International Journal of Obesity (2011) 35, S61–S68

0307-0565/11
www.nature.com/ijo

ORIGINAL ARTICLE

Reproducibility of food consumption frequencies derived from the Children's Eating Habits Questionnaire used in the IDEFICS study

A Lanfer[1], A Hebestreit[1], W Ahrens[1], V Krogh[2], S Sieri[2], L Lissner[3], G Eiben[3], A Siani[4], I Huybrechts[5], H-M Loit[6], S Papoutsou[7], É Kovács[8] and V Pala[2], on behalf of the IDEFICS Consortium

[1]*Department of Epidemiological Methods and Etiologic Research, Bremen Institute for Prevention Research and Social Medicine, University of Bremen, Bremen, Germany;* [2]*Department of Preventive and Predictive Medicine, Nutritional Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy;* [3]*Department of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden;* [4]*Unit of Epidemiology and Population Genetics, Institute of Food Sciences, National Research Council, Avellino, Italy;* [5]*Department of Public Health, Ghent University, Ghent, Belgium;* [6]*Department of Chronic Diseases, National Institute for Health Development, Tallinn, Estonia;* [7]*Research and Education Institute of Child Health, Strovolos, Cyprus and* [8]*Department of Paediatrics, University of Pécs, Pécs, Hungary*

Objective: To investigate the reproducibility of food consumption frequencies derived from the food frequency section of the Children's Eating Habits Questionnaire (CEHQ-FFQ) that was developed and used in the IDEFICS (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) project to assess food habits in 2- to 9-year-old European children.
Design and methods: From a subsample of 258 children who participated in the IDEFICS baseline examination, parental questionnaires of the CEHQ were collected twice to assess reproducibility of questionnaire results from 0 to 354 days after the first examination. Weighted Cohen's kappa coefficients (κ) and Spearman's correlation coefficients (r) were calculated to assess agreement between the first and second questionnaires for each food item of the CEHQ-FFQ. Stratification was performed for sex, age group, geographical region and length of period between the first and second administrations. Fisher's Z transformation was applied to test correlation coefficients for significant differences between strata.
Results: For all food items analysed, weighted Cohen's kappa coefficients (κ) and Spearman's correlation coefficients (r) were significant and positive ($P<0.001$). Reproducibility was lowest for diet soft drinks ($\kappa=0.23$, $r=0.32$) and highest for sweetened milk ($\kappa=0.68$, $r=0.76$). Correlation coefficients were comparable to those of previous studies on FFQ reproducibility in children and adults. Stratification did not reveal systematic differences in reproducibility by sex and age group. Spearman's correlation coefficients differed significantly between northern and southern European countries for 10 food items. In nine of them, the lower respective coefficient was still high enough to conclude acceptable reproducibility. As expected, longer time (>128 days) between the first and second administrations resulted in a generally lower, yet still acceptable, reproducibility.
Conclusion: Results indicate that the CEHQ-FFQ gives reproducible estimates of the consumption frequency of 43 food items from 14 food groups in European children.
International Journal of Obesity (2011) **35,** S61–S68; doi:10.1038/ijo.2011.36

Keywords: reproducibility; reliability; food frequency questionnaire; children; Europe; proxy

Introduction

Assessing diet is a central task in epidemiological studies that aim at investigating associations between dietary habits and health status. In adults, a variety of feasible and valid dietary assessment methods have been developed for research purposes.[1] Assessing diet in children is of increasing interest as more epidemiological studies are investigating childhood diet and its relation to the onset of diet-related diseases. However, to accomplish this relation on a European basis, there is a need for harmonised and standardised dietary assessment methods in children.[2] Therefore, it was one of the main issues of the diet module of the IDEFICS (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) study to develop and provide standardised instruments for assessing

Correspondence: Professor Dr W Ahrens, Department of Epidemiological Methods and Etiologic Research, Bremen Institute for Prevention Research and Social Medicine, University of Bremen, Achterstrasse 30, 28359 Bremen, Germany.
E-mail: ahrens@bips.uni-bremen.de

dietary habits and intake in European children.[3,4] Assessing diet in children from different cultural backgrounds holds special challenges. For instance, inter-country differences in commonly consumed foods present a challenge to researchers aiming to administer comparable food frequency questions in culturally different countries. However, comparability is the basis for successful analysis of dietary data from different countries. In addition, the limited cognitive capacity of children questions the application of conventional self-reporting instruments of dietary intake and makes it necessary to approach parents or other proxies as a substitute for the child.[5] Proxy interviews in turn bear the problem that children can consume a considerable amount of food without the proxy's presence; for instance, during school meals.

Against the background of these considerations a unique and new set of dietary assessment instruments was designed for the IDEFICS study, including three complementary methods. First, a parent or another proxy living with the child completed a Children's Eating Habits Questionnaire (CEHQ) that contained a qualitative food frequency section (CEHQ-FFQ). The latter recorded the consumption frequency of 43 pan-European food items that have the potential of affecting children's health status. Second, at least one 24-h dietary recall was collected for every child within the IDEFICS study. Third, the 24-h dietary recall was complemented with an observation of consumption during kindergarten or school meals. It was necessary to adapt these methods to survey languages and conditions in eight countries: Italy, Estonia, Cyprus, Belgium, Sweden, Germany, Hungary and Spain.

All instruments were evaluated for various aspects of validity and reproducibility during development and application. Here, we report the results of a reproducibility study that was conducted during the baseline examination in six of the eight survey countries to allow assessment of the test–retest reliability of the CEHQ-FFQ. Special attention is paid to possible modifying factors on reproducibility such as sex, age, geographical region and length of time period between the first and second administrations.[1,6–8]

Subjects and methods

The Children's Eating Habits Questionnaire—food frequency section

The CEHQ is a screening instrument that aims at investigating food consumption frequency and behaviours associated with overweight, obesity and general health in children. The CEHQ includes a food frequency section (CEHQ-FFQ) in which parents or another proxy living with the child reports the frequency of their child's consumption of selected food items in a typical week during the preceding 4 weeks, outside the school canteen or childcare meal provision settings. Assessing longer intervals such as several months or years can usually provide better assessments of habitual intake. Nevertheless, we used 4 weeks as reference period as a limited time span was likely to be more accurate when assessing growing children's diets. A long reference period would not be able to detect alterations in diet because of development. In addition, the opportunity to choose the more typical week among the previous four and not just the previous one allowed us to reduce the probability of a 'special week', for instance, due to holiday or illness. Consequently the week-by-week variability was reduced. The CEHQ-FFQ was complemented by nine sections assessing family eating contexts, parental attitudes and behaviours regarding food and nutrition, special eating habits, for example, vegetarianism, and a question on the number of meals per week consumed at home.

In the CEHQ-FFQ, the same general description of each food item was used in each country to guarantee standardisation among all eight surveyed countries. However, country-specific food examples for the given food items were used in such a way that the proxies would consistently classify foods to the same food group among all countries. To facilitate filling out the questionnaire, the same response scale was used for all CEHQ-FFQ dietary items. Response options displayed from left to right were as follows: 'Never/less than once a week', '1–3 times a week', '4–6 times a week', '1 time per day', '2 times per day', '3 times per day', '4 or more times per day' and 'I have no idea'. This scale was adopted from the proxy eating habits questionnaire of the Early Childhood Longitudinal Survey of the United States Department of Agriculture.[9] Frequencies of intake were assessed without attempting to quantify portion sizes.

The CEHQ was subject to pretesting before the IDEFICS baseline examination in all eight countries of investigation.[10] The final CEHQ-FFQ asked for the consumption frequency of 43 pan-European food items of 14 food groups.[11] It was used in the whole sample of the IDEFICS baseline examination from September 2007 to May 2008 and was completed for more than 15 000 children between 2 and 9 years of age from the eight survey centres.

Test on reproducibility

A CEHQ reproducibility study was conducted as a substudy during the IDEFICS baseline examination period from September 2007 to May 2008. Participation in the reproducibility study was voluntary for the survey centres. Six of the eight IDEFICS survey centres agreed to participate (Italy, Estonia, Cyprus, Belgium, Sweden and Hungary). It was the minimum requirement for each participating centre to continuously collect a convenience sample of children who participated in the IDEFICS baseline survey during the baseline examination period and up to 2 months after its end. To avoid an intervention effect, all questionnaires were collected before the intervention activities of the IDEFICS study started in the fall of 2008. The parents or, when necessary, other proxies living with these children were asked to fill in a second CEHQ and a second IDEFICS

parental questionnaire that included information on age and sex of the child and of the proxy who completed the questionnaire. There was no fixed time period between the first and second administration to enable the assessment of reproducibility for a range of time intervals. For inclusion in the analysis, both questionnaires had to be filled in for the same child. This was achieved by comparing the questions about sex and date of birth of the child between the first and second questionnaires. For both questionnaire administrations, the parents or guardians in general and not a specific proxy were addressed to fill in the questionnaires. Therefore, it was possible that the first and second questionnaires were answered by different persons. This complies with the study design of the IDEFICS surveys in which reporting by different proxies during baseline and follow-up surveys is possible. Therefore, these cases were also not excluded from the reproducibility analyses to enable assessment of reproducibility under realistic conditions.

The convenience sample collected by the six centres consisted of 276 families. A total of 18 questionnaires were excluded from the final data set, as reported sex or date of birth of the child differed between the first and second questionnaires. The final reproducibility subsample that was used in the analyses comprised the first and second questionnaires of 258 children.

Statistics

On the basis of the answers to the CEHQ-FFQ, consumption frequencies per week were calculated for each food item. When the proxy reported having 'no idea' for a given food in the first or repeat questionnaire, consumption frequency could not be calculated and the data were not used in the analysis of the respective food item.

Consumption frequencies of the first questionnaires of the reproducibility subsample were compared with those of the total sample of the baseline examinations of those countries that assessed reproducibility by applying Mann–Whitney *U*-tests. This was performed to check for possible differences between the samples due to selection bias as the reproducibility study was based on a convenience sample. For reproducibility analyses, Spearman's rank correlation coefficients and weighted Cohen's kappa coefficients were calculated for each food item in the reproducibility subsample to determine the degree of agreement between the first and second reporting. The Cicchetti–Allison weighting scheme was used to calculate the weighted kappa coefficients.[12] In addition, the Wilcoxon signed-rank test was applied for each food item to detect general differences in consumption frequencies between the first and second administrations.

All coefficients were also calculated after stratification by age group (2–5 years vs 6–9 years), sex of the child, southern vs northern European countries (Belgium, Sweden, Estonia vs Hungary, Italy and Cyprus) and by time period between the first and second administration (equal or less than the median period of 128 days vs more than the median period). Fisher's Z transformation was used to test for differences of Spearman's coefficients between strata using the SAS COMPCORR Macro.[13] All analyses were conducted using SAS version 8.02 (SAS Institute, Cary, NC, USA).

Ethical approval

In each country, the participating centres obtained the ethical approval of their responsible authority. All children and their parents provided oral and/or written informed consent for all examinations and/or for the collection of samples, subsequent analysis and storage of personal data and collected samples.

Results

Basic characteristics of the reproducibility subsample are displayed in Table 1. In the sample, both sexes were almost equally represented, which is comparable to the sex distribution in the full IDEFICS baseline examination sample.[14] However, there were differences between countries with respect to the proportion of old and young children. Age was more equally distributed between countries in the IDEFICS full sample.[14] This was partly due to the selection of the reproducibility subsample as a convenience sample.

Table 1 Basic characteristics of the reproducibility subsample

	N (*full IDEFICS sample*[a])	N (*subsample*)	*Number of days between first and second CEHQ-FFQ*		*Sex*				*Age (in years)*			
			Mean (± s.d.)	*Range*	*Boys*		*Girls*		*2–<6*		*6–<10*	
					N	%	N	%	N	%	N	%
Italy	2250	55	12 (±37)	0–281	28	50.9	27	49.1	10	18.2	45	81.8
Estonia	1719	40	133 (±26)	56–179	18	45.0	22	55.0	35	87.5	5	12.5
Cyprus	2380	22	157 (±26)	145–266	8	36.4	14	63.6	2	9.1	20	90.9
Belgium	1926	38	167 (±65)	105–354	16	42.1	22	57.9	7	18.4	31	81.6
Sweden	1809	53	135 (±55)	10–244	29	54.7	24	45.3	30	56.6	23	43.4
Hungary	2567	50	150 (±44)	0–236	14	28.0	36	72.0	18	36.0	32	64.0
Total	12 651	258	117 (±73)	0–354	113	43.8	145	56.2	102	39.5	156	60.5

Abbreviations: CEHQ-FFQ, Children's Eating Habits Questionnaire-food frequency section; IDEFICS, Identification and prevention of dietary- and lifestyle-induced health effects in children and infants. [a]Full IDEFICS sample as described in ref. 14 without German and Spanish sample.

Table 2 Reproducibility of consumption frequency per week between first (FFQ1) and second administration (FFQ2) of the CEHQ-FFQ

Food item	N[a]	*Mean consumption frequency FFQ1*	*Mean consumption frequency FFQ2*	*Cohen's weighted kappa coefficient*[c]	*Spearman's correlation coefficient*[c]	*Rank by weighted kappa coefficient*[d]	*Rank by Spearman's correlation coefficient*[d]
Vegetable, cooked	251	4.1	4.2	0.49	0.59	22	23
Potatoes, fried	245	0.9	1.2[b]	0.33	0.40	41	42
Vegetable, raw	215	5.2	5.5	0.46	0.72	26	6
Fruit, fresh, without sugar	254	8.0	8.0	0.45	0.58	27	25
Fruit, fresh, sugar added	220	1.9	1.8	0.28	0.47	42	37
Water	240	22.4	22.5	0.49	0.59	23	22
Fruit juice	249	5.6	5.4	0.54	0.73	12	3
Soft drink, sugar added	253	2.2	2.0	0.54	0.53	13	32
Soft drink, diet	242	0.4	0.3	0.23	0.32	43	43
Breakfast cereals, sugar added	248	2.3	2.1	0.55	0.70	7	9
Breakfast cereals, no sugar	235	1.9	2.1	0.57	0.71	5	7
Milk, no sugar	240	8.2	7.8	0.60	0.74	3	2
Milk, sugar added	237	2.9	2.8	0.68	0.76	1	1
Yoghurt, no sugar	224	1.5	1.5	0.50	0.60	20	19
Yoghurt, sugar added	252	2.6	2.7	0.51	0.63	17	13
Fish, not fried	239	1.1	1.1	0.45	0.49	28	36
Fish, fried	244	1.0	1.0	0.38	0.41	40	41
Cold cuts, sausage	247	3.4	3.1	0.44	0.53	33	34
Meat, not fried	246	2.3	2.2	0.42	0.54	35	30
Meat, fried	242	1.9	2.0	0.54	0.63	11	15
Egg, fried	247	0.7	0.8	0.40	0.46	36	38
Egg, boiled	245	0.8	0.8	0.39	0.45	38	39
Mayonnaise	247	0.6	0.6	0.54	0.60	10	20
Meat replacement products	242	0.2	0.2	0.44	0.60	31	21
Cheese	220	5.1	5.2	0.50	0.69	19	10
Honey, jam	250	1.5	1.5	0.51	0.61	18	17
Chocolate-, nut-based spread	253	1.7	1.6	0.62	0.72	2	5
Butter, margarine on bread	244	3.7	3.5	0.51	0.70	16	8
Butter, margarine on bread, low fat	237	3.0	2.7	0.55	0.66	9	12
Ketchup	249	1.4	1.4	0.56	0.63	6	14
Bread, white	254	6.6	5.8[b]	0.49	0.69	21	11
Bread, wholemeal	243	3.6	3.7	0.43	0.62	34	16
Pasta, rice	251	3.2	3.1	0.55	0.57	8	28
Cereals, milled	240	0.6	0.9	0.48	0.53	24	33
Pizza, main dish	245	0.4	0.4	0.52	0.54	15	29
Hamburger, hot dog, falafel	252	2.0	2.0	0.60	0.72	4	4
Nuts, seeds, dried fruit	248	1.1	0.9	0.45	0.57	29	26
Salty snacks	253	0.8	0.9	0.39	0.41	39	40
Savoury pastries	248	0.7	0.8	0.47	0.54	25	31
Chocolate	254	1.7	1.6	0.44	0.57	32	27
Candy, non-chocolate	244	1.8	1.5[b]	0.52	0.61	14	18
Cake, pudding, cookies	253	2.4	2.3	0.45	0.59	30	24
Ice cream	255	1.5	1.9[b]	0.40	0.51	37	35

Abbreviation: CEHQ-FFQ, Children's Eating Habits Questionnaire-food frequency section. [a]Cases were excluded from analyses for the respective food group when 'I have no idea' or nothing was reported in first or second administration. [b]Significant difference in consumption frequency at $P<0.05$ between first and second administration assessed by Wilcoxon signed-rank test. [c]All coefficients were significantly different from zero at $P<0.0001$. [d]1 = highest coefficient, 43 = lowest coefficient.

As there was no fixed time between questionnaire administrations, a high variation in time between the first and second administrations of the questionnaire was achieved that ranged between zero days to almost 1 year later. On average, the time span between the first and second administration was almost 4 months. Except for Italy, where the average time between the first and second administration was only 10 days, the average time span was comparable between countries. The vast majority of questionnaires were completed by a female proxy of the child both times (87%), most likely the mother.

Comparison of consumption frequencies between the full sample and the reproducibility subsample revealed significantly lower consumption frequencies in the subsample for 'sugar added fresh fruit', 'fruit juice', 'chocolate-, nut-based spread', 'salty snacks' and 'chocolate' and a higher frequency for 'water'.

Results of the reproducibility analysis are summarised in Table 2. Wilcoxon signed-rank tests revealed significant differences in mean consumption frequency between first and second administrations for 'fried potatoes', 'white bread', 'non-chocolate candies' and 'ice cream' only. Reproducibility coefficients ranged from 0.23 to 0.68 for weighted kappa coefficients and from 0.32 to 0.76 for Spearman's correlation coefficients. Reproducibility was lowest for 'diet soft drinks' and highest for 'sweetened milk'. All weighted kappa and correlation coefficients were significantly different from zero

($P<0.0001$). Reproducibility was related to consumption frequency. For those 21 food items with the lowest mean consumption frequencies, mean Cohen's kappa and Spearman's correlation coefficients were also lower ($\kappa=0.44$ and $r=0.53$) than for the food items with higher mean consumption frequencies ($\kappa=0.52$ and $r=0.65$). We did not observe a general tendency in seasonal foods such as vegetables, fruit, drinks, nuts and dried fruits and ice cream for lower reproducibility compared with non-seasonal foods such as bread (Table 2).

In general, stratified analyses by sex and age did not reveal systematic differences in reproducibility between strata (data not shown). Spearman's correlation coefficients were significantly higher for boys than for girls for 'boiled eggs' ($r=0.58$ vs $r=0.35$) and 'chocolate- and nut-based spread' ($r=0.84$ vs $r=0.63$) and significantly lower for 'meat replacement products' ($r=0.43$ vs $r=0.69$), 'ketchup' ($r=0.47$ vs $r=0.76$), 'pasta, rice' ($r=0.46$ vs $r=0.64$), 'pizza as main dish' ($r=0.44$ vs $r=0.64$) and 'hamburger, hot dogs and falafel' ($r=0.65$ vs $r=0.78$). Spearman's coefficients were significantly higher for older children than for younger ones for 'sweetened yoghurt' ($r=0.76$ vs $r=0.41$), 'cheese' ($r=0.75$ vs $r=0.57$) and 'chocolate- and nut-based spread' ($r=0.76$ vs $r=0.49$) and lower for 'meat replacement products' ($r=0.53$ vs $r=0.74$) and 'unsweetened breakfast cereals' ($r=0.54$ vs $r=0.76$).

Results for the stratified analyses by region and time between interviews are shown in Table 3. Spearman's coefficients significantly differed by geographical region for 10 food items. However, there was no trend of generally lower or higher coefficients in any of the two groups. As expected, Spearman's coefficients tended to be higher when time between the two administrations was short (Table 3).

Discussion

Together with validity, reproducibility is an important indicator for the performance of a food frequency questionnaire (FFQ). We assessed reproducibility of the CEHQ-FFQ by administering the CEHQ two times to a subsample of IDEFICS study participants during the IDEFICS baseline survey and subsequently tested the derived consumption frequencies for agreement. The results show that the CEHQ-FFQ developed for a multicultural setting was reproducible for a convenience subsample of IDEFICS participants for all food items.

Although several studies on reproducibility of FFQs have been conducted in adults, there exists no definition of what constitutes a 'good' and 'bad' reproducibility. For kappa statistics, Landis and Koch[15] have suggested to define Kappa coefficients of 0.21–0.40 as 'fair' agreement, 0.41–0.60 as 'moderate', 0.61–0.80 as 'substantial' and 0.81–1.00 as 'almost perfect'. According to this classification, reproducibility was 'fair' for eight food items, 'moderate' for 32 food items and 'substantial' for three food items. However, this categorisation was developed to evaluate inter-observer agreement and might not be equally applicable to FFQ reproducibility.

When assessing the reproducibility of nutrient intakes derived from FFQs in adults, correlation coefficients appear to typically range between 0.5 and 0.7.[1,16] In our study, Spearman's correlation coefficients were higher than 0.5 for 35 (81%) food items. Given the fact that correlation coefficients of food intakes and consumption frequencies are usually more variable than those of nutrient intakes,[1] this indicates that reproducibility of consumption frequencies derived from the CEHQ-FFQ is comparable to that of FFQs used in adults. Previous studies from the United States of America, New Zealand and Belgium that have reported reproducibility of food intakes derived from FFQs in children show a wide range of correlation coefficient values ranging from 0.24 to 0.82.[17–20] Design of the FFQs, statistical methods, data management and the time between first and second administrations vary in these studies, making it difficult to directly compare results. Nevertheless, they give some indication of the typical range of correlation coefficients. The Spearman's correlation coefficients of our study all lay within that range (Table 2).

In our study, Spearman's correlation coefficients were generally higher than Cohen's kappa coefficients. Both statistics reflect different aspects of reliability. Correlation coefficients depict the degree to which the two administrations are generally related.[21] Cohen's kappa statistics measure agreement between administrations, discounting the proportion of agreement that is expected by chance alone.[22] Therefore, the values of both coefficient types cannot be directly compared. However, both measures generally resulted in similar ranking of food items by reproducibility (Table 2).

Dietary reproducibility reflects both reporting errors and variability in diet. Therefore, foods that do not contribute to a regular eating pattern should be more prone to low reproducibility than those foods that are frequently consumed. In line with this, kappa and Spearman's coefficients in our study tended to be higher for frequently consumed foods, which could also be seen in other FFQ reproducibility studies.[17]

Our results did not show systematic differences in reproducibility by sex of the child. There have been reports on sex/gender differences in reproducibility of FFQs in adults and self-reporting children and adolescents; however, in general, results are not conclusive.[7,8,17,18,23] One possible explanation of our results might be the fact that we used proxy reports on diet. When parents and other proxies report on the children's intakes, it is possible that the sex of the proxy is more important than the sex of the child. Unfortunately, the number of men filling out the FFQ for their child was too small in our sample to investigate possible differences in reporting due to the sex of the proxy and these assumptions could thus not be investigated.

Stratification by age group revealed no consistent age trend in reproducibility. This suggests that reproducibility of reports by proxies on child consumption frequency is also

Table 3 Spearman's correlation coefficients comparing consumption frequency per week between first (FFQ1) and second administration (FFQ2) stratified by length of time between FFQ1 and FFQ2 and geographical region

Food item	*Time period*			*Region*		
	Short[a]	*Long*[a]	P-*value*[b]	*North*[a]	*South*[a]	P-*value*[b]
Vegetable, cooked	0.65	0.52	0.139	0.55	0.49	0.517
Potatoes, fried	0.51	0.28	0.032	0.36	0.48	0.274
Vegetable, raw	0.80	0.60	0.005	0.61	0.72	0.160
Fruit, fresh, without sugar	0.66	0.48	0.038	0.62	0.53	0.269
Fruit, fresh, sugar added	0.52	0.43	0.393	0.40	0.53	0.232
Water	0.66	0.51	0.077	0.58	0.51	0.434
Fruit juice	0.73	0.73	0.919	0.78	0.69	0.127
Soft drink, sugar added	0.67	0.35	<0.001	0.51	0.55	0.660
Soft drink, diet	0.31	0.31	0.973	0.34	0.27	0.533
Breakfast cereals, sugar added	0.71	0.67	0.618	0.65	0.73	0.261
Breakfast cereals, no sugar	0.76	0.67	0.168	0.76	0.42	<0.001
Milk, no sugar	0.77	0.70	0.207	0.73	0.69	0.579
Milk, sugar added	0.87	0.63	<0.001	0.60	0.75	0.033
Yoghurt, no sugar	0.72	0.50	0.008	0.61	0.61	0.982
Yoghurt, sugar added	0.71	0.55	0.038	0.56	0.67	0.152
Fish, not fried	0.65	0.26	<0.001	0.38	0.59	0.036
Fish, fried	0.42	0.36	0.624	0.44	0.38	0.538
Cold cuts, sausage	0.57	0.43	0.145	0.43	0.63	0.026
Meat, not fried	0.67	0.39	0.002	0.46	0.55	0.327
Meat, fried	0.69	0.56	0.092	0.65	0.49	0.058
Egg, fried	0.59	0.32	0.008	0.30	0.49	0.087
Egg, boiled	0.48	0.40	0.427	0.46	0.42	0.657
Mayonnaise	0.65	0.53	0.131	0.62	0.58	0.607
Meat replacement products	0.71	0.49	0.005	0.53	0.77	<0.001
Cheese	0.67	0.71	0.557	0.65	0.74	0.237
Honey, jam	0.65	0.57	0.307	0.63	0.60	0.698
Chocolate-, nut-based spread	0.82	0.57	<0.001	0.77	0.65	0.043
Butter, margarine on bread	0.79	0.51	<0.001	0.64	0.77	0.046
Butter, margarine on bread, low fat	0.81	0.48	<0.001	0.72	0.57	0.045
Ketchup	0.76	0.48	<0.001	0.62	0.62	0.994
Bread, white	0.70	0.67	0.618	0.68	0.62	0.430
Bread, wholemeal	0.67	0.53	0.085	0.52	0.50	0.844
Pasta, rice	0.64	0.45	0.035	0.56	0.56	0.958
Cereals, milled	0.56	0.49	0.445	0.47	0.58	0.234
Pizza, main dish	0.56	0.45	0.247	0.11	0.59	<0.001
Hamburger, hot dog, falafel	0.72	0.70	0.776	0.70	0.72	0.799
Nuts, seeds, dried fruit	0.65	0.47	0.036	0.57	0.57	0.993
Salty snacks	0.47	0.35	0.244	0.40	0.41	0.949
Savoury pastries	0.54	0.54	0.980	0.45	0.52	0.508
Chocolate	0.55	0.57	0.862	0.45	0.61	0.094
Candy, non-chocolate	0.67	0.53	0.100	0.48	0.74	0.001
Cake, pudding, cookies	0.71	0.43	<0.001	0.48	0.62	0.109
Ice cream	0.52	0.49	0.763	0.45	0.56	0.239

[a]Short, ⩽128 days; Long, >128 days. North: Sweden, Estonia, Belgium; South: Hungary, Italy, Cyprus. [b]*P*-values from Fisher's *Z* transformation testing for differences in correlation coefficients by strata.

not systematically influenced by the age of the child. In a previous study, differences in reproducibility of food item intakes were detected between 1- to 4- and 5- to 9-year-old children.[18] However, according to the authors, differences in reproducibility were likely, as for the younger age group parents completed the questionnaire, whereas for the older age group most of the children had filled out the questionnaire themselves with the assistance of an adult. In line with our results, a previous study that assessed reproducibility of food intakes derived from an FFQ in 9- to 18-year-old adolescents observed no consistent pattern in reproducibility according to age group.[17]

Because of the study design, a large variability in time period between first and second administrations could be achieved and reproducibility for different time spans could be investigated. Correlation coefficients were lower when the period between the first and second administrations was more than 128 days (~4 months). This was to be expected as—next to reporting errors and general variability in diet—dietary reproducibility is also influenced by real changes in diet, and the probability of variability and real changes in diet increases with time. In addition, changes in season between two questionnaire administrations can further increase variability in diet.

Although it may be argued that it is a disadvantage of dietary reproducibility studies that they fail to separate reporting errors from real changes in diet, this may be practically of minor importance. Both sources of variation may affect the results of studies that aim at investigating the relationship between diet and health outcome as they both contribute to a misclassification of long-term dietary intakes.[1] Therefore, low reproducibility coefficients are indicators of misclassification regardless of whether they are due to reporting error or dietary changes. In our study, long-term Spearman's correlation coefficients ranged between 0.26 and 0.73 (Table 3). As these values lie within the range of correlation coefficient values of previous studies with $r=0.24$ up to $r=0.82$,[17–20] we concluded that long-term reproducibility was still acceptable. We did not observe generally lower reproducibility in seasonal foods than in non-seasonal foods. This either implies that seasonal foods are replaced by other foods of the same category in changing season, for instance strawberries by oranges, or reflects the fact that seasonal foods can be purchased in all seasons given the modern food supply systems.

Unfortunately, separate analyses for each country were not feasible because of too small sample sizes. Stratification was therefore carried out by broader geographical region. Significant differences mainly occurred in food items with relatively high correlation coefficients in both strata (Table 3). For instance, the correlation coefficient for 'unsweetened breakfast cereals' was significantly lower for southern European countries. Nevertheless, it had a value of $r=0.42$, which implies acceptable reproducibility compared with previous studies.[17–20] Only for 'Pizza as main dish' was the correlation coefficient for northern European countries significantly different from southern European countries and below the minimum value of 0.24 of previous studies.

One limitation of our study lies in the fact that the proxies selected for second administrations were collected as a convenience sample, which may not be representative for the whole IDEFICS study population. For reproducibility studies, convenience sampling may introduce bias, as individuals that are willing to fill in a second questionnaire may also find the study more important and give more effort to report accurately. This could lead to an overestimation of reproducibility.

Comparison of the age and sex distribution between the subsample and the full IDEFICS examination population revealed some differences, especially with regard to the age distribution. Theoretically, this might have had an effect on reproducibility. However, in our study, reproducibility was not systematically influenced by sex and age of the child.

To give some indication of diet-related differences between the reproducibility subsample and the full IDEFICS sample, we compared the consumption frequencies of the subsample and the full sample. There was a general tendency in the reproducibility subsample to report lower consumption frequencies for the less favourable foods with regard to achievement of a healthy diet and higher frequencies for foods that describe healthy food choices. However, significant differences in consumption frequencies between the reproducibility subsample and the full sample were only detected for 6 of 43 food items (Table 2). Against this background, it seems as if the reproducibility subsample either followed a slightly more favourable diet or was more prone to yield socially desirable answers. Although social desirability bias or a more healthy diet does not necessarily influence reproducibility, it should be kept in mind that because of the convenience sampling of the study with no information on response rates there may be a bias in the data towards higher reproducibility estimates.

Nevertheless, to our knowledge, this is the first study to assess reproducibility of a pan-European FFQ in children from different cultural and national backgrounds. Therefore, it contributes to the development of dietary instruments that can be applied in different countries and facilitates comparability of results. Furthermore, a variety of possible influencing factors on reproducibility could be investigated.

In summary, reproducibility of the CEHQ-FFQ was similar to other FFQs in children, adolescents and adults, despite the multinational application of the questionnaire. Reproducibility was acceptable in both northern and southern European countries, although differences in reproducibility could be detected by geographical region. No systematic differences in reproducibility were detected by sex and age and reproducibility was still acceptable in a subsample of questionnaires with more than 4 months between the first and second administration of the CEHQ-FFQ.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This study was conducted as part of the IDEFICS study (http://www.idefics.eu). We gratefully acknowledge the financial support of the European Community within the Sixth RTD Framework Programme Contract No. 016181 (FOOD) and the grant support from the EU for the IDEFICS study.

The information in this document reflects the author's view and is provided as it is.

Statement of ethics

We certify that all applicable institutional and governmental regulations regarding the ethical use of human volunteers were followed during this research. Approval by the appropriate ethics committees was obtained by each of the eight centres engaged in the fieldwork. Participants were not subjected to any study procedure before both the children and their parents gave consent for examinations, collection of samples,

subsequent analysis and storage of personal data and collected samples. The participating children and their parents could consent to single components of the study while refraining from others.

References

1 Willett WC. *Nutritional Epidemiology, Monographs in Epidemiology and Biostatistics*. 2nd edn. Oxford University Press: New York, 1998.
2 Lambert J, Agostoni C, Elmadfa I, Hulshof K, Krause E, Livingstone B *et al*. Dietary intake and nutritional status of children and adolescents in Europe. *Br J Nutr* 2004; **92/2**: S147–S211.
3 Bammann K, Peplies J, Sjöström M, Lissner L, Henauw S, Iacoviello L *et al*. Assessment of diet, physical activity and biological, social and environmental factors in a multi-centre European project on diet- and lifestyle-related disorders in children (IDEFICS). *J Public Health* 2006; **14/5**: 279–289.
4 Ahrens W, Bammann K, de Henauw S, Halford J, Palou A, Pigeot I *et al*. Understanding and preventing childhood oesity and related disorders—IDEFICS: European multilevel epidemiological approach. *Nutr Metab Cardiovas* 2006; **16/4**: 302–308.
5 Treiber FA, Leonard SB, Frank G, Musante L, Davis H, Strong WB *et al*. Dietary assessment instruments for preschool children: reliability of parental responses to the 24-h recall and a food frequency questionnaire. *J Am Diet Assoc* 1990; **90/6**: 814–820.
6 Northstone K, Emmet PM. Are dietary patterns stable throughout early mid-childhood? A birth cohort study. *Br J Nutr* 2008; **100/5**: 1069–1076.
7 McPherson R, Hoelscher DM, Alexander M, Scanlon KS, Serdula MK. Dietary assessment methods among school-aged children: validity and reliability. *Prev Med* 2000; **31**: 11–33.
8 Stevens J, Metcalf PA, Dennis BA, Shimakawa T, Tell GS, Folsom AR. Reliability of a food frequency questionnaire by ethnicity, gender, age and education. *Nutr Res* 1996; **16/5**: 735–745.
9 USDA/ORC Macro. *Developing Effective Wording and Format Options for a Children's Nutrition Behavior Questionnaire for Mothers of Children in Kindergarten*. Report no 10. USDA: Washington, 2005.
10 Suling M, Hebestreit A, Peplies J, Bammann K, Nappo A, Eiben G *et al*. Design and results of the pretest of the IDEFICS study. *Int J Obes* 2011; **35** (Suppl 1): S30–S44.
11 Huybrechts I, Börnhorst C, Pala V, Moreno LA, Barba G, Lissner L *et al*. Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children. *Int J Obes* 2011; **35** (Suppl 1): S69–S78.
12 Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *Am J EEG Technol* 1971; **11**: 101–109.
13 SAS Knowledge Base. Compare correlations using Fisher's Z transformation. http://support.sas.com/kb/24/995 html 2010.
14 Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, Iacoviello L *et al*. The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes* 2011; **35** (Suppl 1): S3–S15.
15 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33/1**: 159–174.
16 Cade JE, Burley VJ, Warm DL, Thompson RL, Margetts BM. Food-frequency questionnaires: a review of their design, validation and utilisation. *Nutr Res Rev* 2004; **17/1**: 5–22.
17 Rockett HR, Wolf AM, Colditz GA. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. *J Am Diet Assoc* 1995; **95/3**: 336–340.
18 Metcalf PA, Scragg RK, Sharpe S, Fitzgerald ED, Schaaf D, Watts C. Short-term repeatability of a food frequency questionnaire in New Zealand children aged 1–14 y. *Eur J Clin Nutr* 2003; **57/11**: 1498–1503.
19 Huybrechts I, De Baker G, De Bacquer D, Maes L, De Henauw S. Relative validity and reproducibility of a food-frequency questionnaire for estimating food intakes among Flemish preschoolers. *Int J Environ Res Public Health* 2009; **6/1**: 382–399.
20 Buzzard IM, Stanton CA, Figueiredo M, Fries EA, Nicholson R, Hogan CJ *et al*. Development and reproducibility of a brief food frequency questionnaire for assessing the fat, fiber, and fruit and vegetable intakes of rural adolescents. *J Am Diet Assoc* 2001; **101/12**: 1438–1446.
21 Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires—a review. *Public Health Nutr* 2002; **5/4**: 567–587.
22 Fleiss JL. *Statistical Methods for Rates and Proportions*. John Wiley & Sons: Chichester, 1981.
23 Ibiebele TI, Parekh S, Mallitt KA, Hughes MC, O'Rourke PK, Webb PM. Reproducibility of food and nutrient intake estimates using a semi-quantitative FFQ in Australian adults. *Public Health Nutr* 2009; **12/12**: 2359–2365.