British Journal of Nutrition (2008), **100**, 1060–1068
© The Authors 2008
doi:10.1017/S0007114508971300

British Journal of Nutrition

Under-reporting of energy intake is more prevalent in a healthy dietary pattern cluster

Fernanda B. Scagliusi[1]*, Eduardo Ferriolli[2], Karina Pfrimer[2], Cibele Laureano[1], Caroline S. F. Cunha[1], Bruno Gualano[1], Bárbara Lourenço[3] and Antonio H. Lancha Jr[1]

[1]*Laboratory of Nutrition and Applied Metabolism, Department of Biodynamics, School of Physical Education and Sport, University of São Paulo, Av. Prof. Mello Moraes 65, São Paulo, SP, Brazil*
[2]*Division of General Internal and Geriatric Medicine, Department of Internal Medicine, School of Medicine of Ribeirão Preto, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP, Brazil*
[3]*School of Public Health, University of São Paulo, Av. Dr Arnaldo 715, São Paulo, SP, Brazil*

(Received 16 October 2007 – Revised 26 February 2008 – Accepted 28 February 2008 – First published online 1 April 2008)

The aim of the present study was to determine whether under-reporting rates vary between dietary pattern clusters. Subjects were sixty-five Brazilian women. During 3 weeks, anthropometric data were collected, total energy expenditure (TEE) was determined by the doubly labelled water method and diet was measured. Energy intake (EI) and the daily frequency of consumption per 1000 kJ of twenty-two food groups were obtained from a FFQ. These frequencies were entered into a cluster analysis procedure in order to obtain dietary patterns. Under-reporters were defined as those who did not lose more than 1 kg of body weight during the study and presented EI:TEE less than 0·82. Three dietary pattern clusters were identified and named according to their most recurrent food groups: sweet foods (SW), starchy foods (ST) and healthy (H). Subjects from the healthy cluster had the lowest mean EI:TEE (SW = 0·86, ST = 0·71 and H = 0·58; $P=0·003$) and EI − TEE (SW = −0·49 MJ, ST = −3·20 MJ and H = −5·08 MJ; $P=0·008$). The proportion of under-reporters was 45·2 (95 % CI 35·5, 55·0) % in the SW cluster; 58·3 (95 % CI 48·6, 68·0) % in the ST cluster and 70·0 (95 % CI 61·0, 79) % in the H cluster ($P=0·34$). Thus, in Brazilian women, under-reporting of EI is not uniformly distributed among dietary pattern clusters and tends to be more severe among subjects from the healthy cluster. This cluster is more consistent with both dietary guidelines and with what lay individuals usually consider ‘healthy eating’.

Under-reporting: Energy intake: Doubly labelled water: Dietary pattern analysis

Traditionally, in epidemiological studies concerning diet and health, nutritional exposure is assessed as intake of a single or a few nutrients or foods. However, diet is something far more complex and ends up being artificially reduced and simplified when a single nutrient or food is analysed separately. Thus, this approach may have several limitations. First, individuals do not eat isolated foods and nutrients. They eat meals composed of diverse foods which contain a combination of several nutrients. It is likely that there is an interaction and synergy between these foods and nutrients, which are disregarded in the single food or nutrient approach(1). Second, the effect of one nutrient may be confounded by the effect of the eating pattern(2). Third, there is a high correlation between nutrients, since one single food can provide many nutrients(3). Finally, the effect of a single food may be too small and undetectable, but the cumulative effect of various foods may be sufficiently large(4). The dietary pattern analysis can overcome these issues because it takes into account the existing interactions and collinearity between foods and nutrients in the diet. This analysis considers how foods and nutrients are consumed in combination, providing a pattern of habitual intake that more closely resembles the way individuals actually eat(4).

Dietary patterns can be assessed by means of *a priori* techniques (such as a diet-quality index) or by the use of empirical (data-derived) techniques (such as clusters and factorial analysis) applied to self-reports of food intake. There is an increasing interest in determining the validity of dietary pattern analysis, but there is no standard protocol for doing this(5). Most authors are using one of the following procedures to assess validity of the dietary pattern analysis: (a) comparing the patterns based on one dietary assessment method with the nutrient intake obtained through another dietary assessment method(3,6); (b) comparing dietary patterns obtained by different dietary assessment methods(7–9); (c) correlating dietary patterns to concentration biomarkers(7,10,11); (d) determining the risk of developing a disease according to dietary patterns(3,6,12–14). It is a known fact that dietary assessment methods present serious limitations. Recently, the doubly labelled water (DLW) method emerged as the ‘gold standard’ marker of the validity of self-reports of energy intake (EI). This technique clearly showed that all methods tend to largely

---

Abbreviations: DLW, doubly labelled water; EI, energy intake; TEE, total energy expenditure.

* **Corresponding author:** Dr Fernanda B. Scagliusi, fax +55 11 38135921, email fernanda.scagliusi@gmail.com

underestimate EI and to be heavily biased(15–19). Dietary pattern analysis is also based on these methods that tend to produce selective under-reporting of foods and nutrients(20–25). However, little is known about the way under-reporting may affect dietary pattern analysis. Bailey *et al.* conducted a cluster analysis to determine dietary patterns in a sample of 179 elderly individuals(26). Two dietary patterns were obtained; one was nutritionally inadequate, while the other was more balanced and consistent with dietary guidelines. Under-reporting subjects were defined as those who presented EI (assessed by 5 d of 24 h diet recalls) 55 % lower than total energy expenditure (TEE) (estimated by equations). When the under-reporters were not included in the sample, the same dietary patterns were found but the proportion of subjects belonging in the healthier pattern decreased from 40·2 to 33·8 %. When the entire sample was analysed the inadequate pattern was characterised by smaller intake of milk, fish, poultry and some vegetables. When under-reporters were not included in the sample, however, this difference disappeared and the inadequate pattern showed a higher intake of nuts and starchy vegetables. Thus, the presence of under-reporters produced spurious differences and, at the same time, attenuated real differences. It also placed some subjects in the wrong pattern. Nevertheless, this study had some limitations. These were: (a) not using DLW, the gold standard technique for identifying under-reporting; (b) having a very specific sample (community-dwelling elderly, aged 66–87 years), so it is not known if these results are valid for other samples; (c) not presenting under-reporting rates (as proportion of under-reporters, ratio of EI to energy expenditure and difference between these two measurements) for the dietary pattern clusters. Thus, the present study aimed to overcome these limitations by investigating whether under-reporting rates vary between dietary pattern clusters in a sample of adult women. Since under-reporters tend to under-report foods considered ‘unhealthy’ or ‘fattening’, we hypothesised that under-reporting would be more prevalent in the healthier cluster.

Subjects and methods

Subjects

In two universities, students, teachers, workers and patrons were invited by means of flyers to join the research. To participate in the study, subjects completed a questionnaire developed for this research with questions regarding their health, medication use, lifestyle, weight variations and dieting practices. Subjects had to be healthy, weight-stable and literate females with a BMI between 18·5 and 39·9 kg/m^2. Dieters and smokers were excluded as well as those taking diuretics, appetite suppressants, orlistat, thyroid hormone medication and topiramate. We enrolled subjects with a wide range of BMI because one of the aims of the research was to assess influence of BMI on under-reporting (these results are still under analysis and will be published elsewhere). The sample was composed of sixty-five adult females between the ages of 18 and 57 years, living in the city of São Paulo (south-eastern Brazil). The sample size was determined considering Cade *et al.* guidelines for validation studies(27) and calculations made by OpenEpi software (AG Dean, KM Sullivan and MM Soe, Rollins School of Public Health, Emory University, Atlanta, GA, USA; updated June 2007). According to this software, in order to compare mean TEE with mean EI obtained by the FFQ, with 80 % of power and 5 % of significance, thirty-six subjects would be necessary. Data were collected between December 2004 and July 2006.

Participation in the study lasted 3 weeks. The study design is depicted in Table 1. Five subjects (7·7 % of the sample) had their TEE measured by DLW a second time, 1 month apart from the first test. With these data, we obtained a within-subject CV in TEE measured by DLW of 8·8 %.

Ethnicity was self-assigned by subjects based on skin colour. Subjects were asked to classify themselves as black or mulatto, white, or Asian-Brazilian. Height was measured with a stadiometer to the nearest 0·5 cm. Subjects were weighed (in their underwear) with a digital scale to the nearest 0·1 kg. Fat and fat-free mass were estimated by a skinfold measurement protocol (which measured thigh, supra-iliac and sub-scapular skinfolds) developed for Brazilian women(28). Considering that 73·2 % of fat-free mass is water(29), fat-free mass was multiplied by 0·732 in order to determine total body water (which is necessary to calculate the DLW dose).

The School of Physical Education and Sport Ethics Committee approved the study protocol. Subjects were informed about the experimental procedures and of any possible risks before giving their written informed consent to participate. The study was funded by The State of São Paulo Research Foundation (FAPESP).

Total energy expenditure measurement

To determine TEE over the 10 d free-living period, a two-point DLW method was employed. The technique involves the administration of dose water composed of the stables isotopes ^{2}H and ^{18}O. The ^{2}H is eliminated from the body as water and the ^{18}O is eliminated as both water and carbon dioxide. The difference in these elimination rates, after adjusting for total body water and for isotopic fractionation, is a measure of carbon dioxide production rate. More details about this method can be found elsewhere(30). In the morning of day 1, subjects arrived at the laboratory after a 10 h fast and provided a baseline urine sample. DLW was given orally at a dose of 0·12 g 99·8 % ^{2}H-labelled water and 2 g 10 % ^{18}O-labelled

Table 1. Study design: data collection realised between December 2004 and July 2006 with sixty-five Brazilian women

Procedures	Weeks 1 Wednesday	 2 Monday	 3 Wednesday
Anthropometric evaluation (height and body weight)	X		X
Skinfold measurement (thigh, supra-iliac and sub-scapular)	X		
FFQ			X
Doubly labelled water administration		X	
Collection of urine samples		X	X
Collection of blood sample		X	

water per kg estimated total body water, along with a subsequent 50 ml water rinse of the dose bottle. Urine specimens were collected 2, 3, 4 and 5 h after the dose was administered. In the morning of day 10, subjects returned to the laboratory and provided two urine specimens.

Enrichment of ^{2}H and ^{18}O in the specimens was analysed by isotope ratio MS (ANCA 20-20; Europa Scientific, Crewe, Cheshire, UK). The laboratory where the analysis was conducted – Mass Spectrometry Laboratory of the School of Medicine of Ribeirão Preto – was evaluated and approved for total body water and energy expenditure analysis by the International Atomic Energy Agency (Vienna, Austria) in 2007. Tap water was collected and analysed and all calculations were adjusted for the content of isotopes in the drinking water. Isotope dilution spaces were calculated using the baseline specimen and the specimen collected 5 h after dosing(29). Elimination rates of ^{2}H and ^{18}O were calculated using the specimens mentioned above and the last specimen collected on day 10(31). Rate of CO_2 production was calculated according to Schoeller(17), and TEE was determined using the modified Weir equation(32), assuming a respiratory coefficient of 0·85.

Dietary assessment

A FFQ estimated habitual dietary intake within the last month. It was semi-quantitative, self-administered and included seventy-three food items. For each food item, subjects identified serving size (small, medium, or large) and frequency of consumption. Eight frequency options were given, ranging from ‘never or almost never’ to a maximum of ‘more than 3 times per day’. Complete details of this instrument are provided elsewhere(33). In the original validation study of the FFQ, the Pearson correlation coefficient between EI measured by the FFQ and four 24 h diet recalls was 0·44 (P=0·0001)(33).

All answered questionnaires were checked by dietitians, and when an item was not clear the subject was asked to clarify it. The coding and analysis of the FFQ were performed by a trained dietitian. The Brazilian software Virtual Nutri 1.0 (ST Philippi, SC Szarfarc and AR Latterza, Public Health School, University of São Paulo, São Paulo, SP, Brazil) was used to convert portion sizes into weight (g). EI was calculated using Brazilian food composition tables(34,35) and, when necessary, data from the United States Department of Agriculture National Nutrient Database for Standard Reference (release 17) were also used(36).

Food groupings

To examine dietary patterns, the seventy-three individual food items listed on the FFQ were reduced to twenty-two predefined food groups (shown in Table 2), based on the similarity of energy and macronutrient content. The food groups were consistent with the subgroups presented in the Brazilian Food Pyramid(37). For example, the foods belonging to the group ‘pasta, grains and root vegetables’ present a range of 372–481 kJ/100 g and 18·4–29·5 g carbohydrate/100 g(34,35). As conducted by Hu *et al.* (1), we did not group some foods that presented very specific nutrient profiles (such as eggs, soft drinks, sugar, hamburgers and pizza). The variables employed to perform the analysis were daily frequencies of each food group/1000 kJ.

Table 2. Food groups used in the dietary pattern analysis

Food groups	Individuals food items
Bread	Bread (sliced or baguette)
Pasta, grains and root vegetables	Pasta, rice, corn, popcorn, potatoes (not fried), cassava (cooked and as cassava flour) and yam
Beans	Kidney beans, lentils, peas and chick-peas
Fruits	Juices, oranges/tangerines, banana, papaya, apples, watermelon/melon, pineapple, avocado, mango, lemon, passion fruit, grapes, guava and pear
Vegetables	French beans, okra, carrots, eggplant, cauliflower, lettuce, cabbage, kale, endive, tomato, guinea pepper, chayote, pumpkin, summer squash and cucumber
Eggs	Eggs
Regular dairy	Whole milk, whole yogurt, cheese spread and other types of cheese
Fat-reduced dairy	Skimmed or low-fat milk, skimmed or low-fat yogurt, reduced-fat cheese spread and other types of low-fat cheese
Bovine meat	Beef with bone, beef without bone and viscera
Swine meat	Pork, sausages and bacon
Poultry	Chicken
Fish	Fresh fish, tuna/sardine and shrimps
Cookies	Sweet and salty biscuits
Confectionery products/candies	Cakes, ice creams, sweets, chocolate powder, chocolate (bars or pieces) and puddings
Sugar	Sugar
Regular soft drinks	Regular soda (i.e. with sugar)
Diet soft drinks	Diet soda (i.e. with artificial sweeteners)
Spreads	Butter, margarine and mayonnaise
Fat-reduced spreads	Reduced-fat margarine
Fried foods	Chips, crisps and other Brazilian snacks (deep-fried appetisers or snacks with various fillings, such as meat, cheese and chicken)
Pizza	Pizza
Hamburger	Hamburger

Definition of under-reporters of energy intake

Subjects were identified as under-reporters of EI based on the 95 % confidence limits of the expected EI:TEE ratio of 1·0. The 95 % confidence limits were calculated from the published equation(38):

$$\pm 2 \times \sqrt{(\mathrm{CV}_{\mathrm{EI}}^2/\mathrm{d}) + \mathrm{CV}_{\mathrm{TEE}}^2}.$$

$\mathrm{CV}_{\mathrm{EI}}$ is the within-subject CV for EI, which was 23 %. $\mathrm{CV}_{\mathrm{TEE}}$ is the within-subject CV for TEE measured by DLW, which was 8·8 %. The number of days of dietary assessment is expressed by its abbreviation d. Since the FFQ refers to the habitual intake, the number of days could be considered as infinite and, in this case, the expression of $\mathrm{CV}_{\mathrm{EI}}$ disappears(39). The equation becomes:

$$\pm 2 \times \sqrt{0 + 77{\cdot}44} = \pm 2 \times \sqrt{77{\cdot}44} = \pm 2 \times 8{\cdot}8$$

$$= 17{\cdot}6\% \text{ or } 0{\cdot}176.$$

Then, 1·0 (total concordance between EI and TEE) minus 0·176 is equal to 0·82. Thus under-reporters were defined as those subjects who did not lose at least 1 kg of body weight during the study and presented EI:TEE ratios smaller than 0·82. Using this cut-off value, the proportion of under-reporters was identified. This cut-off is very similar to those employed by Svendsen & Tonstad(40) and Andersen *et al.* (39). The limit of 1 kg body weight was set because under-reporting is, by definition, the report of an EI lower than TEE without weight loss. However, the literature does not define exactly what ‘without weight loss’ means in terms of g or kg, and it is a known fact that mild variations in body weight are normal. We established this limit considering that a loss greater than 1 kg probably is not derived from normal variations and, instead, it represents a real weight loss caused by an energy deficit.

Statistical methods

Analyses were performed with the Statistical Package for Social Sciences release 12.0 (SPSS Inc., Chicago, IL, USA). The significance level adopted was 0·05.

Cluster analysis was employed to derive dietary patterns from the dataset. The analysis was performed using the SPSS QUICK CLUSTER procedure, which is a K-means method that employs the Euclidean distances between observations to empirically estimate clusters. The itineration was conducted, being that its maximum number was 10. In this procedure the number of clusters must be established *a priori*. Since no information was available on the appropriate number of clusters, several solutions with a varying number of clusters (from two to five) were examined. We chose a solution with three clusters, because it provided more separated clusters (according to the ANOVA that compared the food groups’ variables between the clusters, for each solution), more homogeneous cluster sizes and greater ease in interpreting results.

The sample’s characteristics were presented as mean values and standard deviations. When means are being compared, 95 % CI and minimum and maximum values (ranges) are also provided. ANOVA was run to compare log-transformed EI, ratio of EI:TEE and difference between EI and TEE (EI − TEE) between the dietary pattern clusters. A χ^2 test compared proportion of under-reporters between the dietary pattern clusters. We also provided the 95 % CI for the proportion of under-reporters obtained by each cluster. The χ^2 test also compared proportion of under-reporters between BMI categories (normal-weight, overweight and obese subjects).

Results

Of the sixty-five participants, twenty-eight (43·1 %) were normal weight, ten (15·4 %) were overweight and twenty-seven (41·5 %) were obese. Approximately 66·1 % (*n* 43) were white, 26·2 % (*n* 17) were black or mulatto and 7·7 % (*n* 5) were Asian-Brazilians. Mean age was 33·7 (SD 10·8) years, while mean monthly income per capita was US$ 642·5 (SD 611·5). Mean body weight was 73·7 (SD 17·4) kg, while mean height was 1·63 (SD 0·07) m and mean BMI was 27·9 (SD 6·7) kg/m^2. Mean TEE was 10·97 (SD 2·05) MJ. The proportion of under-reporters was 53·6 % among normal-weight women, 40·0 % among overweight women and 59·3 % among obese women ($\chi^2(2) = 1·1$; $P=0·58$). Since this difference was not significant, we did not analyse obese individuals separately.

We obtained a solution with three clusters which were named according to their most recurrent food groups. Table 3 shows the mean frequency of intake of each food group per 1000 kJ obtained by each cluster. The first cluster consisted of thirty-one subjects and was named ‘sweet foods’, since it had higher mean intake of regular soda and confectionery products/sweets. The second cluster was composed of twenty-four participants and was entitled ‘starchy foods’ for having a higher mean intake of beans and pasta, grains and root vegetables. The third cluster consisted of ten subjects and was denominated ‘healthy’, since it presented higher mean intake of fruits, vegetables and poultry.

Table 4 shows EI, TEE, EI:TEE and EI − TEE obtained by each dietary pattern cluster and also the results from the ANOVA comparing these values between the clusters. Approximately 58·3 % of the sample was classified as under-reporters. Proportion of under-reporters was 45·2 (95 % CI 35·5, 55·0) % in the ‘sweet foods’ cluster, 58·3 (95 % CI 48·6, 68·0) % in the ‘starchy foods’ cluster and 70·0 (95 % CI 61·0, 79·0) % in the ‘healthy’ cluster ($\chi^2(2) = 2·2$; $P=0·34$).

Table 5 compares the characteristics of the subjects belonging to each dietary pattern cluster.

Discussion

The present study was designed to determine if under-reporting rates vary according to dietary pattern clusters. In order to do so, it was necessary to identify biased reports of EI at an individual level. Thus we used the procedures recommended by Black & Cole(38): (a) use of DLW to determine TEE; (b) assessment of within-variation in TEE measured by DLW; (c) calculation of the confidence limits of agreement between TEE and EI. We found that most of the subjects under-reported their EI and that under-reporting was more intense and frequent among the participants from the ‘healthy’ cluster.

The rationale for this research came from the observation that under-reporting tends to be greater for some specific foods and nutrients(20–25,41). Nevertheless, these studies identified under-reporters according to EI:RMR equations, N excretions, heart rate monitoring or by direct observation in a metabolic unit. One of the strengths of the present study is identifying under-reporters according to TEE measured by DLW, the only gold standard for this. Mela & Aaron asked some subjects what would be their attitude if they were asked to report their intake(42). Approximately 18·6 % said that they would reduce fat intake, 30·8 % would decrease consumption of sweets and 42·9 % would increase fruit and vegetable intake. If these observations are valid, it is possible that dietary pattern analysis is affected by under-reporting considering it is based on the same data.

We obtained a solution with three clusters. The ‘sweet foods’ cluster had the highest number of subjects (47·7 % of the sample) and it was the unhealthiest cluster. The ‘starchy foods’ cluster (with 36·9 % of the sample) and the ‘healthy’ cluster (with 15·4 % of the sample) were more consistent with dietary guidelines. Furthermore, the ‘healthy’ cluster

Table 3. Daily frequency of intake of food groups by each dietary pattern cluster*
(Mean values and 95 % confidence intervals)

	Dietary pattern clusters							
	Cluster 1: 'sweet foods' (*n* 31)		Cluster 2: 'starchy foods' (*n* 24)		Cluster 3: 'healthy' (*n* 10)			
Food groups (daily frequency/1000 kJ)	Mean	95 % CI	Mean	95 % CI	Mean	95 % CI	*F*†	*P*†
Bread	0·120	0·088, 0·151	0·139	0·103, 0·172	0·120	0·079, 0·16	0·4	0·69
Pasta, grains and root vegetables‡	0·163	0·143, 0·182	0·301	0·237, 0·363	0·229	0·165, 0·292	11·7	0·0001
Beans‡	0·065	0·048, 0·081	0·155	0·110, 0·203	0·060	0·024, 0·098	10·7	0·0001
Fruits‡	0·261	0·201, 0·320	0·306	0·239, 0·373	0·641	0·452, 0·837	17·1	0·0001
Vegetables‡	0·263	0·213, 0·311	0·320	0·272, 0·368	0·724	0·519, 0·93	33·0	0·0001
Eggs	0·022	0·012, 0·029	0·024	0·014, 0·031	0·022	0·005, 0·041	0·1	0·91
Regular dairy	0·088	0·057, 0·117	0·062	0·036, 0·086	0·065	0·024, 0·105	1·0	0·36
Fat-reduced dairy	0·084	0·057, 0·110	0·093	0·062, 0·206	0·134	0·022, 0·165	0·6	0·54
Bovine meat	0·072	0·057, 0·086	0·084	0·036, 0·086	0·060	0·053, 0·115	0·7	0·52
Swine meat	0·026	0·019, 0·033	0·033	0·024, 0·043	0·038	0·010, 0·065	1·0	0·38
Poultry‡	0·041	0·031, 0·05	0·038	0·022, 0·053	0·074	0·022, 0·124	3·1	0·05
Fish	0·014	0·010, 0·022	0·026	0·010, 0·05	0·029	0·012, 0·041	2·1	0·14
Cookies	0·062	0·036, 0·088	0·062	0·033, 0·096	0·053	0·031, 0·074	0·1	0·91
Confectionery products/candies‡	0·318	0·253, 0·382	0·103	0·072, 0·136	0·141	0·076, 0·203	18·7	0·0001
Sugar	0·122	0·079, 0·165	0·086	0·002, 0·120	0·057	0·045, 0·124	1·7	0·19
Regular soft drinks	0·048	0·017, 0·076	0·010	0·012, 0·019	0·022	0·002, 0·048	2·6	0·07
Diet soft drinks	0·033	0·010, 0·057	0·017	0, 0·033	0·012	0·001, 0·024	1·1	0·34
Spreads	0·091	0·055, 0·124	0·124	0·065, 0·186	0·098	0·031, 0·167	0·6	0·54
Fat-reduced spreads	0·007	0·001, 0·014	0·002	0·001, 0·007	0·017	0·005, 0·041	1·9	0·17
Fried foods	0·029	0·019, 0·038	0·024	0·007, 0·038	0·022	0·010, 0·033	0·29	0·75
Pizza	0·014	0·019, 0·038	0·014	0·007, 0·038	0·019	0·010, 0·033	0·4	0·7
Hamburger	0·012	0·007, 0·014	0·007	0·005, 0·012	0·019	0·002, 0·038	2·5	0·09

* For details of food groups and procedures, see Tables 1 and 2 and Subjects and methods.
† *F* and *P* values were obtained by means of an ANOVA that compared daily frequency of intake of food groups between the clusters.
‡ Significant difference between clusters.

was also more consistent with what lay individuals usually consider 'healthy eating'(43,44).

Under-reporting of EI was present among subjects from all clusters, but we may state that it was smaller in the 'sweet foods' cluster, intermediate in the 'starchy foods' cluster and higher in the 'healthy' cluster. The 'healthy' cluster obtained the lowest means for EI:TEE and EI − TEE. It can also be seen in Table 3 that, for these variables, this cluster presented the lowest minimum and maximum values. According to the χ^2 test, there was no significant difference regarding the proportion of under-reporters between the clusters. However, it is important to remember that this test is very sensitive to the size of the subgroups. The 95 % CI show that there is a significant difference in the proportion of under-reporters between the subjects from the 'sweet foods' cluster and those from the 'healthy' cluster. Wirfält *et al.* observed that subjects reporting EI:BMR $\leq$ 1·35 have a greater chance of belonging to healthiest clusters, characterised by intake of high-fibre and low-fat foods(45). Nevertheless, their study could not verify if under-reporting was greater among these clusters, because they used calculations and cut-off points that do not individually identify under-reporters. Moreover, they did not evaluate weight change during the study. This is extremely important because subjects with low EI can only be classified as under-reporters if they report eating less than TEE and do not lose weight. Nevertheless, the authors found that it was highly unlikely for subjects from the cluster with the highest intake of fat and sweets to report an EI:BMR $\leq$ 1·35. Studying obese subjects, Svendsen & Tonstad found that intake of sweets and desserts was positively correlated with reporting accuracy and it was one of the few multivariate determinants of accuracy(40). These results corroborate our findings, although the former was obtained in a sample composed only by obese subjects and we recruited subjects with a wide range of BMI. However, in their research and in several aforementioned, food intake was expressed in absolute values, without energy adjustment. Another strength of the present study is that even after this adjustment was made we still found that under-reporters tended to report a healthier dietary pattern. Thus, the present results show that under-reporting of EI is not uniformly distributed among dietary pattern clusters. To our knowledge, the present study is the first to show this.

The present study also has some limitations. We employed a sample of Brazilian women, who had a large range of BMI and income. Nevertheless, the sample's size may limit the generalisability to other populations, although it was suitable for the statistical methods used. Also, it was not possible to repeat the analysis after the exclusion of under-reporters because this number was so high that after eliminating these subjects the sample would be too small to carry out the cluster analysis. However, this analysis is very important, since only one study has conducted it so far and has presented inconclusive findings. Moreover, it would be even more relevant to examine in cohort studies if dietary patterns obtained with and without under-reporters present different correlations with health effects. This research would indicate if under-reporting of EI really impairs inferences concerning dietary patterns and health outcomes. Another possible limitation is the use of a FFQ. Marks *et al.* showed that FFQ and diet records produce different reports of some food groups

Table 4. Values of total energy expenditure (TEE), energy intake (EI), ratio of energy intake:total energy expenditure (EI:TEE) and difference between EI and TEE (EI − TEE) obtained by the dietary pattern clusters*

(Mean values, 95 % confidence intervals and minimum and maximum values)

	Cluster 1: ‘sweet foods’ (*n* 31)			Cluster 2: ‘starchy foods’ (*n* 24)			Cluster 3: ‘healthy’ (*n* 10)				
	Mean	95 % CI	Minimum and maximum	Mean	95 % CI	Minimum and maximum	Mean	95 % CI	Minimum and maximum	*F*	*P*
TEE (MJ/d)	10·42	9·72, 11·12	6·60, 14·10	11·73	9·99, 13·48	6·40, 15·20	11·37	10·55, 12·20	7·60, 16·20	2·39	0·10
EI (MJ/d)	8·93	7·67, 10·18	4·60, 18·80	8·18	6·70, 9·65	3·30, 15·90	6·66	4·21, 9·11	3·40, 14·00	2·5	0·08
EI:TEE	0·86	0·76, 0·96	0·45, 1·50	0·71	0·59, 0·83	0·42, 1·44	0·58	0·36, 0·8	0·30, 1·36	6·5	0·003
EI − TEE (MJ/d)	−1·49	−2·63, −0·35	−6·69, 6·31	−3·20	−4·52, −1·87	−8·04, 4·87	−5·08	−7·69, −2·46	−8·55, 3·71	5·3	0·008

* For details of food clusters and procedures, see Tables 1 and 3 and Subjects and methods.

Table 5. Characteristics of the subjects belonging to each dietary pattern cluster*

(Mean values and standard deviations, medians and minimum and maximum values)

	Cluster 1: ‘sweet foods’ (*n* 31)				Cluster 2: ‘starchy foods’ (*n* 24)				Cluster 3: ‘healthy’ (*n* 10)					
	Mean	SD	Median	Minimum and maximum	Mean	SD	Median	Minimum and maximum	Mean	SD	Median	Minimum and maximum	*F*	*P*
Age (years)	31	10	27	18, 55	40	8	42	24, 53	34	12	34	18, 57	2·81	0·07
BMI (kg/m^2)	25·8	6·1	24·2	18·5, 39·9	31·1	6·7	32·9	20·7, 39·1	29·3	6·7	31·0	18·9, 39·2	3·55	0·04
Monthly income per capita (US$)	797·4	727·8	500	100, 3000	613·3	556·8	542	58, 2000	462	411	344	38, 1500	2·17	0·12
Education (years)	13·5	2·0	15·0	8·0, 15·0	13·3	2·1	14·0	10·0, 15·0	12·1	2·6	13·0	4·0, 15·0	2·51	0·09
Protein density (g protein/1000 kJ)	9·4	1·5	9·4	5·5, 12·0	9·8	1·5	9·6	7·6, 13·2	10·0	1·9	9·8	6·7, 13·6	1·1	0·34
Carbohydrate density (g carbohydrate/1000 kJ)	30·8	4·5	31·6	21·1, 40·7	32·5	4·9	32·9	22·0, 39·8	32·1	6·2	32·9	19·9, 43·2	0·60	0·55
Lipid density (g lipid/1000 kJ)	8·7	1·5	8·8	4·5, 11·7	8·1	1·5	8·2	6·3, 11·2	8·3	1·8	8·0	4·9, 11·8	0·74	0·48

* For details of food clusters and procedures, see Tables 1 and 3 and Subjects and methods.

(such as fruits, eggs and vegetables)[46]. Nevertheless, authors such as Willett state that carefully developed FFQ can provide measures of dietary intake compatibles with those obtained by more detailed dietary assessment methods(47). We chose this method because it is able to capture habitual intake, which is necessary for dietary pattern analysis, and also because it is the most used method in epidemiological studies(48). Thus, it is important to investigate the errors associated with a method that is widely used.

Although dietary pattern analysis is a promising technique for nutrition studies, several issues (besides influence of under-reporting) remain. Newby & Tucker affirmed that few studies have investigated its validity and that it is not clear how validation should be conducted(5). Comparing the patterns based in one dietary assessment method with the nutrient intake obtained through another dietary assessment method or comparing dietary patterns obtained by different dietary assessment methods falls back into the same dilemma that researchers face when it is necessary to validate a single dietary assessment method. All dietary assessment methods have limitations and many of them present errors that are dependent on true intake and correlated with the errors from other dietary assessment methods. A better way would be to correlate dietary patterns with risk factors for diseases and biomarkers, but one should note that most of these studies conducted to this date were cross-sectional. Additionally, many studies establish an '*a priori*' premise, such as: 'we know that intake of nutrient X protects against disease Y, so if my analysis is valid, the dietary pattern with the highest intake of X must have a lower risk of developing Y'. This reasoning is logical and intuitive, but it ignores the fact that we do not know all relationships between diet and health and that some of the known relationships may be biased by errors. So, if the data contradict the premise, it is possible that the dietary pattern analysis is not valid, but it is also possible that the premise is false.

The dietary patterns found in the present study are similar to others found in the literature(10,49). Even in very different countries such as Iran(50), Uruguay(51), the USA(52) and Japan(53), it is common to observe at least two dietary patterns: one healthier or more prudent, characterised by the intake of fruits, vegetables, reduced-fat products, and one characterised by consumption of meat, fat, fast-food and sweets, which is frequently named the 'Western' pattern. This shows that dietary patterns tend to have high reproducibility across populations. Nevertheless, one question arises: do the individuals that report a healthy pattern actually eat healthily or merely report doing so? Most of the studies above mentioned observed that the healthy pattern protects against several diseases, which supports the idea that the healthy reporters are actually healthy eaters. However, two of our main findings may contradict that idea: (a) those who reported a healthy pattern under-reported more their EI and (b) those subjects tended to be heavier, which is counterintuitive since Newby *et al.* (10,49) and Esmaillzadeh & Azadbakht(50) found that a pattern characterised by a higher intake of fruits and vegetables protects against overweight. It is possible that our subjects from the healthy cluster actually consumed more fruits, vegetables and poultry (as they reported) and under-reported the intake of other foods. To solve this question, the use of some biomarkers that are associated with fruit and vegetable intake may be helpful(54). Unfortunately, we do not have such biomarker data in this research, but we recommend its use in further studies.

In summary, under-reporting of EI is not uniformly distributed among dietary pattern clusters. As we hypothesised, it was more severe in subjects from the healthiest cluster, probably due to selective under-reporting. Researchers should remember that dietary pattern analysis is derived from biased methods. More research is necessary in order to investigate how reporting errors affect dietary pattern analysis and also what is the best way to validate this technique, especially before dietary recommendations are made based on them.

Acknowledgements

F. B. S. was the principal investigator of the project. F. B. S., E. F. and A. H. L. Jr designed the study. F. B. S., K. P., C. L., C. S. F. C. and B. G. collected the data. F. B. S., E. F., K. P., C. L. and C. S. F. C. conducted data analysis. All authors wrote the manuscript. Funding for the study was provided through a grant from The State of São Paulo Research Foundation (FAPESP). The researchers were independent from the Foundation. The Foundation did not have any role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the article for publication. None of the authors had any personal or financial conflicts of interest. The authors thank Sandra Ribeiro, Camila Maria de Melo, Fabiana Benatti, Desire Coelho, Viviane Polacow, Bruna São Pedro, Aline Kazurayama and Geisa Gastal for their help in data collection and Juliana Pasqual for assisting in manuscript preparation.

The institution where the work was performed was University of São Paulo, School of Physical Education and Sport, Department of Biodynamics, Laboratory of Nutrition and Applied Metabolism (São Paulo/SP, Brazil).

References

1. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L & Willett WC (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* **69**, 243–249.
2. Khani BR, Ye W, Terry P & Wolk A (2004) Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *J Nutr* **134**, 1541–1545.
3. Millen BE, Quatromoni PA, Copenhafer DL, Demissie S, O'Horo CE & D'Agostino RB (2001) Validation of a dietary pattern approach for evaluating nutrition risk: the Framingham Nutrition Studies. *J Am Diet Assoc* **101**, 187–194.
4. Hu FB (2002) Dietary patterns analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
5. Newby PK & Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 177–203.
6. Quatromoni PA, Copenhafer DL, Demissie S, D'Agostino RB, O'Horo CE, Nam BH & Millen BE (2002) The internal validity of a dietary pattern analysis: the Framingham Nutrition Studies. *J Epidemiol Community Health* **56**, 381–388.
7. Newby PK, Hu FB, Rimm EB, Smith-Warner AS, Feskanich D, Sampson L & Willett WC (2003) Reproducibility and validity

of the Diet Quality Index Revised as assessed by use of a food-frequency questionnaire. *Am J Clin Nutr* **78**, 941–949.

8. McNaughton SA, Mishra GD, Bramwell G, Paul AA & Wadsworth ME (2005) Comparability of dietary patterns assessed by multiple dietary assessment methods: results from the 1946 British Birth Cohort. *Eur J Clin Nutr* **59**, 341–352.
9. Newby PK, Muller D & Tucker KL (2004) Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *Am J Clin Nutr* **80**, 759–767.
10. Newby PK, Muller D, Hallfrisch J, Andres R & Tucker KL (2004) Food patterns measured by factor analysis and anthropometric changes in adults. *Am J Clin Nutr* **80**, 504–513.
11. Kant AK & Graubard BI (2005) A comparison of three dietary pattern indexes for predicting biomarkers of diet and disease. *J Am Coll Nutr* **24**, 294–303.
12. Nettleton JA, Steffen LM, Mayer-Davis EJ, Jenny NS, Jiang R, Herrington DM & Jacobs DR Jr (2006) Dietary patterns are associated with biochemical markers of inflammation and endothelial activation in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr* **83**, 1369–1379.
13. Fung TT, McCullough ML, Newby PK, Manson JE, Meigs JB, Rifai N, Willett WC & Hu FB (2005) Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. *Am J Clin Nutr* **82**, 163–173.
14. Newby PK, Weismayer C, Akesson A, Tucker KL & Wolk A (2006) Long-term stability of food patterns identified by use of factor analysis among Swedish women. *J Nutr* **136**, 626–633.
15. Hill RJ & Davies PSW (2001) The validity of self-reported energy intake as determined using doubly labelled water technique. *Br J Nutr* **85**, 415–430.
16. Livingstone MBE & Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133**, 895S–920S.
17. Schoeller DA (2002) Validation of habitual energy intake. *Public Health Nutr* **5**, 883–888.
18. Lissner L, Heitmann BL & Lindroos AK (1998) Measuring intake in free-living human subjects: a question of bias. *Proc Nutr Soc* **57**, 333–339.
19. Trabulsi J & Schoeller DA (2001) Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. *Am J Physiol Endocrinol Metab* **281**, E891–E899.
20. Heitmann BL, Lissner L & Osler M (2000) Do we eat less fat or just report so? *Int J Obes Relat Metab Dis* **24**, 435–442.
21. Heitmann BL & Lissner L (1995) Dietary under-reporting by obese individuals: is it specific or non-specific? *BMJ* **311**, 986–989.
22. Scagliusi FB, Polacow VO, Artioli GG, Benatti FB & Lancha AH Jr (2003) Selective under-reporting of energy intake in women: magnitude, determinants, and effect of training. *J Am Diet Assoc* **103**, 1306–1313.
23. Krebs-Smith SM, Graubard BI, Kahle LL, Subar AF, Cleveland LE & Ballard-Barbash R (2000) Low energy reporters *vs.* others: a comparison of reported food intakes. *Eur J Clin Nutr* **54**, 281–287.
24. Lafay L, Mennen L, Basdevant A, Charles MA, Borys JM, Eschwege E & Romon M (2000) Does energy intake under-reporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Santé (FLVS) study. *Int J Obes Relat Metab Dis* **24**, 1500–1506.
25. Pomerleau J, Ostbye T & Bright-See E (1999) Potential under-reporting of energy intake in the Ontario Health Survey and its relationship with nutrient and food intakes. *Eur J Epidemiol* **15**, 553–557.
26. Bailey RL, Mitchell DC, Miller C & Smiciklas-Wright H (2007) Assessing the effect of under-reporting energy intake on dietary patterns and weight status. *J Am Diet Assoc* **107**, 64–71.
27. Cade J, Thompson R, Burley V & Warm D (2002) Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutr* **5**, 567–587.
28. Guedes DP (1994) *Composição Corporal: Princípios, Técnicas e Aplicações (Body Composition: Principles, Techniques and Applications)*, 2nd ed. Londrina, Brazil: APEF.
29. Schoeller DA (1996) Hidrometry. In *Human Body Composition*, pp. 25–43 [AF Roche, SB Heymsfield and TG Lohman, editors]. Champaign, IL: Human Kinetics.
30. Schoeller DA (2002) Validation of habitual energy intake. *Public Health Nutr* **5**, 883–888.
31. Speakman JR (1998) The history and theory of the doubly labeled water technique. *Am J Clin Nutr* **68**, 932S–938S.
32. Weir JB (1949) New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol* **109**, 1–9.
33. Sichieri R & Everhart JE (1998) Validity of a Brazilian food frequency questionnaire against dietary recalls and estimated energy intake. *Nutr Res* **18**, 1649–1659.
34. Lajolo FM & Menezes EW (1997) Tabela Brasileira de Composição de Alimentos (Brazilian Food Composition Table). http://www.fcf.usp.br/tabela (accessed July 2002).
35. Núcleo de Estudos e Pesquisas em Alimentação (2004) *TACO – Tabela Brasileira de Composição de Alimentos (TACO – Brazilian Food Composition Table)*. Campinas, Brazil: NEPA-UNICAMP.
36. United States Department of Agriculture (2004) National Nutrient Database for Standard Reference. http://www.nal.usda.gov/fnic/cgibin/nut_search.pl (accessed November 2003).
37. Philippi ST (2005) Brazilian food pyramid. *Nutr Today* **40**, 79–83.
38. Black AE & Cole TJ (2000) Within- and between-subject variation in energy expenditure measured by the doubly-labelled water technique: implications for validating reported dietary energy intake. *Eur J Clin Nutr* **54**, 386–394.
39. Andersen LF, Tomten H, Haggarty P, Lovo A & Hustvedt BE (2003) Validation of energy intake estimated from a food frequency questionnaire: a doubly labelled water study. *Eur J Clin Nutr* **57**, 279–284.
40. Svendsen M & Tonstad S (2006) Accuracy of food intake reporting in obese subjects with metabolic risk factors. *Br J Nutr* **95**, 640–649.
41. Pryer JA, Vrijheid M, Nichols R, Kiggins M & Elliot P (1997) Who are the “low energy reporters” in the dietary and nutritional survey of British adults? *Int J Epidemiol* **26**, 146–153.
42. Mela DJ & Aaron JI (1997) Honest but invalid: what subjects say about recording their food intake. *J Am Diet Assoc* **97**, 791–793.
43. Rozin P, Ashmore M & Markwith M (1996) Lay American conceptions of nutrition: dose insensitivity, categorical thinking, contagion, and the monotonic mind. *Health Psychol* **15**, 438–447.
44. Oakes ME & Slotterback CS (2002) The good, the bad, and the ugly: characteristics used by young, middle-aged, and older men and women, dieters and non-dieters to judge healthfulness of foods. *Appetite* **38**, 91–97.
45. Wirfält E, Mattisson I, Gullberg B & Berglund G (2000) Food patterns defined by cluster analysis and their utility as dietary exposure variables: a report from the Malmö Diet and Cancer Study. *Public Health Nutr* **3**, 159–173.
46. Marks GC, Hughes MC & van der Pols JC (2006) Relative validity of food intake estimates using a food frequency questionnaire is associated with sex, age, and other personal characteristics. *J Nutr* **136**, 459–465.
47. Willett W (2001) Invited Commentary: a further look at dietary questionnaire validation. *Am J Epidemiol* **154**, 1100–1102.

48. Thompson FE & Byers T (1994) Dietary assessment resource manual. *J Nutr* **124**, 2245S–2317S.
49. Newby PK, Muller D, Hallfrisch J, Qiao N, Andres R & Tucker KL (2003) Dietary patterns and changes in body mass index and waist circumference in adults. *Am J Clin Nutr* **77**, 1417–1425.
50. Esmaillzadeh A & Azadbakht L (2008) Major dietary patterns in relation to general obesity and central adiposity among Iranian women. *J Nutr* **138**, 358–363.
51. De Stefani E, Boffetta P, Ronco AL, Deneo-Pellegrini H, Acosta G & Mendilaharsu M (2007) Dietary patterns and risk of laryngeal cancer: an exploratory factor analysis in Uruguayan men. *Int J Cancer* **121**, 1086–1091.
52. Gao X, Chen H, Fung TT, Logroscino G, Schwarzschild MA, Hu FB & Ascherio A (2007) Prospective study of dietary pattern and risk of Parkinson disease. *Am J Clin Nutr* **86**, 1486–1494.
53. Okubo H, Sasaki S, Murakami K, Kim MK, Takahashi Y, Hosoi Y & Itabashi M (2007) Three major dietary patterns are all independently related to the risk of obesity among 3760 Japanese women aged 18–20 years. *Int J Obes* (Epublication ahead of print version 25 September 2007).
54. Brantsaeter AL, Haugen M, Rasmussen SE, Alexander J, Samuelsen SO & Meltzer HM (2007) Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Public Health Nutr* **10**, 838–847.