Validation of a picture book to be used in a pan-European dietary survey

Sofia Vilela[1],*, Carla Lopes[1,2], Sofia Guiomar[3], Milton Severo[1,2], Lalka Rangelova[4], Stefka Petrova[4], Zuzsanna Horváth[5], Júlia Cseh[5], Antje Schweter[6], Oliver Lindtner[6], Árpád Ambrus[5] and Duarte Torres[7]

[1]EPIUnit – Instituto de Saúde Pública, Universidade do Porto, Rua das Taipas nº 135, 4050-600 Porto, Portugal: [2]Departamento de Ciências da Saúde Pública e Forenses, e Educação Médica, Faculdade de Medicina, Universidade do Porto, Alameda Prof. Hernâni Monteiro, Porto, Portugal: [3]National Health Institute Doutor Ricardo Jorge, Lisbon, Portugal: [4]Department of Food and Nutrition, National Center of Public Health and Analyses, Sofia, Bulgaria: [5]National Food Chain Safety Office, Directorate for Food Safety Risk Assessment, Budapest, Hungary: [6]Federal Institute for Risk Assessment, Berlin, Germany: [7]Faculdade de Ciências de Nutrição e Alimentação, Universidade do Porto, Porto, Portugal

Abstract

Objective: To validate a picture book for estimation of food portion sizes using two approaches: (i) 'perception' of food portions by comparison with a series of food photos; and (ii) 'conceptualization and memory', using the same photos to estimate the amount of served food one hour after self-served food portions.
Design: Each partner developed a country-specific picture book based on the so-called EPIC-Soft picture book. Representative and common photo series were chosen achieving approximately 25 % of the original picture book (*n* 23). Three portions from each photo series were randomly selected.
Setting: The study was performed within the Pilot study in the view of a Pan-European dietary survey – Adolescents, adults and elderly (PILOT-PANEU) project.
Subjects: A sample of adolescents and adults was recruited in five countries: Bulgaria (*n* 103), Finland (*n* 34), Germany (*n* 69), Hungary (*n* 62) and Portugal (*n* 77).
Results: Among the portions of the corresponding photo series and depending on the type of food, from 18 % (cheese) to 96 % (ratatouille) of participants chose the correct portions. In the perception study, agreement between the portions shown and reported was substantial (intraclass correlation coefficient (ICC) = 0·805) and the mean difference was very low. In the memory study, agreement between the served and reported portions was lower than in the perception study (ICC = 0·536). Agreement also seemed to decrease as the appearance of food on the plate differed from food in the picture.
Conclusions: Overall, the picture series selected can be applied in future intake surveys to quantify foods similar to those depicted in the pictures.

In recent years, the EU has highlighted the importance of encouraging healthier and safer living among the population. Related to this, the European Food Safety Authority has a central role in promoting and coordinating standardized data collection and harmonized risk assessment methodologies as a basis to ensure safe food consumption and healthy diets in Europe[1]. Food consumption data from dietary surveys are available in most European countries; however, data obtained at national level cannot be directly compared due to differences in how information is collected. In February 2010, members of the European Food Safety Authority's Advisory Forum signed a declaration supporting the establishment of a pan-European food consumption survey[2]. The EU-Menu project is one of the results of recognizing the need for having harmonized and accurate data collection on food consumption across Europe, allowing improvement of the consistency and reliability of exposure assessments.

In 2002, the EU-funded European Food Consumption Survey Method (EFCOSUM) project published recommendations on methods for monitoring European food consumption in national representative samples of adults[3]. More recently, the European Food Consumption Validation (EFCOVAL) project (http://www.efcoval.eu)

Corresponding author: Email sofia.vilela@ispup.up.pt

was funded as a continuation of the EFCOSUM project with the main purposes to develop further and validate a European food consumption survey method using standardized repeated 24 h dietary recall, applying the EPIC-SOFT[4] software as a tool for assessing dietary intake in European countries. The EFCOSUM group highlighted that estimation of the portion size consumed is one of the sources of error in collecting food intake data[5]. Weighing of the served portion is considered the most accurate method for measuring food intake. However, this method has some disadvantages such as being time consuming, costly and requiring a high level of cooperation from respondents[6]. An alternative to weighing the food portion size is to use a variety of visual aids to help participants estimate the amounts of food consumed more accurately during dietary recall interview. These quantification methods include household measures, abstract shapes and food photographs, among others[7–9]. Being easily adaptable to local conditions, cheap, reproducible and transportable are the main advantages of food photographs, in comparison to other portion size estimation tools[10], and the benefits of using photographs to help individuals estimating portion sizes are reported in several studies[11–14]. Three individual psychological elements can affect the reported portion size: perception, conceptualization and memory[15]. Perception involves the ability of a subject to relate the real quantity of a presented food to an amount depicted in a photograph. Conceptualization concerns a subject's ability to make a mental construct of an amount of previously consumed food and to relate that to a portion displayed on a photograph. Memory affects the precision and validity of the conceptualization.

Photo series for food portion estimation should be validated before being included in a picture book. Despite several validation methods having been described in the literature, no gold standard method to validate food picture books has been established[8,13,16–20].

The present study was performed within the Pilot study in the view of a Pan-European dietary survey – Adolescents, adults and elderly (PILOT-PANEU) project[21]. The project's goals were to develop, test and evaluate the applicability of tools and procedures for conducting a dietary survey. The PANEU partners collected 24 h dietary recall data within a pilot survey and developed country-specific picture books for assisting estimation of the food consumption amount, by selecting up to eighty photo series out of the 140 from the original EPIC-Soft picture book[22]. Each country could add new photo series if needed. The picture book was developed in the context of the European Prospective Investigation into Cancer and Nutrition (EPIC) study in 1994/95[22]. The foods and portions included were selected using data from ten European countries involved in the EPIC study. Data used to develop the picture book, including the portions, were based on pilot studies developed at the beginning of the EPIC study. For example, pilot data were used to choose the lower and higher weight extremes for each series of photos, based on the daily consumption[23].

The aim of the present study was to validate the EPIC-Soft picture book for portion size estimation among adolescents, adults and elderly. Two approaches were addressed: (i) 'perception', where prepared food portions on standard plates were estimated using EPIC-Soft food photo series; and (ii) 'memory', using the same photos series to estimate the food portions self-served (and in most of the cases also consumed) one hour before the estimation.

Methods

Subjects

As previously described[21], five PANEU countries were involved in the validation study: Bulgaria, Finland, Germany, Hungary and Portugal. Participants were invited by the most convenient way in each country. We aimed to include adolescents, adults and elderly. For recruitment, researchers' networks were used (e.g. friends, social networks, office intranet). Each invited person was asked to bring persons (friends or relatives) to participate in the study. To recruit adolescents, schools were also involved. In Bulgaria, participants were schoolchildren from the 19th High School of Sofia, university students from the Department of Hygiene, Medical Ecology and Nutrition at the Medical University of Sofia, employees from the National Centre of Public Health and Analyses, and individuals from the lists of general practitioners from the city of Sofia. In Finland, volunteers from a working unit at the Division of Welfare and Health Promotion were invited to participate. The invitation also addressed the children or adolescents belonging to the family of the staff. In Germany, participants were recruited via personal contact, posters and leaflets in the canteen of the Federal Institute for Risk Assessment; adolescents from schools in the vicinity, from the 'Girls and Boys Day' of the Federal Institute for Risk Assessment and from Berliner Schüler Unternehmen were also recruited. In Hungary, participants were invited from the Budapest Business School and from the Hungarian Food Safety Office staff. Adolescents were also recruited from a general school. In Portugal, volunteers were recruited by an email sent to students, professors and workers on a mailing list of the University of Minho. They were invited to participate and were asked to bring friends and family of different age groups to their own.

Sample size was calculated through the equation proposed by Nelson and Haraldsdottir[16]. To detect a difference of 25 % with a significance of 5 %, power of 80 % and assuming a CV of 50 %, a minimum of thirty-two independent portion estimates (individuals carrying out the estimation) were calculated for the perception method and sixty-four for the memory method. The final sample

size by country was: Bulgaria (*n* 103), Finland (*n* 34), Germany (*n* 69), Hungary (*n* 62) and Portugal (*n* 77).

Picture book

For validation purposes, twenty-three food items associated to twenty-three photo series (approximately 25 % of the photos in the picture book) were allocated to the five participating countries to reduce the number of validations needed. Food items were allocated to each country according to their degree of association with the country's typical gastronomy and by checking food items listed in the Food Propensity Questionnaire developed by the PILOT-PANEU project. Each partner validated seven photo series (most of them were tested in two countries), corresponding to seven food items. For the portion estimation based on perception, three portion sizes from each food series were randomly selected. For the portion estimation based on memory, the same food items were grouped into two meals. The weight in grams of the portions for each photo series can be seen in the online supplementary material, Supplemental Table 1. A validation protocol, registration form and record sheets were developed to assure comparable data collection in the five countries.

Procedures for validation

In Bulgaria, Finland and Portugal the food portions were estimated by using a scale[24] presenting the options 1–6, each number corresponding to a portion in the EPIC-Soft picture book. The scale had a middle point between adjacent portions (e.g. 3·5 represents a portion between the 3rd and 4th portion) and the option of choosing a portion lower or higher than the smallest or biggest portion of the photo series, respectively. In Germany and Hungary the participants used a decimal scale between adjacent food portions, as described before[25]. The scale was extended with five decimals below the smallest quantity and five decimals above the biggest portion.

Perception test

Weighed food servings were presented to subjects who were asked to compare them with food portions shown in photo series and mark the results on a record sheet. Each food portion was served on a white plate of 26 cm diameter according to the International Agency for Research on Cancer's guidelines for the preparation of the EPIC-Soft picture book[26]. The twenty-one portions were randomly distributed on a table, except in the case of Hungary, which placed the three test portions together. The respective printed photo series were placed on the table next to the food item (see online supplementary material, Supplemental Figures 1 and 2). For each food item, the weight of the food serving was set in advance and was kept constant throughout the study. For the photo series that presented two or more different formats of the same food ('cheese, pieces', 'pizza', 'mushrooms', 'grapes' and 'omelette/scrambled eggs'), only one format was displayed on the plate. For the photo series for 'fish fillets' and 'fish slices', the foods displayed were not exactly the same (e.g. the size of the fish was different).

Memory test

The memory method tests were conducted in Germany, Hungary and Portugal. The food items validated through this method were the same ones used in the perception method. Participants were instructed to serve themselves the amount of food that they would normally consume. The served food was weighed using a digital weighing scale and recorded by the study researchers. After that, the participants were involved in different activities for about an hour, followed by estimation of the amount of food served with the help of the same picture series that were used for the perception method, and marked on the record form.

Statistical analyses

Perception method

To test components of variance in the reported portions we performed mixed-effects models[27]. By calculating the variance components, it is possible to determine which factors explain the difference between observers. The components of the model were the specific portion size shown to the participant in each country (designated by 'plate'), the observer (participant) and the type of food. Because among different foods, different portions were shown to participants, the mixed-effect models were also adjusted for the mean of the portions shown to the participants for each food.

To assess validity of the study, the mean difference between the portion reported by the participant and the real portion of the presented food servings was calculated (reported portion minus shown portion). To test if the differences were acceptable, the mean and SD were calculated, as well as the percentage of observers who selected the correct, adjacent or distant picture. The differences were considered acceptable if the mean difference between reported and shown portions was equal to or lower than 0·5 and if the SD was equal to or lower than 1. The portion was classified has being overestimated (or underestimated) if the mean difference was higher than 0·5 (or lower than −0·5) and if the SD was equal to or lower than 1. If the difference had an SD higher than 1 portion, the estimation was considered 'not precise'. To assess agreement in the reported portions among observers, for each food, the intraclass correlation coefficient (ICC) by observer and the respective standard deviation of measurement error (SDME) were calculated.

Memory method

To test the components of variance in the memory study we also performed mixed-effects models, adjusted for the mean portion of each food served by the participants.

The variance components were the observer (participant) and the type of food included in the study. To test the validity in the memory study, the mean and SD of the reported and served food portions were calculated, as well as the difference between the reported and served portions. To test if the differences were statistically significant, the paired Wilcoxon test was used. The ICC and the respective SDME were also calculated to assess the agreement between the reported portions and true pictures, for each food.

All statistical analyses were performed using the statistical software package IBM SPSS Statistics version 19.0 and the software R version 2.12.1.

Criteria for classification of bias

To analyse the mean difference between the real portions and reported portions, the following criteria were defined: differences lower than 0·25 portion indicate no bias; differences between 0·25 and 0·5 portion indicate low bias; differences between 0·5 and 1·0 portion indicate moderate bias; and differences higher than 1·0 portion indicate large bias. Guidelines for interpreting ICC statistics suggest that values of 0·81–1·00 indicate almost perfect agreement, 0·61–0·80 substantial agreement, 0·41–0·60 moderate agreement, 0·21–0·40 fair agreement, and values less than 0·21 indicate a poor or slight agreement[28]. If SDME is lower than 0·5 portion it indicates no error, if it is between 0·5 and 1·0 portion it indicates moderate error, and if it is higher than 1·0 portion it indicates large error.

Results

Participant characteristics

Table 1 shows the characteristics of the whole sample participating in the validation study. A total of 345 individuals (59 % female) participated in the perception study and 208 individuals (58 % female) participated in the memory study. The mean age of participants was 34 years, ranging from 8 to 77 years. About 70 % of participants had high education.

The weight in grams of the food shown (perception study) or served (memory study) as well as weight estimated by the participants can be checked in the online supplementary material (Supplemental Table 2).

Perception study

Table 2 presents the percentage of participants choosing the correct, adjacent or distant picture when comparing reported and shown portions. The percentage of participants choosing the correct picture varied from 18 % for pieces of cheese to almost all for estimations of ratatouille (96 %). Between 0 and 30 % of the participants chose a distant picture (difference of more than 1 portion for each food).

Sources of variance in reported portions

Table 3 presents the components explaining the variance found in the reported portion. The main source of variance in the measurement was the plate (the specific portion size shown to the participant in each country; 67·8 %). The results suggested a small systematic higher or lower estimate for some foods even after adjusting for the mean portion shown for each food (5·3 %). The SDME (= 0·611) can be classified as moderate. The interaction between food and observer explained a small part (3·8 %) of the total variance.

Validity: agreement between shown and reported portions

Overall, the agreement between the portions shown and reported was substantial (ICC = 0·805). Table 4 describes the portions shown and reported, the respective difference and the agreement among observers by food. For several foods, differences between reported and shown mean portions (underestimation and overestimation) were detected. For the majority of foods, the observer ICC

Table 1 Characteristics of the sample (all countries), by age and sex groups

	All*		10–17 years Adolescents		18–64 years Adults		≥ 65 years Elderly	
	Female (n 200)	Male (n 136)	Female (n 56)	Male (n 47)	Female (n 123)	Male (n 71)	Female (n 21)	Male (n 18)
	%	%	%	%	%	%	%	%
Age (years)								
Mean	33·5	34·0	12·5	12·9	36·8	38·8	70·4	70·4
SD	19·8	21·2	2·7	2·3	13·6	14·0	3·6	3·9
Education†								
No education	2·2	0·0	–	–	1·2	0·0	16·7	0·0
Low	5·6	8·7	–	–	6·0	2·6	0·0	20·0
Medium	22·2	19·6	–	–	19·0	15·8	66·7	40·0
High	70·0	71·7	–	–	73·8	81·6	16·7	40·0

*Nine participants with missing information on age.
†The level of education was asked only of participants aged ≥18 years.

Table 2 Percentage of participants choosing the correct, adjacent or distant picture when comparing food portions with photo series (perception study)

ID	Food	n	Correct (%) 0	Adjacent (%) −1	Adjacent (%) 1	Distant (%) >−1	Distant (%) >1
1	Potatoes, boiled	72	46·4	3·55	44·9	1·73	3·40
3	French fries	77	58·7	34·4	0·87	3·90	2·17
8	Green salad	179	85·0	1·90	11·5	0·00	1·70
13	Carrots	77	54·6	42·4	0·40	0·90	1·70
15	Tomatoes	56	28·3	0·90	63·8	1·21	5·79
25	Apple compote	67	69·1	0·50	24·4	0·00	5·90
32	Cheese, pieces	76	17·8	49·7	2·63	29·4	0·43
34	Pasta	84	28·8	56·6	2·80	11·6	0·20
36	Rice	85	82·4	11·3	4·43	1·30	0·33
42	Steak	58	64·5	25·8	4·00	5·30	0·42
51	Fish, fillets	83	33·7	2·42	52·3	0·35	11·2
52	Fish, slices	77	79·1	9·60	7·40	0·00	3·90
64	Fruit cake	48	52·5	1·80	42·5	0·00	3·20
74	Pizza	77	31·3	0·40	39·1	0·00	29·2
75	Spaghetti	72	80·3	9·10	8·50	1·00	4·80
79	Soup	99	47·8	33·3	1·90	16·7	0·40
91	Mushrooms	48	19·2	50·2	0·00	30·1	0·48
92	Grapes	102	63·5	16·7	2·84	16·3	0·78
104	Stew	72	81·0	5·60	12·8	0·30	0·30
302	Omelette/scrambled eggs	84	69·9	16·6	9·80	1·30	2·50
402	Risotto	77	55·0	2·20	21·2	2·20	19·9
507	Ratatouille	68	96·0	1·00	2·97	0·00	0·00
517	Creamed spinach	72	61·1	28·1	7·30	6·40	1·60

ID, identification number of each photo series in the EPIC-Soft picture book.

Table 3 Components of variance in the reported portion after adjusting for the mean portion shown in each food (perception study)

Effect	n	Variance	SD	% of total variance*
Main effects				
Plate	159	1·298	1·139	67·8
Observer	344	0·067	0·259	3·5
Food	29	0·101	0·318	5·3
Food × Observer	2834	0·073	0·270	3·8

Effect	n	Variance	SDME	% of total variance*
Residual				
Food × Observer	9134	0·373	0·611	19·5

SDME, standard deviation of measurement error.
*Variance of each component divided by the total variance.

(agreement among observers) was substantial or excellent and the SDME for each food ranged from 0·5 to 1·0 (moderate error). The pictures evaluated were considered acceptable and precise enough (no or small bias) to estimate portion size for green salad, apple compote, rice, steak, slices of fish, spaghetti, stew, omelette/scrambled eggs and ratatouille. Photo series depicting French fries, carrots, pieces of cheese, pasta, soup, mushrooms, grapes and creamed spinach were all underestimated. On the other hand, pictures showing boiled potatoes, tomatoes, fish fillets, fruit cake, pizza and risotto were all overestimated. Although boiled potatoes, French fries, carrots, tomatoes, fruit cake, soup, grapes, risotto and creamed

spinach showed a moderate bias, a moderate to almost perfect agreement among observers was described. With the exceptions of the picture series of green salad, carrots, pieces of cheese, pasta, fish filets, fish slices, pizza, mushrooms and omelette/scrambled eggs, the agreement among observers was substantial to almost perfect ($0·61 < \text{ICC} < 1·00$). Additionally, except for pictures of grapes, all picture series that had the same portion of food presented in different formats (e.g. round and square pizza slice) presented low agreement among observers.

Memory study

Table 5 shows that the major source of variance after adjusting for the served mean portion was the food (49·0 %); these results suggested a large systematic higher or lower estimation for different foods and that the observer explained a small part of the variance (8·1 %).

Stratification of the results by food confirmed that several foods showed low agreement between served and reported portions (Table 6). The overall agreement between the served and the reported portions was lower than in the perception study ($\text{ICC} = 0·536$). Only carrots, apple compote, steak, fish slices, fruit cake and creamed spinach showed substantial agreement, while green salad, fish fillets, spaghetti, stew, omelette/scrambled eggs and ratatouille showed moderate agreement. In general, overestimated (or underestimated) foods in the perception approach were similarly overestimated (or underestimated) in the memory study. Considering the mean differences between reported and served portions for the

Table 4 Shown and reported food portions, the respective difference and the agreement among observers by food (perception study)

ID	Food	No. of portions (n)	No. of observations (n)	Shown mean	Reported mean	Shown SD	Reported SD	Difference (reported portion minus shown portion) Mean	SD	Evaluation*	Bias†	ICC observer	SDME observer
1	Potatoes, boiled	6	96	3·61	4·19	1·20	1·44	0·58	0·24	Overestimated	Moderate bias	0·699	0·477
3	French fries	3	77	3·17	2·58	1·03	1·18	−0·59	0·15	Underestimated	Moderate bias	0·678	0·721
8	Green salads‡	6	180	1·50	1·66	0·41	0·68	0·16	0·27	Acceptable	No bias	0·403	0·473
13	Carrots	3	77	2·17	1·42	0·62	0·84	−0·75	0·22	Underestimated	Moderate bias	0·523	0·612
15	Tomatoes	9	171	2·72	3·42	1·41	1·08	0·70	0·33	Overestimated	Moderate bias	0·668	0·578
25	Apple compote	6	68	1·87	2·31	0·86	1·32	0·44	0·46	Acceptable	Low bias	0·856	0·467
32	Cheese, pieces§	3	77	2·83	1·53	0·63	0·94	−1·30	0·31	Underestimated	Large bias	0·148	0·510
34	Pasta	9	171	3·49	2·45	1·00	1·12	−1·04	0·12	Underestimated	Large bias	0·527	0·656
36	Rice	9	171	2·89	2·65	0·96	1·16	−0·24	0·20	Acceptable	No bias	0·724	0·564
42	Steak	6	96	3·34	3·16	1·07	1·28	−0·18	0·21	Acceptable	No bias	0·694	0·510
51	Fish, fillets‡	9	171	2·75	3·69	0·77	1·03	0·94	0·26	Overestimated	Moderate bias	0·372	0·686
52	Fish, slices§	3	77	1·50	1·61	0·41	0·91	0·11	0·50	Acceptable	No bias	0·104	0·763
64	Fruit cake	6	96	2·30	2·83	0·96	1·43	0·53	0·47	Overestimated	Moderate bias	0·816	0·498
74	Pizza	3	96	2·13	3·49	0·55	1·15	1·36	0·60	Overestimated	Large bias	0·471	0·623
75	Spaghetti	6	77	3·06	2·98	1·57	1·70	−0·08	0·13	Acceptable	No bias	0·867	0·541
79	Soup	9	171	2·81	2·03	1·71	1·11	−0·78	0·60	Underestimated	Moderate bias	0·779	0·488
91	Mushrooms	6	96	4·62	3·05	1·32	1·12	−1·57	0·20	Underestimated	Large bias	0·593	0·383
92	Grapes	9	171	2·88	2·35	1·49	1·18	−0·53	0·31	Underestimated	Moderate bias	0·675	0·665
104	Stew	6	96	2·73	2·89	1·32	1·50	0·16	0·18	Acceptable	No bias	0·915	0·374
302	Omelette/scrambled eggs‡	9	171	1·76	1·67	0·59	0·75	−0·09	0·16	Acceptable	No bias	0·322	0·564
402	Risotto	3	77	3·17	3·88	1·44	1·52	0·71	0·08	Overestimated	Moderate bias	0·827	0·738
507	Ratatouille	6	96	2·11	2·18	0·83	0·91	0·07	0·08	Acceptable	No bias	0·824	0·347
517	Creamed spinach	6	68	3·37	2·81	1·47	1·49	−0·56	0·02	Underestimated	Moderate bias	0·817	0·520

ID, identification number of each photo series in the EPIC-Soft picture book; ICC, intraclass correlation coefficient; SDME, standard deviation of measurement error.
*Established series category based on mean and SD of difference: acceptable, │mean difference│ ≤0·5 and SD ≤1; overestimation, │mean difference│ >0·5 and SD ≤1; underestimation, │mean difference│ <−0·5 and SD ≤1; not precise, SD >1.
†Classification of bias: no bias, │mean difference│ <0·25 portion; low bias, │mean difference│ between 0·25 and 0·5 portion; moderate bias, │mean difference│ between 0·5 and 1·0 portion; large bias, │mean difference│ >1·0 portion.
‡Fair agreement among observers.
§Poor agreement among observers.

Table 5 Components of variance in the reported portion after adjusting for the served mean portion of each food (memory study)

Effect	n	Variance	SD	% of total variance*
Main effects				
Observer	198	0·1008	0·3175	8·1
Food	23	0·6088	0·7803	49·0

Effect	n	Variance	SDME	% of total variance*
Residual				
Food × Observer	921	0·5316	0·7291	42·8

SDME, standard deviation of measurement error.
*Variance of each component divided by the total variance.

memory study, the pictures were considered acceptable and precise enough to estimate portion size for French fries, green salad, carrots, apple compote, steak, slices of fish, fruit cake, spaghetti, grapes, stew and ratatouille.

Discussion

Accurate portion size estimation is a key to a successful data collection on food intake. However, estimation of consumed portion sizes is a challenging task for subjects. The present study evaluated the validity of twenty-three food photo series. The results of this validation study carried out in five European countries showed that overall the picture series from the so-called EPIC-Soft picture book could be used to estimate food consumption in dietary surveys. Overall, the agreement decreased as the appearance of the food displayed differed from the food in the picture (e.g. fish, slices). The correct picture chosen by the participants ranged from 18 % (cheese) to 96 % (ratatouille). Between 0 and 30 % of the participants chose a distant picture. Currently there are no existing guidelines regarding the minimum level of acceptable accuracy for a picture to be used. A previous study[29] concluded that there is still a lack of consensus concerning expression of error rates and that precision in portion size estimation is not yet a realistic expectation. Our results seem to be close to the previous studies. In another study validating a food picture book to be used among children in pan-European national dietary surveys, it was described that between 37 and 71 % of the participants chose the correct picture and between 1 and 16 % of the participants chose a distant picture[18]. In another study among Mexican adolescents, 28·7 % of the foods were correctly quantified using a food picture book[30]. In Europe, a study serving predefined portions showed that 50 % of all estimations performed by adult volunteers (aged 25–65 years) were exactly correct[17].

Regarding the overall results only a tendency of under- or overestimation of portions was not observed, contrary to some previous studies that described a clear pattern of under- or overestimation[11,29,31]. However, analysing each photo series, it was found that the overestimation occurred more in smaller portions, while the underestimation occurred more in larger portions. This may simply be the result of having a greater number of photos for smaller portions than for larger portions in general and the tendency that subjects have of choosing something in the middle.

Concerning the perception study, the agreement found between the portions shown and reported was substantial and the bias was low, meaning that the food pictures were in general validly perceived. Typically, overestimated foods in the perception approach were similarly overestimated in the memory study. The same was true for underestimated foods. In the perception study, the main source of variance in the measurement method was the plates, which indicates that the major source of variance in the measurement was due to differences between portions on each plate. Consequently, observers were substantially successful in discriminating different portions. The minor differences found across countries did not seem to have affected the results of the validation study.

The variance in portion size estimation was small, reflecting small systematic differences among observers. This result highlights that there might not be significant differences between countries. On the other hand, the variance related to the measurement error was relatively large, suggesting that other important factors could explain the variance in the reported portions. One possible explanation for the low agreement is the fact that, for most foods, the mean portions tested in the perception study were far from the middle pictures of the photo series (between 2·5 and 3·5). From a total of ten food items with portions shown between 2 and 3, eight food items had an ICC higher than 0·41. From the remaining food items only four in thirteen had an ICC higher than 0·41. This results in a relative risk of 2·6 (95 % CI 1·09, 6·22) of a picture having a higher agreement if the mean portion shown has been situated between 2 and 3. This means that the original portions from the picture book presented a deviation since the portions of each food were chosen to quantify daily consumption and not quantity by meal occasion[22,23].

The differences between real and reported portions in both the perception and memory study were compared and overall overestimated foods in the perception approach were similarly overestimated in the memory study. The same is true for underestimated foods. The interaction between food and observer explained a small part of the total variance; this means that observers did not systematically estimate better (or worse) some foods than others.

The perception method involved only the direct comparison of the portions of food, prepared in advance, with the portions shown on the pictures. The results of the memory method provided additional information for estimation of the uncertainty of a dietary recall methodology[15]. Data obtained from the latter type of study provide

Table 6 Served and reported food portions, the respective difference and the agreement between served and reported portions by food (memory study)

ID	Food	No. of plates (n)	Served mean	Reported mean	Served SD	Reported SD	Difference (reported portion minus served portion) Mean	SD	Evaluation*	Bias†	ICC agreement	SDME agreement
1	Potatoes, boiled‡	53	2·46	3·61	0·80	1·13	1·15	0·33	Overestimated	Large bias	0·352	0·623
3	French fries‡	32	1·52	1·11	0·81	0·55	−0·41	0·26	Acceptable	Low bias	0·376	0·521
8	Green salad	30	0·96	0·83	0·82	0·42	−0·13	0·40	Acceptable	No bias	0·473	0·472
13	Carrots	43	1·21	0·98	0·64	0·45	−0·23	0·19	Acceptable	No bias	0·605	0·327
15	Tomatoes‡	31	1·31	2·21	0·66	0·88	0·90	0·22	Overestimated	Moderate bias	0·382	0·469
25	Apple compote	34	3·36	3·68	1·48	2·09	0·32	0·61	Acceptable	Low bias	0·754	0·886
32	Cheese, pieces‡	28	1·61	1·05	0·76	0·79	−0·56	0·03	Underestimated	Moderate bias	0·391	0·556
34	Pasta§	35	4·17	1·57	1·21	0·80	−2·60	0·41	Underestimated	Large bias	0·180	0·509
36	Rice§	34	3·54	1·56	0·50	0·71	−1·98	0·21	Underestimated	Large bias	0·083	0·426
42	Steak	52	2·44	2·69	1·12	1·12	0·25	0·00	Acceptable	No bias	0·739	0·554
51	Fish, fillets	31	1·93	2·75	0·93	1·32	0·82	0·39	Overestimated	Moderate bias	0·447	0·761
52	Fish, slices	43	2·56	2·62	1·01	1·04	0·06	0·03	Acceptable	No bias	0·727	0·534
64	Fruit cake	53	2·70	3·15	1·30	1·17	0·45	0·13	Acceptable	Low bias	0·711	0·615
74	Pizza‡	32	2·00	3·17	0·90	1·13	1·17	0·23	Overestimated	Large bias	0·392	0·606
75	Spaghetti	53	3·11	3·15	1·47	1·21	0·04	0·26	Acceptable	No bias	0·571	0·879
79	Soup‡	34	2·26	1·59	0·59	0·72	−0·67	0·13	Underestimated	Moderate bias	0·303	0·487
91	Mushrooms§	53	4·71	2·57	1·72	0·84	−2·14	0·88	Underestimated	Large bias	0·174	1·056
92	Grapes‡	34	2·06	1·98	1·34	0·78	−0·08	0·56	Acceptable	No bias	0·321	0·899
104	Stew	52	2·88	2·89	1·09	0·99	0·01	0·10	Acceptable	No bias	0·560	0·688
302	Omelette/scrambled eggs	34	2·50	1·77	1·30	0·79	−0·73	0·51	Underestimated	Moderate bias	0·530	0·639
402	Risotto§	43	1·11	1·76	0·32	0·48	0·65	0·16	Overestimated	Moderate bias	0·195	0·309
507	Ratatouille	34	1·88	2·09	0·71	0·68	0·21	0·03	Acceptable	No bias	0·512	0·477
517	Creamed spinach	53	3·75	2·88	1·89	1·41	−0·87	0·48	Underestimated	Moderate bias	0·616	0·918

ID, identification number of each photo series in the EPIC-Soft picture book; ICC, intraclass correlation coefficient; SDME, standard deviation of measurement error.
*Established series category based on mean and SD of difference: acceptable, |mean difference| ≤ 0·5 and SD ≤ 1; overestimation, |mean difference| > 0·5 and SD ≤ 1; underestimation, |mean difference| < −0·5 and SD ≤ 1; not precise, SD > 1.
†Classification of bias: no bias, |mean difference| < 0·25 portion; low bias, |mean difference| between 0·25 and 0·5 portion; moderate bias, |mean difference| between 0·5 and 1·0 portion; large bias, |mean difference| > 1·0 portion.
‡Fair agreement among observers.
§Poor agreement among observers.

a different sort of information and a more realistic estimation for the uncertainty of reported portion sizes during surveys. This method is intended to simulate the dietary assessment method used in surveys, where people recall the foods eaten 24 h before. In the memory study, there were significant differences when comparing served and reported food portion sizes, where some foods such as potatoes, tomatoes and pizza were overestimated and others like carrots, pasta and mushrooms were underestimated.

Compared with perception results, the variance explained in the memory study by portion was lower and the error was higher. The error was higher in the memory study, increasing from 0·59 to 0·95 portions, and the agreement between served and reported portions was lower.

Picture series with low agreement and large differences (difference between reported and real portions greater than 1 portion) need to be replaced or other quantification method should be applied to these foods. With exception of the picture series of green salad, carrots, pieces of cheese, fish, pizza, mushrooms and omelette/scrambled eggs, the agreement among observers was in general substantial. The agreement seemed to decrease as the appearance of the food presented differed from the food in the picture (e.g. slices of fish). Additionally, except for grapes, a decreased agreement among observers was observed in all picture series that had the same portion of food presented in different formats (e.g. round and square pizza slice or different pieces of cheese).

Strengths and limitations

The described multicentre validation study was conducted in five different countries in Europe. This makes it more acceptable to generalize the results to other European countries. The inclusion of the conceptualization and memory dimensions, in addition to perception, is also a methodological strength as the food picture series are intended to be used to aid food quantification during recall methods. As there is no defining criterion for the level of acceptable accuracy in the evaluation of the picture series, we opted to evaluate the performance of each picture series based on agreement among observers and bias between reported and real portions. Another strength was to choose three random portions from each picture series that allowed to validate both pictures depicting small and large portions and to describe the tendency to over- or underestimation.

The study should have been performed with representative samples of adolescents, adults and elderly to have a clear idea about the range of bias in overall dietary intake using the food picture books. As that was not feasible in all study centres, convenience samples were drawn. Elderly and males in general were under-represented in our samples, which might have influenced the results. Also, the majority of the participants had high education, which could also have some influence on the accuracy of reported data. However, as the variance in the portion size's estimation was small among observers, we believe that these differences in sample did not greatly influence the results. Moreover, under-reporting might be an issue among obese participants. We had information regarding BMI of the participants for Germany (13 % of obesity) and Bulgaria (9 % of obesity), and performed a sensitivity analysis by excluding the obese participants in the memory study. We found similar results, and the mean differences between reported and real portion sizes and agreement between participants did not change significantly. Taking this into account, and that there were no statistical differences between the two countries regarding the distribution of BMI categories, we believe that the under-reporting among obese individuals might have only a minimum effect in our results.

Conclusions

The present results confirmed that overall the picture series included in the EPIC-Soft picture book can be applied in future intake surveys for foods similar to those depicted in the pictures. However, foods with substantially different shape cannot be quantified correctly by applying the current picture series. In such cases, we recommend that new picture series should be prepared, or some corrections should be applied. Other quantification methods can be used to estimate amount of foods accurately, such as household measures or standard units.

Acknowledgements

Germany did not complete a food survey and took part only in picture book validation where no approval was necessary). The project was also approved by the Portuguese Authority of Data Protection. All participants were informed about the benefits and potential discomfort and written informed consent forms were signed by all participants or by legal representatives in the case of children.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1368980017004153

References

1. European Food Safety Authority (2009) General principles for the collection of national food consumption data in the view of a pan-European dietary survey. *EFSA J* **7**, 1435.
2. European Food Safety Authority (2010) Declaration of the Advisory Forum on the Pan-European Food Consumption Survey. What's on the Menu in Europe? – A Pan-European Food Consumption Survey (EU-Menu). http://www.efsa.europa.eu/en/datexfoodcdb/datexeumenu.htm (accessed December 2016).
3. Brussaard JH, Lowik MR, Steingrimsdottir L *et al.* (2002) A European food consumption survey method – conclusions and recommendations. *Eur J Clin Nutr* **56**, Suppl. 2, S89–S94.
4. Slimani N, Ferrari P, Ocke M *et al.* (2000) Standardisation 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* **54**, 900–917.
5. Biro G, Hulshof KF, Ovesen L *et al.* (2002) Selection of methodology to assess food intake. *Eur J Clin Nutr* **56**, Suppl. 2, S25–S32.
6. Rutishauser IH (2005) Dietary intake measurements. *Public Health Nutr* **8**, 1100–1107.
7. Byrd-Bredbenner C & Schwartz J (2004) The effect of practical portion size measurement aids on the accuracy of portion size estimates made by young adults. *J Hum Nutr Diet* **17**, 351–357.
8. Cypel YS, Guenther PM & Petot GJ (1997) Validity of portion-size measurement aids: a review. *J Am Diet Assoc* **97**, 289–292.
9. Venter CS, MacIntyre UE & Vorster HH (2000) The development and testing of a food portion photograph book for use in an African population. *J Hum Nutr Diet* **13**, 205–218.
10. Huybregts L, Roberfroid D, Lachat C *et al.* (2008) Validity of photographs for food portion estimation in a rural West African setting. *Public Health Nutr* **11**, 581–587.
11. Frobisher C & Maxwell SM (2003) The estimation of food portion sizes: a comparison between using descriptions of portion sizes and a photographic food atlas by children and adults. *J Hum Nutr Diet* **16**, 181–188.
12. Lillegaard IT, Overby NC & Andersen LF (2005) Can children and adolescents use photographs of food to estimate portion sizes? *Eur J Clin Nutr* **59**, 611–617.
13. Subar AF, Crafts J, Zimmerman TP *et al.* (2010) Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *J Am Diet Assoc* **110**, 55–64.
14. Martin CK, Nicklas T, Gunturk B *et al.* (2014) Measuring food intake with digital photography. *J Hum Nutr Diet* **27**, Suppl. 1, 72–81.
15. Nelson M, Atkinson M & Darbyshire S (1994) Food photography. I: The perception of food portion size from photographs. *Br J Nutr* **72**, 649–663.
16. Nelson M & Haraldsdottir J (1998) Food photographs: practical guidelines I. Design and analysis of studies to validate portion size estimates. *Public Health Nutr* **1**, 219–230.
17. Ovaskainen ML, Paturi M, Reinivuo H *et al.* (2008) Accuracy in the estimation of food servings against the portions in food photographs. *Eur J Clin Nutr* **62**, 674–681.
18. Trolle E, Vandevijvere S, Ruprich J *et al.* (2013) Validation of a food quantification picture book targeting children of 0–10 years of age for pan-European and national dietary surveys. *Br J Nutr* **110**, 2298–2308.
19. Naska A, Valanou E, Peppa E *et al.* (2016) Evaluation of a digital food photography atlas used as portion size measurement aid in dietary surveys in Greece. *Public Health Nutr* **19**, 2369–2376.
20. Bouchoucha M, Akrout M, Bellali H *et al.* (2016) Development and validation of a food photography manual, as a tool for estimation of food portion size in epidemiological dietary surveys in Tunisia. *Libyan J Med* **11**, 32676.
21. Ambrus Á, Horváth Z, Farka Z *et al.* (2013) Pilot Study in the View of a Pan-European Dietary Survey – Adolescents, Adults and Elderly. http://www.onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2013.EN-508/pdf (accessed March 2017).
22. Van Kappel A, Amoyel J, Slimani N *et al.* (1995) *EPIC-Soft Picture Book for Estimation of Food Portion Sizes*. Lyon: International Agency for Research on Cancer.
23. Kaaks R, Slimani N & Riboli E (1997) Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European prospective investigation into cancer and nutrition. *Int J Epidemiol* **26**, Suppl. 1, S26–S36.
24. EFSA/EU Menu Guidance (2013) Example of a Protocol for Validation of a Food Picture Book Based on the PILOT-PANEU Project. http://www.efsa.europa.eu/sites/default/files/efsa_rep/blobserver_assets/3944A-6-3-3.pdf (accessed March 2017).
25. Szenczi-Cseh J, Horváth Z & Ambrus Á (2017) Validation of a food quantification picture book and portion sizes estimation applying perception and memory methods. *Int J Food Sci Nutr* **68**, 960–972.
26. Crispim S, Nicolas G, Freisling H *et al.* (2011) *Preparation of Photos for Food Portion Quantification in EPIC-Soft – Reference Guidelines*. Lyon: IARC.
27. McGraw KO & Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* **11**, 30–46.
28. Landis JR & Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.
29. Hernández T, Wilder L, Kuehn D *et al.* (2006) Portion size estimation and expectation of accuracy. *J Food Compost Analysis* **19**, Suppl, S14–S21.
30. Bernal-Orozco MF, Vizmanos-Lamotte B, Rodriguez-Rocha NP *et al.* (2013) Validation of a Mexican food photograph album as a tool to visually estimate food amounts in adolescents. *Br J Nutr* **109**, 944–952.
31. Turconi G, Guarcello M, Berzolari FG *et al.* (2005) An evaluation of a colour food photography atlas as a tool for quantifying food portion size in epidemiological dietary surveys. *Eur J Clin Nutr* **59**, 923–931.