Validation of a food quantification picture book targeting children of 0–10 years of age for pan-European and national dietary surveys

Ellen Trolle[1]*, Stefanie Vandevijvere[2], Jiří Ruprich[3], Majken Ege[1], Marcela Dofková[3], Evelien de Boer[4] and Marga Ocké[4]

[1]*National Food Institute, Technical University of Denmark (DTU), Moerkhoej Bygade 19, 2860 Soeborg, Denmark*
[2]*Scientific Institute of Public Health (SIPH), Brussels, Belgium*
[3]*National Institute of Public Health (NIPH), Brno, Czech Republic*
[4]*National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands*

Abstract

The aim of the present study was to validate thirty-eight picture series of six pictures each developed within the PANCAKE (Pilot study for the Assessment of Nutrient intake and food Consumption Among Kids in Europe) project for portion size estimation of foods consumed by infants, toddlers and children for future pan-European and national dietary surveys. Identical validation sessions were conducted in three European countries. In each country, forty-five foods were evaluated; thirty-eight foods were the same as the depicted foods, and seven foods were different, but meant to be quantified by the use of one of the thirty-eight picture series. Each single picture within a picture series was evaluated six times by means of predefined portions. Therefore, thirty-six pre-weighed portions of each food were evaluated by convenience samples of parents having children aged from 3 months to 10 years. The percentages of participants choosing the correct picture, the picture adjacent to the correct picture or a distant picture were calculated, and the performance of individual pictures within the series was assessed. For twenty foods, the picture series performed acceptably (mean difference between the estimated portion number and the served portion number less than 0·4 (sd <1·1)). In addition, twelve foods were rated acceptable after adjustment for density differences. Some other series became acceptable after analyses at the country level. In conclusion, all picture series were acceptable for inclusion in the PANCAKE picture book. However, the picture series of baby food, salads and cakes either can only be used for foods that are very similar to those depicted or need to be substituted by another quantification tool.

Key words: Food picture books: Food intake estimation: Dietary surveys: Children

Reliable and harmonised food consumption data are an essential element for exposure assessment. Especially, data on food consumption among children are of interest because children are at higher risk due to relatively higher levels of exposure. Until now, national dietary surveys have provided food consumption data for the risk assessment tasks of the European Food Safety Authority (EFSA). However, the existing data were collected with different dietary assessment methodologies, and there is a lack of data from some countries and some population groups, e.g. children and infants. Therefore, the EFSA is planning to organise a first pan-European food consumption survey, EU Menu, in collaboration with Member States during the next 10 years.

The EFCOSUM (European food consumption survey method) and the EFCOVAL (European Food Consumption Validation) projects[1–4] have provided recommendations on the design and overall methodology of surveys for adults and children, respectively. Several methodological challenges have to be addressed still within the work of implementing a harmonised data collection process across European countries. One important issue is how to ensure the best estimation of the amount of foods consumed. The weighed record method is considered too burdensome for large-scale national surveys where a representative sample and a high response rate are important. A combination of various tools has been used instead. For example, sixteen of twenty adult national dietary surveys included in the EFSA comprehensive food consumption database have reported the use of a picture book besides the use of household measures and standard units[5]. As recommended by the EFCOVAL project, these tools need to be country- and age-specific and have to be developed based on knowledge about foods on the national

market and food preferences of children and on using weighed records identifying ranges of portion sizes for different foods and recipes[4]. Some studies have shown that providing children with food photographs depicting age-specific portion sizes substantially improved the accuracy of portion size estimates[6,7].

The PANCAKE (Pilot study for the Assessment of Nutrient intake and food Consumption Among Kids in Europe) project aimed to develop, test and evaluate protocols and tools for a pan-European food consumption survey for infants, toddlers and children aged 0–10 years. One objective of the PANCAKE project was to develop a picture book for these age groups, aiming to ensure the best possible estimates of portion sizes. Another objective was to provide the participants within the PANCAKE pilot studies (parents of children or other proxies) with a visual quantification tool (in addition to g/ml, household measures, standard portions and standard units) for use with the developed food diaries. Several factors may influence the apparent size of a food in a picture: the size and colour of the surroundings such as the size of the food relative to the plate and relative to the picture size; the position of the food on the plate; and the actual weight, volume and characteristics of the food, such as shape, colour and firmness[8,9]. The aim of the present study was to validate the perception of the portion sizes of the foods in the picture series among parents of children aged from 3 months to 10 years in three culturally diverse countries in Europe.

Methods

The picture book

The PANCAKE picture book included picture series for foods commonly eaten in the countries of the PANCAKE consortium, derived from population-based data, foods that cannot be estimated well using household measures or units and that differ considerably from other foods in shape or size[8]. In order to guarantee the practical use of the PANCAKE picture book for the participants, the number of picture series was limited to thirty-eight common food picture series, including three series of baby foods; six series of butter and 'fillings/spreads' on bread; eight series of breakfast cereals, rice, pasta and potatoes; nine series of meat and fish; eight series of vegetables and four series of cakes and pies (Table 1). Each country could then add country-specific picture series, if needed.

The aim was to develop common age-appropriate picture series to be shared by the countries in the pilot study. However, only one picture book was developed to cover the portion sizes for children within the whole age range of 0–10 years, because it would reduce the risk of interviewer bias and because it was possible to cover the range of portion sizes for this age group with six pictures within a series. The 5th and 95th percentiles of Danish weighed portion sizes of foods consumed by children in the relevant age range were used for defining the range from the smallest to the largest portion of each series. Weighed data were not available for Belgium and the Czech Republic. Foster et al.[10] used the equal increments on a log scale with reference to Weber's law, which is based on perceptual research – the just

Table 1. Food items from the PANCAKE (Pilot study for the Assessment of Nutrient intake and food Consumption Among Kids in Europe) picture book served and presented during the validation sessions

Food items presented together on a plate
1. Chicken, rice and peas
2. Spaghetti, meat sauce and squash
3. Fish, boiled potatoes and grated carrots
4. Hamburger and French fries with a mixed salad
5. Meatballs, pasta (other shapes) and tomato slices
6. Meat cubes, cauliflower and cherry tomatoes
7. Steak, French fries (other shapes) and green lettuce
8. Lasagne, cucumber and carrots
9. Casserole, dark, with mashed potatoes and maize
10. Casserole, light, with boiled potatoes and broccoli

Food items presented on a separate plate
11. Butter on toast
12. Butter – as a block with a knife
13. Cheese, hard, sliced on toast
14. Cheese, spread on toast
15. Jam on toast
16. Mayonnaise salad on toast
17. Cornflakes
18. Oats and muesli
19. Porridge (baby)
20. Fruit purée (baby)
21. Vegetable purée (baby)
22. Cake, round and tall
23. Cake, square and flat (as brownies)
24. Cake, round (as a tart)
25. Cake, rectangular

notable difference gets bigger, usually in proportion to the stimulus magnitude (when the portion sizes get bigger). However, pre-trials to the present study indicated that the log scale might be in favour when the picture series should cover a wide range and the smallest portion size was above a certain size since it would ensure the visibility of the differences between the largest portions. The equidistant approach seemed to be more suitable for ensuring the visibility of the differences between the smallest portions for foods with a narrower range to cover, starting with a small portion. For approximately half of the common picture series, the portion sizes were distributed according to the equidistant approach (Table 4, equidistant (e) and almost equidistant (varying distance) (e_irreg)), while a mix between a log scale and the equidistant approach was used for the other half, using smaller increments between the smallest portion sizes than between the largest portion sizes (Table 4, increasing distance (i) and mixed (m)).

The photos were taken by a professional photographer in a special photo set-up to ensure the right technical parameters of pictures (e.g. resolution, lighting, contrast and colour balance, besides, for example, photo angles). The tableware was neutral, white, and of a commonly used size. These technical specifications have been described in detail elsewhere (PANCAKE web-library www.kostvaner.dk/pancake), based on the previous work with photo series in Denmark[11,12].

The country-specific bread shapes and pictures of glasses, cups and bowls were not validated in the present study.

Validation study

Subjects were selected using convenience sampling of parents having children aged from 3 months to 10 years. Participants were preferably those who were responsible for the child's food preparation. The aims were to include 108 parents from each country with a minimum of 25 % among each age group, infants (3–11 months), toddlers (1–3 years) and children (4–10 years), and to cover different parental educational levels. In Denmark, parents were recruited at three worksites. In Belgium, parents were recruited via the National Register, via letters distributed to the parents of children and infants in the surrounding schools and day-care centres, and via posters and leaflets distributed in the canteen of a university, a scientific institute, a train station and supermarkets. In the Czech Republic, parents were recruited via leaflets distributed in blocks of flats, supermarkets, schools and kindergartens in Brno city. The potential participants were invited by an explanatory letter. Incentives in Denmark were two cinema tickets of the participant's own choice, in Belgium a voucher of €20 was offered, and participants in the Czech Republic were given 500 CZK (€20). The validation sessions took place in the canteen of worksites or research institutes. The present study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the ethical committee of the University of Ghent, Belgium. Verbal informed consent was obtained from all subjects. Verbal consent was witnessed and formally recorded.

To ensure comparable data collection in the three countries, common materials were prepared, such as registration forms and files, coding system and labels, templates for data and instructions for the research staff.

The basic design was that the thirty-eight picture series were evaluated using forty-five foods. Of these, thirty-eight foods were similar to the depicted foods, whereas the seven additional foods – squash, root vegetables, pasta (in other shapes), cauliflower, fried potatoes, mixed salad and maize – were used to validate the picture series for cucumber, potatoes (boiled), spaghetti, broccoli, potatoes (boiled), green salad and peas, respectively. For every food, each of the six pictures in a series was evaluated six times in each country, by serving predefined portion sizes: two portions smaller, two portions equal and two portions larger than the depicted portion: i.e. $\pm 1/3$ of the amount between the picture and the adjacent picture. In total, $45 \times 6 \times 6 = 1620$ pre-weighed servings were evaluated in each country.

In each country, 108 participants evaluated one serving size of fifteen foods each, meaning that in total $108 \times 15 = 1620$ evaluations were carried out. The serving sizes were randomly allocated between the participants. For practical reasons, the validation sessions were planned on three different days covering fifteen foods per d, with all eighteen serving sizes per food being evaluated at two almost identical sessions with eighteen participants each time, by thirty-six participants per food item.

The research staff dished up all the food items by weighing and coding all the predefined portion sizes. Two picture series differed from this procedure: poultry and tomatoes (whole), for which it was not possible to get exactly predefined portion sizes. Each of the six pieces of chicken shown in the picture series was served three times, the pieces were weighed while dishing up and the actual weight of the portions was noted. These eighteen portions were used in two sessions, implying that in all thirty-six evaluations were conducted. The same procedure was performed for eighteen whole tomatoes of different sizes within the range of 7 and 205 g.

For some food items, such as mixed salads and cakes, densities can vary substantially. However, for the appearance of the portion size, both the volume of the food and the weight of the food are important. Therefore, food volume measurements of all food items were performed in all three countries by a water displacement method described elsewhere (PANCAKE web-library www.kostvaner.dk/pancake). Densities of the food items were calculated from the weight in grams and the volume in millilitres. In addition, a photograph of all portion sizes was taken in all three countries.

The validation sessions started with a short introduction and instruction, which was given individually in Belgium, while in the Czech Republic and Denmark, it was given collectively to the participants. The participants were assigned a table where the plates with the weighed and coded food portions were placed together with the picture book. They evaluated the portion sizes of the different food items in a predefined random order. After evaluating the fifteen portions, each participant completed a short questionnaire about the difficulties encountered while using the picture series, relevance of the pictures, presentation of the pictures and the perception of the range of the pictures in the picture book (data not shown).

Statistical analyses

The pictures of each series were given numbers from 1 (smallest amount) to 6 (largest amount). For each food item, the agreement between the picture chosen by the participant and the picture depicting the correct portion size was assessed, and the percentages of participants choosing the correct picture, the picture adjacent to the correct picture or a distant picture were calculated, as in other studies[9,13].

In addition, the mean and standard deviation of difference between the portion number of the picture chosen by the participants and the portion number of the correct picture for each series were calculated. A SD of 1 was used as the limit of variance since estimations with a random error of ± 1 picture are considered as acceptable. For example, a SD equal to 1 is obtained if all estimates are $+1$ or -1, or when not more than 25 % of the estimates are $+2$ or -2 and 75 % are correctly estimated (values $= 0$), or when not more than 11 % are $+3$ or -3 and the remaining estimates are 0. The 99·997 % CI of the estimation error based on a normal distribution was used as the limit for bias of the mean. The high percentages of the CI were chosen because they give a broader range of acceptance where the mean value is considered not significantly different from 0 than a lower CI. For data obtained from all countries (n approximately 108 and SD $= 1$), the limit for bias of the mean is approximately 0·4, and for the individual countries separately (n approximately 36 and

SD = 1), the limit for bias of the mean is approximately 0·7, based on the 99·997 % CI of the estimation error. Differences between the countries were tested by the Kruskal–Wallis rank sum test. Spearman's correlation coefficients between the chosen picture number and the correct picture number were calculated for data obtained from all countries together and for those obtained from the individual countries separately. Data were re-analysed after adjustment for density by dividing the actual weight by its density.

The agreement between the chosen picture number and the correct picture number of the individual pictures of the series was shown in mosaic plots, which presented how well the individual pictures within a series performed. Estimation of the extent of overall over- and underestimation of the series and of the individual pictures within the series was done for all the evaluations separately by the estimated differences in grams and in gram percentage (estimated weight – served weight in grams – and as percentages of the served weight). Spearman's correlation coefficients between the weight of the chosen portion in grams and that of the served portion were calculated. Statistical calculations were carried out using the software package JMP 8.0 (SAS Institute, Inc.).

Results

The characteristics of the participants are listed out in Table 2. Parents of infants were under-represented in the present study, especially in Belgium and Denmark. Furthermore, higher educated women were over-represented.

Table 3 summarises for all food items the performance of the corresponding picture series, with and without adjustment for density differences, respectively. With regard to performance, the series were divided into four categories based on the mean

and SD of difference between the portion number of the picture chosen by the participants and the portion number of the correct picture: series that were rated as 'acceptable' (mean ≤ |0·4| and SD < 1·1); series that were 'acceptable after density adjustment' (mean > |0·4| or SD > 1·1 before density adjustment and mean ≤ |0·4| and SD < 1·1 after density adjustment); series that were misestimated even after density adjustment (overestimated: mean > 0·4 and SD < 1·1; underestimated: mean < −0·4 and SD < 1·1); series that did not give precise results (SD > 1·1).

Picture series that performed well

Among the twenty 'acceptable without adjustment' series, the percentage of participants who selected the correct picture was between 39 and 71 %, the percentage of participants who selected a distant picture varied between 1 and 16 %, and Spearman's correlation coefficients between the chosen picture number and the correct picture number were between 0·73 and 0·92. Significant differences with regard to the mean differences between the estimated portion number and the correct portion number were found between the countries for some of these series, but all means at the country level were within the acceptable range (mean < 0·7 and SD approximately 1), except the Danish mean values for mayonnaise filling and fried potatoes, which were −0·91 and 0·8, respectively, and the Belgian mean value for butter on toast, which was 1·09 (data at the country level not shown).

There were twelve series rated as 'acceptable after density adjustment' (Table 3). After adjustment for density, the percentage of participants who selected the correct picture in these series was between 37 and 67 %, the percentage of participants who selected a distant picture was between 4

Table 2. Characteristics of the participating parents in the validation study

(Number of subjects and percentages)

	Czech Republic		Belgium		Denmark		Total	
	n	%	n	%	n	%	n	%
Total	108	100	108	100	106	100	322	100
Sex								
Male	14	13	27	25	30	28	71	22
Female	94	87	80	74	76	72	250	78
Missing	0	0	1	1	0	0	1	0
Level of education								
10 years or below	4	4	3	3	2	2	9	3
10–12 years	55	51	10	9	5	5	70	22
13–15 years	14	13	19	18	27	25	60	19
Above 15 years	35	32	60	56	72	68	167	52
Missing	0	0	16	15	0	0	16	5
Age of child								
3–11 months	33	31	12	11	11	10	56	17
1–3 years	41	38	34	31	40	38	115	36
4–10 years	34	31	24	22	55	52	113	35
Missing	0	0	38	35	0	0	38	12
Used picture book before								
Yes	5	5	7	6	29	27	41	13
No	99	92	100	94	77	73	276	87
Missing	4	4	1	1	0	0	5	< 0·5

Table 3. Percentages of participants choosing the correct, adjacent or distant picture when comparing food items on plates with pictures within the PANCAKE (Pilot study for the Assessment of Nutrient intake and food Consumption Among Kids in Europe) picture book; Spearman's coefficients of correlation between the estimated picture number and the actual portion number; mean and standard deviation (SD) of the difference between the estimated picture number minus the actual portion number; established series category based on the mean and standard deviation; and Kruskal–Wallis test of significance of difference between the countries†

Dishes/foods	n	Correct 0 %	Adjacent −1 %	Adjacent 1 %	Distant <−1 %	Distant >1 %	Spearman's coefficients of correlation	Difference Mean	Difference SD	Series category Overall evaluation	Difference between the countries Kruskal–Wallis
Cornflakes	106	57	29	4	9	1	0·92	−0·4	0·8	Acceptable	
Cheese (hard) on toast bread	108	65	20	5	9	1	0·90	−0·3	0·8	Acceptable	*
Cucumber	108	52	34	6	5	4	0·85	−0·3	0·9	Acceptable	
Mayonnaise filling on bread	106	39	27	18	12	4	0·83	−0·3	1·0	Acceptable	***
Casserole (light)	108	58	23	13	5	1	0·90	−0·2	0·8	Acceptable	
Broccoli	108	58	28	12	1	1	0·90	−0·1	0·8	Acceptable	*
Rice	108	57	26	13	2	2	0·91	−0·1	0·7	Acceptable	
Potatoes (boiled)	108	49	21	21	7	2	0·84	−0·1	0·9	Acceptable	
Fruit purée	106	51	27	16	2	4	0·88	−0·1	0·8	Acceptable	*
Meat sauce	106	48	28	23	1	0	0·91	−0·1	0·7	Acceptable	**
Steak	106	42	23	24	7	5	0·82	−0·1	1·1	Acceptable	
Meatballs	108	71	9	14	5	1	0·91	0·0	0·7	Acceptable	***
Mashed potatoes	108	44	25	18	6	7	0·79	0·0	1·1	Acceptable	*
Jam on toast bread	108	57	13	23	5	2	0·86	0·0	0·8	Acceptable	
Root vegetables (pieces)	106	54	17	22	4	4	0·82	0·1	0·9	Acceptable	*
Casserole (dark)	108	60	16	21	0	3	0·92	0·1	0·7	Acceptable	*
Fried potatoes	102	51	13	27	3	6	0·86	0·2	0·9	Acceptable	***
Butter block	107	49	15	28	3	6	0·87	0·2	0·9	Acceptable	
Cheese spread on toast bread	108	49	12	25	4	10	0·84	0·3	1·0	Acceptable	*
Butter on toast bread	104	50	11	29	0	11	0·73	0·4	0·8	Acceptable	***
Hamburger	108	30 (44)	39 (28)	2 (16)	30 (11)	0 (2)	0·84 (0·79)	−1·07 (−0·35)	1·0 (1·04)	Acceptable after adjustment	(***)
Courgette	105	43 (67)	34 (16)	4 (13)	18 (2)	1 (2)	0·87 (0·88)	−0·70 (−0·05)	0·90 (0·70)	Acceptable after adjustment	
Maize	108	51 (54)	34 (29)	3 (7)	11 (9)	1 (1)	0·90 (0·88)	−0·53 (−0·38)	0·79 (0·79)	Acceptable after adjustment	
Tomatoes (slices)	108	56 (56)	25 (27)	6 (8)	13 (9)	0 (0)	0·86 (0·87)	−0·50 (−0·41)	0·93 (0·88)	Acceptable after adjustment	
Peas	108	56 (57)	31 (31)	3 (6)	9 (7)	0 (0)	0·91 (0·90)	−0·49 (−0·40)	0·75 (0·75)	Acceptable after adjustment	
Spaghetti	106	49 (50)	42 (42)	2 (4)	6 (4)	1 (1)	0·90 (0·89)	−0·49 (−0·42)	0·72 (0·77)	Acceptable after adjustment	* (**)
Cauliflower	108	40 (53)	51 (22)	2 (19)	4 (2)	4 (4)	0·86 (0·89)	−0·48 (0·03)	0·87 (0·77)	Acceptable after adjustment	(***)
Carrot (slices)	108	47 (56)	37 (47)	4 (5)	9 (8)	3 (3)	0·85 (0·85)	−0·45 (−0·32)	0·92 (0·88)	Acceptable after adjustment	(**)
Meat cubes/pieces	107	41 (37)	26 (15)	15 (32)	13 (6)	5 (10)	0·78 (0·76)	−0·33 (0·23)	1·13 (1·13)	Acceptable after adjustment	*** (***)
Oats or muesli	108	46 (45)	19 (21)	18 (21)	15 (9)	3 (3)	0·73 (0·77)	−0·31 (−0·17)	1·15 (1·04)	Acceptable after adjustment	*** (**)
Cake, round and tall (piece)	107	40 (59)	6 (10)	43 (21)	3 (4)	8 (7)	0·87 (0·87)	0·47 (0·14)	0·91 (0·89)	Acceptable after adjustment	* (***)
Fish	105	43 (53)	4 (11)	33 (24)	1 (1)	19 (10)	0·83 (0·85)	0·70 (0·31)	1·06 (0·96)	Acceptable after adjustment	** (**)

British Journal of Nutrition

Table 3. Continued

Dishes/foods	n	Correct	Adjacent		Distant		Spearman's coefficients of correlation	Difference		Series category Overall evaluation	Difference between the countries Kruskal–Wallis
		0	−1	1	<−1	>1		Mean	SD		
		%	%	%	%	%					
Pasta, penne or fusilli	108	40	47	2	11	0	0·92	−0·7	0·7	Underestimating	
French fries	108	54	21	7	18	1	0·84	−0·5	1·0	Underestimating	
Carrots (grated)	106	48	37	6	9	1	0·89	−0·5	0·8	Underestimating	***
Lasagne	108	45	6	33	5	11	0·79	0·5	1·1	Overestimating	*
Cake, square and flat	108	37	5	39	5	15	0·79	0·6	1·1	Overestimating	
Tart (piece)	106	43	6	33	2	16	0·84	0·6	1·0	Overestimating	***
Salad (lettuce)	105	26	27	1	47	0	0·67	−1·6	1·3	Not precise	*
Vegetable purée	108	31	35	5	28	1	0·74	−1·0	1·2	Not precise	
Porridge	108	35	25	11	28	1	0·78	−0·8	1·2	Not precise	***
Salad (mixed)	108	28	25	11	31	6	0·62	−0·8	1·4	Not precise	***
Cake (rectangular pieces)	105	30	26	18	15	10	0·67	−0·3	1·5	Not precise	***

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.
† Values obtained after density adjustment are indicated in parentheses.

and 16 %, and Spearman's correlation coefficients between the chosen picture number and the correct picture number were between 0·76 and 0·90. Of the twelve series, eight performed significantly differently in the three countries, but all means and SD were within the acceptable range, except three SD values exceeding 1·1 (Belgian meat cubes and oats/muesli and Czech fish) (country-specific data not shown).

Problematic picture series

Even after density adjustment, picture series for six food items still seemed to either underestimate portion size (overall mean < -0.45) – pasta (penne or fusilli), French fries and carrots (grated) – or overestimate portion size (overall mean > 0.45) – lasagne, tart and cake (square and flat) (data not shown). The underestimation was reflected in the higher percentage of participants selecting a picture smaller than the correct one (between 39 and 58 %) and the lower percentage selecting the correct picture (only 40 to 54 %) (Table 3). However, data at the country level were within the acceptable range for these series (mean > -0.7 and SD ≤ 1). Spearman's correlation coefficients between the chosen picture number and the correct picture number were 0·91–0·95, 0·84–0·85 and 0·89–0·94 for pasta (penne or fusilli), French fries and carrots, grated, respectively. The overall results for lasagne, tart and cake (square and flat) were borderline to overestimation (mean = 0·5, 0·6 and 0·6, respectively) and to 'not precise' (SD = 1·1, 1·0 and 1·1). Data obtained from the individual countries show that only in the Czech evaluation, the picture series of lasagne and cake seemed to perform well. The picture series of 'tart' seemed to perform well in Belgium, but not in the Czech Republic and Denmark.

Five series gave overall problematic results and were identified as 'not precise' since SD > 1.1: salad (lettuce), salad (mixed), vegetable purée, porridge and cake (rectangular pieces). For these series, only between 26 and 35 % of the participants selected the correct picture and 25 and 47 % selected a distant picture. Spearman's correlation coefficients between the chosen picture number and the correct picture number were between 0·62 and 0·78. After density adjustment, data obtained from the individual countries showed that only in the Danish evaluation the picture series of salad (lettuce), porridge and cake (rectangular pieces) performed well, and Spearman's correlation coefficients of these Danish data were between 0·78 and 0·89 (country-specific data not shown).

Difficulties with individual pictures within the series

Table 4 indicates that portion/picture number 6 was most often difficult to detect for the participants (defined as 17 % or more of the participants having chosen a distant picture, which is comparable to the overall results of the acceptable series where less than 16 % of the participants selected a distant picture), while for portion/picture number 1, a distant picture was chosen only in very few cases. From these observations, it can be derived that within the acceptable series, picture numbers 3 and 4 performed best and more pictures in the series categorised as 'overestimating' or 'not precise'

Table 4. Overview of the performance of the individual pictures of the series: pictures with >16% distant choices, differences between the estimated portion weight and the actual portion weight (g) and differences in grams related to the actual portion size (g %), and Spearman's correlation coefficients between the estimated weight and the actual weight

Dishes/foods	Picture number (actual)* (x indicates >16% distant choices)						Spearman's correlation coefficients	Series rating	Category increments between the pictures
	1	2	3	4	5	6			
Broccoli							0·91	Acceptable	i
Rice							0·91	Acceptable	m
Fruit purée							0·89	Acceptable	e
Meat sauce							0·92	Acceptable	m
Jam on toast bread							0·87	Acceptable	m
Casserole (dark)							0·92	Acceptable	m
Cornflakes						x	0·92	Acceptable	e_irreg
Cucumber						x	0·85	Acceptable	m
Casserole (light)						x	0·92	Acceptable	m
Potatoes (boiled)						x	0·83	Acceptable	e
Meatballs						x	0·92	Acceptable	e
Root vegetables (pieces)						x	0·83	Acceptable	e
Fried potatoes						x	0·87	Acceptable	e
Cheese (hard) on toast bread					x	x	0·89	Acceptable	m
Steak				x			0·82	Acceptable	e
Butter on toast bread		x					0·72	Acceptable	e
Mayonnaise filling on bread			x	x		x	0·83	Acceptable	m
Mashed potatoes			x	x		x	0·82	Acceptable	e
Cheese spread on toast bread			x	x			0·85	Acceptable	m
Butter block		x	x				0·88	Acceptable	i
Spaghetti						x	0·91	Acceptable after adjustment	m
Cauliflower						x	0·87	Acceptable after adjustment	m
Carrot (slices)						x	0·85	Acceptable after adjustment	m
Hamburger					x	x	0·86	Acceptable after adjustment	e
Courgette					x	x	0·88	Acceptable after adjustment	m
Maize					x	x	0·90	Acceptable after adjustment	m
Tomatoes (slices)					x	x	0·87	Acceptable after adjustment	e_irreg
Peas					x	x	0·90	Acceptable after adjustment	m
Oats or muesli					x	x	0·72	Acceptable after adjustment	e_irreg
Meat cubes/pieces		x		x	x	x	0·78	Acceptable after adjustment	e
Fish		x	x	x			0·83	Acceptable after adjustment	i
Cake, round and tall (piece)	x		x	x	x		0·87	Acceptable after adjustment	m
Pasta, penne or fusilli						x	0·93	Underestimating	m
Carrots (grated)						x	0·90	Underestimating	m
French fries					x	x	0·85	Underestimating	e
Cake, square and flat		x	x	x			0·79	Overestimating	e_irreg
Lasagne	x	x	x	x			0·79	Overestimating	e
Tart (piece)	x	x	x	x			0·85	Overestimating	i
Vegetable purée				x	x	x	0·76	Not valid	e
Salad (lettuce)			x	x	x	x	0·68	Not valid	e
Porridge			x	x	x	x	0·79	Not valid	e
Salad (mixed)			x	x	x	x	0·62	Not valid	e
Cake (rectangular pieces)		x	x	x	x	x	0·67	Not valid	e
Tomatoes, whole†							0·75	Others	
Chicken†							0·83	Others	

i, increasing distance; m, mixed (increasing between the smallest portions or between the smallest and medium portions, or differently for the smallest portions compared with the largest portions); e, equidistant; e_irreg, almost equidistant (varying distance).

* Chicken values are not shown per estimated portion number, since we assumed that the participants were able to detect which part of the chicken they had on the plate in front of them.

† The series of tomatoes and chicken were constructed differently from the others. Whole tomatoes of increasing sizes were depicted. Different parts of the chicken were depicted (1 = 145 g, 2 = 20 g, 3 = 40 g, 4 = 205 g, 5 = 70 g and 6 = 130 g).

were problematic, compared with the acceptable series. Table 4 also summarises the performance with regard to over- and underestimation by the calculation of the mean difference in grams between the estimated portion and the served portion and the difference in percentages of the served weight. These values varied a lot for the individual pictures (data not shown). Expressed in absolute amounts, the largest differences were found for the largest portion sizes, while the largest differences expressed as percentages were observed for the smallest portion sizes. The problematic series had at least one picture in the series where between 44 and 94 % had chosen a distant picture, while for the acceptable series, it varied between 6 and 67 % (data not shown).

Alternative constructed series: tomatoes (whole) and chicken

The series of whole tomatoes and chicken were constructed differently from the others. Whole tomatoes of six increasing sizes were depicted. Different parts of the chicken were depicted. For these series, estimation errors of the mean difference in grams between the estimated portion and the served portion and the difference in percentages of the served weight were only measured, showing a general overestimation of chicken, mean difference in g % = 46, and especially of the largest portion sizes, and overestimation of the smallest tomatoes and underestimation of the largest tomatoes. Spearman's correlations coefficients based on the estimated weight compared with the actual weight were 0·75 and 0·83 for tomatoes (whole) and chicken, respectively (Table 4).

Discussion

The results obtained for all the three countries showed that thirty-two of the forty-five food items could be estimated well using the thirty-eight food picture series, however, twelve of these foods only after adjustment for density differences. This indicated that density differences between the same foods in the three countries influenced the results, but that it was possible to distinguish between the portion sizes of each picture within a series. Between 37 and 71 % of the participants chose the correct picture, with a mean of 50 % including all 'acceptable' and 'acceptable after density adjustment' series. Between 1 and 16 % of the participants chose a distant picture. Lillegaard et al.[13] found that 60 % (number of estimates = 2019) selected the correct picture in a similar study with predefined portions in the series with four pictures (children and adolescents aged 9–19 years). In another study serving predefined portions among adults, 50 % chose the correct picture using series with five pictures each[14]. A study with pre-weighed portions among adolescents (series with nine pictures) found that only 28 % selected the correct picture and 74 % the correct or an adjacent photograph[15], while in an earlier study[16], 49 % selected the correct picture and 70 % the correct or adjacent picture; in both the studies, the participants were allowed to choose among seven portion sizes, three depicted and four virtual representing portions between the pictures and below the smallest picture and above the largest

picture. Compared with these results, most of the picture series of the present study performed well.

Three food items (French fries, penne or fusilli and carrots, grated) seem to be underestimated by the picture series having a mean bias < -0.4. However, since the mean and SD for the individual countries without adjustments were within the acceptable limits, these series were acceptable. The picture series of lasagne actually performed well in the Czech Republic but not in Belgium and Denmark, where they overestimated. Examination of the photos taken during the sessions showed that lasagne was distinctly overcooked in Belgium and Denmark (dark brown and dry), and exclusion of these results led to the conclusion that the series was acceptable, but a new validation session is suggested to confirm this.

With regard to cake, square and flat, examination of the photos and densities indicated that the cake from Czech Republic, which also performed well, was more similar to the cake depicted in the picture series regarding height and heaviness than the others. Correspondingly, similar results were obtained for the tart used in Belgium and the rectangular cake in Denmark. Therefore, it seems crucial that the cake used for the pictures is similar to the cake being estimated. Adding an object of a known size to the picture might also make it easier for the user to realise the size of the depicted cake.

It was remarkable that the series of porridge and vegetable purée were rated as 'not precise', whereas the fruit purée series was rated acceptable. These series were very similar, depicting the same portion sizes and using the same plate. Examination of the photos of the food items from the field-work showed that these picture series must be reserved for soft porridge and purées only, not for firm vegetable mash, as was served during the validation sessions.

Finally, the estimations of the portion sizes of green and mixed salads were problematic. Only the Danish green salad (lettuce) performed acceptably after adjustment. From the photos from the sessions, we know that none of the green salad servings appeared as depicted in the picture series. The mixed salad also appeared differently from country to country, although a common recipe for a mixed salad was used in all the countries. Portions with an amorphous shape-like salad have been described as problematic by others[17], but no suggestion for improvement has been made.

The three series identified as problematic, the cake (rectangular piece), the green salad and the mixed salad, also had lowest Spearman's correlation coefficients between the estimated mean portion number and the correct portion number of all, 0·67, 0·67 and 0·62, respectively (Table 3), while Spearman's correlation coefficients of the acceptable series typically were greater than 0·8. Other studies have found correlation coefficients within the same range[18,19].

The estimation errors of the alternative series of whole tomatoes showed values corresponding to those of the acceptable series, while the values from the series of chicken, in general, were overestimated. It is suggested that only the series of chicken be used in future surveys if the portion sizes reflect the most common size of chicken eaten by the target group of the survey.

Performance of individual pictures within the series

It appeared that the pictures with the largest portion sizes most often exhibited problematic performance. In the thirty-two acceptable series, this was the case of picture number 5 in nine series and picture number 6 in twenty series. Only in few of the acceptable series, picture numbers 3 and 4 turned out to be problematic. Other studies have found that smaller portions were more often correctly estimated than larger portions[14,15]. This might reflect a general hesitation to select the largest portions. However, in the present study, this may have occurred due to general difficulties with perception of the volume of the large portions, and this is probably more likely to happen when the users do not dish up themselves. The portion sizes were randomly assigned to the participants, and from the qualitative evaluation question-naires, we found that about one-third of the participants believed that the portions, in general, were too large compared with the portions that their children usually consumed.

When picture series are used in dietary surveys, participants having children consuming large portions might perform better when choosing among the largest portions. Further investigations are needed to confirm this.

Measures of estimation errors

We observed no clear association between the performance of the series and how the series were constructed with regard to the increments of the portion sizes (Table 4). A one-way ANOVA and least square mean differences Student's *t* tests for comparison between means and between SD of the difference between the estimated picture number minus the actual portion number (both non-adjusted and adjusted values) showed only significant differences between the means of series with equidistant and increasing increments. Corre-spondingly, only the difference between the SD of series with equidistant and mixed increments was significant. There was a tendency for the equidistant series to cause more difficulties while estimating the large portions, while the series with increasing increments were more problematic while estimating the smallest and medium-sized portions. This indicates that, in general, the mixed series performed best and that the performance of the series might benefit from taking equal increments on a log scale as a point of departure and modify from there.

The seven food items that were estimated by a picture series from another similar food item showed different results. For two foods, the series performed acceptably, three were accep-table when adjusted, one was underestimating and one was 'not precise'.

Density differences

The present study clearly showed that the difference in density of the food items depicted in the picture series and that of the food items on the plate has an impact on the bias of the esti-mates. However, adjusted values have to be used with caution. Measurement of the volume of the foods as presented on the

plate may seem simple but involves great uncertainties. Especially, measurement uncertainties will have a large impact on the density factor for foods with low densities such as (mixed) salad, grated carrots, cornflakes, oats and muesli, and some cakes. However, when carrying out a dietary survey, density adjustment can improve the quantification when densities of the foods depicted in the picture series are known from measurements and the densities of the most common foods eaten by the target group of the survey are known. For instance, it is possible to take density differences directly into account, when picture series are used for foods other than those depicted, but densities of food items eaten by the individuals are rarely known. Further development in the methods of measuring the volume of foods, such as use of image analysis, is needed before it is possible to benefit from adjustments at an individual level.

Strengths and limitations

The foods used in the present study were intended to appear as similar to the foods depicted in the picture series. In prac-tice, this was difficult as the study was conducted at three study centres. Therefore, it was a strength of the present study that the densities of the foods were measured and a photo of each food as dished on the plate was taken during the validation sessions. It was a tool to further investigate con-tradicting results and to avoid drawing wrong conclusions.

The study investigated the perception of the participants with regard to the portion sizes of the series in the sense of their ability to correctly link an amount of food presented on a plate to the correct picture in a series. This is most rel-evant when the picture series are going to be used in relation to a food diary. Conceptualisation and memory skills also influence the accuracy of portion size estimations in dietary surveys based on recall methods[20], but the present study was not designed to take this into account. Moreover, there are no existing guidelines regarding the level of acceptable accuracy for a picture to be used. Hernández *et al.*[21] pointed out that there is a lack of consensus of expression of error rates and that precision in portion size estimation is not yet a realistic expectation. It was, therefore, a challenge in the pre-sent study to define criteria for acceptability in the evaluation of the picture series. The design of the present study made it possible to determine whether each picture series performed adequately/sufficiently based on mean difference between the estimated portion number and served portion number and the SD.

An additional strength was that each picture of all picture series was evaluated six times each. This made it possible to identify problematic (or less valid) picture series and problematic pictures within the series. However, with the pre-weighed portions in the study, only relative tendency to overestimate and underestimate is indicated. The study design does not give insights into the absolute over- or underestimation of portions eaten by the target group. For the latter, a study should be organised among participants representative for the target group and preferably in real-life situations where they dish up their own realistic portions. Only after such a study is

conducted, inferences can be made about the extent of error on overall dietary intake in food consumption surveys.

It was a strength of the present study that it was conducted in three different countries in Europe. This makes it more acceptable to generalise the results to other European countries. However, the under-representation of parents with infants and over-representation of higher educated women in the present study might have influenced the results. Although Nelson et al.[22] have suggested sex and age as potential confounders, other studies have indicated that this might not be important for drawing conclusions[14,15,18,23,24]. The missing information about the participants in Belgium was due to the fact it was not registered during the first two validation sessions.

In the effort of providing comparable food consumption data and harmonising the methodology of data collection, it is regarded as an advantage to use a common tool such as the PANCAKE picture book, developed using science-based principles and practical experiences. Differences in dietary habits between the European countries do, however, necessitate variation as to which picture series are to be included in the picture book at the national level. Before using the series in another country, it is recommended that careful consideration be given as to whether the foods in the series cover the most commonly eaten foods and the relevant range of portion sizes in the new country. Otherwise, new country-specific series are needed. The PANCAKE picture book and the repository on the web, http://www.kostvaner. dk/pancake, offer the needed flexibility, since it is possible to choose between the picture series and add new series when needed. Since the picture book for the food record method is going to be used by the participants when they register their diet on their own, the number of picture series had to be restricted; we suggest a maximum of 45.

Conclusion

We recommend that all picture series in the PANCAKE picture book be included for the future EU Menu survey among infants, toddlers and children. In general, the performance of the picture series was best when the food item to be estimated was similar to the depicted food with regard to appearance and density. However, some specific series should be used with caution:

(1) The picture series of baby food – fruit purée, vegetable purée and porridge – are not suitable for firm and dry porridge and mashed fruits and vegetables, but suitable only for soft and smooth variants.
(2) The picture series of green salad and cakes are suitable only for food items that are very similar to the depicted foods.
(3) It is suggested that only the series of chicken be used in future surveys to identify the eaten part of chicken and the country-specific most common size of chicken eaten by the target group of the survey be used for the estimation of the amount eaten.

New and improved series should be developed or other quantification tools should be sought in order to cover the need for estimating portion sizes of various cakes and salads.

Furthermore, the performance of both large and small picture numbers in the accepted series might be further improved by defining the portions sizes for the pictures with increasing increments of the portion sizes.

For the development of new series, the procedure of the PANCAKE picture book is recommended. It is recommended that the range of portions be based on data from representative weighed data of the target group and increasing increments of the portions of the series be used. It is suggested that equal increments on a log scale be taken as a point of departure. The density of the food items within the series should be measured.

It is recommended that new pictures be validated using a study design and a method for analyses that are the same as those used in the present study. Additional validation is suggested for gaining insights into the over- and underestimation of the amounts as used in the target groups.

References

1. Biró G, Hulshof KFAM, Ovesen L, *et al.* for the EFCOSUM Group (2002) Selection of methodology to assess food intake. *Eur J Clin Nutr* **56**, Suppl. 2, S25–S32.
2. Brussaard JH, Löwik MRH, Steingrímsdóttir L, *et al.* (2002) A European food consumption survey method – conclusions and recommendations. *Eur J Clin Nutr* **56**, Suppl. 2, S89–S94.
3. de Boer EJ, Slimani N, van't Veer P, *et al.* on behalf of the EFCOVAL consortium (2011) The European Food Consumption Validation (EFCOVAL) Project: conclusions and recommendations. *Eur J Clin Nutr* **65**, Suppl. 1, 102–107.
4. Trolle E, Amiano P, Ege M, *et al.* on behalf of the EFCOVAL consortium (2011) Feasibility of 2 × 24-h recalls combined with a food-recording booklet, using EPIC-Soft, among schoolchildren. *Eur J Clin Nutr* **65**, Suppl. 1, S65–S76.
5. European Food Safety Authority (2011) Use of the EFSA comprehensive European food consumption database in exposure assessment. *EFSA J* **9**, 2097–2130.
6. Foster E, Matthews JNS, Nelson M, *et al.* (2006) Accuracy of estimates of food portion size using food photographs – the importance of using age-appropriate tools. *Public Health Nutr* **9**, 509–514.
7. Foster E, Adamson AJ, Anderson AS, *et al.* (2009) Estimation of portion size in children's dietary assessment: lesson learnt. *Eur J Clin Nutr* **63**, S45–S49.
8. Nelson M & Haraldsdóttir J (1998) Food photographs: practical guidelines II. Development and use of photographic atlases for assessing food portion size. *Public Health Nutr* **1**, 231–237.
9. Nelson M & Haraldsdóttir J (1998) Food photographs: practical guidelines 1. Design and analysis of studies to validate portion size estimates. *Public Health Nutr* **1**, 219–230.
10. Foster E, Matthews JNS, Lloyd J, *et al.* (2009) Children's estimates of food portion size: the development and evaluation of three portion size assessment tools for use with children. *Br J Nutr* **99**, 175–184.
11. Nielsen TH, Biltoft-Jensen A & Ygil KH (2011) *Development of Picture Series for the Danish National Survey of Diet and Physical Activity (in Danish)*. Soeborg: National Food Institute, Technical University of Denmark.
12. Nielsen TH, Biltoft-Jensen A, Christensen T, *et al.* (2011) *Study of Food and Beverage Portion Sizes Among Danes in the Age 4 to 75 Years (in Danish)*. Soeborg: National Food Institute, Technical University of Denmark.
13. Lillegaard ITL, Øverby NC & Andersen LF (2005) Can children and adolescents use photographs of food to estimate portion sizes? *Eur J Clin Nutr* **2005**, 611–617.
14. Ovaskainen M-L, Pautri 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.
15. Vereecken C, Dohogne S, Covents M, *et al.* (2010) How accurate are adolescents in portion-size estimation using the computer tool Young Adolescents' Nutrition Assessment on Computer (YANA-C)? *Br J Nutr* **103**, 1844–1850.
16. Lucas F, Niravong M, Villeminot S, *et al.* (1995) Estimation of food portion size using photographs: validity, strengths, weaknesses and recommendations. *J Hum Nutr Diet* **8**, 65–74.
17. 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.
18. De Keyzer W, Huybrechts I, De Maeyer M, *et al.* (2011) Food photographs in nutritional surveillance: errors in portion size estimation using drawings of bread and photographs of margarine and beverages consumption. *Br J Nutr* **105**, 1073–1083.
19. 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.
20. Nelson M, Atkinson M & Darbyshire S (1994) Food photography I: the perception of food portion size from photographs. *Br J Nutr* **72**, 649–663.
21. Hernández T, Wilder L, Kuehn D, *et al.* (2006) Portion size estimation and expectation of accuracy. *J Food Comp Anal* **19**, S14–S21.
22. Nelson M, Atkinson M & Darbyshire S (1996) Food photography II: use of food photographs for estimating portion size and the nutrient content of meals. *Br J Nutr* **76**, 31–49.
23. Robson PJ & Livingstone MBE (2000) An evaluation of food photographs as a tool for quantifying food and nutrient intake. *Public Health Nutr* **3**, 183–192.
24. Faggiano F, Vineis P, Cravanzola D, *et al.* (1992) Validation of a method for estimation of food portion size. *Epidemiology* **3**, 379–382.