A meta-analysis of the validity of FFQ targeted to adolescents

Garden Tabacchi[1,*], Anna Rita Filippi[1], Emanuele Amodio[1], Monèm Jemni[2], Antonino Bianco[3], Alberto Firenze[1] and Caterina Mammina[1]

[1]Department of Sciences for Health Promotion and Mother Child Care 'G. D'Alessandro', University of Palermo, Via Del Vespro 133, 90127 Palermo, Italy: [2]School of Science, University of Greenwich at Medway, Chatham Maritime, Kent, UK: [3]Sport and Exercise Sciences Unit, University of Palermo, Palermo, Italy

Abstract

Objective: The present work is aimed at meta-analysing validity studies of FFQ for adolescents, to investigate their overall accuracy and variables that can affect it negatively.

Design: A meta-analysis of sixteen original articles was performed within the ASSO Project (Adolescents and Surveillance System in the Obesity prevention).

Setting: The articles assessed the validity of FFQ for adolescents, compared with food records or 24 h recalls, with regard to energy and nutrient intakes.

Subjects: Pearson's or Spearman's correlation coefficients, means/standard deviations, kappa agreement, percentiles and mean differences/limits of agreement (Bland–Altman method) were extracted. Pooled estimates were calculated and heterogeneity tested for correlation coefficients and means/standard deviations. A subgroup analysis assessed variables influencing FFQ accuracy.

Results: An overall fair/high correlation between FFQ and reference method was found; a good agreement, measured through the intake mean comparison for all nutrients except sugar, carotene and K, was observed. Kappa values showed fair/moderate agreement; an overall good ability to rank adolescents according to energy and nutrient intakes was evidenced by data of percentiles; absolute validity was not confirmed by mean differences/limits of agreement. Interviewer administration mode, consumption interval of the previous year/6 months and high number of food items are major contributors to heterogeneity and thus can reduce FFQ accuracy.

Conclusions: The meta-analysis shows that FFQ are accurate tools for collecting data and could be used for ranking adolescents in terms of energy and nutrient intakes. It suggests how the design and the validation of a new FFQ should be addressed.

Keywords
Meta-analysis
Validity
FFQ
Adolescent

Semi-quantitative FFQ are valid and reliable dietary assessment methods used worldwide on adolescents and are suggested as appropriate tools for the collection of dietary intake data in large-scale surveys[1,2], since they have the advantages of ease of administration, saving of economic resources and ability to assess dietary intake over an extended period of time[3]. Among all the used FFQ, large variations in design characteristics have been highlighted[4], such as number of food items or consumption interval.

Our recent systematic literature review[5] identified the FFQ used in adolescents and validated during the last decade throughout the world. One of the aspects emphasized by the review is that there is an ongoing need for the refinement of existing approaches, especially ones that can be used in large epidemiological studies.

When preparing the tools for dietary data collection, the specific design and validation issues of the data collecting instrument have to be taken into account. There are many factors that may affect the accuracy of a dietary questionnaire such as respondent characteristics, questionnaire design and quantification, adequacy of the reference data, quality control and data management[6], including the statistical analyses of validation data. This leads to the necessity to further characterize or create new FFQ targeted to adolescents to address the need for a valid, reproducible, user-friendly, fast, cost-effective, standardized method of accurately assessing nutrient intakes in adolescents.

The ASSO Project (Adolescents and Surveillance System for the Obesity prevention), funded by the Italian Ministry of Health and involving different national and international

partners, aims at developing an innovative web-based system for a standardized collection of data on food consumption and lifestyles in adolescents[7]. To this purpose, valid and reliable instruments are envisaged to be developed within the Project[8], including a questionnaire for the assessment of food consumption and nutrient intakes. Our previously mentioned review suggested the development of a new semi-quantitative FFQ that could fit the purposes of the ASSO Project. In order to establish the design of an appropriate FFQ that could provide valid data, the present work was aimed at conducting a meta-analysis of the validity studies of FFQ specifically addressed to adolescents. The overall degree of correlation and agreement between FFQ and the reference method was assessed and variables that can affect FFQ validity identified.

Methods

Systematic literature review

A systematic literature review was recently performed by the authors on studies describing dietary assessment methods in adolescents published worldwide between 2001 and 2012[5]. The electronic databases MEDLINE, EMBASE, ISI Web of Science and Cochrane were explored. In the MEDLINE and Cochrane databases, besides free text terms, Medical Subject Headings (MeSH) and MeSH Major Topics were included in the syntax. A sensitivity check was executed by deleting terms in the syntax systematically to see if important articles were missed with the current syntax. Publication language was restricted to the English, Italian, Spanish and French idioms. Key search terms, used alone and in combination, included the following: terms referred to the type of dietary method (questionnaire, 24-HR, 24 h recall, 24-h recall, FFQ, history, record, diary); terms including diet, nutrition, food, intake; terms related to the validation of the methods (validity, validation, accuracy, accurate). Additional searches were carried out on websites of national and international organizations (e.g. university websites and relevant professional societies or organizations) and the grey literature was also considered. The studies that used biomarkers were not considered since they often reflect status rather than intake, short-term rather than long-term intakes and are invasive and expensive[9]. The reference lists of articles retrieved for inclusion in the review were hand-searched to identify other relevant articles.

Studies that met all of the following inclusion criteria were included in the review: describing dietary assessment methods developed for epidemiological purposes; targeting adolescent populations in the age range 13–17 years; and reporting the validity and/or reproducibility of the method *v.* one reference method.

The retrieved records were sent to Endnote® (version X 4·02). After removing all duplicates, title and abstracts were screened. When a title or abstract could not be rejected with certainty, the paper was included in the eligibility papers and the full text was further evaluated. The following exclusion criteria were applied: population age not in the range 13–17 years; non-healthy subjects; hospitalized or not free-living subjects; pregnant adolescent women; refugees; vulnerable populations such as low-income or rural; specific ethnicity; overweight/obese subjects; athletes; vegetarians; dietary instrument specific only to certain nutrients (folate, vitamins, calcium, fats, proteins, etc.), specific only to certain foods (alcohol, beverages, fruit and vegetables, sugar snacks, seafood, etc.) or specific only to energy and fast-food consumption; feeding study or intervention study; subjects with eating disorders; study relative to eating or health behaviour; psychometric tests (e.g. for craving); subjects with food allergies; study relative to particular substances intake (acrylamide, etc.); questionnaire only for physical activity assessment; questionnaire only for nutrition knowledge assessment; study aimed at perceptions; study where only parental reporting on their children was considered; study with only food insecurity measurement; and study with only portion size estimation.

The full texts of the articles assessed for eligibility were then examined. Some articles and the relative full version of the questionnaires were obtained through direct contact with the author.

The literature search and the systematic review were conducted by two independent investigators, after a standardization of the procedure. In the case of any incongruity, the two investigators came to an agreement after further analysis and discussion. Further details on the systematic literature review can be found in the published paper[5].

Data extraction

Data indicating correlation and agreement between the FFQ and the reference method were considered, from each retrieved study, in relation to energy and the following nutrient intakes: protein, carbohydrate, sugar, fibre, starch, total fat, SFA, MUFA, PUFA, cholesterol, thiamin, riboflavin, niacin, vitamin B_6, vitamin B_{12}, folic acid, vitamin C, vitamin A, carotene, vitamin D, vitamin E, Ca, Mg, P, K, Na, Fe, Zn, Cu and iodine.

In detail, Pearson's or Spearman's correlation coefficients, means and standard deviations, kappa agreement, percentiles and mean agreement/limits of agreement (LOA) estimated through the Bland–Altman method were extracted and analysed. Prior to the extraction, a data extraction form was developed, which was filled by two independent reviewers after an informal training exercise.

Meta-analysis of correlation coefficients and of means/standard deviations

To determine the overall degree of correlation between FFQ and reference method, the correlation coefficients

were extracted. In addition, means and standard deviations were also extracted and meta-analysed in order to assess the overall agreement derived from pooling together the populations from different studies. All data were analysed by using the statistical software package STATA/MP 12·1, with the 'metan' command used for meta-analysis[10].

Pooled estimates were calculated using both fixed-effects and DerSimonian and Laird[11] random-effects models (that estimates the mean of a distribution of effects), weighting individual study results by the inverse of their variances. Forest plots were used to visually assess the pooled estimates and corresponding 95 % confidence intervals across studies. A test of heterogeneity was performed using a χ^2 test[12] at significance level of $P < 0.05$ and reported with the I^2 statistic, in which cut-offs of 25 %, 50 % and 75 % indicate low, moderate and high heterogeneity, respectively[13].

When the test showed significant heterogeneity, the sources of heterogeneity were explored with a meta-regression analysis, through a stratification by the following characteristics of the FFQ and of the validation study: reference method, divided into the two categories of food record (FR) and 24 h recall (24-HR); number of food items, with the two classes <114 and ≥114 (where 114 is the median value of the number of food items extracted from the FFQ); administration mode, which includes interviewer-administered (IW) and self-administered (SA) modes; collection setting, as school and non-school environment; consumption interval, with the two categories considered being previous year/6 months and previous month/week of consumption; portion size estimation method, with household units and visual serving sizes; number of subjects, with number ≤80 and >80; and study quality, where low-quality studies were compared with high-quality studies. In order to judge the methodological quality of studies based on the validation characteristics, the authors carried out a study quality assessment[5], according to the summary score described by Serra-Majem *et al.*[6], which classified studies as very good, good, acceptable/reasonable or poor. Since the number of studies is limited, in order to have variables with two modalities, the high and low categories were chosen for the meta-regression.

Sensitivity analyses were conducted to examine the contribution of each individual study by evaluating the impact of the outlier studies (e.g. observations that deviate so much from other observations as to arouse suspicion that they were generated by a different mechanism), eliminating each study from the meta-analysis and comparing the point estimates including and excluding the study.

To assess the potential of publication bias, the Egger test[14,15] was performed for examining the relative symmetry of individual study estimates around the overall estimate, displaying the results in a Galbraith plot (where

the standard normal deviate of the intervention effect estimate is plotted against its precision). To overcome the limit of the Egger test due to the presence of small studies, evidence of asymmetry was set on $P < 0.1$ and intercepts have been presented with 90 % confidence interval, as suggested by Egger *et al.*[15]. According to the suggestion that the use of this test is not reasonable for fewer than ten studies, the analysis included fourteen studies, and the outcome measures were energy, all macronutrient, Ca and Fe intakes.

In detail, the meta-analysis of correlation coefficients was conducted by retrieving all effect sizes in the form of Pearson's or Spearman's correlation coefficients, and estimating the pooled effect for energy and each nutrient considered. Pearson's or Spearman's correlation coefficients were used respectively when the sample distribution was normal (or transformed into a normal one) and when it was skewed. In some studies the correlation was considered raw; in some others the presentation of results included the adjustment of nutrients for total energy intakes using regression techniques (energy-adjusted values) and/or values de-attenuated from the weakening effect of measurement error. Thus, for each identified FFQ, raw and de-attenuated/energy-adjusted (de-att/E-adj) Pearson's and Spearman's correlation coefficients were extracted and the effect sizes of both the raw and the de-attenuated and/or energy-adjusted correlation coefficients were estimated. Following the recommendation by Hunter and Schmidt[16], correlation coefficients were not transformed into Fisher's Z scores as this transformation produces an upward bias in the mean estimation of the correlation because of the larger weights given to the larger correlations. On the other hand, this upward bias is usually higher than the negligible downward bias produced by untransformed correlations.

Cohen's rule of thumb for interpretation of the correlation coefficients was followed: a value of 0·1 indicates a small effect, a value of 0·25 indicates a medium effect and a value of 0·4 a large effect[17].

The nutrients with less than three correlation coefficient values reported (vitamin D, Cu, iodine, starch, alcohol) were excluded from the analysis. The sex-specific correlation coefficients between FFQ and the reference method were not stated in most studies, therefore we did not include them in the study comparison; when two correlation coefficients were available for males and females their mean was used as the representative value.

With regard to the meta-analysis of means and standard deviations, values for energy, macronutrient and micronutrient intakes were extracted for the test (FFQ) and reference methods (FR and 24-HR) in all the retrieved studies. They were incorporated in a meta-analysis study to estimate the overall effect, expressed as the standardized mean difference (SMD). The SMD was used since the studies all assessed the same outcome (energy and nutrients) but measured it by using instruments with

different characteristics. It expresses the size of the intervention effect in each study relative to the variability observed in that study. Cohen's rule of thumb for interpretation of the SMD statistic was followed: a value of 0·2 indicates a small effect, a value of 0·5 indicates a medium effect and a value of 0·8 or larger indicates a large effect[17].

Analysis of kappa agreement, percentiles and mean agreement/limits of agreement

Weighted kappa (κ_w), which were used as a measure of agreement[18] between FFQ and the reference method, were extracted. The agreement was classified with the following thresholds established by Landis and Koch[19]: $\kappa_w \leq 0$ indicates less than chance agreement; $\kappa_w = 0\cdot01–0\cdot20$ indicates slight agreement; $\kappa_w = 0\cdot21–0\cdot40$ fair agreement; $\kappa_w = 0\cdot41–0\cdot60$ moderate agreement; $\kappa_w = 0\cdot61–0\cdot80$ substantial agreement; $\kappa_w = 0\cdot81–0\cdot99$ indicates almost perfect agreement.

The proportions of individuals classified into percentiles (quintiles, quartiles and tertiles) were extracted in order to evaluate the ability of the FFQ in ranking subjects across levels of nutrient intake.

Mean agreement and LOA estimated through the Bland–Altman method[20] were also analysed from some studies. This method permits determining the direction of error and estimating heteroscedasticity. If differences are approximately normally distributed and not related to the magnitude of the measures (homoscedasticity), the systematic bias is estimated by the mean of the differences and the random error is estimated by the standard variation of the differences.

Results

Fourteen original articles retrieved through the mentioned systematic literature review and two more papers updated to May 2015[21–36] were identified as studies assessing the validation of FFQ against reference dietary instruments, translating the food intakes into nutrient intakes and targeting adolescent populations in the age range 13–17 years (Table 1). A high variability was highlighted between the studies[5].

Meta-analysis study of correlation coefficients

The meta-analysis of both raw and de-att/E-adj correlation coefficients showed fair/high correlation between FFQ and FR or 24-HR for energy and all nutrients (Table 2): the overall raw effect estimate was high (correlation coefficient > 0·4) for most nutrients, while it was fair (correlation coefficient = 0·25–0·39) for sugar, PUFA, cholesterol, vitamin C, vitamin A, carotene, vitamin E and Zn; the overall de-att/E-adj effect size was high for most nutrients, and fair for sugar, MUFA, PUFA, vitamin A and Na.

However, the heterogeneity was high for energy and most nutrients in raw correlation coefficients, and for half of the nutrients in de-att/E-adj correlation coefficients (Table 2). Homogeneity was found only for raw values of vitamin B$_{12}$ ($I^2 = 0\cdot0\,\%$, $P = 0\cdot962$) and de-att/E-adj values of SFA ($I^2 = 0\cdot0\,\%$, $P = 0\cdot984$); moderate heterogeneity was found for raw values of carotene ($I^2 = 35\cdot3\,\%$, $P = 0\cdot186$) and for de-att/E-adj values of protein, sugar, MUFA, Mg, P and K (Table 2).

Taking into account both the correlation coefficients and I^2 values, these values were plotted (data not shown) to obtain values with fair/high correlation (>0·25) and low/moderate heterogeneity ($I^2 < 50\,\%$) at the same time: for SFA, MUFA, protein, Mg, K and P, the two methods were well correlated and studies were quite homogeneous.

In order to investigate the factors influencing the high heterogeneity of the de-att/E-adj values, we stratified by the characteristics of the study and of the FFQ. For energy and vitamin A, the stratified analysis did not show any heterogeneity reduction; this indicates that other not observed variables, different from the characteristics of the study and FFQ, could have generated heterogeneity, such as sex, which could not be evaluated in our stratification as very few studies provided data separately for males and females.

For the other nutrients, the heterogeneity was explained mainly by the following variables: IW administration mode and number of food items ≥114. Noteworthy, for total fat, the stratification by administration mode highlighted the IW mode as a source of heterogeneity (Fig. 1).

Meta-analysis study of means and standard deviations

Table 3 shows the effect estimate with the 95 % confidence interval, heterogeneity and P value for energy and each nutrient. A significant very small effect (SMD < 0·20) of the FFQ compared with the reference method was found for protein, total fat, PUFA, cholesterol, vitamin A, vitamin E, thiamin, niacin, vitamin B$_6$, folic acid, Na and Fe; a small effect (SMD = 0·21–0·50) was found for energy, carbohydrate, SFA, MUFA, riboflavin, vitamin B$_{12}$, Ca and P.

For sugar, carotene and K, a significant SMD value between 0·51 and 0·80 was found in the direction of an overestimation. A large effect (SMD > 0·81) was not found in any of the nutrients.

Sugar, fibre, vitamin C, carotene, Mg, K and Zn showed significant overestimation when measured by the FFQ compared with the reference instrument. PUFA, cholesterol and thiamin showed an underestimation, but the SMD was not significant.

Results referring to the heterogeneity indicated that it was very high for all nutrients except for sugar and vitamin B$_6$ (Table 3).

We explored the sources of heterogeneity for energy and nutrients through stratification by the methodological

Table 1 Overview of the retrieved sixteen studies assessing the validation of FFQ against reference dietary instruments

Study	No. of food items on FFQ	Consumption interval	FFQ type	Administration mode	Administration duration (min)	No. of subjects	Portion size estimation method	Collection setting	Reference method	Statistics used for validation	Quality level*	Outcome
Ambrosini et al. (2009)[21]	212	Previous year	PB	IW	NR	785	Household units	Hospital	FR (3 d)	Pearson's CC; Bland–Altman method; tertile method	Good	Nutrient intakes
Araujo et al. (2010)[22]	93	NR	PB	IW	NR	169	Household units	School	FR (3 d, 2 weekdays and 1 weekend day)	Pearson's CC; Bland–Altman method; κ_w; quartile method	Good	Nutrient intakes
Bertoli et al. (2005)[23]	136	NR	PB	IW	30	19	Visual serving size	School	FR (7 d)	Pearson's CC; Bland–Altman method; κ_w; tertile method	Good	Nutrient intakes
Cullen et al. (2008)[24]	72	Previous week	PB	IW	NR	83	Visual serving size	Home (telephone)	24-HR (two, 1 weekday and 1 weekend day; over a 7 d period)	Pearson's CC; Bland–Altman method	Good	Food and nutrient intakes
Deschamps et al. (2009)[25]	124	Previous year	PB	IW	NR	17	Visual serving size	Home	24-HR (four, 1 each season)	Pearson's CC; quintile method	Good	Food and nutrient intakes
Hong et al. (2010)[26]	170	Previous 6 months	PB	IW	30	180	Visual serving size	School	24-HR (four, 3 weekdays and 1 weekend day)	Pearson's CC; Bland–Altman method; κ_w; quintile method	Good	Nutrient intakes
Lietz et al. (2002)[27]	130	Previous year	PB	IW	NR	50	Household units	School	FR (7 d)	Spearman's CC; Bland–Altman method; tertile method	Good	Nutrient intakes
Martinez et al. (2013)[28]	50	Previous year	PB	IW	NR	109	NR	School	24-HR (four)	Pearson's CC; κ_w; tertile method	Good	Food and nutrient intakes
Nurul-Fadhilah et al. (2012)[29]	124	Previous year	PB	IW	NR	170	Visual serving size and household units	School	24-HR (3 d, 2 weekdays and 1 weekend day)	Pearson's CC; quartile method	Acceptable/ reasonable	Nutrient intakes
Papadopoulou et al. (2008)[30]	108	Previous week	PB	IW	NR	250	Household units	School	FR (3 d)	Pearson's CC	Acceptable/ reasonable	Nutrient intakes
Rockett et al. (2007)[31]	26	NR	PB	SA	NR	261	Household units	All settings	(i) FFQ (YAQ) full list (ii) 24-HR (three, every 4 months; over a 1 year period)	Pearson's CC	Acceptable/ reasonable	Nutrient intakes
Shatenstein et al. (2010)[32]	78	Previous year	PB	SA	30	65	Visual serving size	Hospital	FR (3 d, 2 weekdays and 1 weekend day)	Spearman's CC; Bland–Altman method; quartile method	Acceptable/ reasonable	Nutrient intakes
Slater et al. (2003)[33]	76	Previous 6 months	PB	IW	NR	106	Household units	School	24-HR (three, 2 weekdays and 1 weekend day, applied at intervals of 45 d)	Pearson's CC; quartile method	Very good/ excellent	Nutrient intakes
Vereecken et al. (2010)[34]	137	NR	WB	IW	NR	55	Visual serving size	Home	24-HR (YANA-C) (four, 3 weekdays and 1 weekend day; over a 2-month period)	Spearman's CC; Bland–Altman method	Good	Food and nutrient intakes

Table 1 Continued

Study	No. of food items on FFQ	Consumption interval	FFQ type	Administration mode	Administration duration (min)	No. of subjects	Portion size estimation method	Collection setting	Reference method	Statistics used for validation	Quality level*	Outcome
Watanabe et al. (2011)[35]	82	Previous month	PB	SA	30	63	Visual serving size	School	FR (7 d) (7 d-FRRI)	Pearson's and Spearman's CC	Poor	Food and nutrient intakes
Watson et al. (2009)[36]	120	Previous 6 months	PB	SA	NR	113	Household units	School	FR (1 d)	Pearson's and Spearman's CC; Bland–Altman method; κ_w; quintile method	Good	Nutrient intakes

PB, paper-based; WB, web-based; IW, interviewer-administered; SA, self-administered; NR, not reported; FR, food record; 24-HR, 24 h recall; YAQ, Youth/Adolescent Questionnaire; YANA-C, Young Adolescents' Nutrition Assessment on Computer; 7 d-FRRI, 7 d weighed food record; CC, correlation coefficient; κ_w, weighted kappa.
*According to Serra-Majem et al.[6].

characteristics of the study and of the instrument used. The sensitivity analysis revealed that the exclusion of outliers from the analyses in most cases influenced the overall results.

With regard to energy, the heterogeneity after stratification remained always high. When the sensitivity analysis by excluding an outlier study[25] and the stratification were performed, the SMD remained low/medium and the heterogeneity was annulled in SA studies (Fig. 2), was reduced in low-quality studies (SMD = 0·29, 95 % CI 0·06, 0·52; $I^2 = 49·6$ %, $P = 0·138$) and in FFQ asking for consumption in the previous month/week (SMD = 0·11, 95 % CI −0·11, 0·32; $I^2 = 36·6$ % $P = 0·207$).

The results for almost all nutrients also showed a significant heterogeneity across the studies. The initial overall effect for carbohydrates was 0·45 and studies showed high heterogeneity. The exclusion of outliers[22,31] improved the SMD (0·28, 95 % CI 0·10, 0·46) and decreased heterogeneity, even though it remained at high levels. Stratifying to investigate the sources of heterogeneity, the FFQ with SA mode of administration had lower heterogeneity ($I^2 = 35·1$ %, $P = 0·214$) compared with the IW mode ($I^2 = 81·7$ %, $P = 0·000$), even though the effect was higher than in the IW (SMD = 0·58, 95 % CI 0·34, 0·81 v. SMD = 0·18, 95 % CI −0·01, 0·36). The studies with <80 subjects were moderately heterogeneous ($I^2 = 44·8$ %, $P = 0·107$) even though SMD was 0·60. Reference method, collection setting, portion size estimation method and quality did not affect the overall effect.

For the intake of fibre, evaluated after excluding the outliers[22,29,31], the consumption interval of the previous month/week showed a low effect and a medium heterogeneity (SMD = 0·23, 95 % CI −0·02, 0·48; $I^2 = 50·8$ %, $P = 0·131$). Stratifying by food items, the FFQ with fewer than 114 items had SMD of 0·21 (95 % CI 0·01, 0·41) and a reduced heterogeneity ($I^2 = 64·1$ %, $P = 0·016$). The low-quality studies showed SMD of 0·37 (95 % CI 0·17, 0·56) and $I^2 = 33·1$ %, $P = 0·225$.

Concerning protein, heterogeneity remained high even after eliminating the outliers[27,31]. The stratification analysis revealed that SA FFQ had a fair SMD of 0·37 (95 % CI 0·11, 0·64) and low heterogeneity ($I^2 = 49·0$ %, $P = 0·139$) compared with the IW ones. Also study quality influenced heterogeneity, with high-quality studies explaining the heterogeneity (Fig. 3).

For total fat intake, after the exclusion of three outliers[22,24,31], the heterogeneity was reduced for the consumption interval of the previous month/week (SMD = 0·09, 95 % CI −0·13, 0·30; $I^2 = 31·4$ %, $P = 0·227$).

Analysing the intake of SFA, after eliminating the outliers[21,31], an overall medium SMD (0·40) was observed and heterogeneity decreased ($I^2 = 52·0$ %, $P = 0·100$).

In the sensitivity analysis for PUFA, the exclusion of one study[31] did not modify the overall heterogeneity. Stratifying by study quality, low-quality studies had a low effect and were homogeneous (SMD = 0·25, 95 % CI 0·09, 0·41; $I^2 = 0·0$ %, $P = 0·595$).

Table 2 Pooled effect estimates (ES) and heterogeneity of raw and de-attenuated/energy-adjusted (de-att/E-adj) correlation coefficients (CC) for energy and nutrients

| Nutrient	Raw CC					De-att/E-adj CC				
	No. of studies	Pooled ES	95 % CI	I^2 %	P	No. of studies	Pooled ES	95 % CI	I^2 %	P
Energy	16	0·51	0·38, 0·64	96·1	<0·001	–	–	–	–	–
Protein	15	0·46	0·33, 0·58	93·6	<0·001	11	0·40	0·33, 0·46	39·3	0·087
Carbohydrate	15	0·47	0·37, 0·58	92·0	<0·001	11	0·50	0·37, 0·62	88·0	<0·001
Sugar	4	0·37	0·23, 0·51	68·7	0·022	3	0·38	0·21, 0·56	56·9	0·098
Fibre	15	0·40	0·33, 0·48	80·1	<0·001	12	0·46	0·37, 0·54	79·9	<0·001
Total fat	16	0·49	0·36, 0·61	93·9	<0·001	12	0·49	0·40, 0·59	79·8	<0·001
SFA	7	0·48	0·34, 0·63	93·1	<0·001	4	0·44	0·38, 0·50	0·0	0·984
MUFA	7	0·40	0·24, 0·57	93·0	<0·001	4	0·35	0·25, 0·45	47·4	0·127
PUFA	8	0·29	0·10, 0·48	95·0	<0·001	5	0·26	0·15, 0·38	65·8	0·020
Cholesterol	10	0·38	0·30, 0·44	68·5	0·001	7	0·46	0·33, 0·59	82·5	<0·001
Thiamin	8	0·44	0·36, 0·51	69·7	0·002	5	0·47	0·37, 0·56	62·2	0·032
Riboflavin	8	0·49	0·34, 0·63	94·4	<0·001	5	0·47	0·37, 0·57	63·7	0·041
Niacin	8	0·42	0·29, 0·55	90·8	<0·001	5	0·41	0·31, 0·52	71·3	0·008
Vitamin B_6	3	0·51	0·29, 0·74	92·8	<0·001	NA	NA	NA	NA	NA
Vitamin B_{12}	3	0·48	0·42, 0·55	0·0	0·962	NA	NA	NA	NA	NA
Folic acid	7	0·41	0·29, 0·52	82·1	<0·001	4	0·46	0·29, 0·65	86·7	<0·001
Vitamin C	10	0·39	0·28, 0·49	86·4	<0·001	7	0·48	0·23, 0·72	97·2	<0·001
Vitamin A	9	0·39	0·26, 0·51	89·7	<0·001	6	0·37	0·23, 0·50	87·6	<0·001
Carotene	5	0·29	0·23, 0·36	35·3	0·186	4	0·42	0·32, 0·52	61·9	0·049
Vitamin E	4	0·39	0·25, 0·53	76·1	0·003	NA	NA	NA	NA	NA
Ca	15	0·52	0·42, 0·62	91·1	<0·001	11	0·55	0·48, 0·61	54·8	0·014
Mg	8	0·50	0·36, 0·64	93·1	<0·001	5	0·51	0·43, 0·58	39·8	0·156
P	5	0·51	0·33, 0·89	94·8	<0·001	4	0·50	0·44, 0·57	36·0	0·196
K	7	0·46	0·29, 0·63	92·2	<0·001	5	0·42	0·35, 0·50	46·6	0·112
Na	7	0·41	0·22, 0·60	92·5	<0·001	5	0·35	0·11, 0·59	91·2	<0·001
Fe	12	0·45	0·36, 0·55	86·8	<0·001	9	0·44	0·33, 0·55	84·2	<0·001
Zn	8	0·38	0·27, 0·49	82·7	<0·001	5	0·40	0·27, 0·52	81·0	<0·001

NA, not available.
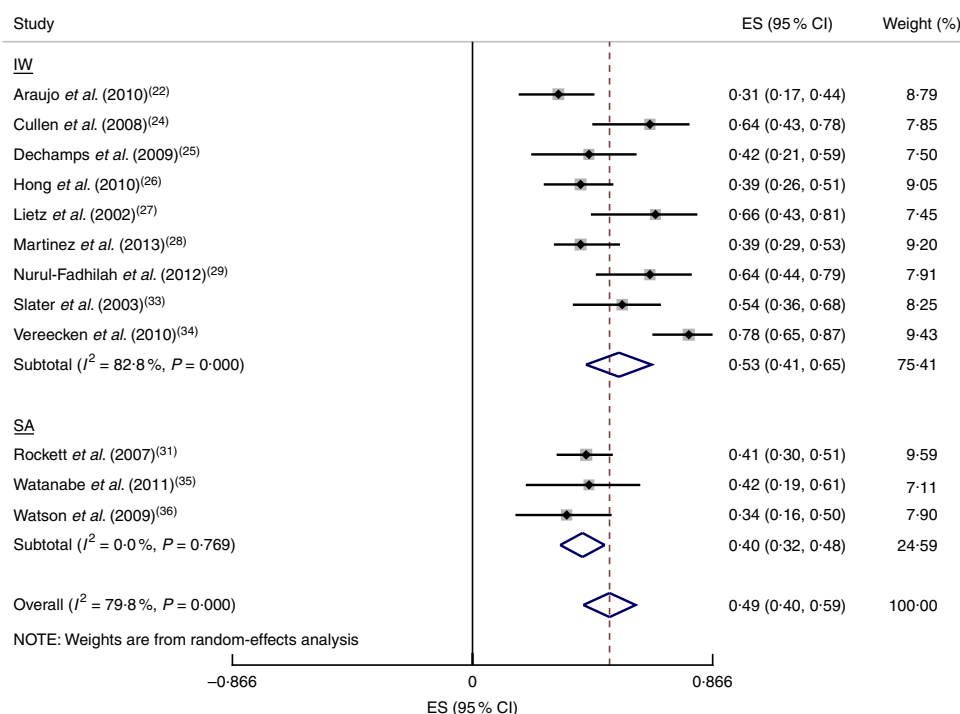
Fig. 1 Forest plot of effect estimates (ES) for the correlation coefficients of total fat intake in adolescents estimated by FFQ compared with a reference dietary instrument of food records or 24 h recalls, by administration mode (IW, interviewer-administered; SA, self-administered). The study-specific ES and 95 % CI are represented by the black diamond and horizontal line, respectively; the area of the grey square is proportional to the specific-study weight to the overall meta-analysis. The centre of the open diamond presents the pooled ES and its width represents the pooled 95 % CI

Table 3 Pooled effect estimates (standardized mean differences (SMD)) and heterogeneity of the means and standard deviations of energy and nutrients

Nutrient	No. of studies	SMD	95 % CI	I^2 %	P
Energy	16	0·22	−0·10, 0·54	96·4	<0·001
Protein	15	0·18	−0·08, 0·45	94·5	<0·001
Carbohydrate	15	0·45	−0·04, 0·94	98·4	<0·001
Sugar	4	0·59	0·44, 0·73	30·1	0·232
Fibre	15	0·40	0·05, 0·76	97·0	<0·001
Total fat	16	0·15	−0·25, 0·54	97·5	<0·001
SFA	7	0·25	−0·67, 1·17	99·3	<0·001
MUFA	7	0·41	−0·47, 1·29	99·1	<0·001
PUFA	8	−0·03	−0·58, 0·52	98·1	<0·001
Cholesterol	10	−0·04	−0·38, 0·29	95·9	<0·001
Thiamin	8	−0·03	−0·24, 0·31	92·8	<0·001
Riboflavin	7	0·35	−0·11, 0·81	97·5	<0·001
Niacin	7	0·12	−0·63, 0·87	99·0	<0·001
Vitamin B_6	2	0·16	−0·01, 0·32	5·6	0·303
Vitamin B_{12}	3	0·33	−0·25, 0·91	95·3	<0·001
Folic acid	5	0·02	−0·43, 0·46	96·2	<0·001
Vitamin C	9	0·32	0·06, 0·58	92·9	<0·001
Vitamin A	7	0·10	−0·18, 0·37	90·8	<0·001
Carotene	5	0·56	0·29, 0·85	91·9	<0·001
Vitamin E	4	0·18	−0·14, 0·49	86·9	<0·001
Ca	15	0·26	−0·01, 0·54	94·8	<0·001
Mg	7	0·36	0·11, 0·60	89·6	<0·001
P	5	0·33	−0·33, 0·96	97·8	<0·001
K	6	0·60	0·11, 1·09	95·1	<0·001
Na	6	0·15	−0·51, 0·80	97·2	<0·001
Fe	12	0·15	−0·17, 0·47	96·2	<0·001
Zn	8	0·34	0·09, 0·58	91·9	<0·001
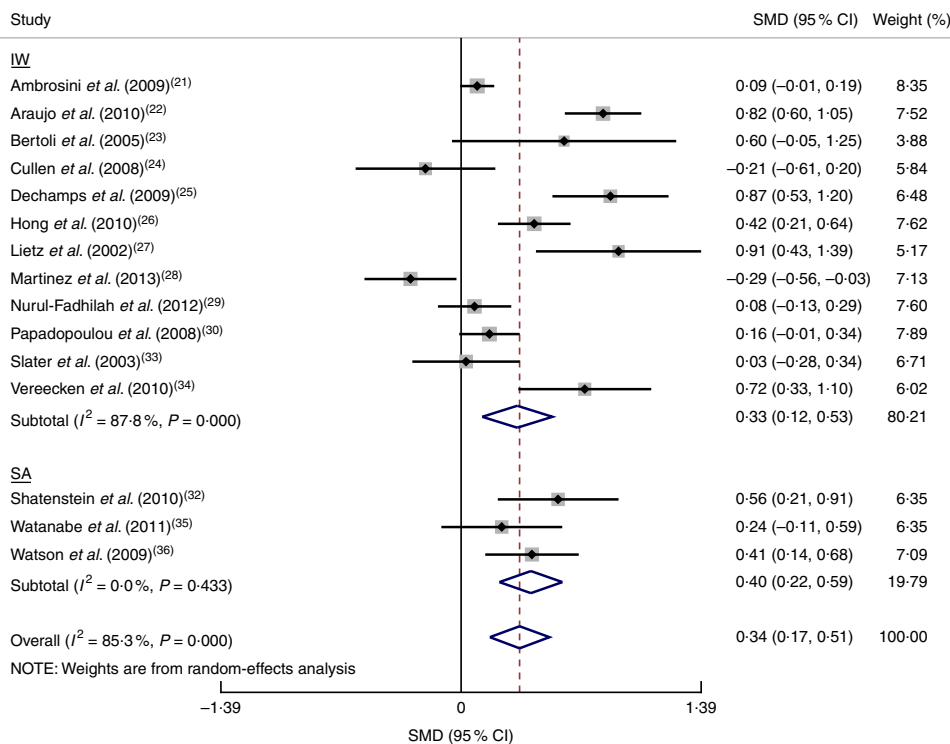
Fig. 2 Forest plot of standardized mean differences (SMD) of the energy intake in adolescents estimated by FFQ compared with a reference dietary instrument of food records or 24 h recalls, by administration mode (IW, interviewer-administered; SA, self-administered). The study-specific SMD and 95 % CI are represented by the black diamond and horizontal line, respectively; the area of the grey square is proportional to the specific-study weight to the overall meta-analysis. The centre of the open diamond presents the pooled SMD and its width represents the pooled 95 % CI
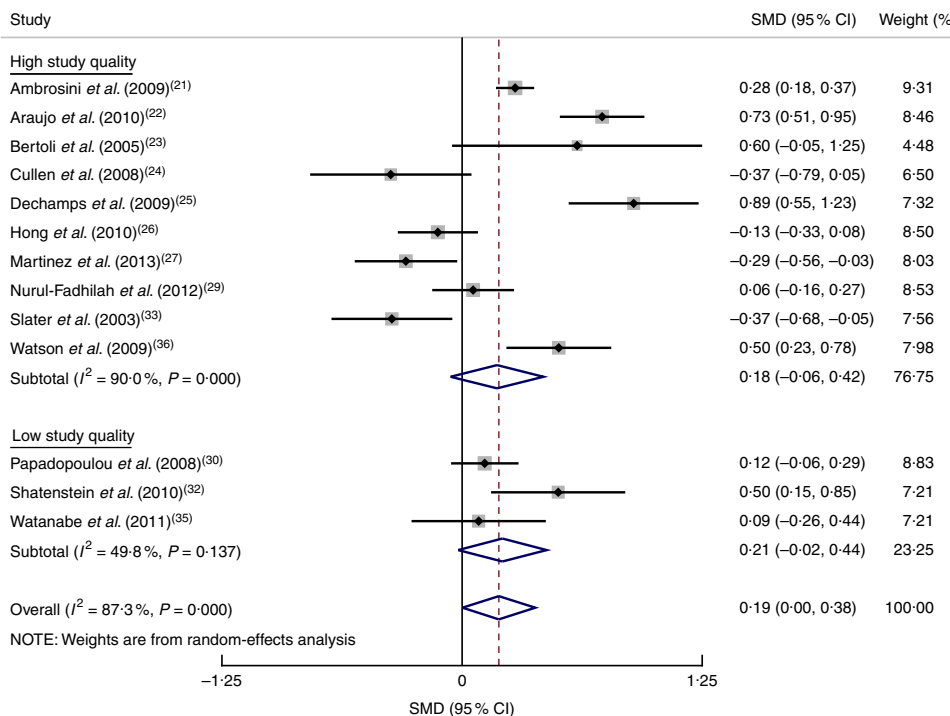
Fig. 3 Forest plot of standardized mean differences (SMD) of the protein intake in adolescents estimated by FFQ compared with a reference dietary instrument of food records or 24 h recalls, by study quality. The study-specific SMD and 95 % CI are represented by the black diamond and horizontal line, respectively; the area of the grey square is proportional to the specific-study weight to the overall meta-analysis. The centre of the open diamond presents the pooled SMD and its width represents the pooled 95 % CI

After exclusion of the outliers[28,31] and sensitivity analysis for MUFA, the effect remained low/medium and the heterogeneity decreased for SA administration mode (SMD = 0·31, 95 % CI −0·12, 0·73; $I^2 = 72·7$ %, $P = 0·055$) and household units (SMD = 0·23, 95 % CI 0·13, 0·33; $I^2 = 3·6$ %, $P = 0·308$).

In the cholesterol analysis, the stratification by excluding the outlier[31] showed that studies using FR as reference method and low-quality studies became homogeneous (SMD = 0·25, 95 % CI 0·17, 0·33; $I^2 = 0·0$ %, $P = 0·958$ and SMD = 0·27, 95 % CI 0·11, 0·43; $I^2 = 0·0$ %, $P = 0·857$, respectively).

With respect to the vitamins, the sensitivity analysis showed that studies where the number of food items was ≥114 were homogeneous and with a low effect estimate for thiamin (SMD = −0·11, 95 % CI −0·20, 0·03; $I^2 = 0·0$ %, $P = 0·513$). After excluding the outlier[29], riboflavin showed less heterogeneity and a low SMD in studies using FR as the reference method (SMD = 0·26, 95 % CI 0·12, 0·41; $I^2 = 52·6$ %, $P = 0·097$), in FFQ having number of food items ≥114 (SMD = 0·22, 95 % CI 0·13, 0·31; $I^2 = 3·9$ %, $P = 0·353$), in FFQ being IW (SMD = 0·25, 95 % CI 0·14, 0·36; $I^2 = 36·1$ %, $P = 0·209$) and administered within the school environment (SMD = 0·32, 95 % CI 0·20, 0·45; $I^2 = 8·2$ %, $P = 0·337$). Vitamin C showed low heterogeneity and low effect in FFQ with <80 subjects (SMD = 0·76, 95 % CI 0·50, 1·02; $I^2 = 0·00$ %, $P = 0·662$). For folic acid, the school environment showed SMD of 0·27 (95 % CI 0·15, 0·39) with $I^2 = 0·0$ % and $P = 0·554$ (Fig. 4). For vitamin A, the SA mode had SMD of 0·59 (95 % CI 0·28, 0·90; $I^2 = 48·3$ %, $P = 0·164$).

With regard to minerals, the analysis of Fe, after excluding the outlier[22], showed low heterogeneity and low effect when the FR was used as the reference method (SMD = 0·25, 95 % CI 0·12, 0·38; $I^2 = 47·4$ %, $P = 0·107$). For Mg intake, different variables explained the heterogeneity, even though the SMD was always significant: the 24-HR method (SMD = 0·18, 95 % CI 0·02, 0·35; $I^2 = 18·4$ %, $P = 0·268$); number of food items <114 (SMD = 0·43, 95 % CI 0·17, 0·369; $I^2 = 58·7$ %, $P = 0·089$; Fig. 5); the previous month/week (SMD = 0·32, 95 % CI 0·11, 0·52; $I^2 = 27·1$ %, $P = 0·241$); the SA method (SMD = 0·71, 95 % CI 0·44, 0·98; $I^2 = 50·4$ %, $P = 0·133$); and number of subjects <80 (SMD = 0·58, 95 % CI 0·32, 0·83; $I^2 = 0·0$ %, $P = 0·485$).

Correlation coefficients and standardized mean differences

Finally, plotting the correlation coefficients v. the SMD for energy and nutrients (Fig. 6), a high agreement between the effect estimate derived from the two meta-analyses (correlation coefficient > 0·40 and SMD < 0·20) was present for protein, total fat, thiamin, niacin, vitamin B_6, folic acid, Fe and Na, thus indicating that these nutrients are well assessed by the FFQ. Sugar and carotene intakes had, instead, both low correlation coefficient and high

SMD (Fig. 6). All the other nutrients showed low/moderate SMD and fair correlation coefficients.

Analysis of kappa agreement, percentiles and mean agreement/limits of agreement

These data were reported as measures of FFQ validity by some of the considered studies.

With regard to kappa agreement, five studies reported the related values for macronutrients[22,23,26,28,34] ranging from fair to moderate agreement (Table 4), with a mean κ_w of 0·43 for energy, 0·29 for protein, 0·40 for carbohydrate, 0·29 for total fat and 0·36 for Ca. Lower κ_w values were found only for PUFA (0·15)[28], protein (0·16)[23], SFA (0·18) and vitamin C (0·17)[28] intakes.

An overall good ranking ability was evidenced by data of percentiles. Eleven studies calculated the percentage of subjects' ranking[21–23,25–31,34] through the quintile, quartile or tertile method, reporting good ranges of agreement and low ranges of disagreement (Table 4).

Other studies reported good/acceptable estimates of mean agreement and LOA[23,26,32,36], except for retinol in the study by Hong et al.[26] and for Ca in the study by Watson et al.[36] that showed wide LOA. Other studies[21,22,24,27,34] showed, on the contrary, low values of agreement, thus stating that the examined FFQ are not able to assess the absolute intake of nutrients in adolescents.

Discussion

The present analysis showed a good overall correlation and agreement between the FFQ and the reference method in collecting data on energy and nutrient intakes in studies on adolescents. It provided information on the factors that could negatively affect the accuracy of an FFQ, namely IW administration mode, consumption interval of the previous year/6 months and high number of food items.

Moreover, the study added indications on what nutrients should be taken particularly into account when assessing their intake through an FFQ, such as sugar, carotene and K, whose intake was on average significantly overestimated by the use of FFQ.

When examining the degree of correlation, all the retrieved studies reported correlation coefficient values, and the overall correlation resulted fair/high for all nutrients considered. The heterogeneity was high for raw correlation coefficients, while it decreased in de-att/E-adj values, thus suggesting that it is important to correct from the weakening effect of measurement error and for energy intake when performing statistical analysis in these kinds of study. After exploring sources of heterogeneity, two variables were shown mainly to affect FFQ: IW administration mode and number of food items ≥114. Therefore, the SA mode could be considered a valid approach of questionnaire administration, as it is inexpensive, quick, well suited for
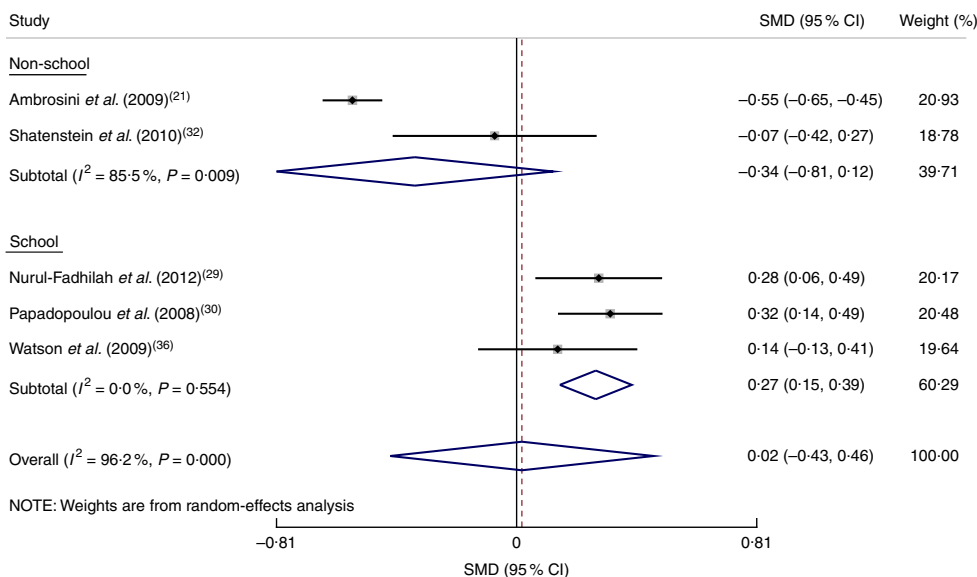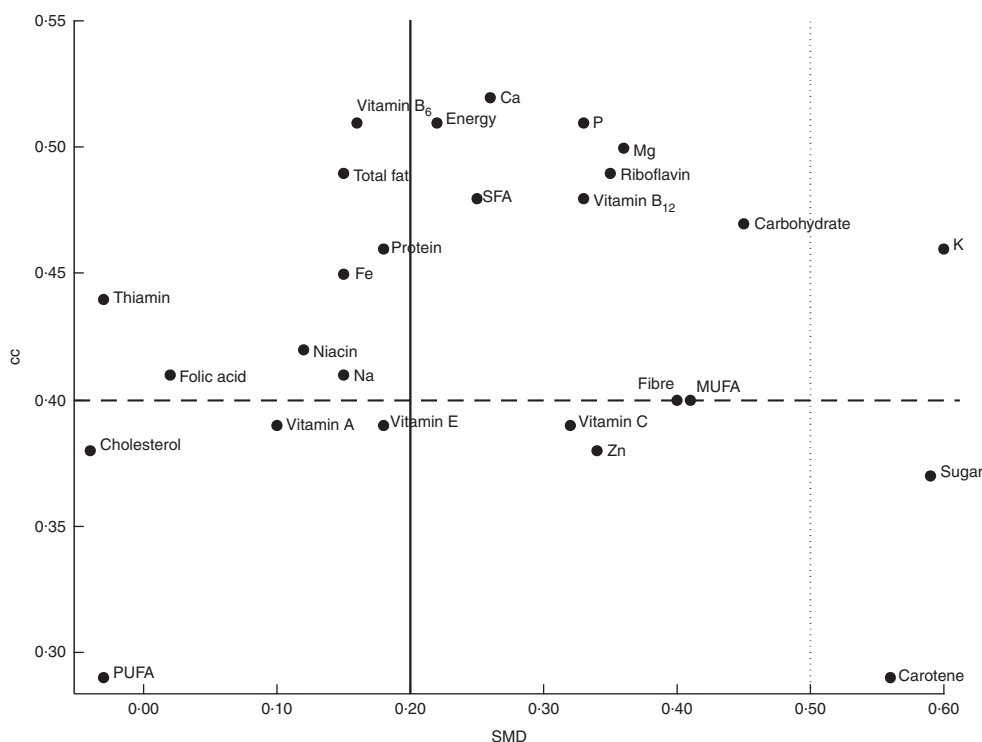
Fig. 4 Forest plot of standardized mean differences (SMD) of the folate intake in adolescents estimated by FFQ compared with a reference dietary instrument of food records or 24 h recalls, by data collection setting. The study-specific SMD and 95 % CI are represented by the black diamond and horizontal line, respectively; the area of the grey square is proportional to the specific-study weight to the overall meta-analysis. The centre of the open diamond presents the pooled SMD and its width represents the pooled 95 % CI
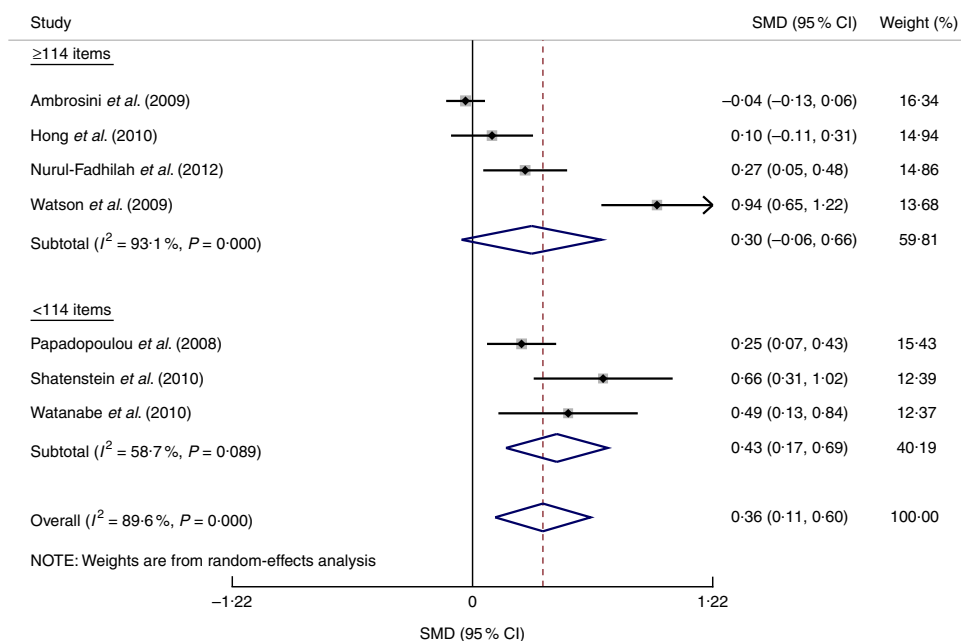
Fig. 5 Forest plot of standardized mean differences (SMD) of the magnesium intake in adolescents estimated by FFQ compared with a reference dietary instrument of food records or 24 h recalls, by number of food items on the FFQ. The study-specific SMD and 95 % CI are represented by the black diamond and horizontal line, respectively; the area of the grey square is proportional to the specific-study weight to the overall meta-analysis. The centre of the open diamond presents the pooled SMD and its width represents the pooled 95 % CI

simple questionnaires, and allows by-passing the issue of confidentiality and the engagement of human resources when administration by an interviewer is performed. Similarly, a not too long FFQ could provide accurate information on nutrient intake, as adolescents can better focus on their intake. With regard to the meta-analysis of means and standard deviations, a very small or small effect was found for over- or underestimation, thus revealing that the FFQ could be considered an accurate instrument for assessing intakes of energy and most nutrients in adolescents. Only sugar, carotene and K (and their food sources) should be taken into account when assessing their intake through an existing or a new FFQ, since despite fair/high correlation coefficients

Fig. 6 Correlation coefficient (CC) *v.* standardized mean difference (SMD) of intake for energy and nutrients in adolescents estimated by FFQ compared with a reference dietary instrument of food records or 24 h recalls, showing agreement between the effect estimates derived from the two meta-analyses (– – –, CC = 0·4 indicates a large effect; ———, SMD = 0·2 indicates a small effect; · · · · ·, SMD = 0·5 indicates a medium effect)

found between the FFQ and reference method, their intake was not assessed well through the examined FFQ.

The overestimation of sugar is probably due to the difficulty in the evaluation of sugar in the different foods such as soft drinks, biscuits, cakes, ice creams, chocolate, sweets or candies; overestimates can occur because the added sugars in the pyramid tip include oligosaccharides[37]. The carotene and K overestimates are not easy to elucidate. It is likely that the overestimation could be higher with items least frequently reported. This suggests that careful consideration must be given to the measurement of their dietary sources when a new FFQ has to be developed. The main sources of carotene are carrots, dark green leafy vegetables, melons and squashes, peas, broccoli, and tree fruits such as sour cherries and apricots. The main sources of K are fruit (dried apricots, avocados and bananas), dark leafy greens, legumes such as white beans, and cereals.

Studies assessing sugar and vitamin B6 intakes were found to be quite homogeneous. However, for vitamin B6 the intake was assessed only in two studies, and since a limit of the meta-analysis is that it has low power when studies are few, this result should be handled carefully. We could also consider that this result may be related to other possibilities, i.e. that vitamin B6 is found in high concentrations in a few food items that are generally consumed in small quantities, such as Marmite or seeds.

Even though combining crude data of mean and standard deviation resulted in an initial high heterogeneity for energy and all nutrients, this variability was explained by some characteristics of the FFQ and of the study design. The strongest contributors to the heterogeneity for all the other nutrients were the IW administration mode and consumption interval of previous year/6 months, which should be carefully considered when developing and validating a new FFQ. The meta-analysis of correlation coefficients partially confirmed these findings, indicating as powerful source of heterogeneity the IW administration mode. The SMD, however, provides a clearer indication on the difference between the intakes assessed through the FFQ and the reference method, while the analysis of correlation coefficients provides only the degree of association between the FFQ and the reference method and is not appropriate to assess validity[38,39]. The meta-analysis approach for the comparison of means and standard deviations, thus, better allows predicting the accuracy of the examined instrument.

In the means and standard deviations analysis, the number of food items did not reveal a clear direction in influencing the validity of the FFQ. It was suggested not to reduce the length of the food list too much when developing FFQ to rank persons according to nutrient intake[4], as short FFQ lack details on some food intake. On the other side, there is some evidence that overestimation increases

Table 4 Agreement degree and ability to rank subjects according to energy and nutrient levels of the FFQ examined in the sixteen retrieved articles

Study	Agreement (estimated through κ_w)	Ability to rank subjects (percentiles)	Agreement (estimated through the Bland–Altman method)
Ambrosini et al. (2009)[21]		Good (tertiles)	Low
Araujo et al. (2010)[22]	Fair	Good (quartiles)	Low
Bertoli et al. (2005)[23]	Fair	Good (tertiles)	Good
Cullen et al. (2008)[24]			Low
Deschamps et al. (2009)[25]		Good (quintiles)	
Hong et al. (2010)[26]	Moderate	Good (quintiles)	Good
Lietz et al. (2002)[27]		Good (tertiles)	Low
Martinez et al. (2013)[28]	Moderate	Good (tertiles)	
Nurul-Fadhilah et al. (2012)[29]		Good (quartiles)	
Papadopoulou et al. (2008)[30]			
Rockett et al. (2007)[31]			
Shatenstein et al. (2010)[32]		Good (quartiles)	Acceptable
Slater et al. (2003)[33]		Good (quartiles)	
Vereecken et al. (2010)[34]			Low
Watanabe et al. (2011)[35]			
Watson et al. (2009)[36]	Fair	Good (quintiles)	Fairly good

κ_w, weighted kappa.

with the length of the food list[40]; long and extensive FFQ may contribute to lower response rates since subjects may require long times to answer and become fatigued and frustrated, thus contrasting with the purpose of developing a fast and easy FFQ. Therefore, we think that the number of food items of a potential new FFQ should be no longer than 114 items.

One important issue when considering the validity of an FFQ is the food composition database that is used to convert foods into nutrients. Even though the influence of the use of different databases could not be evaluated by the current meta-analysis, it could be interesting to evaluate, beyond the number, also the allocation of food items, and compare them in the different FFQ. One common procedure when developing a new FFQ is that the composition database is arranged according to the way the foods are grouped. Different FFQ often gather food groups in different ways, thus leading to a variable conversion into nutrient intakes and to loss of information. An indication for future studies could be to evaluate how the foods are grouped and whether the different FFQ contain all the important food items.

The estimation of portion size is difficult for adults and children and is potentially a large source of error in dietary assessment; food models appear to be less accurate than photographs for estimating portion size[40]. In line with the study from Molag *et al.*[4], the portion size estimation method was found not as affecting validity in one specific direction; then we decided that the portion size estimation method of the ASSO-FFQ will be based on photographs and on household units when necessary. However, it is important that the portion size photographs are age appropriate, in order to reduce overestimation[41].

Although a high heterogeneity across studies was initially shown, information on the sources of heterogeneity was obtained from the subgroup analysis, from the sensitivity analysis and the exploration of publication bias.

Kappa agreement, the percentile method and the Bland–Altman method are suitable to assess the accuracy of a questionnaire, but not all the retrieved studies reported them as measures of their FFQ validity.

An overall fair/moderate agreement between FFQ and reference method measured by κ_w was reported in five studies, this confirming that the FFQ is able to fairly assess intakes of nutrients. The ability to rank subjects according to levels of nutrient intakes is always present on FFQ, as evidenced by the values of percentiles reported in nine of the considered studies. Five out of nine studies, instead, reported a low agreement estimated through the Bland–Altman analysis; this not confirming an overall absolute validity of the examined FFQ. It should be specified that the method of Bland and Altman that includes the LOA remains the one suggested to assess the absolute validity of an FFQ, but unfortunately it is used in few studies.

A limitation of the present meta-analysis is that it is based on observational studies; therefore, many confounding factors that might affect the correlation of energy and nutrient intakes between FR and FFQ could not be controlled. Moreover, since our selection criteria excluded articles before the year 2000 and papers analysing the validity concerning specific nutrients, this could have influenced our results in different ways. For example, we could have collected more data on some nutrients such as starch, vitamin D, Cu and iodine, performing the analysis on them as well; results on vitamin B_6, which were found only
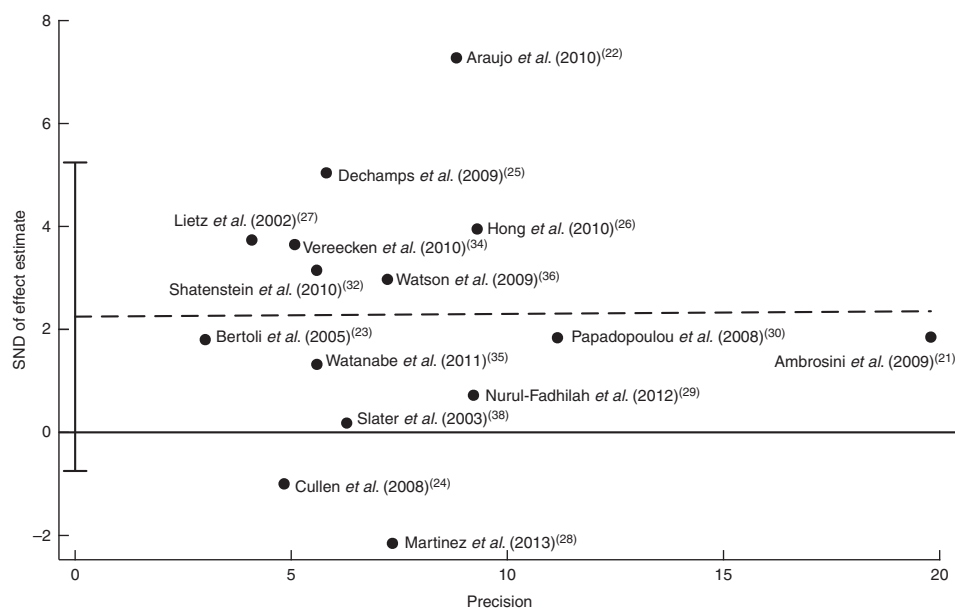
Fig. 7 Galbraith plot of the standard normal deviate (SND) of effect estimate *v.* precision for energy intake in the examined studies. Regression line and 95 % CI for the intercept are represented by — — — — and the error bar, respectively

in two studies, could have been affected by the presence of other data on that vitamin.

Another limitation is due to the fact that we could not remove the effect of sex by conducting separate analyses for males and females, since very few studies provided data separately for males and females. Anyway, it is known that females generally better evaluate their food intake. Moreover, Galbraith plots revealed asymmetry for energy (Fig. 7) and carbohydrates, indicating the presence of publication bias. Thus, results obtained for energy and carbohydrates should be handled carefully. For all the other nutrients no publication bias was present. For some nutrients, such as starch, vitamin D, Cu, iodine and alcohol, intake was not collected in all studies, making it more difficult to draw conclusions.

There could be other limitations due to other factors that could not be analysed within the present meta-analysis, such as the adolescents' level of understanding of the questions. All the FFQ we have analysed are specifically addressed to adolescents, so it is supposed that FFQ should be age specific, with questions easily understandable by the students. Actually, this could be tested in a small sample before administering the FFQ to the population, in order to understand whether it is suitable for the target population. Moreover, it would need evaluating whether the FFQ is culturally specific, as this could influence the accuracy and precision of the instrument.

Finally, too few studies were found for web-based FFQ and therefore we could not analyse the strength of the web-based method, even though our recent review[5] showed that the FFQ from Matthys *et al.*[42], the 24-HR 'Synchronised Nutrition and Activity Program™' (SNAP™)[43],

the 24-HR Young Adolescents' Nutrition Assessment on Computer (YANA-C)[44,45], the Health Behaviour in School-aged Children (HBSC) FFQ[46] and the Healthy Lifestyle by Nutrition in Adolescence (HELENA) FFQ[34], all being web-based, could fit the purpose.

The present analysis of the combination of different studies on FFQ developed worldwide confirms that FFQ are robust instruments for ranking adolescents according to energy and nutrient intake levels, even though their absolute validity has not always been demonstrated.

Specific variables that can negatively affect the validity of an FFQ in relation to energy and nutrient intakes were identified, such as the IW administration method, a high number of food items and the consumption interval requested being a long interval, and some nutrients were recognized not to be well assessed by FFQ (sugar, carotene, K), thus suggesting to the scientific community how the design and the validation of a new FFQ might be addressed.

Acknowledgements

of data and contributed in drafting the article. E.A. and A.B. contributed in drafting the article and revising it critically. M.J., A.F. and C.M. revised the article critically. *Ethics of human subject participation:* The ethical approval was given by the ethics committee of the 'Azienda Ospedaliera Universitaria Policlinico Paolo Giaccone' (approval code number 9/2011).

References

1. Ortiz-Andrellucchi A, Henríquez-Sánchez P, Sánchez-Villegas A *et al.* (2009) Dietary assessment methods for micronutrient intake in infants, children and adolescents: a systematic review. *Br J Nutr* **102**, Suppl. 1, S87–S117.
2. Cade JE, Burley VJ, Warm DL *et al.* (2004) Food frequency questionnaires: a review of their design, validation and utilisation. *Nutr Res Rev* **17**, 5–22.
3. Subar AF (2004) Developing dietary assessment tools. *J Am Diet Assoc* **104**, 769–770.
4. Molag ML, de Vries JHM, Ocké MC *et al.* (2007) Design characteristics of food frequency questionnaires in relation to their validity. *Am J Epidemiol* **166**, 1468–1478.
5. Tabacchi G, Amodio E, Di Pasquale M *et al.* (2014) Validation and reproducibility of dietary assessment methods in adolescents: a systematic literature review. *Public Health Nutr* **17**, 2700–2714.
6. Serra-Majem L, Frost Andersen L, Henríque Sánchez P *et al.* (2009) Evaluating the quality of dietary intake validation studies. *Br J Nutr* **102**, Suppl. 1, S3–S9.
7. Tabacchi G (2011) Asso project: a challenge in the obesity prevention context. *J Sport Sci Law* **1–3**, issue III, 2.
8. Tabacchi G & Bianco A (2011) Methodological aspects in the development of the lifestyle surveillance toolkit in the ASSO project. *J Sport Sci Law* **IV**, issue 4, 96–100.
9. Lampe JW & Rock CL (2008) Biomarkers and their use in nutrition intervention. In *Nutrition in the Prevention and Treatment of Disease*, 2nd ed., pp. 187–201 [AM Coulston and CJ Boushey, editors]. San Diego, CA: Academic Press.
10. Bradburn MJ, Deeks JJ & Altman DG (1999) sbe24: metan – an alternative meta-analysis command. *Stata Tech Bull* **44**, 4–15.
11. DerSimonian R & Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177–188.
12. Fleiss JL (1993) The statistical basis of meta-analysis. *Stat Methods Med Res* **2**, 121–145.
13. Higgins JP, Thompson SG, Deeks JJ *et al.* (2003) Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560.
14. Sterne JA & Egger M (2001) Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* **54**, 1046–1055.
15. Egger M, Davey Smith G, Schneider M *et al.* (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629–634.
16. Hunter JE & Schmidt FL (1990) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: SAGE.
17. Cohen J (1998) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.
18. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* **70**, 213–220.
19. Landis JR & Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.
20. Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.
21. Ambrosini GL, de Klerk NH, O'Sullivan TA *et al.* (2009) The reliability of a food frequency questionnaire for use among adolescents. *Eur J Clin Nutr* **63**, 1251–1259.
22. Arajuo MC, Yokoo EM & Pereira RA (2010) Validation and calibration of a semi quantitative food frequency questionnaire designed for adolescents. *J Am Diet Assoc* **110**, 1170–1177.
23. Bertoli S, Petroni ML, Pagliato E *et al.* (2005) Validation of food frequency questionnaire for assessing dietary macronutrients and Ca intake in Italian children and adolescents. *J Pediatr Gastroenterol Nutr* **40**, 555–560.
24. Cullen KW, Watson K & Zakeri I (2008) Relative reliability and validity of the Block Kids Questionnaire among youth aged 10 to 17 years. *J Am Diet Assoc* **108**, 862–866.
25. Deschamps V, De Lauzon-Guillain B, Lafay L *et al.* (2009) Reproducibility and relative validity of a food-frequency questionnaire among French adults and adolescents. *Eur J Clin Nutr* **63**, 282–291.
26. Hong TK, Dibley MJ & Sibbritt D (2010) Validity and reliability of an FFQ for use with adolescents in Ho Chi Minh City, Vietnam. *Public Health Nutr* **13**, 368–375.
27. Lietz G, Barton KL, Longbottom PJ *et al.* (2002) Can the EPIC food-frequency questionnaire be used in adolescent populations? *Public Health Nutr* **5**, 783–789.
28. Martinez MF, Philippi ST, Estima C *et al.* (2013) Validity and reproducibility of a food frequency questionnaire to assess food group intake in adolescents. *Cad Saude Publica* **29**, 1795–1804.
29. Nurul-Fadhilah A, SzeTeo P & Huat Foo L (2012) Validity and reproducibility of a food frequency questionnaire (FFQ) for dietary assessment in Malay adolescents in Malaysia. *Asia Pac J Clin Nutr* **21**, 97–103.
30. Papadopoulou SK, Barboukis V, Dalkiranis A *et al.* (2008) Validation of a questionnaire assessing food frequency and nutritional intake in Greek adolescents. *Int J Food Sci Nutr* **59**, 148–154.
31. Rockett HR, Berkey CS & Colditz GA (2007) Comparison of a short food frequency questionnaire with the Youth/Adolescent Questionnaire in the Growing Up Today Study. *Int J Pediatr Obes* **2**, 31–39.
32. Shatenstein B, Amre D, Jabbour M *et al.* (2010) Examining the relative validity of an adult food frequency questionnaire in children and adolescents. *J Pediatr Gastroenterol Nutr* **51**, 645–652.
33. Slater B, Philippi ST, Fisberg RM *et al.* (2003) Validation of a semi-quantitative adolescent food frequency questionnaire applied at a public school in São Paulo, Brazil. *Eur J Clin Nutr* **57**, 629–635.
34. Vereecken CA, De Bourdeaudhuij I & Maes L (2010) The HELENA online food frequency questionnaire: reproducibility and comparison with four 24-hour recalls in Belgian-Flemish adolescents. *Eur J Clin Nutr* **64**, 541–548.
35. Watanabe M, Yamaoka K, Yokotsuka M *et al.* (2011) Validity and reproducibility of the FFQ (FFQW82) for dietary assessment in female adolescents. *Public Health Nutr* **14**, 297–305.
36. Watson JF, Collins CE, Sibbritt DW *et al.* (2009) Reproducibility and comparative validity of a food frequency questionnaire for Australian children and adolescents. *Int J Behav Nutr Phys Act* **6**, 62.
37. Sigman-Grant M & Morita J (2003) Defining and interpreting intakes of sugars. *Am J Clin Nutr* **78**, 815–826.
38. Hebert JR & Miller DR (1991) The inappropriateness of conventional use of the correlation coefficient in assessing validity and reliability of dietary assessment methods. *Eur J Epidemiol* **7**, 339–343.
39. Chinn S (1990) The assessment of methods of measurement. *Stat Med* **9**, 351–362.

40. Medical Research Council (n.d.) Dietary assessment – Quantification. http://dapa-toolkit.mrc.ac.uk/dietary-assessment/quantification/index.php (accessed August 2015).

41. Foster E, Matthews JNS, Nelson M et al. (2006) Accuracy of estimates of food portion size using food photographs – the importance of providing age-appropriate tools. *Public Health Nutr* **9**, 509–514.

42. Matthys C, Pynaert I, De Keyzer W et al. (2007) Validity and reproducibility of an adolescent web-based food frequency questionnaire. *J Am Diet Assoc* **107**, 605–610.

43. Moore HJ, Ells LJ, McLure SA et al. (2008) The development and evaluation of a novel computer program to assess previous-day dietary and physical activity behaviours in school children: The Synchronised Nutrition and Activity Program™ (SNAP™). *Br J Nutr* **99**, 1266–1274.

44. Vereecken CA, Covents M, Matthys C et al. (2005) Young adolescents' nutrition assessment on computer (YANA-C). *Eur J Clin Nutr* **59**, 658–667.

45. Vereecken CA, Covents M, Sichert-Hellert W et al. (2008) Development and evaluation of a self-administered computerized 24-h dietary recall method for adolescents in Europe. *Int J Obes* **32**, Suppl. 5, S26–S34.

46. Vereecken CA & Maes L (2006) Comparison of a computer-administered and paper-and-pencil administered questionnaire on health and lifestyle behaviors. *J Adolesc Health* **38**, 426–432.