The use of multiple imputation method for the validation of 24-h food recalls by part-time observation of dietary intake in school

Emil Kupek[1]* and Maria Alice A. de Assis[2]

[1]*Department of Public Health, Center of Health Science, Federal University of Santa Catarina, Campus Universitário – Trindade, Florianópolis, SC, 88040-900, Brazil*
[2]*Center of Health Science, Federal University of Santa Catarina, Campus Universitário – Trindade, Florianópolis, SC, 88040-900, Brazil*

Abstract

External validation of food recall over 24 h in schoolchildren is often restricted to eating events in schools and is based on direct observation as the reference method. The aim of this study was to estimate the dietary intake out of school, and consequently the bias in such research design based on only part-time validated food recall, using multiple imputation (MI) conditioned on the information on child age, sex, BMI, family income, parental education and the school attended. The previous-day, web-based questionnaire WebCAAFE, structured as six meals/snacks and thirty-two foods/beverage, was answered by a sample of 7–11-year-old Brazilian schoolchildren (n 602) from five public schools. Food/beverage intake recalled by children was compared with the records provided by trained observers during school meals. Sensitivity analysis was performed with artificial data emulating those recalled by children on WebCAAFE in order to evaluate the impact of both differential and non-differential bias. Estimated bias was within ±30 % interval for 84·4 % of the thirty-two foods/beverages evaluated in WebCAAFE, and half of the latter reached statistical significance ($P < 0·05$). Rarely (<3 %) consumed dietary items were often under-reported (fish/seafood, vegetable soup, cheese bread, French fries), whereas some of those most frequently reported (meat, bread/biscuits, fruits) showed large overestimation. Compared with the analysis restricted to fully validated data, MI reduced differential bias in sensitivity analysis but the bias still remained large in most cases. MI provided a suitable statistical framework for part-time validation design of dietary intake over six daily eating events.

Key words: Diet: Online dietary assessment tools: Validation: Multiple imputation: Children

The need for surveillance and monitoring of diets of schoolchildren and physical activity has been frequently voiced as an essential aid in preventing child obesity[1]. To this end, web-based food questionnaires are a low-cost means, but their validity needs to be established in order to qualify such instruments for decision making (e.g. school food policy, dietary advice to parents and children). The imperative to validate food questionnaires is challenging given the lack of a usable gold standard and high cost of validation, even only for a sample subset, with a reference method such as 7-d food diaries[2].

The accuracy of the 24-h dietary recall in children has been evaluated through observation of school meals comparing foods recalled with foods either observed as eaten or foods actually weighed[3–9]. These studies have demonstrated varying levels of inaccuracy because of both under-reporting and over-reporting, related to the difficulties of young schoolchildren (<12 years old) in recalling and quantifying the foods consumed[10–12].

For children, the school environment is normally more restrictive regarding food variability, whereas parental control and peer pressure can strongly impact the food choices outside school. Therefore the food eaten at school may not be representative of the 24-h dietary intake, thus highlighting the need to validate this part too.

Food choices are known to be heavily influenced by personal preferences, environmental restrictions and purchasing power[10,13]. In addition, children's age, sex and weight status, as well as parents' income, education and BMI, have been shown to correlate with child-reported dietary recall[10,13]. For each combination of these variables, observed dietary intake provides the benchmark for estimating likely dietary intake for situations where the reference method application is not feasible. By combining observed and estimated values of the reference method for all subjects over all eating events during 24 h, a complete set of reference values can be provided, thus allowing the validation of self-reported dietary intake over the entire previous day.

Abbreviation: MI, multiple imputation.

* Corresponding author: Professor E. Kupek, fax +55 48 3721 9542, email emil.kupek@ufsc.br

Recently, we developed WebCAAFE (Schoolchildren Food Consumption and Physical Activity), an online, previous-day questionnaire aimed at Brazilian children attending second to fifth grades of elementary school. Usability tests showed very good acceptability and child capacity to understand and respond to WebCAAFE[14]. WebCAAFE dietary items have been extensively tested for internal[15] and external validity[4,5,9] in Brazil with very good results. In a subsequent study, the viability of WebCAAFE for decision making at school level was evaluated and proved to be a valid questionnaire for screening compliance with dietary recommendations for medium and sometimes even small groups of children (e.g. in a classroom)[16]. Although these studies tackled important methodological issues, only food consumption in school was validated.

In this methodological study, we applied the multiple imputation (MI) technique to generate reference values for the whole day. It is possible to estimate likely values of the reference method when it is missing for some subjects and/or on some occasions (e.g. out-of-school meals). The association between self-reported food intake at school and a reference method such as dietary intake observed in school, as well as with other predictors (sex, school grade, family income, parent's educational level) established in the previous research[9], allows the validation of self-reported dietary intake over the six meals from the previous day.

The objectives of this study were to estimate unobserved components of daily food intake such as out-of-school dietary intake, to sum them with the results of observed food intake in school and compare this sum with self-reported, 24-h dietary intake over six daily meals/snacks obtained by the WebCAAFE questionnaire. Such comparison provides an estimate of the WebCAAFE bias over six daily eating events, which is the focus of this study. We do not attempt substantial discussion of the results regarding specific dietary items.

Methods

Sampling and instruments

An intentional sample of five public schools with children attending second to fifth grades (7–11-year-olds) was selected in the city of Florianopolis, southern Brazil. School selection was based on the need to include low, intermediate and high level of information technology resources within the school, such as the number and the quality of personal computers and Internet connection at child disposal, as well as to cover the geographical regions of the municipality (north, south, east, west and central); six classes were randomly selected from the alphabetic list of the eligible classes within each school grade (a, b, c, d, etc.) by systematically choosing every fifth from the list without replacement. All students within the selected classes were invited to take part in the study (n 778). Both child and parental consent was obtained for 708 children. Of these, 106 children were excluded because they were either absent from school during the observation of the meals and/or when WebCAAFE was applied on the next day. The final sample included 602 children. The school schedule in these public schools was either in the morning (08.00–12.00 hours) or in the

afternoon (13.00–17.00 hours), with 45 min of teaching classes and 15 min of break between the third and the fourth class.

WebCAAFE is a self-reported, online food and physical activity questionnaire regarding the previous-day (24 h) recall[9,14]. Of the thirty-two dietary items whose images appear on the computer screen for each of the six daily eating events (breakfast, mid-morning snack, lunch, afternoon snack, dinner and evening snack), those selected by clicking or dragging were counted as consumed. A robot-like avatar guides children responding to the questionnaire. A print screen of the questionnaire, including its food images, is available on http://www.caafe.ufsc.br/public/uploads_midias/1381079027.pdf. Validity tests of the food consumption section, using direct observation at school meals as the reference method, showed 43 % of matches, 29 % of intrusions and 28 % of omissions[9], placing this questionnaire's accuracy close to that of other similar instruments[3,8].

Direct observation of dietary intake in school was carried out by trained observers who checked the consumed items on a previously prepared protocol sheet. During the training, at least thirty children were observed under supervision, and the average agreement between the trainees and their supervisors was 96 %[9]. Each observer was assigned at most five children to monitor in school on the day before WebCAAFE was applied. A more detailed description of research procedures has been published elsewhere[9].

Before initiating data collection with schoolchildren, their parents completed a self-administered questionnaire reporting their household income and the highest educational level achieved, along with a signed consent form. Both household income and mother's education were categorised according to the Brazilian Statistical Office classification. In addition, monetary units and the years of school completed were also provided for these categories in order to facilitate their comparison with other countries' classifications of income and educational level.

The school management provided information on child age and sex.

Anthropometric measurements of children were performed in each school by a trained physical education teacher at most 15 d before the direct observation and the application of WebCAAFE. Weight and height were measured with the children wearing light clothes and barefoot, following standard techniques[17]. A digital, solar, 180-kg scale (Marte®, model PP; Marte Scale and Precision Equipment) was used to measure weight, whereas height was measured using a metal stadiometer (Seca). BMI was computed as weight (kg) divided by height squared (m^2) and categorised into quintiles.

Statistical analyses

In all, two sources of missing data were identified: subject and item non-response. The former occurred when a child was absent from school either on the day of the direct observation of school meals or on the following day when WebCAAFE was applied. As no plausible relationship between missing from school on 1 or 2 consecutive days and usual diet could be conceived in apparently healthy children, a missing- completely- at- random mechanism[18] was assumed for this missing pattern. Consequently, no bias should be brought about by

excluding the children following this pattern, although the variance estimates are reduced in this way.

As for item non-response, it should be kept in mind that breakfast, dinner and evening snack were never observed, and thus these meals were missing by design in the observation data. Among 602 children who responded to WebCAAFE and were observed during at least one school meal, 48% were observed both during morning snack and lunch, 35·3% were observed only during the mid-morning snack and 16·6% were observed only during the afternoon snack. No missing data existed in the food section of WebCAAFE as the dietary items not selected by children were considered not consumed. MI covariates included child characteristics (age, sex, BMI), family background (income, mother's highest educational level), meal/snack type (breakfast, mid-morning snack, lunch, afternoon snack, dinner and evening snack) and the school attended. The missing values for these variables were relatively rare (<10%) and treated as special response categories for MI, and thus these were never missing from the MI perspective.

WebCAAFE food recalls and observed dietary intake in school measured the same behaviour by two different methods and with different accuracy. Therefore, both sources of the data for dietary intake over each of the six meals/snacks on the previous day were assumed to belong to the same population, including those missing by design (i.e. unobserved out-of-school dietary intake). For each unobserved dietary intake, a missing value was imputed to provide an estimate of the reference method, that is, whether a dietary intake of the questionnaire items would have been observed if the researchers had the opportunity to use this method for all children over all meals/snacks of the day before WebCAAFE was applied.

MI covariates were chosen on both theoretical and empirical grounds as important predictors of food consumption as established by previous studies[9,16], such as child sex, age, BMI and the school attended, as well as family income and educational level. Missing-at-random (MAR) mechanism[19] was assumed to estimate the unobserved components of daily food intake. In other words, the data available (WebCAAFE food recalls, partially observed dietary intake in school and selected covariates) should provide unbiased estimates for the missing data on both in-school and out-of-school food intakes.

Sensitivity analysis was performed to check the robustness of MI results under differential and non-differential bias by generating the following missing-not-at-random (MNAR)[19] scenarios: self-reported dietary intake was underestimated by 20, 30 and 50% for all children, as well as 20% for those <9 years of age, 30% for boys and 70% among girls, and overestimated by 50%. Most of the scenarios represent underestimation in varying levels of rarely consumed (<1, 1–2 and 3–4%) foods as a well-established problem in nutritional epidemiology[20,21]. True values of these foods and MI covariates were assumed to be equal to those reported by WebCAAFE and then modified in two steps: (a) the food items were biased according to the aforementioned MNAR scenarios, and (b) missing values were substituted for the same subjects and meals as in the real WebCAAFE data. Both modifications were concatenated in a single data set with two records per subject: one for the biased WebCAAFE report and the other for partially observed dietary intake during school meals as in the main MI analysis.

For both main MI and sensitivity analyses, the imputation was performed by predictive mean matching[19] with five nearest neighbours and repeated thirty times following the recommendations to achieve a suitable trade-off between bias and variance of the MI estimates[22]. A fully conditional model[23,24] was estimated by so-called 'chained equations' in Stata software version 12.0[25,26]. After both MI analyses, bivariate probit regression with MI estimates as dependent and WebCAAFE self-reports as independent variables was used to estimate the reporting bias for each food. The regression used so-called Rubin's rules to account for variability in MI estimates and robust ('sandwich') estimators to account for within-subject clustering. Convergence of MI iterations was verified graphically by trace plots[27] for five separate chains with ten burn-in iterations each.

The statistical software package Stata, version 12.1, was used for all calculations. Stata code for MI is available from the authors upon request.

This study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the Federal University of Santa Catarina Human Research Ethics Committee (protocol 2250/11). Oral consent was obtained from participating children, and written informed consent was given by their parents and educators.

Results

Sample characteristics showed about 6% more girls than boys, as well as similar percentages of children across age bands of primary interest (6·5–11·5 years) and about 3% for each extreme of age distribution (Table 1). Almost half of the children lived in families with annual income of up to 7·236 US dollars, and almost 30% of the parents attended only elementary school.

With about 52% of the children observed during one and 48% during two school meals, no validation by direct observation was available for the remaining meals, resulting in a large proportion (77–78%) of unobserved meals (Table 2) with non-monotone missing data patterns.

No apparent trends were found in trace plots for five separate chains, thus suggesting convergence of the predictive mean matching algorithm.

Mean frequency of the 24-h dietary intake ranged between 0·034 (nuggets) and 1·062 (rice) for WebCAAFE reports (Table 2). As direct observation of school meals covered only a part of daily food intake, direct comparison with WebCAAFE mean cannot be made. However, MI estimates for unobserved food intake can be summed with observed food intake in school (the column 'MI 24 h' in Table 2) and compared with self-reported food consumption over 24 h. This comparison is the WebCAAFE bias estimate (the mean under 'Percentage difference' heading in Table 2) whose magnitude ranged from 3·5% for instant pasta to 46% for green leaves. Among thirty-two foods presented on the WebCAAFE computer screen, twenty-seven (84·4%) had the bias within ±30% interval. Half of the foods showed statistically significant bias ($P < 0.05$).

Table 1. Sample (*n* 602) characteristics used as covariates for multiple imputation

Characteristics	%
Child sex*	
Boys	46·3
Girls	53·6
School grade (mean age)*	
2nd (7·8 years)	18·7
3rd (8·8 years)	23·2
4th (9·9 years)	31·5
5th (10·9 years)	26·4
Child BMI (quintiles)†	
1st (mean 14·6 kg/m^2)	19·7
2nd (mean 15·9 kg/m^2)	19·2
3rd (mean 17·2 kg/m^2)	19·9
4th (mean 19·2 kg/m^2)	19·1
5th (mean 23·9 kg/m^2)	18·7
Household income in US dollars per year‡	
Very low (<3·619)	7·9
Low (3·620–7·236)	41·5
Intermediate (7·237–10·856)	18·2
High (10·857–18·056)	7·5
Very high (18·057–28·947)	3·8
Highest educational level (years of school completed) among parents‡	
Elementary school not completed (<8 years)	19·1
Elementary school completed (8 years)	10·1
High School not completed (9–10 years)	12·1
High School completed (11 years)	22·9
College not completed (12–15 years)	6·6
College completed (>15 years)	8·6

* Informed by school administration.
† Anthropometric measurements by trained researchers.
‡ Informed by parents; income and education levels follow the Brazilian Statistical Office classification.

The largest WebCAAFE over-reporting was estimated for green leaves (46 %), fruits (40 %), manioc flour (33 %), sweets (33 %) and vegetables (30 %), whereas the largest under-reporting was found for fish/seafood (−28 %), vegetable soup (−27 %) and cheese bread (−25 %). The intake of rarely consumed foods tended to be under-reported by WebCAAFE (e.g. fish/seafood, vegetable soup, cheese bread, French fries, nuggets), and the largest over-reporting was found for most frequently consumed foods such as bread/biscuits, meat and fruits (Fig. 1).

Simulated data showed large biases under MNAR models for both complete case and MI analyses, particularly in the case of differential biases regarding the effects of sex and age on the outcomes (Table 3). However, the bias magnitude of the MI estimates was clearly lower compared with complete case analysis in seven out of eight models tested. Although the 95 % CI of this bias contained true bias in five of eight models with MI compared with only two of eight with complete case analysis, the bias magnitude was still large.

Discussion

The present study showed the viability of the MI method in estimating dietary intake bias in a simplified web-based questionnaire for children when only a subsample was validated by direct observation of the school meals, thus missing information on the children's diet outside the school. In nutritional epidemiology, this is a novel application of a well-established statistical method whose substantial results are discussed below.

Intakes of fruits, vegetables and green leaves may be overestimated because of social desirability of these items as part of healthy diet recommendations[28,29], reiterated by school teachers and some parents. Manioc flour often accompanies meat in a traditional Brazilian meal and the same goes for maize and potatoes in the geographical region analysed. Frequently consumed foods are easier to recall compared with those rarely consumed, which require higher amount of searching through memory to mark dietary items consumed on the previous day[29,30]. Another facilitator of food recall may be its preference, so that highly valued foods (e.g. sweets) are easier to recall[28,29]. In Brazil, children often consume sweets and biscuits and bread with milk (e.g. instant chocolate drinks). Preference for these items and the fact that they are often consumed together may have also led to their overestimation as suggested by MI.

More recent events are generally easier to recall, and thus a more recent meal is more accurately reported than the one with longer retention interval[7,12]. About three times higher proportion of the WebCAAFE reports validated by direct observation in school for the afternoon snack compared with the morning snack[9] was in line with the aforementioned rule. Consequently, dietary items with longer retention interval may be underestimated both in terms of mean and variance because the two are equal in a Poisson model for counts. This gives rise to heteroscedastic memory error[31] observed for similar tasks that require searching through past events[32]. Episodic or rarely consumed foods would also be underestimated according to the same mechanism.

Cognitive difficulties in identifying components of mixed foods have been cited to cause under-reporting[10,11,29]. In addition, recognition of food images on a computer screen may be affected by specific brand and packaging a child is used to, so that less-specific images of dietary items may be difficult to identify correctly[10,28]. These factors may combine with lesser accuracy of reporting the meals with longer retention interval – for example, a child may omit breakfast cereal in the WebCAAFE report because it did not correspond to a specific brand/packaging he or she ate the day before.

Among many imputation techniques, predictive mean matching was chosen for its robustness when non-linear relationships were suspected[33]. The latter are plausible for food consumption items as personal preferences and environmental restrictions exert strong influences and impose food acceptance and availability thresholds – for example, by not eating strongly disliked foods or those not available in school, thus leading to skewed distributions. Predictive mean matching combines linear regression and the nearest-neighbour technique to find a subset of most likely values from which to perform repeated random draws within the range of regression-predicted values. Although extreme values and/or rare dietary patterns can be of interest for child health, the main purpose of the imputation here was to verify the validity of WebCAAFE reports for major population groups (e.g. boys/girls, age groups) as represented by their mean frequency of dietary intake over 24 h.

Table 2. Mean frequency of dietary intake obtained by WebCAAFE by direct observation and estimated by multiple imputation (MI)
(Mean values and standard deviations; upper, lower 95 % confidence intervals)

| | WebCAAFE (24 h) | | Direct observation of school meals | | | MI (24 h) | Percentage difference of WebCAAFE v. MI | | | |
| | | | | | | | | 95 % CI | | |
Foods	Mean	SD	Mean frequency (per meal)	SD	% Missing	Mean	Mean	Lower	Upper	P
Rice	1·060	0·834	0·187	0·390	78·06	1·170	−9·0	−20·0	1·8	0·102
Vegetables	0·255	0·536	0·156	0·371	77·50	0·196	30·0	18·0	41·0	0·001
Green leaves	0·167	0·403	0·136	0·342	78·03	0·114	46·0	31·0	61·0	0·001
Vegetable soup	0·182	0·427	0·008	0·081	78·03	0·249	−27·0	−55·0	1·4	0·062
Cooked beans	0·707	0·800	0·126	0·331	78·03	0·822	−14·0	−26·0	−2·5	0·018
Manioc flour	0·174	0·458	0·088	0·281	78·03	0·131	33·0	20·0	46·0	0·001
Maize, potatoes	0·091	0·318	0·020	0·134	78·03	0·077	19·0	−5·8	44·0	0·130
Pasta	0·413	0·604	0·162	0·368	78·03	0·353	17·0	7·8	27·0	0·001
Instant pasta	0·109	0·340	0·008	0·081	78·03	0·105	3·5	−20·0	28·0	0·770
French fries	0·180	0·484	0·001	0·000	78·03	0·222	−19·0	−39·0	0·4	0·054
Meat	0·880	0·824	0·329	0·470	78·03	0·715	23·0	13·0	33·0	0·001
Sausages	0·194	0·479	0·001	0·000	78·03	0·246	−21·0	−41·0	−1·0	0·040
Eggs	0·111	0·332	0·021	0·140	78·03	0·101	10·0	−12·0	33·0	0·368
Fish, seafood	0·142	0·389	0·003	0·041	78·03	0·197	−28·0	−51·0	−5·1	0·017
Fruits	0·383	0·677	0·310	0·560	78·03	0·274	40·0	30·0	51·0	0·001
Bread, biscuits	0·923	0·848	0·257	0·437	78·03	0·763	21·0	11·0	31·0	0·001
Cheese bread	0·086	0·300	0·005	0·058	78·03	0·115	−25·0	−49·0	−1·2	0·040
Cream cookies	0·446	0·668	0·161	0·366	78·03	0·409	9·0	−2·9	21·0	0·136
Breakfast cereal	0·089	0·299	0·001	0·000	78·03	0·116	−23·0	−49·0	4·1	0·095
Porridge	0·072	0·287	0·020	0·134	77·50	0·066	9·7	−14·0	33·0	0·410
Cheese	0·074	0·295	0·026	0·156	77·50	0·063	17·0	−3·3	36·0	0·101
Coffee and milk	0·481	0·770	0·053	0·221	77·50	0·394	22·0	10·0	34·0	0·001
Milk	0·172	0·449	0·003	0·041	78·03	0·145	19·0	2·1	36·0	0·028
Yogurt	0·295	0·602	0·041	0·196	78·01	0·369	−20·0	−40·0	−0·5	0·045
Chocolate milk	0·420	0·678	0·086	0·279	78·03	0·512	−18·0	−32·0	−5·0	0·008
Fresh fruit juice	0·476	0·730	0·186	0·388	78·03	0·454	4·8	−7·8	17·0	0·449
Sodas	0·302	0·618	0·003	0·041	78·01	0·336	−10·0	−28·0	7·7	0·260
Sweets	0·154	0·382	0·074	0·260	78·03	0·117	32·0	15·0	49·0	0·001
Fast food	0·091	0·302	0·043	0·200	78·03	0·080	14·0	−6·6	35·0	0·180
Fried snacks	0·292	0·566	0·083	0·274	78·03	0·298	−1·9	−23·0	19·0	0·858
Nuggets	0·034	0·176	0·001	0·000	78·03	0·039	−13·0	−45·0	20·0	0·439
Chocolate cake	0·270	0·533	0·073	0·283	77·50	0·241	12·0	−6·1	29·0	0·195
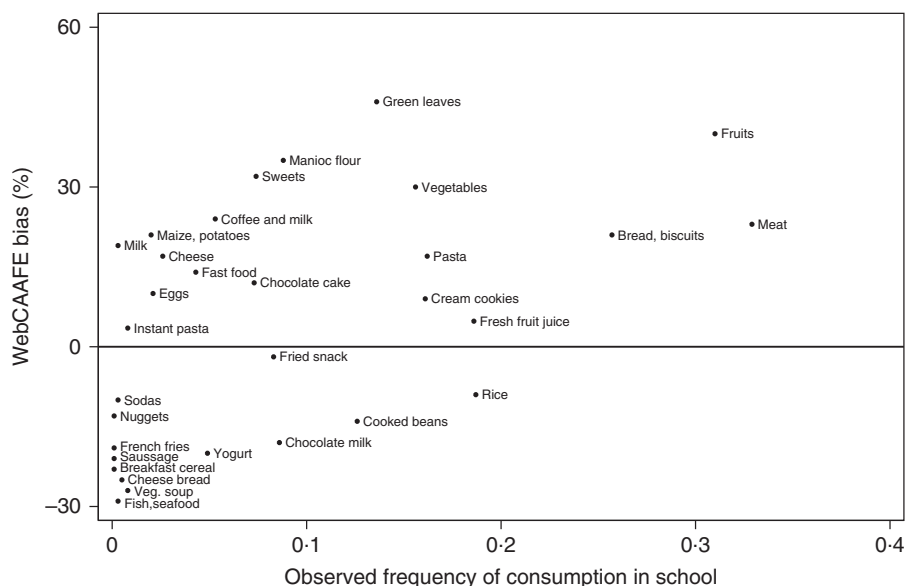
Fig. 1. WebCAAFE bias for 24-h dietary intake v. mean frequency of dietary consumption estimated by multiple imputation.

Table 3. Simulated data comparison of WebCAAFE bias in complete case v. multiple imputation analysis under various missing-not-at-random scenarios of systematic self-reporting error
(Bias estimates and 95 % confidence intervals)

Simulated data (n 660 meals)			Complete case analysis			MI analysis		
True MFDI*	True bias in WebCAAFE	% Missing	Bias estimate*	95 % CI	P	Bias estimate*	95 % CI	P
24·4	+50 %	39·03	93·0	65·8, 126·0	<0·001	10·4	0·1, 21·8	0·049
1·2	−30 %	39·01	−73·6	−80·1, −60·0	<0·001	−45·7	−58·3, −9·3	<0·001
1·6	−50 %	39·52	172·5	15·6, 542·3	0·022	17·1	−13·8, 58·9	0·311
2·0	−20 %	39·01	480·0	90·8, 1662·8	0·002	34·8	−7·8, 97·1	0·123
2·2	−70 % girls	39·03	−91·1	−44·7, 137·9	0·712	−5·1	−40·6, 51·6	0·825
2·9	−30 % boys		−57·1	−73·6, −30·9	0·001	−53·8	−65·2, −38·7	0·207
2·6	−20 % if age 7–8 years	39·04	14·7	−44·7, 137·9	0·712	−23·7	−48·2, 12·4	0·171
3·7	None if age 9–12 years		165·7	19·9, 1023·4	0·024	−77·8	−84·9, −7·29	<0·001

MFDI, mean frequency of 24-h dietary intake; MI, multiple imputation by predictive mean matching.
* Multiplied by 100.

Validity of the MI estimates largely depends on the plausibility of the MAR model assumed – that is, to what extent the covariates predict unbiased estimates of food intake in and out of school. In school, the food environment was more restricted than out of school, and a representative sample of children was observed to validate their reports by WebCAAFE. Outside school, however, no such validation was feasible and the food options were likely more diversified. The latter are enhanced by higher child age and family income and education, all of which were available for analysis in this study. Nevertheless, some other factors that influence child food intake at home were not available for analysis, such as parental control over selection and timing of dietary intake based on their health beliefs[34]. The impact of unmeasured predictors on MI bias diminishes as the strength of their associations with the measured predictors increases. Therefore, a fairly comprehensive coverage of the latter provides some reassurance against MI bias.

Among the predictors of unobserved food intake, the role of meal type (breakfast, morning snack, lunch, afternoon snack,

dinner and evening snack) was prominent as food choices are strongly meal dependent. For example, eating rice and beans is part of a typical Brazilian lunch but not of breakfast or snacks between main meals. Therefore, conditioning MI estimates on meal type enhances the chances of filling in the food choices observed in a sample of children during school hours, whereas other covariates had a larger influence in determining the estimates for unobserved meals (breakfast, dinner and evening snack) and were likely to produce more variable estimates. From a Bayesian perspective, conditioning on meal type represents common knowledge of its large impact on food choices while considering individuals 'exchangeable' (equivalent to random effects in mixed models).

Strengths of this study include the use of well-established and tested statistical principles used in MI with non-monotone missing data patterns and subsequent analysis based on their estimates. In addition, thirty repetitions of MI should provide good account of the MI estimates' precision[24,33,35], which in turn implies more realistic estimates of between-subject variability. In addition, a comprehensive coverage of

socio-demographic characteristics, widely recognised for their association with food consumption in the literature and used in MI, provides solid empirical and theoretical grounds for reducing bias in the estimates obtained[2,23]. Finally, previously established coherence between self-reported and observed dietary intakes in schoolchildren[4,5,9] enhances the plausibility of the MI model used in this study.

Study limitations include a large percentage of missing values in observed data and uncertainty regarding the assumed MAR model. Despite the above-mentioned confidence in the relevance of MI covariates, there is no guarantee that they can fully describe the missing data mechanism or the functional form of the causal influence, assumed to be a Poisson process in this case. Sensitivity analysis showed that MI reduced WebCAAFE reporting bias under most MNAR scenarios considered but this may not be good enough for some purposes (e.g. individual classification into healthy v. unhealthy diet category). Reporting bias was quite large for some foods and requires further investigation.

Although biomarkers can be used as the gold standard for total energy expenditure[36] and a few micronutrients, their use in population surveys is prohibitively expensive, and therefore food records remain the best available methods to measure population dietary intake[37]. The latter include 24-h dietary intake recall and food diaries, preferably covering all days of the week. However, food records require highly motivated participants, and thus may induce a self-selection bias that is difficult to account for, especially in primary school children. On the other hand, direct observation of child dietary intake in school has the advantage of random assignment of the participants and has been successfully applied in this age group[4,5,9]. The WebCAAFE food questionnaire was constructed on 7–11-year-old children in Brazil and validated by direct observation in school in several studies[4,5,9] without provoking significant child reactivity. Moreover, the very high agreement between trained observers of child dietary intake in school and their supervisors[9] justifies the use of direct observation as the reference method. All these elements suggest that the sample of children observed in the present study was representative of typical dietary intake in school and that suitable screening and reference methods were applied.

Despite the confidence placed on the reference method, its coverage was not only partial (i.e. applied only to a sample of children who responded WebCAAFE) but also absent for all children out of school. Coupled with already-mentioned differences between eating at school and at home, the imputed values for the reference standard must be considered with due caution, especially in the light of the sensitivity analysis results (Table 3), which pointed to a wide range of bias estimates when MAR assumption was violated. Another sensitivity analysis with food records also showed high dependency of effect estimates on model assumptions regarding the covariance structure of measurement errors[2]. In addition, although MI bias reduction in comparison with complete case analysis has been corroborated in other studies[22,23], the bias magnitude was still large in the present study. Finally, the large number of variables with considerable percentage of missing values makes MI estimates vulnerable to bias[38].

To the authors' knowledge, this is the first external validation of a previous-day food recall in schoolchildren for the whole of that period, thus allowing an estimation of the food questionnaire bias on per-day basis. The extrapolation of dietary intake based on partially observed school meals to all meals was based on MI. The latter has been scarcely applied in nutritional epidemiology, despite its value being pointed out in a recent review of statistical methods in this area[39]. MI is suitable of addressing both non-differential (random) and differential (systematic) measurement errors in exposure and their effect on the outcome of interest. Chained equations were indicated for estimation when a validation sample was available as in the present study, and sensitivity analysis was recommended to assess the impact of systematic errors. However, so far, most of the MI applications in nutritional epidemiology have focused on correcting the effect of partially measured exposure on outcome in case–control studies with non-differential measurement error[2,37,39]. A different MI model was used in the present study, which imputed partially observed outcomes given a set of covariates, thus providing a novel approach to the difficult problem of external validation of dietary intake in the absence of both a gold standard method and complete coverage of the period within which it occurred (e.g. out-of-school period).

To gauge the scope of WebCAAFE in nutritional epidemiology, the difference between 24-h recall methods and short-form FFQ should be considered. WebCAAFE was designed as a previous-day recall ('what did you eat yesterday?') of the frequency of the markers of (un)healthy diet as opposed to their quantity (weight) employed in the 24-h recall method. WebCAAFE is a structured short-form FFQ with six daily eating events and thirty-two food and beverage items, whereas the 24-h recall method asks about all types and quantities of food and beverages consumed in the last 24 h, often applying the multiple-pass method[8]. Short-form FFQ for children avoids the difficulties associated with the assessments of portion size and simplifies the memory task by prompting only the most relevant diet markers of the previous day[10–12,29]. In line with FFQ of this kind validated in other countries[40,41], WebCAAFE keeps the questionnaire relatively brief and easy for children to self-complete, with minimal assistance in the school setting.

From a nutritional surveillance perspective, WebCAAFE can be used to repeatedly measure a large cohort of schoolchildren and validate some samples by direct observation of school meals or by food diaries at an acceptable cost. MI is certainly less expensive than field data collection[39] and cost-efficient for estimating exposure–disease association, especially when a surrogate of the exposure is available from the full cohort data[37]. The lack of accuracy in online food questionnaires may be substantially reduced by their partial validation, thus providing a viable means for monitoring nutritional development of schoolchildren.

Acknowledgements

References

1. Pan American Heath Organization (2014) *Plan of Action for the Prevention of Obesity in Children and Adolescents.* Washington, DC: Pan American Heath Organization.

2. Keogh RH, White IR & Rodwell SA (2013) Using surrogate biomarkers to improve measurement error models in nutritional epidemiology. *Stat Med* **32**, 3838–3861.

3. Baranowski T, Islam N, Baranowski J, *et al.* (2002) The food intake recording software system is valid among fourth-grade children. *J Am Diet Assoc* **102**, 380–385.

4. de Assis MAA, Kupek E, Guimarães D, *et al.* (2008) Test-retest reliability and external validity of the previous day food questionnaire for 7-to 10-year-old school children. *Appetite* **51**, 187–193.

5. Assis MA, Benedet J, Kerpel R, *et al.* (2009) Validation of the third version of the previous day food questionnaire (PDFQ-3) for 6-to-11-years-old schoolchildren. *Cad Saude Publica* **25**, 1816–1826.

6. Baxter SD, Hardin JW, Smith AF, *et al.* (2009) Twenty-four hour dietary recalls by fourth-grade children were not influenced by observations of school meals. *J Clin Epidemiol* **62**, 878–885.

7. Medin AC, Astrup H, Kåsin BM, *et al.* (2015) Evaluation of a web-based food record for children using direct unobtrusive lunch observations: a validation study. *J Med Internet Res* **17**, e273.

8. Diep CS, Hingle M, Chen TA, *et al.* (2015) The automated self-administered 24-hour dietary recall for children, 2012 version, for youth aged 9 to 11 years: a validation study. *J Acad Nutr Diet* **115**, 1591–1598.

9. Davies VF, Kupek E, de Assis MA, *et al.* (2015) Validation of a web-based questionnaire to assess the dietary intake of Brazilian children aged 7–10 years. *J Hum Nutr Diet* **28**, Suppl. 1, 93–102.

10. Livingstone MB, Robson PJ & Wallace JM (2004) Issues in dietary intake assessment of children and adolescents. *Br J Nutr* **92**, S213–S222.

11. Baxter SD (2009) Cognitive processes in children's dietary recalls: insight from methodological studies. *Eur J Clin Nutr* **63**, S19–S32.

12. Baxter SD, Hitchcock DB, Guinn CH, *et al.* (2014) A validation study concerning the effects of interview content, retention interval, and grade on children's recall accuracy for dietary intake and/or physical activity. *J Acad Nutr Diet* **114**, 1902–1914.

13. Cullen K, Baranowski T, Watson K, *et al.* (2007) Food category purchases vary by household education and race/ethnicity: results from grocery receipts. *J Am Diet Assoc* **107**, 1747–1752.

14. da Costa FF, Schmoelz CP, Davies VF, *et al.* (2013) Assessment of diet and physical activity of Brazilian schoolchildren: usability testing of a web-based questionnaire. *JMIR Res Protoc* **2**, e31.

15. Davies VF, Kupek E, de Assis MA, *et al.* (2015) Qualitative analysis of the contributions of nutritionists to the development of an online instrument for monitoring the food intake of schoolchildren. *J Hum Nutr Diet* **28**, Suppl. 1, 65–72.

16. Kupek E, de Assis MA, Bellisle F, *et al.* (2016) Validity of WebCAAFE questionnaire for assessment of schoolchildren's dietary compliance with Brazilian Food Guidelines. *Public Health Nutr* (Epublication ahead of print version 6 April 2016).

17. Lohman TG, Roche AF & Martorell R (1988) *Anthropometric Standardization Reference Manual.* Champaign, IL: Human Kinetics.

18. Little RJA (1988) Missing-data adjustments in large surveys. *J Bus Econ Stat* **6**, 287–296.

19. Keogh RH & White IR (2011) Allowing for never and episodic consumers when correcting for error in food record measurements of dietary intake. *Biostatistics* **12**, 624–636.

20. Pérez A, Zhang S, Kipnis V, *et al.* (2012) Intake_epis_food(): an R Function for Fitting a bivariate nonlinear measurement error model to estimate usual and energy intake for episodically consumed foods. *J Stat Softw* **46**, 1–17.

21. Schenkel N & Taylor JMG (1996) Partially parametric technique for multiple imputation. *Comput Stat Data Anal* **22**, 425–446.

22. Lee KJ & Carlin JB (2010) Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol* **171**, 624–632.

23. Seaman SR & Keogh RH (2015) Handling missing data in matched case-control studies using multiple imputation. *Biometrics* **71**, 1150–1159.

24. van Buuren S (2007) Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* **16**, 219–242.

25. StataCorp (2011) *Stata Multiple-Imputation Reference Manual: Release 12.* College Station, TX: Stata Press.

26. Abayomi K, Gelman A & Levy M (2008) Diagnostics for multivariate imputations. *J R Stat Soc Ser C Appl Stat* **57**, 273–291.

27. Latif H, Watson K, Nguyen N, *et al.* (2011) Effects of goal setting on dietary and physical activity changes in the Boy Scout badge projects. *Health Educ Behav* **38**, 521–529.

28. Sharman SJ, Skouteris H, Powell MB, *et al.* (2016) Factors related to the accuracy of self-reported dietary intake of children aged 6 to 12 years elicited with interviews: a systematic review. *J Acad Nutr Diet* **116**, 76–114.

29. Kolodziejczyk JK, Merchant G & Norman GJ (2012) Reliability and validity of child/adolescent food frequency questionnaires that assess foods and/or food groups. *J Pediatr Gastroenterol Nutr* **55**, 4–13.

30. Halford JC, Boyland EJ, Cooper GD, *et al.* (2008) Children's food preferences: effects of weight status, food type, branding and television food advertisements (commercials). *Int J Pediatr Obes* **3**, 31–38.

31. Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl Cognit Psychol* **5**, 213–236.

32. Kupek E (1999) Estimation of the number of sexual partners for the nonrespondents to a large national survey. *Arch Sex Behav* **28**, 233–242.

33. Royston P & White IR (2011) Multiple imputation by chained equations (MICE): implementation in Stata. *J Stat Softw* **45**, 1–20.

34. Couch SC, Glanz K, Zhou C, *et al.* (2014) Home food environment in relation to children's diet quality and weight status. *J Acad Nutr Diet* **114**, 1569–1579.

35. Kenward MG & Carpenter JR (2007) Multiple imputation: current perspectives. *Stat Methods Med Res* **16**, 199–218.
36. Burrows TL, Martin RJ & Collins CE (2010) A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc* **110**, 1501–1510.
37. Keogh RH & White IR (2014) A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Stat Med* **33**, 2137–2155.
38. Sterne JA, White IR, Carlin JB, *et al.* (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338**, b2393.
39. Keogh RH, Park JY, White IR, *et al.* (2012) Estimating the alcohol-breast cancer association: a comparison of diet diaries, FFQs and combined measurements. *Eur J Epidemiol* **27**, 547–559.
40. 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.
41. Magarey A, Golley R, Spurrier N, *et al.* (2009) Reliability and validity of the Children's Dietary Questionnaire: a new tool to measure children's dietary patterns. *Int J Pediatr Obes* **4**, 257–265.