British Journal of Nutrition (2013), **109**, 1881–1891 doi:10.1017/S0007114512003868

British Journal of Nutrition

Dietary patterns obtained through principal components analysis: the effect of input variable quantification

Andrew D. A. C. Smith[1], Pauline M. Emmett[1], P. K. Newby[2,3,4] and Kate Northstone[1]*

[1]*School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK*

[2]*Department of Pediatrics and Program in Graduate Medical Nutrition Sciences, Boston University School of Medicine, 88 East Newton Street, Vose Hall 308, Boston, MA 02188, USA*

[3]*Department of Epidemiology, Boston University School of Public Health, 88 East Newton Street, Vose Hall 308, Boston, MA 02188, USA*

[4]*Program in Gastronomy, Culinary Arts, and Wine Studies, Metropolitan College at Boston University, Boston, MA 02215, USA*

(Submitted 15 February 2012 – Final revision received 9 July 2012 – Accepted 31 July 2012 – First published online 6 September 2012)

Abstract

Principal components analysis (PCA) is a popular method for deriving dietary patterns. A number of decisions must be made throughout the analytic process, including how to quantify the input variables of the PCA. The present study aims to compare the effect of using different input variables on the patterns extracted using PCA on 3-d diet diary data collected from 7473 children, aged 10 years, in the Avon Longitudinal Study of Parents and Children. Four options were examined: weight consumed of each food group (g/d), energy-adjusted weight, percentage contribution to energy of each food group and binary intake (consumed/not consumed). Four separate PCA were performed, one for each intake measurement. Three or four dietary patterns were obtained from each analysis, with at least one component that described 'more healthy' and 'less healthy' diets and one component that described a diet with high consumption of meat, potatoes and vegetables. There were no obvious differences between the patterns derived using percentage energy as a measurement and adjusting weight for total energy intake, compared to those derived using gram weights. Using binary input variables yielded a component that loaded positively on reduced fat and reduced sugar foods. The present results suggest that food intakes quantified by gram weights or as binary variables both resulted in meaningful dietary patterns and each method has distinct advantages: weight takes into account the amount of each food consumed and binary intake appears to describe general food preferences, which are potentially easier to modify and useful in public health settings.

Key words: Dietary patterns: Principal components analysis: Avon Longitudinal Study of Parents and Children

The use of dietary patterns to explore the effects of diet on a variety of health outcomes is now well established as a method that complements examining individual foods and nutrients. Dietary patterns allow the assessment of the whole diet, accounting for the fact that foods/nutrients are consumed in combination and are therefore highly correlated. Principal components analysis (PCA), a form of factor analysis, is a popular method for deriving dietary patterns. It makes use of the correlations between food intakes to identify underlying patterns in the data. There are several subjective decisions that must be made when using PCA. A particularly important one, which is often overlooked, is how to quantify the input variables. Depending on the source of dietary data, a number of different variables could be considered. For example, data from diet diaries can be quantified continuously as gram weights or percentage energy from food groups or dichotomously (i.e. whether each food group was consumed or not).

The input variables used in PCA vary across studies[1] and include frequency of consumption, gram weights, energy-adjusted weight, daily percentage energy contribution and binary variables. Many studies based on diet diaries use weight of foods consumed as the input variable[2–5]. Energy adjustment using the residual method[6] is often applied in studies based on diet diaries and diet recalls[7–9], as well as studies based on FFQ data[10–12]. Percentage energy is another

---

Abbreviations: A1, A2, A3, components derived from weights (g/d) adjusted for total energy intake using the residual method; ALSPAC, Avon Longitudinal Study of Parents and Children; B1, B2, B3, B4, components derived from binary variables; P1, P2, P3, P4, components derived from percentage contribution of each food to total energy intake; PCA, principal components analysis; W1, W2, W3, components derived from weights (g/d).

* **Corresponding author:** Dr K. Northstone, fax +44 117 3310080, email Kate.Northstone@bristol.ac.uk

potential input variable[13] and a few studies[14,15] have dichotomised intakes into binary variables. Most studies select one strategy for dietary patterns analyses, but seldom justify the decision, and only a few studies have made comparisons between the different input variables but with no formal conclusions[14,16,17]. There are no studies to our knowledge that have compared all four strategies and no studies have made comparisons in children.

In order to facilitate comparisons across studies, it is vital that researchers are as informed as possible about the decisions that they need to make and use the best evidence available. Therefore, the aim of the present study is to derive dietary patterns using PCA and using four different input variables – weight (g/d), energy-adjusted weight, percentage energy contribution and binary variables (consumed or not consumed) – and compare the interpretability of the patterns among children participating in the Avon Longitudinal Study of Parents and Children (ALSPAC).

Methods

Participants

The ALSPAC is an ongoing longitudinal cohort study designed to investigate determinants of development, health and disease during and after childhood. Eligible participants were pregnant women resident in the former Avon Health Authority, in South West England, due to deliver between 1 April 1991 and 31 December 1992. Further details are given elsewhere[18] and can be found on the website http://www.bris.ac.uk/alspac. The study includes children from the core ALSPAC sample, consisting of 14 541 pregnancies, and an additional 542 eligible pregnancies not in the core sample, invited to participate at a later date. The present study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects/patients were approved by the ALSPAC Law and Ethics Committee and the Local Research Ethics Committees. Written informed consent was obtained from all subjects/patients.

Dietary assessment

The study children were invited to attend a clinic when they were 10 years old, and a diet diary was sent with their confirmation to be completed prior to their visit. Children and their care-givers recorded, in household measures, all food and drink consumed by the child over two (not necessarily consecutive) weekdays and one weekend day. During clinic attendance, the children were interviewed to ensure the quality of the diary (e.g. clarifying portion size or omitted details on the types of food and drinks consumed). If the child did not bring a diary to the clinic, the fieldworker conducted a 24-h recall to record all food and drink consumed by the child in the previous day. Further details are given elsewhere[19]. The completed diaries were entered into the Diet In Data Out computer program[20], which generated the weight and energy contribution of every food consumed by each child. For the purposes of the present study, the average daily intake of food weight and energy were used.

Each food consumed was initially allocated to one of ninety-five food groups that were based on those used in FFQ that had previously been administered to the ALSPAC cohort[21]. Sugar-free confectionery, alcohol, herbs and spices were removed from the analysis, as very few children consumed these foods and, thus, they did not contribute meaningfully to any dietary patterns. The remaining food items were combined into sixty-two groups, based on similarities between foods (e.g. nuts, peanuts and peanut butter were combined), to reduce the number of input variables and prevent infrequently consumed foods from diluting the dietary patterns. The appendix describes the food groups in detail.

Statistical methods

Dietary patterns were derived using PCA. Principal components are linear combinations of the input variables and explain as much of the variation in the data as possible. Each component describes a dietary pattern and the linear combination allows the calculation of a component score for each child; the higher the score, the more likely this pattern is present in an individual's diet. The patterns described by each component may be interpreted by its factor loadings, which are the correlations between the component and each input variable. Large positive or negative factor loadings indicate the foods that are important in that component; loadings with magnitude of at least 0·2 were considered when describing dietary patterns. Scree plots[22] and the interpretability of each component were also used to determine the appropriate number of components to select. Varimax rotation[23] was employed to aid the interpretation of components. The purpose of the present study was to compare the different dietary patterns obtained using each of the input variables; therefore, the patterns were given alphanumeric labels rather than descriptive names to aid reporting.

Four separate analyses were carried out using four different input variables. The first used the weight (g/d) of each food consumed. The variables were standardised prior to entry into the PCA to prevent components being dominated by the foods that are consumed in the highest quantities, such as water. The second analysis adjusted the mean weight for total energy intake, using the residuals method[6]. Specifically, the PCA input variables were the standardised residuals from a linear regression of mean weight on mean daily energy intake. Regression was only performed on non-zero values, and both weight and energy were log-transformed before regression and transformed back before standardisation. The third analysis used the percentage contribution of each food to the daily energy intake as input variables. These percentage energy input variables were also standardised prior to entry into the PCA to prevent the components being dominated by the foods that provide the highest percentage energy. In the fourth analysis, the input variables were dichotomised into binary variables (consumed or not consumed), as food intake variables were highly skewed and many children did

https://doi.org/10.1017/S0007114512003868 Published online by Cambridge University Press

not consume some of the food groups. The PCA was performed directly on their covariance matrix for this fourth method (as opposed to the correlation matrix for the previous three methods), as standardisation is not appropriate for binary variables. For each of the four PCA, scores were calculated for each subject, for each pattern derived by summing the products of each standardised input variable and their corresponding coefficient in the component (or dichotomised in the case of binary variables).

Agreement between the derived patterns was assessed in two ways. Agreement between component scores was assessed by calculating Pearson's sample correlation coefficients. Congruence coefficients[24] were also calculated for pairs of matrices of component coefficients in order to assess the difference between the coefficients assigned to individual foods by each component.

Results

Of the 11 868 children eligible to attend the clinic, a total of 7557 (63·7 %) attended and 7473 of these (98·9 %) provided dietary information. Of these, 5769 (77·2 %) provided 3 d of dietary records. Girls, white children, children with older, more educated, non-smoking mothers and children from homes that were owned or mortgaged were more likely to provide data (all $P < 0·001$; data not shown).

When gram weights were used as input variables, three principal components were retained and explained 10·4 % of the variation in the sample. Factor loadings are shown in Table 1. The first component (W1) had high positive loadings on non-white bread, fruit and vegetables, cooked pasta, tuna and oily fish, cheese, yoghurt, high energy density sauce (e.g. mayonnaise), fruit juice and water. There were high negative loadings on processed meat, coated poultry, tinned pasta/baked beans, chips (French fries), crisps (potato chips) and carbonated sweet drinks (non-diet soda). The second component (W2) had high positive loadings on meat, roast potatoes, batter/pastry products, vegetables, puddings and low energy density sauce (e.g. gravy, ketchup) and a high negative loading on chips. The third component (W3) had high positive loadings on white bread, margarine, cheese, cold meats, salty flavourings, crisps, biscuits (cookies) and diet squash/cordial.

As can be seen in Table 2, energy adjustment did not have a discernible effect on the dietary patterns when compared with those using unadjusted weights: the factor loadings were almost identical, differing by no more than 0·084.

Four components were obtained when percentage energy contribution was used as the input variable, explaining 12·3 % of the variation in the sample. Factor loadings are shown in Table 3. The first three components, labelled P1, P2 and P3, had high loadings on the same foods that loaded highly on components W1, W2 and W3, with the exception that water loaded highly on W1 but not P1; vegetarian products, legumes and nuts loaded highly on P1 but not W1; and diet squash/cordial loaded highly on W3 but not P3. The fourth component (P4) had high positive loadings on reduced fat milk, yoghurt, breakfast cereal and biscuits and high negative loadings on rice, other breads (e.g. pitta), poultry, eggs, butter, salad, legumes and carbonated sweet drinks.

When PCA was performed on binary variables, four components were obtained, explaining 17·3 % of the variation in the sample. Table 4 shows factor loadings for these four components. The first component (B1) had high loadings on meat, roast potatoes, batter/pastry products, vegetables and low energy density sauces. The second component (B2) had high positive loadings on non-white bread, fruit, nuts, salad, vegetarian foods and vegetable dishes, potatoes, pasta, tuna and oily fish, cheese, yoghurt, eggs, butter, high energy density sauce, sweet spreads (e.g. jam), dairy puddings, cakes, chocolate, fruit juice, regular squash/cordial and water. There were high negative loadings on diet squash/cordial, and roast potatoes. The third component (B3) had high loadings on processed meat, coated poultry, tinned pasta/baked beans, white bread, margarine, vegetable oil, chips, crisps, chocolate, sweets (candy), sweet spreads (jams), sugar, cakes, dairy puddings, biscuits, carbonated sweet drinks and diet squash/cordial. The fourth component (B4) had high positive loadings on reduced fat milk, margarine, diet carbonated drinks and diet squash/cordial. It also had high negative loadings on their alternatives, i.e. full-fat milk, butter, carbonated sweet drinks and regular squash/cordial. It also had a high positive loading on breakfast cereals.

Table 5 shows the correlations between the component scores, and Table 6 shows congruence coefficients between components. The components generated from gram weights and energy-adjusted weight input variables are very similar, as assessed by correlations between component scores and the congruence coefficient between these components. The first three components from the analysis with percentage energy input variables were also similar to those generated from gram weights: the correlations among P1, P2, P3 and W1, W2, W3 were at least 0·907. The components generated by binary input variables share partial similarities with the other components. In terms of component scores, B1 was positively correlated with W2, B2 with W1 and B3 was negatively correlated with W1.

Discussion

The present study of dietary diary data from 10-year-old children compared dietary patterns derived from PCA using four strategies for quantifying input variables. When continuous variables were used (gram weights, energy-adjusted weight and percentage energy contribution), the first three components extracted had similar loadings and described similar dietary patterns: one contrasting 'more healthy' foods with 'less healthy' foods, one with high loadings on meat, potatoes and vegetables and one with high loadings on lunch and snack foods. The fourth component, present only when intake was measured as percentage energy, was difficult to interpret. When binary variables were used, the four components extracted described slightly different dietary patterns: the component with high loadings on meat, potatoes and vegetables was still present, but the component with positive loadings on 'more healthy' foods and negative loadings on

Table 1. Factor loadings from principal components analysis of diet diary data on 7473 children aged 10 years, where input variables are weights (g/d)

Factor (variance explained)	W1 (3·80 %)	W2 (3·60 %)	W3 (3·10 %)
Full-fat milk	−0·056	0·004	−0·055
Reduced fat milk	0·158	0·023	−0·012
Cheese	0·309*	−0·123	0·261*
Yoghurt and fromage frais	0·208*	−0·030	0·157
Butter and animal fat	0·162	−0·081	−0·099
Margarine	0·061	0·035	0·712*
Vegetable oil	−0·083	−0·058	−0·057
High-fibre bread	0·334*	−0·099	−0·063
Low-fibre bread	−0·012	−0·040	0·707*
Other bread (e.g. pitta)	0·223*	−0·077	−0·101
Batter and pastry products	0·032	0·277*	0·081
Breakfast cereal	0·098	−0·036	−0·175
Rice	0·164	−0·016	−0·156
Pasta	0·249*	−0·045	−0·086
Baked beans and tinned pasta	−0·210*	−0·117	−0·099
Pizza	−0·036	−0·176	−0·090
Eggs	0·089	−0·074	−0·030
Coated and fried chicken	−0·310*	−0·132	−0·104
Poultry	0·052	0·223*	−0·047
Ham and bacon	−0·008	0·006	0·235*
Red meat	0·023	0·233*	−0·056
Meat pies and pasties	−0·142	0·059	−0·037
Processed meat	−0·295*	−0·042	−0·015
Coated and fried white fish	−0·087	−0·087	−0·100
White fish and shellfish	0·095	−0·026	−0·098
Tuna and oily fish	0·276*	−0·099	−0·042
Vegetarian products	0·190	−0·069	−0·036
Chips (French fries)	−0·512*	−0·224*	−0·176
Roast potatoes	−0·149	0·678*	0·013
Other potatoes	0·152	0·159	−0·027
Root vegetables	0·114	0·251*	−0·025
Carrots	0·104	0·610*	0·002
Green leafy vegetables	0·124	0·527*	−0·032
Peas, broad beans and sweet corn	0·031	0·249*	−0·096
Other cooked vegetables and dishes	0·234*	0·179	−0·069
Salad and tomatoes	0·443*	−0·149	−0·035
Legumes	0·235*	−0·086	−0·068
Soup	0·134	−0·081	−0·012
Nuts, seeds and peanut butter	0·193	−0·039	0·041
Fresh fruit	0·427*	−0·007	0·048
Tinned and dried fruit	0·143	0·007	−0·026
Puddings	0·012	0·240*	−0·169
Dairy puddings	−0·098	0·218*	−0·180
Cakes	0·095	0·041	−0·053
Chocolate	−0·133	−0·023	0·015
Sweets (candy)	−0·149	−0·012	0·028
Sugar	−0·093	0·053	−0·008
Sweet spreads (e.g. jam)	0·105	0·042	0·174
Biscuits (cookies)	−0·116	0·040	0·245*
Crackers and crisp breads	0·145	−0·038	0·106
Crisps (potato chips)	−0·207*	−0·035	0·333*
Low energy density sauce (e.g. gravy, ketchup)	0·014	0·599*	−0·008
High energy density sauce (e.g. mayonnaise)	0·302*	−0·138	−0·010
Salty flavouring (e.g. yeast extract)	0·110	−0·035	0·345*
Water and flavoured water	0·304*	−0·016	−0·128
Carbonated sweet drinks (soda)	−0·246*	−0·076	−0·090
Carbonated diet drinks (diet soda)	−0·226*	0·050	0·079
Regular squash and cordial	0·046	−0·065	0·002
Diet squash and cordial	−0·184	0·083	0·289*
Fruit juice	0·263*	−0·055	−0·029
Flavoured milk drinks	−0·034	−0·003	0·006
Tea and coffee	−0·034	0·093	0·079

W1, W2, W3, components derived from weights (g/d).
* Factor loadings with magnitude greater than 0·2.

https://doi.org/10.1017/S0007114512003868 Published online by Cambridge University Press

https://doi.org/10.1017/S0007114512003868 Published online by Cambridge University Press

British Journal of Nutrition

Table 2. Factor loadings from principal components analysis of diet diary data on 7473 children aged 10 years, where input variables are weights (g/d) adjusted for total energy intake using the residual method

Factor (variance explained)	A1 (3·80 %)	A2 (3·60 %)	A3 (3·10 %)
Full-fat milk	−0·062	−0·014	−0·126
Reduced fat milk	0·154	0·022	−0·034
Cheese	0·309*	−0·131	0·211*
Yoghurt and fromage frais	0·202*	−0·038	0·108
Butter and animal fat	0·155	−0·104	−0·133
Margarine	0·058	0·037	0·713*
Vegetable oil	−0·079	−0·072	−0·085
High-fibre bread	0·333*	−0·104	−0·096
Low-fibre bread	−0·020	−0·047	0·718*
Other bread (e.g. pitta)	0·224*	−0·083	−0·093
Batter and pastry products	0·024	0·271*	0·048
Breakfast cereal	0·101	−0·047	−0·220*
Rice	0·172	−0·020	−0·121
Pasta	0·251*	−0·041	−0·082
Baked beans and tinned pasta	−0·212*	−0·121	−0·111
Pizza	−0·039	−0·183	−0·100
Eggs	0·091	−0·082	−0·051
Coated and fried chicken	−0·308*	−0·143	−0·105
Poultry	0·058	0·223*	−0·025
Ham and bacon	−0·006	0·003	0·224*
Red meat	0·020	0·230*	−0·072
Meat pies and pasties	−0·142	0·054	−0·064
Processed meat	−0·297*	−0·048	−0·046
Coated and fried white fish	−0·092	−0·095	−0·103
White fish and shellfish	0·098	−0·029	−0·095
Tuna and oily fish	0·269*	−0·101	−0·032
Vegetarian products	0·187	−0·068	−0·026
Chips (French fries)	−0·515*	−0·241*	−0·194
Roast potatoes	−0·148	0·676*	0·008
Other potatoes	0·149	0·157	−0·044
Root vegetables	0·112	0·249*	−0·009
Carrots	0·105	0·606*	−0·004
Green leafy vegetables	0·125	0·521*	−0·036
Peas, broad beans and sweet corn	0·029	0·245*	−0·101
Other cooked vegetables and dishes	0·240*	0·176	−0·060
Salad and tomatoes	0·442*	−0·152	−0·039
Legumes	0·241*	−0·087	−0·054
Soup	0·136	−0·081	−0·011
Nuts, seeds and peanut butter	0·191	−0·047	0·024
Fresh fruit	0·422*	−0·014	0·012
Tinned and dried fruit	0·130	0·001	−0·057
Puddings	0·010	0·228*	−0·180
Dairy puddings	−0·099	0·198	−0·264*
Cakes	0·093	0·030	−0·129
Chocolate	−0·142	−0·040	−0·052
Sweets (candy)	−0·155	−0·022	−0·016
Sugar	−0·096	0·044	−0·083
Sweet spreads (e.g. jam)	0·098	0·026	0·132
Biscuits (cookies)	−0·130	0·023	0·177
Crackers and crisp breads	0·138	−0·042	0·079
Crisps (potato chips)	−0·217*	−0·048	0·295*
Low energy density sauce (e.g. gravy, ketchup)	0·017	0·597*	−0·016
High energy density sauce (e.g. mayonnaise)	0·304*	−0·150	−0·012
Salty flavouring (e.g. yeast extract)	0·106	−0·027	0·353*
Water and flavoured water	0·307*	−0·019	−0·109
Carbonated sweet drinks (soda)	−0·255*	−0·095	−0·132
Carbonated diet drinks (diet soda)	−0·223*	0·055	0·080
Regular squash and cordial	0·038	−0·083	−0·023
Diet squash and cordial	−0·193	0·091	0·262*
Fruit juice	0·266*	−0·079	−0·066
Flavoured milk drinks	−0·035	−0·011	−0·042
Tea and coffee	−0·037	0·089	0·048

A1, A2, A3, components derived from weights (g/d) adjusted for total energy intake using the residual method.
* Factor loadings with magnitude greater than 0·2.

https://doi.org/10.1017/S0007114512003868 Published online by Cambridge University Press

Table 3. Factor loadings from principal components analysis of diet diary data on 7473 children aged 10 years, where input variables are percentage contribution of each food to total energy intake

Factor (variance explained)	P1 (3·50 %)	P2 (3·20 %)	P3 (3·00 %)	P4 (2·60 %)
Full-fat milk	−0·054	−0·041	−0·134	−0·004
Reduced fat milk	0·151	0·007	−0·082	0·492*
Cheese	0·306*	−0·137	0·184	0·062
Yoghurt and fromage frais	0·203*	−0·075	0·054	0·258*
Butter and animal fat	0·143	−0·099	−0·118	−0·230*
Margarine	0·066	0·014	0·720*	0·078
Vegetable oil	−0·076	−0·086	−0·086	−0·038
High-fibre bread	0·331*	−0·123	−0·106	0·163
Low-fibre bread	−0·019	−0·025	0·740*	−0·174
Other bread (e.g. pitta)	0·235*	−0·097	−0·117	−0·340*
Batter and pastry products	0·060	0·207*	0·013	−0·136
Breakfast cereal	0·099	−0·043	−0·263*	0·556*
Rice	0·183	−0·003	−0·114	−0·312*
Pasta	0·255*	−0·013	−0·096	0·060
Baked beans and tinned pasta	−0·172	−0·136	−0·071	0·057
Pizza	−0·071	−0·174	−0·095	0·074
Eggs	0·084	−0·078	−0·049	−0·214*
Coated and fried chicken	−0·315*	−0·129	−0·079	−0·037
Poultry	0·063	0·246*	−0·035	−0·302*
Ham and bacon	−0·039	0·064	0·153	−0·150
Red meat	−0·013	0·363*	−0·069	0·032
Meat pies and pasties	−0·163	0·032	−0·064	0·065
Processed meat	−0·323*	−0·029	−0·046	−0·071
Coated and fried white fish	−0·115	−0·093	−0·073	0·144
White fish and shellfish	0·076	−0·048	−0·096	−0·047
Tuna and oily fish	0·263*	−0·109	−0·061	−0·050
Vegetarian products	0·271*	−0·125	−0·008	−0·004
Chips (French fries)	−0·558*	−0·210*	−0·154	−0·089
Roast potatoes	−0·121	0·679*	0·009	−0·052
Other potatoes	0·146	0·091	−0·045	0·098
Root vegetables	0·192	0·175	−0·011	−0·148
Carrots	0·134	0·588*	−0·006	0·047
Green leafy vegetables	0·138	0·534*	−0·020	0·037
Peas, broad beans and sweet corn	0·024	0·204*	−0·111	0·034
Other cooked vegetables and dishes	0·160	−0·046	−0·073	−0·140
Salad and tomatoes	0·203*	−0·111	−0·065	−0·208*
Legumes	0·272*	−0·110	−0·075	−0·234*
Soup	0·130	−0·097	−0·015	−0·061
Nuts, seeds and peanut butter	0·204*	−0·062	0·048	0·009
Fresh fruit	0·389*	0·001	−0·004	0·088
Tinned and dried fruit	0·205*	−0·066	−0·037	0·054
Puddings	0·017	0·189	−0·171	−0·050
Dairy puddings	−0·082	0·147	−0·249*	−0·033
Cakes	0·084	0·021	−0·145	−0·080
Chocolate	−0·145	−0·049	−0·047	−0·120
Sweets (candy)	−0·162	−0·015	−0·018	−0·066
Sugar	−0·078	0·017	−0·108	0·112
Sweet spreads (e.g. jam)	0·086	−0·004	0·130	0·027
Biscuits (cookies)	−0·120	0·004	0·159	0·259*
Crackers and crisp breads	0·147	−0·077	0·065	0·054
Crisps (potato chips)	−0·208*	−0·021	0·301*	0·043
Low energy density sauce (e.g. gravy, ketchup)	0·082	0·407*	−0·047	−0·043
High energy density sauce (e.g. mayonnaise)	0·279*	−0·163	−0·020	−0·141
Salty flavouring (e.g. yeast extract)	0·105	−0·009	0·394*	−0·013
Water and flavoured water	0·005	0·029	−0·043	−0·050
Carbonated sweet drinks (soda)	−0·280*	−0·098	−0·087	−0·255*
Carbonated diet drinks (diet soda)	−0·220*	0·138	0·074	−0·062
Regular squash and cordial	0·012	−0·065	0·024	−0·036
Diet squash and cordial	−0·071	0·033	0·068	0·165
Fruit juice	0·271*	−0·088	−0·052	−0·091
Flavoured milk drinks	−0·022	−0·022	−0·057	−0·011
Tea and coffee	0·012	−0·010	−0·043	−0·052

P1, P2, P3, P4, components derived from percentage contribution of each food to total energy intake.
* Factor loadings with magnitude greater than 0·2.

Table 4. Factor loadings from principal components analysis of diet diary data on 7473 children aged 10 years, where intakes are expressed as binary (consumed/not consumed) variables

Factor (variance explained)	B1 (5·20 %)	B2 (5·00 %)	B3 (3·90 %)	B4 (3·20 %)
Full-fat milk	0·065	0·043	0·106	− 0·654*
Reduced fat milk	− 0·032	0·077	− 0·002	0·773*
Cheese	− 0·091	0·424*	0·025	0·089
Yoghurt and fromage frais	− 0·019	0·264*	0·057	0·200
Butter and animal fat	− 0·030	0·298*	− 0·069	− 0·282*
Margarine	0·086	− 0·080	0·230*	0·320*
Vegetable oil	− 0·088	0·171	0·313*	− 0·009
High-fibre bread	− 0·068	0·341*	− 0·109	0·112
Low-fibre bread	0·045	0·042	0·230*	0·040
Other bread (e.g. pitta)	− 0·001	0·171	− 0·060	− 0·020
Batter and pastry products	0·392*	− 0·003	0·132	− 0·029
Breakfast cereal	0·015	0·164	0·089	0·204*
Rice	0·050	0·173	− 0·120	− 0·014
Pasta	− 0·052	0·348*	− 0·147	0·075
Baked beans and tinned pasta	− 0·053	− 0·064	0·386*	− 0·047
Pizza	− 0·151	0·101	0·154	0·029
Eggs	− 0·035	0·249*	0·078	− 0·017
Coated and fried chicken	− 0·070	− 0·099	0·386*	− 0·043
Poultry	0·369*	0·032	0·029	0·024
Ham and bacon	0·084	0·097	0·128	0·094
Red meat	0·425*	0·032	− 0·056	0·023
Meat pies and pasties	0·083	− 0·066	0·080	0·006
Processed meat	0·039	− 0·092	0·368*	− 0·030
Coated and fried white fish	− 0·027	− 0·013	0·147	− 0·018
White fish and shellfish	− 0·012	0·146	− 0·044	− 0·014
Tuna and oily fish	− 0·047	0·306*	− 0·079	0·040
Vegetarian products	− 0·080	0·203*	− 0·069	− 0·019
Chips (French fries)	− 0·096	− 0·187	0·551*	− 0·091
Roast potatoes	0·761*	− 0·210*	0·040	− 0·053
Other potatoes	0·166	0·238*	0·028	0·022
Root vegetables	0·233*	0·181	− 0·078	− 0·011
Carrots	0·700*	0·062	− 0·074	− 0·022
Green leafy vegetables	0·579*	0·078	− 0·113	− 0·025
Peas, broad beans and sweet corn	0·368*	0·098	0·057	0·013
Other cooked vegetables and dishes	0·343*	0·266*	− 0·135	0·034
Salad and tomatoes	− 0·069	0·594*	− 0·086	− 0·001
Legumes	− 0·031	0·190	− 0·090	− 0·034
Soup	− 0·017	0·137	− 0·048	− 0·025
Nuts, seeds and peanut butter	− 0·043	0·236*	− 0·023	0·012
Fresh fruit	0·041	0·459*	0·010	0·082
Tinned and dried fruit	0·011	0·298*	0·021	0·003
Puddings	0·193	0·123	0·064	− 0·040
Dairy puddings	0·157	0·227*	0·292*	− 0·045
Cakes	0·054	0·267*	0·243*	− 0·020
Chocolate	− 0·021	0·210*	0·349*	− 0·031
Sweets (candy)	0·011	0·137	0·367*	− 0·054
Sugar	− 0·013	0·129	0·342*	0·051
Sweet spreads (e.g. jam)	− 0·044	0·299*	0·254*	− 0·022
Biscuits (cookies)	0·062	0·127	0·222*	0·067
Crackers and crisp breads	− 0·016	0·170	0·043	0·021
Crisps (potato chips)	0·048	− 0·015	0·216*	0·063
Low energy density sauce (e.g. gravy, ketchup)	0·507*	0·040	0·085	0·011
High energy density sauce (e.g. mayonnaise)	− 0·082	0·362*	− 0·058	0·037
Salty flavouring (e.g. yeast extract)	0·010	0·123	0·000	0·017
Water and flavoured water	0·015	0·336*	− 0·140	− 0·053
Carbonated sweet drinks (soda)	− 0·020	0·050	0·213*	− 0·321*
Carbonated diet drinks (diet soda)	0·036	− 0·127	0·252*	0·241*
Regular squash and cordial	− 0·070	0·241*	0·123	− 0·258*
Diet squash and cordial	0·123	− 0·244*	0·296*	0·346*
Fruit juice	− 0·043	0·410*	0·049	0·007
Flavoured milk drinks	0·030	0·122	0·129	0·017
Tea and coffee	0·074	− 0·031	0·121	0·067

B1, B2, B3, B4, components derived from binary variables.
* Factor loadings with magnitude greater than 0·2.

British Journal of Nutrition

Table 5. Correlations between component scores obtained from different input variables*

	W1	W2	W3		B1	B2	B3	B4
A1	0·995	0·105	−0·066		0·143	0·652	−0·430	0·159
A2	0·101	0·996	−0·040		0·765	0·058	−0·052	0·075
A3	−0·051	−0·059	0·962		−0·043	−0·151	0·050	0·241
P1	0·931	0·142	−0·023		0·159	0·599	−0·413	0·199
P2	0·061	0·918	−0·050		0·708	0·010	−0·086	0·068
P3	−0·056	−0·084	0·907		−0·069	−0·154	0·029	0·160
P4	0·003	−0·010	−0·076		0·044	0·043	−0·006	−0·392
B1	0·145	0·767	−0·026		P1	P2	P3	P4
B2	0·653	0·074	−0·102	A1	0·942	0·065	−0·063	0·004
B3	−0·420	−0·037	0·119	A2	0·144	0·937	−0·078	−0·018
B4	0·156	0·061	0·219	A3	−0·011	−0·034	0·962	−0·045

* W, components derived from weights (g/d); A, components derived from weights (g/d) adjusted for total energy intake using the residual method; P, components derived from percentage contribution of each food to total energy intake; B, components derived from binary variables.

'less healthy' foods was replaced by two components: one with high loadings on the 'more healthy' foods and the other with high loadings on the 'less healthy' foods. The fourth component had positive loadings for reduced-fat and reduced-sugar foods and negative loadings on their alternatives.

There are strong similarities between patterns in the presence and absence of energy adjustment, the main differences being in the relative loadings of high- and low-fibre bread, and full- and low-fat milk. In a comparison of energy-adjusted and unadjusted analyses of data from FFQ administered to the ALSPAC mothers[16], five components appear in the unadjusted analysis, but four components suffice under energy adjustment; the missing component described a 'processed' dietary pattern. A study[17] comparing gram weights and percentage energy as input variables, in PCA of FFQ data from Irish adults, concludes that gram weights give more interpretable patterns than percentage energy.

In the present study, the patterns obtained when gram weights were used as the input variables were the most interpretable. Weight is a clear, quantitative way to measure food consumption and can be easily linked to portion sizes. A drawback of using gram weights (unadjusted and adjusted for energy) and percentage energy was that they potentially led to skewed input variables, with many zeroes for foods that were not frequently consumed. This resulted in component scores with skewed distributions. Adjusting the weight for energy intake did not alter the dietary patterns, agreeing with research in adults[14]. These results suggest that energy-adjusting the input variables does not offer any specific benefit when determining dietary patterns, using PCA, from diet diaries administered to children. It may be more appropriate to perform energy adjustment later in the analytic process, as this allows for more accurate assessment of the effect of energy itself. A similar conclusion was reached when obtaining dietary patterns using PCA in the ALSPAC mothers, although this was based on the FFQ data[16].

In agreement with other research (in adults)[17], using percentage energy as an input variable led to patterns that were harder to interpret than those derived from gram weights. In the present study, the percentage energy strategy led to components in which water did not load highly, as it does not contribute to energy intake. This could be considered an inherent limitation of this approach, given non-energy-containing foods (e.g. water, coffee, tea and diet soda) often contribute meaningfully to dietary patterns. This is shown in the present study, in which water loaded highly on the components obtained when gram weights were used as the input variable strategy, whether energy-adjusted or unadjusted. These results indicate that variation in water intake is an important part of childhood diet and is missed when using the percentage energy method. Percentage energy is an attractive concept, as it considers one's overall dietary composition. However, it is harder to comprehend when dealing with individual food groups, which provide relatively small contributions to total energy intake when considered on their own (i.e. in contrast to considering, say, the macronutrient composition of the diet).

Few studies have used binary input variables to derived dietary patterns using PCA. Using this method, they overcame the issues of skewness and the sometimes large numbers of non-consumers of food groups, and led to interpretable dietary patterns. A study of data from an FFQ administered to adults in four European cohorts[14] showed no effect of dichotomisation of input variables on dietary patterns. However, in the present study, the patterns were different from those obtained from continuous variables; binary

Table 6. Congruence coefficients between components obtained from different input variables*

First set	Second set	Congruence
W1, W2, W3	A1, A2, A3	0·994
W1, W2, W3	P1, P2, P3	0·954
W1, W2, W3	B2, B1, B3	0·624
A1, A2, A3	P1, P2, P3	0·964
A1, A2, A3	B2, B1, B3	0·579
P1, P2, P3, P4	B2, B1, B3, B4	0·505

* W, components derived from weights (g/d); B, components derived from binary variables; A, components derived from weights (g/d) adjusted for total energy intake using the residual method; P, components derived from percentage contribution of each food to total energy intake.

https://doi.org/10.1017/S0007114512003868 Published online by Cambridge University Press

(consumed/not consumed) variables are easy to understand and conceptually represent choices and/or preferences of food rather than quantities consumed. This was evident in component B4, which seemed to differentiate among individuals who chose reduced fat, reduced sugar foods and those who chose the regular (full fat, full sugar) options for those foods. Food choices are potentially easier to modify, but it must be recognised that people consume food in different quantities, and dichotomising food intakes does not capture the complexity of eating behaviour.

The findings of the present study are strengthened by the large sample size. However, the sample is biased towards higher socio-economic status. As well, the present study has not assessed the effect of different input variables on a specific diet–disease association. As the patterns obtained with different strategies were similar, the effect of input variables on a given diet–disease association may be similar, although this is an important next step to further this literature and needs to be examined. Another input variable that could be considered is the number of servings per day, which is commonly used in studies that assess diet using an FFQ. However, as the present study made use of diet diaries, considered a 'gold standard' method of self-reported dietary assessment, we elected not to consider this semi-quantitative approach commonly used in FFQ, given the level of detail we have in the diet diaries.

In conclusion, the present study is the first to comprehensively compare different input variables used in dietary pattern analysis obtained using PCA. The present results indicate that there appears to be no benefit associated with energy adjustment, given results were similar to those when unadjusted. We also showed that patterns based on percentage energy did not capture meaningful dietary intakes, completely missing some items consumed such as water, and were also harder to interpret. Thus, while the final choice of input variable treatment may depend on the purpose of a particular analysis, the use of food weights and binary variables appeared to be the best approaches to quantify input variables in the present study among children. More research is needed to see whether input variable treatment has an impact on diet–disease associations, as understanding the role of diet on health outcomes is the ultimate objective of nutritional epidemiological studies. However, for the purposes of describing the underlying patterns of diet in a population, we would recommend using weights of foods; binary input variables would be a complementary approach to this in which specific dietary choices can be identified.

Acknowledgements

We are extremely grateful to all the families who took part in the present study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council, the Wellcome Trust and the University of Bristol provide core support for ALSPAC. The present work was supported by the World Cancer Research Fund grant number 2009/23. K. N. and P. M. E. designed the study; A. D. A. C. S. performed the statistical analysis; K. N. had primary responsibility for final content. All authors contributed to writing the manuscript and approved the final version. The authors declare no conflict of interest.

References

1. Newby PK & Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 177–203.
2. Cucó G, Fernández-Ballart J, Sala J, *et al.* (2006) Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. *Eur J Clin Nutr* **60**, 364–371.
3. Hamer M, McNaughton SA, Bates CJ, *et al.* (2010) Dietary patterns, assessed from a weighed food record, and survival among elderly participants from the United Kingdom. *Eur J Clin Nutr* **64**, 853–861.
4. Mikkilä V, Räsänen L, Raitakari OT, *et al.* (2005) Consistent dietary patterns identified from childhood to adulthood: The Cardiovascular Risk in Young Finns Study. *Br J Nutr* **93**, 923–931.
5. Yannakoulia M, Yiannakouris N, Melistas L, *et al.* (2008) A dietary pattern characterized by high consumption of whole-grain cereals and low-fat dairy products and low consumption of refined cereals is positively associated with plasma adiponectin levels in healthy women. *Metabolism* **57**, 824–830.
6. Willett WC, Howe GR & Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65**, Suppl., 1220S–12288S.
7. Kesse-Guyot E, Vergnaud A, Fezeu L, *et al.* (2010) Associations between dietary patterns and arterial stiffness, carotid artery intima–media thickness and atherosclerosis. *Eur J Cardiovasc Prev Rehabil* **17**, 718–724.
8. McNaughton SA, Mishra GD, Bramwell G, *et al.* (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. Okubo H, Murakami K, Sasaki S, *et al.* (2010) Relative validity of dietary patterns derived from a self-administered diet history questionnaire using factor analysis among Japanese adults. *Public Health Nutr* **13**, 1080–1089.
10. Bamia C, Orfanos P, Ferrari P, *et al.* (2005) Dietary patterns among older Europeans: the EPIC – Elderly study. *Br J Nutr* **94**, 100–113.
11. Martínez-Ortiz JA, Fung TT, Baylin A, *et al.* (2005) Dietary patterns and risk of nonfatal acute myocardial infarction in Costa Rican adults. *Eur J Clin Nutr* **60**, 770–777.
12. Velie EM, Schairer C, Flood A, *et al.* (2005) Empirically derived dietary patterns and risk of postmenopausal breast cancer in a large prospective cohort study. *Am J Clin Nutr* **82**, 1308–1319.
13. Newby PK, Muller D, Hallfrisch J, *et al.* (2004) Food patterns measured by factor analysis and anthropometric changes in adults. *Am J Clin Nutr* **80**, 504–513.
14. Balder HF, Virtanen M, Brants HAM, *et al.* (2003) Common and country-specific dietary patterns in four European cohort studies. *J Nutr* **133**, 4246–4251.
15. Guinot C, Latreille J, Malvy D, *et al.* (2001) Use of multiple correspondence analysis and cluster analysis to study dietary behaviour: food consumption questionnaire in the SU.VI.-MAX. cohort. *Eur J Epidemiol* **17**, 505–516.
16. Northstone K, Ness AR, Emmett PM, *et al.* (2008) Adjusting for energy intake in dietary pattern investigations using principal components analysis. *Eur J Clin Nutr* **62**, 931–938.

17. Hearty AP & Gibney MJ (2009) Comparison of cluster and principal components analysis techniques to derive dietary patterns in Irish adults. *Br J Nutr* **101**, 598–608.
18. Golding J, Pembrey M, Jones R, *et al.* (2001) ALSPAC – The Avon Longitudinal Study of Parent and Children. I. Study methodology. *Paediatr Perinat Epidemiol* **15**, 74–87.
19. Cribb VL, Jones LR, Rogers IS, *et al.* (2011) Is maternal education level associated with diet in 10-year-old children? *Public Health Nutr* **14**, 2037–2048.
20. Price GM, Paul AA, Key FB, *et al.* (1995) Measurement of diet in a large national survey: comparison of computerised and manual coding of records in household measures. *J Hum Nutr Diet* **8**, 417–428.
21. Northstone K, Emmett P & The ALSPAC Study Team (2005) Multivariate analysis of diet in children at four and seven years of age and associations with socio-demographic characteristics. *Eur J Clin Nutr* **59**, 751–760.
22. Cattell RB (1966) The scree test for the number of factors. *Multivariate Behav Res* **1**, 245–276.
23. Kline P (1994) *An Easy Guide to Factor Analysis*. London: Routledge.
24. Harman HH (1976) *Modern Factor Analysis*. Chicago, IL: University of Chicago Press.

British Journal of Nutrition

https://doi.org/10.1017/S0007114512003868 Published online by Cambridge University Press

Appendix: Food groups and their components

Full-fat milk	Full-fat cow's, sheep's or goat's milk
Reduced fat milk	Skimmed or semi-skimmed cow's milk
Cheese	Hard, soft, cream or cottage cheese
Yoghurt and fromage frais	Plain or fruit yoghurt, fromage frais
Butter and animal fat	Butter, dripping, ghee, lard, suet
Margarine	Hard or soft margarine or spread
Vegetable oil	Canola/rapeseed, coconut, cod liver, corn, olive, peanut, safflower, sesame, soya or sunflower oil
High-fibre bread	White bread, hamburger buns, bagels
Low-fibre bread	Brown, wholemeal, granary or rye bread
Other bread	Pitta or naan bread, ciabatta, chapattis, papadums, tortillas
Batter and pastry products	Breadcrumbs, brioche, croissants, pancakes, pastry, scones, stuffing, Yorkshire pudding
Breakfast cereal	Bran, corn, rice or oat-based cereal or sweetened cereal
Rice	Brown, white, risotto or pilau rice
Pasta	Pasta, spaghetti, macaroni, lasagna, noodles, couscous
Baked beans and tinned pasta	Baked beans, canned spaghetti or ravioli, macaroni cheese, pasta salad, gnocchi, cannelloni, pot snacks
Pizza	Pizza, lunchbox snacks
Eggs	Hen's, duck's or quail's eggs, quiche, omelette, Scotch eggs
Coated and fried chicken	Chicken or turkey burgers, fingers, Kiev, nuggets or in crumbs
Poultry	Chicken, turkey, duck, rabbit, grouse, pheasant
Ham and bacon	Ham, gammon, bacon
Red meat	Beef, lamb, pork, veal, venison, haggis, liver, kidney
Meat pies and pasties	Beef, chicken or pork pie, sausage rolls
Processed meat	Sausages, burgers, luncheon meat
Coated and fried white fish	Cod, haddock, plaice, skate all in batter or breadcrumbs
White fish and shellfish	Cod, coley, haddock, hake, halibut, monkfish, plaice, sea bass, snapper, sole, clams, crab, cockles, mussels, scallops, scampi, squid, prawns
Tuna and oily fish	Tuna, anchovies, herring, kipper, mackerel, pilchards, salmon, sardines, swordfish, trout
Vegetarian products	Vegetable or bean burgers/sausages, Quorn, soya
Chips (French fries)	Chips, fried potatoes, potato waffles or croquettes
Roast potatoes	Old potatoes, roasted in fat
Other potatoes	New and old potatoes, boiled or baked
Root vegetables	Artichoke, beetroot, garlic, onion, parsnip, swede, turnip, yam
Carrots	Carrots
Green leafy vegetables	Broccoli, Brussels sprouts, cabbage, kale, spinach
Peas, broad beans and sweet corn	Peas, broad beans, sweet corn, mange-tout
Other cooked vegetables and dishes	Asparagus, cauliflower, celery, courgette, green or French beans, leek, marrow, peppers, pumpkin, squash, vegetable flans or pastries, cauliflower cheese
Salad and tomatoes	Raw vegetables, tomatoes
Legumes	Beans, lentils
Soup	Soup
Nuts, seeds and peanut butter	Nuts, peanuts, seeds, peanut butter
Fresh fruit	Citrus or other fruit
Tinned and dried fruit	Tinned or dried fruit
Puddings	Cheesecake, Christmas pudding, crumble, flan, fruit pie, jelly, Pavlova, sponge, trifle
Dairy puddings	Blancmange, bread and butter pudding, cream, custard, ice cream, mousse, rice pudding
Cakes	Buns, cakes, pastries
Chocolate	Chocolate confectionery
Sweets (candy)	Sugar confectionery
Sugar	Sugar, icing
Sweet spreads	Jam, honey, chocolate spread, lemon curd, marmalade
Biscuits (cookies)	Biscuits, fully-coated chocolate biscuits
Crackers and crisp breads	Crackers, oatcakes, water biscuits, cheese biscuits, rice cakes
Crisps (potato chips)	Potato crisps, corn snacks, pretzels
Low energy density sauce	Bread/cheese/tomato sauces, gravy, mustard, vinegar. Energy density below 8 kJ/g (2 kcal/g)
High energy density sauce	Mayonnaise, salad cream, chutney. Energy density above 8 kJ/g (2 kcal/g)
Salty flavouring	Yeast extract, stock cubes, table salt
Water and flavoured water	Water, flavoured water
Carbonated sweet drinks (soda)	Cola, lemonade, ginger ale, tonic water, energy drinks
Carbonated diet drinks (diet soda)	Diet cola, lemonade or energy drinks
Regular squash and cordial	Fruit squash or cordial*
Diet squash and cordial	Low sugar fruit squash or cordial*
Fruit juice	Fruit juice
Flavoured milk drinks	Flavoured milk
Tea and coffee	Tea, coffee†, herbal tea
Foods not included‡	Sugar-free sweets/jelly/mints/chewing gum, artificial sweetener, black treacle, instant dessert powder, diabetic jam/chocolate
	Alcoholic drinks
	Herbs, spices

* Weight of undiluted squash was multiplied by five to obtain equivalent diluted weight.
† Weight of coffee granules was multiplied by 190 to obtain equivalent liquid weight.
‡ Due to infrequency of consumption and lack of importance in any extracted component.