A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis

Kate Northstone* and Pauline M. Emmett

Department of Social Medicine, University of Bristol, 24 Tyndall Avenue, Bristol BS8 1TQ, UK

Few studies have examined the longitudinal nature of dietary patterns obtained using principal components analysis (PCA); the methods used are inconsistent. This paper investigates the methodologies used to assess stability and changes in such patterns. Pregnant women recorded frequency of consumption of various food items as part of regular self-completed questionnaires in the Avon Longitudinal Study of Parents and Children. This was repeated when their children were 4 years of age; 8953 women provided data at both times. Dietary patterns were identified using PCA and component scores were calculated at each time point. Additional 'applied' scores were created at 4 years using the loadings obtained from the PCA on the pregnancy data. Correlations were similar for each component across the time points, though slightly larger using the applied method. The applied scores were considerably lower on average than those obtained from separate PCA at 4 years. Women's scores decreased on 'health conscious' and 'confectionery' components while 'processed' and 'vegetarian' scores both increased over the 4-year period. In contrast, applied scores were systematically lower for all components. When split into quintiles, weighted κ was slightly higher between pregnancy and applied 4-year scores compared to the separate scores. In this cohort it was felt that the 'applied' method to obtain scores at the second time point was inappropriate, primarily due to the differences in FFQ between the two time points. We recommend that future studies using such 'applied' scores compare them with cross-sectional scores and consider the implications of any differences.

Dietary patterns: Pregnancy: Principal components analysis: ALSPAC: Diet

There has been a substantial increase in the use of multivariate statistical methods, such as principal components analysis (PCA), to assess patterns of diet. These patterns are being used in addition to individual nutrients or food intakes in disease-risk studies. PCA is useful in identifying the underlying dimensions in a large number of variables and is based on the inter-correlations between these variables.

Despite the popularity of this empirical method, only a handful of studies have examined dietary patterns longitudinally and the methods that have been employed have been inconsistent. There are a number of reasons why it would be of interest to investigate the longitudinal nature of dietary patterns. From a methodological point of view, several subjective decisions need to be taken by the researcher in order to derive dietary patterns by PCA and concern has been raised regarding the lack of generalisability of dietary patterns obtained using PCA across studies[1,2]. As a result, it would be beneficial to assess the stability, reproducibility, and therefore validity, of dietary patterns within the same population[3]. Furthermore, changes in dietary patterns over time are increasingly being used to assess disease-risk[4–6]. Such changes in patterns may reflect the modification of diet by individuals over time as a result of nutritional advice which is constantly being updated, or changes in the food supply; for example, newly available foods such as soya products and probiotics. Changes may also be due to a major life event, such as becoming pregnant and having children. It will be important to examine the implications of these changes to health.

Differing methodologies have been employed to investigate the stability of dietary patterns and changes over time. The most popular to date involves performing separate PCA at each time point and calculating component scores independently of each other[5–15]. Alternatively two studies which found that the dietary patterns obtained were similar across time points (in terms of the foods describing them and size of loadings), applied the loadings from one time point to the data at another thereby obtaining scores using the same formulae[14,15]. To our knowledge only three studies have assessed any change in dietary pattern scores over time. Again different methodologies have been used: regression analyses to obtain parameter estimates representing change in overall score[8,14] or movement across categorical groups[12]. Only one study has examined both continuous and categorical scores[15].

The aim of this study was to investigate the methods used to assess stability and/or changes in dietary patterns, which have been obtained using PCA, in a cohort study of pregnant women who provided further dietary information at 47 months after birth.

Methods

The Avon Longitudinal Study of Parents and Children (ALSPAC) is an ongoing longitudinal cohort study designed to investigate the determinants of development, health, and disease during childhood and beyond[16]. Pregnant women were eligible for the study if they had an expected date of delivery between 1 April 1991 and 31 December 1992, and were resident in the former Avon health authority in south-west England. More detailed information on the study can be found on the website (www.alspac.bris.ac.uk). A total of 14 541 women enrolled. Ethical approval for the study was obtained from the ALSPAC Law and Ethics committee and the three local research ethics committees. The primary source of data collection was via self-completed postal questionnaires.

The present study used data from questionnaires completed at 32 weeks gestation and 47 months (approximately 4 years) after the birth of the child. Both questionnaires contained a set of questions enquiring about the frequency of consumption of a wide variety of foods and drinks. Previous analysis of the pregnancy food frequency data[17] showed this questionnaire to produce mean nutrient intakes similar to those obtained for women in the British National Diet and Nutritional survey for adults[18]. The specific question on the frequency of oily fish consumption in pregnancy has been validated by comparison with the erythrocyte fatty acid composition of pregnancy blood samples. The leucocyte DHA content increased significantly with increasing frequency of consumption of oily fish, the predominant source of DHA ($P < 0.001$)[19].

In both questionnaires the woman was given the following options to indicate how often she was currently consuming a variety food type: (i) never or rarely; (ii) once in 2 weeks; (iii) 1–3 times per week; (iv) 4–7 times per week; (v) more than once daily. The woman was also asked to record how many cups of tea or coffee, the number of glasses of cola and the number of slices of bread she usually consumed daily. The usual type of bread (white or other) she used was also recorded. The second questionnaire had been modified slightly in the light of analysis of the first questionnaire and some changes in the foods available. For example, separate categories were created for coated poultry and fish products, vegetarian pies and tuna, which had previously been encompassed in other categories. Therefore, at 4 years a number of additional foods and drinks (see Table 1) were included in the questionnaire. We chose to keep this extra information separate rather than combine into existing food groups created from the pregnancy questionnaire.

The frequency of consumption data were numerically transformed into times consumed per week, in order to apply quantitative meaning to the frequency categories, as follows: (i) 0; (ii) 0·5; (iii) 2; (iv) 5·5 and (v) 10 times per week. All data were standardized by subtracting the mean and dividing by the standard deviation for each variable; this was necessary because tea, coffee, cola and bread were measured on a different scale from the other variables.

Statistical methods

PCA with varimax rotation[20,21] was performed on the forty-four standardized food items from the pregnancy questionnaire

and has been described in detail elsewhere[22]. An identical procedure was used for the fifty-two standardized food items from the 4-year questionnaire. The number of components that best represented the data was primarily chosen on the basis of the scree plot[23] and the interpretability of the components. Women were excluded from each PCA if they had more than ten dietary items missing from the respective questionnaire. We made the assumption that if ten or fewer items were missing, the woman did not consume those items and they were given a value of 0. Most (92 %) of the women gave complete responses to the FFQ at both time points. Of those with incomplete data, 85 % omitted only one item and 8 % omitted two items. Foods with loadings above 0·3 on a component were considered to have a strong association with that component and were deemed to be the most informative in describing the dietary patterns. We have chosen to give each component a label; these do not perfectly describe each underlying pattern but aid in the report and discussion of the results.

A component score was created for each woman for each of the components identified at both time points by multiplying the factor loadings by the corresponding standardized value for each food and summing across the food items. In line with previous studies[14,15], an additional set of scores were created for the 47-month data using the loadings obtained from the PCA on the pregnancy data; to aid reporting we have chosen to call these scores 'applied'.

All component scores were approximately normally distributed. Pearson's correlation coefficients were calculated to measure the associations between the dietary pattern scores obtained at the two time points and using the two different methods. Paired t tests were applied to assess the change in mean scores over the 4-year period between questioning. Limits of agreement (95 %)[24] were calculated as the mean

Table 1. Baseline characteristics of the study sample (n 8935)

Characteristic	n	%
Level of education		
No school qualifications at age 16	2245	25·2
School qualification obtained at age 16	3175	35·6
Education beyond age 16	3500	39·2
Maternal age at birth (years)		
< 20	233	2·6
20–24	1322	14·8
25–29	3580	40·0
≥ 30	3818	42·6
Housing tenure during pregnancy		
Owned or mortgaged	7042	80·2
Council or housing association	991	11·3
Private rented or other	746	8·5
Parity		
0	4016	45·9
1	3107	35·5
2 +	1627	18·6
Ethnicity		
White	8734	98·3
Non-white	151	1·7
Vegetarian		
Yes	491	5·6
No	8278	94·4
Smoked in last trimester of pregnancy		
Yes	1509	16·9
No	7442	83·1

Table 2. Factor loadings of various food items in the principal dietary components identified (loadings above ± 0.3 are shown in bold) in women at 32 weeks of pregnancy and when their children were 4 years of age based on food frequency questionnaires

Food item (Variance explained)	'Health conscious' Pregnancy (10·6%)	4 years (9·0%)	'Traditional'* Pregnancy (8·2%)	'Processed' Pregnancy (4·9%)	4 years (7·5%)	'Confectionery' Pregnancy (4·0%)	4 years (4·8%)	'Vegetarian' Pregnancy (3·6%)	4 years (3·8%)
White bread	**−0·535**	**−0·294**	0·075	**0·367**	**0·550**	0·080	0·003	−0·018	−0·080
Non-white bread	**0·615**	**0·331**	−0·049	**−0·323**	**−0·546**	−0·057	0·077	0·032	0·124
Bran based cereal	**0·365**	0·287	0·092	−0·126	−0·196	−0·004	0·064	0·009	−0·029
Oat based cereal	0·297	0·272	0·113	−0·039	−0·152	0·050	0·104	0·140	0·160
Other breakfast cereal	−0·110	−0·041	−0·015	0·139	0·171	0·221	0·258	−0·082	−0·042
Biscuits	0·108	−0·029	0·023	−0·007	0·025	**0·603**	**0·674**	−0·108	−0·089
Crispbreads/crackers	0·218	0·261	0·088	−0·010	−0·024	0·052	−0·032	0·156	0·102
Puddings	0·265	0·101	0·064	0·124	0·121	**0·389**	**0·475**	−0·112	−0·017
Yoghurts	NA	**0·353**	NA	NA	−0·165	NA	0·260	NA	0·041
Ice cream	NA	0·013	NA	NA	0·097	NA	**0·480**	NA	−0·003
Cakes/buns	0·202	0·035	0·004	0·086	0·002	**0·559**	**0·620**	−0·080	−0·030
Poultry	0·270	**0·359**	0·223	0·121	0·087	0·023	0·031	**−0·535**	**−0·502**
Red meat	0·147	0·207	0·219	0·166	0·168	0·101	0·090	**−0·596**	**−0·635**
Cold meats	NA	0·092	NA	NA	0·073	NA	0·186	NA	**−0·458**
Coated poultry products	NA	−0·011	NA	NA	**0·434**	NA	0·137	NA	−0·119
Meat pies	−0·105	−0·165	0·032	**0·538**	**0·520**	0·087	0·149	−0·118	−0·064
Offal	0·087	0·108	0·091	0·248	0·177	−0·066	−0·035	0·087	−0·069
Sausages, burgers	−0·091	−0·146	−0·062	**0·565**	**0·458**	0·029	0·194	−0·169	−0·131
Fried foods	−0·094	NA	0·001	**0·574**	NA	0·164	NA	−0·009	NA
Pizza	0·233	0·014	−0·105	**0·349**	0·277	0·104	0·250	0·105	0·274
Breaded/battered white fish	NA	0·096	NA	NA	**0·413**	NA	0·173	NA	0·073
Fish†	**0·457**	**0·479**	0·155	0·133	−0·019	−0·075	−0·009	−0·018	0·076
Eggs	0·278	0·232	0·090	**0·403**	0·201	−0·027	0·111	−0·016	0·037
Cheese	**0·443**	0·203	0·078	0·053	−0·011	0·122	0·286	0·026	0·199
Vegetarian pies	NA	0·092	NA	NA	0·190	NA	0·020	NA	**0·530**
Meat substitutes (soya, tofu etc)	0·180	0·203	0·066	0·124	0·039	−0·028	−0·021	**0·577**	**0·624**
Pulses	**0·356**	**0·365**	0·146	0·006	−0·116	−0·055	−0·044	**0·565**	**0·489**
Nuts	0·278	0·183	0·116	0·051	−0·002	0·052	0·137	**0·531**	**0·391**
Chips	−0·255	−0·148	−0·057	**0·561**	**0·566**	0·235	0·284	−0·036	0·089
Roast potatoes	−0·271	0·036	0·225	**0·388**	**0·467**	0·154	0·117	−0·165	−0·205
Potatoes (not chips)	0·254	**0·387**	**0·321**	0·104	0·136	0·070	0·058	−0·219	−0·130
Pasta	**0·578**	**0·422**	0·045	0·136	−0·113	−0·070	0·039	0·121	0·215
Rice	**0·543**	**0·421**	0·078	0·125	−0·040	−0·120	0·015	0·063	0·113
Baked beans‡	0·004	0·113	0·049	**0·413**	**0·473**	0·081	0·135	0·045	0·112
Leafy green vegetables	0·045	**0·667**	**0·809**	0·011	0·124	−0·015	−0·150	0·041	−0·130
Other green vegetables	0·147	**0·697**	**0·799**	−0·043	0·109	−0·004	−0·149	0·054	−0·128
Carrots	0·178	**0·660**	**0·704**	−0·020	0·095	0·023	−0·053	0·008	−0·139
Other root vegetables	0·084	**0·549**	**0·606**	0·018	0·162	0·003	−0·131	0·106	−0·053
Peas	0·174	0·297	**0·352**	0·190	**0·344**	0·063	0·060	−0·104	−0·039
Sweetcorn	NA	**0·412**	NA	NA	0·216	NA	−0·040	NA	0·113
Salad	**0·420**	**0·395**	0·212	−0·078	0·146	−0·022	−0·011	0·100	0·157
Fresh fruit	**0·518**	**0·491**	0·182	−0·229	−0·247	0·090	0·130	0·005	0·039
Fruit juice	**0·488**	0·252	0·079	−0·090	−0·209	0·085	0·242	0·057	0·091
Fizzy drinks	NA	−0·052	NA	NA	0·184	NA	0·219	NA	−0·109
Cola	−0·209	NA	−0·081	0·221	NA	0·142	NA	0·051	NA
Tea	−0·100	0·014	0·078	0·156	0·108	0·029	0·003	−0·037	0·041

Dietary pattern changes

1101

Table 2. Continued

Food item (Variance explained)	'Health conscious'		'Traditional'*	'Processed'		'Confectionery'		'Vegetarian'	
	Pregnancy (10.6 %)	4 years (9.0 %)	Pregnancy (8.2 %)	Pregnancy (4.9 %)	4 years (7.5 %)	Pregnancy (4.0 %)	4 years (4.8 %)	Pregnancy (3.6 %)	4 years (3.8 %)
Coffee	−0.161	0.002	0.053	0.105	0.008	0.002	−0.084	−0.037	−0.048
Herbal tea	0.186	0.155	0.068	−0.085	−0.123	−0.057	−0.024	0.302	0.205
Water	NA	0.232	NA	NA	−0.103	NA	0.128	NA	0.066
Flavoured milk drinks	NA	0.087	NA	NA	0.130	NA	0.165	NA	0.070
Squash	NA	0.031	NA	NA	0.134	NA	0.323	NA	−0.078
Sweets	−0.098	−0.044	0.071	0.069	0.092	0.514	0.371	0.061	−0.077
Chocolate§	0.000	−0.053	0.022	0.036	0.019	0.717	0.578	0.058	−0.026
Chocolate bars	−0.085	NA	−0.020	0.096	NA	0.749	NA	0.021	NA
Crisps	−0.101	−0.077	−0.041	0.292	0.177	0.381	0.474	0.004	−0.002

NA, question not asked.

* Traditional pattern not obtained post-partum.

† In pregnancy all fish combined. At 47 months, fish other than coated.

‡ At 47 months includes tinned pasta.

§ At 47 months includes all chocolate (chocolate bars combined).

difference between the pregnancy and 4-year scores plus or minus twice the standard deviation of the differences; these enabled us to assess the extent of agreement between the time points and provide an idea of the spread of the variation of scores between the time points. In order to make comparisons with other studies[8,13] all component scores were then split into quintiles and were compared across time using weighted κ[25]. Weighted κ was employed due to the ordered nature of the categorical data; weighted κ takes into account partial agreement between groups. Finally in an attempt to assess the stability of the patterns over time, the dietary pattern scores were split into quintiles. Cross-tabulations between the pregnancy score quintiles and the two sets of 4-year score quintiles are presented.

All analyses were performed using SPSS for Windows version 12.0.1, with the exception of weighted κ, which was obtained using STATA for Windows version 9.2.

Results

A total of 12 436 women returned the questionnaire completed at 32 weeks gestation (85·5 % of the original sample; many of these had already been lost due to miscarriage). Of these, 12 053 (96·9 %) had sufficient dietary data available for the PCA. 9504 (65·4 % of the original sample) women returned the questionnaire 4 years after birth and a total of 8953 women (61·6 %) had PCA performed at both time points. Table 1 details the basic characteristics of the sample and it can be seen that those under study are more likely to be older, better educated, live in owner-occupied accommodation and be pregnant for the first time. Only 1·7 % of the sample was of non-white ethnicity. 5·6 % of the women reported themselves to be vegetarian and 16·9 % were smoking at 32 weeks gestation. We present elsewhere a detailed analysis of the differences in dietary pattern scores and socio-demographic factors[22].

Five dietary components were chosen to best describe the dietary patterns of the women in pregnancy, explaining 31·3 % of the variance. Table 2 shows the factor loadings obtained. At 4 years after birth, four components were preferred, explaining 25·1 % of the variance. It should be noted that these proportions are not directly comparable due to the differences in the number of food items entered into the two PCA. At both time points the first component loaded highly on salad, fresh fruit, rice, pasta, fish, pulses, and non-white bread and was labelled 'health conscious'. The second component from the pregnancy data had high loadings for all types of vegetables and to some extent red meat and poultry and was labelled 'traditional' in line with the familiar British 'meat and two veg' diet. This component was not extracted from the 4-year data. The foods that described this component at the first time point loaded highly on the 'health conscious' component at the second time point. As foods such as vegetables, poultry and fish are recommended in healthy eating regimes, we kept the label 'health conscious' for the first component at the second time point.

The three remaining components at each time point were virtually identical in the foods that loaded highly on them resulting in 'processed', 'confectionery' and 'vegetarian' components. As described earlier, the additional foods loading highly on each of these components at 4 years did not require a change in the name from pregnancy. The 'processed'

component loaded highly on meat pies, sausages and burgers, fried foods, pizza and chips. The 'confectionery' component was characterized by high intakes of chocolate, sweets, biscuits, cakes and other puddings. The 'vegetarian' component loaded highly on meat substitutes, pulses, nuts and herbal tea and high negative loadings were seen with all meats at both time points.

Since the 'traditional' pattern was not repeated in the 4-year data, this component was not included in any further comparative analyses. Table 3 presents the correlations between the dietary patterns. The first row compares the distinct scores obtained over the 4-year period while the second row correlates the applied scores. It can be seen that the correlations are similar though slightly larger across the scores using the applied method. Table 3 also presents the mean scores with paired t tests to assess the change in scores, for each of the dietary components. The 4-year applied scores are considerably lower on average than those obtained from the separate 4-year PCA, however the standard deviations are much larger for the former, reflecting the inflated maximum values using this method. Using the scores obtained from the two separate PCA it can be seen that, overall, women decreased their scores on the 'health conscious' component (mean difference -0.284; $P < 0.0001$). The 'processed' and 'vegetarian' scores both increased over the 4-year period (mean differences 0.042 ($P < 0.0001$) and 0.024 ($P = 0.021$) respectively). In contrast, applying the factor loadings from the pregnancy data resulted in systematic decreases in all scores at the later time point, with the biggest differences seen in the 'health conscious' and 'confectionery' pattern scores (mean differences -0.284 and -0.108; both $P < 0.0001$). Comparing scores from the pregnancy data with the applied 4-year data produced considerably wider limits of agreement than comparison with the separate 4-year scores suggesting greater variation in the applied scores compared to the separate scores from the 4-year data.

A reasonable level of agreement was seen between the categorized component scores from the pregnancy data and the 4-year data, ranging from 0.267 to 0.306 (Table 4). The values of κ were slightly higher for the applied 4-year data, with the exception of the 'vegetarian' component. The highest levels of agreement were seen been the sets of scores obtained from the 4-year data (ranging from 0.586 for the 'vegetarian' component to 0.707 for the 'health conscious' component).

Table 5 assesses the stability of the scores for two of the components 'health conscious' and 'processed' over time. The dietary pattern scores were split into quintiles and the Table presents the cross-tabulations of the quintiles from the 32-week data against the two sets of 4-year data. It can be seen that agreement was slightly better for the 'applied' score of the 'health conscious' component compared to that directly obtained from the 4-year data. For example 50·1% who were in the lowest quintile for the 'health conscious' component during pregnancy remained in this quintile at 4 years using the applied scores, compared to 45·5% using the separate score; while 48·3% of mothers remained in the top quintile using the applied scores compared to 37·5% using the separate score. This was not the case for the 'processed' component where the applied score appeared to be much less stable compared to the separate score. For every quintile of the 'processed' component in pregnancy, approximately

Table 3. Pearson's correlation coefficients, mean values and standard deviations, mean change and limits of agreement for dietary pattern scores obtained at 32 weeks gestation and when children were 4 years of age

	'Health conscious'		'Processed'		'Confectionery'		'Vegetarian'	
Correlations								
Pregnancy v. 4-year†	0.42		0.42		0.42		0.50	
Pregnancy v. applied 4-year†	0.56		0.44		0.45		0.50	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pregnancy score (min, max)	0.082 (−3.27, 10.42)	0.99	−0.052 (−3.02, 16.16)	0.95	0.008 (−2.13, 6.33)	0.97	−0.027 (−3.81, 14.41)	1.01
4-year score* (min, max)	0.007 (−2.57, 11.51)	0.99	−0.009 (−2.88, 11.84)	0.99	0.004 (2.81, 8.68)	1.00	−0.003 (−4.88, 10.36)	0.99
Applied 4-year score† (min, max)	−0.202 (−9.08, 31.28)	3.53	−0.071 (−6.10, 28.95)	3.11	−0.100 (−5.47, 21.17)	2.47	−0.100 (−11.12, 24.31)	2.26
4-year score* – Pregnancy score t	−0.075 6.70 (P<0.0001)	1.06	0.042 3.82 (P<0.0001)	1.05	−0.004 0.34 (P=0.737)	1.06	0.024 2.31 (P=0.021)	1.00
Applied 4-year score† – Pregnancy score t	−0.284 8.70 (P<0.0001)	3.09	−0.019 0.55 (P=0.590)	3.31	−0.108 4.12 (P<0.0001)	2.48	−0.073 2.67 (P=0.008)	2.56
Limits of agreement								
Pregnancy score v. 4-year score*	−2.20, 2.05		−2.05, 2.14		2.13, 2.12		−2.89, 2.95	
Pregnancy score v. applied 4-year score†	−6.46, 5.89		−6.63, 6.59		−5.08, 4.87		−5.60, 5.45	

* Scores obtained from separate principal components analysis.
† Score obtained by applying factor loadings obtained from pregnancy data.

Table 4. Weighted κ with 95 % CI for quintiles of dietary pattern scores obtained at 32 weeks gestation and when children were 4 years of age

	Pregnancy *v.* 4-year*		Pregnancy *v.* applied 4-year †		4-year* *v.* applied 4-year †	
	κ	95 % CI	κ	95 % CI	κ	95 % CI
'Health conscious'	0·306	0·304, 0·312	0·400	0·399, 0·415	0·707	0·704, 0·708
'Processed'	0·288	0·283, 0·292	0·305	0·296, 0·310	0·685	0·680, 0·691
'Confectionery'	0·280	0·275, 0·287	0·300	0·296, 0·315	0·674	0·672, 0·676
'Vegetarian	0·267	0·261, 0·276	0·263	0·255, 0·267	0·586	0·580, 0·593

* Scores obtained from separate principal components analysis.
† Score obtained by applying factor loadings obtained from pregnancy data.

80 % migrated to another quintile by 4 years, with, on average, even spread into the remaining four quintiles.

Discussion

We have reported two sets of dietary components obtained using PCA at two time points in the same cohort of women: one in late pregnancy and the other 4 years after the birth. We observed differences in the factorial structures of the components obtained at two time points such that the 'traditional' component was absent at the second time point. Despite this the loadings that were obtained for each set of components were highly comparable. However, additional high loadings were observed on the 'health conscious' component at four years; these were the foods which originally loaded highly on the 'traditional' component in pregnancy. The loss of a component extracted at the second time point can be partly explained by the fact that additional questions were used in the second questionnaire. However, given that the components that were obtained at 4 years were very similar to the other four remaining components from the pregnancy data, the labels applied were appropriate as were the resulting comparisons that we made.

It could be argued that using a pregnant cohort to obtain the first set of dietary components may explain the differences we reported. There is potential for dietary modification during this life event along with other major lifestyle changes. Ideally, pre-conception diet would be compared in parallel with the diet after the birth to determine how much of an effect pregnancy might have on a woman's diet but this data was not available to the study. Cuco *et al.* performed PCA on dietary data collected before conception, at four time points during pregnancy and 6 months postnatally[9]. This study obtained slightly different patterns at each time point but the authors concluded that the patterns did vary significantly over time. Although this study had the advantage of data collection at multiple time-points, it sampled only eighty women and used 7 d dietary records. Potentially PCA performed on data collected from dietary records could be more stable than using data collected from FFQ as we have done. The time lag after birth was very short compared to that reported here and had Cuco *et al.* extended their follow-up time more substantial changes may have become evident. A further limitation of this study is the loss to follow-up. A quarter of women who completed the pregnancy questionnaire failed to complete the follow-up questionnaire. There is potential for bias to have been introduced in the complete-case sample.

However, the women who remained in the study were unlikely to have changed the way they would have completed their FFQ; we doubt that these drop-outs would have a major effect on the dietary patterns obtained.

We found very little difference in the size of the correlations depending on whether the separate or applied scores were used at 4 years. However, the correlations between the two sets of scores in this study are generally lower than those reported by others who have identified quantitatively similar patterns. Borland *et al.*[8] reported correlations of 0·81 and 0·64 for their prudent and high energy patterns respectively in their study of ninety-four young women over a 2-year period. The prudent and western patterns identified by Hu *et al.*[10] had correlations of 0·70 and 0·67 1-year apart in their study of 127 men. However, other studies that had a greater time period (10 years or more) between their dietary assessments reported correlations smaller than ours[5,12]. It would appear from the literature that as the time lag increases the sizes of the correlations decrease. We agree with Weismeyer *et al.*[13] who suggested that: 'studies that are updated only after ≥ 7 years risk the validity of their dietary exposures because complex changes in the diet increase over time'.

In this study we found differences in the means and measures of agreement reported according to whether the applied or separate scores were calculated at the 4-year follow-up. The applied scores were considerably lower on average than those obtained from the separate 4-year PCA, with wider limits of agreement, suggesting greater variability in the applied scores. This is most likely explained by the fact that a greater number of foods were asked about in the 4-year FFQ. Differences were also evident in score changes over time. Based on the scores obtained from the two separate PCA, scores decreased on the 'health conscious' component. The 'processed' and 'vegetarian' scores both increased over the 4-year period. However, using the applied scores from the pregnancy data resulted in systematic decreases in all scores at the later time point. The explanation for this overall decrease is probably due to the fact that women already had lower 'applied' scores. Therefore, using 'applied' scores could be misleading when investigating the associations between real changes in dietary patterns over time and health outcomes. We would recommend that any future studies planning to use 'applied' scores consider performing PCA at each time point to compare separate scores and determine whether there are any apparent differences. Two studies have applied this method previously[12,15]. The earliest of these reported that the factor loadings at follow-up were almost

Table 5. Proportion (%) of women in each quintile of 'health conscious' and 'processed' dietary pattern score obtained at 32 weeks gestation who remained in that quintile when their children were 4 years of age

	Pregnancy dietary component score quintiles									
	4-year dietary component score quintiles					Applied 4-year dietary component score quintiles				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
'Health conscious'										
Quintile 1	45.5	27.5	15.7	8.6	3.7	50.1	24.8	14.6	7.0	3.6
Quintile 2	25.6	24.8	22.5	18.4	8.8	27.8	28.6	21.3	14.6	7.7
Quintile 3	13.8	19.8	23.9	22.8	19.7	14.1	24.1	25.2	21.9	14.7
Quintile 4	9.8	14.9	21.1	24.0	30.3	5.7	15.4	24.7	28.5	25.6
Quintile 5	6.4	13.0	16.9	26.2	37.5	2.3	7.2	14.1	28.0	48.3
'Processed'										
Quintile 1	40.5	25.5	18.7	9.8	5.6	20.3	20.8	18.3	20.4	20.2
Quintile 2	24.3	25.5	22.8	16.1	11.3	21.4	20.6	21.1	18.3	18.7
Quintile 3	16.6	20.5	22.9	24.1	15.9	20.8	20.3	20.5	18.4	19.9
Quintile 4	11.6	16.6	20.2	26.0	25.6	16.5	21.3	18.8	21.9	21.5
Quintile 5	7.0	11.9	15.5	24.0	41.6	20.9	17.0	21.4	21.0	19.7

identical to those derived at baseline but did not discuss the implications of adopting these baseline scores[12]. The later study used a slightly different approach by creating scores based on the number of food items consumed (as opposed to the quantity or frequency of consumption) as reported in 5 d food diaries, entered into a factor analysis at second follow-up[15]. The factor loadings from this cross-sectional analysis were then employed to derive scores at baseline and first follow-up. This approach is appropriate when dietary records are used as it enables a large number of foods to be included from the start. Applying this method to data collected via FFQ would not be as beneficial since the breadth of detail collected by FFQ is limited.

There are theoretical differences between the two methods that we have employed. Use of the 'applied' scores enables the researcher to assess the stability of the dietary patterns that were identified at baseline; however, any new patterns that may have emerged at follow-up would not be identified. Using the two sets of loadings that emerged from separate PCA at each time point enables the researcher to identify the important patterns at each time point. However, if these patterns differed significantly it would not be possible to assess the stability of the patterns obtained at baseline.

In conclusion, the four components obtained from the two time points in this study were mostly consistent in terms of the foods that loaded highly and the sizes of those loadings. However, the loss of one component across the time period meant that exact comparisons could not be made. In this cohort using the 'applied' method to obtain scores at the second time point was not appropriate. We found that there were changes in the overall scores over the 4-year period of follow-up when the scores were computed using separate PCA at the second time point. Work is currently underway in this cohort to determine the factors driving these changes in dietary pattern scores.

Acknowledgements

References

1. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
2. Martinez ME, Marshall JR & Sechrest L (1998) Invited Commentary: factor analysis and the search for objectivity. *Am J Epidemiol* **148**, 17–19.

British Journal of Nutrition

NS

3. 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.

4. Fung TT, Schulze M, Manson JE, Willett WC & Hu FB (2004) Dietary patterns, meat intake and the risk of Type 2 diabetes in women. *Arch Intern Med* **164**, 2235–2240.

5. Newby P, Weismayer C, Akesson A, Tucker K & Wolk A (2006) Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. *J Nutr* **136**, 2580–2587.

6. Schulze M, Fung T, Manson J, Willett W & Hu F (2006) Dietary patterns and changes in body weight in women. *Obesity* **14**, 1444–1453.

7. Van Dam RM, Rimm EB, Willett WC, Stampfer MJ & Hu FB (2002) Dietary patterns and the risk for type 2 diabetes mellitus in US men. *Ann Intern Med* **136**, 201–209.

8. Borland SE, Robinson SM, Crozier SR, Inskip HM & SWS Study Group (2007) Stability of dietary patterns in young women over a 2-year period. *Eur J Clin Nutr*, AOP: doi:10.1038/sj.ejcn.1602684.

9. Cuco G, Fernandez-Ballart J, Sala J, Viladrich C, Iranzo R, Vila J & Arija V (2006) Dietary patterns and associated lifestyles in preconception, pregnancy and postpartum. *Eur J Clin Nutr* **60**, 364–371.

10. Hu FB, Rimm E, Smith-Warner SA, Feskanich MJ, Stampfer MJ, Ascherio A, Sampson L & Willett WC (1999) Reproducibilty and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* **69**, 243–249.

11. 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.

12. Mikkila V, Rasanen L, Raitakari OT, Pietinen P & Viikari J (2005) Consistent dietary patterns identified from childhood to adulthood: the cardiovascular risk in young Finns Study. *Br J Nutr* **93**, 923–931.

13. Weismayer C, Anderson JG & Wolk A (2006) Changes in the stability of dietary patterns in a study of middle-aged Swedish women. *J Nutr* **136**, 1582–1587.

14. Prevost AT, Wichelow MT & Cox BD (1997) Longitudinal dietary changes between 1984–5 and 1991–2 in British adults: associations with socio-demographic, lifestyle and health factors. *Br J Nutr* **78**, 873–888.

15. Mishra G, McNaughton SA, Bramwell GD & Wadsworth MEJ (2006) Longitudinal changes in dietary patterns during adult life. *Br J Nutr* **96**, 735–744.

16. Golding J, Pembrey M & Jones R and the ALSPAC Study Team (2001) ALSPAC - The Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol* **15**, 74–87.

17. Rogers I & Emmett P and the ALSPAC Study Team (1998) Diet during pregnancy in a population of pregnant women in South West England. *Eur J Clin Nutr* **52**, 246–250.

18. Gregory J, Foster K, Tyler H & Wiseman M (1990) Classification and types of diet. In *The Dietary and Nutritional Survey of British Adults. Office of Population Censuses and Surveys*, chapter 13. London: HMSO.

19. Williams C, Birch EE, Emmett PM & Northstone K and the ALSPAC 'Children in Focus' Study Team (2001) Steroacuity at 3·5 years of age in children born full term is associated with prenatal and postnatal dietary factors; an early report from a population based study. *Am J Clin Nutr* **73**, 316–322.

20. Gorsuch RL (1974) *Factor Analyses*. Philadelphia: W.B.Saunders.

21. Kline P (1994) *An Easy Guide to Factor Analysis*. London: Routledge.

22. Northstone K, Emmett P & Rogers I (2007) Dietary patterns in pregnancy and associations with socio-demographic and lifestyle factors. *Eur J Clin Nutr*, AOP: doi:10.1038/sj.ejcn.1602741.

23. Cattell RB (1966) The scree test for the number of factors. *Multivariate Behav Res* **1**, 245–276.

24. Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.

25. Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**, 213–220.