Nutrition, Metabolism & Cardiovascular Diseases (2019) **29**, 127–134

Available online at www.sciencedirect.com

Nutrition, Metabolism & Cardiovascular Diseases

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

Identifying sources of measurement error in assessing dietary intakes – Results of a multi-country ring-trial

A. Gavrieli [a], A. Trichopoulou [a,*], L.M. Valsta [b,1], S. Ioannidou [b], R. Berry [c], M. Roe [c], L. Harvey [c], P. Finglas [c], M. Glibetic [d], M. Gurinovic [d], A. Naska [a,e]

[a] *Hellenic Health Foundation, Athens, Greece*
[b] *Evidence Management (DATA) Unit, European Food Safety Authority, Parma, Italy*
[c] *Quadram Institute Bioscience, Norwich, UK*
[d] *Institute for Medical Research, Centre of Research Excellence in Nutrition and Metabolism, University of Belgrade, Serbia*
[e] *Dept. of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Greece*

Received 12 June 2018; received in revised form 24 October 2018; accepted 29 October 2018
Handling Editor: A. Siani
Available online 8 November 2018

KEYWORDS
24 h dietary recalls;
Food records;
Dietary assessment;
Ring-trial;
Nutrition surveys

Abstract *Background and aims:* Epidemiological investigations include dietary intakes as primary exposures or potential confounders. To reduce bias, data collection protocols include the administration of questionnaires together with measurements of biomarkers. Some error, however, remains and needs to be considered in the analysis and interpretation of results. The European Food Safety Authority supported a ring-trial to compare the precision and reproducibility of dietary assessment methods applied in Europe.
Methods and results: Software applications used to collect 24-hour recalls and food records in six countries (Estonia, Italy, Latvia, Portugal, Spain, and Sweden) were assessed. The intake of 256 foods was identically reported to each method. Experienced interviewers participated and were instructed to repeat national protocols closely. The error in recording quantities, compared with reference values, was variable but in about 60% of recorded quantities was in the range of ±20%. Errors were however unsystematic and independent of the food type or quantification method used - although food pictures performed better. The reproducibility of some tools was limited. The methods generally captured additional ingredients (usually flavoring agents), but not sweetening agents or fortification and failed to record packaging information in about 60% of the cases.
Conclusion: In a design that eliminated respondent bias, this study indicates that softwares, supporting databases and interviewers generally introduce random error in dietary assessments. The inclusion of large sample sizes and food pictures to quantify portions, together with enhanced attention on interviewers' training, standardisation of procedures and regular tool upgrades are essential in assuring a study's quality and comparability.

* Corresponding author. Kaisareias 13 & Alexandroupoleos GR-115 27, Athens, Greece. Fax: +30 210 777 0771.
E-mail address: atrichopoulou@hhf-greece.gr (A. Trichopoulou).
1 National Institute for Health and Welfare, Department of Public Health Solutions, Helsinki, Finland

https://doi.org/10.1016/j.numecd.2018.10.011
0939-4753/

Introduction

Epidemiological investigations on associations between environmental factors and health-related outcomes include dietary intakes either as the primary exposure or as potential confounders. Methods to collect self-reported dietary data include: *food frequency questionnaires (FFQs)*, traditionally applied in studies assessing diet–disease associations; *single or multiple 24-hour dietary recalls (24-HDRs) and food records*, which are used in nutritional assessment and surveillance [1,2]. Data collection often relies on computerized systems (interviewer-assisted or self-administered), which aim to standardize the process and probe into details of food consumption in a harmonized manner.

Sources of uncertainty are inherent in all dietary assessment methods [3,4]. To overcome this, current practice recommends the concurrent administration of different dietary questionnaires and the collection of biological specimens to estimate levels of biomarkers of intake. However, even when the best possible method or combination of methods is selected, some measurement error remains and needs to be accounted for in the analysis and interpretation of results [5,6]. Over the years, attempts to understand the structure of the measurement error in dietary research has focused on the bias introduced through the participant ("intake-related" and "person-specific") [7] and less attention has been given to error introduced through the interviewers and/or the operational characteristics of the tools used to collect data.

In view of a pan-European survey, the European Food Safety Authority (EFSA) launched the EU-Menu [8,9], which recommends the administration of at least two 24-HDRs and a non-quantitative FFQ in surveys addressing adolescents and adults [9]. However, several European countries have considerable experience in implementing large-scale dietary surveys and have already applied in-house protocols [10]. In order to compare data collected through different protocols, EFSA supported a ring trial [11,12]. The trial provided insight to sources of uncertainty different from those introduced by the study participants. The present manuscript describes this test, thereby providing an insight on sources of measurement error other than those introduced through the subject's self-reporting.

Methods

The ring-trial was designed and organized by the EFSA Evidence Management Unit (DATA) following the principles of an inter-laboratory proficiency testing study [13]. The trial was analyzed by a Consortium of three Academic Institutions (Hellenic Health Foundation, Athens, Greece; Institute of Food Research (current Quadram Institute Bioscience), Norwich, UK; Institute for Medical Research, University of Belgrade, Serbia) [12]. The ring-test assessed computerized and/or web-based tools that collect dietary information through either the 24-HDR method or food records. The performance of the different software was evaluated based on results from 256 food entries introduced as ten consumption days, which were reported in an identical manner to each one of the tools assessed.

Participants

Invitations were sent to all research teams in Europe involved in national dietary surveys. Groups from ten European countries expressed their interest to participate and six (Estonia, Italy, Latvia, Portugal, Spain, and Sweden) fulfilled the criteria for participation [12]. The tools included in the trial and a summary of their characteristics are provided in Table 1, by country in alphabetical order. In brief, four 24-HDRs and three food records (one tool represented both methods) were included in the ring-trial. In both methods, food images (on screen or in print) were mainly used for quantification purposes. Five tools provided functions to allow the interviewer to handle recipes and composite foods. All tools included information on food composition and five tools could record information on the product's brand name.

The ring-trial

In the preparatory phase, the six teams provided supporting documents, background material and information in order to standardize food reporting during the trial. EFSA compiled 256 food entries divided into ten consumption days of fictitious male and female individuals. In general, the selection of food items was random. In certain instances, foods commonly consumed in the participating countries were also intentionally included. From 113 items reported, 61 were consumed once and 52 were consumed in repetition (i.e. several times within the same or different days). In total, the intake of 256 foods, beverages and food supplements was reported over the ten consumption days, together with details on portion sizes and food descriptors. Seven out of the ten days were defined as 'usual' and the remaining three referred to special occasions (i.e. buffet dinner and restricted eating because of disease).

EFSA staff, who were native or fluent speakers of the languages of the participating countries and not involved in the evaluation of results, acted as the "interviewee/survey participant". Each acting "interviewee" was given a script describing in detail the meals of the ten consumption days in their national language and was trained to report intakes and related information in a uniform manner. In particular "interviewees" were provided with print-outs of interview menus including exactly the same foods and related descriptors, as well as clear instructions on which food image to indicate in each national picture book. Moreover, before the trial's execution, "interviewees" were thoroughly trained in order to keep the procedures consistent and potentially unclear terminology was thoroughly discussed. Furthermore, the "interviewees" kept notes during the interviews on food details that were not inquired (e.g. milk fat content). Trained interviewers took part in the ring-test and were instructed to repeat, as closely as possible, the data collection

Table 1 Countries and dietary assessment tools participating in the ring-trial.[a]

Country/Name of the software	Estonia/Nutridata pro	Italy/INRAN-DIARIO-MPS 1.0 [21]	Latvia/Food Consumption Information System (FCIS)	Portugal/EPIC-SOFT [20]	Spain/Software for Assessing Diets and Food Calculations DIAL [22]	Sweden/Riksmaten 2010 [23]
Method	24-HDR or food diary	Food diary	24-HDR	24-HDR	24-HDR	Food diary
Validated tool	Yes	No	No	Yes	No	Yes
Mode of administration	Self- & Interviewer-administered	Self-administered	Self-administered	Interviewer-administered	Interviewer-administered	Web-based, Self-administered
Food pictures: validated	Yes	No	Not known	Yes	No	Yes
Food pictures: number of items	60	62	144	85	144	25/318[b]
Food pictures: number of pictures per item	3–8	1–4	2–6	4–6	1–3	1–7
Number of integrated databases	14	9	3	9	11	12
Possibility to expand existed incorporated databases (foods, recipes, supplements)	Yes	Yes	Yes	Yes	Yes	Yes
Possibility to import databases	No	No	Yes	Yes	No	Yes
Quality controls for missing and/or implausible values (total number of controls)	Yes (10)	Yes (9)	Yes (6)	Yes (10)	No	Yes (3)
Number of methods for food quantification	7	6	2	7	7	5
Structured interviewing/data entry	Yes	Yes	No	Yes	No	Yes
Application of probing questions	Yes	No	No	Yes	No	No
Questions about forgotten items	Yes	No	No	Yes	Yes	No
International recipes included	Yes	No	Yes	Yes	No	Yes

[a] Tools may have been updated since this information was captured 24-HDR: 24 h dietary recall.
[b] 25 pictures in the print-out version and 318 pictures in the electronic version.

protocols that are routinely applied in their national settings. The procedures usually applied in the case of self-administered food records were also respected - phone interviews were carried out to clarify the recordings.

The ring-trial was organized as a two day data collection among interviewees, interviewers and the researchers responsible for data analysis at EFSA premises, according to a pre-determined schedule (Fig. 1). All participants signed a confidentiality agreement before the data collection started. Members of the research team observed the data collection and asked for clarifications about the procedures after the ring-test was completed. The food record inputs and checks were performed parallel with the interviews during the same data collection days.

Statistical analysis

Data were analyzed qualitatively and quantitatively. The qualitative analysis assessed the precision through which characteristics of the meal (e.g. place) and each food (e.g. fortification agent) were recorded. Overall 573 food characteristics (descriptors) were recorded and were grouped

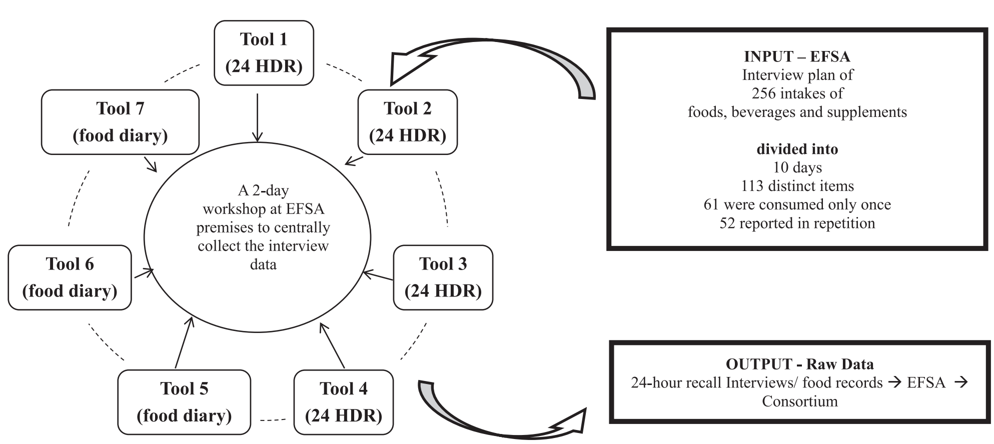

Figure 1 Schematic presentation of the ring-trial methodology.

in 22 categories (facets), based on the EFSA FoodEx2 classification system [14,15].

The quantitative analysis included an evaluation of the *relative precision* through which the reported quantities were captured. The percentage deviation of the recorded quantity from the reference food quantity was calculated. Box-and-whisker plots per food group were generated to present the variability of percent relative errors in recorded quantities within all tools combined. Data normality was assessed through the Kolmogorov–Smirnov test. Since the distribution of the reported intakes was right-skewed and remained skewed after log transformation, Spearman's rank correlation coefficients were estimated.

Data analysis additionally included the evaluation of consistency among identical repetitions of measurements (*reproducibility*). Fifteen items were reported identically three times, 21 items five times and one item was repeated 11 times during the 10 consumption days. Percent reproducibility was calculated as the proportion of times the most frequently recorded quantity appeared in the data output. Because the conditions of repeated administrations are never identical, reproducibility was calculated with and without accepting a ±10% difference between the recorded quantities. All statistical analyses were performed with the STATA, v.11 package and significance was set as $p < 0.05$.

Results

The relative precision of the recorded data

In all dietary assessment tools, the recorded quantities were significantly positively correlated with the respective reference quantities (Spearman's ρ above 0.90, p-value < 0.001 in all assessments). Table 2 presents frequencies of items for which the difference between the recorded and the reference quantity was 0, 10, 15 and 20%, by tool. The proportion of items which were precisely recorded by the tools (0% difference from the reference quantity) was generally low and ranged between 16% and 38% in the 24-HDRs and between 21% and 33% in the diaries. When a 20% difference from the reference quantity was accepted, the relative precision ranged between 50 and 62% for the 24-HDRs and between 38 and 62% for the diaries. In general, food pictures led to more precise estimations of portion sizes (Table 3).

Fig. 2 presents box-and-whisker plots reflecting the spread of percent relative errors in recording the mean intakes of nine food groups. The variability of relative errors in assessing mean consumption was generally large, except for the consumption of bread, fresh pork, processed meat, milk and dairies. Negative relative errors (in pasta, bread, soft drinks and processed meat) indicated that the estimated average intakes were smaller than the average reference quantities and positive errors (in potatoes, chicken and salmon of different types) indicate the opposite. The relative error in capturing intake was inversely correlated with the number of times a food item was reported ($r = -0.60$, p-value = 0.09) and was rather balanced in the case of processed meat, milk and diaries (median relative error = 0%). Considering the fact that EFSA-related risk assessment often relies on intake values generated through different assessment tools, these figures indicate that sources of systematic or random bias, other than that introduced through study participants, also contribute to the generated error.

The reproducibility of the recorded data

The consistency in capturing repeated measurements varied among tools (reproducibility index ranged between

Table 3 Percentage of food quantities captured with a difference of 0% or ±10% between the recorded and the reference (expected) quantity by tool and quantification means (food pictures vs. other methods).

Tool	Quantities (%) recorded through pictures (n/n_{total})	Quantities (%) not recorded through pictures (n/n_{total})
Tool 1 (24-HDRs)	52 (37/71)	18 (21/119)
Tool 2 (24-HDRs)	63 (55/88)	47 (51/109)
Tool 3 (24-HDRs)	47 (34/72)	33 (42/126)
Tool 4 (24-HDRs)	37 (32/87)	38 (42/112)
Tool 6 (food diaries)	37 (19/52)	30 (45/152)
Tool 7 (food diaries)	78 (38/49)	32 (24/76)

*Tool 5 is not presented since all portion sizes were recorded through food pictures.

Table 2 Number (%) of quantities recorded by each tool, with a difference of 0%, ±10%, ±15% and ±20% between the recorded and the reference (expected) quantity.

Tool	Missing data	Number of recorded quantities which were equal or close to the respective reference quantities			
		Equal	Equal ±10% difference[a]	Equal ±15% difference[a]	Equal ±20% difference[a]
Tool 1 (24-HDRs)	14 (6.8)	30 (15.6)	60 (31.3)	79 (41.1)	96 (50.0)
Tool 2 (24-HDRs)	9 (4.4)	75 (38.1)	106 (53.8)	116 (58.9)	123 (62.4)
Tool 3 (24-HDRs)	8 (3.9)	35 (17.7)	76 (38.4)	89 (44.9)	106 (53.5)
Tool 4 (24-HDRs)	7 (3.4)	47 (23.6)	74 (37.2)	89 (44.7)	109 (54.8)
Tool 5 (Food diaries)	1 (0.5)	43 (21.0)	62 (30.2)	70 (34.1)	78 (38.0)
Tool 6 (Food diaries)	2 (1.0)	44 (21.6)	64 (31.4)	68 (33.3)	89 (43.6)
Tool 7 (Food diaries)	4 (3.1)	41 (32.8)	62 (49.6)	74 (59.2)	78 (62.4)

[a] Difference between the recorded and the respective reference quantity.

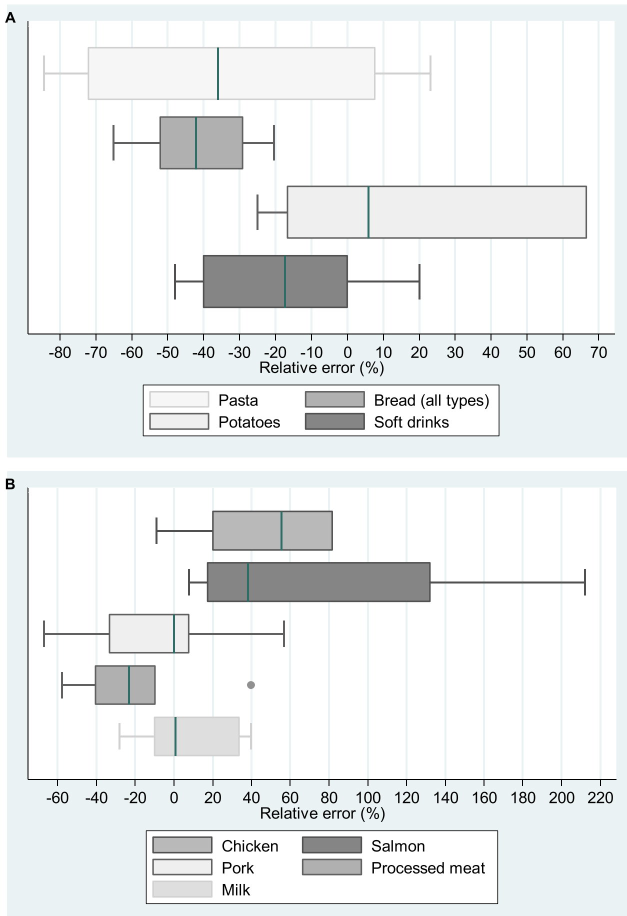

Figure 2 Box-and-whisker plots of percent relative errors in estimating the intake of the indicated food groups and soft drinks; (A) Plant foods (B) Animal foods. The left and right side of the box indicate the first and third quartiles of the errors and the band inside the box presents the median. The ends of the whiskers represent the minimum and maximum of the tools' error in capturing the reported intake and individual points denote outliers.

40 and 100%). Considering that reproducibility provides an indication of random errors in data collection, almost all tools provided reproducible recordings for alcoholic and non-alcoholic beverages.

Capturing characteristics (descriptors) of the food items and meals

Along with hazard identification, risk assessment requires a precise estimation of the levels of exposure to particular attributes of the items consumed. Table 4 presents the number of food characteristics correctly recorded per category and tool. Taking into consideration characteristics that can be of particular interest in risk assessment, the tools captured precisely the information on: (a) food source (e.g. cow, sheep, etc. for milk and dairies; wheat, barley etc. for cereals) in 53% of the reports, (b) additional ingredients (e.g. fruit in yoghurt) in 78% of reports, (c) type of sweetening agent (39%), (d) fortification (27%), (e) process technology (e.g. UHT) (24%), (f) treatment (e.g. squeezed) (52%), (g) preservation method (canned, dried, etc.) (57%) and (h) on packaging (43%), with tools more efficiently capturing information on the packaging format than material (Table 4). When different protocols were compared, food dairies were generally less efficient in capturing food characteristics. For instance, food diaries failed to capture information on production methods (e.g. organic) and the technology used for processing. In 24-HDRs inconsistencies were rather unsystematic. For instance, Tool 1 allowed the precise recording of the products' fat content and fortification (79% and 71% of the reports, respectively), but failed to capture the type of sweetening agent used (33%). The discrepancy observed could have been generated by the supporting databases, inadequate training of the interviewers or the tools' limited capacity to standardise procedures. Tools were also assessed in relation to how accurately they captured the time and type of each meal as well as the eating place. Results are presented in the trial's final report [12].

Discussion

In an attempt to understand sources of errors that are not directly introduced by the study participants, we performed a ring-trial testing six dietary assessment tools, using ten consumption days as the reference material (including three days of special occasions) and for a total of 256 food items reported. In general, the relative error introduced by the methods, the protocols and the interviewers was variable and acceptable recordings barely exceeded 60%, even when a 20% deviation from the reference quantity was allowed. Nonetheless, the recorded quantities were highly correlated with the reported ones and the over- or underestimation of the recorded quantities appeared to be random and independent of the type of food or the method used to quantify intake - although food images generally performed better. Hence, overall, the sum of quantities of solid foods reported in each one of the 10 days was $\pm$100 g from the sum of the corresponding reference quantities, which ranged from 866 g (day 2) to 2555 g (day 6), the percent deviation ranging between $-$43.3% and +2.4% over the 10 days.

Most tools adequately recorded specific characteristics of the meal occasions, but were not reproducible in capturing intakes. In addition, tools could not precisely record the majority of the items' qualitative characteristics (e.g. additional ingredients, sweetening agents, fortification, packaging) in about 60% of the cases. As the study design eliminated the error introduced by the respondent, this finding probably indicates that the supporting databases and/or related operations need improvement, since detailed information on attributes of dietary exposure is essential to risk assessors and public health officials.

Traditionally, the attempts to understand and model the error in dietary reporting have focused on subject-specific bias [7]. This ring-trial, however, aimed to eliminate the variation introduced by the study participants and assess the nature and magnitude of the variation generated through data collection, i.e. the interview process and the interviewer, the data entry in self-administered

Table 4 Distribution (N, %) of food characteristics (descriptors) captured by each tool and per category (facet) of characteristics.

Facets/Categories	No of food item characteristics	Characteristics captured						
		24-HDRs				Food diaries		
		Tool 1	Tool 2	Tool 3	Tool 4	Tool 5	Tool 6	Tool 7[a]
1. Source of the food (e.g. cow milk, rye bread, wheat bread)	44	25 (56.8)	28 (63.6)	26 (59.1)	18 (40.9)	26 (59.1)	16 (36.4)	14 (50.0)
2. Raw source of derivatives (e.g. still, tap water)	32	26 (81.3)	27 (84.4)	31 (96.9)	7 (21.9)	9 (28.1)	24 (75.0)	5 (20.8)
3. Additional ingredients (e.g. fruits and nuts in yogurt)	6	4 (66.7)	5 (83.3)	5 (83.3)	6 (100.0)	6 (100.0)	2 (33.3)	4 (100.0)
4. Physical state (e.g. liquid, small pieces, tablet)	38	27 (71.1)	23 (60.5)	33 (86.8)	22 (57.9)	27 (71.1)	9 (23.7)	13 (54.2)
5. Flavor (e.g. orange taste, lemon aroma)	11	4 (36.4)	4 (36.4)	10 (90.9)	7 (63.6)	5 (45.5)	0 (0.0)	1 (14.3)
6. Fat content	14	11 (78.6)	3 (21.4)	0 (0.0)	0 (0.0)	5 (35.7)	4 (28.6)	8 (88.9)
7. Sweetening agent (e.g. sugar, aspartame)	9	3 (33.3)	6 (66.7)	6 (66.7)	2 (22.2)	3 (33.3)	1 (11.1)	2 (33.3)
8. Fortification	14	10 (71.4)	0 (0.0)	5 (35.7)	2 (14.3)	0 (0.0)	6 (42.9)	8 (61.5)
9. Qualitative indication of various characteristics (e.g. full fat, normal fat, light, no sugar added)	52	42 (80.8)	27 (51.9)	31 (59.6)	28 (53.8)	37 (71.2)	20 (38.5)	26 (81.3)
10. Alcohol content	12	7 (58.3)	0 (0.0)	5 (41.7)	0 (0.0)	5 (41.7)	12 (100.0)	6 (100.0)
11. Dough-mass (e.g. yeast leavened)	1	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100.0)
12. Process Technology (e.g. UHT, parboiled)	12	5 (41.7)	1 (8.3)	8 (66.7)	0 (0.0)	0 (0.0)	3 (25.0)	1 (12.5)
13. Treatment (e.g. minced, squeezed, peeled)	20	17 (85.0)	9 (45.0)	17 (85.0)	6 (30.0)	11 (55.0)	2 (10.0)	9 (64.3)
14. Cooking method (e.g. baked, boiled)	38	21 (55.3)	19 (50.0)	22 (57.9)	2 (5.3)	15 (39.5)	9 (23.7)	18 (72.0)
15. Preservation method (e.g. dried, canned)	17	14 (82.4)	7 (41.2)	9 (52.9)	11 (64.7)	8 (47.1)	9 (52.9)	8 (72.7)
16. Packaging Format (e.g. container, bottle, can)	63	29 (46.0)	3 (4.8)	51 (81.0)	10 (15.9)	26 (41.3)	2 (3.2)	25 (58.1)
17. Packaging Material (e.g. glass, plastic)	54	32 (59.3)	1 (1.9)	41 (75.9)	0 (0.0)	0 (0.0)	0 (0.0)	33 (82.5)
18. Production method (e.g. organic)	4	4 (100.0)	2 (50.0)	3 (75.0)	3 (75.0)	0 (0.0)	0 (0.0)	0 (0.0)
19. Place of preparation/production (e.g. homemade, commercial)	6	4 (66.7)	5 (83.3)	5 (83.3)	0 (0.0)	3 (50.0)	0 (0.0)	5 (100.0)
20. Target consumer groups (e.g. adults, toddlers)	11	3 (27.3)	2 (18.2)	0 (0.0)	3 (27.3)	3 (27.3)	0 (0.0)	3 (33.3)
21. Intended use (e.g. raw, cooked)	4	4 (100.0)	3 (75.0)	3 (75.0)	1 (25.0)	1 (25.0)	3 (75.0)	2 (100.0)
22. Brand name	37	24 (64.9)	12 (32.4)	32 (86.5)	12 (32.4)	34 (91.9)	6 (16.2)	22 (91.7)

[a] In the case of Tool 7, data collection included seven instead of ten-day records and the number of reported items was generally smaller.

methods and the operational characteristics of the tool used for the assessment [7]. Interviewers were recognized as a major contributor to the variance in the collected data. Due to the open-ended nature of the 24-HDR method, interviewers are subjected to a memory burden [16]. In addition, their background and degree of training on how to ask questions, probe for clarifications and react to responses may affect data quality. The relationship which interviewers establish with the participants, the attention they pay and their motivation may also impact the data quality [16]. In the present study, several of the discrepancies observed could have been avoided if the interviewer was intensively trained to probe for details and be able to distil the key features of the reported items.

Dietary intake is a component of human behavior, and may vary according to cultural norms, personal beliefs and preferences. For this reason, extensive standardization should be limited and training sessions should be based on the previous experience of the fieldwork team. Self-administered tools have been proposed to overcome the interviewer bias, but they could introduce other types of error related to non-response, altered behavior and social desirability bias [17,18].

Characteristics of the software tools used, such as the type and number of the integrated databases, explained in part the differences observed in how the tools recorded food consumption. The selection of integrated descriptors could be related to the history and purposes of the method development. In order to accommodate EFSA needs, adjustments in food lists and supporting databases may need to be applied before data collection. The between-tools variation observed in this study raises concerns when national data are pooled to be analyzed collectively for inter-country comparisons. The incorporation of common facets and descriptors in the tools is certainly a prerequisite to reduce uncertainty in the assessment of food-related risks and the updated FoodEx2 system aims to accomplish this [19].

Crispim et al. [20] provided a very detailed and comprehensive description of the quality controls included in a 24-HDR tool before, during and after the interview. These controls refer to all steps of fieldwork preparation, data collection, entry, calculation and analysis and have been shown to improve the precision, completeness and comparability of the generated dietary data. The use of tools with long lists of facets and food descriptors is generally expected to enhance the precision of the data. Nevertheless, in the present study, less equipped tools performed equally well or even better than the more sophisticated and flexible tools. In reality, in their selection of a method, researchers need to balance between serving the study's scope and reducing the burden introduced to participants.

To the best of our knowledge, this is the first time that a ring-test method (usually developed to evaluate the performance of analytical laboratories) was applied in a dietary survey setting. In designing and conducting this study, the participant (respondent-specific) bias was controlled as much as possible so that other sources of error would emerge and be recorded. Nonetheless, although the interviewees had been instructed to follow exactly the guidelines that ensured the standardised reporting, an interaction between the interviewer and the respondents could not be ruled out. This is an innovative approach to understand a genuinely complex issue, such as dietary reporting, which is directly related to several aspects of public health, including —but not limited to-consumers' safety, disease prevention and health promotion. This study assisted EFSA to expand knowledge on dietary tools available in EU and the results were considered in the development of the EU Menu guidance document [9].

The dietary assessment tools and the data collection procedures evaluated in the present study generally led to a random introduction of errors in the generated data. Several sources of variation could explain the discrepancies observed between the recorded and the reference data with interviewer being a major source of the inaccurate recordings. In conclusion, dietary surveys that employ methods of questionnaire reporting should include: (a) sample sizes adequately large to address random errors; (b) a thorough training of the interviewers (or study participants in the case of self-administered methods); (c) the use of food images for quantification purposes; (d) the standardization and automatization of the data collection procedures; and (e) the regular upgrade of the tools so as to be flexible, dynamic and up-to-date. These measures could contribute to improving the comparability of the generated information in multi-centre studies.

Conflict of interest

None to declare.

Funding

This work was supported by EFSA, Italy [grant number CT/EFSA/DCM/2012/02] and was coordinated by the Hellenic Health Foundation, Greece. The opinions expressed are those of the Contractor only and do not represent the Authority's official position.

Acknowledgments

Authors would like to thank the personnel of Institutes providing the dietary assessment applications for their collaboration, as well as the EFSA staff who acted as "survey participants". Authors would also like to acknowledge Ms. Petra Gergelova (EFSA staff) in the preparation and implementation of the ring-trial and Ms. Charitini Konstantinidi (HHF staff) in the statistical analysis.

References

[1] Elmadfa I, Meyer AL. Developing suitable methods of nutritional status assessment: a continuous challenge. Adv Nutr 2014;5:590S–8S.

[2] Shim JS, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36:e2014009. https://doi.org/10.4178/epih/e2014009.

[3] Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, et al. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. J Am Diet Assoc 2006;106:1640–50. https://doi.org/10.1016/j.jada.2006.07.011.

[4] Lissner L, Troiano RP, Midthune D, Heitmann BL, Kipnis V, Subar AF, et al. OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. Int J Obes (Lond) 2007;31:956–61. https://doi.org/10.1038/sj.ijo.0803527.

[5] Kubena KS. Accuracy in dietary assessment: on the road to good science. J Am Diet Assoc 2000;100:775–6. https://doi.org/10.1016/S0002-8223(00)00226-1.

[6] Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. J Natl Cancer Inst 2011;103:1086–92. https://doi.org/10.1093/jnci/djr189.

[7] Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. Publ Health Nutr 2002;5:915–23. https://doi.org/10.1079/PHN2002383.

[8] EFSA. Guidance of EFSA general principles for the collection of national food consumption data in the view of A pan-European dietary survey. EFSA Journal 2009;7:1–51.

[9] EFSA. Guidance on the EU menu methodology. EFSA Journal 2014;12:3944–77. https://doi.org/10.2903/j.efsa.2014.3944.

[10] Merten C, Ferrari P, Bakker M, Boss A, Hearty A, Leclercq C, et al. Methodological characteristics of the national dietary surveys carried out in the European Union as included in the European Food Safety Authority (EFSA) comprehensive european food consumption database. Food Addit Contam Part A Chem Anal Control Expo Risk Assess 2011;28:975–95. https://doi.org/10.1080/19440049.2011.576440.

[11] EFSA. Guidance on uncertainty in EFSA scientific assessment EFSA scientific committee. 2016. https://www.efsa.europa.eu/sites/default/files/160321DraftGDUncertaintyInScientificAssessment.pdf. Accessed 12/05/2018.

[12] Gavrieli ANA, Berry R, Roe M, Harvey L, Finglas P, Glibetic M, et al. Dietary monitoring tools for risk assessment. EFSA Journal 2014. EN-607:287pp.

[13] FAO. Things to know about the ring test. 2015. http://www.fao.org/ag/againfo/home/documents/2015_Announcement_Ring_test.pdf. Accessed 12/05/2018.

[14] EFSA. A standardised system for classifying and describing food makes it easier to compare data from different sources and perform more detailed types of data analysis. 2012. http://www.efsa.europa.eu/en/data/data-standardisation. accessed 12/05/2018.

[15] EFSA. The food classification and description system FoodEx2 (revision 2). EFSA Journal 2015. EN-804:90pp.

[16] Slimani N, Ferrari P, Ocke M, Welch A, Boeing H, Liere M, et al. Standardization of the 24-hour diet recall calibration method used in the european prospective investigation into cancer and nutrition (EPIC): general concepts and preliminary results. Eur J Clin Nutr 2000;54:900–17.

[17] Falomir Z, Arregui M, Madueno F, Corella D, Coltell O. Automation of food questionnaires in medical studies: a state-of-the-art review and future prospects. Comput Biol Med 2012;42:964–74. https://doi.org/10.1016/j.compbiomed.2012.07.008.

[18] Illner AK, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol 2012;41:1187–203. https://doi.org/10.1093/ije/dys105.

[19] EFSA. Report on the development of a Food Classification and Description System for exposure assessment and guidance on its implementation and use. EFSA Journal 2011;9: 2489–573.

[20] Crispim SP, Nicolas G, Casagrande C, Knaze V, Illner AK, Huybrechts I, et al. Quality assurance of the international computerised 24 h dietary recall method (EPIC-Soft). Br J Nutr 2014; 111:506–15. https://doi.org/10.1017/S0007114513002766.

[21] Leclercq C, Arcella D, Piccinelli R, Sette S, Le Donne C, Turrini A. The Italian National Food Consumption Survey INRAN-SCAI 2005-06: main results in terms of food consumption. Publ Health Nutr 2009;12:2504–32. https://doi.org/10.1017/S1368980009005035.

[22] Ortega RM, López-Sobaler AM, Andrés P, Requejo AM, Aparicio A, Molinero LM. DIAL software for assessing diets and food calculations (version 3.0.0.5). 2013. http://www.alceingenieria.net/nutricion.htm. Accessed 12/05/2018.

[23] Bjermo H, Sand S, Nalsen C, Lundh T, Enghardt Barbieri H, Pearson M, et al. Lead, mercury, and cadmium in blood and their relation to diet among Swedish adults. Food Chem Toxicol 2013; 57:161–9. https://doi.org/10.1016/j.fct.2013.03.024.