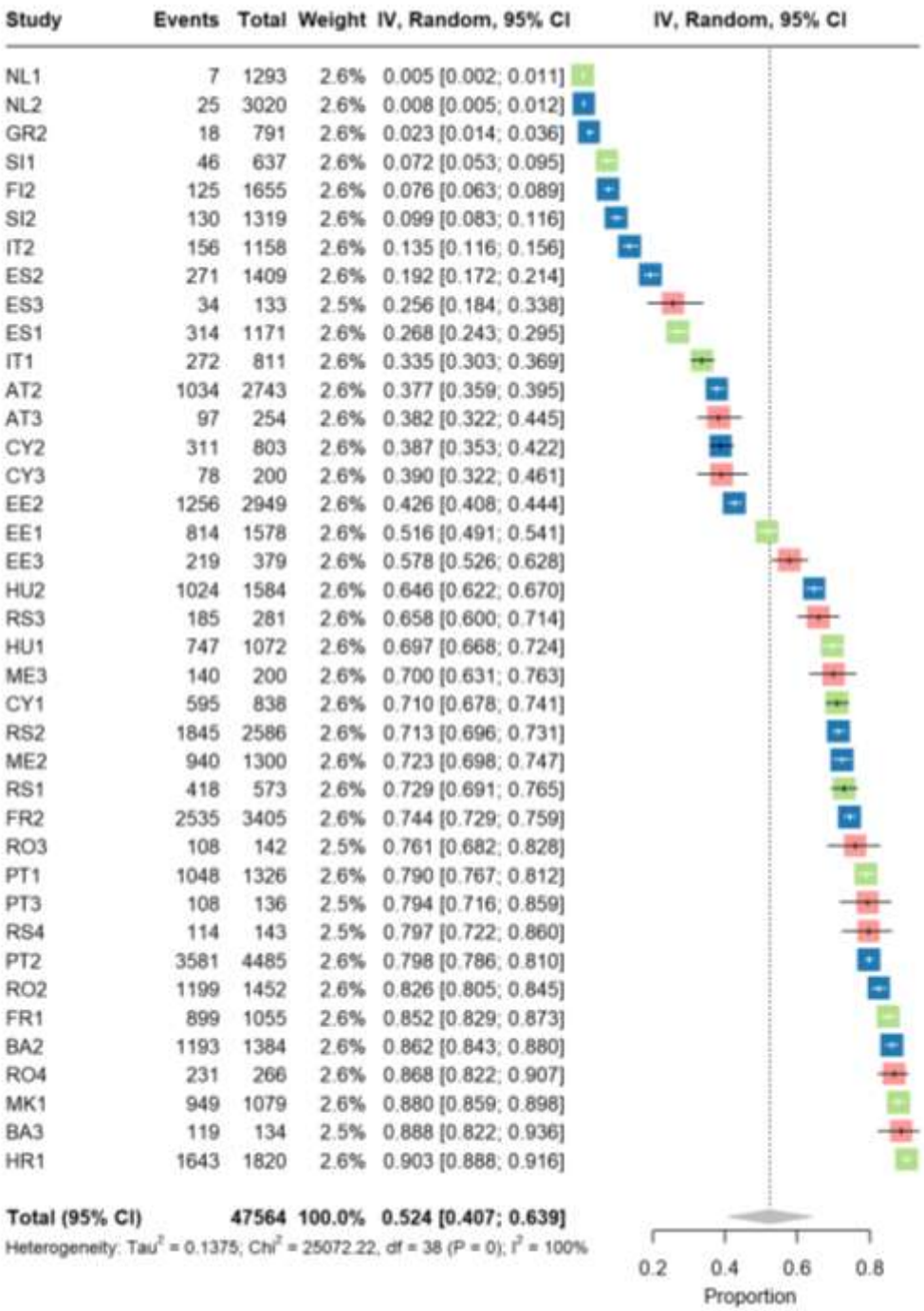

Figure 8. Indicator D2.2.1 – Proportion of participants with interview gap between 8 and 15 days.

The indicator D2.4.1 Duration of 24-hours recall (24h-R) and total interview was evaluated based on the information in the survey reports, which were not always available. The 24h-R duration ranged from 13-45' and total interview from 31-64' (Figure A 11 and Figure A 12, Appendix A), which can include many different survey dimensions, but was not always clear

in the reports. Further harmonization is necessary, especially concerning the duration of 24h-R, as eight survey units report 24h-R with less than 30 minutes.

3.1.5 **Dimension E –** Data processing: software tools and validation procedures

A validated tool for dietary data collection complying with the multipass method was adopted by most surveys (71.4% <10 years; 72.2% ≥10 years; 63.6% ad-hoc), as well as at least one automatic checking and probing questions (e.g., missing entries, amounts, easily forgotten foods, outliers) (85.7% <10 years; 88.9% ≥10 years; 90.9% ad-hoc), as shown in Table 8. Nonetheless, the reports provided varying and frequently missing information regarding the availability automatic quality controls and probing questions used in the different software tools. In lack of further information, we considered it missing rather than a negative answer. Nevertheless, this suggests possible heterogeneity in the software tools used in each survey.

Table 8. Summary statistics for categorical indicators of Dimension E – Data processing: software tools and validation procedures.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
E1.1	Use of a validated dietary assessment tool	no	1	7.1	1	5.6	1	9.1
		yes	9	64.3	11	61.1	7	63.6
		missing	*4*	*28.6*	*6*	*33.3*	*3*	*27.3*
E1.2	Use of the multi-pass method (or similar)	no	1	7.1	1	5.6	1	9.1
		yes	10	71.4	13	72.2	7	63.6
		missing	*3*	*21.4*	*4*	*22.2*	*3*	*27.3*
E1.3	Monitoring of the interview time	no	0	0.0	0	0.0	0	0.0
		yes	9	64.3	12	66.7	8	72.7
		missing	*5*	*35.7*	*6*	*33.3*	*3*	*27.3*
E1.4.1.1	Automatic checking for empty food consumption occasions	no	0	0.0	0	0.0	0	0.0
		yes	10	71.4	14	77.8	8	72.7
		missing	*4*	*28.6*	*4*	*22.2*	*3*	*27.3*
E1.4.1.2	Automatic checking for minimum/maximum accepted quantities per food type	no	0	0.0	0	0.0	0	0.0
		yes	8	57.1	11	61.1	9	81.8
		missing	*6*	*42.9*	*7*	*38.9*	*2*	*18.2*
E1.4.1.3	Automatic checking for missing quantities	no	0	0.0	0	0.0	0	0.0
		yes	10	71.4	13	72.2	10	90.9
		missing	*4*	*28.6*	*5*	*27.8*	*1*	*9.1*
E1.4.2	Automatic probing questions for easily forgotten foods	no	0	0.0	0	0.0	0	0.0
		yes	12	85.7	16	88.9	10	90.9
		missing	2	14.3	2	11.1	1	9.1
E1.4.3	Calculation of energy and macronutrient intake and	no	2	14.3	4	22.2	4	36.4
		yes	11	78.6	12	66.7	6	54.5
		missing	*1*	*7.1*	*2*	*11.1*	*1*	*9.1*

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
	identification of outliers at the end of interviews							
E2.1.1	Use of a validated food picture book	no	0	0.0	0	0.0	0	0.0
		yes	13	92.9	17	94.4	11	100.0
		missing	*1*	*7.1*	*1*	*5.6*	*0*	*0.0*
E2.1.2	If the picture book has been updated, has it been validated again?	no	0	0.0	0	0.0	0	0.0
		not applicable	7	50.0	8	44.4	8	72.7
		yes	4	28.6	6	33.3	1	9.1
		missing	*3*	*21.4*	*4*	*22.2*	*2*	*18.2*
E2.2	Minimum number of pictures per photo series in picture book (≥4)	no	1	7.1	3	16.7	1	9.1
		yes	8	57.1	8	44.4	6	54.5
		missing	*5*	*35.7*	*7*	*38.9*	*4*	*36.4*
E2.4	Availability of food standard units as a quantification method	no	0	0.0	0	0.0	0	0.0
		yes	12	85.7	16	88.9	10	90.9
		missing	*2*	*14.3*	*2*	*11.1*	*1*	*9.1*
E2.5	Availability of default quantities as a quantification method	no	2	14.3	2	11.1	1	9.1
		yes	8	57.1	7	38.9	4	36.4
		missing	*4*	*28.6*	*9*	*50.0*	*6*	*54.5*
E3.1.1	Application of a FPQ or FFQ	no	0	0.0	0	0.0	0	0.0
		yes	14	100.0	18	100.0	11	100.0
E3.1.2	Self-administered FPQ/FFQ subsequently checked for completeness	no	0	0.0	0	0.0	0	0.0
		not applicable	7	50.0	5	27.8	2	18.2
		yes	2	14.3	4	22.2	1	9.1
		missing	*5*	*35.7*	*9*	*50.0*	*8*	*72.7*
E3.2.1	FPQ: includes less frequently eaten foods	no	2	14.3	2	11.1	1	9.1
		yes	7	50.0	10	55.6	7	63.6
		missing	*5*	*35.7*	*6*	*33.3*	*3*	*27.3*
E3.2.2	FPQ: includes foods w/ higher contamination potential	no	3	21.4	2	11.1	2	18.2
		yes	6	42.9	10	55.6	5	45.5
		missing	*5*	*35.7*	*6*	*33.3*	*4*	*36.4*
E3.2.3	FPQ: includes dietary supplements	no	1	7.1	1	5.6	0	0.0
		yes	12	85.7	16	88.9	11	100.0
		missing	*1*	*7.1*	*1*	*5.6*	*0*	*0.0*
E3.2.4	FPQ: designed to capture seasonal variation of foods	no	0	0.0	0	0.0	0	0.0
		yes	4	28.6	5	27.8	4	36.4
		missing	*10*	*71.4*	*13*	*72.2*	*7*	*63.6*

As before, some survey reports present missing information on indicators of this dimension. The missing patterns are not influenced by the survey target group but seem to be determined by the tools used in each country (heatmap, Figure A 13, Appendix A).

Despite all surveys reporting applying a FPQ, a high proportion of data from the survey reports regarding its specific characteristics is missing, which hindered the quality evaluation.

Furthermore, it was not possible to assess the quality of the FPQs from data-driven indicators since the EU Menu datasets do not include FPQ-related variables.

Out of the 43 surveys, 41 reported using a picture book for portion size estimation, and 36 included information of the number of picture series, which ranged from 21 to 333, showing high heterogeneity (Figure 9).

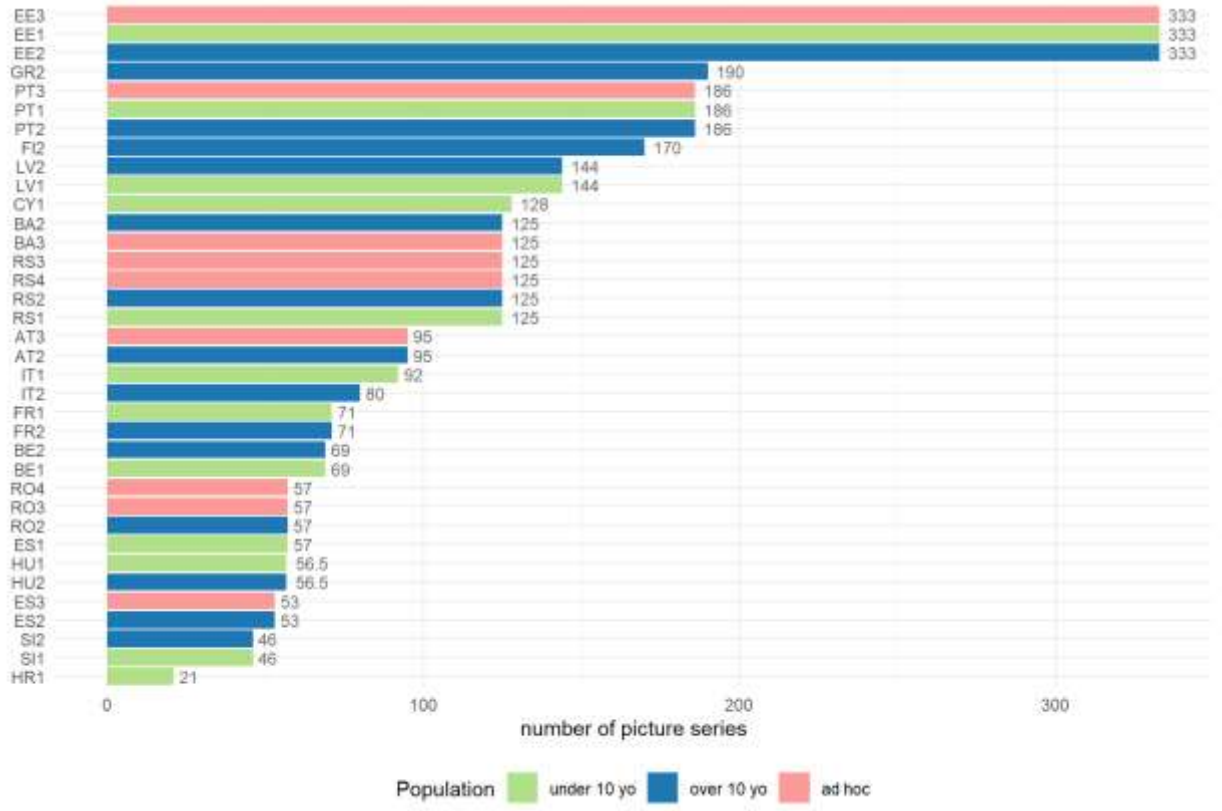

Figure 9. Indicator E2.3 – Total number of picture series in the picture book.

3.1.6 **Dimension F –** Non-dietary data collection

Table 9 and Figure A 14 show the results of categorical indicators of Dimension F. Anthropometric data were measured using standardized procedures in most surveys. In the case of pregnant women, more than 80% of surveys do not report whether the self-reported weight before pregnancy was collected. More than 90% of surveys also do not report whether possible deviations from the protocol for measuring weight and height were monitored. In general (>80% of surveys), the interviewers were trained regularly through repeated anthropometric measurements but most surveys do not report whether the equipment used for anthropometric measurements had been regularly checked (e.g. proper calibration of stadiometer or body scale on a regular basis)..

The assessment of physical activity levels using the IPAQ was highly prevalent in surveys ≥10 years (77.8%) and ad-hoc (72.7%), but only one survey collected accurate measurements through accelerometers in a subsample. Regarding surveys on participants <10 years, only

50.0% reported assessing physical activity with a self-reporting method, and even fewer (14.3%) reported having collected accurate measurements using accelerometers.

Table 9. Summary statistics for categorical indicators of Dimension F – Non-dietary data collection

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
F1.1.3	Pregnant - Self-reported weight before pregnancy collected	no	0	0.0	0	0.0	0	0.0
		not applicable	14	100.0	0	0.0	1	9.1
		yes	0	0.0	1	5.6	1	9.1
		missing	*0*	*0.0*	*17*	*94.4*	*9*	*81.8*
F1.2.1	Standardized procedures for anthropometric measurements - Adults	no	0	0.0	0	0.0	0	0.0
		not applicable	14	100.0	2	11.1	5	45.5
		yes	0	0.0	16	88.9	2	18.2
F1.2.2	Standardized procedures for anthropometric measurements - Children	no	1	7.1	0	0.0	0	0.0
		not applicable	0	0.0	18	100.0	11	100.0
		yes	12	85.7	0	0.0	0	0.0
		missing	*1*	*7.1*	*0*	*0.0*	*0*	*0.0*
F1.2.3	Report of deviations from the protocol for measuring weight and height	no	0	0.0	0	0.0	0	0.0
		not applicable	0	0.0	2	11.1	4	36.4
		yes	1	7.1	1	5.6	0	0.0
		missing	*13*	*92.9*	*15*	*83.3*	*7*	*63.6*
F1.2.4	Regular training for interviewers for anthropometric measurement	no	0	0.0	0	0.0	0	0.0
		not applicable	0	0.0	2	11.1	4	36.4
		yes	12	85.7	13	72.2	4	36.4
		missing	*2*	*14.3*	*3*	*16.7*	*3*	*27.3*
F1.2.5	Regular check and calibration of equipment for anthropometric measurements	no	0	0.0	0	0.0	0	0.0
		not applicable	0	0.0	2	11.1	4	36.4
		yes	5	35.7	7	38.9	1	9.1
		missing	*9*	*64.3*	*9*	*50.0*	*6*	*54.5*
F2.1.1	Physical activity assessment - IPAQ adults	no	0	0.0	2	11.1	2	18.2
		not applicable	14	100.0	0	0.0	0	0.0
		yes	0	0.0	14	77.8	8	72.7
		missing	*0*	*0.0*	*2*	*11.1*	*1*	*9.1*
F2.1.2	Physical activity assessment - reporting method children	no	6	42.9	0	0.0	0	0.0
		not applicable	0	0.0	18	100.0	11	100.0
		yes	7	50.0	0	0.0	0	0.0

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
		missing	*1*	*7.1*	*0*	*0.0*	*0*	*0.0*
F2.2	Collection of accurate physical activity measurements (e.g. accelerometers)	no	9	64.3	17	94.4	11	100.0
		yes	2	14.3	1	5.6	0	0.0
		missing	*3*	*21.4*	*0*	*0.0*	*0*	*0.0*

As depicted in the F1.1 Percentage of individuals with anthropometric measurements indicator figures in appendix (Figure A 15 and Figure A 16), weight and height present similar results. For both measures there are a group of surveys (5 for weight and 3 for height) that have almost 100% of individuals with self-report measurement, and the meta-analyses indicate that in average more than 80% of individuals have standardized measurements (84.2% for weight and 89.6% for height). Out of the 43 survey target groups, 15 are below the meta-average of measured weight, where six are ad-hoc, 8 are ≥10 years, and only one <10 years (meaning a better performance in children). For measured height, 13 survey units are below the meta-average, and no patterns are observed dependent on the survey target group. As expected, the surveys with more self-reporting in anthropometric variables also exhibited higher digit preference (Indicator F1.4, Figure A 17 to Figure A 20, Appendix A).

3.1.7 **Dimension G –** Data completeness

Indicators from Dimension G reflect how surveys described and characterized the reported foods, supplements and recipes and the missing information in the datasets shared with EFSA.

Apart from two surveys (one <10 years and one ≥10 years), all assessed food supplements within the 24h-R/food diaries (Indicator G1.2.1). However, for half of surveys, it was not reported whether supplements quantification in grams was available (Indicator G1.2.2). Moreover, except for two surveys (one <10 years and one ≥10 years), all have a recipe database (Indicator G1.3.1). Most surveys (34 out of 43) reported having updated the database during fieldwork (9 out of 43 with missing information), as evaluated in the Indicator G1.3.2 (Appendix A Table A 1 and Heatmap Figure A 21).

Missing and or unclassified information in the dataset variables was evaluated in indicators G1.5 (missings in foods' nutrient composition), G2.2 (unclassified values in consumption dataset variables) and G3.1 (missings and unclassified values in subjects dataset variables). Regarding G1.5 (Figure 10), various surveys do not report some (n=16) or even all (n=15) energy, macronutrients, water and alcohol contents in the Foods dataset, mainly because these variables were not mandatory.

In the consumption dataset (Figure 11), low prevalence of unclassified values in the mandatory variables are observed: only 5 surveys from two countries present more than 30% of missings in the variable "Place" and 3 surveys from the same country present around 60% of missings in "Exception day". These values may result from a limitation of the catalogue of options to classify these variables. In many cases the unclassified values were explained in a commentary variable. However, this variable was not available for our team to assess.

As far as the G3.1 (Figure 12), several variables (non-mandatory) present 100% of missings in many surveys. Some examples include reports of Energy intake, misreporters, profession-related, labour-related variables and ethnicity. Also, education-related variables are missing for 100% of participants in three surveys ≥10 years (self-education variable: EDUCATIONS) and two surveys <10 years (parental education variables: EDUCATIONM and EDUCATIONF). Physical activity information is also lacking in most surveys under and over 10 years but it is reported in most ad-hoc surveys. Non-mandatory variables may exhibit a multitude of missing values, even if the surveys possess that information. However, due to their non-mandatory nature, the surveys may have chosen not to provide the information. Regarding mandatory variables from the Subjects datasets, in this analysis, no unclassified values were found, thus we decided to not present the plot.

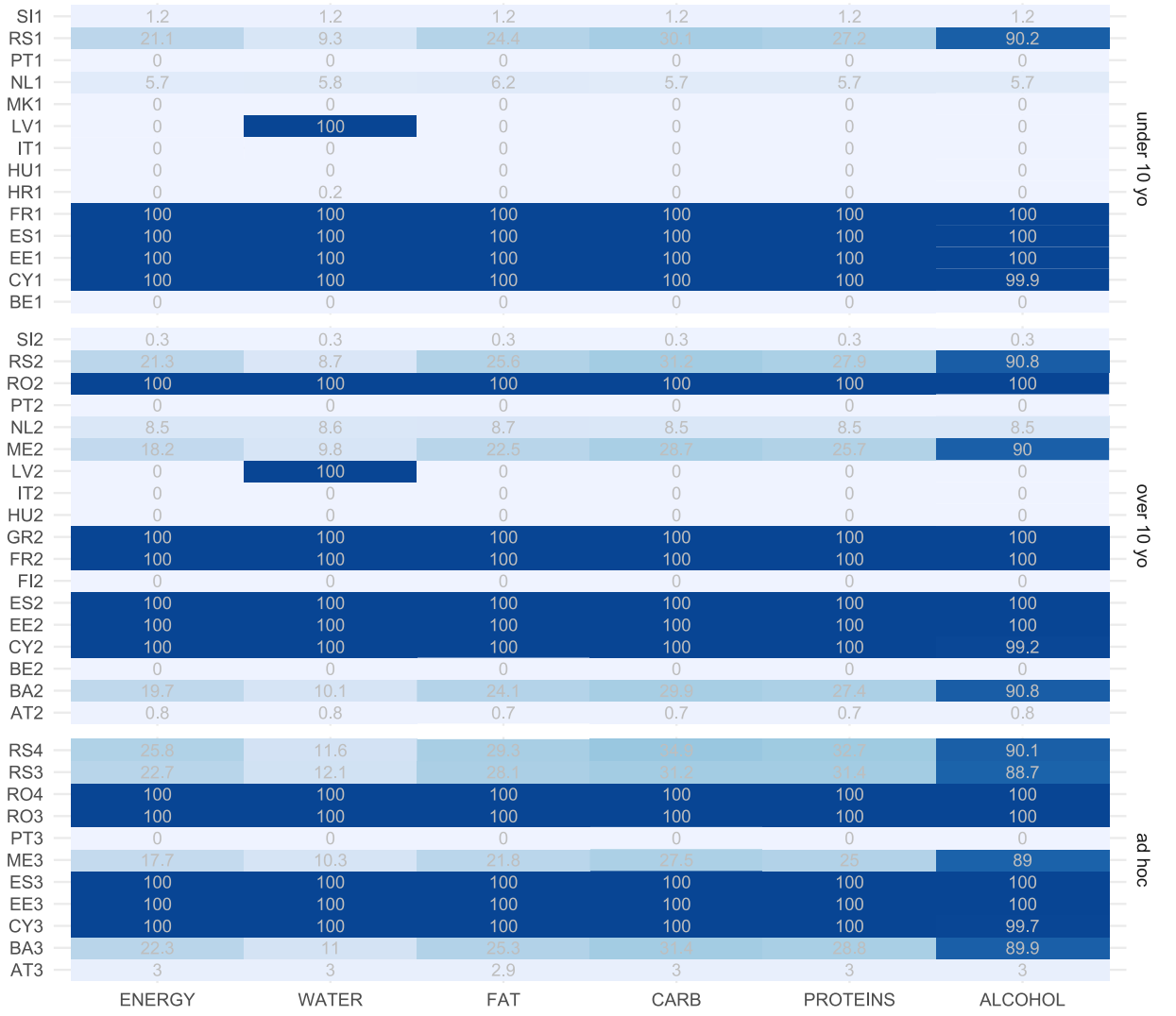

Figure 10. Indicator G1.5 – Missing values in energy, macronutrients, water and alcohol (non-mandatory) per 100g of food (calculated as % of food items with missing information in the variables mentioned).

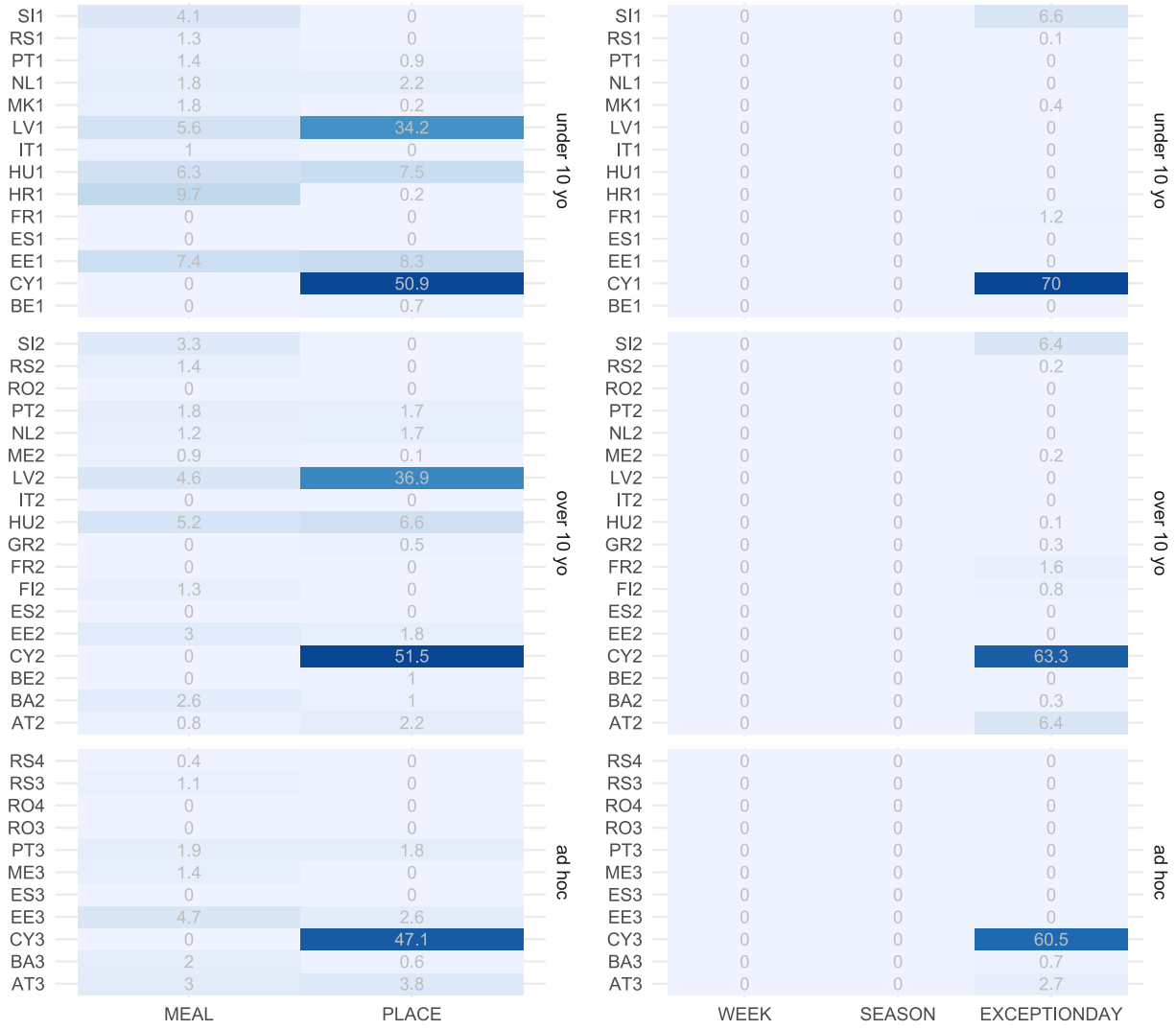

Figure 11. Indicator G2.2 – Unclassified values in mandatory variables: Consumption dataset.

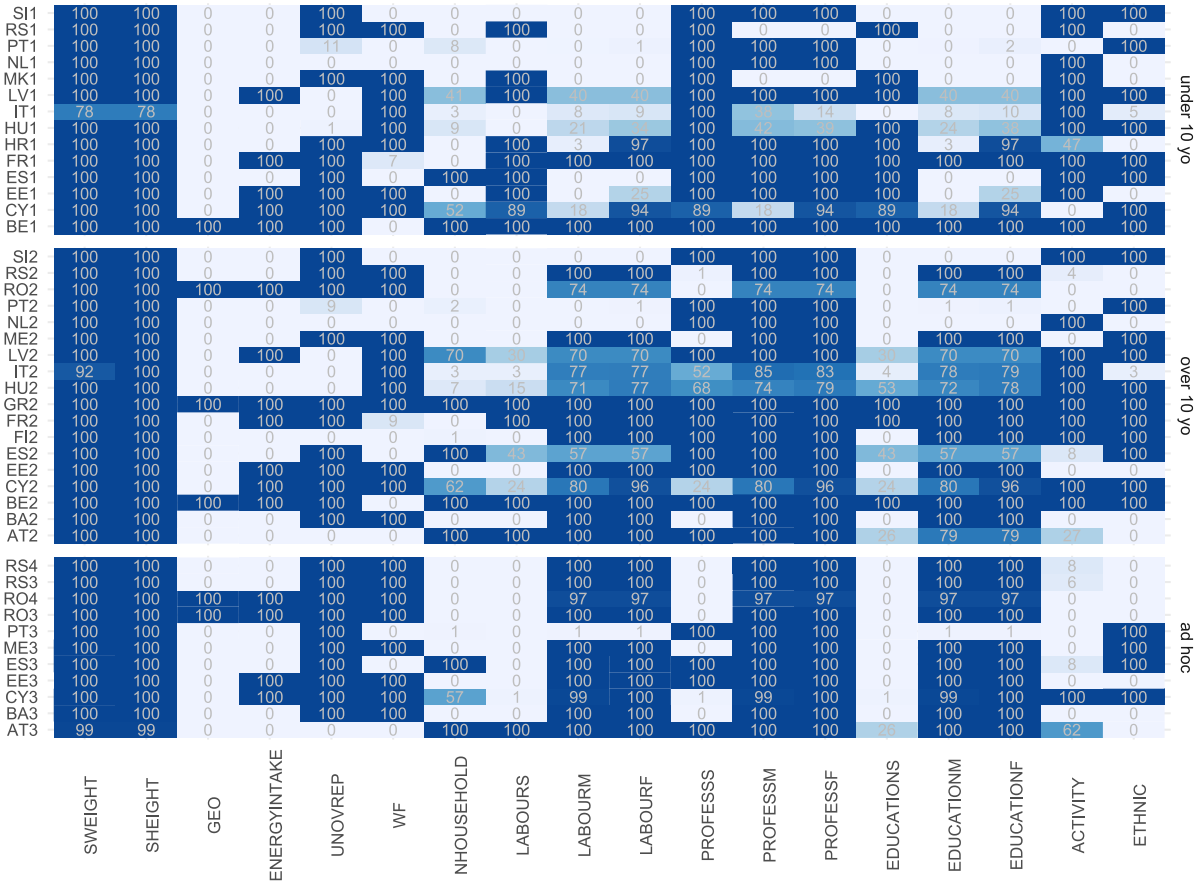

Figure 12. Indicator G3.1 – Missing values in non-mandatory variables: Subjects dataset

The use of the FoodEx2 classification system was evaluated through several indicators in this dimension, particularly regarding the use of facets and specificity of codes according to the FoodEx2 exposure hierarchy. Before describing the indicators and the respective results, in view of its interpretation, it is important to mention that the facet use was evaluated based on the added facets (i.e., excluding the implicit facets) and by considering each food report in the consumption file independently. This approach allows to accurately assess the indicators based on each separate report of each food item, controlling for the possible bias that could arise from assumptions on the similarity of food items, if a list of the unique consumed food items was considered. Moreover, using this approach allows for more frequently consumed foods to drive the results and these are weighed for often consumed foods.

Indicator G1.4 evaluated the proportion of reported foods with only the FoodEx2 base term, considering the complete list of single food items from the consumption file. Figure 13 shows the results from this indicator that varies mostly between 6.4% and 68% but goes up to 90% for 3 surveys from the same country.

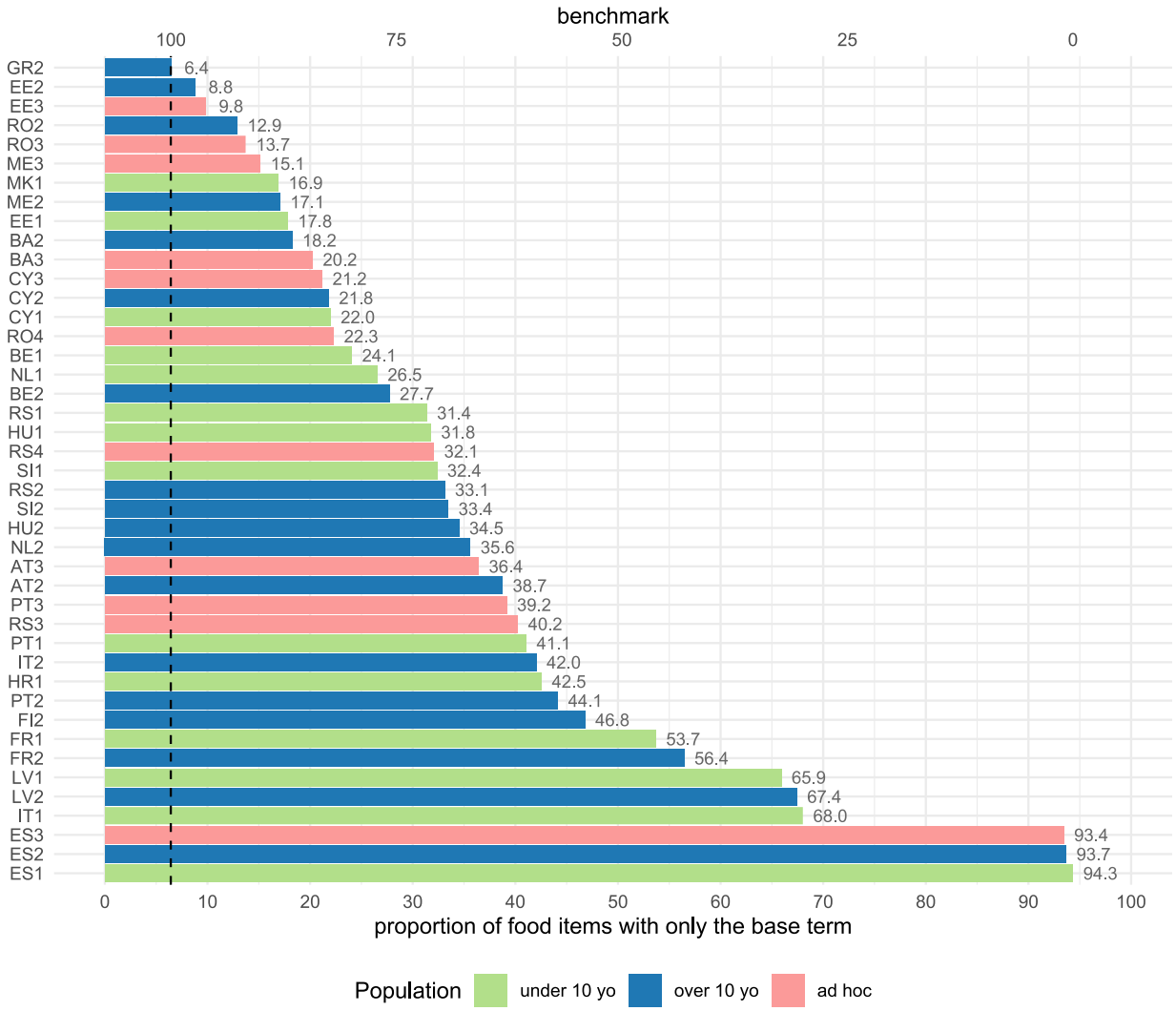

Figure 13. Indicator G1.4 – Food items with only FoodEx2 base term.

In general, facets use depends on the foods, but also on the software used by each country. There are seven facets considered relevant according to the EFSA Guidance, namely F01 (source), F08 (sweetening agent), F09 (fortification agent), F10 (qualitative-info), F19 (packaging-material), F20 (part-consumed-analysed) and F28 (processing). Figure 14 presents the proportion of foods from each survey using these facets (extrinsically). Overall, this plot shows that even the recommended facets are used scarcely in the surveys. However, not all foods are expected to include all these facets. Thus, in the next paragraphs from this report, extrinsic facet use will be mainly described according to food groups.

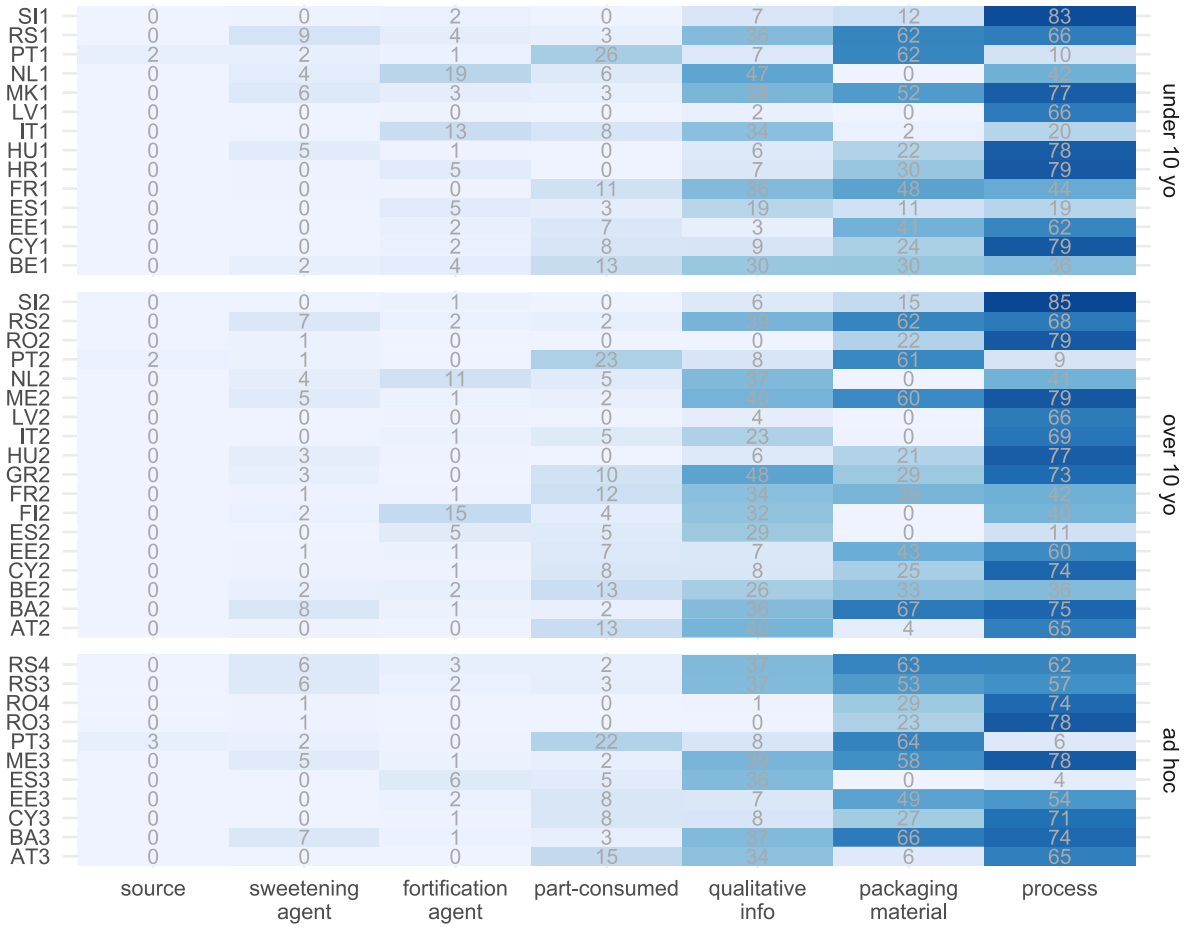

Figure 14. Proportion of foods reported with each recommended facet per survey. Calculated using the consumed FoodEx2 codes, without implicit facets.

F01, "Source", defines the origin of raw commodities and it is mainly used as an implicit facet. Thus, as only the added facets are being considered in this assessment, this facet is generally not reported by the participants, as shown in Figure A 26 presented in the appendix. **F08, "Sweetening agent"**, is used mainly in beverages (e.g., coffee, cocoa and other hot drinks, tea and infusions, soft drinks) or dairy (e.g., fermented milk and dairy desserts). However, even in these food groups, where sweetening agents are expected, the use of this facet is heterogeneous, with many surveys presenting 0% of foods from these groups with this facet while in other surveys the percentages may go up to 100%, whereas the median of report does not exceed the 10%, as shown in Figure 15.

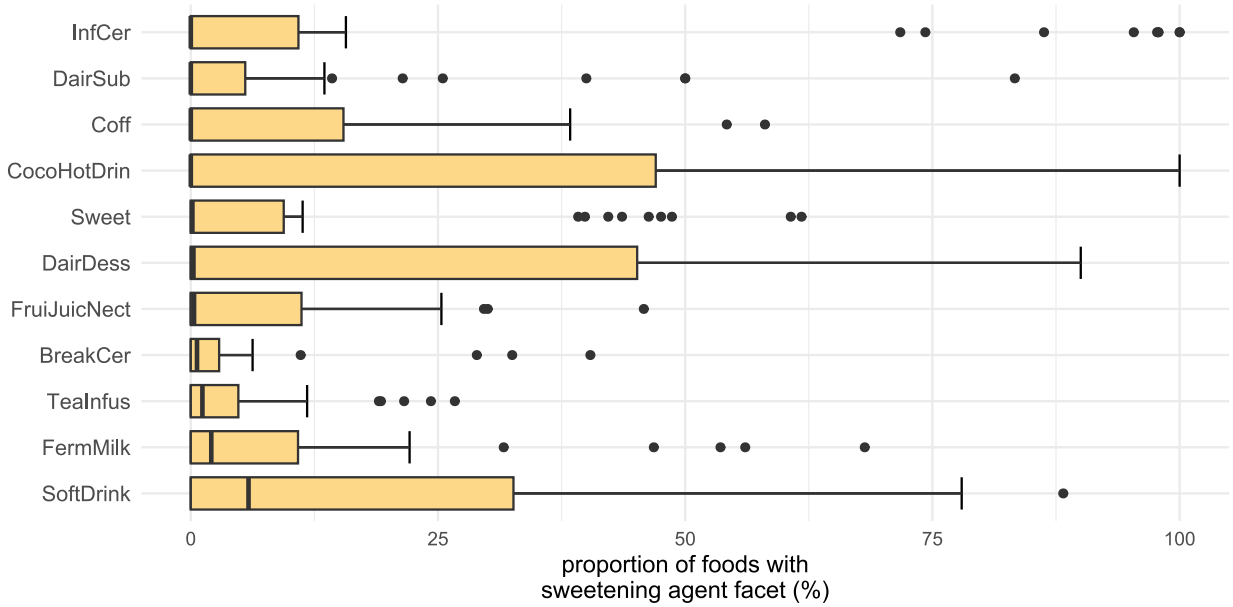

Figure 15. Summary of F08 (sweetening agent) use in the most relevant food groups, calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **BreakCer**: Breakfast cereals; **FermMilk:** Fermented milk products; **DairDess:** Dairy dessert and similar; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **SoftDrink:** Soft drinks and energy drinks; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfCer:** Processed cereal-based food for infants and young children; **DairSub:** Dairy imitates

The use of **F09, "Fortification agent"**, also depends on the food group, but its use is once again heterogeneous and scant. For instance, it is expected that most infant cereals are fortified. However, even for this food group, the median does not reach the 5%, and it is possible to observe a wide range of F09 use (0-100%). The Dairy Substitutes and Margarines are other examples of heterogeneous and limited use of this facet.

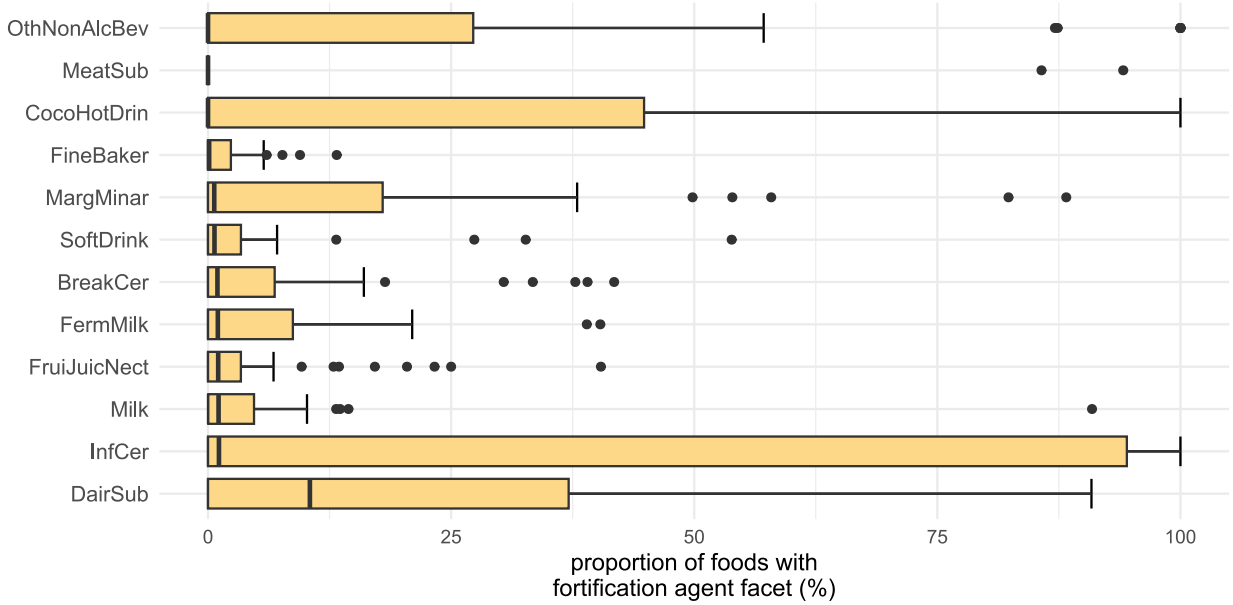

Figure 16. Summary of F09 (fortification agent) use in the most relevant food groups, calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Milk:** Milk; **FermMilk:** Fermented milk products; **MargMinar:** Margarines and minarines; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **OthNonAlcBev:** Other non-alcoholic beverages; **CocoHotDrin:** Cocoa and other hot drinks; **InfCer:** Processed cereal-based food for infants and young children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **SoftDrink:** Soft drinks and energy drinks

F10 refers to qualitative information of foods, namely related to the energy, fat, sugar content, and in general can be applicable to all foods. This facet is used more often than the previous one but also not in a harmonized way across surveys, as shown in the boxplot (Figure 17), for the food groups where it is more frequently used.

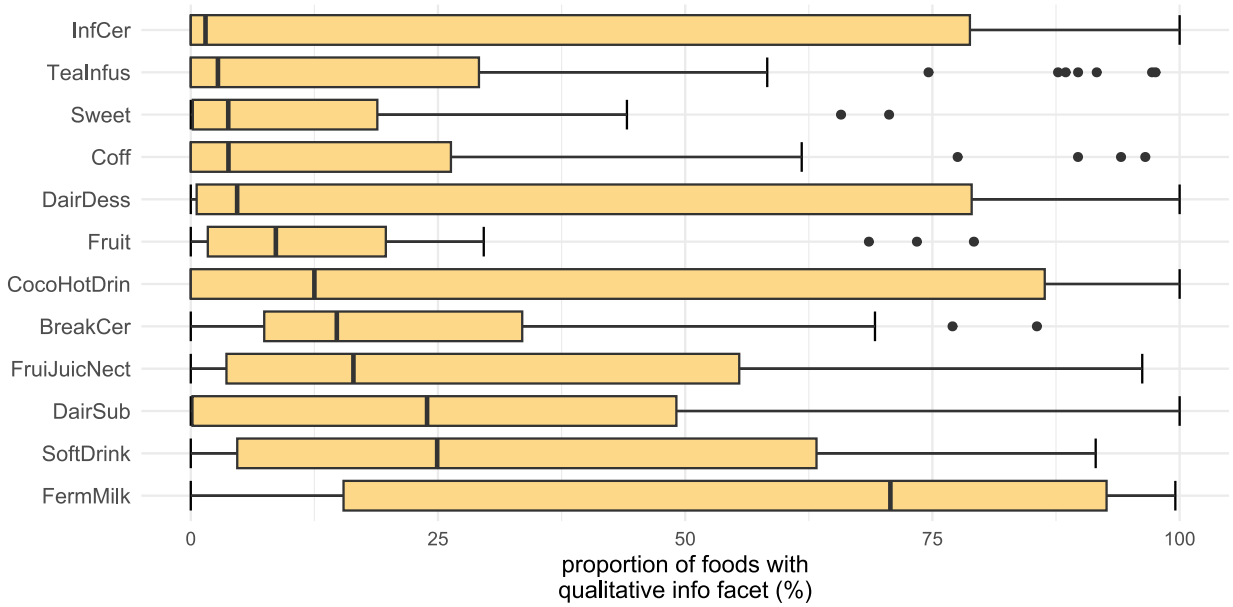

Figure 17. Summary of F10 (qualitative-info) use in the most relevant food groups, calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **BreakCer**: Breakfast cereals; **Fruit:** Fruit and fruit products; **FermMilk:** Fermented milk products; **DairDess:** Dairy dessert and similar; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **SoftDrink:** Soft drinks and energy drinks; **Coff:** Coffee ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfCer:** Processed cereal-based food for infants and young children; **DairSub:** Dairy imitates.

Packaging material facet (F19) is expected to be used in an extensive share of foods in the surveys as many foods consumed are packaged. Indeed, it is possible to see from the boxplot (Figure 18) that it is used in many food groups across surveys. Nonetheless, once again this is not homogeneous and even in food groups where we expected to have around 100% of report, such as fermented milks, dairy substitutes, soft drinks, among others, the median does not surpass the 80% (in the best case). For instance, 9 surveys do not use this facet in almost all food groups, as evident in the corresponding heatmap (Figure A 30).

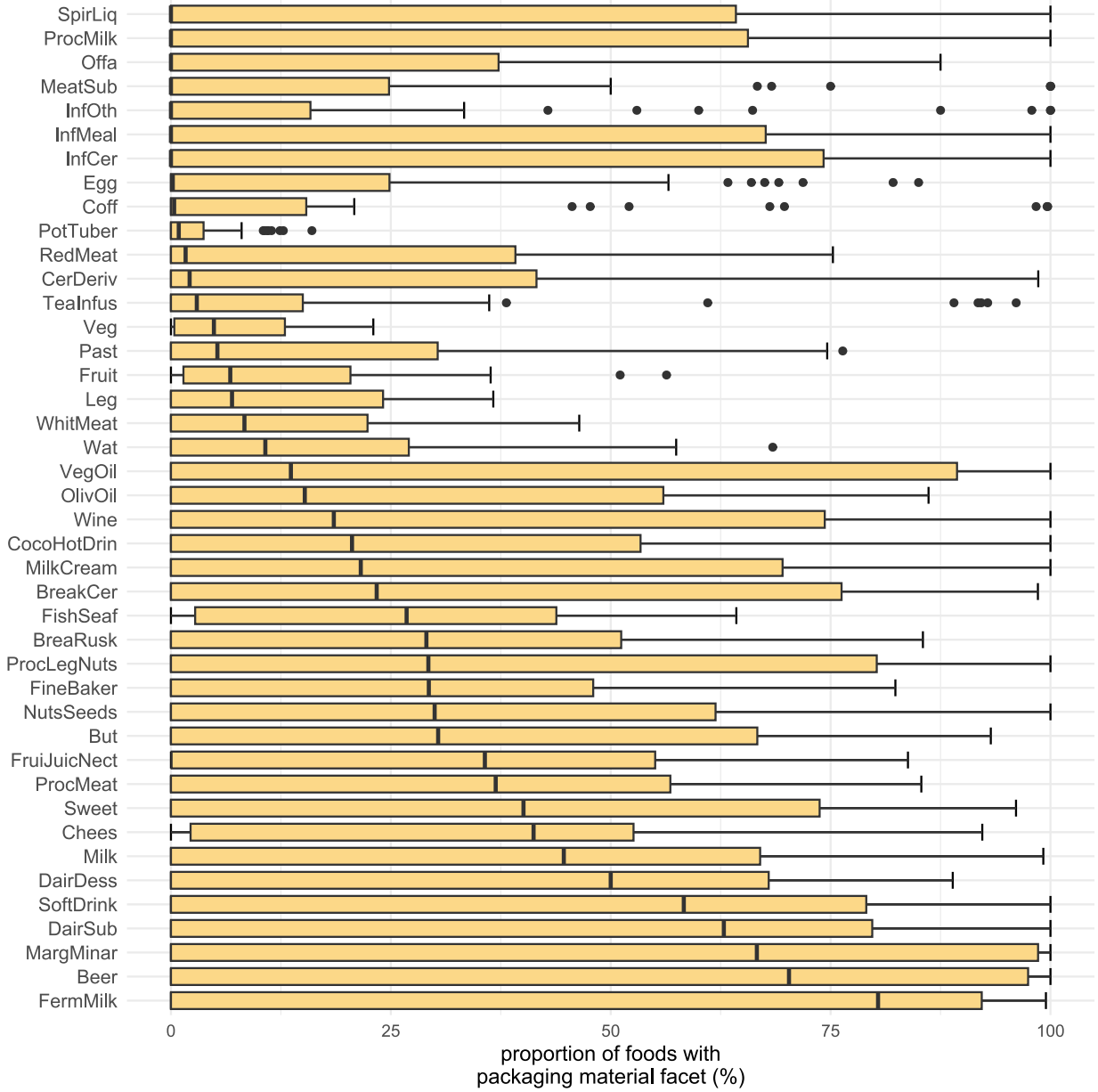

Figure 18. Summary of F19 (packaging material) use in the most relevant food groups, calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **Coff:** Coffee ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other

hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates.

F20 refers to the foods' part-consumed-analysed, which is relevant in some cases, to better specify the sub-part of a food that is referred to. Thus, it is not expected in a wide proportion of foods or in all food groups. It is mainly used in foods such as potatoes and tubers, fruits or vegetables, when it is specified whether the food was consumed with or without peel or husks or in animal foods, such as fish and seafood or meat and poultry (i.e., consumed with or without skin for example). Nevertheless, even when applicable, the use is very heterogeneous across different surveys, which is evident from the boxplot (Figure 19).

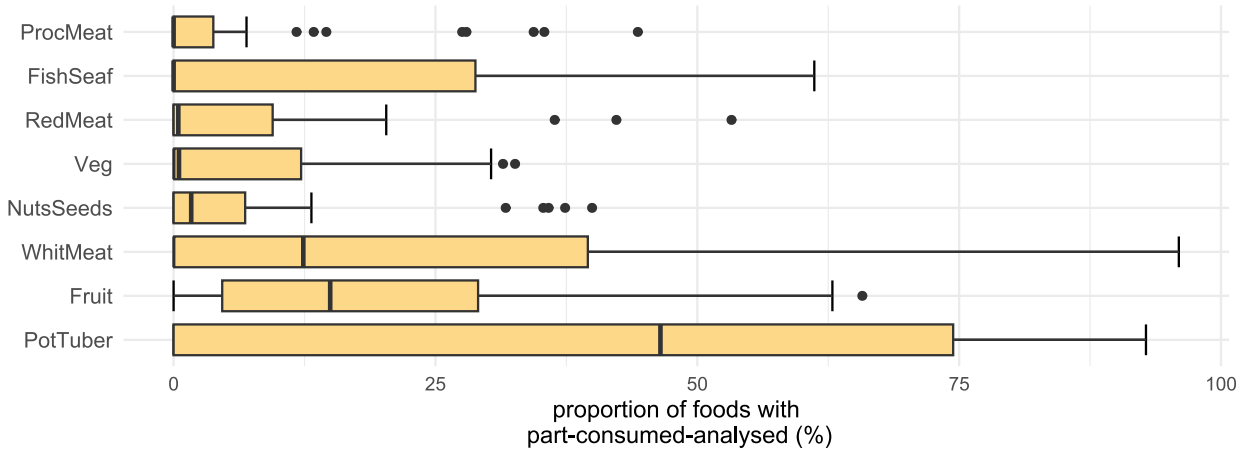

Figure 19. Summary of F20 (part-consumed-analysed) use in the most relevant food groups, calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **NutsSeeds:** Nuts, oilseeds and oilfruits; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates

Processing, F28, is the most added facet. In general, discarding the implicit F28, foods such as meat, fish, rice, pasta, potatoes, pulses, among other examples, are not consumed raw, implying the need of using a processing facet. Thus, a wide proportion of foods from these groups include this facet in most surveys. Figure 20 shows that meat, pasta, pulses, potatoes and tubers, and cereal derivatives (e.g., rice) are mostly reported with F28 (median >90% of foods). Still, even for these food groups, some surveys report close to 0% with processing facet. For the remaining food groups, the distribution of F28 use is wider, thus more heterogenous.

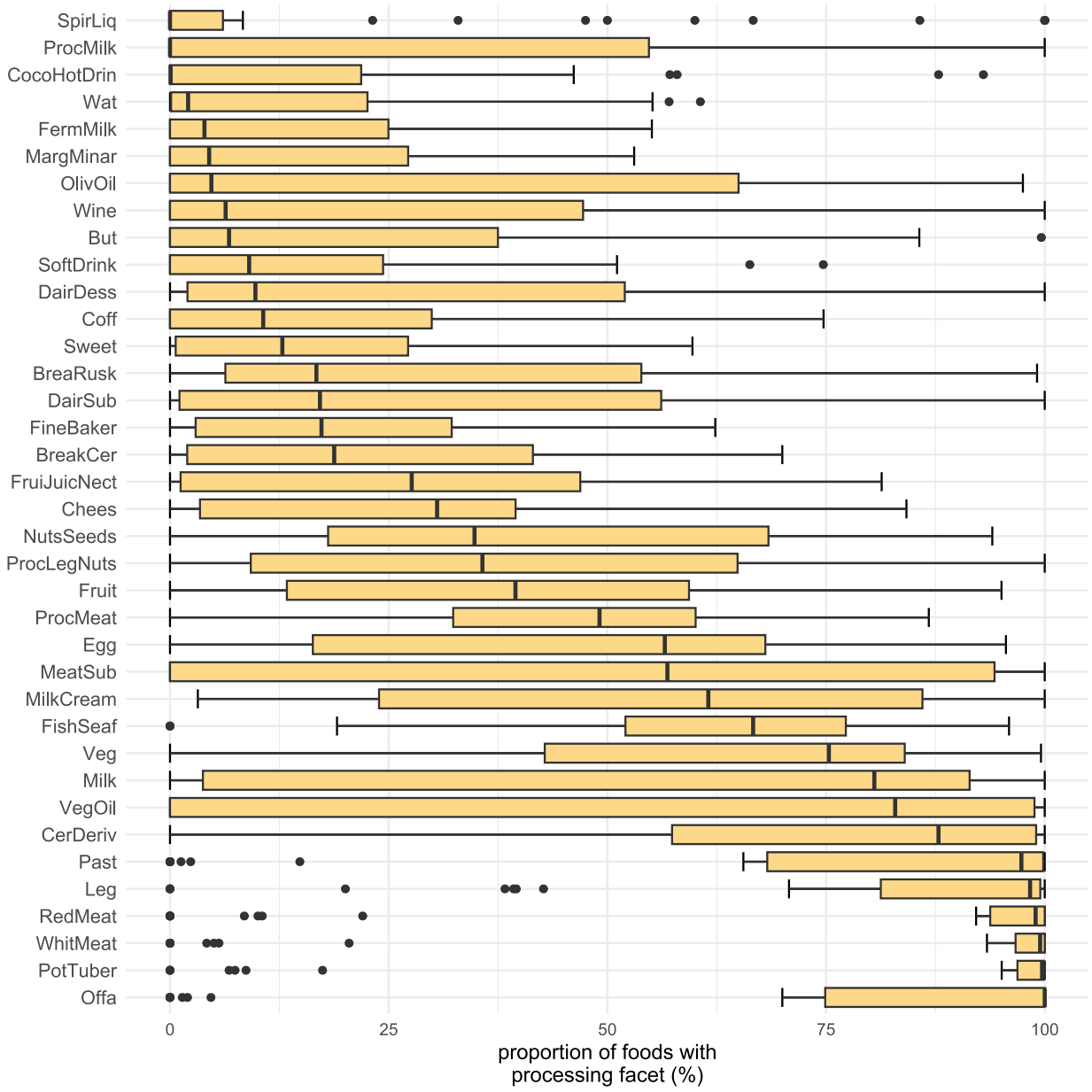

Figure 20. Summary of F28 (processing) use in the most relevant food groups, calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **Coff:** Coffee ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant

and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates.

Extra figures with results from indicators G2.3 – Total number of facets and G2.4 – Minimum recommended facets and heatmaps for facet use, can be consulted in Figure A 24, Figure A 25 and from Figure A 26 to Figure A 32 from Appendix A.

Foods reported with FoodEx2 codes from higher hierarchical levels are more specific. Thus, for more detailed and unambiguous information on food consumption, foods should be reported with codes above level 4 from the Foodex2 exposure hierarchy or above level 3 if accompanied by a facet. Indicator G2.5 assesses the use of specific levels of FoodEx2 classification (Figure 21), with most surveys presenting all food items coded ≥level 4 or ≥level 3 plus facet. Nonetheless, there are five surveys with more than 5% of foods reported with FoodEx2 codes from lower hierarchical levels. Some of these surveys are also the ones with smaller unique food reports (Indicator G1.1, Appendix A, Figure A 22) and with higher proportion of foods without facets (Indicator G1.4, Figure 13).

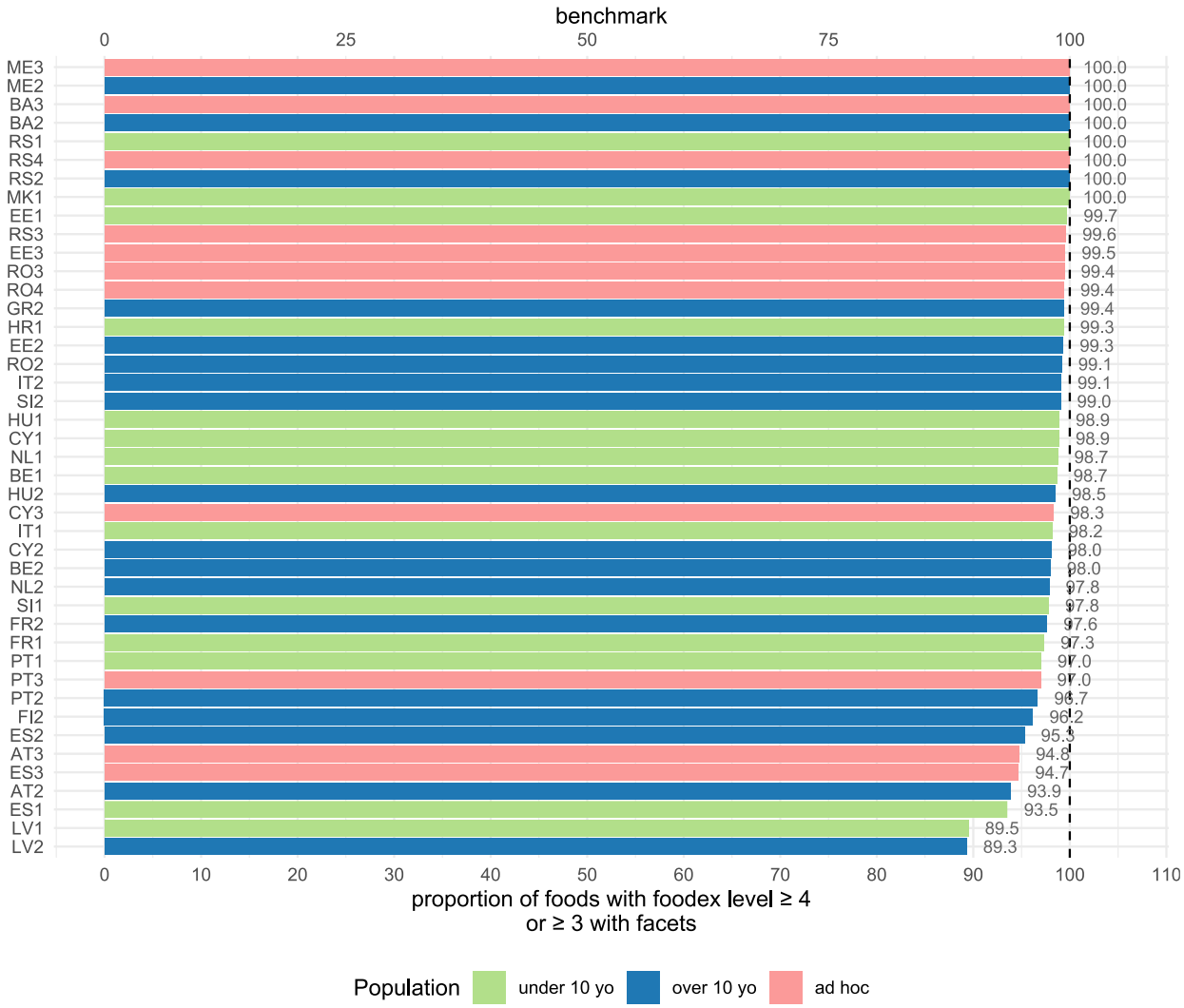

Figure 21. Indicator G2.5 – Prevalence of foods classified as level 4 (or above) in the FoodEx2 hierarchy or level 3 plus facet. This indicator reflects higher specificity in the FoodEx2 codes selection, according to the exposure hierarchy.

In this dimension, the indicator G2.1.1 – Percentage of composite dishes in Foodex2 codes was also evaluated, and the results are presented in Figure A 23 from Appendix A. This indicator represents the percentage of food items in the consumption datasets classified with a FoodEx2 code from the Composite Dishes node of the FoodEx2 Exposure hierarchy. In summary, because the EU Menu methodology implies the disaggregation of recipes into ingredients, the percentage of composite dishes reported as single food items is very low in all surveys (max 2%). Even though the EU Menu methodology requires the disaggregation of recipes, it would be important to include a variable with the recipe codes to identify foods consumed together as part of a recipe (Indicator G2.1.2). There is a variable in the databases that could allow to identify this, but it is not harmonized across surveys. Thus, it was not possible to assess Indicator G2.1.2 in a structured manner.

3.1.8 **Dimension H** – Data analyses

Indicators from dimension H measure several quality aspects of data analyses. Results from the categorical indicators extracted from survey reports for dimension H can be consulted from the corresponding Table A 2 and heatmap (Figure A 33) from Appendix A. In summary, in general surveys do not use biomarkers to validate dietary intake. Only a small proportion (14%) report using biomarkers in subsamples. The Goldberg method (Goldberg et al., 1991), updated by Black (Black, 2000), used to identify misreporting of energy intake, is reported in 16 survey units, but many surveys do not report this information at all, although it can be considered in subsequent analyses of the survey.

Regarding outliers in energy intake, two cut-offs were considered depending on the age groups included in the surveys: 3500 kcal for ≥10 years and ad-hoc (Willett, 2012) and 2400 kcal for <10 years (kcal cut-off defined to match the quantile of 3500 kcal in the ≥10 years energy intake distribution), as presented in Figure 22. The meta-average obtained from this indicator was around 2% of energy intake outliers for both age groups. Surveys with more outliers in both <10 years and ≥10 years tend to belong to the same countries.

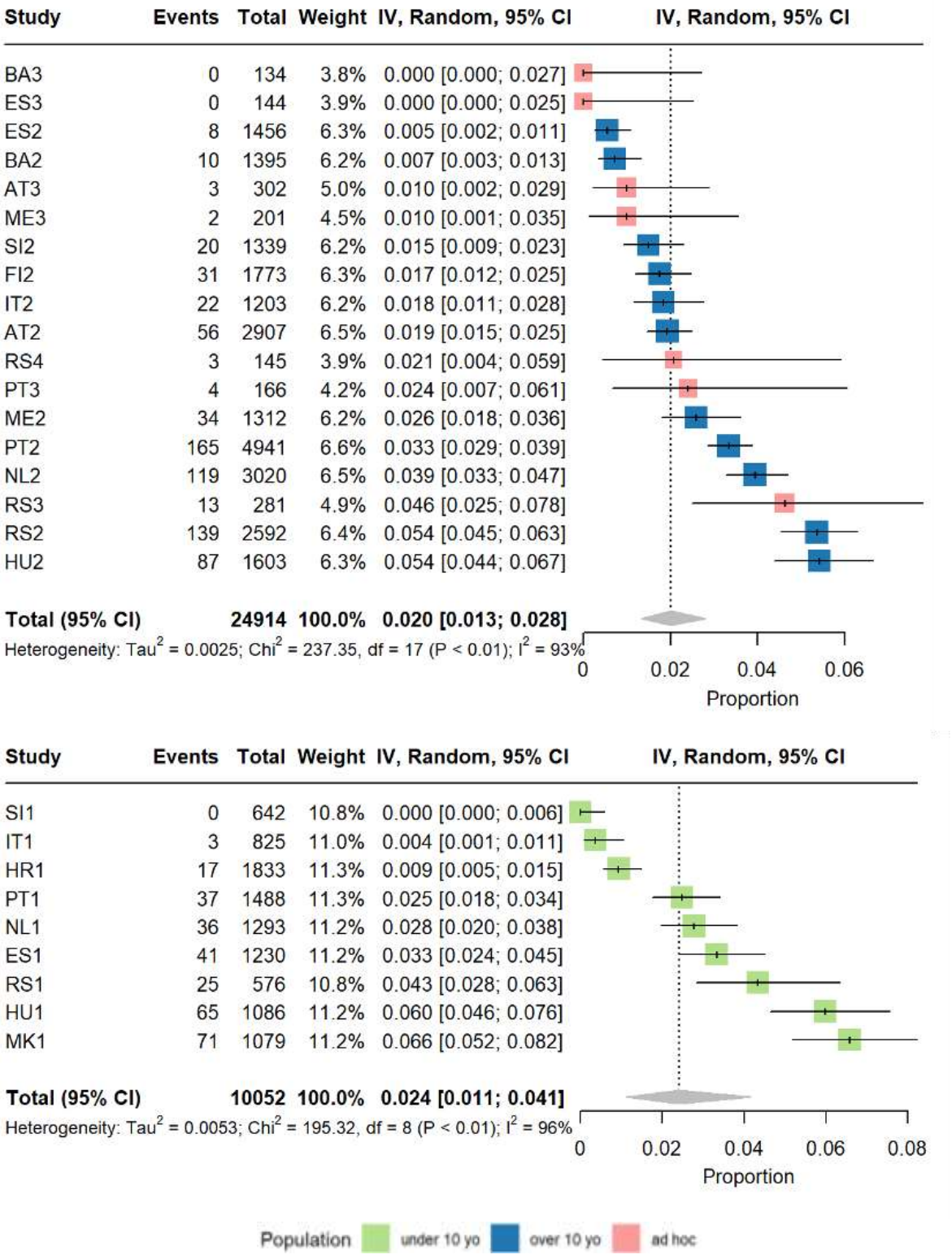

Figure 22. Indicator H1.1 – Energy intake outliers, above 3500 kcal for surveys ≥10 years and ad-hoc, or above 2400 kcal for surveys <10 years.

Results from other indicators evaluating outliers (energy<500 kcal and food groups), as well indicators evaluating the prevalence of incomplete interviews, the average number of food items reported per interview or the digit preference in food amounts can be consulted in Figures Figure A 34 to Figure A 39 from the Appendix A.

Indicator H2.1 evaluates the intraclass correlation coefficient (ICC) for food groups and nutrients represents the proportion of total variance explained by differences between individuals. Thus, it measures the reliability of food groups' consumption and nutrient intake estimates, and it is one of the most valuable quality indicators evaluated. According to our results, some food groups, such as pulses, fish or offal, present ICC lower than 0.2, compromising the reliability of estimating these foods using only two reporting days (as it happens in 2x24hR). In contrast, food groups more frequently consumed, such as coffee, milk, bread, fruit and vegetables are measured more consistently in all surveys, with two days of report. As expected, the infant food groups (infant cereals, formulas) are more reliable among children surveys, whereas alcoholic beverages are more reliable among adult surveys. However, in the case of alcoholic beverages type, there are some specificities depending on the country's consumption pattern. Alcohol is more reliably measured in adult surveys. All the macronutrients, water and energy are reliably measured in surveys, whenever available. Figure 23 and Figure 24 summarize the information for this indicator illustrating some examples of food groups and nutrients. Extended information on this indicator can be found in Figure A 40 and Figure A 41 from Appendix A.

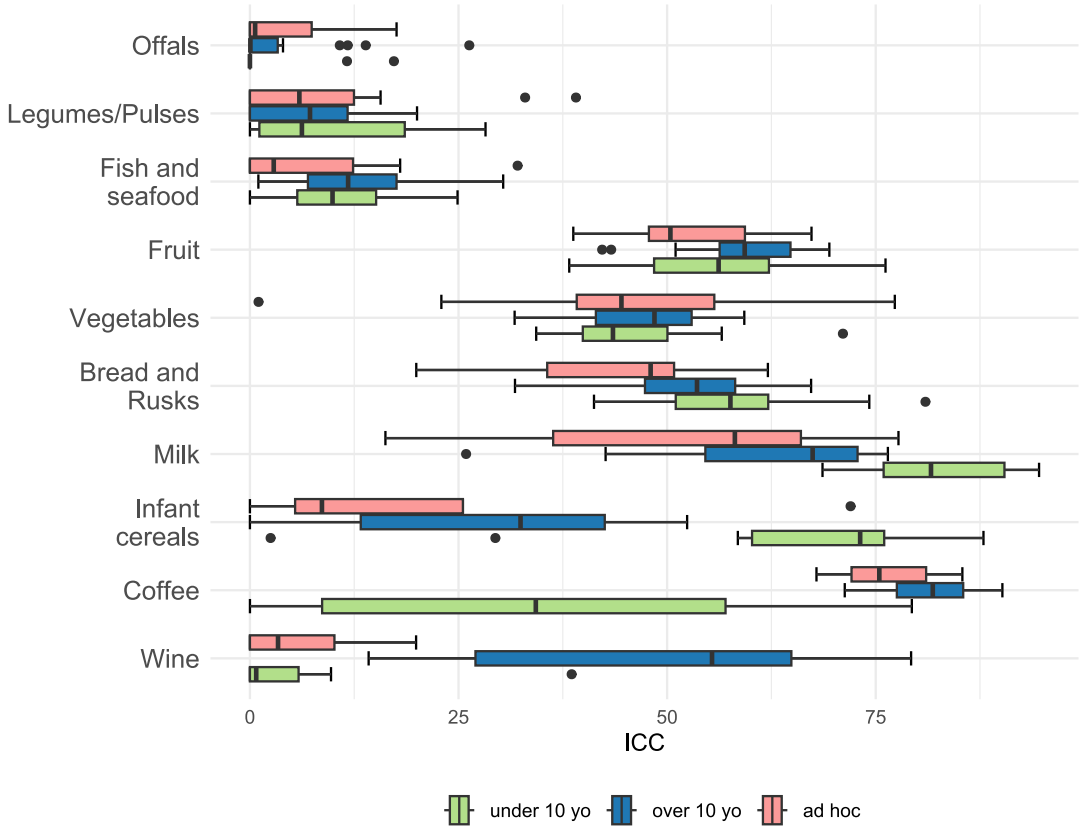

Figure 23. Summary of indicator H2.1 Proportion of total variance explained by differences between individuals, for some food groups.

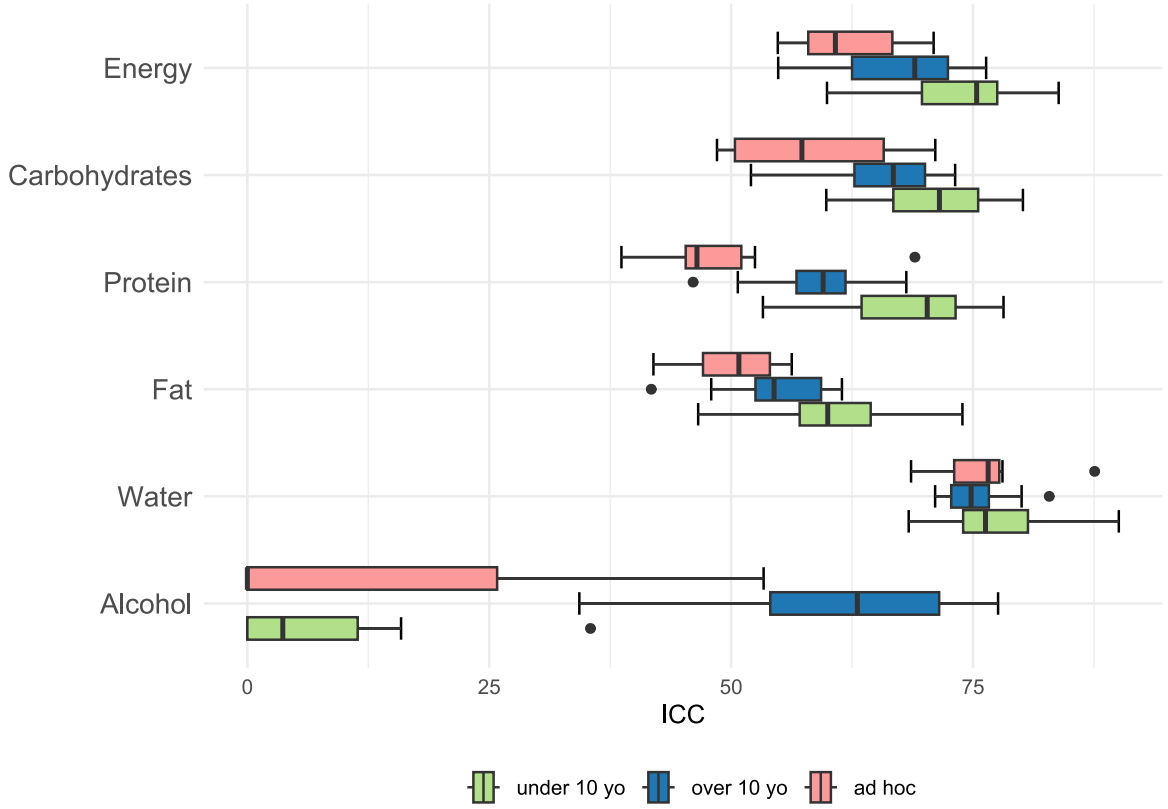

Figure 24. Summary of indicator H2.1 Proportion of total variance explained by differences between individuals, for energy, macronutrients, water and alcohol.

The meta-average of the prevalence of plausible energy reporters among participants aged over 10 years old was 80.8% varying between 65.9% and 91.8%. This indicator was computed only for surveys ≥10 years with available estimates for energy intake. The estimation of energy misreport was based on the EFSA Guidance 2014 protocol for misreport estimation along with the respective average age-specific physical activity levels (PAL) values (low activity level was considered) and Coefficients of variation (CV). It's important to note that the results generated through this approach may diverge from a given country's own estimates regarding the prevalence of misreporting. Such discrepancies may arise due to variations in the input data used for the calculations.

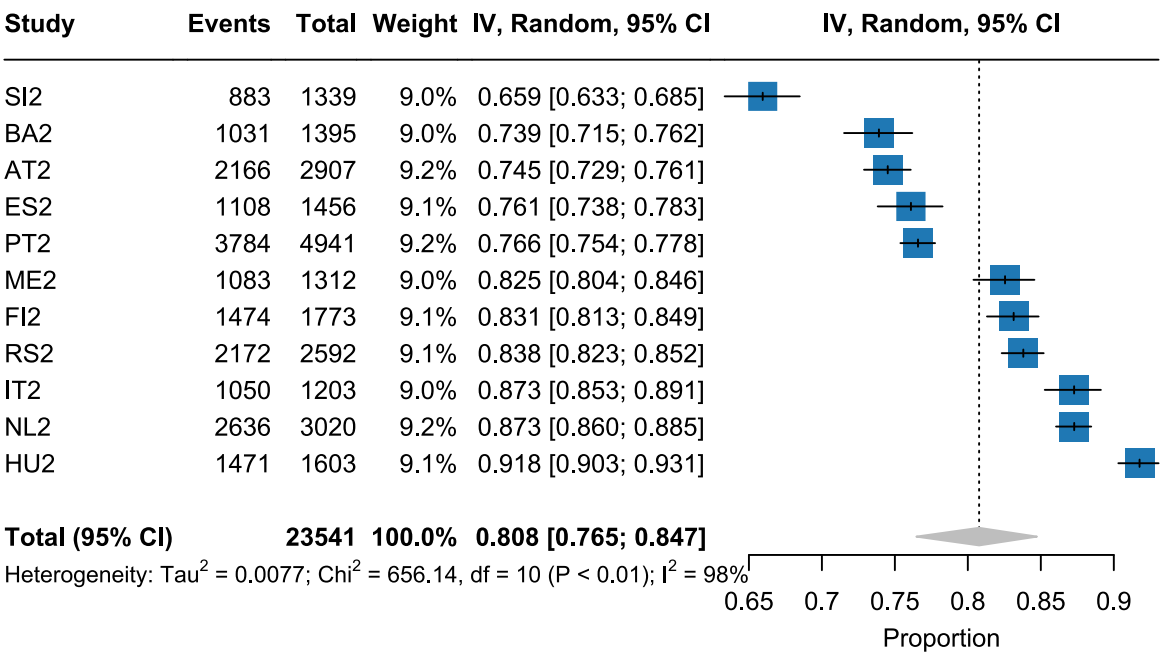

Figure 25. Indicator H4.1.2 Proportion of plausible reporters of energy intake. This indicator was computed only for surveys ≥10 years with available estimates for energy intake.

3.1.9 **Dimension I** – Results reporting

The EFSA survey reports mainly describe the survey protocols and methodological considerations, neglecting the reporting of results in many cases, which are presented in extra reports, most of them in the countries own language. Consequently, because only the EFSA reports were used in this assessment, a large proportion of information for the indicators from this dimension is lacking, compromising its quality assessment, as presented in the Table 10 and heatmap (Appendix A, Figure A 42).

This lack of information in EFSA survey reports could be overcome in the future with a better harmonized guidance focused also in results's reporting.

Table 10. Summary statistics for categorical indicators of Dimension I – Results reporting.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
I1.1	Weighted results to ensure the representativeness	no	3	21.4	3	16.7	0	0.0
		not applicable	0	0.0	0	0.0	11	100.0
		yes	8	57.1	11	61.1	0	0.0
		missing	*3*	*21.4*	*4*	*22.2*	*0*	*0.0*

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
I2.1	Adjustment of nutritional intake for the intra-individual variability	no	2	14.3	1	5.6	0	0.0
		yes	4	28.6	5	27.8	2	18.2
		missing	*8*	*57.1*	*12*	*66.7*	*9*	*81.8*
I2.2	Usual intake estimated using the food propensity/frequency questionnaire	no	3	21.4	3	16.7	4	36.4
		not applicable	2	14.3	1	5.6	0	0.0
		yes	1	7.1	3	16.7	1	9.1
		missing	*8*	*57.1*	*11*	*61.1*	*6*	*54.5*
I3.1	Sensitivity analysis excluding misreporters of energy intake	no	3	21.4	2	11.1	2	18.2
		yes	3	21.4	6	33.3	4	36.4
		missing	*8*	*57.1*	*10*	*55.6*	*5*	*45.5*

3.2 Summary quality scores and associated factors

3.2.1 Summary quality scores

Summary quality scores according to the different survey dimensions were created using Principal Component Analysis (PCA). The scores help to correlate different survey quality aspects more concisely given the extensive list of individual quality indicators assessed. Each dimension was converted into one or two summary quality scores, resulting from the sum of the multiple indicators within the dimension, after standardization and according to the respective weights, that correspond to the factor loadings from the PCA, presented in

Table 11. Higher indicator factor loadings (especially >0.4) correspond to higher weights for the indicator within the dimension score according to this data-driven approach, also indicating higher correlation with the remaining indicators from the same dimension. The proportion of variance of the components ranged from 17.9% to 49.0%. As mentioned in the methodology section, indicators from dimension B were not summarized using the PCA and only the reported response rate was considered. Moreover, all indicators were converted to reflect higher quality (e.g. Indicator on prevalence of missing values was inverted and it is now reflecting lower prevalence of missings).

An interpretation for each summary score obtained is presented in the paragraphs below.

- Dimension A - *Sampling Plan* covers 11 indicators and only one presented a factor loading <0.4 (*A2.2 Sampling by waves*). Thus, **higher Dimension A score** reflects (i) the preferred use of **national registers** with appropriate **coverage of the target population** as sampling frame, (ii) an adequate (**probabilistic**) **sampling design with stratification** for relevant variables, (iii) achieving a sufficient sample size to obtain **low relative standard errors** for key variables and iv) a higher sample **representativeness** of the target population.

- Dimension C - *Training and supervision of interviewers* includes 11 indicators and five of them presented factor loadings <0.4 (*C1.1 Background in Nutrition/Dietetics and/or interviewing experience*, *C2.2.1 Training: conducted at regular basis, C3.2.1 Observer bias: explained variance (%) for key variables compared between interviewers during the fieldwork, C3.2.2 Observer bias: proportion of item non-response, by interviewer, continuously monitored and C4.1.1 Pilot Study conducted*). Briefly, **higher Dimension C score indicates** (i) **good training procedures** (such as appliance of standard operating procedures and higher training duration); and (ii) an **effective monitoring of interviewers** through a dynamic strategy between coordinators and the remaining staff to address fieldwork constraints and reporting solutions during fieldwork.

- Dimension D - *Interview administration procedures* covers 9 indicators and two were less meaningful to the overall score (*D1.1 At least one face-to-face interview, D2.2.1 Proportion of participants with interview gap 8-15*). **Higher Dimension D score denotes** (i) a more **uniform distribution** of dietary interviews per **weekdays and seasons**, (ii) a higher proportion of **complete participants** (≥2 interviews), (iii) an **interview duration closer to the benchmark** (30 min for 24hR and 50 min for total interview) and (iv) definition of the **interview setting according to the participants' preference**. On the contrary, having the same mode of administration to all participants at each interview moment was negatively associated with the dimension (negative factor loading), supporting that having higher flexibility in the interview administration mode can improve the other indicators from this dimension.

- Dimension E - *Software tools and validation procedures* comprises 17 indicators of which eight (*E1.4.2 Automatic probing questions for easily forgotten foods, E2.1.2 If the picture book has been updated, has it been validated again?, E2.3 total number of*

picture series in the picture book, E2.5 Availability of default quantities as a quantification method, E3.2.2 FPQ: includes foods with higher contamination potential, E3.2.3 FPQ: includes dietary supplements and E3.2.4 FPQ: designed to capture seasonal variation of foods) presented lower factor loadings below 0.4. **Higher E score** implies (i) using a **proficient software tool** that is validated, uses the multipass method and has adequate quality control procedures implemented, such as several automatic checking for empty and implausible values; (ii) have adequate and multiple **quantification methods available** (namely validated picture book and default quantities); and apply a **FPQ** that includes **less frequently eaten foods**.

- Dimension F - *Non-dietary data collection* includes 11 indicators and 5 presented factor loadings <0.4 (*F1.1.3 Pregnant: Self-reported weight before pregnancy been collected, F1.2.1 & F1.2.2 Standardized procedures for anthropometric measurements, F1.2.4 Regular training for interviewers for anthropometric measurement, F1.2.5 Regular check and calibration of equipment for anthropometric measurements and F2.2 Collection of accurate physical activity measurements (e.g. accelerometers*)) for score F. Accordingly, **higher score F entails** (i) **higher proportion of objective measurements of anthropometric variables** and consequently **lower digit preference**; as well as (ii) **physical activity assessment** through the IPAQ (adults) or other reporting methods (children). The indicator *F1.2.3 Report of deviations from the protocol for measuring weight and height* was negatively associated with the remaining indicators from this dimension, suggesting that a higher proportion of measured anthropometrics implies less deviations from the protocol.

- Dimension G - *Data completeness* was summarized into two scores, named as G_a *Food & Recipe description* and G_b *Missing data.* From the 13 original indicators included, 5 did not reach a meaningful factor loading in any component (*G1.1 Number of food items reported; G1.2.1 Availability of a food supplements database; G1.3.1 Availability of a recipe composition database, G2.1.1 Percentage of composite dishes; G2.2 Unclassified values in variables from the Consumption dataset*). **Higher G_a score** indicates better **Food & Recipe description**, namely due to a more specific **FoodEx2 classification** use (higher levels of the hierarchy), **improved facet** use and **updated recipe database**. **Higher G_b score** represents **lower prevalence of missing values** in non-mandatory variables (from the Foods and Subjects datasets), but it is also associated with an adequate quantification of food supplements.

- Dimension H - *Data analyses* was also summarized in two scores, named as H_a *Data reliability and* H_b *Outliers.* Three out of the 10 indicators proposed for dimension H presented factor loadings below the cutoff (*H1.1 Energy outliers <500 kcal, H1.5 Digit preference in food amounts and H4.1.1* Calculation of misreporting of energy intake through recommended methods). **Higher H_a score** denotes (i) **higher ICC for foods and nutrients**, (ii) **higher proportion of complete interviews**, and (iii) **lower number of food items reported per interview** (Indicator H1.4). **Higher H_b score** indicates (i) **lower proportion of positive energy outliers** (>3500 adults/2400 children) and (ii) **lower levels of food group outliers**. Indicator *H3.1 Availability of data, such as biomarkers, to validate dietary intake* was unexpectedly negatively

associated with H_b score. However, the value of this indicators results is questionable as only very few surveys report it, and only in subsamples and for specific nutrients.

- Dimension I - *Results reporting* score reflects the performance of all 4 indicators included. **Higher I score** implies (i) the presentation of **results with appropriate weighing** to ensure the representativeness, (ii) estimating the **usual intake** of foods and nutrients by adjusting for the intra-individual variability and using the information from the FPQ for less frequently eaten foods and (iii) conducting **sensitivity analysis excluding misreporters** of energy intake. This information, however, was difficult to ascertain from the survey reports, which mainly focused on methodological procedures rather than results presentation.

Table 11. Factor loadings for each quality indicator, obtained from the Principal Component Analysis (PCA). In bold, are highlighted the factor loadings >0.4, more meaningful to the dimension summary score.

Quality indicators		Factor loadings
Sampling plan: methods and coverage		**A**
A1.1.1	Sampling frame - National population register (updated and accessible)	**0.41**
A1.2.1	Sampling frame covering the entire target population	**0.63**
A2.1.1	Probabilistic sampling design	**0.96**
A2.1.2	Sampling stratified by age classes, sex, and/or other important characteristics	**0.79**
A2.2	Sampling by waves	-0.25
A3.1.2	Minimum target sample size defined by EFSA achieved for all age groups	**0.81**
A4.1 BMI	(Lower) Relative standard error for key estimates – BMI	**0.87**
A4.1 EI	(Lower) Relative standard error for key estimates – Energy intake	**0.50**
A4.1 FG	(Lower) Relative standard error for key estimates – Food groups	**0.93**
A5.1	Study sample with similar distribution as target population	**0.76**
A5.2	Weighting procedures applied	**0.47**
Proportion of Variance		*49.9%*
Training and supervision of interviewers		**C**
C1.1	Background in Nutrition/Dietetics and/or interviewing experience	0.28
C2.1.1	Training: standard operating procedures (SOP) in place	**0.87**
C2.1.2	Training: conducted according to SOP	**0.87**
C2.1.3	Training: conducted during the pilot	**0.49**
C2.2.1	Training: conducted at regular basis	0.22
C2.2.2	Training duration of interviewers per survey	**0.48**
C3.1.4	Interviewer monitoring: Dynamic strategy to address issues emerging in the field	**0.73**
C3.1.5	Interviewer monitoring: coordinator timely report solutions to all staff	**0.47**
C3.2.1	Observer bias: explained variance (%) for key variables compared between interviewers during the fieldwork	0.08
C3.2.2	Observer bias: proportion of item non-response, by interviewer, been continuously monitored?	-0.37
C4.1.1	Pilot study conducted	0.21
Proportion of Variance		*27.7%*
Data collection: Interview administration procedures		**D**
D1.1	At least one face-to-face interview	-0.26

Quality indicators		Factor loadings
D1.2.1	Same mode of administration to all participants at each interview moment	**-0.70**
D2.1.1	Interview setting according to participants preference	**0.45**
D2.2.0	Proportion of complete participants (≥2 interviews) in the databases	**0.68**
D2.2.1	Proportion of participants with interview gap 8-15 days	0.18
D2.3.1	Distribution of dietary data by weekdays	**0.71**
D2.3.2	Distribution of dietary data by season	**0.73**
D2.4.1	Duration of 24-h recall	**0.71**
D2.4.2	Total interview duration	**0.71**
Proportion of Variance		*37.0%*
Data processing: software tools and validation procedures		**E**
E1.1	Use of a validated dietary assessment tool	**0.81**
E1.2	Use of the multi-pass method (or similar)	**0.52**
E1.3	Monitoring of the interview time	**0.42**
E1.4.1.1	Automatic checking for empty food consumption occasions	**0.53**
E1.4.1.2	Automatic checking for minimum/maximum accepted quantities per food type	**0.66**
E1.4.1.3	Automatic checking for missing quantities	**0.61**
E1.4.2	Automatic probing questions for easily forgotten foods	0.35
E1.4.3	Calculation of energy and macronutrient intake and identification of outliers at the end of interviews	**0.82**
E2.1.1	Use of a validated food picture book	**0.57**
E2.1.2	If the picture book has been updated, has it been validated again?	0.08
E2.2	Minimum number of pictures per photo series in picture book (≥4)	**0.40**
E2.3	Total number of picture series in the picture book	-0.31
E2.4	Availability of food standard units as a quantification method	0.21
E2.5	Availability of default quantities as a quantification method	**0.57**
E3.1.2	Self-administered FPQ/FFQ subsequently checked for completeness	0.26
E3.2.1	FPQ: includes less frequently eaten foods	**0.45**
E3.2.2	FPQ: includes foods w/ higher contamination potential	-0.09
E3.2.3	FPQ: includes dietary supplements	0.36
E3.2.4	FPQ: designed to capture seasonal variation of foods	0.38
Proportion of Variance		*24.0%*
Non-dietary data collection		**F**
F1.1 Height	Percentage of individuals with anthropometric measurements - Height	**0.83**

Quality indicators		Factor loadings
F1.1 Weight	Percentage of individuals with anthropometric measurements - Weight	**0.94**
F1.1.3	Pregnant - Self-reported weight before pregnancy been collected	0.35
F1.2.1 & F1.2.2	Standardized procedures for anthropometric measurements	-0.23
F1.2.3	Report of deviations from the protocol for measuring weight and height	**-0.64**
F1.2.4	Regular training for interviewers for anthropometric measurement	-0.14
F1.2.5	Regular check and calibration of equipment for anthropometric measurements	-0.39
F1.4 Height	Lower levels of digit preference in anthropometric measurements - Height	**0.78**
F1.4 Weight	Lower levels of digit preference in anthropometric measurements - Weight	**0.82**
F2.1	Physical activity assessment - IPAQ adults & reporting method children	**0.56**
F2.2	Collection of accurate physical activity measurements (e.g. accelerometers)	0.26
Proportion of Variance		*36.5%*

Data completeness		**G_a** *Food & Recipe description*	**G_b** *Missing Data*
G1.1	Number of food items reported	0.36	-0.08
G1.2.1	Availability of a food supplements database	-0.14	0.33
G1.2.2	Food supplements quantified in grams	0.24	**0.77**
G1.3.1	Availability of a recipe composition database	0.11	0.07
G1.3.2	Recipe database updated during fieldwork	**0.70**	0.16
G1.4	Lower prevalence of food items with only FoodEx2 base term	**0.77**	-0.24
G1.5	Lower prevalence of missing values in energy and macronutrients per 100g of food	0.04	**0.90**
G2.1.1	Lower percentage of composite dishes in Foodex2 codes	0.29	-0.34
G2.2	Lower prevalence of unclassified values in variables from the consumption dataset	0.11	0.32
G2.3	Total number of facets	**0.52**	0.27
G2.4	Minimum recommended facets	**0.80**	-0.16
G2.5	Higher prevalence of foods classified as level 4 (or above) in the FoodEx2 hierarchy or level 3 plus facet	**0.80**	-0.24
G3.1	Lower prevalence of missing values in non-mandatory variables: Subjects dataset	0.12	**0.70**
Proportion of Variance		*21.5%*	*18.5%*

Quality indicators		Factor loadings	
Data analyses		**H_a** *Data reliability*	**H_b** *Outliers*
H1.1	Lower prevalence of Energy outliers (>3500 kcal)	-0.33	**0.90**
	Lower prevalence of Energy outliers (<500 kcal)	0.17	-0.13
H1.2	Lower prevalence of Food groups outliers	0.07	**0.84**
H1.3	Higher proportion of complete interviews (≥2 main meals)	**0.46**	-0.19
H1.4	Number of food items per interview	**-0.61**	-0.50
H1.5	Lower levels of digit preference in food amounts	-0.13	-0.02
H2.1 FG	Proportion of total variance explained by differences between individuals (ICC) – Food groups	**0.51**	0.43
H2.1 Nuts	Proportion of total variance explained by differences between individuals (ICC) – Nutrients	**0.81**	0.32
H3.1	Availability of data, such as biomarkers, to validate dietary intake	0.55	**-0.61**
H4.1.1	Calculation of misreporting of energy intake through recommended methods	-0.31	0.37
Proportion of Variance		*23.2%*	*27.2%*
Results reporting		**I**	
I1.1	Weighted results to ensure the representativeness	**0.58**	
I2.1	Adjustment of nutritional intake for the intra-individual variability	**0.73**	
I2.2	Usual intake estimated using the food propensity/frequency questionnaire	**0.76**	
I3.1	Sensitivity analysis excluding misreporters of energy intake	**0.60**	
Proportion of Variance		*45.0%*	

3.2.2 Associations between summary quality scores

After creating the dimension summary scores, the Pearson correlations between them were estimated and are graphically represented in Figure 26. From the results obtained, some moderate-to-strong correlations stand out, which will be described in the following paragraphs, along with a breakdown of the indicator-indicator correlations (ρ>0.3, p-value<0.05) that likely helps to explain the observed results. Moreover, for some specific individual indicators, a separate analysis is presented to graphically assess its impact on the dimensions' summary scores, using boxplots.

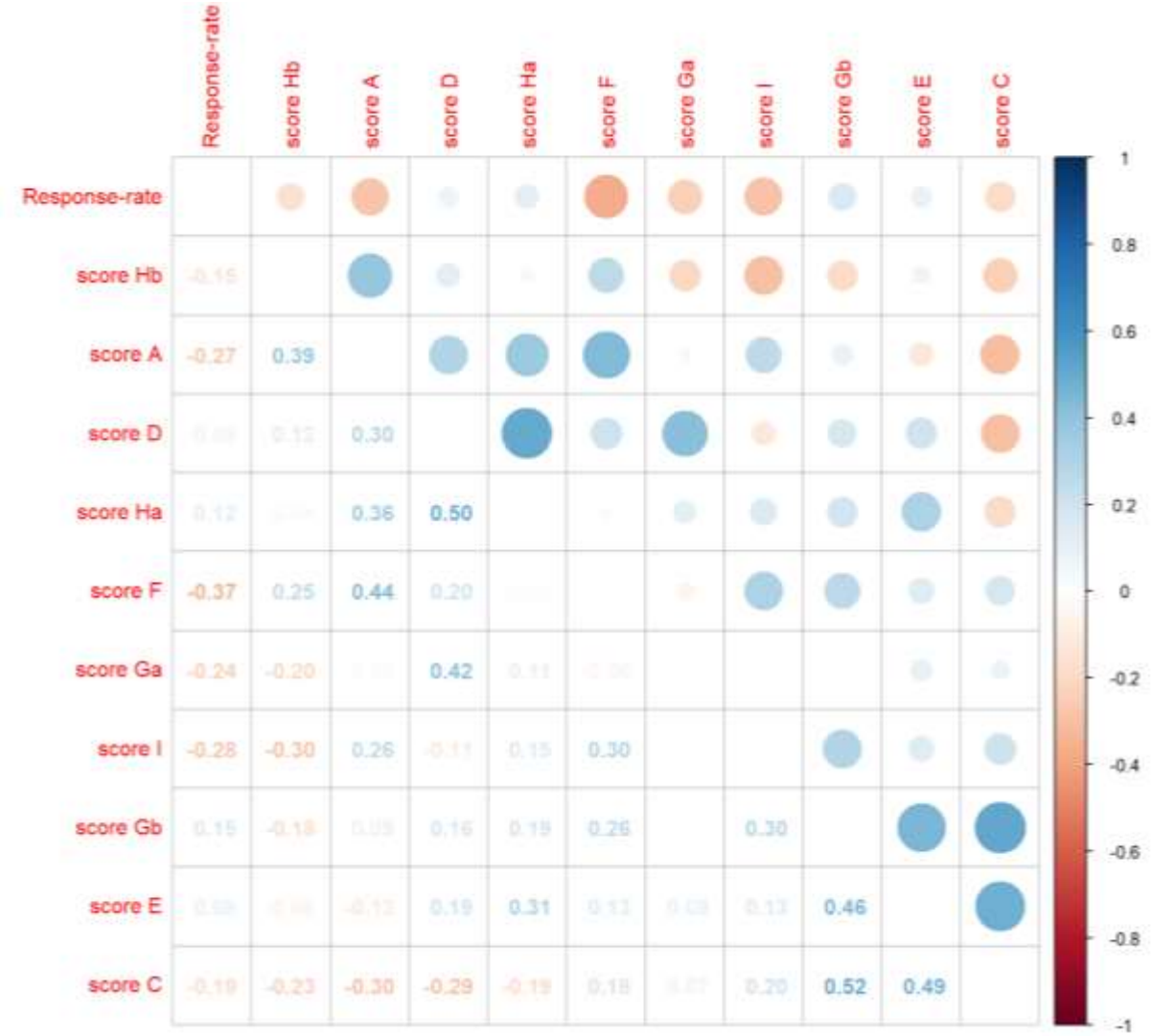

Figure 26. Pearson correlation coefficients between dimension summary scores. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting

Dimensions A. *Sampling plan: methods and coverage* with H_a *Data reliability* (r=0.36, p-value=0.04) and H_b *Outliers* (r=0.39, p-value=0.03)

Having a sampling frame covering the entire target population (A1.2.1), a probabilistic sampling design (A2.1.1), lower RSE for key estimates (A4.1), higher representativeness: study sample with similar distribution as target population (A5.1) and applying weighting procedures (A5.2) are significantly associated with higher ICC values for Food groups and *Nutrients (Indicator H2.1)*.

Moreover, a stratified sampling design by sex, age and other characteristics was associated with a lower proportion of outliers in energy and food groups.

Dimensions A. *Sampling plan: methods and coverage* and F. *Non-dietary data collection* (r=0.44, p-value=0.01)

Surveys with higher A score, characterized by better sampling plan, also presented higher F score, reflecting higher percentage of measured anthropometrics and assessment of physical activity. In general, ad-hoc surveys performed worse in A score, due to the convenience sampling and lack of sample representativeness and presented in many cases lower levels of measured anthropometrics and no physical activity assessment. This might be the most likely reason for this observed moderate correlation.

Dimensions C. Training and supervision of interviewers and E. *Software tools and validation procedures* (r=0.59, p-value=0.005)

Surveys with higher C score, on training and supervision of interviewers also presented higher E score, reflecting the quality of software tools and reported FPQ. Several indicators within these dimensions were moderately-to-strongly associated, and in general this finding suggests that surveys that had a better planning phase (training of interviewers, adequate survey monitoring) also invested in appropriate and validated software, with several automatic validation procedures (probing questions, outlier detection, etc).

Dimensions C. Training and supervision of interviewers and G_b *Missing data* (r=0.52, p-value=0.003)

Surveys with higher C score also tend to have higher G_b score, characterized by lower proportion of missing data. Looking into indicator-indicator correlations, it is possible to observe that training during the pilot and training at a regular basis was associated with lower proportion of missing values.
Longer training duration by itself was not significantly associated with overall dimensions quality, as presented in Figure 27.

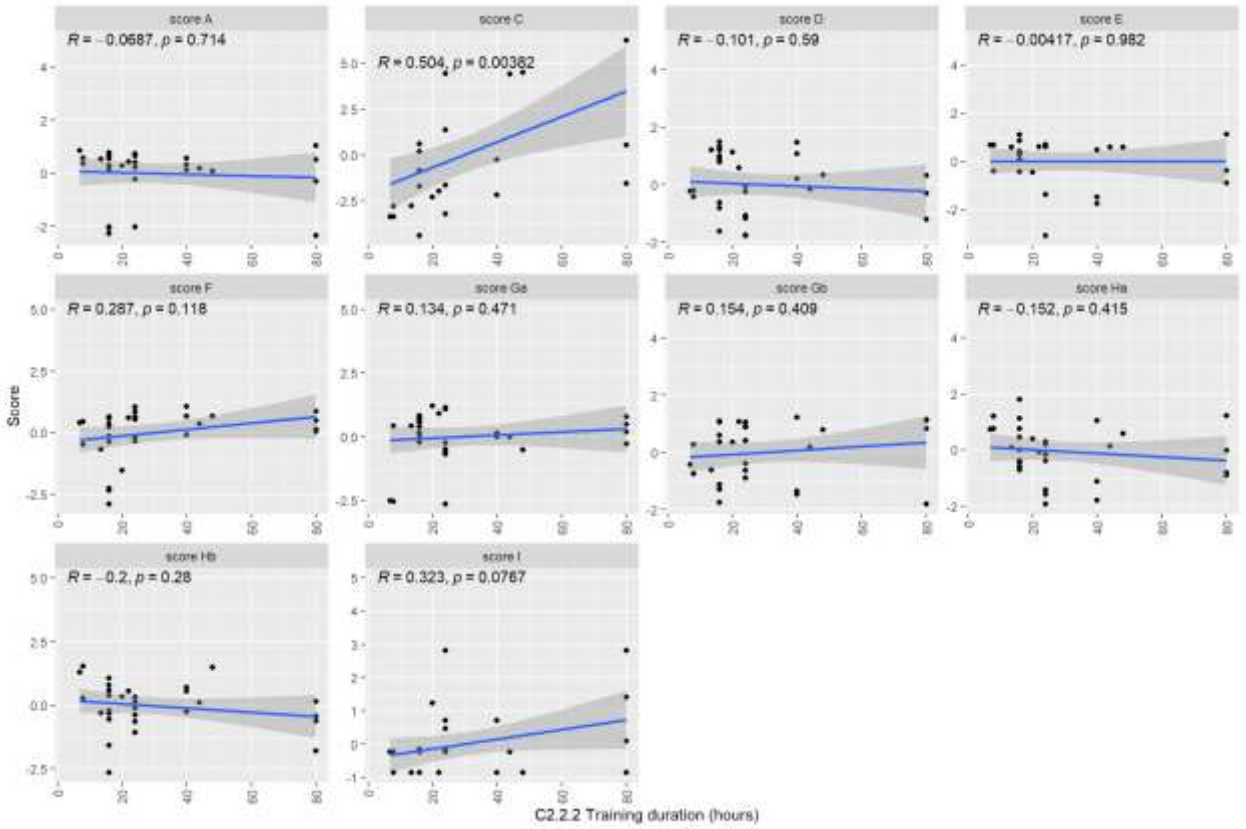

Figure 27. Scatterplots for dimension summary scores according to the indicator C2.2.*2 Training duration (hours)*. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting.

In its turn, having a background in nutrition, despite presenting a factor loading below 0.4 for dimension C, was moderate-strongly associated with the indicators assessing the FoodEx2 use (specificity and facets) that mainly represent the dimension G_a *Food and recipe description*. Accordingly, the boxplot below (Figure 28) shows that having a background in nutrition was associated with higher scores for dimension G_a *Food and recipe description*, and G_b. *Missing data* (lower prevalence).

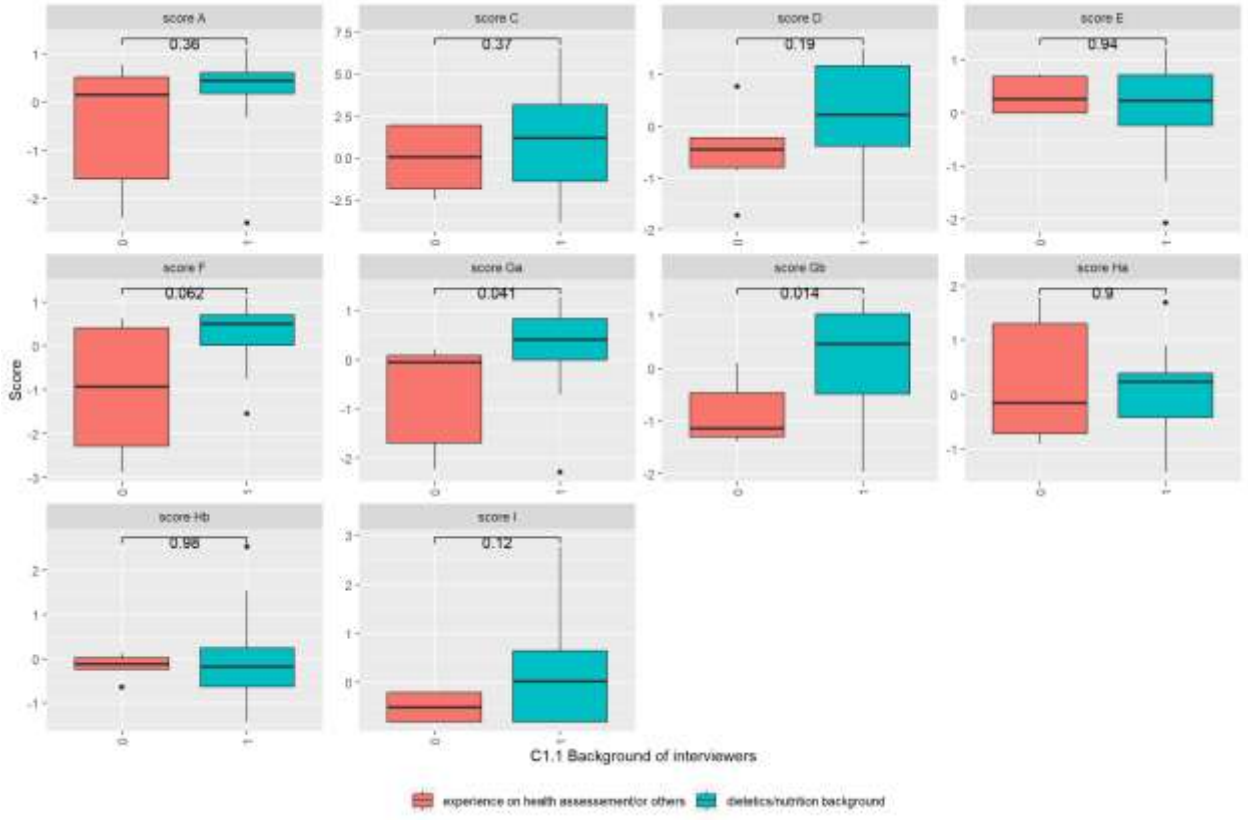

Figure 28. Boxplots for dimension summary scores according to the indicator C1.1 Interviewers background in Nutrition or experience in health assessment. The blue boxes represent the surveys that had interviewers with background in nutrition and dietetics whereas the red boxes surveys with interviewers with different background. The numbers between boxplots are the p-values. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting

Dimensions D. Interview administration and *G_a Food and recipe description* (r=0.42, p-value=0.02)

Higher D score correlated with G_a score suggesting that better interview administration procedures led to better description of the foods and recipes consumed in the dietary surveys. Looking into the specific indicators from these dimensions, it can be observed that the possibility of participants choosing the interview setting was positively correlated with the total use of facets. Moreover, a more comprehensive use of facets (indicators G2.3 and G2.4) was positively correlated with a better distribution of interviews per weekdays and seasons.

Dimensions D. Interview administration and *H_a Data reliability* (r=0.50, p-value=0.004)

Higher D score correlated with H_a score suggesting that overall better interview administration procedures led to higher data reliability. A duration of the interview closer to the benchmark (i.e., 30-45 min for 24h-R) was moderately correlated with higher ICC for food groups report (r=0.36, p-value=0.04). Higher ICC was also found to be correlated with a more uniform distribution of dietary interviews per weekdays.

As shown in the previous paragraphs, Dimension D score correlated with Dimension G_a and H_a. Dimension D indicators relate with the mode of administration of the interview, however within score D, the indicators *D1.1 At least one face-to-face interview and D2.2 Proportion of participants with interview gap 8-15 days* were not meaningful (factor loading <0.4). Moreover, indicator *D1.2.1 Same mode of administration to all participants at each interview moment* was negatively considered within the dimension D score. Thus, these indicators will be addressed isolated to check their relevance to overall data quality, particularly regarding the dimensions G_a and H_a.

Having at least one face-to-face interview was not necessarily correlated with higher quality in most dimensions (Figure 29). Contrarily, surveys that did not have at least one face-to-face interview seem to have higher data reliability (H_a), although not significant (p-value=0.089), according to Figure 29. Regarding indicator *D2.2.2,* the scatterplot (Figure 30) shows that a higher Proportion of participants with interview gap 8-15 days is associated with a higher G_a score (Food and Recipe description) (p-value=0.01). In fact, this indicator is moderately correlated with indicators that reflect improved facet use (G2.4 Minimum recommended facets and G1.4. (Lower) Proportion of foods with only the base term). On the contrary, if the proportion of participants with interview gap ≥8 days is considered (sensitivity analysis), no significant associations with other dimension scores were found, as shown in Figure 31.

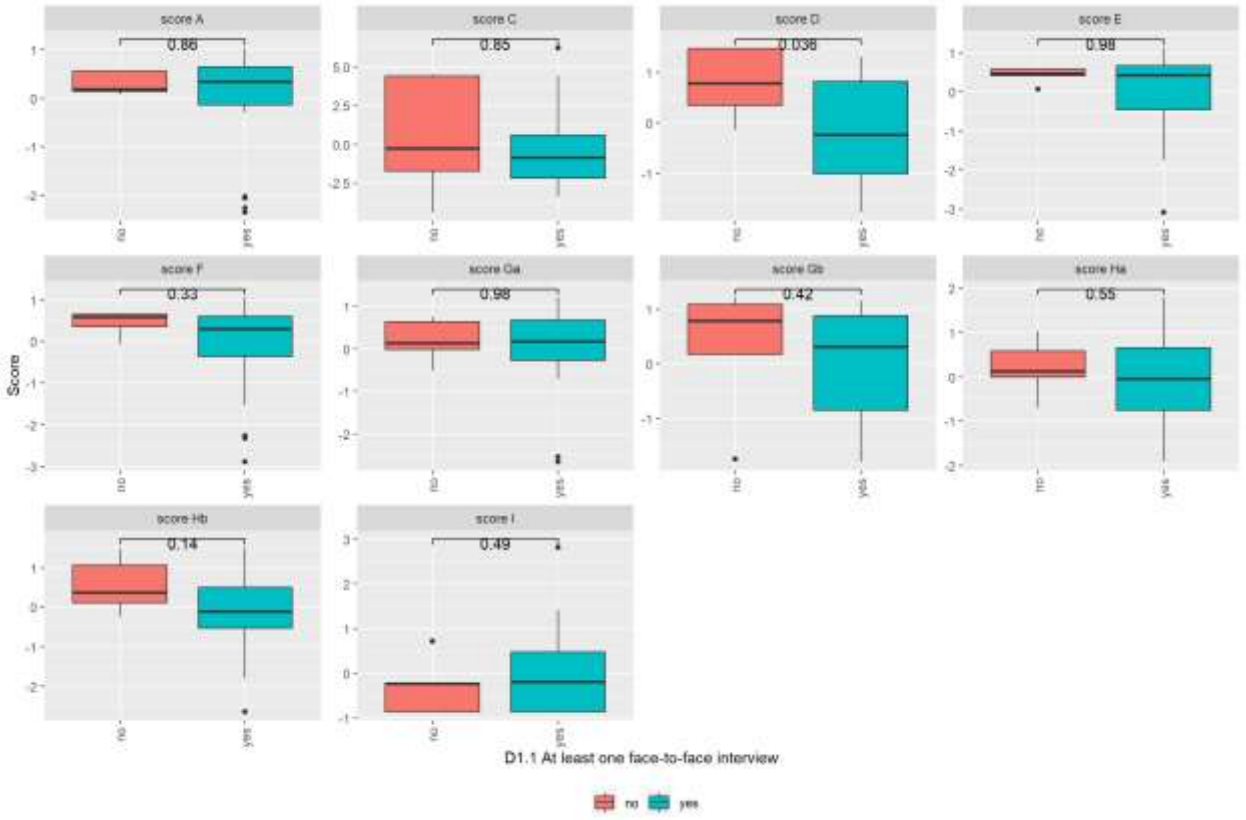

Figure 29. Boxplots for dimension summary scores according to the categories of the indicator D1.1 *At least one face-to-face interview (no; yes)*. The numbers between boxplots are the p-values. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting

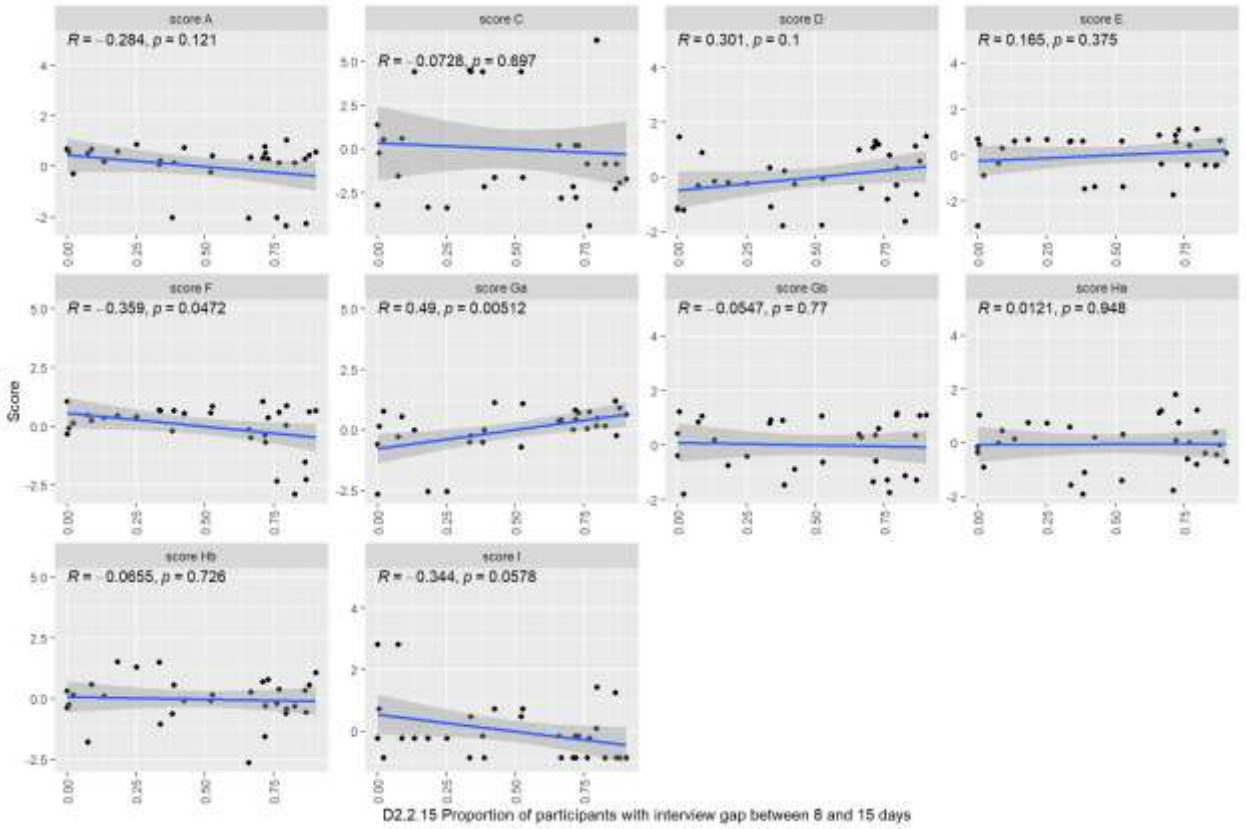

Figure 30. Scatterplots for dimension summary scores according to the indicator D2.2 *2 Proportion of participants with interview gap 8-15 days*. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting.

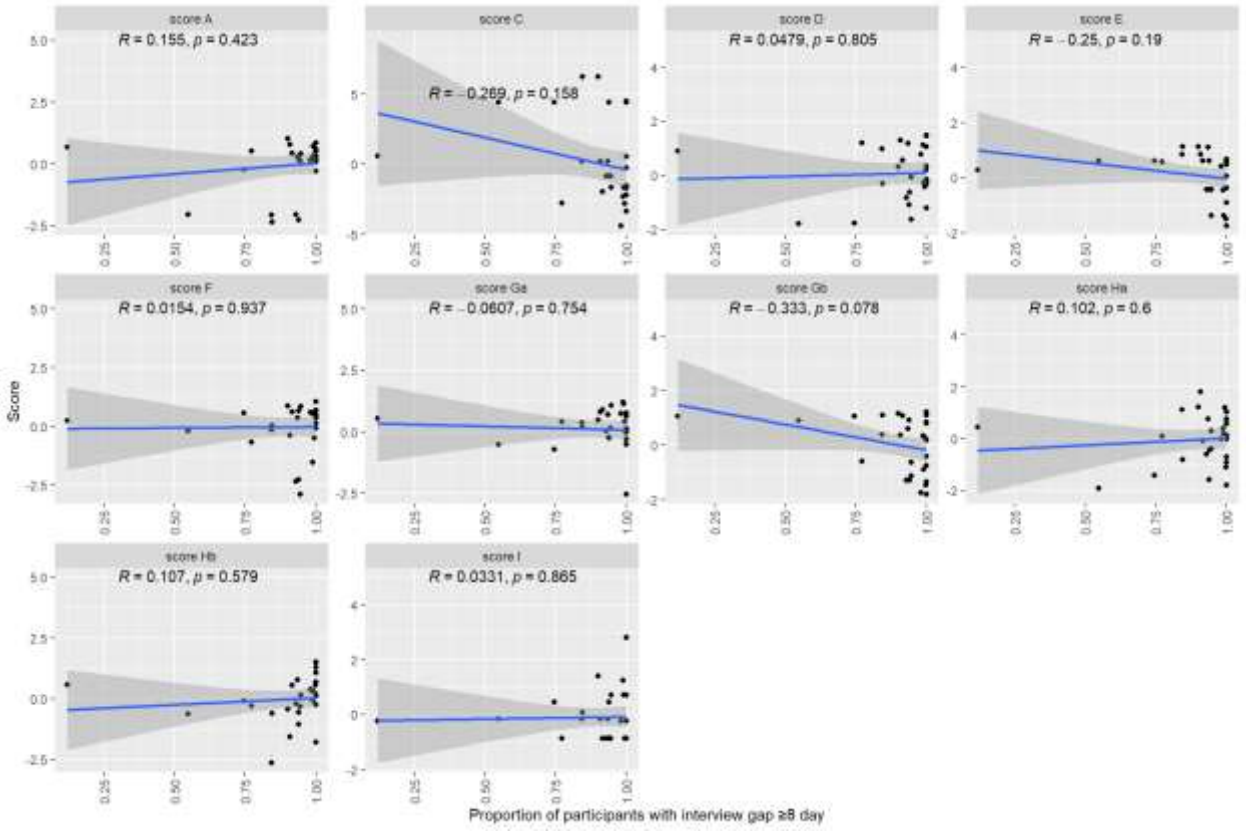

Figure 31. Scatterplots for dimension summary scores according to the sensitivity analysis done for indicator D2.2 *2 Proportion of participants with interview gap ≥8 days*. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting.

Having the same mode of administration of the interview to all participants (e.g. all CAPI vs some CAPI and others CATI) was not clearly associated with higher or lower scores for the other dimensions (Figure 32). The negative correlation with the remaining indicators from the Dimension D, as shown in

Table 11**,** suggests that a higher flexibility in the mode of administration is expected to result in a better distribution of interviews per weekdays and seasons and also a higher percentage of complete participants and a higher compliance with the recommended interview gap.

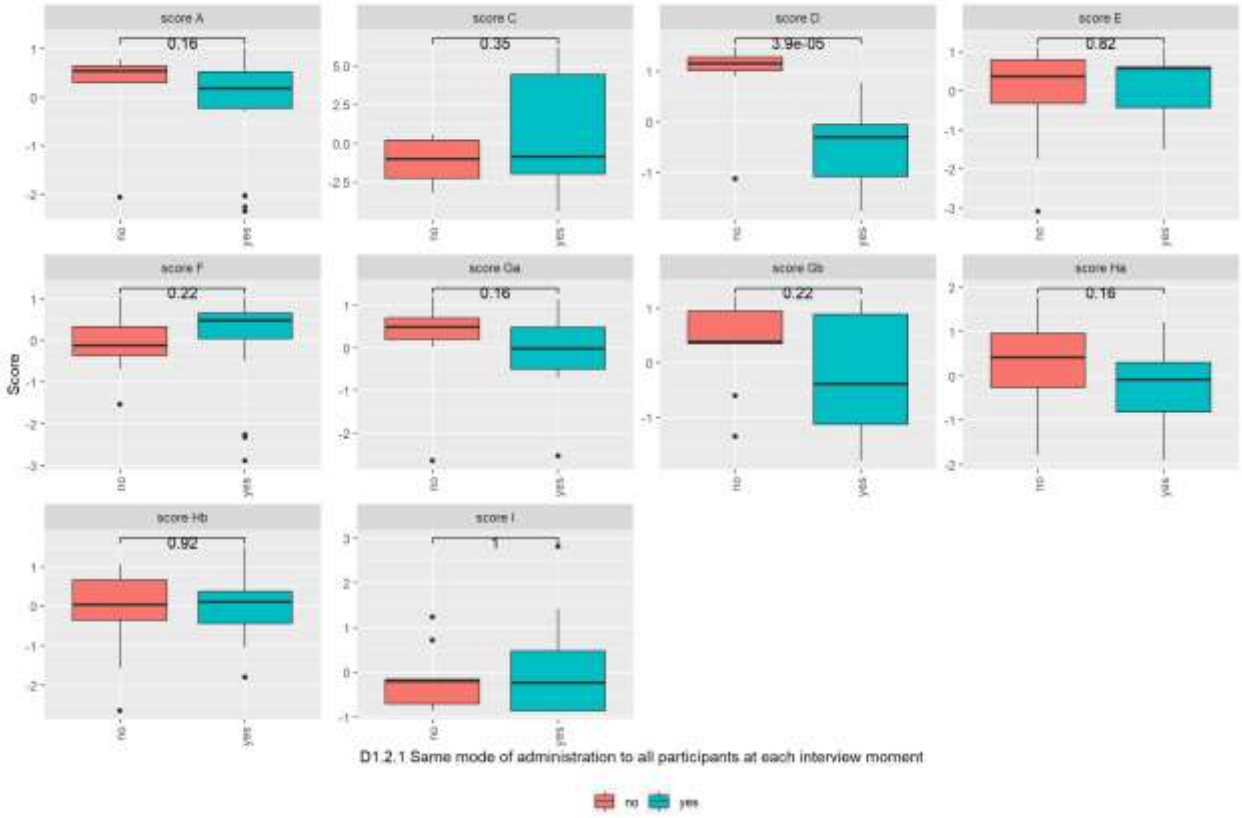

Figure 32. Boxplots for dimension summary scores according to the categories of the indicator D1.2.1 *Same mode of administration to all participants at each interview moment (no; yes)*. The numbers between boxplots are the p-values. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting

Dimensions *E. Software tools and validation procedures* and *G_b Missing data* (r=0.46, p-value=0.01)

Lower proportion of missing data*,* particularly in the foods and consumption dataset (indicators G1.5 and G2.2) was significantly associated with validated software tools using the multipass method and with prompts for minimum and maximum accepted quantities per food type, probing questions for easily forgotten foods, calculation of energy and macronutrients at the end of the interview.

The indicator E2.3 measuring total number of picture series in the picture book was not associated with dimension E score. The scatterplot presented in Figure 33 shows a marginally significant association between this indicator Ga score measuring Food and Recipe description.

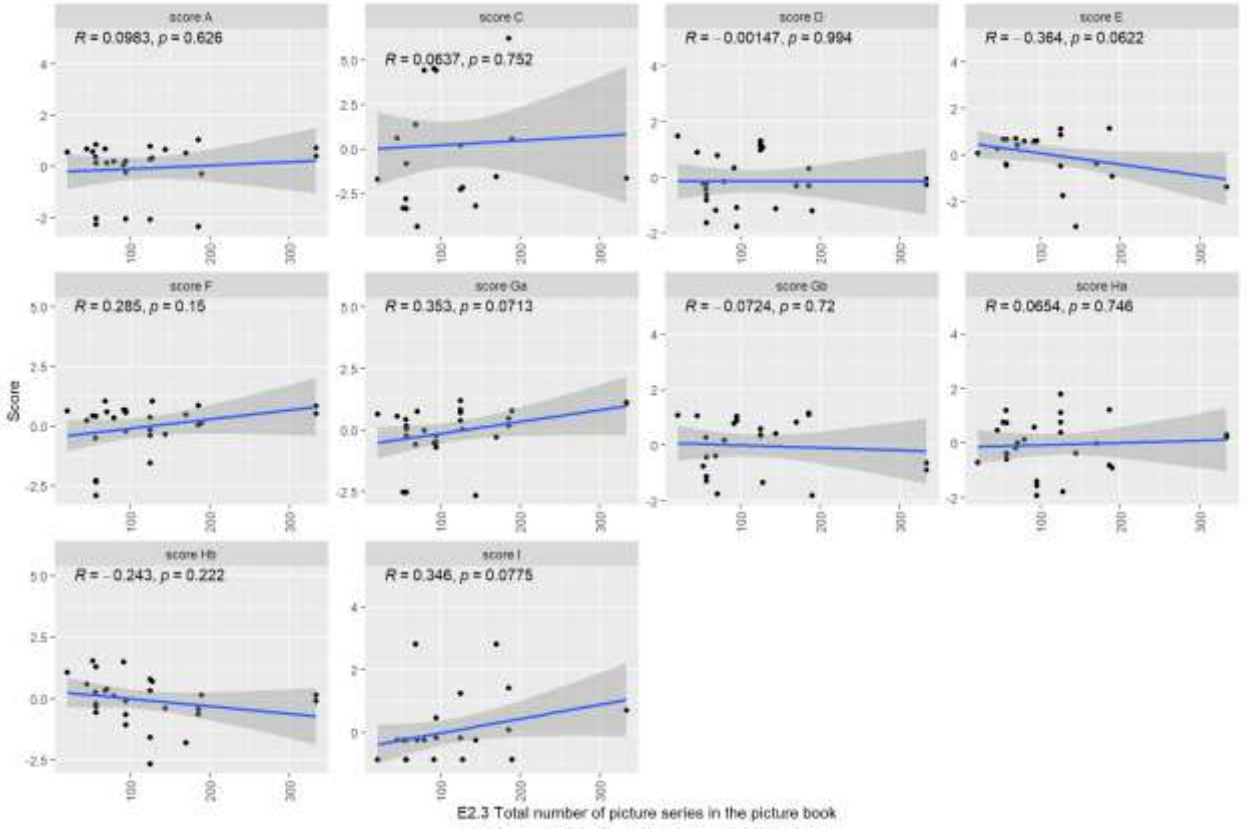

Figure 33. Scatterplots for dimension summary scores according to the indicator E2.3 *Total Number of Picture series in the picture book.* Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b Outliers, I. Results reporting.

A search Directed acyclic graph (DAG) is used to graphically depict and reinforce the associations between summary quality scores, after mutual adjustment (Figure 34). The main findings from the adjusted model, presented in the DAG, are mostly in line with the previous description of correlations. The most relevant conclusions focusing on data quality outcomes (Dimensions G_a, G_b, H_a and H_b) according to the DAG are the following:

a) **Higher quality scores in dimensions A and D** were associated with a **higher score in dimension H_a**, meaning that an effective sampling plan and adequate interview administration procedures such as accounting for seasonality and interview duration led to higher data reliability.

b) **Higher quality scores in dimensions A** were associated with a **higher score in dimension H_b,** indicating that surveys with better sampling presented lower proportion of outliers in energy intake and food groups quantification.

c) **Higher scores in dimensions C and D** were associated with a **higher G_a score** and **higher G_b score**, suggesting that better training of interviewers and appropriate interview administration procedures (seasonality and interview duration) promoted an enhanced food and recipe description, reflected through a higher use of facets and improved FoodEx2 use as well as lower proportion of missing data in the datasets.

d) **Better performance** on the non-dietary data collection **dimension (F)** was positively associated with the dimensions reflecting the **sampling plan (A)** and the **training and supervision of interviewers (C)**.

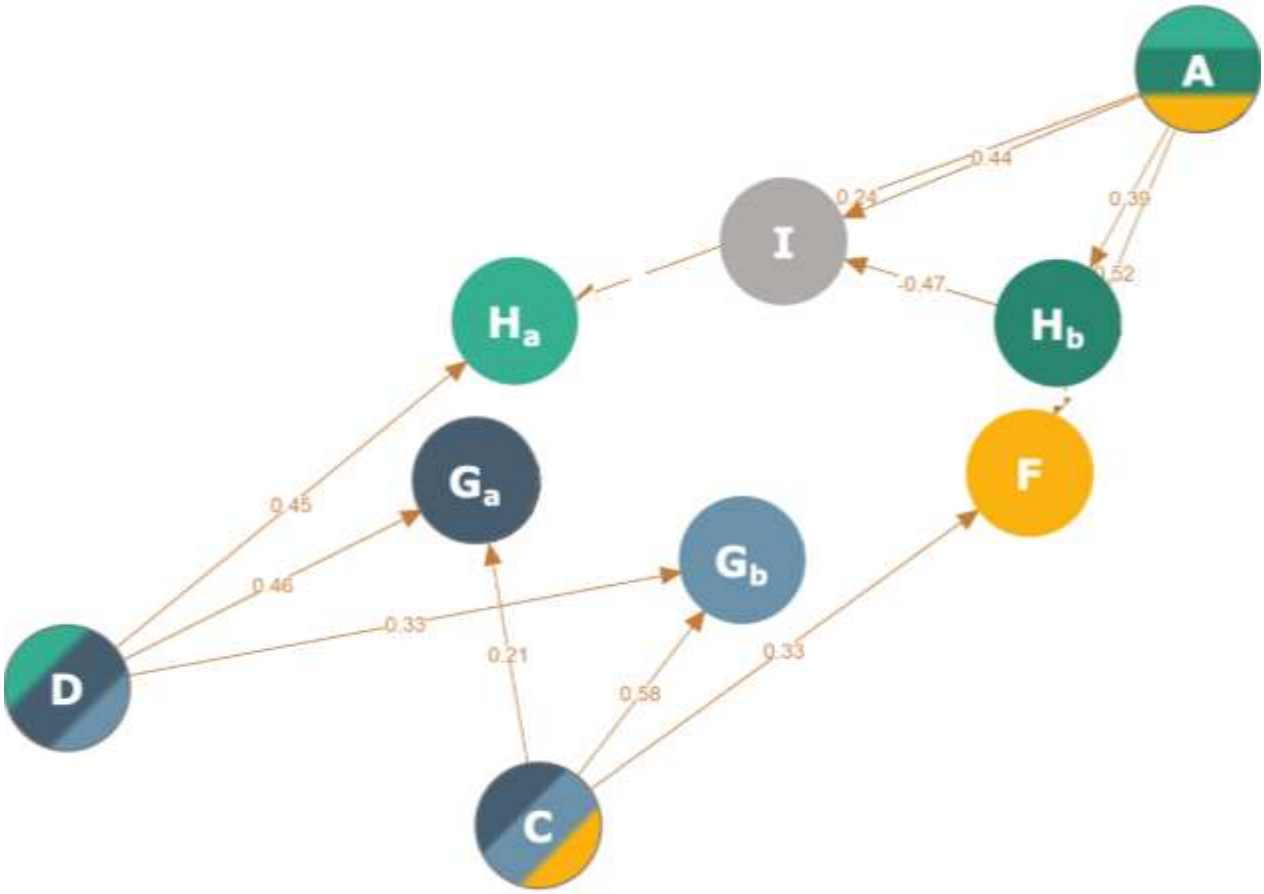

Figure 34. Search Directed Acyclic Graph (DAG) illustrating the network of adjusted associations between survey quality dimensions. Goodness of fit measures: CFI=0.901; TLI=0.857; RMSEA=0.080. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, Ga. Food & Recipe description, Gb. Missing data, Ha. Data reliability, H_b. Outliers, I. Results reporting

Additional summary quality scores results

Plausible reporters according to the Goldberg cut-off method (Goldberg et al., 1991), updated by Black (Black, 2000)

Figure 35 plots the proportion of plausible reporters, according to the Goldberg cut-off method (Goldberg et al., 1991), updated by Black (Black, 2000), and summary quality scores from the several dimensions. This indicator was not included in the PCA analyses because it was only computed for surveys with ≥10 year old participants. Overall, no significant results were found. However, the indicator-indicator correlations showed that a higher proportion of plausible reporters correlated moderately (r>0.3 and p-value<0.05) with a higher ICC for nutrients, higher prevalence of reporting foods from higher FoodEx2 hierarchical levels, and a more uniform distribution of interviews across seasons.

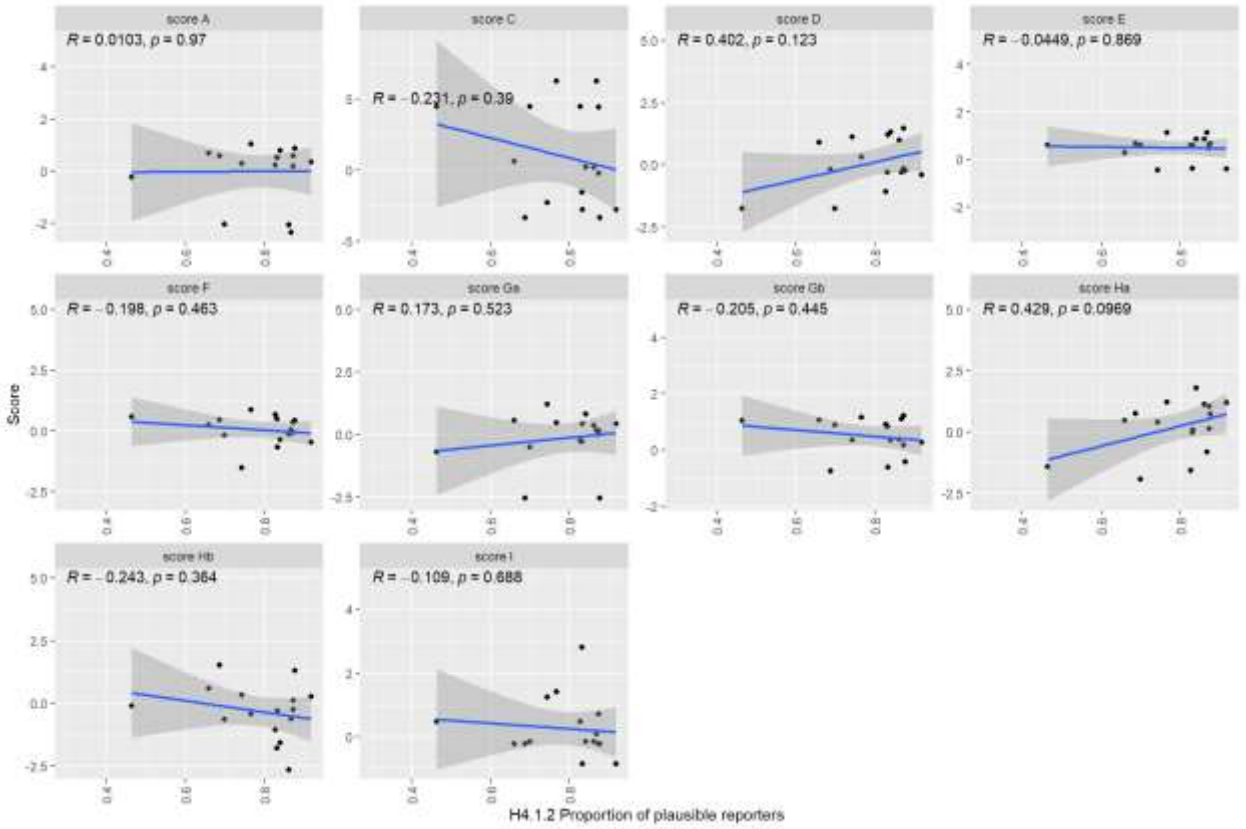

Figure 35. Scatterplots for dimension summary scores according to the indicator H4.1.2 *Proportion of plausible participants*. *Note: This indicator was computed only among participants ≥10 years when energy intake was available.* Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting.

Summary quality scores according to the survey target groups included

Because the original EFSA EU Menu Surveys datasets differed in terms of the target groups included, a boxplot (Figure 36) was created to visualize the overall performance on the summary quality scores of the datasets including only participants <10 years ("child"), ≥10 years ("adult"), all population ("child&adult") and ad-hoc samples ("ad-hoc"). The scores for dimensions A, F, G_b, H_b presented significant differences according to the survey target group. Ad-hoc surveys generally have lower median scores for dimensions A, F and H_b, whereas child surveys presented higher quality in dimension H_b related with outliers. Compared to surveys including only adults, surveys including only children ("child") and all age groups ("child&adult") have higher median scores for dimension G_b (Missing data). For the remaining dimensions no significant differences were observed according to the target group.

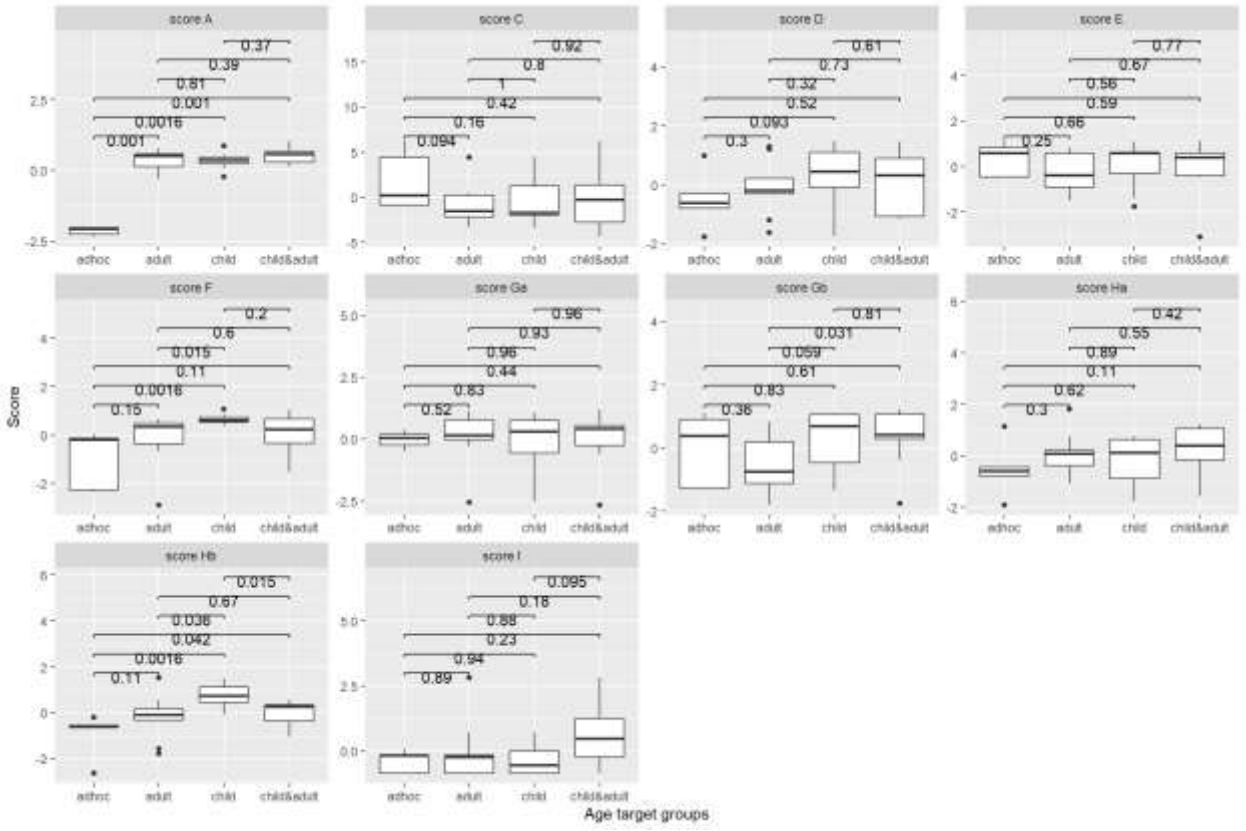

Figure 36. Boxplots for dimension summary scores according to the target groups included in the EFSA EU Menu Survey datasets. "child" refers to datasets including only subjects <10 years; "adult" refers to datasets including only subjects ≥10 years; "child&adult" refers to datasets including all age groups; "ad-hoc" refers to datasets including only ad-hoc (convenience) samples. The numbers between boxplots are the p-values. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers, I. Results reporting.

Summary quality scores across surveys

Figure 37 shows the summary quality scores according to the original survey datasets shared by EFSA. The scores were standardized with a mean of 50 and a standard deviation of 10. Thus, a score of 50 indicates that the survey quality for that dimension is at the mean level and each 10 points difference represents one standard deviation above or below the average. The dimension with higher discrepancies across surveys is dimension C, related with the training and supervision of interviewers, mainly due to differences in missing information in the survey reports. Here, it is also possible to observe that ad-hoc surveys present in general lower scores for dimensions A, F, and H_a. In contrast, for dimensions C, D, E and I, with rare exceptions, surveys from the same country tend to present similar scores.

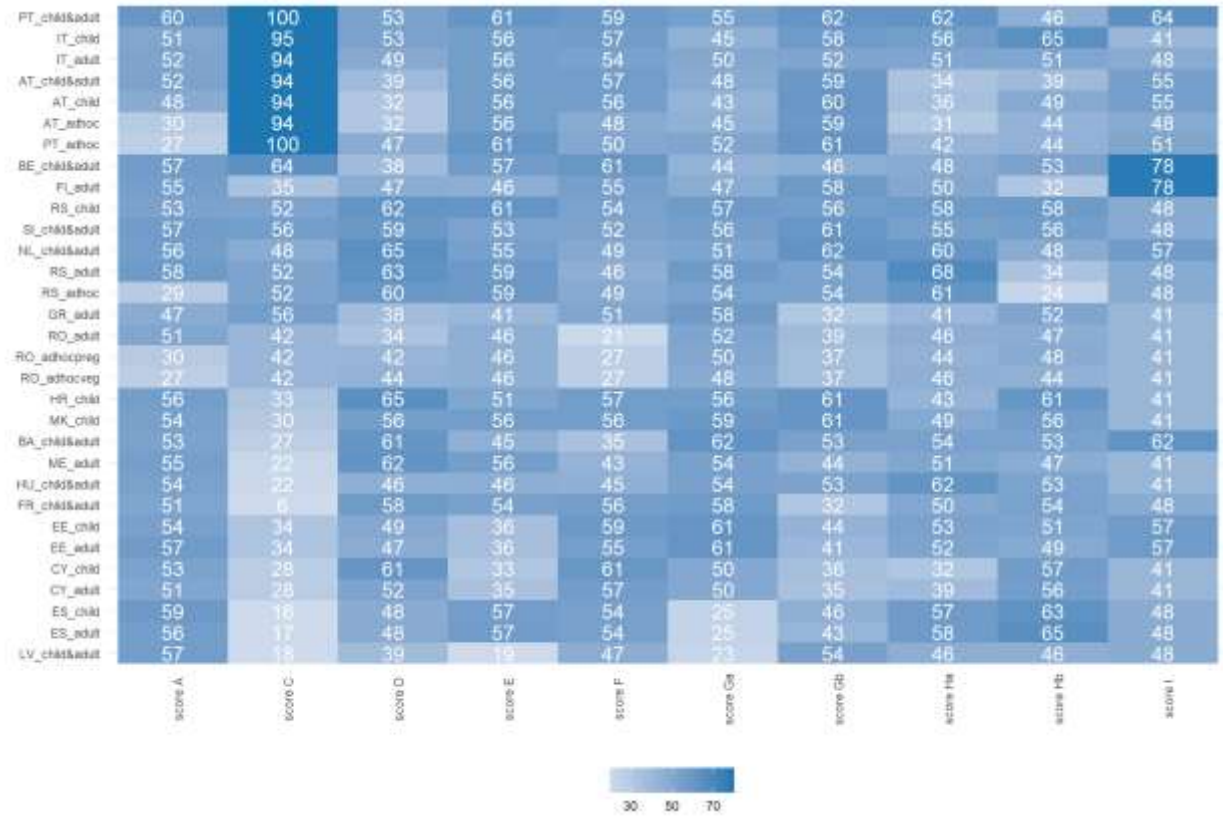

Figure 37. Summary quality scores per EU Menu original survey dataset. The scores were standardized with a mean of 50 and a standard deviation of 10. Dimensions/Score: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, Ga. Food & Recipe description, Gb. Missing data, Ha. Data reliability, Hb. Outliers, I. Results reporting.

3.2.3 Socioeconomic factors associated with summary quality scores

The associations between summary quality scores and socioeconomic factors, at the country level, were explored. The crude correlations are presented in Figure 38. The main findings according to the socioeconomic factors assessed are presented in the following paragraphs.

Education level

The proportion of the population with at least upper secondary attainment was negatively and moderately associated with dimensions C *Training and supervision of interviewers* (r=-0.39, p-value=0.04) and E *Software tools and validation procedures* (r=-0.46, p-value=0.01).

Income

On the other hand, higher income was associated with lower scores in dimensions D *Interview administration procedures* (r=-0.44, p-value=0.01) and H_a *Data reliability* (r=-0.37, p-value=0.04) and higher scores in dimension C *Training and Supervision of interviewers* (r=0.38, p-value=0.04).

Demography (rural vs urban inhabitants)

Finally, surveys in countries with a higher proportion of the population living in rural areas had lower scores in dimensions H_a *Data reliability* (r=-0.46, p-value=0.04), H_b Outliers (r=-0.48, p-value=0.04).

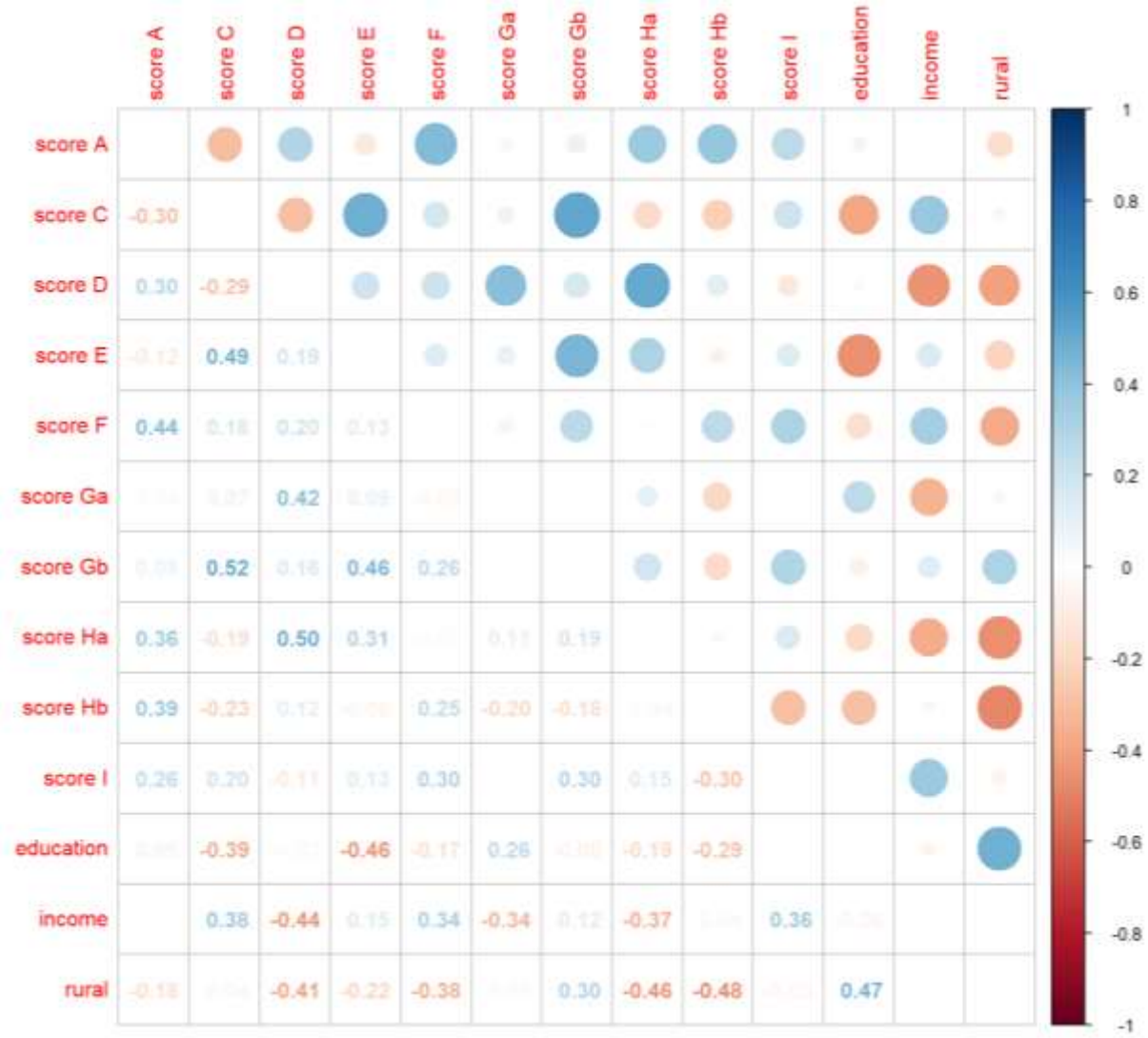

Figure 38. Correlation coefficients between dimension summary scores and socioeconomic factors. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, H_a. Data reliability, H_b. Outliers , I. Results reporting; education: % of the population with at least upper secondary educational attainment; income: GDP per capita, in product purchase standards; and rural (% of the population living in rural areas).

Lastly, a final search DAG is presented as a graphic representation of the mutually adjusted associations to support the findings (Figure 39). This model included the survey dimensions scores, the socioeconomic factors. The most relevant conclusions concerning the effect of socioeconomic factors and targets groups included in the survey dimensions, according to the search DAG (Figure 39) are the following:

a) **Higher proportion of the population living in rural areas** was associated with a **lower score in dimension H_a** that reflects data reliability.

b) Higher proportion of the population with at least upper secondary educational attainment was associated with a **higher score in dimension G_a** suggesting that highly educated populations are likely to better report and describe the foods consumed in the 24h-R. Moreover, this socioeconomic indicator is **negatively associated with dimension C**, suggesting that surveys from highly educated countries may invest less in the training and supervision of interviewers. In fact, adjusting for education of the country, boosted the association between dimension C and G_a, showing the relevance of adequate staff training particularly in surveys from countries with lower proportion of individuals with at least secondary education.

c) The other associations between dimensions that were previously described were overall upheld after adjusting for the socioeconomic factors from countries.

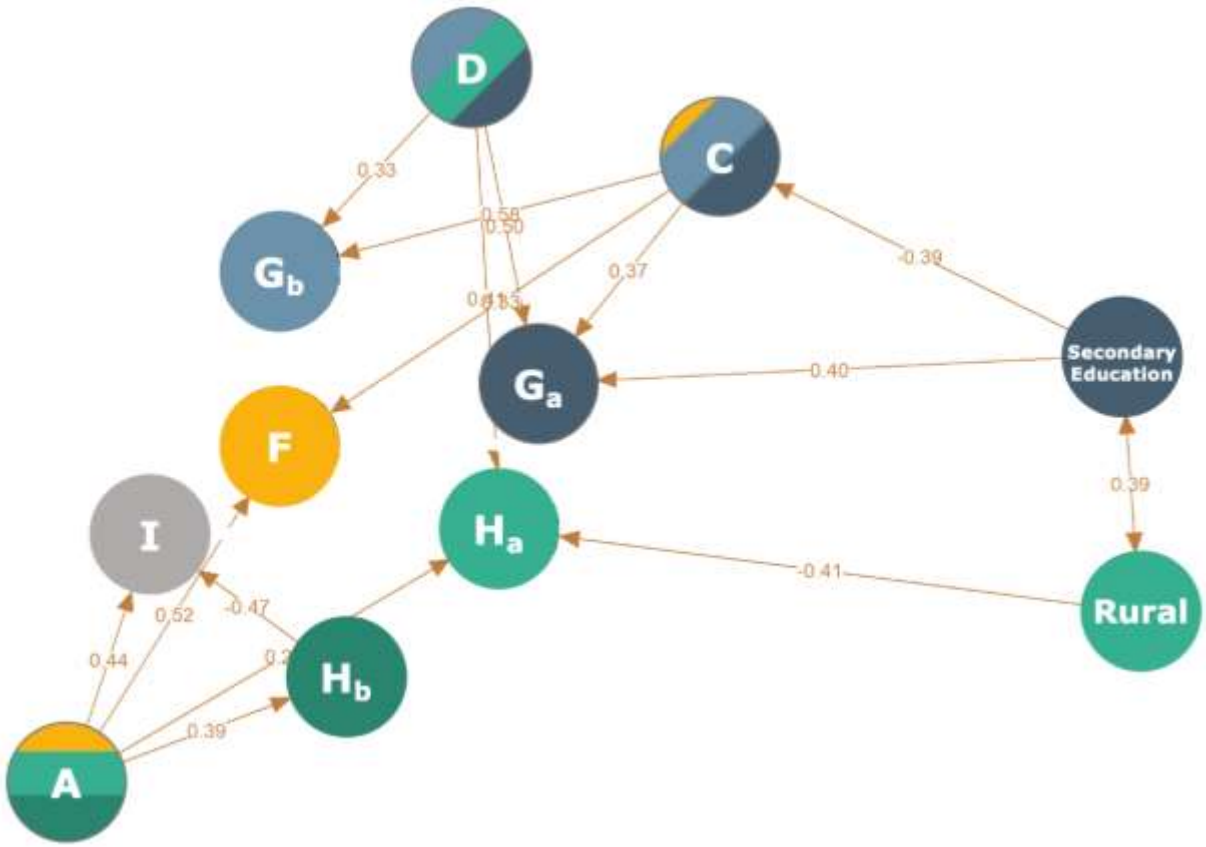

Figure 39. Search Directed Acyclic Graph (DAG) illustrating the network of associations between survey quality dimensions, survey target groups and countries' socioeconomic factors. Goodness of fit measures: CFI=1.000; TLI=1.265; RMSEA=0.000. Dimensions: A. Sampling Plan, B. Recruitment, C. Training and supervision of interviewers, D. Interview administration procedures, E. Software tools and validation procedures, F. Non-dietary data collection, G_a. Food & Recipe description, G_b. Missing data, Ha. Data reliability, H_b. Outliers, I. Results reporting; Secondary education: % of the population with at least upper secondary educational attainment; Rural (% of the population living in rural areas).

4 Discussion and Conclusions

The ERA EU Menu project evaluated the first round of national dietary surveys conducted according to the recommendations defined in the EFSA EU Menu Guidance of 2014. This evaluation followed a comprehensive protocol, which included several quality indicators reflecting the multiple survey dimensions. In the quality assessment process, both the EU MENU survey reports and the datasets were used. A descriptive and exploratory analysis was conducted to examine how various dimensions of the surveys - such as Sampling, Recruitment, Training and supervision of interviewers, Data collection procedures, Software tools, and validation - impacted data quality. Parameters under examination included data completeness, food and recipe description, data reliability, the presence of outliers, and energy misreporting.

In general, a higher quality in the sampling plan was associated to higher data reliability, lower prevalence of outliers and better non-dietary data collection, with ad-hoc surveys presenting lower quality in these dimensions. Moreover, better training of interviewers and appropriate interview administration procedures were associated with an improved food and recipe description, and lower proportion of missing data. The level of education of the country as well as its demography were relevant factors for data quality outcomes in the surveys. Furthermore, our results show that, overall, the surveys adhered to the guidance, resulting in a good level of harmonization and compliance to the recommendations. Nonetheless, some dimensions evaluated in this work presented transversal problems, namely the dimension regarding recruitment where the definitions of participations rate, contact rate and co-operation rate were faultily reported in many cases, or the dimension that measures the food and recipe description, due to the low prevalence of FoodEx2 facets report. Other problems encountered included the absence of essential information in the reports, discrepancies in seasonality during dietary data collection, and heterogeneity in data report to EFSA.
The impact of these issues in data quality were identified in this report throughout the results section and will be further explored and leveraged as valuable lessons in the ERA EU Menu final deliverable, to inform recommendations for future survey rounds. Thus, in this discussion section, we aim to reflect upon the methods employed in this assessment, and address its limitations, strengths and interpretability.

One of the main challenges in the accurate evaluation of survey quality pertains to discrepancies in reporting between surveys due to flexibility in the guidelines. While the reporting guidelines imply certain mandatory elements, they maintain a degree of flexibility in various dimensions, encompassing both the written documentation (i.e., surveys methodological reports) and the datasets shared with EFSA. In this context, many quality indicators relied on the information supposedly available in the survey reports, but its assessment was compromised due to a lack of information provided. Among these, were aspects related to statistics of sample representativeness, to the training and supervision of interviewers, duration of the interviews, characteristics of the software tools and calculation and prevalence of misreporting, among others. Many indicators evaluating these issues presented more than 50% of missing information based on the information retrieved from the reports, hindering the quality assessment. Minimum contract requirements, as for instance the case of season coverage (recommended: 4, but minimum required: 2), might have also impaired the assessment, increasing the complexity in interpretating of results not solely in

alignment with the guidance recommendations but also considering the basic contractual conditions.
Regarding datasets shared with EFSA, stipulated contract requirements established the minimum mandatory criteria for data sharing among surveys. However, the discretion afforded to individual surveys allowed for divergent practices, with some surveys opting to share additional data while others abstained. This variability in data sharing practices potentially introduced bias into specific indicators under evaluation. An example of this is the indicator that assessed the proportion of complete participants (those with ≥2 interviews): because the surveys could choose to share all the data or only the complete participants it is not possible to assess the extent of losses to follow up within the surveys, compromising the interpretability of this indicator. Other examples of such indicators are the socioeconomic variables and nutrients that were recommended or at least covered in the data schema template but were not mandatory.

The results obtained are directly influenced by the choice of quality indicators and their respective assessment methods. A concrete example of this matter is found for FoodEx2 classification and respective facet report. In this work, the FoodEx2 classification was assessed considering the complex consumed food codes as reported by the participants in each food consumption occasion. This involved examining each reported food item individually, such as for estimating the proportion of foods reported without facets (excluding intrinsic facets). This method can yield different results compared to using a unique list of foods instead. Digging deeper into this approach, it is important to consider the structure of the consumption datasets, in which each row represents a single food item reported by a participant for a specific meal on a specific day. The description of the food item in terms of facets not only depends on the unique characteristics of the food itself but also on factors like the software used, the interviewer's skills, and the participant's level of commitment and knowledge on the food consumed. As a result, two food items with the same facets reported by the same or different participants within the consumption dataset may or may not be identical. Our method of treating each food item independently avoids making speculative assumptions about the similarity of foods within the dataset. Instead, we focus on accurately evaluating the objective percentage of facets reported in each food item. In contrast, using a unique list of foods would assume that all food items with the same facets are identical, which is not always the case.
Consequently, for this and other examples, the interpretation of each indicator was carefully considered and formulated according to the indicators' specific construction and assessment, which may differ from results that would have been obtained using different approaches.

The absence of well-established gold-standard cut-offs for the quality indicators as well as the practical unattainability of perfect compliance for most indicators is also a limitation of this work. To mitigate this constraint, we employed a benchmarking approach, which involved either selecting the survey with better performance or computing the meta-average of all surveys. It is important to emphasize that the main objective of this benchmarking exercise was not to rank or compare surveys, but rather to ensure an equitable evaluation of their quality.

Numerous quality indicators were assessed in this evaluation, posing a challenge for drawing concise conclusions and examining the relationships between indicators without a focused

strategy. The vast number of multi-comparisons between individual indicators was difficult to interpret and may have identified several associations by chance. Therefore, a summary-based approach, using PCA, was used to synthesize the evaluation of survey quality for each dimension. This approach facilitates the correlation of diverse dimensions and enables the assessment of how various indicators influence different aspects of data quality reducing the likelihood of finding associations by chance. In this analysis, due to the predominant influence of country-related factors over age groups in most indicators, we opted not to stratify the analysis. Instead, we relied on the original EU Menu datasets, which contained only 31 observations. This relatively small dataset size may imply less power and the PCA might have been more sensitive to the presence of outliers. Nevertheless, to minimize this possible bias, the outliers were identified through multivariate analysis and removed.

Another point relevant for discussion is the way missing values were dealt with in the analyses to identify associations between dimensions and indicators. Missing values in continuous variables were substituted with average values based on the remaining surveys, while categorical indicators with missing values were recategorized as 'no' (or zero). This may have implied misclassification of surveys for some variables. However, even though there is that possibility, it is unlikely because when missing the information is most likely a "no". Moreover, using the average in continuous indicator had most likely attenuated the possible associations with other indicators or dimensions. An attenuation of associations may also exist due to the overall good level of harmonization among surveys. While this harmonization is generally a positive aspect, it may have limited the ability to detect associations between survey dimensions that were otherwise expected.

In this work, a search DAG that uses a specific statistical algorithm was used to find the direction of the associations between scores/dimensions. It is possible that the algorithm failed to find the correct directions, or that a different algorithm could have found slightly different associations. However, the authors verified the causal plausibility of the findings from the final DAGs and presented in this report the most consensual solution obtained.

Considering all these aspects, as well as the fact that the analyses between survey dimension and socioeconomic factors are mainly ecological, the results stemming from the PCA and the subsequent associations were thoughtfully interpreted, and solely pertinent findings, which could be logically interpreted, were highlighted in this document.

As previously mentioned, the main findings of this report along with the results from the Work Package 1 (van Rossum et al., 2022) of the ERA EU Menu project will be used to tailor advice on future recommendations for the next round of national dietary surveys in a pan-european context in a specific report.

References

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. 24, 1119–1130. https://doi.org/10.1038/sj.ijo.0801376

Cattell, R.B., 1966. The scree test for the number of factors. Multivariate Behav. Res. 1, 245–276.

de Boer, E.J., Slimani, N., van 'T Veer, P., Boeing, H., Feinberg, M., Leclercq, C., Trolle, E., Amiano, P., Andersen, L.F., Freisling, H., Geelen, A., Harttig, U., Huybrechts, I., Kaic-Rak, A., Lafay, L., Lillegaard, I.T., Ruprich, J., de Vries, J.H., Ocké, M.C., 2011. Rationale and methods of the European Food Consumption Validation (EFCOVAL) Project. Eur. J. Clin. Nutr. 65, S1–S4. https://doi.org/10.1038/ejcn.2011.82

EFSA, 2014. Guidance on the EU Menu methodology. EFSA J. 12, 1–80. https://doi.org/10.2903/j.efsa.2014.3944

Gibson, R.S., Charrondiere, U.R., 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. 8, 980–991. https://doi.org/10.3945/an.117.016980

Goldberg, G.R., Black, A.E., Jebb, S.A., Cole, T.J., Murgatroyd, P.R., Coward, W.A., Prentice, A.M., 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, 569–581.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2014. Multivariate data analysis, Seventh ed. ed. Pearson Education Limited Harlow, Harlow SE - ii, 734 pages : illustrations ; 28 cm. https://doi.org/LK - https://worldcat.org/title/900353065

Rutishauser, I.H., 2005. Dietary intake measurements. Public Health Nutr. 8, 1100–1107. https://doi.org/10.1079/PHN2005798

Spirtes, P., Glymour, C.N., Scheines, R., 2000. Causation, prediction, and search. MIT press.

van Rossum, C., ter Borg, S., Nawijn, E., Oliveira, A., Carvalho, C., Ocké, M., 2022. Literature review on methodologies and tools for national dietary surveys; results of ERA EU-menu-project. EFSA Support. Publ. 19. https://doi.org/10.2903/sp.efsa.2022.en-7725

Willett, W., 2012. Nutritional epidemiology. Oxford university press.

Glossary and Abbreviations

EFSA	European Food Safety Authority
WHO	World Health Organization
TDI	Tolerable daily intake
24h-R	24-hours recall
BMI	Body Mass Index
DAG	Directed Acyclic Graph

Appendix A – Map surveys according to quality indicators: Extra figures

Dimension A – Sampling plan: methods and coverage

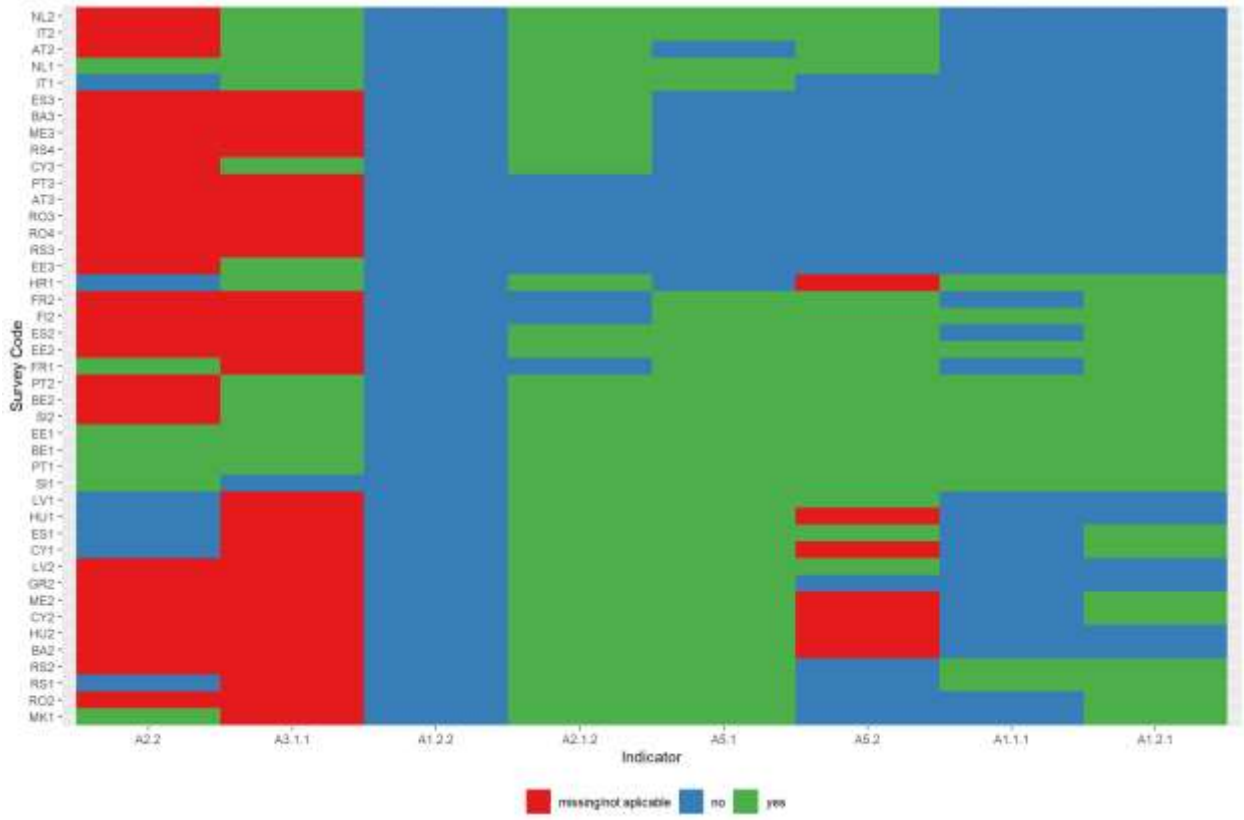

Figure A 1. Heatmap for categorical indicators of Dimension A. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

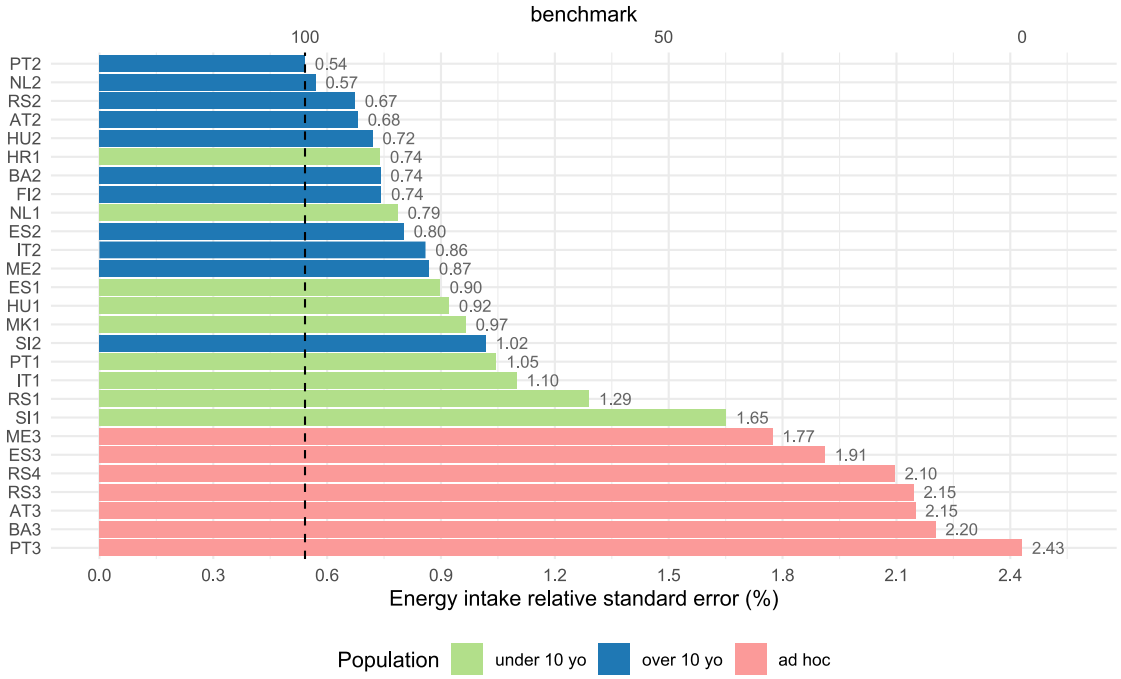

Figure A 2. Indicator A4.1 Relative standard error for Energy Intake.

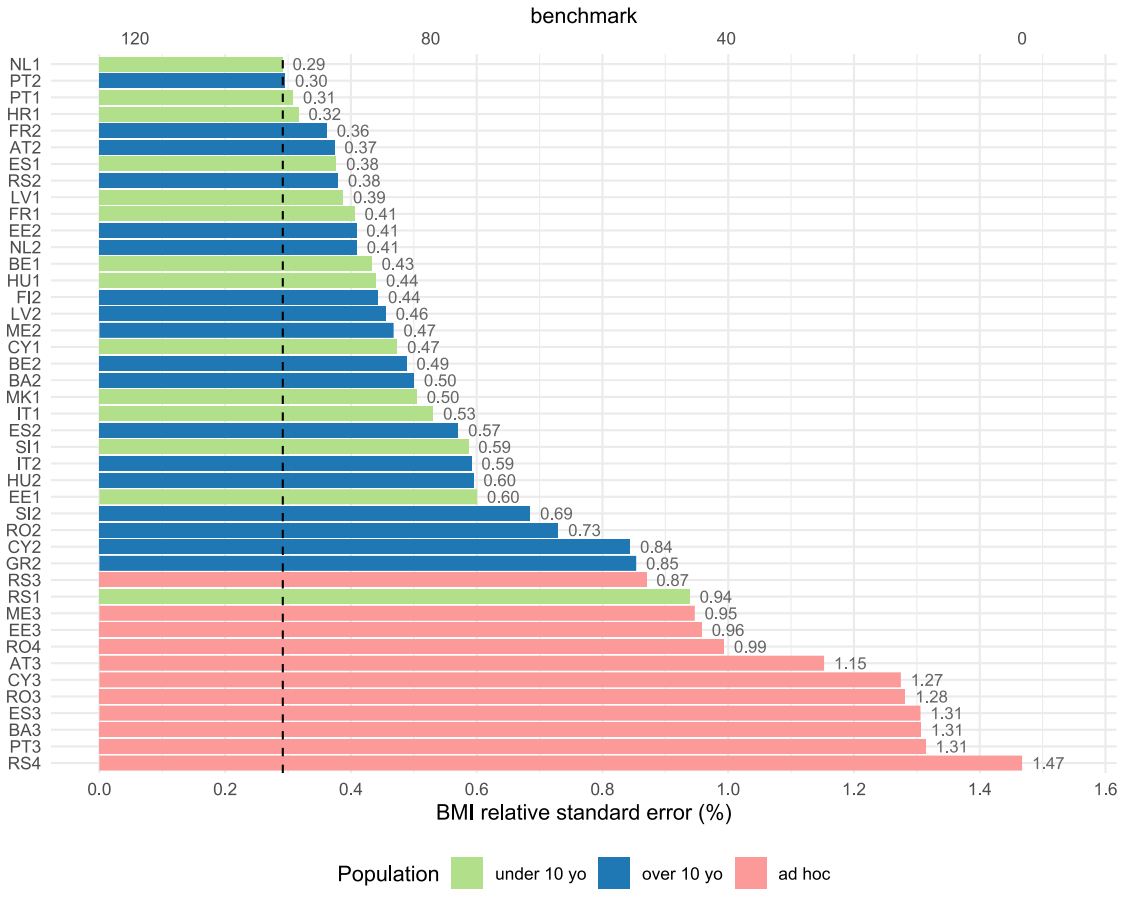

Figure A 3. Indicator A4.1 Relative Standard Error (RSE) for Body Mass Index.

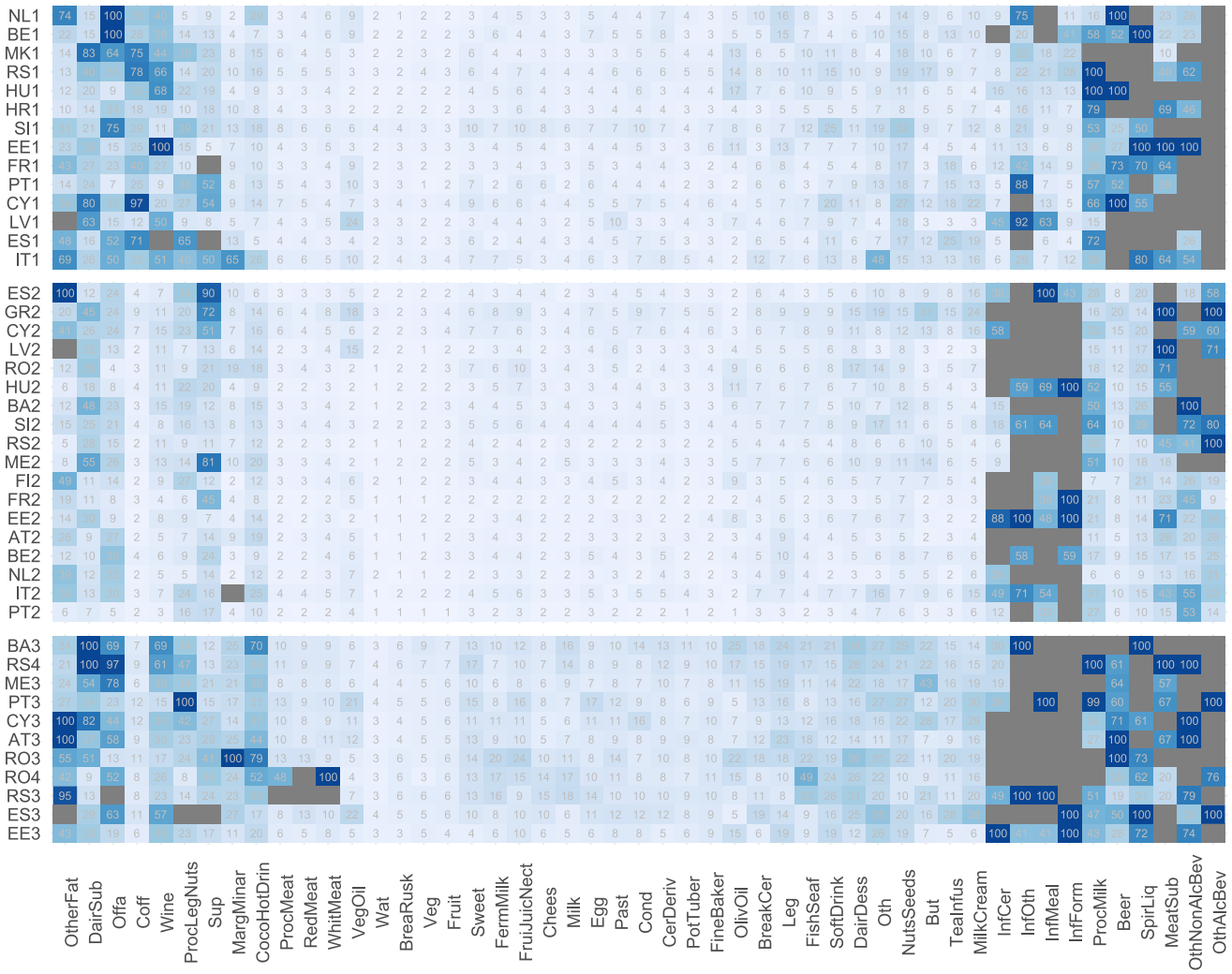

Figure A 4. Indicator A4.1 Relative Standard Error (RSE) for Food Groups. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients.

Dimension B – Recruitment

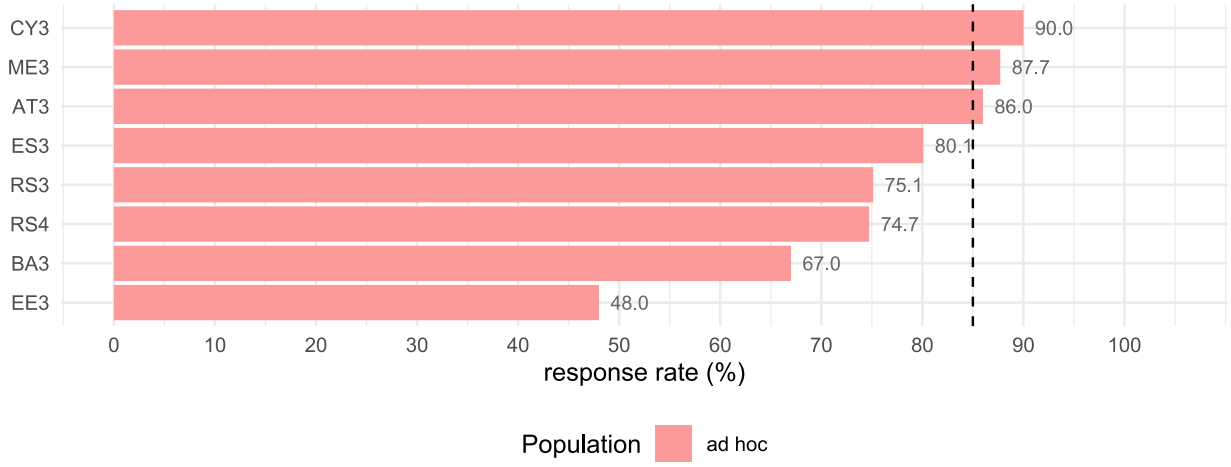

Figure A 5. Indicator B1.1 Response rate for ad-hoc surveys.

Dimension C – Training and supervision of interviewers

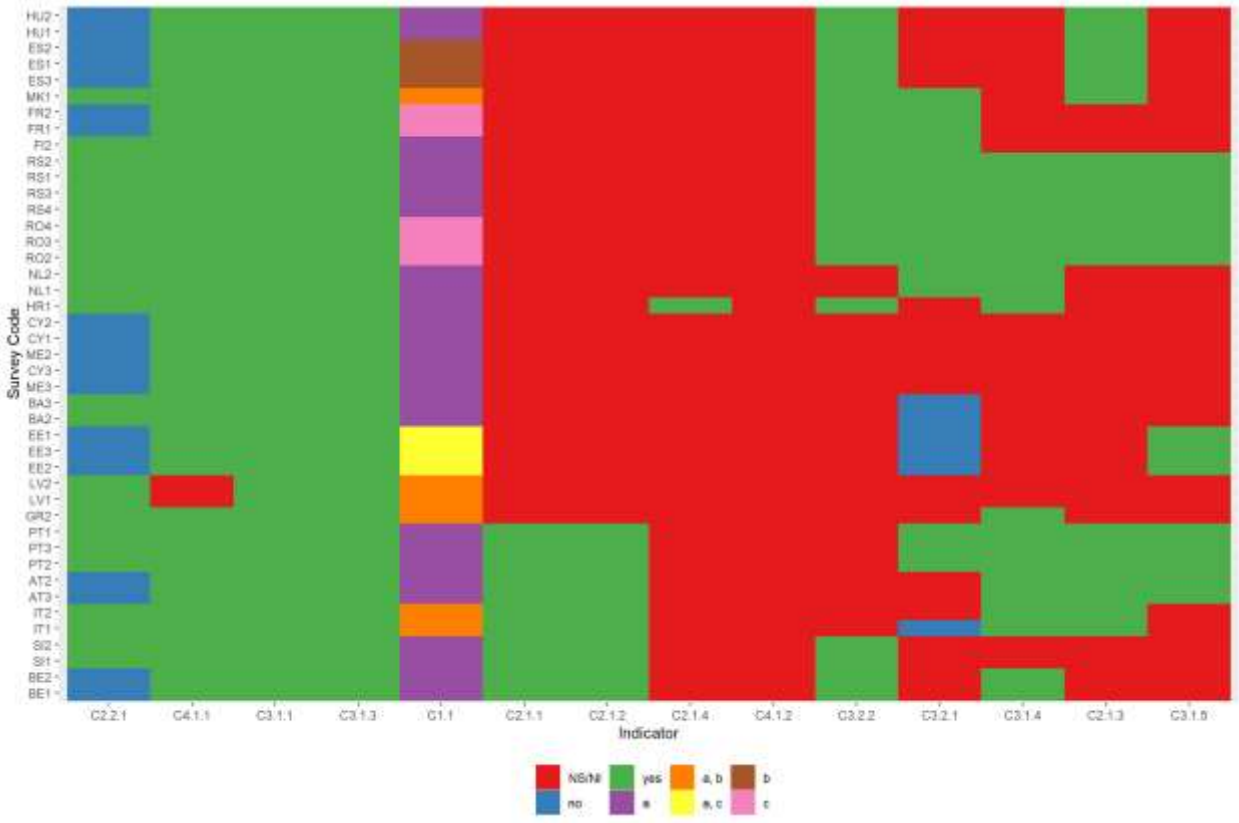

Figure A 6. Heatmap for categorical indicators of Dimension C. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

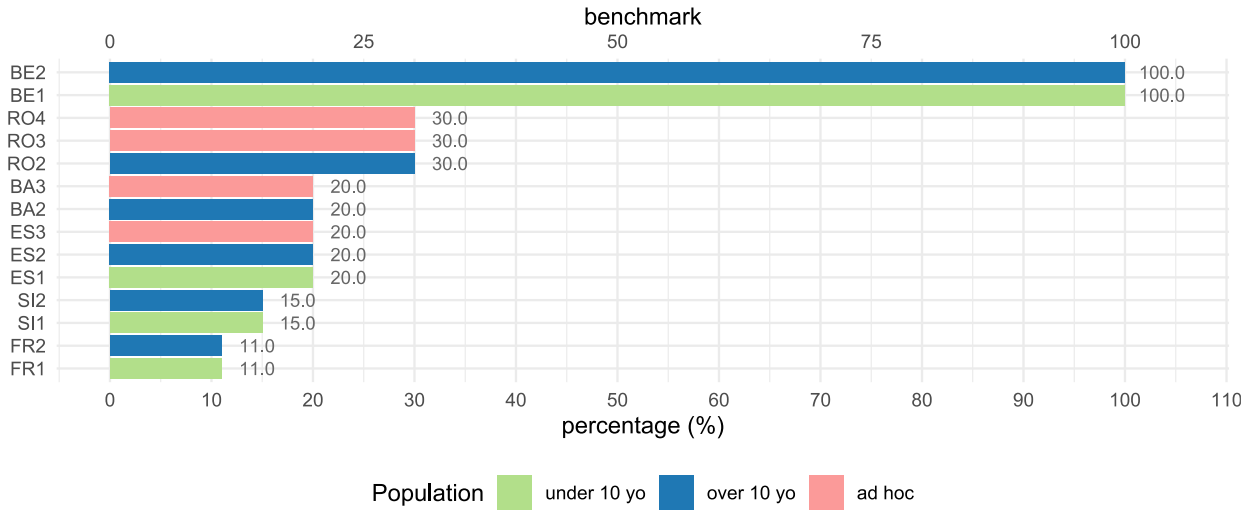

Figure A 7. Indicator C3.1.2 – Percentage of interviews re-contacted via telephone.

Dimension D – Data collection: Interview administration procedures

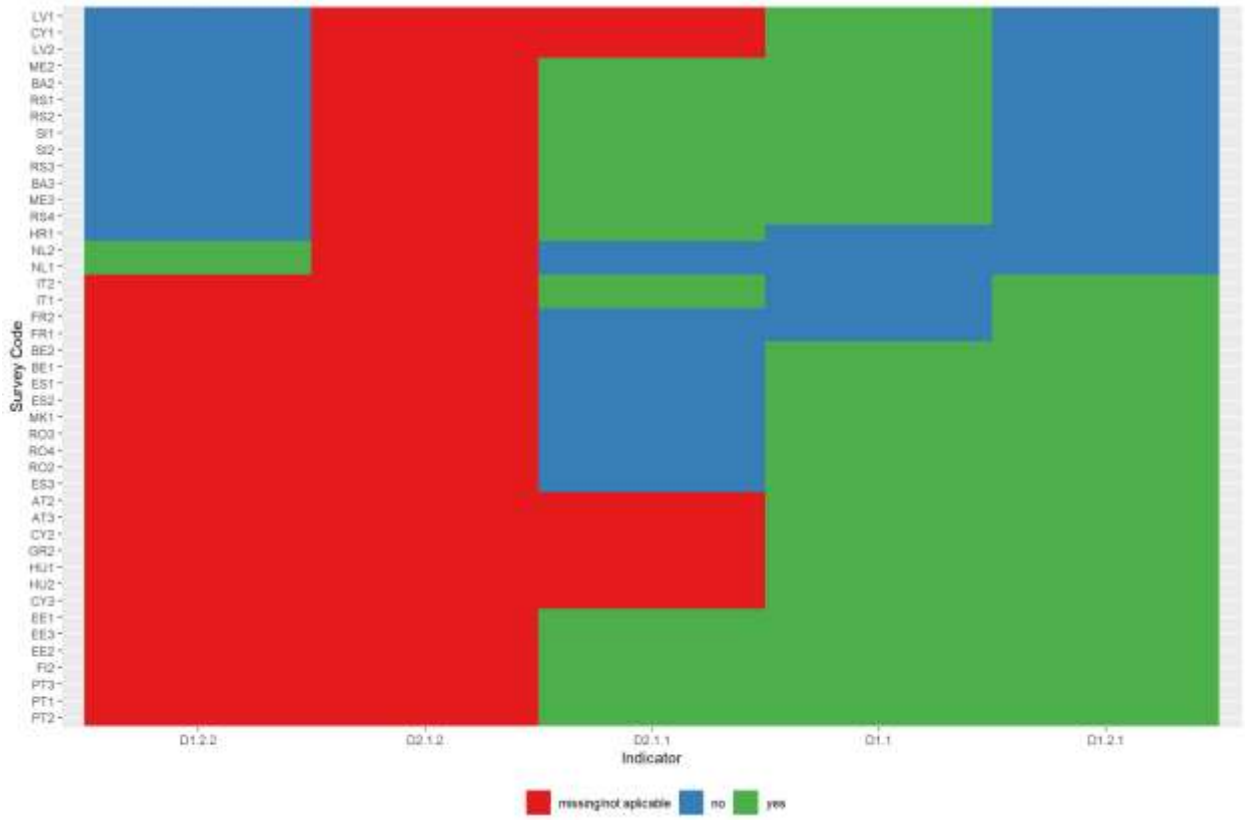

Figure A 8. Heatmap for categorical indicators of Dimension D. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

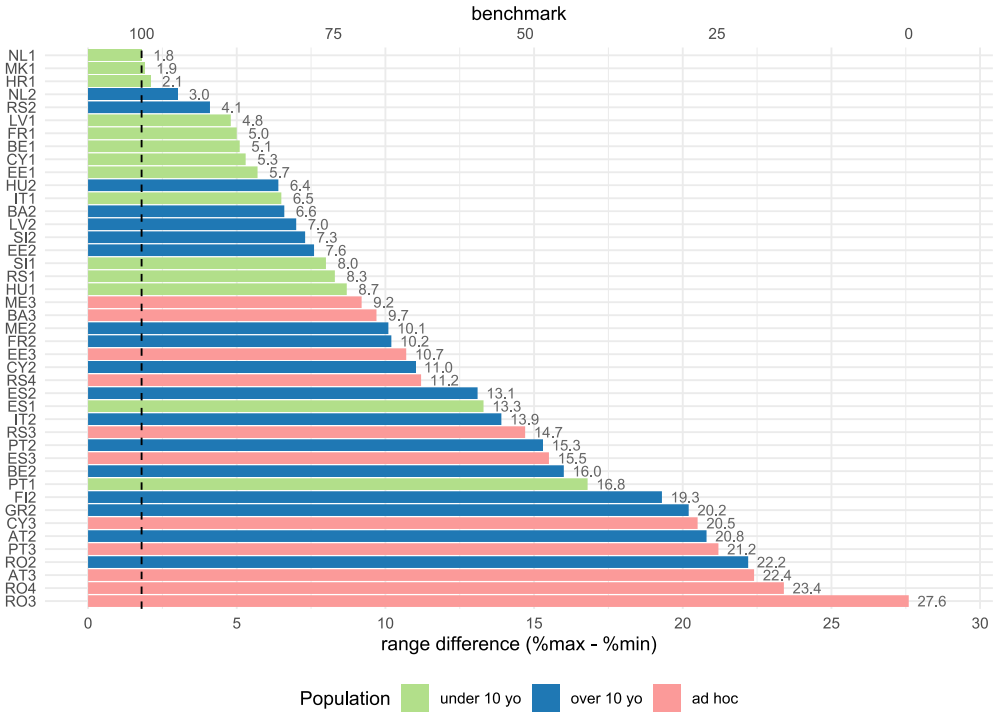

Figure A 9. Indicator D2.3.1 – Distribution of dietary data by weekdays – Benchmark: Difference between weekday with higher proportion of interviews and season with lower proportion per survey.

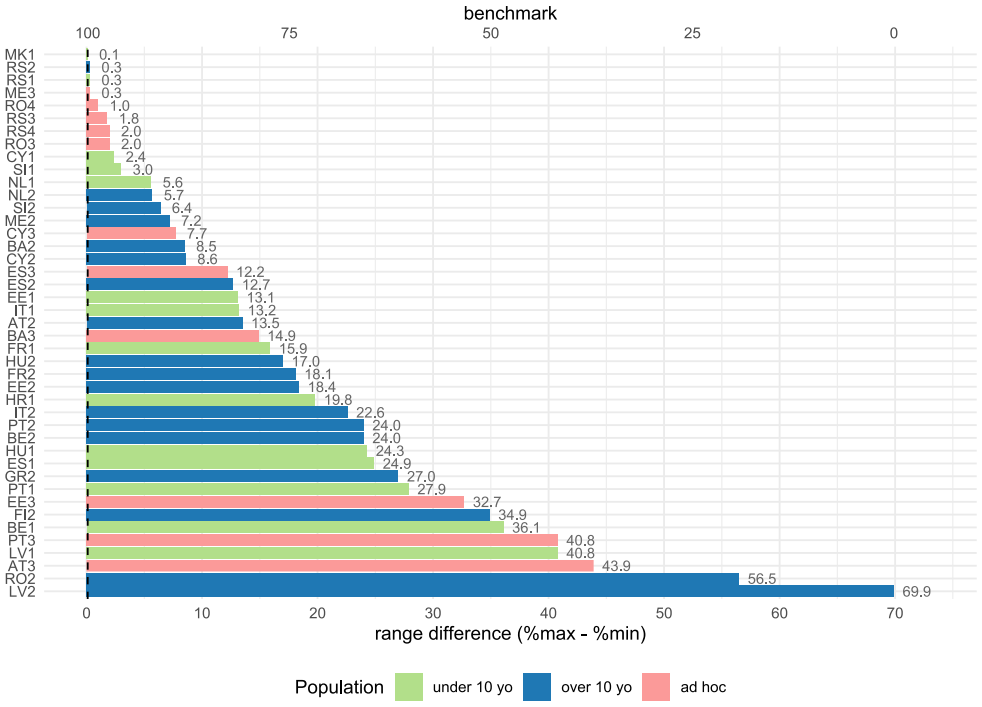

Figure A 10. Indicator D2.3.2 – Distribution of dietary data by season – Benchmark: Difference between season with higher proportion of interviews and season with lower proportion per survey.

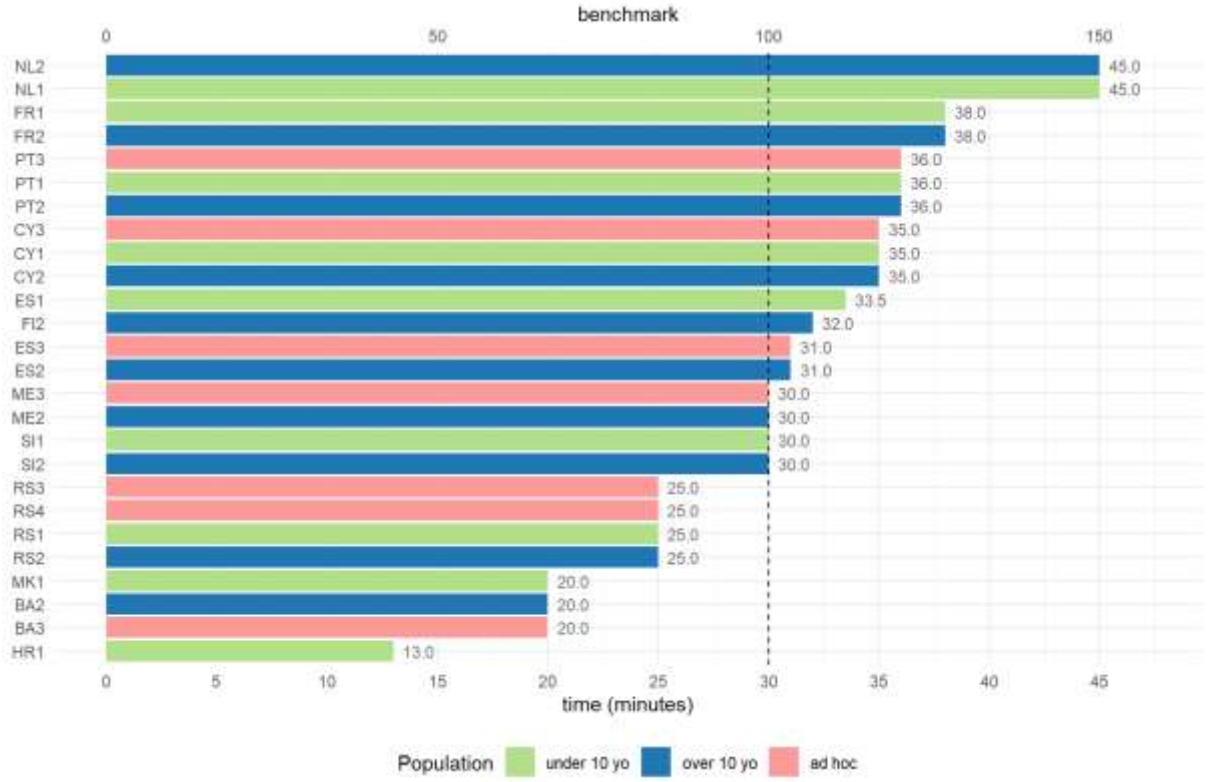

Figure A 11. Indicator D2.4.1 Duration of 24-hours recall.

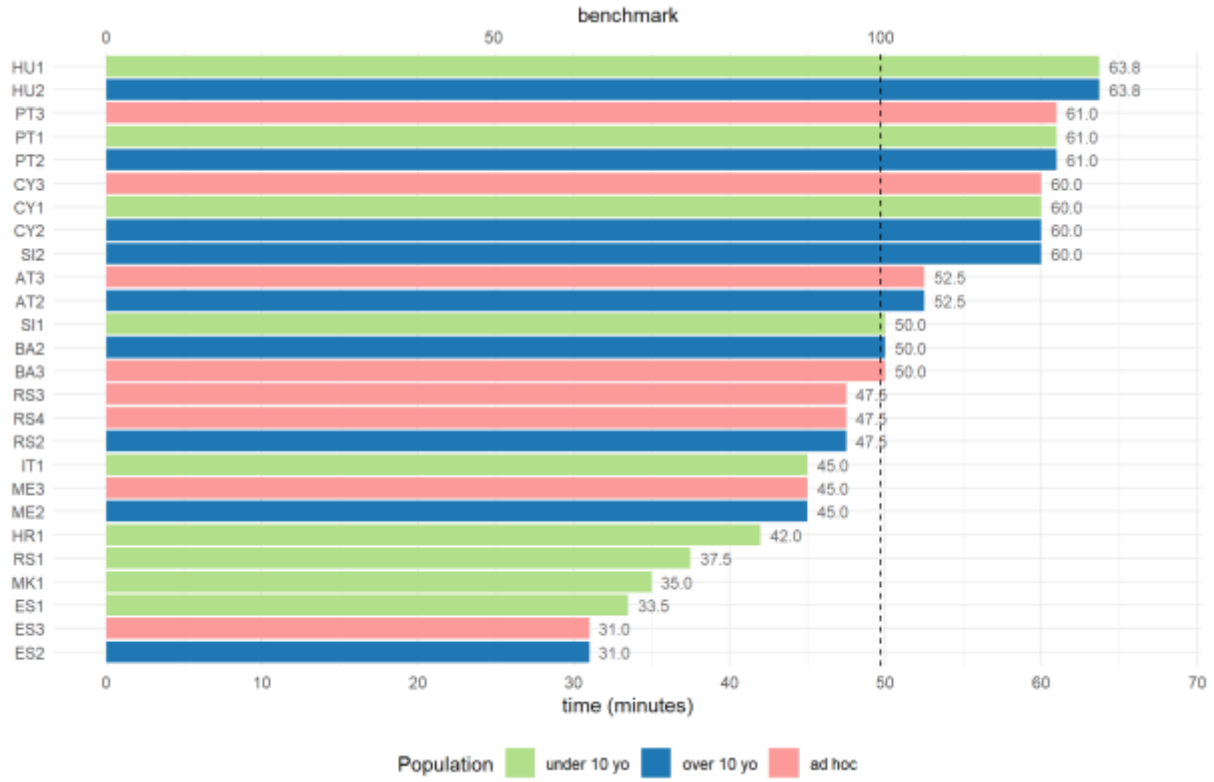

Figure A 12. Indicator D2.4.2 Total interview duration.

Dimension E – Data processing: software tools and validation procedures

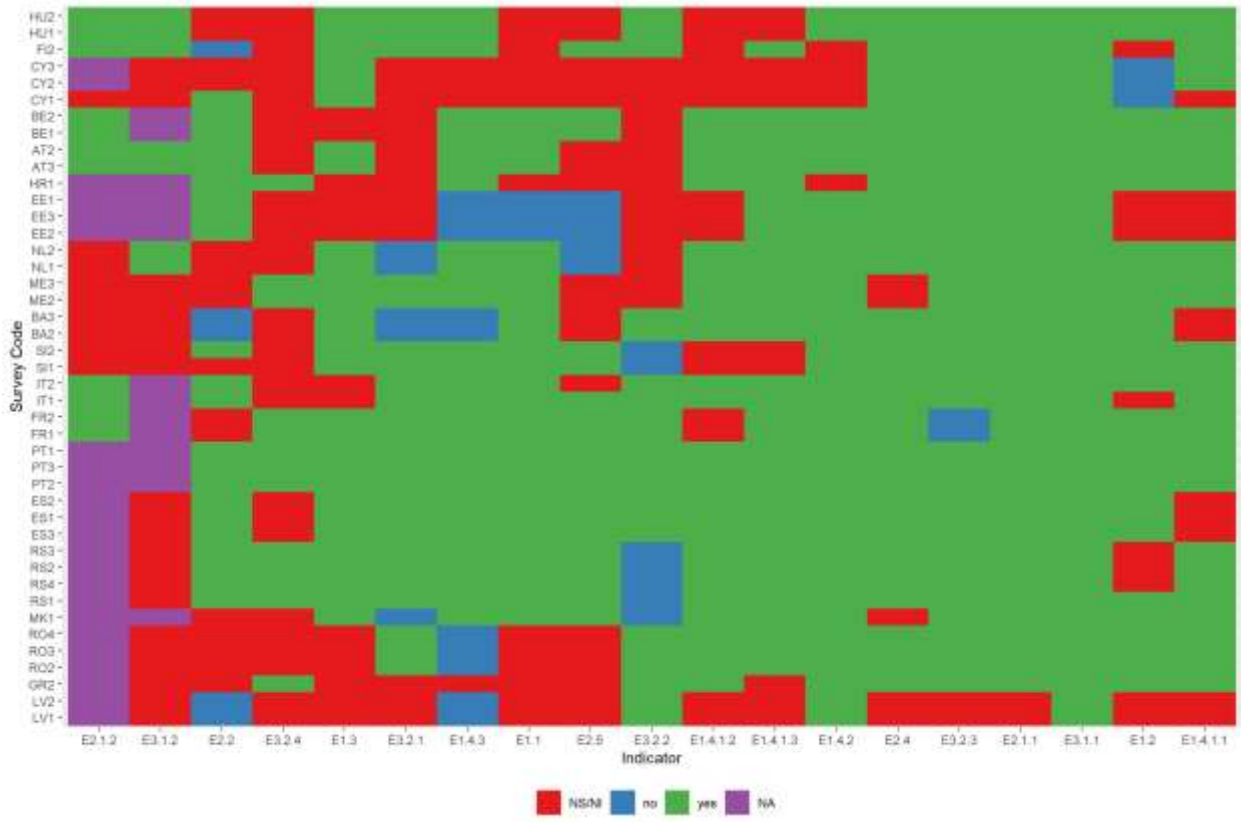

Figure A 13. Heatmap for categorical indicators of Dimension E. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

Dimension F – Non-dietary data collection

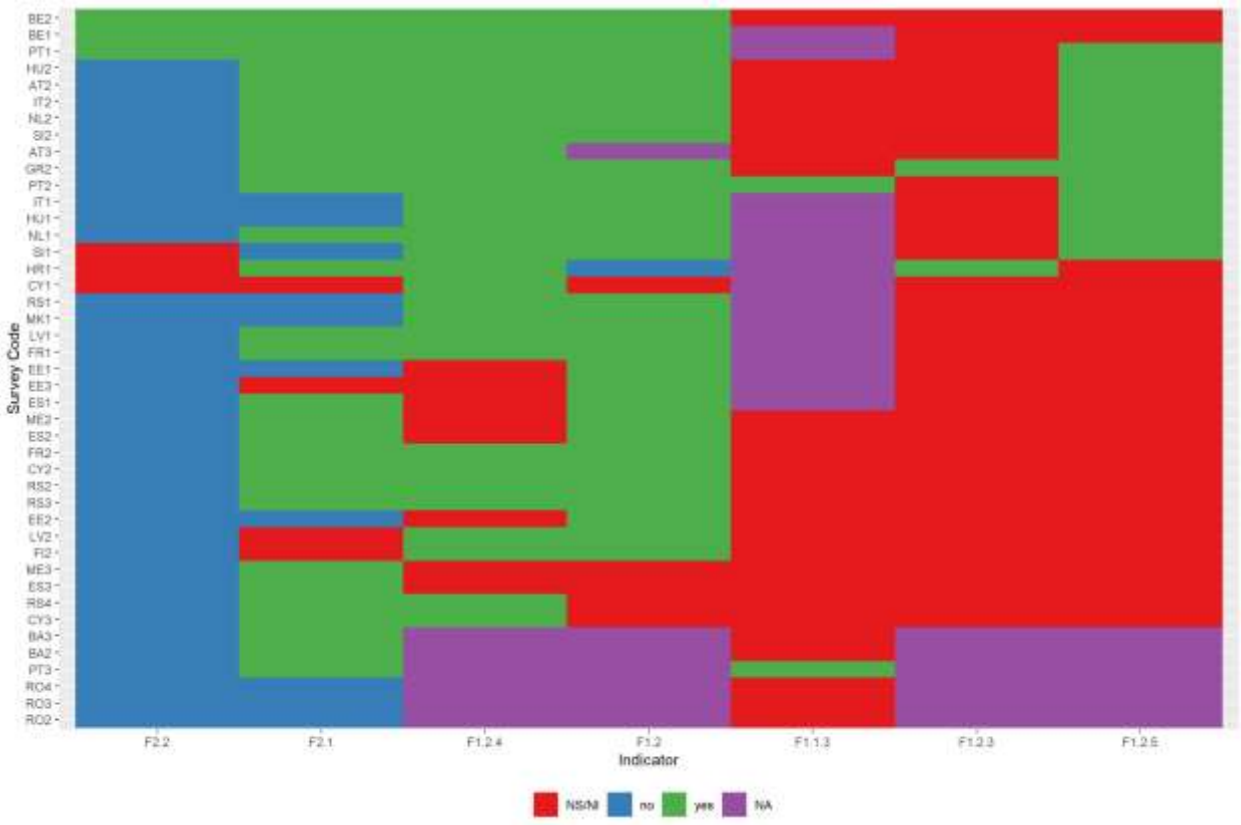

Figure A 14. Heatmap for categorical indicators of Dimension F. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

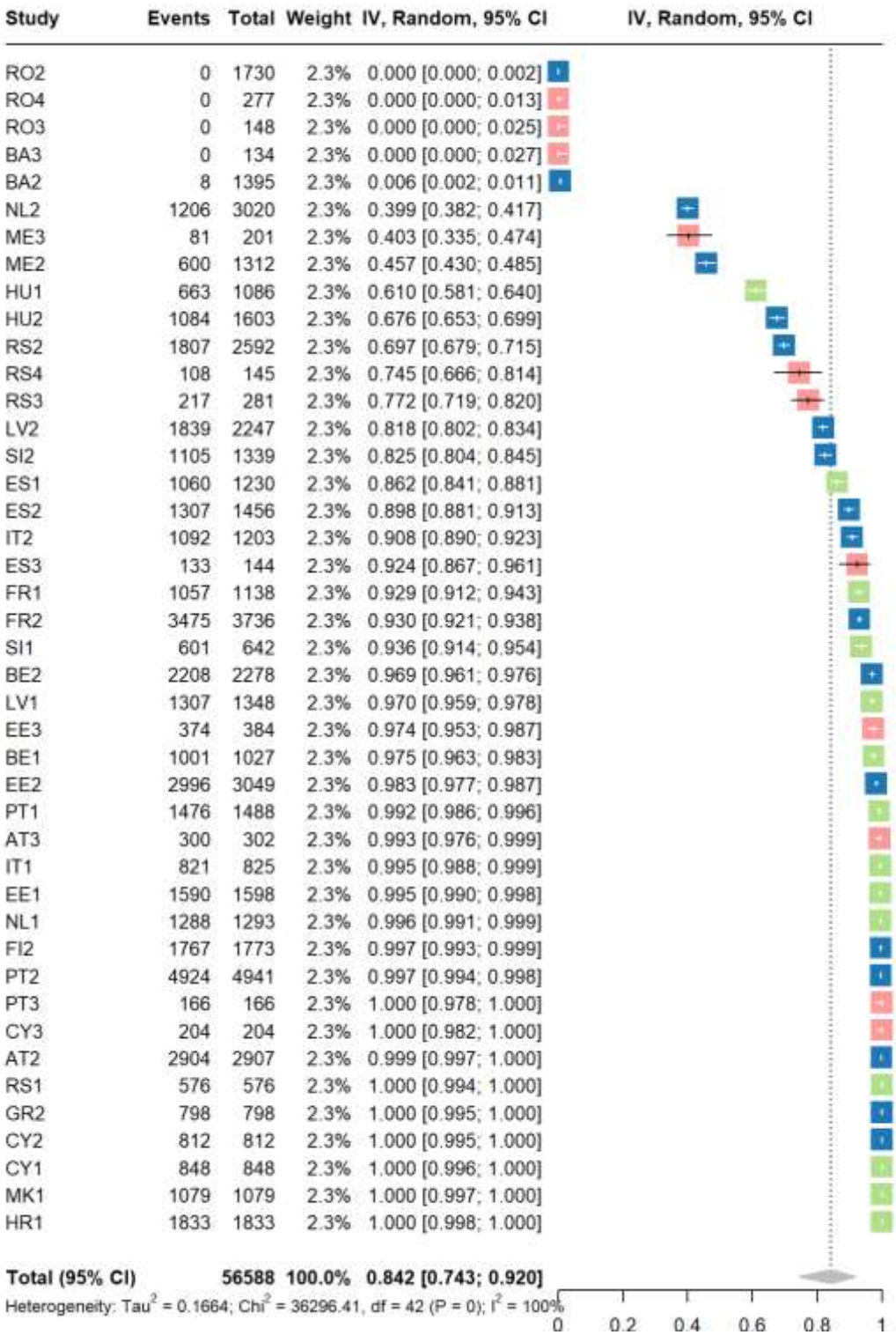

Figure A 15. Indicator F1.1 Percentage of individuals with anthropometric measurements: **weight.**

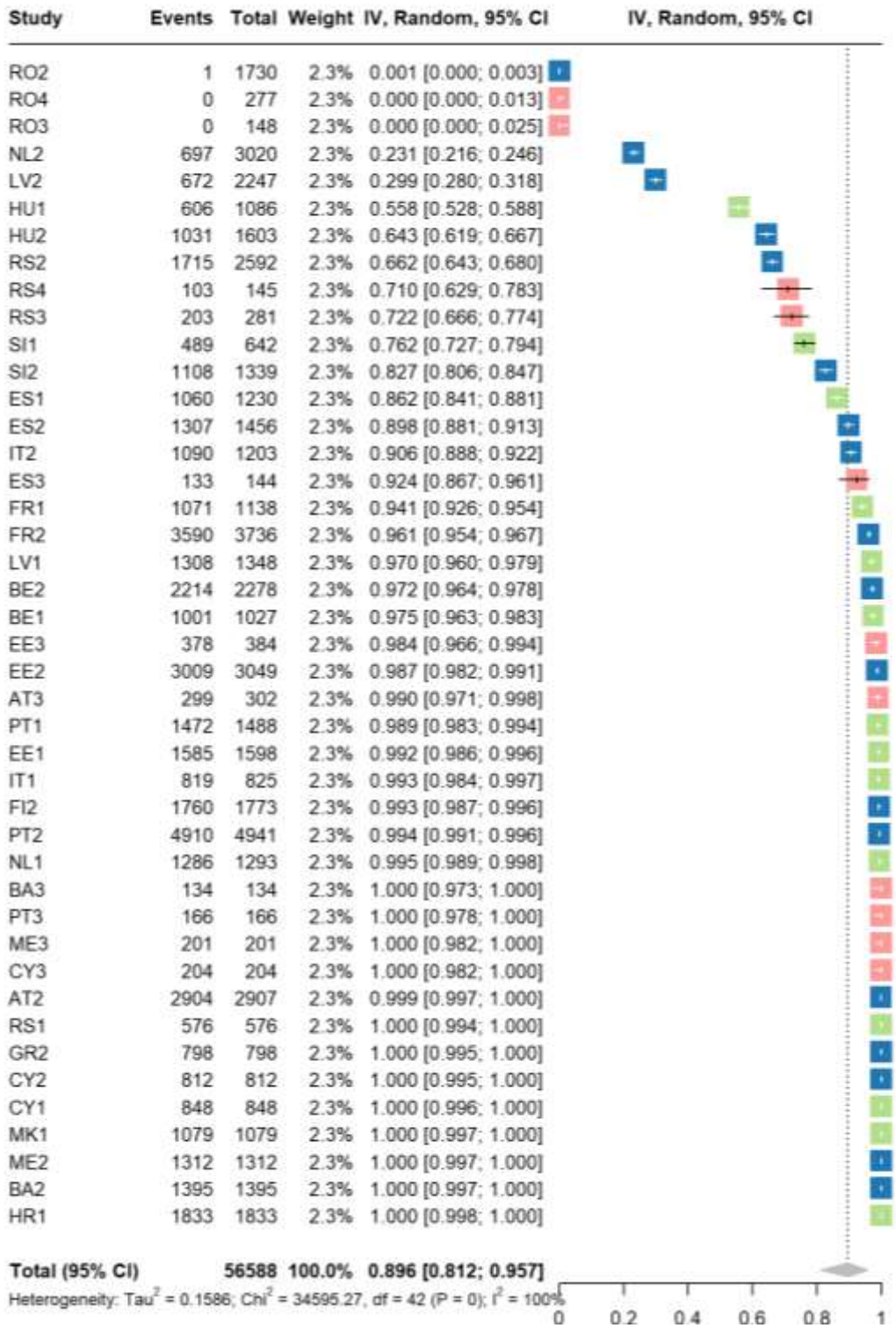

Figure A 16. Indicator F1.1 Percentage of individuals with anthropometric measurements: **height.**

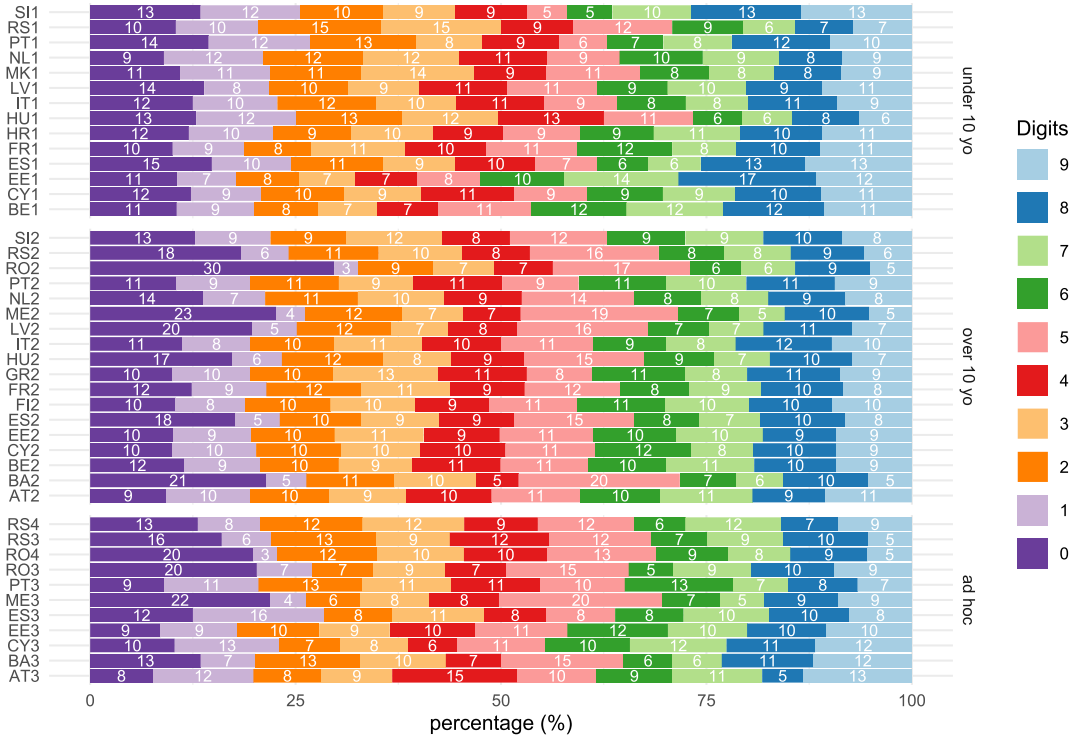

Figure A 17. Indicator F1.4 Digit preference in anthropometric measurements: **weight.** Distribution per digits.

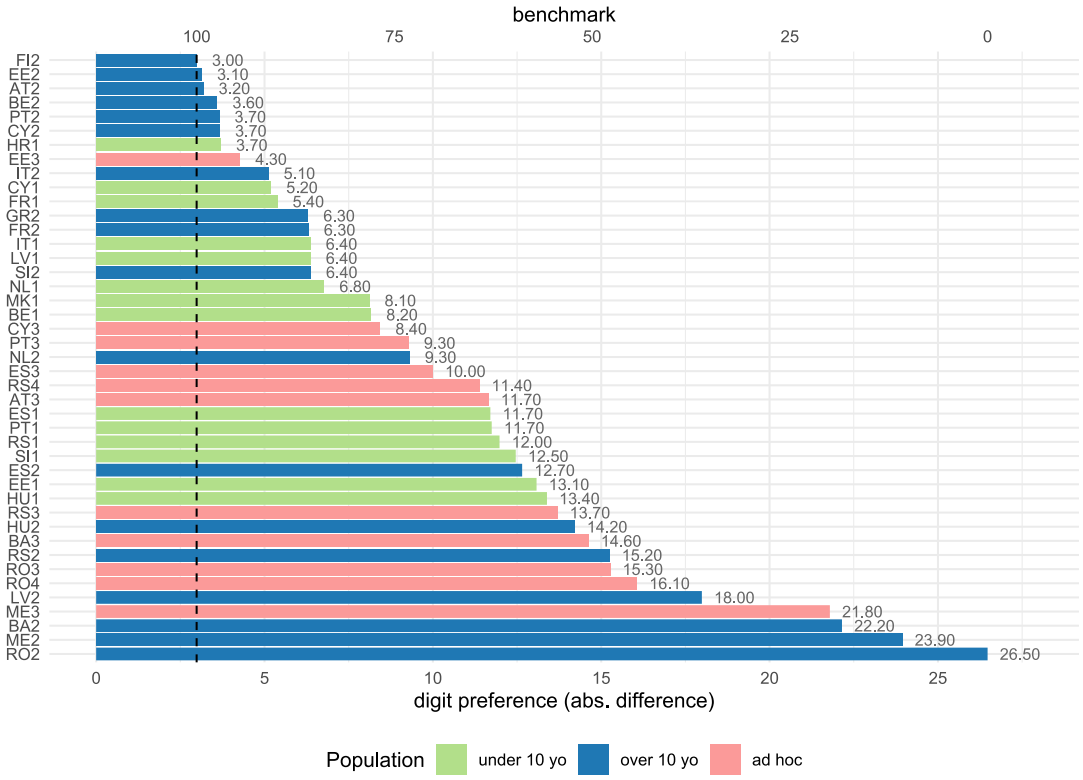

Figure A 18. Indicator F1.4 Digit preference in anthropometric measurements: **weight.** Benchmark.

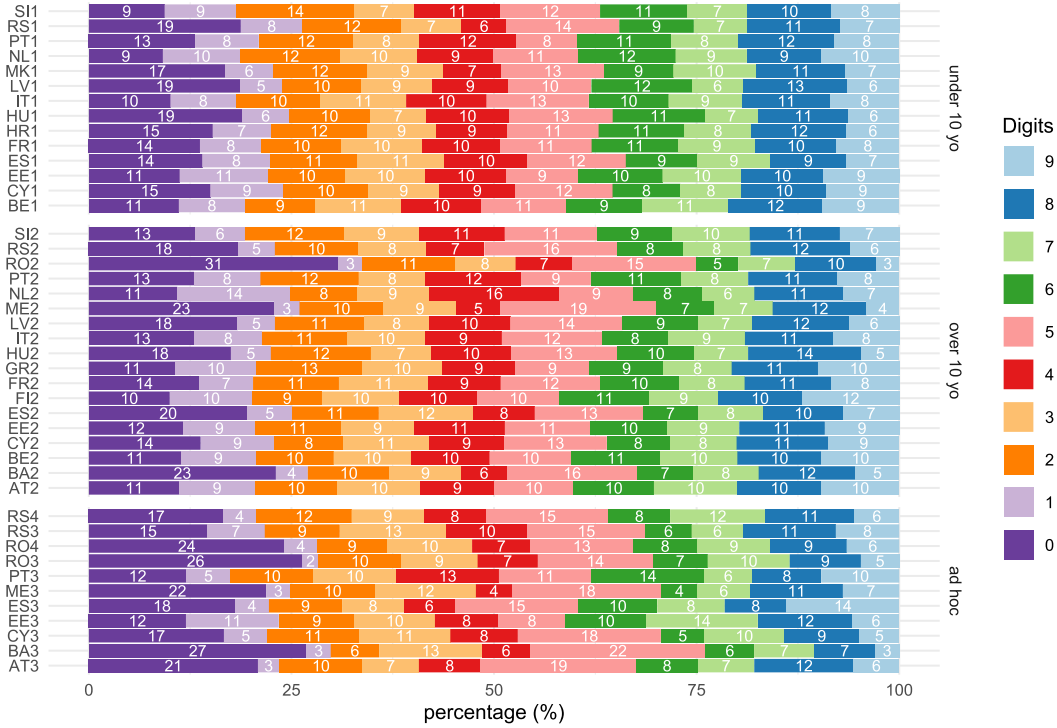

Figure A 19. Indicator F1.4 Digit preference in anthropometric measurements: **height.** Distribution per digits.

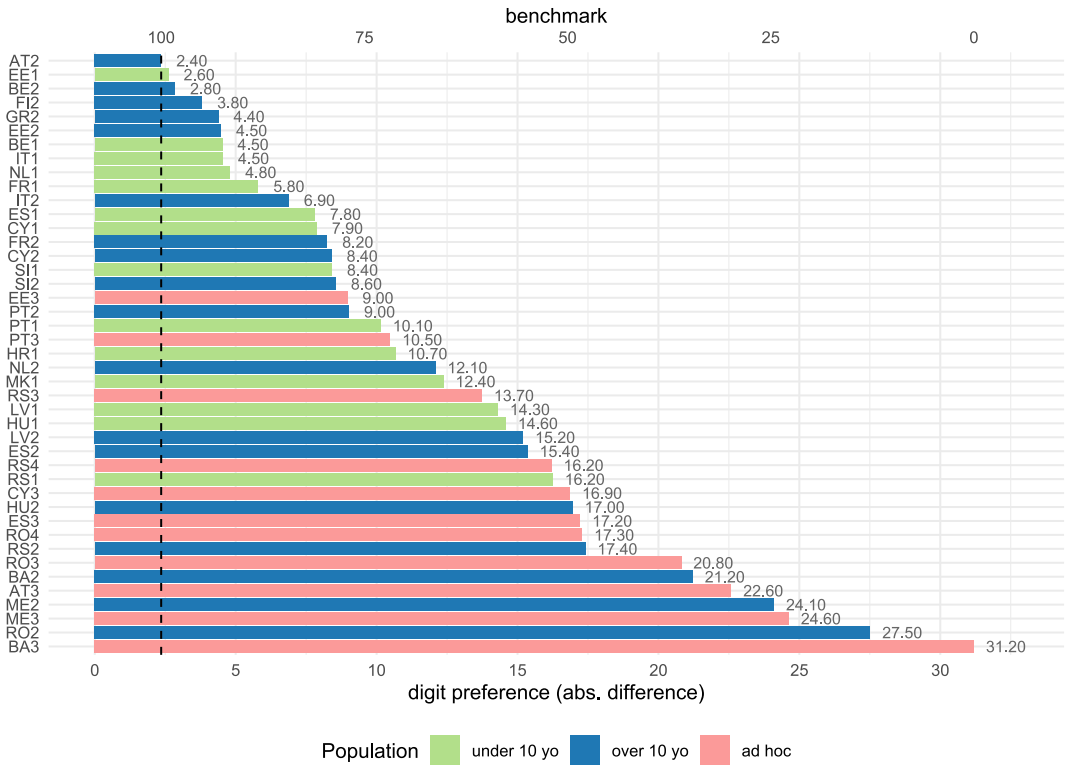

Figure A 20. Indicator F1.4 Digit preference in anthropometric measurements: **height.** Benchmark

Dimension G – Data completeness

Table A 1. Summary statistics for categorical indicators of Dimension G – Data completeness

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
G1.2.1	Availability of a food supplements database	no	1	7.1	1	5.6	0	0.0
		yes	13	92.9	17	94.4	11	100.0
G1.2.2	Food supplements quantified in grams	no	1	7.1	1	5.6	0	0.0
		yes	7	50.0	8	44.4	5	45.5
		missing	*6*	*42.9*	*9*	*50.0*	*6*	*54.5*
G1.3.1	Availability of a recipe composition database	no	1	7.1	1	5.6	0	0.0
		yes	13	92.9	17	94.4	11	100.0
G1.3.2	Recipe database updated during fieldwork	no	0	0.0	0	0.0	0	0.0
		yes	11	78.6	14	77.8	9	81.8
		missing	*3*	*21.4*	*4*	*22.2*	*2*	*18.2*

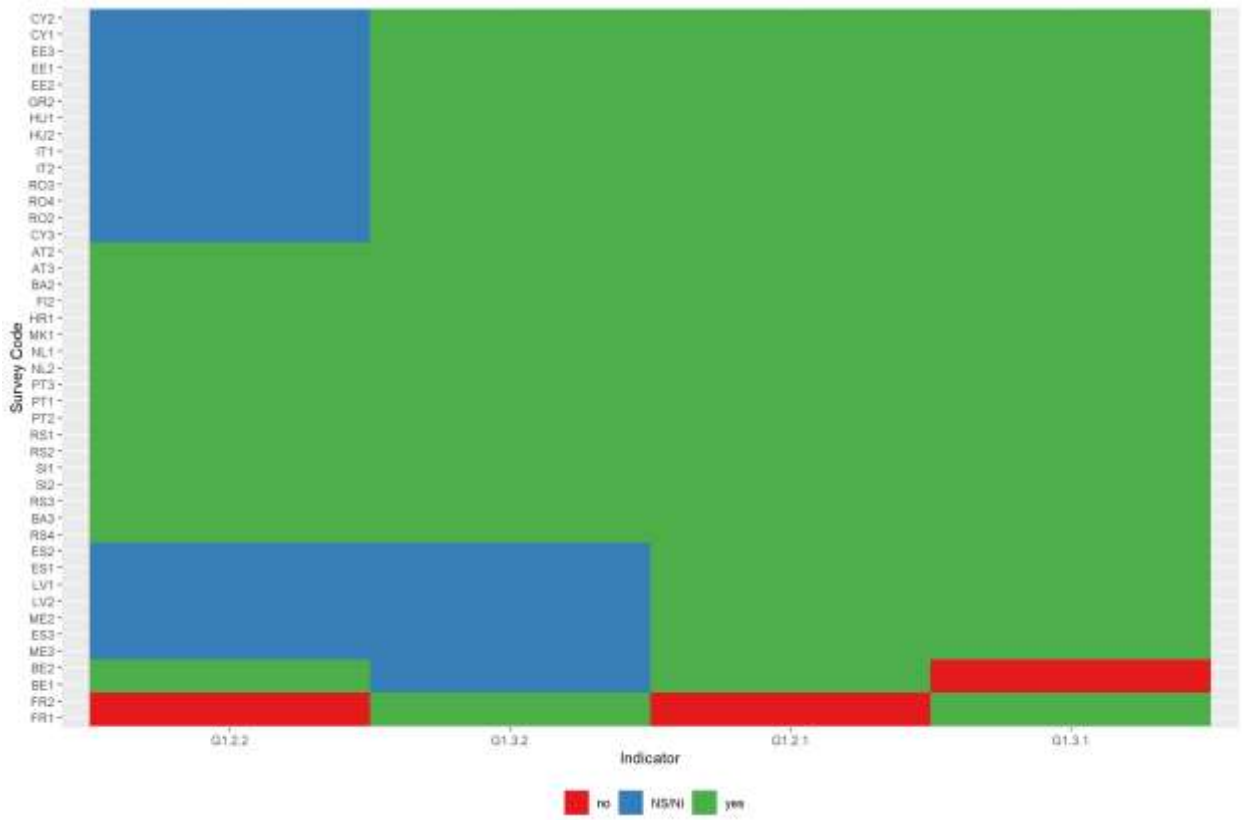

Figure A 21. Heatmap for categorical indicators of Dimension G. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

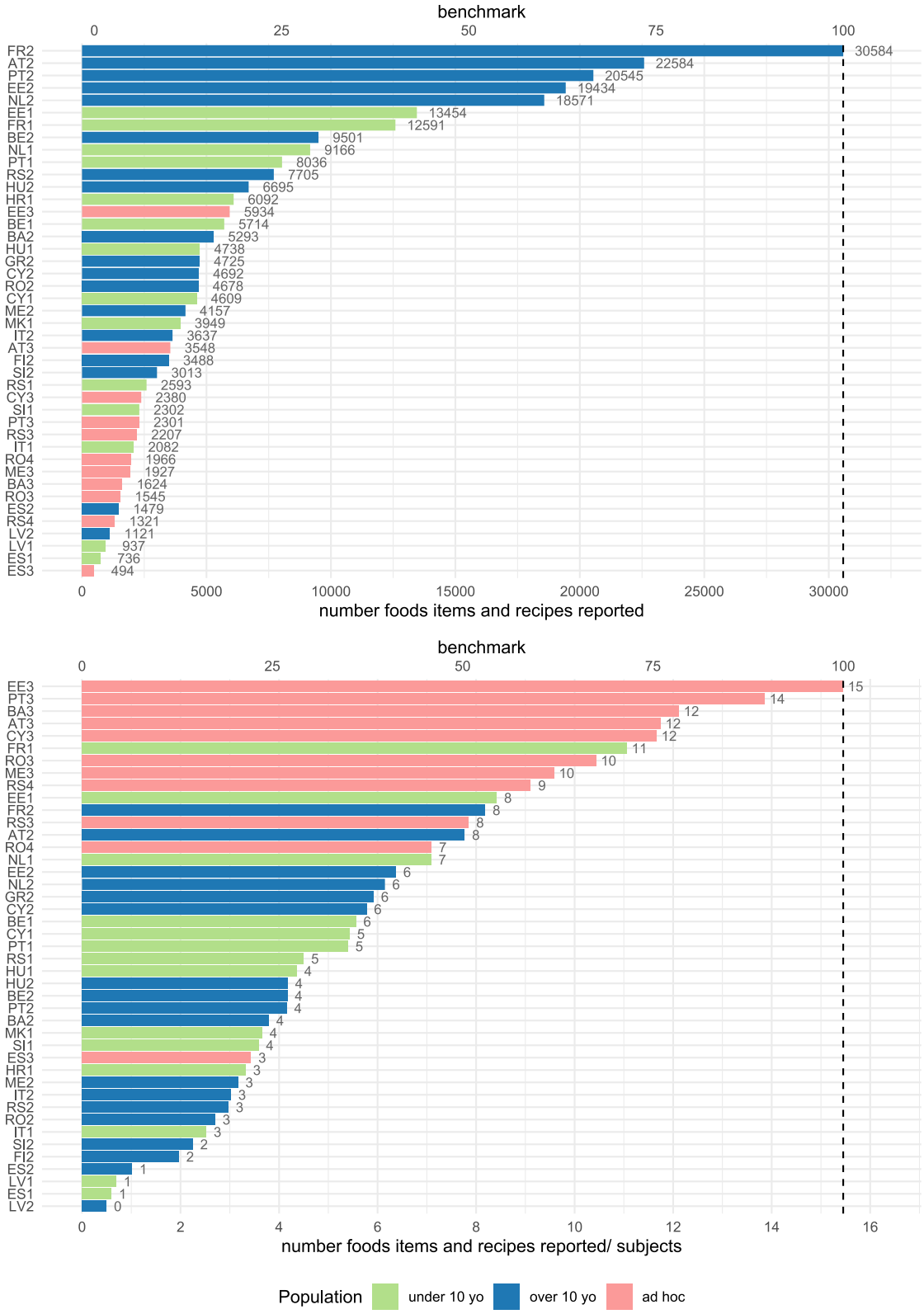

Figure A 22. Indicator G1.1 Number of food items reported: total and per subjects. ***Note:*** *This indicator was computed based on the unique list of foods consumed, including all combinations basic FoodEx2 codes+facets from the consumption file.*

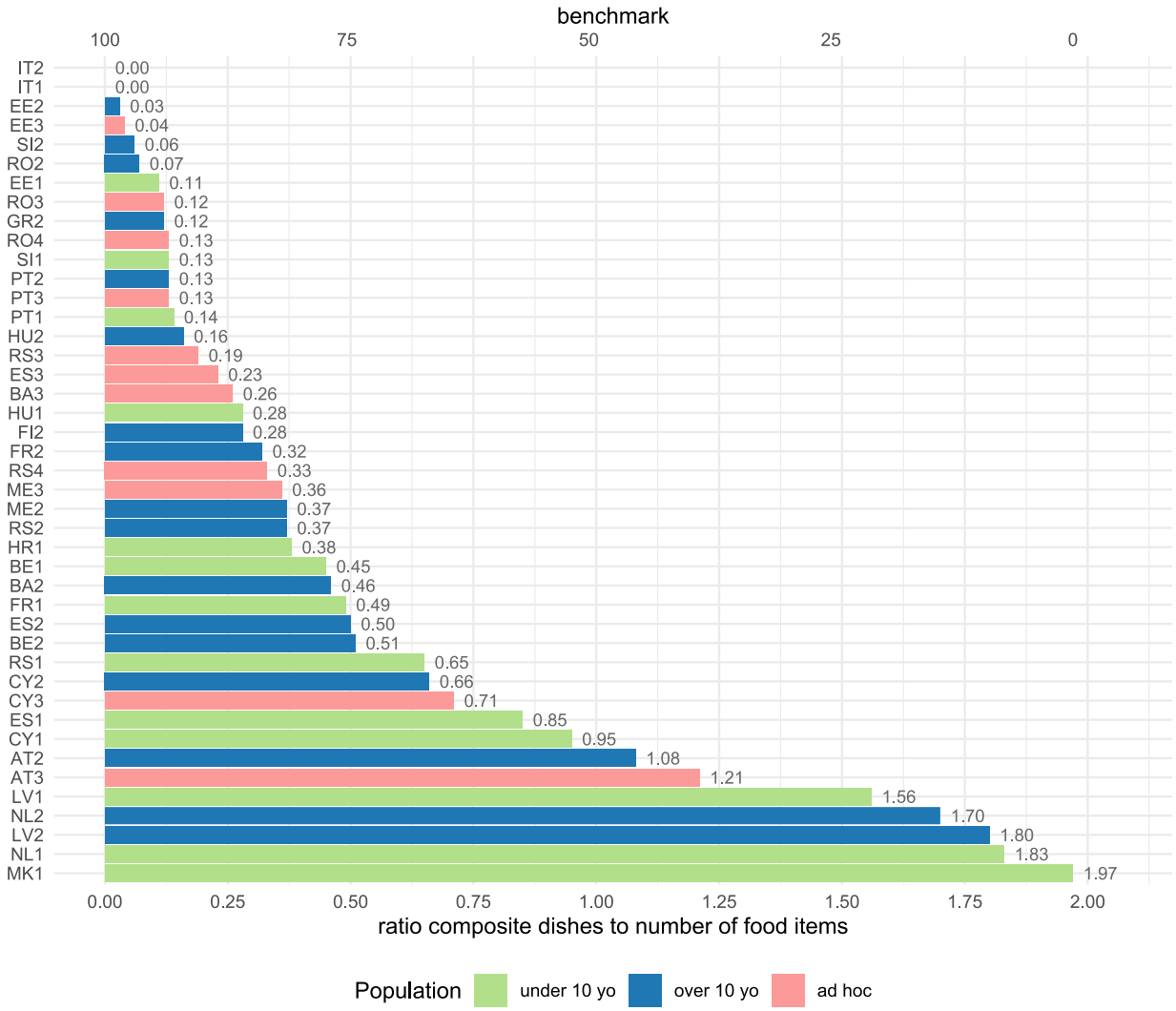

Figure A 23. Indicator G2.1.1 – Percentage of composite dishes in FoodEx2 codes. Note: The composite dishes were identified as the single food items reported within the consumption dataset classified with a FoodEx2 code corresponding to the "Composite dishes" node of the *FoodEx2 Exposure Hierarchy*. The EU Menu methodology implies a disaggregation of recipes (i.e., "composite dishes") into single ingredients. Thus, a very low proportion of food items from this FoodEx2 hierarchical group were expected. Indeed, this plot shows a very low proportion of these foods reported in the Consumption datasets (range: 0 – 1.97%).

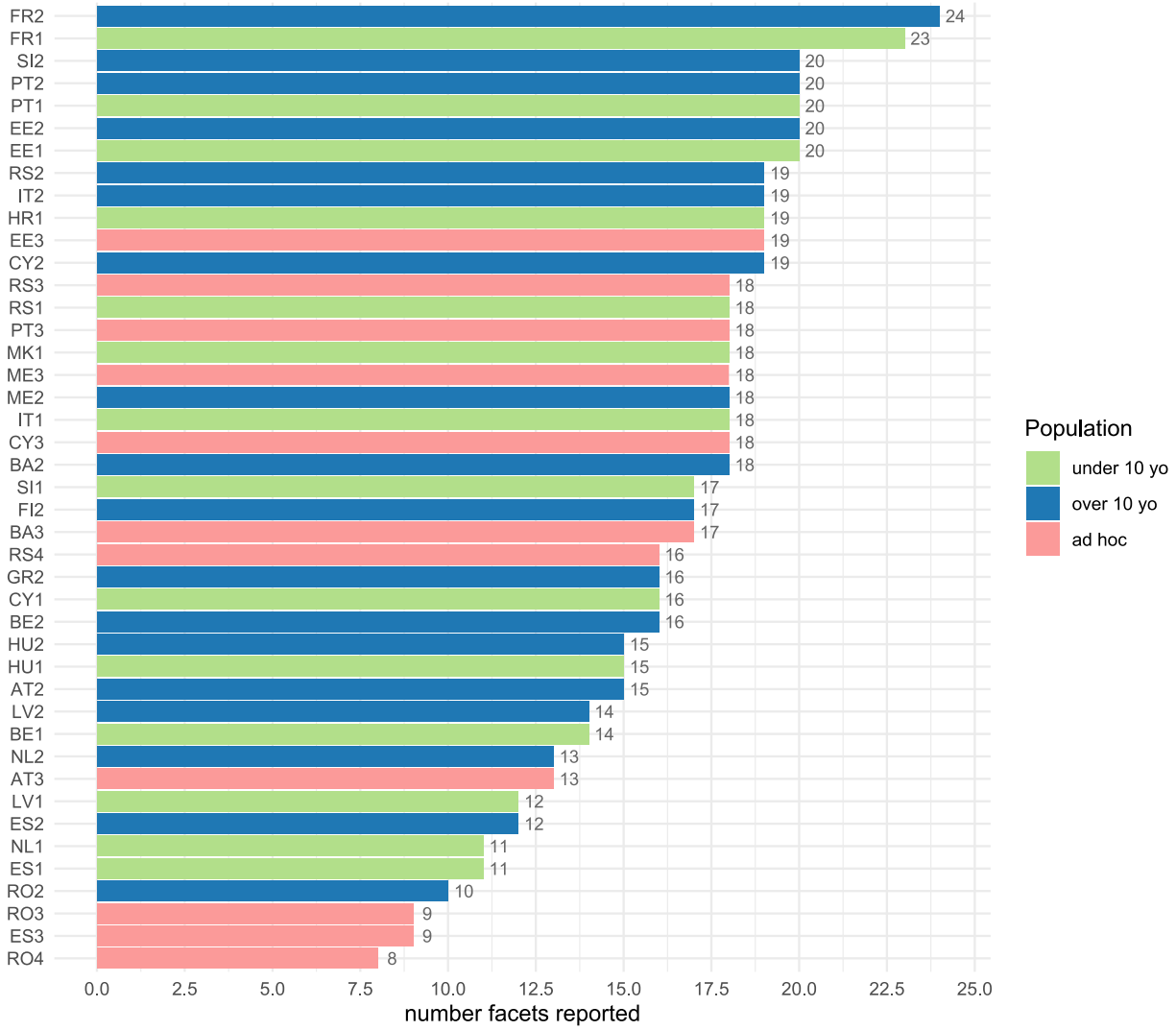

Figure A 24. Indicator G2.3 – Total number of facets used, calculated using the consumed FoodEx2 codes, without implicit facets.

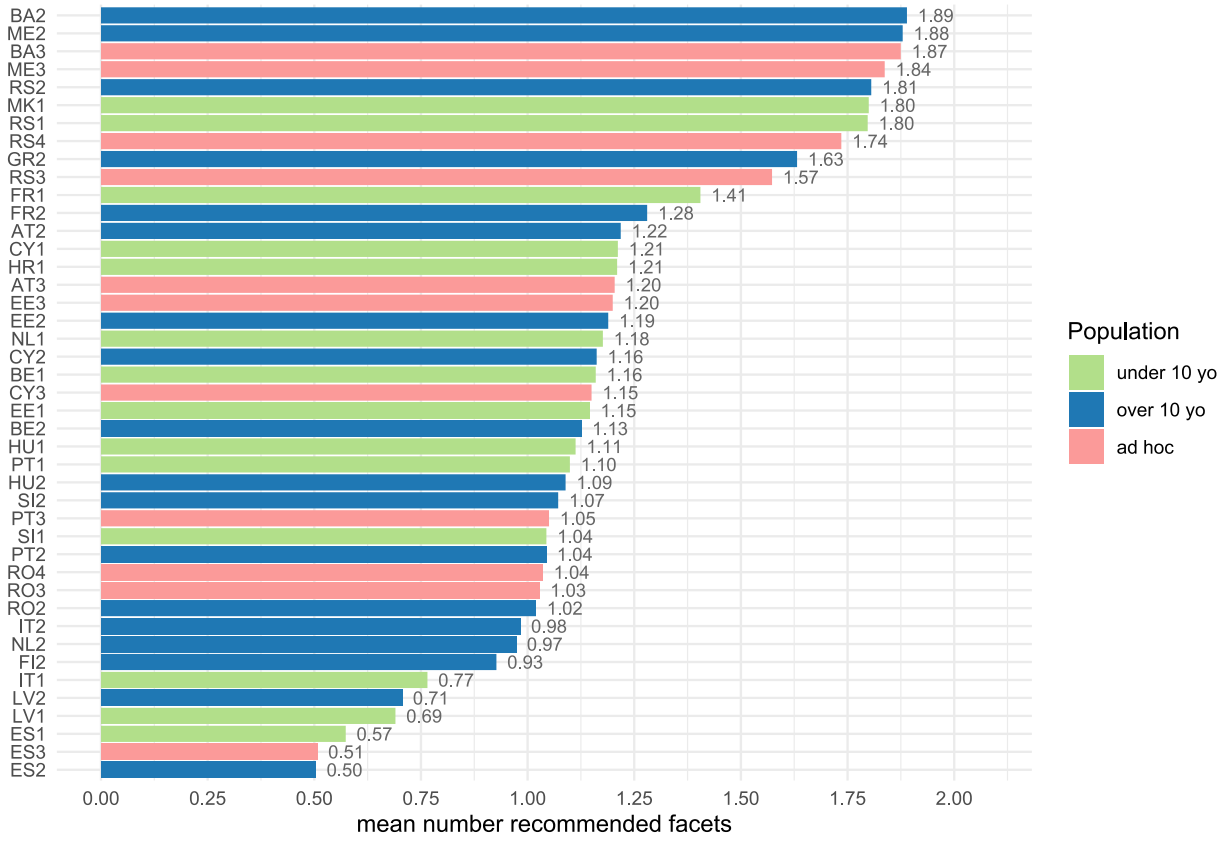

Figure A 25. Indicator G2.4 – Minimum recommended facets: average number of recommended facets used per survey, calculated using the consumed FoodEx2 codes, without implicit facets.

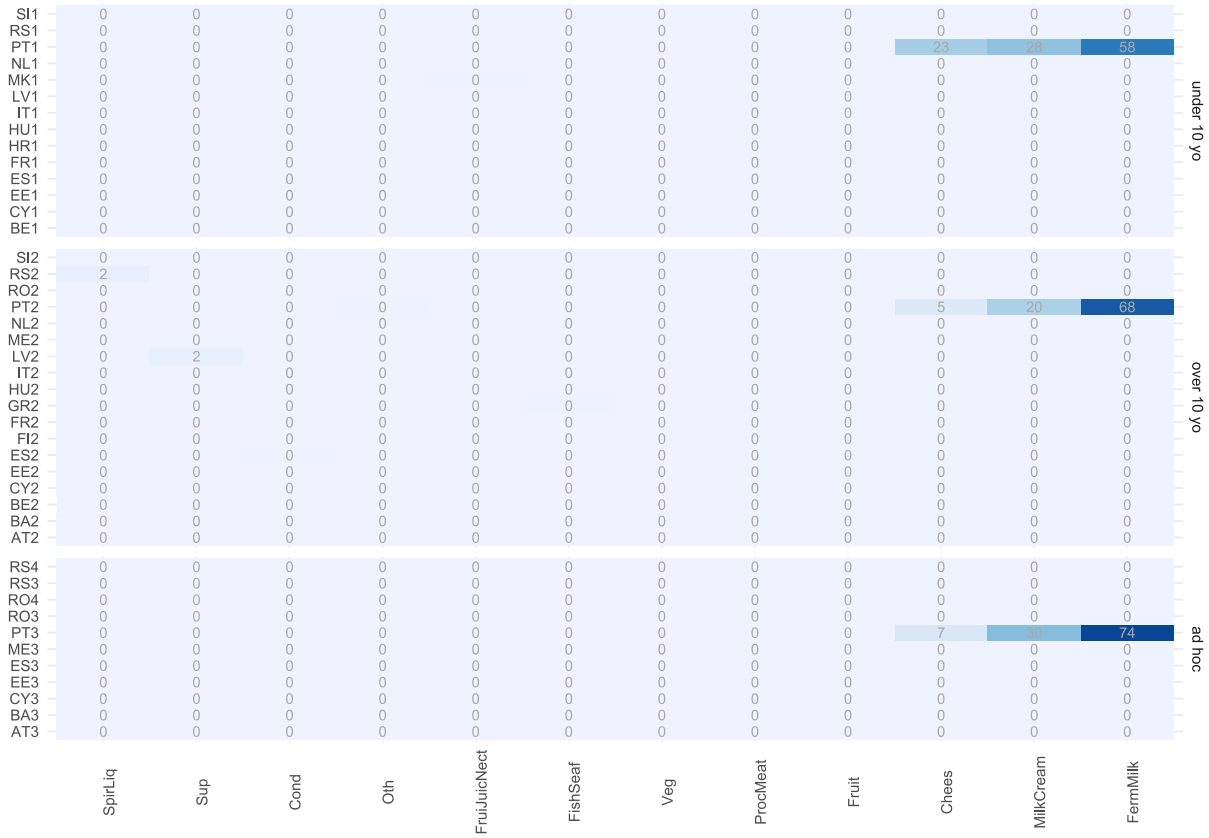

Figure A 26. Proportion of foods from different food groups reported with Facet *F01 Source*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

Figure A 27. Proportion of foods from different food groups reported with Facet *F08 Sweetening Agent*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

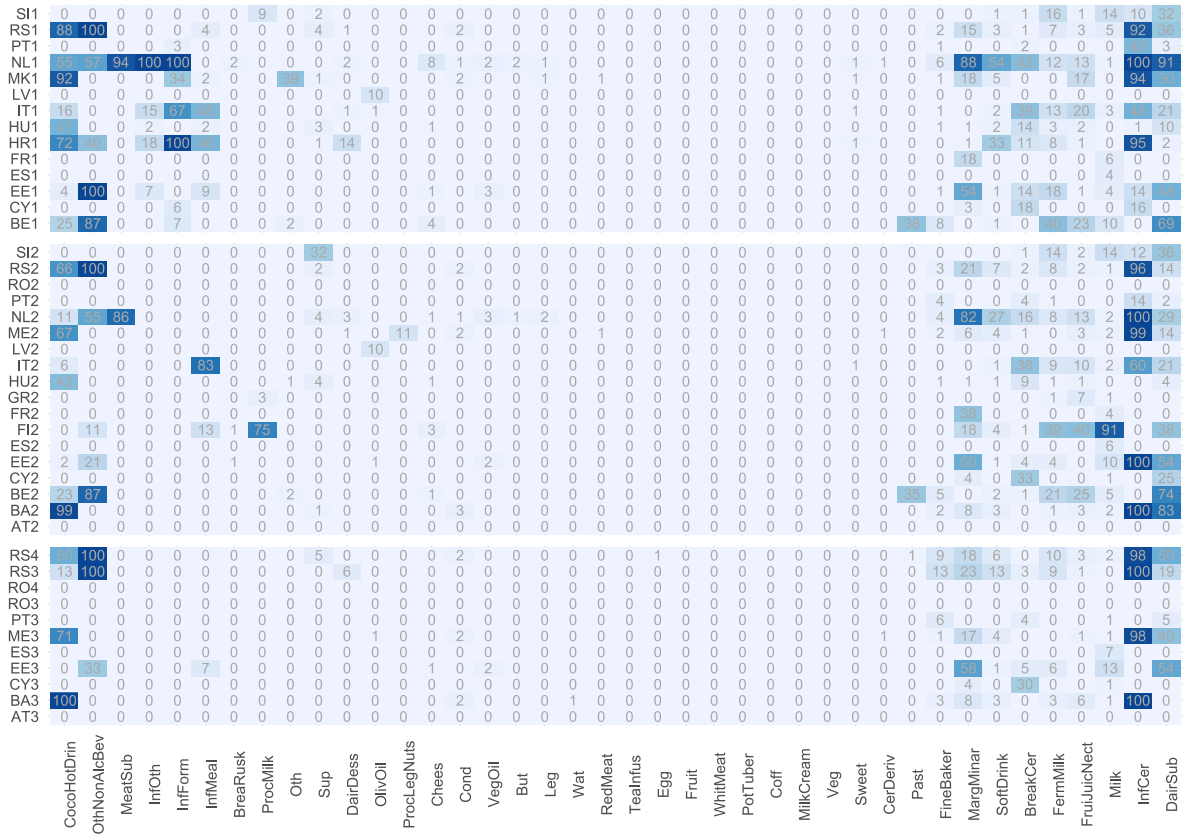

Figure A 28. Proportion of foods from different food groups reported with Facet *F09 Fortification Agent*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

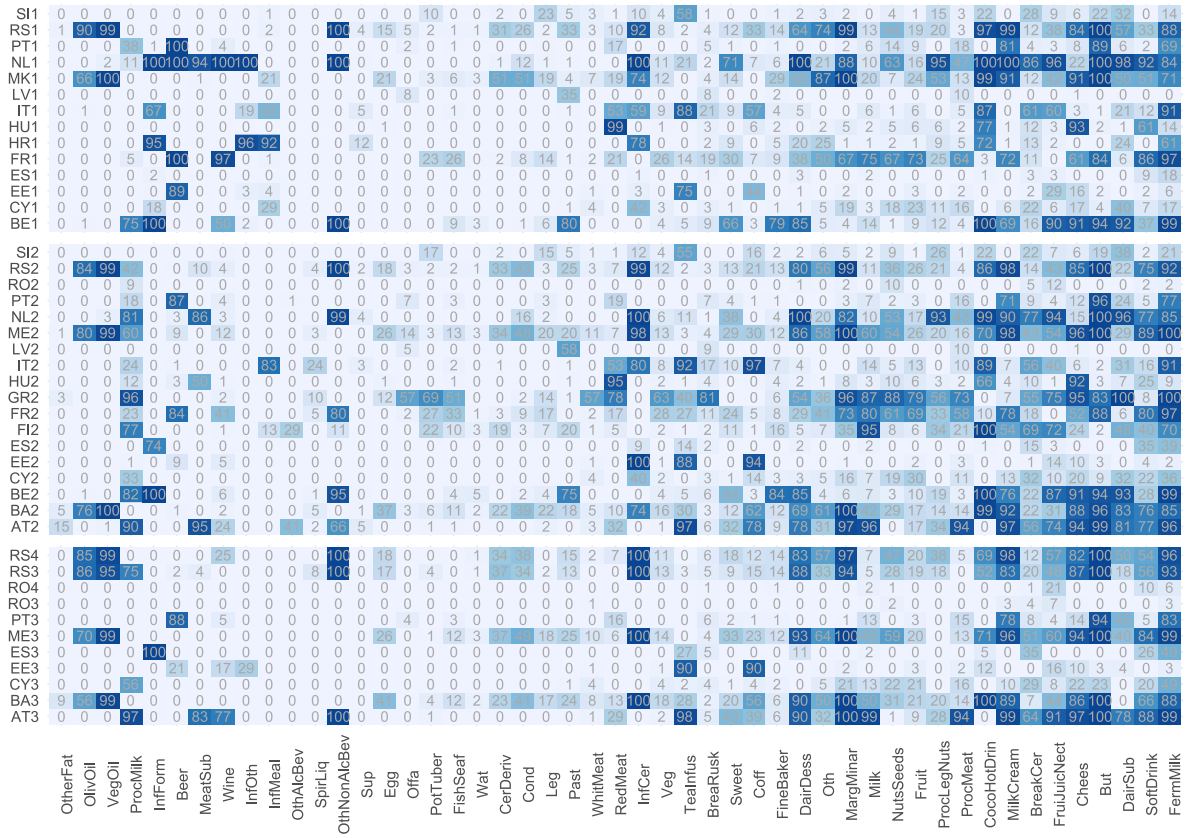

Figure A 29. Proportion of foods from different food groups reported with Facet *F10 Qualitative Information*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

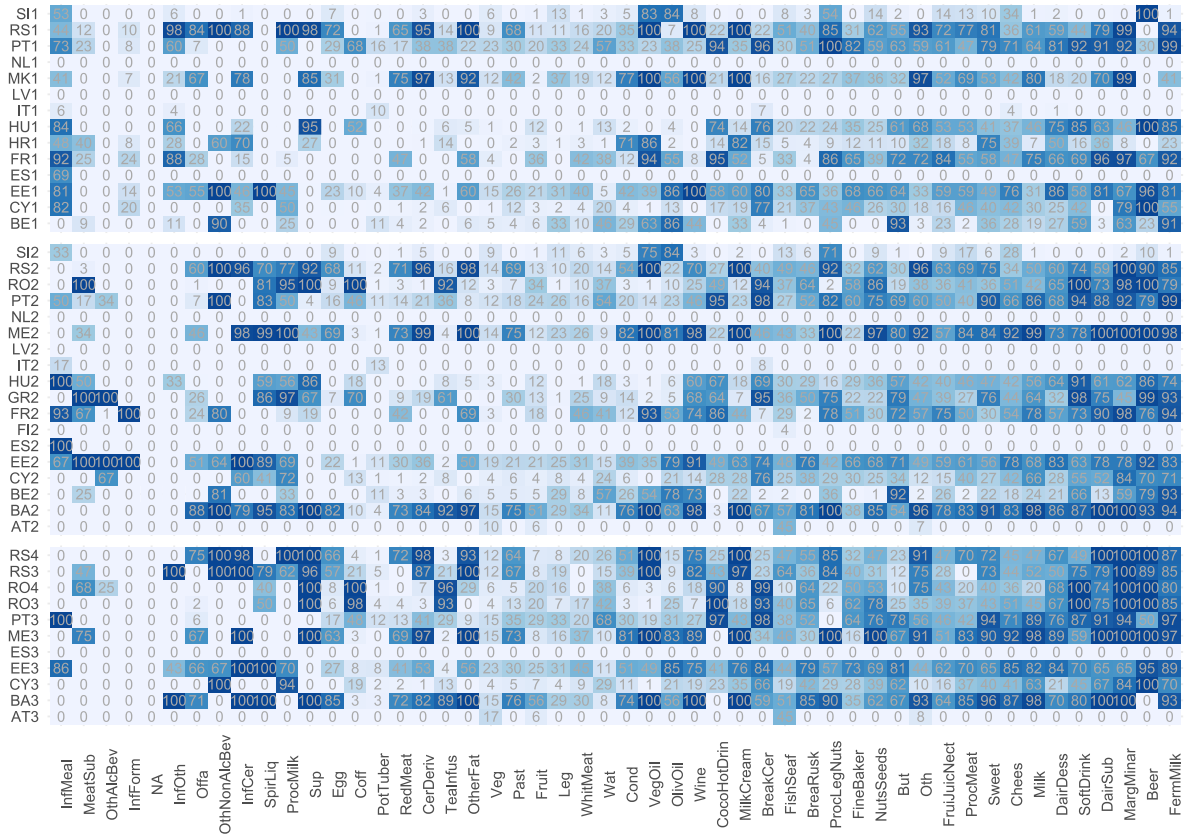

Figure A 30. Proportion of foods from different food groups reported with Facet *F19 Packaging Material*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets.

Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

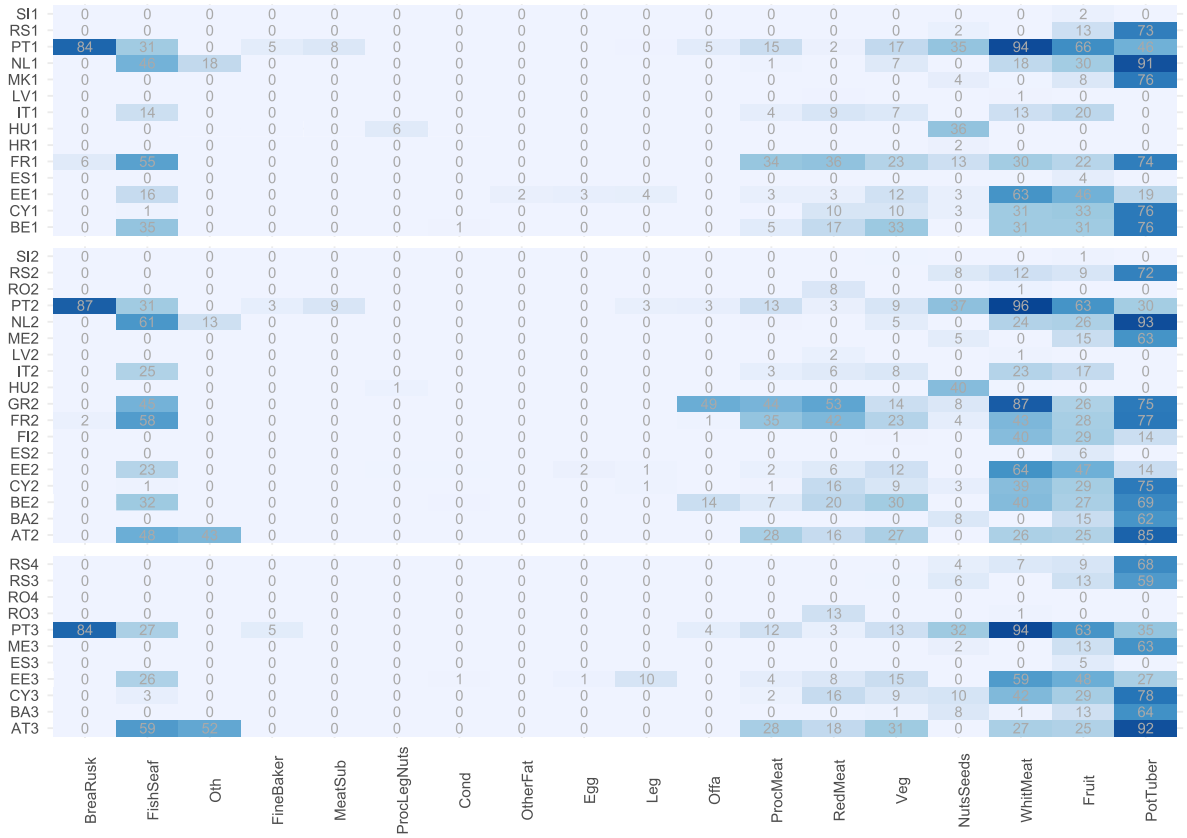

Figure A 31. Proportion of foods from different food groups reported with Facet *F20 Part-consumed-analysed*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

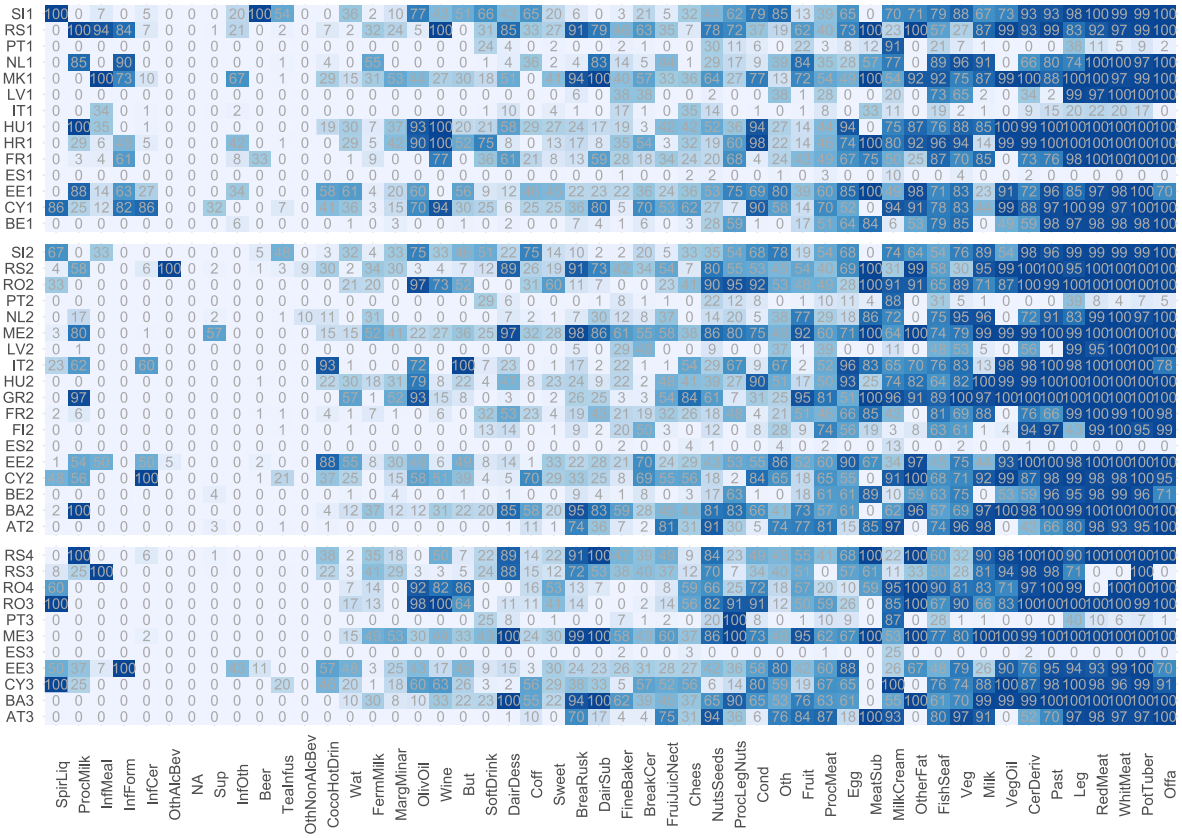

Figure A 32. Proportion of foods from different food groups reported with Facet *F28 Process*, per survey. Calculated using the consumed FoodEx2 codes, without implicit facets. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

Dimension H – Data analyses

Table A 2. Summary statistics for categorical indicators of Dimension H – Data analyses.

Indicator	Label	Value Labels	<10 years		≥10 years		Ad-hoc	
			N	%	N	%	N	%
H3.1	Availability of data, such as biomarkers, to validate dietary intake	no	10	71.4	12	66.7	8	72.7
		yes	1	7.1	3	16.7	2	18.2
		missing	*3*	*21.4*	*3*	*16.7*	*1*	*9.1*
H4.1.1	Calculation of misreporting of energy intake through recommended methods	no	0	0.0	0	0.0	1	9.1
		Goldberg et al., (1991) updated by Black (2000)	6	42.9	8	44.4	2	18.2
		other method	2	14.3	1	5.6	1	9.1
		missing	*6*	*42.9*	*9*	*50.0*	*7*	*63.6*

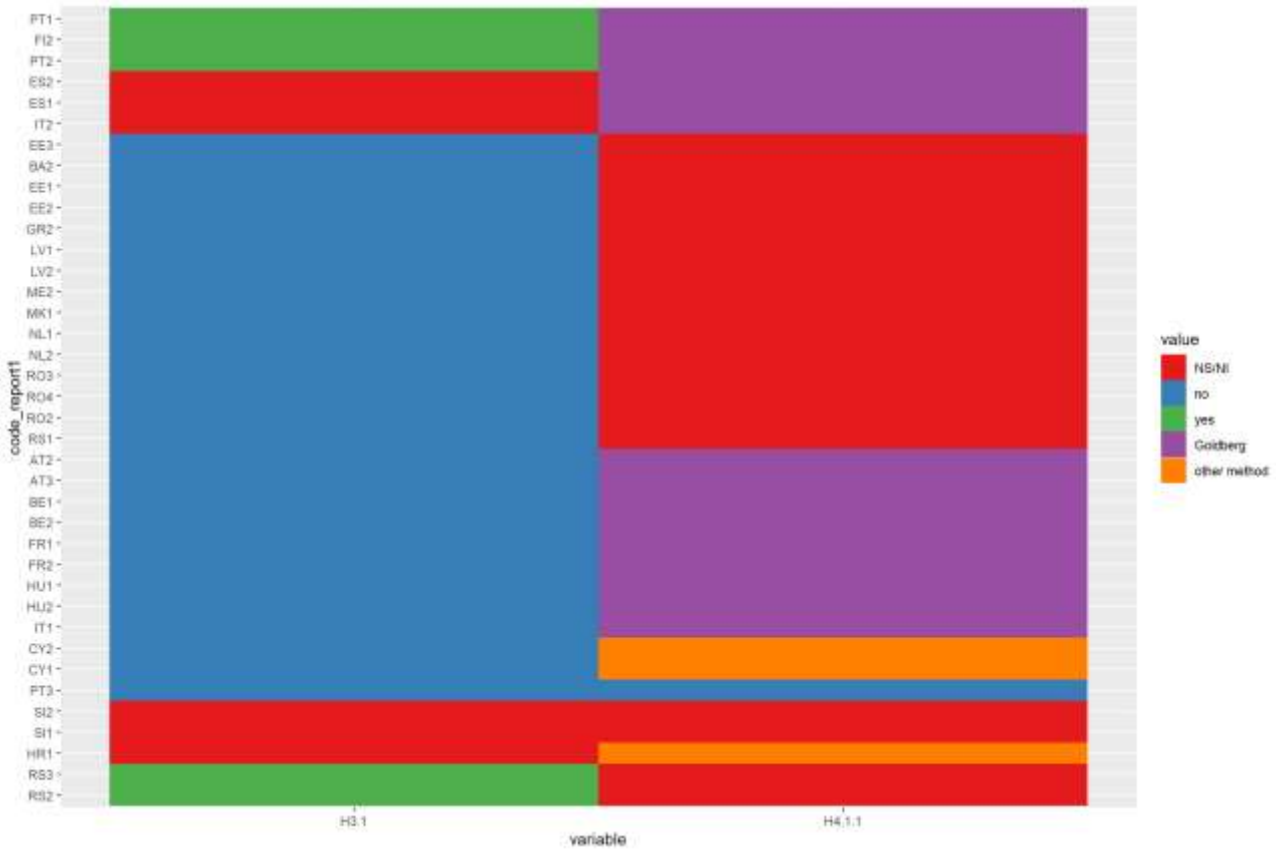

Figure A 33. Heatmap for categorical indicators of Dimension H. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).

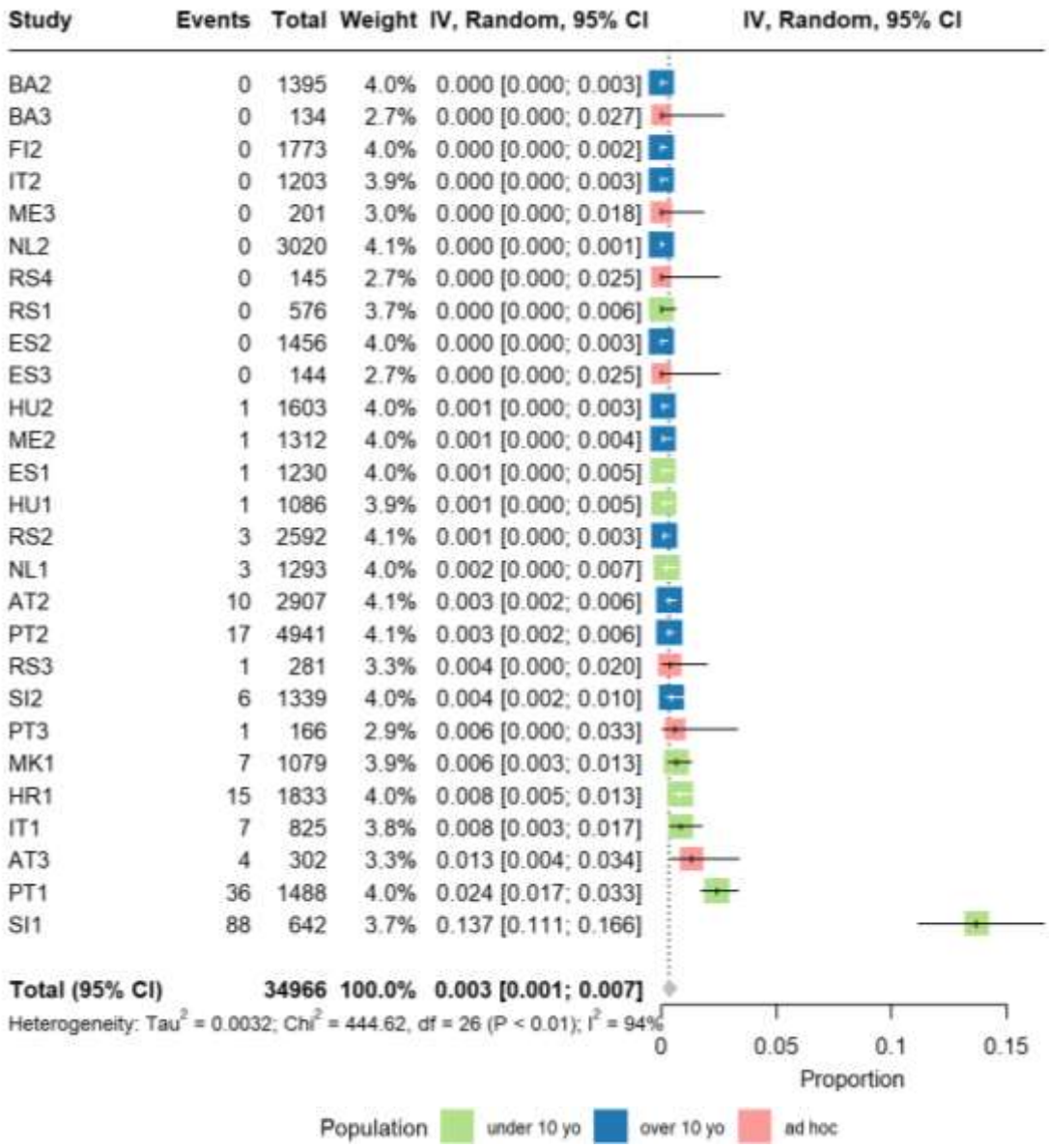

Figure A 34. Indicator H1.1 – Energy outliers, below 500 kcal.

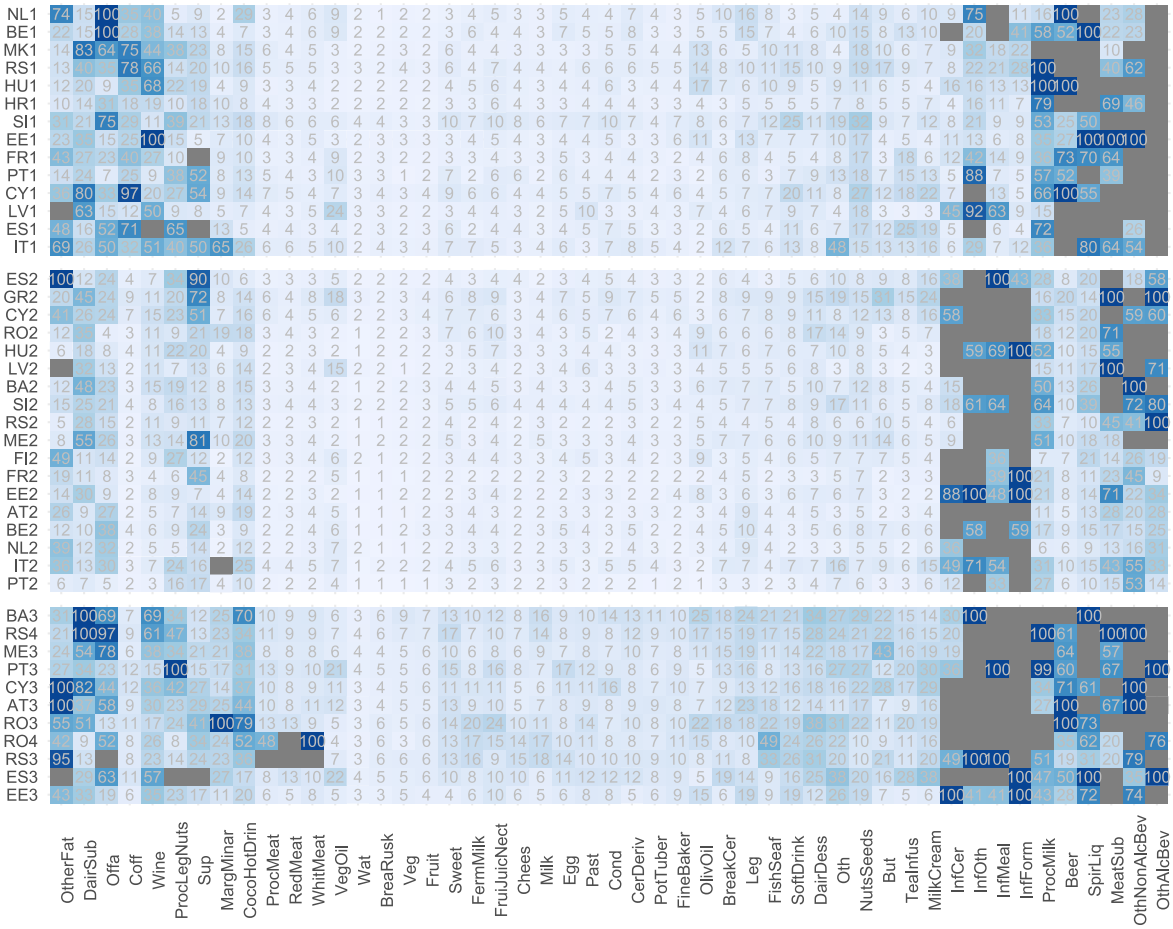

Figure A 35. Indicator H1.2 – Food groups outliers. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients

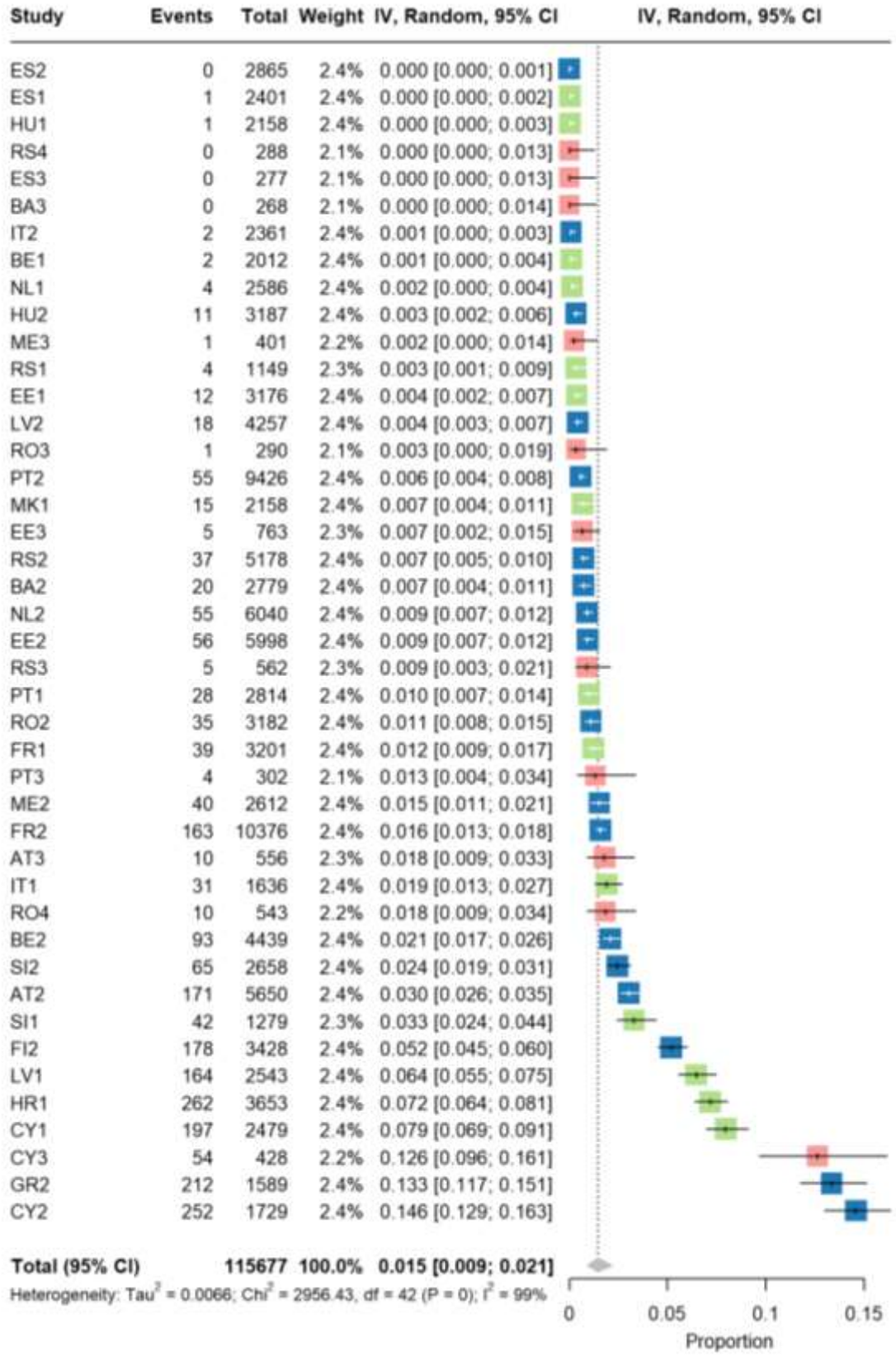

Figure A 36. Indicator H1.3 – Proportion of incomplete interviews (<2 main meals).

Figure A 37. Indicator H1.4 – Number of food items per interview.

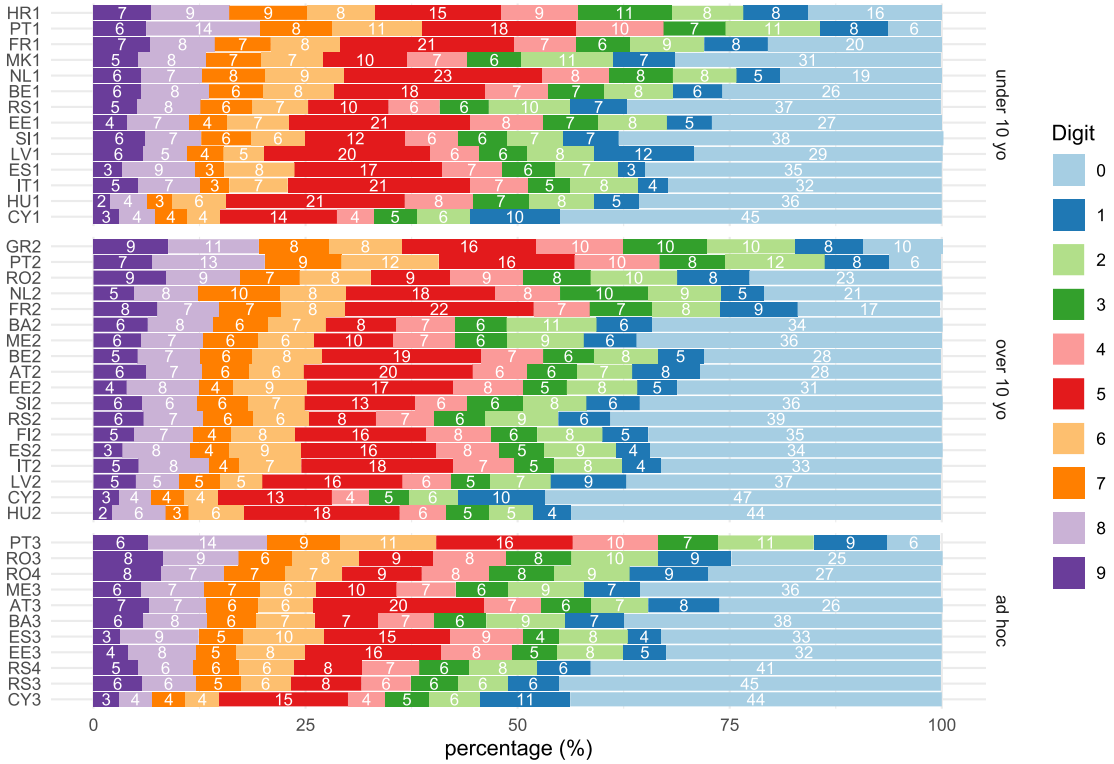

Figure A 38. Indicator H1.5 Digit preference in food amounts. Distribution per digits.

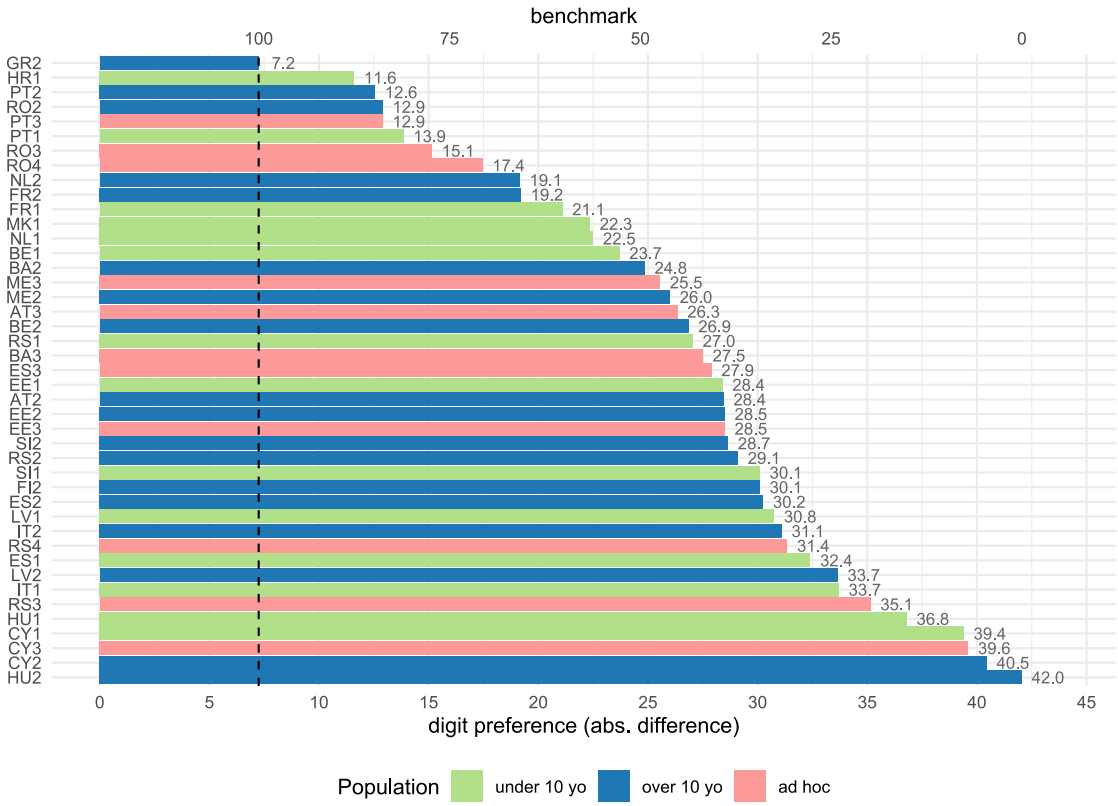

Figure A 39. Indicator H1.5 Digit preference in food amounts. Benchmark.

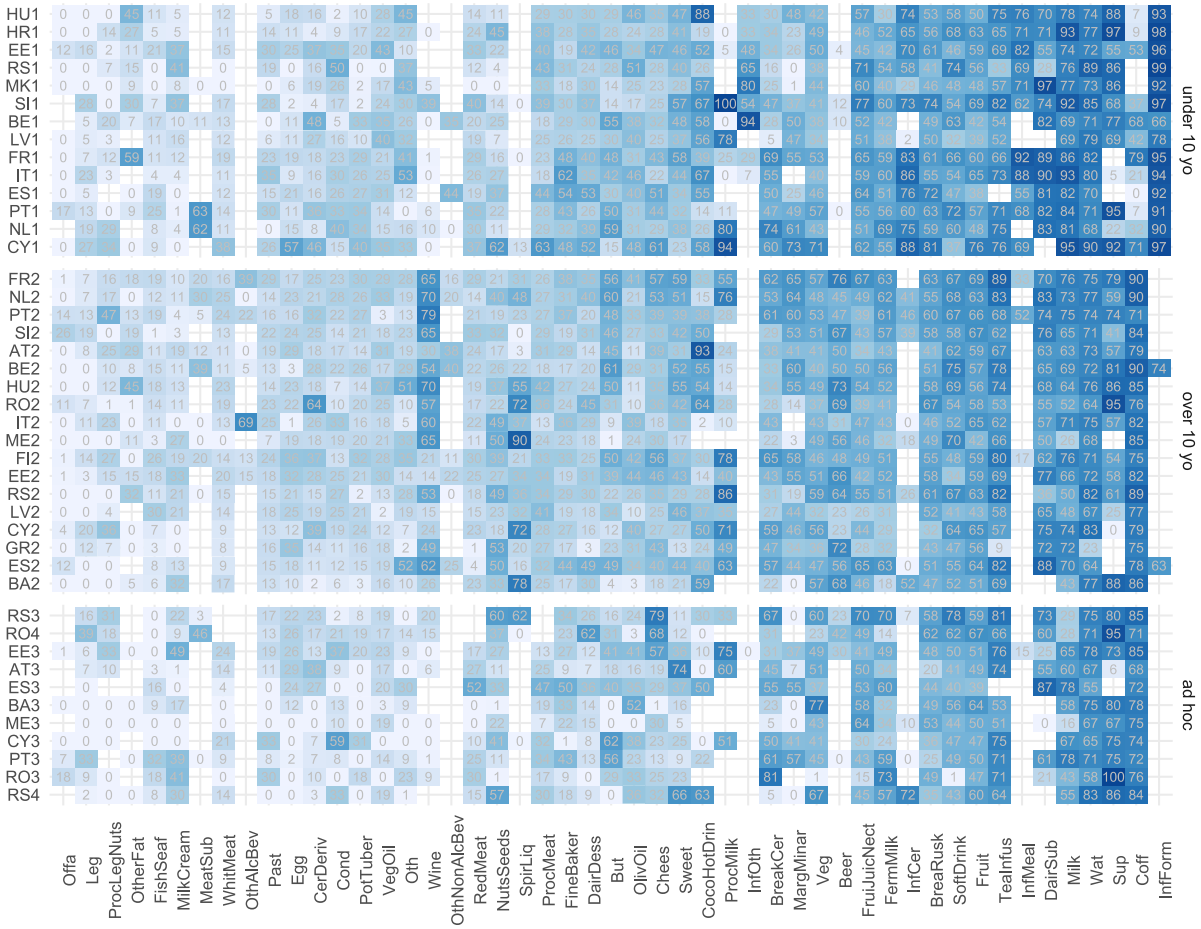

Figure A 40. Indicator H2.1 Proportion of total variance explained by differences between individuals, for food groups. Food groups abbreviations: **CerDeriv:** Cereals and cereal primary derivatives; **BreaRusk:** Bread and similar products; **Past:** Pasta, doughs and similar products; **FineBaker**: Fine bakery wares; **BreakCer**: Breakfast cereals; **Veg:** Vegetables and vegetable products; **PotTuber:** Starchy roots or tubers and products thereof, sugar plants; **Leg:** Legumes; **NutsSeeds:** Nuts, oilseeds and oilfruits; **ProcLegNuts:** Processed legumes, nuts, oilseeds and spices; **Fruit:** Fruit and fruit products; **RedMeat:** Mammals meat; **WhitMeat:** Birds meat; **Offa:** Offal and other slaughtering products; **ProcMeat:** Charcuteriem, sausagges and other processed meats; **FishSeaf:** Fish, seafood, amphibians, reptiles and invertebrates; **Milk:** Milk; **MilkCream:** Dairy cream and products; **FermMilk:** Fermented milk products; **Chees:** Cheese; **ProcMilk:** Dairy products, milk powders and concentrates; **DairDess:** Dairy dessert and similar; **Egg:** Eggs and egg products; **Sweet:** Sugar and similar, confectionery and water-based sweet desserts; **OlivOil:** Olive oils; **VegOil:** Vegetables oils; **But:** Butter; **MargMinar:** Margarines and minarines; **OtherFat:** Other fats; **FruiJuicNect:** Fruit and vegetable juices and nectars (including concentrates); **Wat:** Drinking water; **SoftDrink:** Soft drinks and energy drinks; **OthNonAlcBev:** Other non-alcoholic beverages; **Beer:** Beer and beer-like beverage; **Wine:** Wine and wine-like drinks; **SpirLiq:** Unsweetened spirits and liqueurs; **OthAlcBev:** Other alcoholic beverages; **Coff:** Coffe ingredients and drinks; **TeaInfus:** Tea and infusion ingredients and drinks; **CocoHotDrin:** Cocoa and other hot drinks; **InfForm:** Infant and follow-on formulae; **InfMeal**: Ready-to-eat meal for infants and young children; **InfCer:** Processed cereal-based food for infants and young children; **InfOth:** Other food for infants and children; **MeatSub:** Meat imitates; **DairSub:** Dairy imitates; **Sup:** Food supplements and products for particular diets**; Cond:** Seasoning, sauces, condiments and spices; **Oth:** Other foods and ingredients.

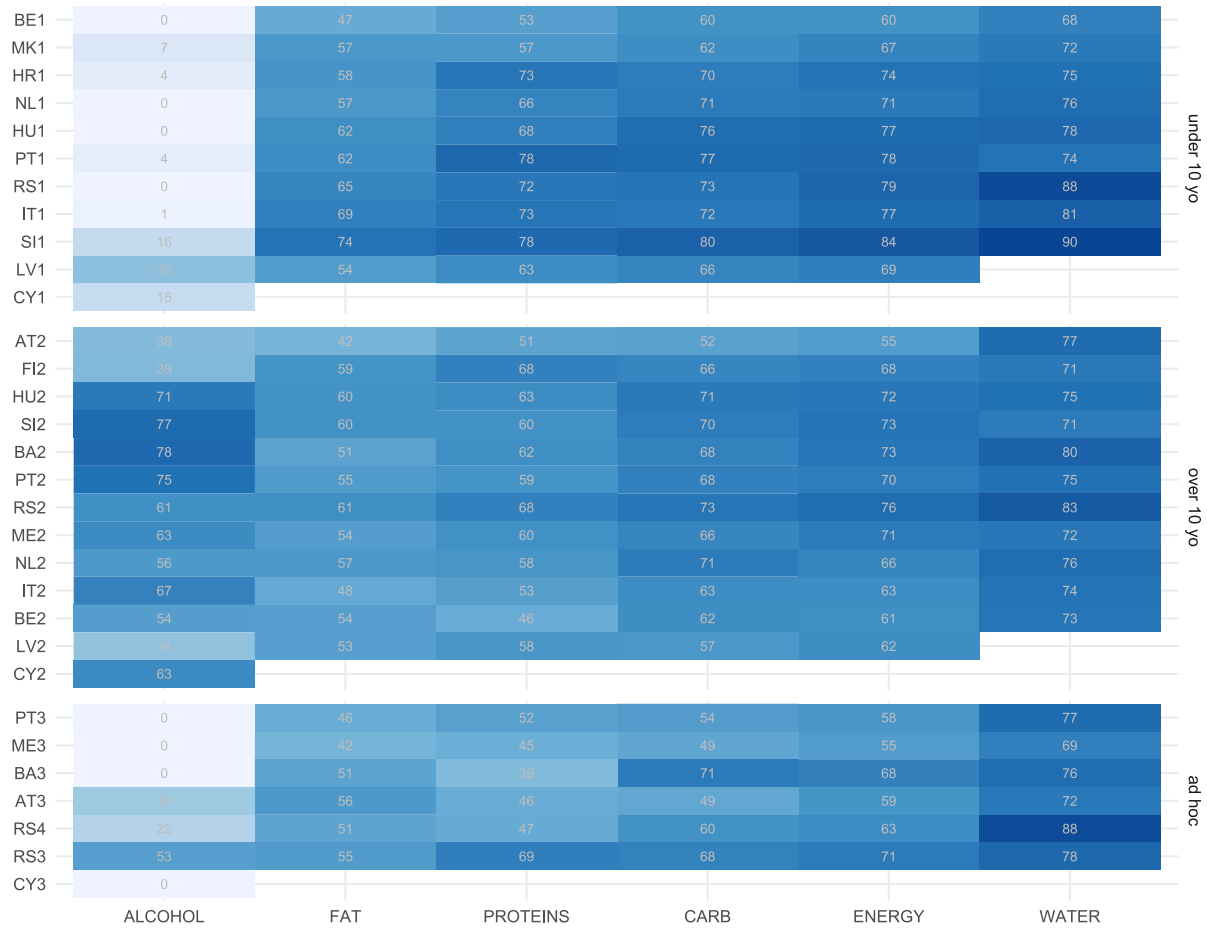

Figure A 41. Indicator H2.1 Proportion of total variance explained by differences between individuals, for energy, macronutrients, water and alcohol.

Dimension I – Results reporting

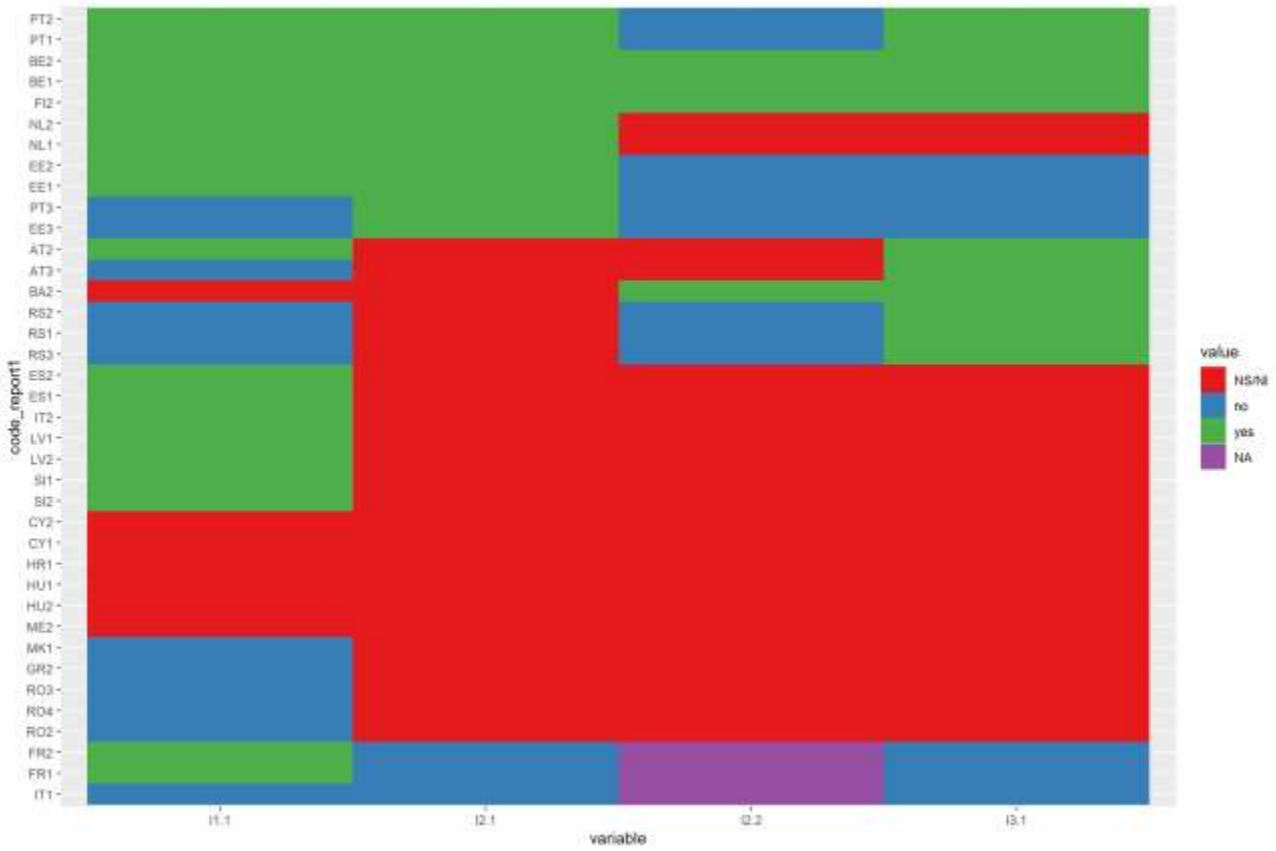

Figure A 42. Heatmap for categorical indicators of Dimension I. Survey codes identify the country (ISO 3166-1 alpha-2 code) and the age target group (1 for <10 years old; 2 for ≥10 years old; 3 and 4 for ad-hoc surveys).