British Journal of Nutrition (2000), **83**, 257–266

A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome

Desmond E. M. Williams, A. Toby Prevost, Margaret J. Whichelow, Brian D. Cox, Nicholas E. Day and Nicholas J. Wareham*

Department of Community Medicine, University of Cambridge, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK

(Received 29 December 1998 – Revised 10 August 1999 – Accepted 30 August 1999)

Previous epidemiological studies have demonstrated relationships between individual nutrients and glucose intolerance and type 2 diabetes, but the association with the overall pattern of dietary intake has not previously been described. In order to characterize this association, 802 subjects aged 40–65 years were randomly selected from a population-based sampling frame and underwent a 75 g oral glucose-tolerance test. Principal component analysis was used to identify four dietary patterns explaining 31·7 % of the dietary variation in the study cohort. These dietary patterns were associated with other lifestyle factors including socio-economic group, smoking, alcohol intake and physical activity. Component 1 was characterized by a healthy balanced diet with a frequent intake of raw and salad vegetables, fruits in both summer and winter, fish, pasta and rice and low intake of fried foods, sausages, fried fish, and potatoes. This component was negatively correlated with central obesity, fasting plasma glucose, 120 min non-esterified fatty acid and triacylglycerol, and positively correlated with HDL-cholesterol. It therefore appears to be protective for the metabolic syndrome. Component 1 was negatively associated with the risk of having undiagnosed diabetes, and this association was independent of age, sex, smoking and obesity. The findings support the hypothesis that dietary patterns are associated with other lifestyle factors and with glucose intolerance and other features of the metabolic syndrome. The results provide further evidence for the recommendation of a healthy balanced diet as one of the main components of chronic disease prevention.

Type 2 diabetes: Dietary patterns: Principal component analysis

Epidemiological studies have a key role in generating greater clarity about the relationship between exposures and disease so that their findings can be translated to public health recommendations. Many population studies of the relationship between diet and chronic diseases have been analysed using the intakes of specific nutrients or food groups of individuals (Hamman, 1992). However, in the absence of data accurate enough to eliminate confounding, it is uncertain whether the observed association is a singular effect of the food or nutrient, or if the food or nutrient acts as a marker for an overall lifestyle and dietary pattern. This is important as the intake of one food or nutrient is often correlated with the intake of another (Barker *et al*. 1990; Kant *et al*. 1995) or with other lifestyle factors (Margetts & Jackson, 1993), which may have a relationship to the risk status of the individual.

Some researchers have attempted, therefore, to examine the relationship between food patterns and disease status rather than individual nutrients (Huijbregts *et al*. 1995; Kant *et al*. 1995; Whichelow & Prevost, 1996; Gittelsohn *et al*. 1998). Three main methods have been used: cluster analysis, scores and principal component analysis (PCA). In this present paper, we report the use of PCA, a means of reducing the complexity of diet to a small number of dimensions. This is analogous to the reduction of dietary complexity to the intake of a few nutrients. However, it has the advantage that the dietary components identified are independent of each other, unlike the intake of different nutrients, and hence confounding between components is diminished. It has the possible disadvantage that it provides no insight into the mechanisms by which diet is associated with disease risk, being driven by patterns of food intake. However, as people consciously consume patterns of food intake rather than individual nutrients, it may be that aetiological information based on patterns could be easier to translate into public health recommendations.

Abbreviations: IGT, impaired glucose tolerance; NEFA, non-esterified fatty acid; PCA, principal component analysis.

* **Corresponding author:** Dr N. J. Wareham, fax +44 (0)1223 330330, email njw1004@medschl.cam.ac.uk

In the UK several studies have used the PCA technique to derive dietary patterns (Barker *et al*. 1990; Whichelow & Prevost, 1996) that were then shown to relate cross-sectionally to other lifestyle factors (Barker *et al*. 1990; Whichelow & Prevost, 1996; Johnson *et al*. 1998) and prospectively to the health circumstance of the subjects (Prevost *et al*. 1997).

The metabolic syndrome, otherwise known as syndrome X or the insulin resistance syndrome, is a cluster of cardiovascular risk factors. These risk factors include hypertriacylglycerolaemia, lowered HDL-cholesterol, glucose intolerance and hypertension (Haffner *et al*. 1992). These abnormalities are often correlated with insulin resistance, and it has been suggested that insulin resistance may be the underlying cause for this clustering (Reaven, 1988). This syndrome has also been shown to be related to central obesity, physical inactivity (Wareham *et al*. 1998), and some nutrient factors (Daly *et al*. 1997; Purnell & Brunzell, 1997). However, few studies have examined the relationship between food patterns and the metabolic syndrome. A study conducted in Canada (Gittelsohn *et al*. 1998) investigated the relationship between broad patterns of food intake, method of preparation and consumption derived from factor analysis and glucose intolerance in a native Canadian population. It showed that some food groups including vegetables, breakfast foods, and hot meals appeared to have a protective effect for impaired glucose tolerance (IGT) and type 2 diabetes.

This present study aims to describe dietary patterns in a middle-aged UK-population-based cohort and to relate these patterns to glucose intolerance and other features of the metabolic syndrome.

Materials and methods

Study population and sampling methods

The Isle of Ely study is a prospective population-based study of the aetiology and pathogenesis of type 2 diabetes and related metabolic disorders. The details of this study have been published elsewhere (Williams *et al*. 1995). In summary, 1122 subjects (486 men and 636 women) aged 40–64 years, who were not known to have diabetes, were randomly selected from a population-based sampling frame. People with previously diagnosed diabetes, those who had moved away from the study area before being contacted, or who were physically unable to come to the study centre were excluded from the study. The response rate was 74 %. Respondents underwent a standard 75 g oral glucose-tolerance test according to World Health Organization recommendations (World Health Organization, 1985) after a 10 h overnight fast and abstinence from smoking, use of a nicotine patch, or chewing nicotine gum on the morning of the test. Blood samples were immediately separated in a cooled centrifuge at 4° and transported to the main laboratory on ice, where they were stored at −70°. Plasma glucose was measured using the hexokinase method (Kunst *et al*. 1983) and total serum cholesterol, HDL and triacylglycerol using the RA 1000 (Bayer Diagnostics, Basingstoke, Hants., UK). Values for LDL-cholesterol concentrations were calculated using the Friedewald formula (Friedewald *et al*. 1972). Estimations were not made if the plasma triacylglycerol was >4 mmol/l. Plasma insulin was measured using an immunoenzymic assay (Medgenix Diagnostics SA, Fleurus, Belgium).

Two trained observers carried out standard anthropometric measurements (Williams *et al*. 1995). Smoking status and alcohol intake were assessed by questionnaire. Subjects were classified as non-, ex- or current smokers. Alcohol consumption was recorded as units of alcohol consumed per week and analysed as a continuous variable and as quartiles of alcohol consumption. Leisure time physical activity was assessed using the Paffenbarger questionnaire (Paffenbarger *et al*. 1993; Wareham *et al*. 1996) with standard coding of specific activities using published energy costs for recreation and sports (Ainsworth *et al*. 1993). These are expressed in the data as metabolic equivalents or MET hours per week, or multiples of the energy cost of resting. Socio-economic group was assessed according to the shortened form of the Registrar General's socio-economic classification (Registrar General, 1980). A collapsed form of this classification was used in which individuals were grouped into non-manual (1, 2, 3) and manual (4, 5, 6) groups. The World Health Organization criteria for diabetes and IGT were used to classify subjects (World Health Organization, 1985). Waist : hip ratio was categorized into sex-specific quartiles. The Cambridge Local Research Ethics Committee approved the study.

Measurement of main exposure

Food consumption was assessed using a modification of the Health and Lifestyle survey questionnaire (Cox *et al*. 1987). There were thirty-five items in the food frequency questionnaire. The questions on frequency of food consumption were taken from those used in similar surveys (Bull, 1965; Yarnell *et al*. 1983; Cox *et al*. 1987). Yarnell *et al*. (1983) reported on the performance of the food frequency questions against weighed intakes. The results from these questions have been shown to discriminate between groups of subjects (Whichelow *et al*. 1991*a*, *b*) and to associate with lifestyle factors (Whichelow & Prevost, 1996). The Health and Lifestyle survey questionnaire was administered as a structured self-completed questionnaire, completed by the study subjects during their visit to the study centre, and was designed to assess food intake over the preceding 12 months. For the main food list, the participants were asked to grade their frequency of consumption of each food item as 'daily', 'most days', 'once or twice a week', 'less than once a week', or 'rarely or never'. Questions aimed at measuring the seasonal variation of salad and raw vegetables and fruits consumption in winter and summer were included in the main food list. Total energy intake was not estimated. Due to missing values in some of the food items, twenty-nine of the thirty-five food items were selected for inclusion in the PCA. The food items excluded due to missing values included cereals, bread type, spreads, preserves, cream, and beverages. Complete information for all twenty-nine food items was given by 802 subjects.

Table 1. Factor loadings of the food items in the four principal dietary components identified; the Ely study (*n* 802)*

	Dietary component			
	1	2	3	4
Variation explained (%)	11·87	7·5	6·63	5·67
Fruit (summer)	*0·686*	0·257	0·193	–0·043
Fruit (winter)	*0·685*	0·223	0·229	–0·022
Salad (summer)	*0·647*	0·056	0·061	0·136
Salad (winter)	*0·636*	0·052	0·154	0·169
Fish (not fried)	*0·453*	0·005	–0·005	0·012
Other vegetables	*0·410*	0·217	–0·258	0·236
Poultry	*0·327*	0·047	–0·184	–0·035
Green vegetables	*0·320*	*0·356*	*–0·563*	–0·054
Pasta/rice	*0·304*	–0·082	0·181	0·217
Ice cream/light desserts	0·280	0·250	0·173	–0·273
Root vegetables	0·143	*0·440*	*–0·587*	–0·068
Nuts	0·121	0·173	0·148	0·250
Cheese	0·107	*0·302*	0·241	*0·366*
Pulses	0·093	*0·377*	*–0·301*	0·113
Tinned fruit	0·091	0·171	0·092	–0·174
Milk, amount	–0·002	0·073	–0·020	0·115
Soft drinks	–0·029	0·211	0·237	0·029
Chocolate	–0·051	*0·337*	*0·445*	–0·017
Eggs	–0·066	0·262	–0·069	*0·594*
Sweets	–0·169	*0·442*	*0·408*	–0·071
Red meat	–0·174	0·245	–0·259	0·062
Puddings/pies	–0·178	*0·391*	0·025	–0·196
Biscuits	–0·226	*0·415*	0·220	*–0·369*
Crisps	*–0·309*	0·241	0·268	0·199
Cakes	*–0·310*	*0·479*	0·136	*–0·326*
Potatoes	*–0·322*	*0·401*	*–0·393*	–0·131
Fried fish	*–0·361*	0·143	0·058	*0·301*
Sausage, processed meat	*–0·376*	0·162	0·083	*0·382*
Fried food	*–0·382*	0·120	–0·082	*0·468*

* Figures in italics represent loadings greater than 0·300 or less than –0·300.

Analysis

Principal component analysis. The data were analysed using Intercooled Stata 5 (Stata Statistical Software Release 5.0, 1997; Stata Corporation, College Station, TX, USA). Data on the frequency of consumption of foods in the main list were collected in five categories and re-coded to 'daily', 'most days', 'once or twice a week', 'less than once a week', and 'rarely or never'. The group of individuals with complete data was not significantly different in respect of age, sex, and prevalence of IGT or type 2 diabetes from those excluded because of missing values (DEMW1). PCA was performed on the correlation matrix of the twenty-nine food items listed in Table 1. This is a data reduction technique that produces successive linear combinations of the variables each accounting for as much of the remaining variance as possible. It therefore describes major food patterns in the data. The FACTOR (pc) command (Intercooled Stata 5) was used to extract the factor loadings for the components and factor scores calculated for each individual based on their initial responses. A higher score indicated that the individual ate the food pattern described by the component more frequently than a person with a lower score. The stability of the generated components was tested by deriving the components for two randomly split sub-samples and correlating the factor loadings. The first four components were highly correlated to their corresponding component derived from each sample but the fifth and sixth components were not, and were then excluded from the remaining analysis. This observation was confirmed by the Scree plot shown in Fig. 1. The Scree plot is a plot of the eigenvalues of the derived components. The occurrence of

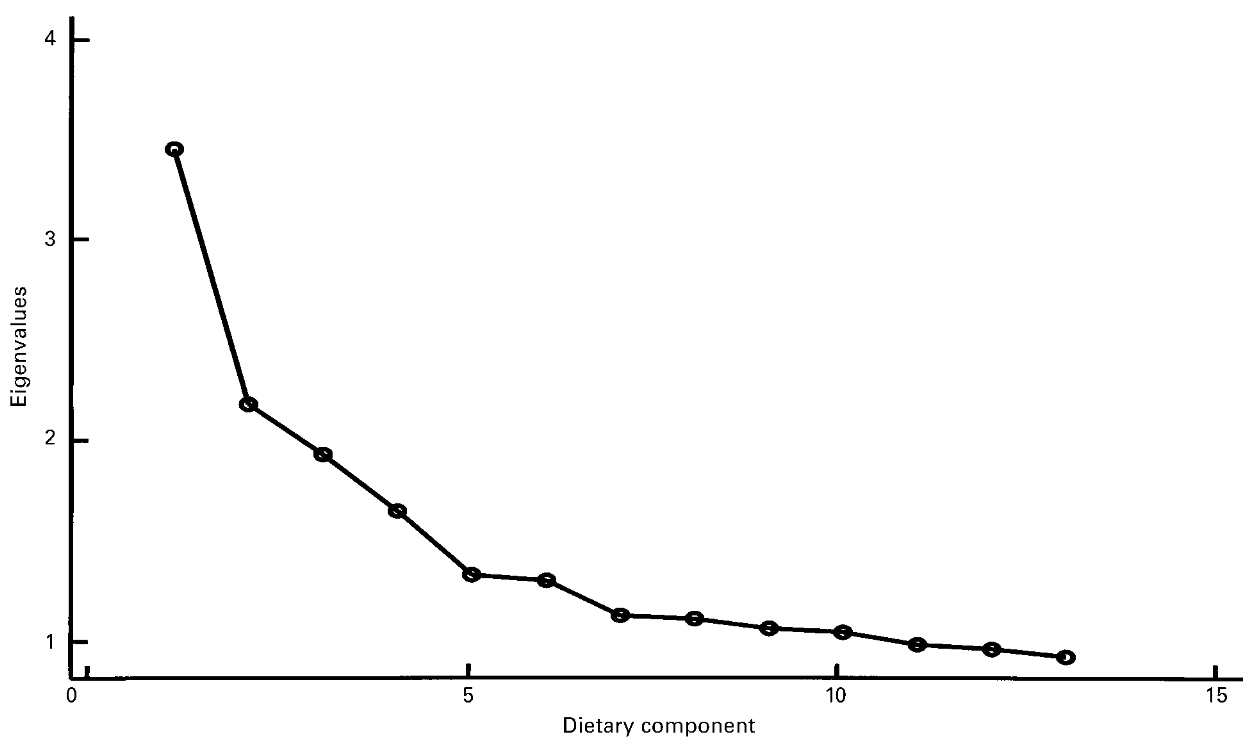

Fig. 1. Scree plot of the eigenvalues of the first fifteen components generated by principal component analysis; the Ely study (*n* 802). For details of study population, see Table 2; for details of principal component analysis, see pp. 259–260.

an elbow or knuckle is usually considered to be where the large eigenvalues stop and the small eigenvalues start and can be used as a visual representation of the number of significant components (Cattell, 1965, 1966). In Fig. 1, two elbows were observed between components 1 and 2, and components 4 and 5. The χ^2 test of isotropic variation (Mardia *et al*. 1979), another technique used in identifying the number of components to retain, also confirmed this observation. This test indicated a significant difference between components 1 and 2 ($P < 0{\cdot}001$) and components 3 and 4 ($P = 0{\cdot}07$) but no difference between components 2 and 3. Both these tests suggested that one or four components should be retained. The first four components were thus retained to describe the dietary patterns observed in this cohort. The relationship between the general characteristics of the subjects and glucose intolerance status was assessed using Student's *t* test and Pearson's Chi-square test (χ^2) to compare the arithmetic means and proportions of these characteristics in the IGT and type 2 diabetes groups against those of the normal glucose tolerance group. ANOVA was used to examine the characteristics of subjects for each of the components identified. The scores of component 1 were categorized into tertiles using the distribution in the normal glucose tolerance group. Spearman's correlation and partial correlation were used to assess the relationship between the dietary components and the variables of the metabolic syndrome. Insulin, 120 min plasma glucose and triacylglycerol concentrations, total alcohol intake and physical activity were skewed and were normalized by logarithmic transformation before the partial correlations were calculated.

The prevalence of IGT and type 2 diabetes were directly age-standardized to the World population of Segi using the truncated standard age range as recommended by King & Rewers (1993). Univariate logistic regression was used to investigate the relationship between the dietary components, IGT and type 2 diabetes as separate outcomes, compared with the normal glucose tolerance group. The effects of potential confounders on the relationship between these dietary components and glucose intolerance were assessed in a series of multiple logistic regression models.

Results

Fifty-one (4·5 %) of the subjects had newly diagnosed type 2 diabetes by the World Health Organization (1985) criteria and 188 (16·8 %) had IGT. The age-standardized prevalence was 2·32 and 11·2 % for previously undiagnosed type 2 diabetes and IGT respectively.

The general characteristics of the three glucose tolerance categories are shown in Table 2. The mean age and BMI of the study population was 54·1 (SD 7·7) years and 26·0 (SD 4·3) kg/m^2 respectively. Subjects with newly diagnosed type 2 diabetes were older, more obese, and less physically active than their normal counterparts, with those with IGT being intermediate. There was a strong significant trend between reported alcohol consumption and the prevalence of IGT and type 2 diabetes, with the prevalence decreasing as reported alcohol consumption increased (χ^2 19·07, $P < 0{\cdot}01$). A greater proportion of subjects with IGT and type 2 diabetes reported a positive family history (at least one or more parent or sibling) than normal subjects. The prevalence of type 2 diabetes and IGT was highest in unskilled and skilled manual workers, socio-economic

Table 2. Anthropometric and lifestyle characteristics by World Health Organization glucose-tolerance category†; the Ely study (*n* 802)

(Mean values and standard deviations in parentheses)

	Normal glucose tolerance	Impaired glucose tolerance	Type 2 diabetes
n (men)	311	54	14
n (women)	334	74	15
Age (years)	52·5 (7·6)	56·1** (7·3)	59·0** (5·1)
BMI (kg/m^2)	25·4 (3·8)	27·5** (4·7)	30·2** (4·4)
Waist : hip ratio			
Men	0·90 (0·05)	0·92* (0·06)	0·96*** (0·06)
Women	0·76 (0·05)	0·78*** (0·06)	0·81*** (0·06)
Reported alcohol consumption (units/week)	5·1 (8·4)	3·7* (7·6)	4·3 (11·4)
Subjects reporting a positive family history of diabetes (%)	16·9	24·2	27·6
Physical inactivity (%) (MET h/week of LTPA)			
<1	35·0	39·8	55·6
1–14	34·1	25·0	33·3
⩾ 15	30·8	35·2	11·1
Smoking status (%)			
Non-smokers	51·6	54·0	41·4
Current smokers	15·9	8·9	10·3
Ex-smokers	32·5	37·1	48·3

MET, metabolic equivalents; LTPA, leisure time physical activity.

Mean values were significantly different from the normal glucose tolerance group: * $P < 0{\cdot}05$, *** $P < 0{\cdot}001$ (Student's *t* test).

† World Health Organization (1985).

group 3 and 6, but this difference was not statistically significant (χ^2 17·37, $P = 0·066$). When the collapsed socio-economic groups were used, those in the manual group had a significantly higher prevalence of IGT and type 2 diabetes than those in the non-manual group (χ^2 11·14, $P < 0·01$).

Four dietary components were identified, collectively explaining 31·7 % of the variation in diet in the study population (Table 1). Component 1 explained 11·9 % of the variation and represented frequent intakes of salad and green vegetables and fruits (in both summer and winter), fish (not fried), pasta or rice, and low intake of fried foods, processed meat, fried fish, potatoes and snacks (biscuits and crisps). Component 2 explained 7·5 % of the variation and included frequent intake of confectionery, cheese, root and green vegetables, pulses, biscuits, chocolate, red meat and potatoes. Component 3 explained 6·6 % of the variation and included frequent intakes of confectionery and chocolate as in component 2 but infrequent intakes of vegetables and potatoes. Component 4 explained 5·7 % of the variation in the data and represented frequent intake of eggs, fried food, processed meat, cheese and fried fish (Table 1).

Higher scores for component 1 were observed in women, subjects who were less than 55 years and in those in the higher socio-economic groups, and lower scores were recorded in current smokers (Table 3). This component was negatively correlated with age, waist : hip ratio,

Table 3. Mean scores for each of the dietary components and lifestyle and anthropometric variables; the Ely study (*n* 802)

	n	Dietary component 1	2	3	4
Sex					
Male	379	–0·633	0·108	–0·063	0·116
Female	423	0·567	–0·097	0·057	–0·104
F test		0·000	0·049	0·223	0·015
Age group (year)					
40–44	149	0·081	0·235	0·088	0·231
45–49	163	0·105	–0·041	0·196	0·180
50–54	149	0·181	–0·014	–0·065	0·147
55–59	135	–0·183	–0·101	–0·090	–0·254
60–64	185	–0·113	–0·070	–0·085	–0·226
65–69	21	–0·502	0·021	–0·352	–0·456
F test		0·331	0·421	0·237	0·000
Socio-economic group					
Professionals	70	0·730	0·118	–0·040	0·462
Managers/executives	283	0·311	–0·162	–0·032	–0·042
Other non-manual	116	0·100	0·014	0·295	–0·031
Skilled manual	191	–0·544	0·130	–0·024	0·044
Semi-skilled manual	82	–0·520	0·157	–0·029	–0·172
Unskilled manual	29	–0·813	–0·397	–0·877	–0·153
F test		0·000	0·133	0·004	0·041
BMI (kg/m²)					
< 25	380	0·208	0·081	0·047	–0·095
25–27	160	–0·384	–0·216	–0·029	0·034
> 27	261	–0·070	0·011	–0·047	0·116
F test		0·002	0·100	0·668	0·115
Waist : hip ratio quartile					
1	216	0·238	0·147	0·109	–0·143
2	170	–0·173	0·051	–0·077	–0·026
3	214	–0·015	–0·011	0·100	0·042
4	198	–0·083	–0·210	–0·135	0·123
F test		0·142	0·096	0·185	0·187
Alcohol intake category (units/week)					
None	185	0·008	0·102	0·117	–0·201
> 0–5	338	0·045	0·055	0·060	–0·147
>5–10	114	0·129	0·029	–0·131	0·246
>10	136	–0·085	–0·289	–0·212	0·433
F test		0·828	0·090	0·108	0·000
Smoking status (%)					
Non-smoker	401	0·151	0·102	0·070	–0·179
Current	113	–0·768	–0·640	–0·273	0·325
Ex-smoker	263	0·169	0·095	0·010	0·112
F test		0·000	0·000	0·069	0·000
Physical inactivity (%) (MET h/week of LTPA)					
0	245	–0·220	0·062	0·072	0·057
1–14	220	0·216	0·004	0·062	–0·001
⩾15	209	0·151	0·020	0·077	0·045
F test		0·021	0·909	0·994	0·879

MET, metabolic equivalents; LTPA, leisure time physical activity.

Table 4. Spearman rank correlation coefficients of the four dietary components with lifestyle, anthropometric and metabolic variables; the Ely study (*n* 802)

	Dietary component			
	1	2	3	4
Age (years)	–0·07*	–0·06	–0·07	–0·16**
BMI (kg/m^2)	–0·06	–0·02	0·00	0·07
Waist : hip ratio	–0·29**	0·00	–0·06	0·11**
Physical activity (total MET)	0·02	–0·03	–0·12*	0·03
Total alcohol intake (units/week)	–0·01	–0·05	–0·07*	0·23**
Fasting plasma glucose (mmol/l)	–0·10**	–0·03	–0·07*	0·02
120 min plasma glucose (mmol/l)	–0·05	–0·09*	0·01	0·00
Fasting plasma insulin concentration (pmol/l)	–0·04	–0·05	–0·01	0·06
120 min plasma insulin concentration (pmol/l)	–0·01	–0·10**	0·03	0·03
Cholesterol (mmol/l)	–0·05	–0·03	0·00	–0·09*
Plasma HDL-cholesterol (mmol/l)	0·16**	–0·01	0·00	0·03
Plasma LDL-cholesterol (mmol/l)	–0·07	–0·03	0·02	–0·11**
Fasting plasma non-esterified fatty acids (mmol/l)	0·05	–0·12**	0·06	–0·06
120 min plasma non-esterified fatty acids (mmol/l)	–0·07*	–0·02	–0·08*	0·00
Plasma triacylglycerol (mmol/l)	–0·19**	–0·07	–0·05	–0·03

MET, metabolic equivalents.
* $P < 0·05$, ** $P < 0·01$.

fasting plasma glucose, 120 min non-esterified fatty acids (NEFA), and triacylglycerol (Table 4). It was positively correlated with HDL-cholesterol. After adjusting for age, component 1 remained negatively correlated with waist : hip ratio and triacylglycerol and positively correlated with HDL-cholesterol (Table 5). Younger subjects and men had higher component 2 scores, but heavy drinkers and current smokers had lower scores (Table 3). Component 2 scores were negatively correlated with fasting NEFA, 120 min insulin and plasma glucose concentration (Table 4). The correlation observed between component 2 and 120 min insulin and fasting NEFA was independent of the effects of age (Table 5). Subjects in non-manual occupations favoured component 3. Component 3 was negatively correlated with fasting plasma glucose, physical activity, total alcohol consumption, and 120 min NEFA concentration (Table 4), although with the exception of physical activity these correlations were not independent of the effect of age (Table 5). Men had higher scores on component 4 as did professionals and skilled manual workers, the younger age groups, heavy drinkers and current smokers. Component 4 was negatively correlated with age and cholesterol (both total and LDL), but positively correlated with waist : hip ratio and total alcohol intake (Table 4). Waist : hip ratio, total alcohol intake and BMI were positively associated with component 4 after adjusting for the effects of age (Table 5).

Subjects with newly diagnosed diabetes had lower scores for component 1 than those who were normal ($P < 0·05$) or had IGT ($P = 0·06$) (Table 6). No significant difference was seen between the three glucose tolerance categories for the other three components. Component 1 scores were categorized into tertiles based on the distribution of the normal glucose tolerance group. Fig. 2 shows the distribution of the

Table 5. Partial correlation coefficients of the four dietary components with lifestyle, anthropometric and metabolic variables adjusting for age; the Ely study (*n* 802)

	Dietary component			
	1	2	3	4
BMI (kg/m^2)	–0·05	–0·02	–0·01	0·09*
Waist : hip ratio	–0·29**	0·01	–0·04	0·15**
Physical activity (total MET)†	–0·02	–0·02	–0·14**	0·01
Total alcohol intake (units/week)†	–0·02	–0·07	–0·06	0·23**
Fasting plasma glucose (mmol/l)	–0·06	–0·03	–0·03	0·06
120 min plasma glucose (mmol/l)†	–0·05	–0·05	0·02	0·03
Fasting plasma insulin concentration (pmol/l)†	–0·03	–0·05	–0·03	0·05
120 min plasma insulin concentration (pmol/l)†	–0·01	–0·09*	0·02	0·06
Plasma cholesterol (mmol/l)	–0·04	0·00	0·02	–0·03
Plasma HDL-cholesterol (mmol/l)	0·19**	0·03	0·00	0·04
Plasma LDL-cholesterol (mmol/l)†	–0·06	0·01	0·04	–0·05
Fasting plasma non-esterified fatty acid (mmol/l)†	0·04	–0·09*	0·07	–0·03
120 min plasma non-esterified fatty acid (mmol/l)†	–0·07	–0·01	–0·07	0·01
Plasma triacylglycerol (mmol/l)†	–0·18**	–0·07	–0·06	0·03

MET, metabolic equivalents.
* $P < 0·05$, ** $P < 0·01$.
† Log transformed values.

Table 6. Dietary component scores by World Health Organization glucose-tolerance category†; the Ely study (*n* 802)

(Mean values and standard deviations)

	Normal glucose tolerance (*n* 645)		Impaired glucose tolerance (*n* 128)		Type 2 diabetes (*n* 29)	
Dietary component	Mean	SD	Mean	SD	Mean	SD
1	0·044	0·073	–0·068	0·164	–0·676*	0·344
2	0·039	0·058	–0·184	0·130	–0·044	0·274
3	–0·006	0·055	0·015	0·123	0·062	0·258
4	0·019	0·051	–0·098	0·113	0·004	0·238

Mean value was significantly different from that for group with normal glucose tolerance: * $P < 0·05$ (paired two-sided *t* test).
† World Health Organization (1985).

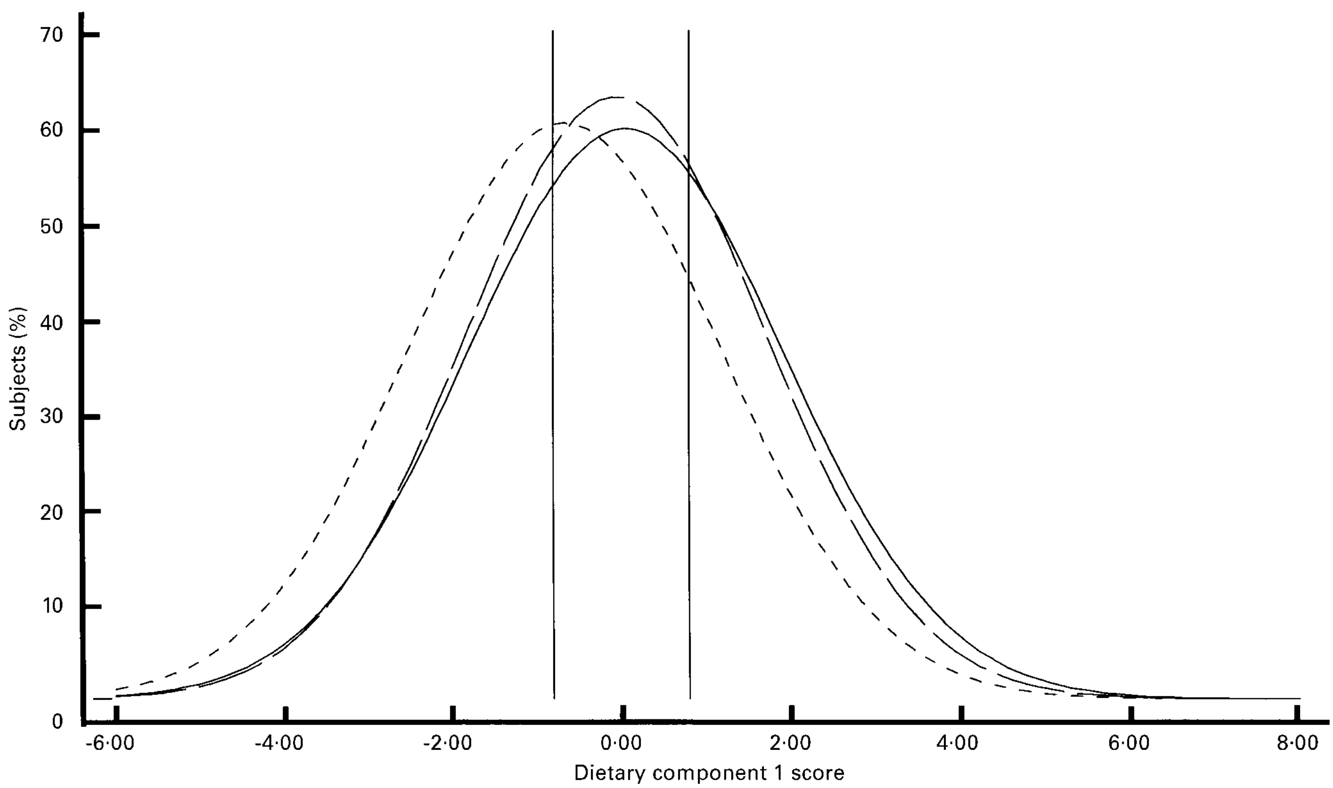

Fig. 2. Distribution of dietary component 1 scores by glucose-tolerance category; the Ely study. (———), Normal glucose tolerance (*n* 645); (— —), impaired glucose tolerance (*n* 128); (- - - - -), type 2 diabetes (*n* 29). For details of study population, see Table 2; for details of prinicipal component analysis, see pp. 259–260. The vertical lines represent the cut-off values defining tertiles in the normal glucose tolerance category

Table 7. Association between component 1 dietary score and previously undiagnosed type 2 diabetes; the Ely study (*n* 802)†

	All subjects (*n* 802)		Subjects ⩾50 years (*n* 396)	
	Odds ratio	95 % CI	Odds ratio	95 % CI
Unadjusted	0·62*	0·38, 0·99	0·55*	0·32, 0·92
Adjusted for age	0·65	0·40, 1·07	0·55*	0·32, 0·94
Adjusted for age, sex, BMI category and smoking status	0·54	0·32, 0·91	0·44*	0·22, 0·88

* $P < 0·05$.
† Reference category was subjects with normal glucose tolerance. Component 1 score was entered as tertiles based on the distribution in the normal glucose tolerance category. Age was entered in 5-year age bands, sex (1, 2), BMI category (<25, 25–29·9, ⩾30) and smoking status (non-, current and ex-smokers).

scores of those with IGT and type 2 diabetes around the tertiles of the normal glucose tolerance group. Almost half of those with type 2 diabetes had scores in the lowest tertile of the distribution. The odds ratio for having diabetes compared with normal glucose tolerance for a tertile change in component 1 was 0·62 (95 % CI 0·38, 0·999). After adjustment for age (entered as an ordered category in five-year age bands), there was a slight change in the point estimate (0·65, 95 % CI 0·40, 1·07) but the upper limit of the 95 % CI exceeded unity. To investigate further the relationship between this component and glucose intolerance, the analysis was restricted to those aged above 50 years. The odds ratio for a tertile change in component 1 was 0·54 (95 % CI 0·32, 0·91) and 0·44 (95 % CI 0·22, 0·88) after adjustment for age, sex, smoking status and obesity (Table 7).

Discussion

By using the PCA technique we have identified four dietary patterns in a UK-based cohort of middle-aged subjects, which show an association with other lifestyle factors and features of the metabolic syndrome. Of note is the striking similarity between the factor loadings observed in this study and those observed in previous UK-based population studies (Gregory *et al*. 1990; Whichelow & Prevost, 1996). The first component described by Whichelow & Prevost (1996) had similarly high loadings for fruits, salads, root vegetables and negative loadings for fried food and processed meat. The dietary and nutritional survey of British adults (Gregory *et al*. 1990) reported similarly high loadings in fruits and vegetables in their first component. However, they reported lower values for pasta, rice, and cereals. The association observed with other lifestyle factors is consistent with the results obtained in previous studies in the UK (Barker *et al*. 1990; Whichelow & Prevost, 1996).

One disadvantage of the PCA technique is the interpretation of the results obtained and the actual naming or describing of a particular food pattern. One approach is to use variables that load above a certain cut-off point (> ±3), but this does not distinguish variables that may have loadings greater than ±3 for different components. Another approach is to plot the loadings for one component against the next in order to identify the variables that discriminate between the two components (Fig. 3). Component 1 loaded heavily for foods that one would assume to be beneficial and came closest to the description of a conventionally accepted balanced healthy diet. With this is mind, it was not surprising to see that it was correlated with most of the features of the metabolic syndrome in the expected direction, i.e. negatively with obesity, plasma glucose, 120 min NEFA and triacylglycerol; and positively with HDL-cholesterol. It was also negatively related to glucose intolerance, with subjects found to have newly diagnosed type 2 diabetes having the lowest scores for component 1. A tertile change in component 1 was associated with a 28 % reduction in the risk of having type 2 diabetes that persisted after restricting the analysis to those above 50 years old and controlling for the effect of sex, age, BMI and smoking.

This result is similar to that found in a Canadian study (Gittelsohn *et al*. 1998) although in that study they included patients known to have diabetes, whose diet may have been modified by the onset of their disease. The present study, which was undertaken in individuals without known diabetes, is not subject to this bias. A similar analysis by Barker *et al*. (1992) identified a cosmopolitan dietary pattern with high loadings for fruits and vegetables that was negatively correlated with total cholesterol, and also

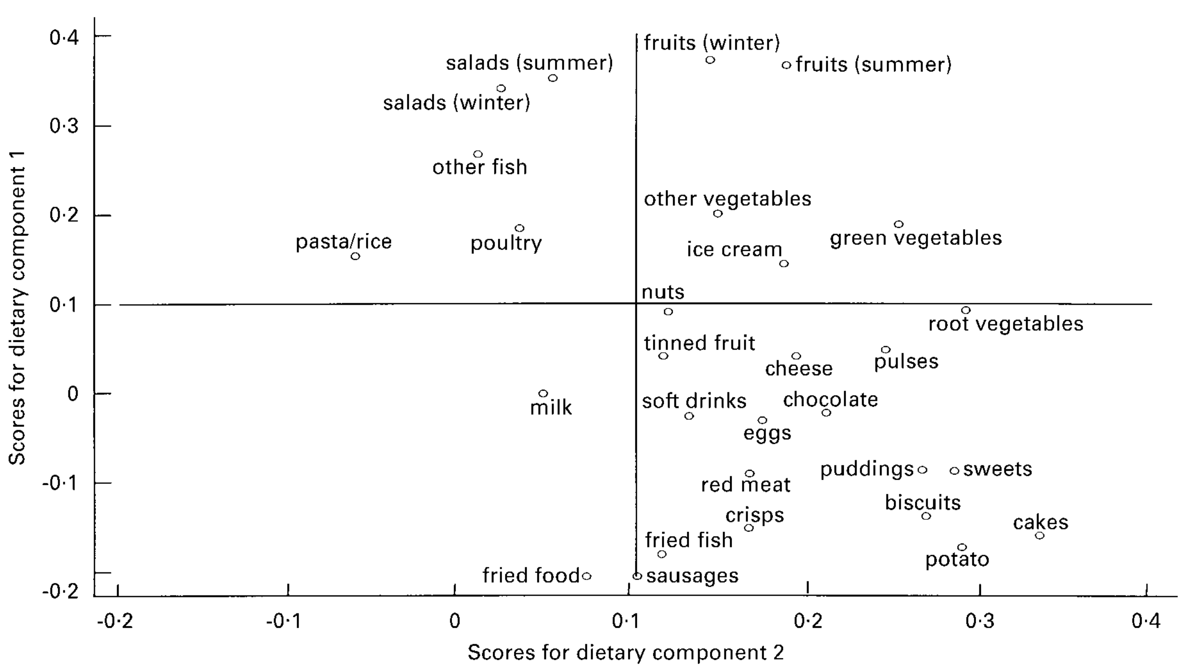

Fig. 3. Plot of the loadings for component 1 *v.* component 2 showing variables that discriminate between the two components. For details of study population, see Table 2; for details of principal component analysis, see pp. 259–260.

observed that dietary patterns containing alcohol were positively associated with HDL-cholesterol.

The metabolic syndrome has been shown to have three main clusters (Meigs *et al*. 1997): one including hyperlipidaemia, raised waist : hip ratio and BMI, and a second associated with hyperglycaemia and hyperinsulinaemia. Our results show two components inversely associated with these two clusters, i.e. component 1 with the lipid cluster and component 2 associated with the glucose cluster (Table 4). This may well indicate that the pathological mechanisms involved in these two clusters may be different and may be influenced by different environmental factors.

In a previous report on data from the Ely study (Williams *et al*. 1999), we have shown that frequent intakes of raw and salad vegetables, especially in the winter, were protective against type 2 diabetes (odds ratio 0·16, 95 % CI 0·04, 0·69). These results now indicate that those findings are probably, in part, a reflection of the overall healthy lifestyle described by component 1. However, the strength of the association with frequent intake of raw and salad vegetables, especially in the winter, appears too large to be explained completely by this relationship, although the confidence intervals are wide. Our present study cannot identify whether a single lifestyle factor is associated with this protective effect, as the dietary components were all associated with other lifestyle factors like physical activity and socio-economic group. Present analytical techniques cannot clearly separate one lifestyle factor from another in the assessment of risks for the occurrence of a particular disease, especially when the lifestyle factors are collinear and are measured imprecisely. It is therefore of interest for nutritional epidemiologists to examine dietary patterns as well as separate nutrient and food effects in the investigation of the effects of diet on disease processes.

We have placed more emphasis on component 1 in this paper because it explained the greatest proportion of the variation and was the most stable component in the two split samples. Furthermore, the factor loadings of component 1 were very similar to those observed in the Health and Lifestyle survey (Whichelow & Prevost, 1996). This emphasis was also justified by the results from the isotropy test which indicated that component 1 was significantly different from the other components. Although we could have omitted the reporting of the other three components we have elected to include their description for completeness but their aetiological relevance was less than that of component 1.

Only 802 subjects with complete information on twenty-nine of the thirty-five food items in the questionnaire were included in this study. In order to determine the effect this exclusion may have had on the derived components, the missing values for subjects who had less than three (*n* 947) and five (*n* 1037) missing food items were imputed and the PCA repeated with these imputed values. There was a very high correlation coefficient ($r > 0{\cdot}94$) between the scores of the components reported in this paper and those obtained using both the three missing items and five missing items imputed food items. The main conclusions of the relationship between component 1, glucose intolerance and the metabolic syndrome remained unchanged.

These results support current recommendations for the adoption of a dietary pattern which includes increased fruit and vegetable consumption and a reduction in the intake of fatty foods and processed meat for protection against the occurrence of the metabolic syndrome and glucose intolerance.

Acknowledgements

The Ely study was financed by the Anglia and Oxford RHA Research and Development Directorate, the MRC and British Diabetic Association. N. J. W. is an MRC Clinician Scientist Fellow. We are grateful to the staff and patients of the St Mary's Street Surgery, Ely and to H. Shannasy, J. Lipscombe, C. Byrne, T. Wang, V. Warren, and D. Brown for their help with the fieldwork.

References

Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montoye HJ, Sallis JF & Paffenbarger RS Jr (1993) Compendium of physical activities: Classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise* **25**, 71–80.

Barker ME, McClean SI, Strain JJ & Thompson KA (1992) Dietary behaviour and health in Northern Ireland: an exploration of biochemical and haematological associations. *Journal of Epidemiology and Community Health* **46**, 151–156.

Barker ME, McClean SI, Thompson KA & Reid NG (1990) Dietary behaviours and sociocultural demographics in Northern Ireland. *British Journal of Nutrition* **64**, 319–329.

Bull NL (1965) Dietary habits of 15 to 25 year olds. *Human Nutrition: Applied Nutrition* **38**A, 1–68.

Cattell RB (1965) Factor analysis: An introduction to essentials. *Biometrics* **21**, 190–215.

Cattell RB (1966) The Scree Test for the number of factors. *Multivariate Behavioral Research* **1**, 245–276.

Cox BD, Blaxter M, Buckle ALJ, Fenner NP, Golding JF, Gore M, Huppert FA, Nickson J, Roth M, Stark J, Wadsworth MEJ & Whichelow M (1987) *The Health and Lifestyle Survey: Preliminary Report*. Cambridge: Health Promotion Research Trust.

Daly ME, Vale C, Walker M, Alberti KG & Mathers JC (1997) Dietary carbohydrates and insulin sensitivity: a review of the evidence and clinical implications. *American Journal of Clinical Nutrition* **66**, 1072–1085.

Friedewald WT, Levy RI & Fredrickson DS (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical Chemistry* **18**, 499–502.

Gittelsohn J, Wolever TMS, Harris SB, Harris-Giraldo R, Hanley AJG & Zinman B (1998) Specific patterns of food consumption and preparation are associated with diabetes and obesity in a Native Canadian community. *Journal of Nutrition* **128**, 541–547.

Gregory J, Foster K, Tyler H & Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*. London: H.M. Stationery Office.

Haffner SM, Valdez RA, Hazuda HP, Mitchell BD, Morales PA & Stern MP (1992) Prospective analysis of the insulin-resistance syndrome (Syndrome X). *Diabetes* **41**, 715–722.

Hamman RF (1992) Genetic and environmental determinants of non-insulin-dependent diabetes mellitus (NIDDM). *Diabetes/Metabolism Reviews* **8**, 287–338.

Huijbregts PPCW, Feskens EJM & Kromhout D (1995) Dietary patterns and cardiovascular risk factors in elderly men: The Zutphen Elderly Study. *International Journal of Epidemiology* **24**, 313–320.

Johnson MF, Nichols JF, Sallis JF, Calfas KJ & Hovel MF (1998) Interrelationships between physical activity and other health behaviors among University women and men. *Preventive Medicine* **27**, 536–544.

Kant AK, Schatzkin A & Ziegler RG (1995) Dietary diversity and subsequent cause-specific mortality in the NHANES I epidemiologic follow-up study. *Journal of the American College of Nutrition* **14**, 233–238.

King H & Rewers M (1993) Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. WHO ad hoc Diabetes Reporting Group. *Diabetes Care* **16**, 157–177.

Kunst A, Draeger B & Ziegenhon J (1983) UV-methods with hexokinase and glucose-6-phosphate dehydrogenase. In *Methods of Enzymatic Analysis*, pp. 163–172 [HU Bergmeyer, editor]. Deerfield, IL: Weinheim Verlag Chemie.

Mardia KV, Kent JT & Bibby JM (1979) *Multivariate Analysis*. London: Academic Press.

Margetts BM & Jackson AA (1993) Interactions between people's diet and their smoking habits: The dietary and nutritional survey of British adults. *British Medical Journal* **307**, 1381–1384.

Meigs JB, D'Agostino RB Sr, Wilson PWF, Cupples LA, Nathan DM & Singer DE (1997) Risk variable clustering in the insulin resistance syndrome: The Framingham Offspring Study. *Diabetes* **46**, 1594–1600.

Paffenbarger RS Jr, Blair SN, Lee I & Hyde RT (1993) Measurement of physical activity to assess health effects in free-living populations. *Medicine and Science in Sports and Exercise* **25**, 60–70.

Prevost AT, Whichelow MJ & Cox BD (1997) Longitudinal dietary changes between 1984–5 and 1991–2 in British adults: associations with socio-demographic, lifestyle and health factors. *British Journal of Nutrition* **78**, 873–888.

Purnell JQ & Brunzell JD (1997) The central role of dietary fat, not carbohydrate, in the insulin resistance syndrome. *Current Opinion in Lipidology* **8**, 17–22.

Reaven GM (1988) Role of insulin resistance in human disease. *Diabetes* **37**, 1595–1607.

Registrar General (1980) *Classification of Occupations*. London: H.M. Stationery Office.

Wareham NJ, Hennings SJ, Byrne CD, Hales CN & Day NE (1998) A quantitative analysis of the relationship between habitual energy expenditure, fitness and the metabolic cardiovascular syndrome. *British Journal of Nutrition* **80**, 235–241.

Wareham NJ, Ness EM, Byrne CD, Cox BD, Day NE & Hales CN (1996) Cigarette smoking is not associated with hyperinsulinemia: Evidence against a causal relationship between smoking and insulin resistance. *Metabolism: Clinical and Experimental* **45**, 1551–1556.

Whichelow MJ, Erzinclioglu SW & Cox BD (1991*a*) Some regional variations in dietary patterns in a random sample of British adults. *European Journal of Clinical Nutrition* **45**, 253–262.

Whichelow MJ, Erzinclioglu SW & Cox BD (1991*b*) A comparison of the diets of non-smokers and smokers. *British Journal of Addiction to Alcohol and Other Drugs* **86**, 71–81.

Whichelow MJ & Prevost AT (1996) Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. *British Journal of Nutrition* **76**, 17–30.

Williams DRR, Wareham NJ, Brown DC, Byrne CD, Clark PMS, Cox BD, Cox LJ, Day NE, Hales CN, Palmer CR, Shackleton JR & Wang TWM (1995) Undiagnosed glucose intolerance in the community: The Isle of Ely Diabetes Project. *Diabetic Medicine* **12**, 30–35.

Williams DE, Wareham NJ, Cox BD, Byrne CD, Hales CN & Day NE (1999) Frequent salad vegetable consumption is associated with a reduction in the risk of diabetes mellitus. *Journal of Clinical Epidemiology* **52**, 329–335.

World Health Organization (1985) *Diabetes Mellitus: Report of a WHO Study Group, Technical Report Series* no. 727. Geneva: WHO.

Yarnell JWG, Fehily AM, Milbank JE, Sweetnam PM & Walker CL (1983) A short dietary questionnaire for use in an epidemiological survey: Comparison with weighed dietary records. *Human Nutrition: Applied Nutrition* **37**, 103–112.