

# Vegetable and fruit consumption associated with anthropometric, dietary and lifestyle factors in Spain

Antonio Agudo\*, Guillem Pera and the EPIC Group of Spain†  
IREC (Institut de Recerca Epidemiològica i Clínica), Mataró, Barcelona, Spain

Submitted 1 September 1998; Accepted 20 March 1999

## Abstract

**Objective:** To evaluate the association of vegetable and fruit intake with several demographic, anthropometric, lifestyle and dietary factors.

**Design and setting:** Cross-sectional analysis. Information on habitual diet was collected by means of the diet history method. Association of vegetable and fruit intakes with other factors was assessed separately by means of multiple lineal regression and the cumulative odds model.

**Subjects:** 39 622 healthy subjects aged 29–69 years from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort in Spain.

**Results:** Fruit intake increased with age, education and physical activity and decreased with intake of saturated fatty acids and cholesterol; smoking and alcohol consumption were also negatively associated with fruit, with a very low consumption for current smokers and heavy drinkers. Vegetable intake increased with education and physical activity and with intake of unsaturated fatty acids, mainly mono-unsaturated. Former smokers consumed more vegetables than never or current smokers and non-consumers of alcohol ate less vegetables than consumers, among whom no differences were observed. All these estimates were adjusted by energy and body mass index (BMI).

**Conclusions:** When assessing the association of fruit and vegetables with chronic diseases it is important to take into account confounding factors. Furthermore, it would be useful to study dietary patterns including several interrelated factors.

**Keywords**  
Vegetables  
Fruit  
Smoking  
Alcohol drinking  
Dietary fats  
Risk factors  
Statistical models  
Regression analysis

Many studies of diet and chronic diseases have shown a protective effect of high vegetable and fruit consumption, mainly against cancer<sup>1</sup> and cardiovascular diseases<sup>2</sup>. The evidence supporting the recommendation for high intake of vegetables and fruit is substantial as well. However, some have suggested that more information is needed, particularly on the identity of protective factors in these foods. Furthermore, it is argued that studies supporting the salutary effects of diets rich in fruits and vegetables are confounded by other, perhaps more important, lifestyle factors<sup>3</sup>. Levels of certain risk factors, mainly tobacco smoking and dietary fat intake, may vary with vegetable and fruit consumption. The association observed with disease might not reflect a causal relationship and may be simply a proxy for behaviours with an important impact on chronic disease risk<sup>4</sup>. In addition to smoking and fat intake, other factors such as physical activity, age, education and alcohol intake may covary with vegetable and fruit intake, and should be evaluated as well. It may be important for both the knowledge of aetiology and prevention of disease.

Studies on diet should also pay attention to other aspects, mainly the accuracy of dietary assessment methods. We present the results of a study carried out on a large population with a high level of intake of both fruits and vegetables<sup>5</sup>, where diet has been measured with a validated instrument. The main purpose was to assess the association of vegetable and fruit consumption with a wide variety of potential risk factors, including demographic, anthropometric and lifestyle factors, as well as dietary intake of calories and several types of fat.

## Subjects and methods

### Subjects

The population studied was the Spanish cohort of EPIC. The cohort included 41 448 individuals of both sexes, aged between 29 and 69 years, recruited between 1992 and 1996 in five regions in Spain, three from the north (Asturias, Guipúzcoa and Navarra) and two from the south (Granada and Murcia). They were

\* Corresponding author: Email agudo@csm.scs.es

† This paper presents findings from the EPIC (European Prospective Investigation into Cancer and Nutrition) studies, for details of the group, see the end of the paper.

healthy volunteers, who were mainly blood donors. The overall participation rate varied among centres from 55% to 60%. Each subject provided a blood sample, anthropometric measurements, a questionnaire on lifestyle factors and a dietary questionnaire.

### ***Dietary information***

The information on habitual intake over the previous year was collected by trained interviewers using a computerized questionnaire based on the diet history method, which had been previously validated<sup>6</sup>. It was structured by meals and the weekly frequency of consumption was reported for each food consumed at least twice a month, taking into account seasonal consumption. Although the questionnaire was open, it contained a list of commonly consumed foods and recipes to facilitate reporting. In addition to frequency, the preparation and amount consumed of each food were recorded. The portion size was assessed by means of photo series, natural units and household measures. The final quantity derived from the weekly frequency and portion size consumed was divided by seven to give the daily intake in grams. As in many studies, vegetables did not include tubers (potatoes), legumes or cereals. The complete list of foods included as vegetables and fruits is given in the Appendix. Energy, alcohol consumption and nutrient intakes (total fat, fatty acids and cholesterol) were estimated using a conversion table in a computerized database especially compiled for the EPIC study in Spain<sup>7</sup>.

### ***Anthropometric measurements***

Height, weight and waist and hip circumferences were taken following a standard protocol by previously trained interviewers. Each subject was measured barefoot wearing light clothes. Height (to the nearest 0.1 cm) was measured using a movable stadiometer and weight was measured to the nearest 100 g after urination, using an electronic, regularly calibrated weighing machine. BMI was calculated in the usual way as weight in kilograms divided by the square height in metres. Subjects with a BMI < 20 kg m<sup>-2</sup> were considered 'lean', from 20.00 to 24.99 were classified as 'normal', from 25.00 to 29.99 were considered as being 'overweight' and those with a BMI ≥ 30.00 were considered 'obese'.

### ***Lifestyle and other factors***

This information was collected by personal interview using a general questionnaire. Physical activity section consisted of questions on occupational and leisure activities. Current job was classified into sedentary, standing, manual and heavy manual work. Leisure activities included questions on time spent during the last year walking, cycling, gardening, practicing hobbies, sports and housework. Total leisure activities

were computed as the sum of the mean number of hours per week in summer and winter of all activities excluding sports and housework, which were considered separately. The reproducibility and validity of this part of the questionnaire had been evaluated within the EPIC study<sup>8</sup>. Consumption of any type of tobacco was used to classify subjects as never, current and former smokers. Women were also asked if they had ever used oral contraceptives. Finally, education was categorized into four levels, from uncompleted primary school to university.

### ***Data analysis***

Three approaches were used to assess the relationship between vegetable and fruit consumption and other factors. First, the intakes of vegetables and fruits were considered as continuous variables; since both variables were skewed, the log<sub>e</sub> (grams + 0.01) was used to normalize the distributions. Using multiple regression analysis<sup>9</sup> several models were developed for each food group, with the intake (log-transformed) as the dependent variable and all other demographic, anthropometric, lifestyle and dietary factors as independent variables. Categorical factors were introduced in the model as dummy variables. The variables representing dietary intake were categorized in quartiles, except alcohol which was categorized in three levels: non-consumers and two levels of consumption. To assess which variables or interaction terms should be kept or dropped from a model, the variation in R-square was considered in addition to statistical significance of change in the deviance. Graphical study of the residuals was used to assess the adequacy of the model. All these analyses were carried out separately for vegetables and fruits.

Multiple linear regression was also used to calculate adjusted geometric means of vegetable (or fruit) consumption for different categories of a variable, taking into account all other relevant factors. It was achieved by computing an 'adjusted intake', adding to the overall mean (in the log-scale) the residual of the model, with vegetable (or fruit) intake as a dependent variable and all other factors as covariates. This 'adjusted intake' was used to calculate an adjusted mean and confidence limits for each category of the factor of interest, which were then exponentiated to produce interpretable measures. It is analogous to the adjusted mean obtained by the analysis of covariance<sup>9</sup>.

In the second approach, the dependent variable representing either vegetable or fruit consumption was taken as an ordered categorical variable instead of a continuous one. Several models for ordinal responses have been proposed; among them, the cumulative odds<sup>10</sup> must be the first choice when an underlying continuous outcome may be assumed and the order of categories can be specified a priori. It is interpreted in

terms of odds ratios (OR), which are familiar to epidemiologists. In our study we categorized vegetable consumption in three levels: consumers of  $< 150 \text{ g day}^{-1}$ , consumer of  $150\text{--}324 \text{ g day}^{-1}$  and consumers of  $\geq 325 \text{ g day}^{-1}$ . For fruits, the values applied were  $< 200$ ,  $200\text{--}449$  and  $\geq 450 \text{ g day}^{-1}$ . These cut-off points corresponded to values close to the 25th and 75th percentiles, so that the intermediate category may be seen as the 'normal' intake in this population and subjects in the two extreme categories may be seen as 'low' and 'high' consumers. The model assumes that the OR for moving from the low category of intake to the normal-high is the same as the OR for moving from the low-intermediate category to the high intake level; this common OR is the one estimated by the model. This assumption of proportionality (also called parallelism) may be formally tested or may be examined by graphical methods<sup>11</sup>.

Finally we computed the proportion of individuals who ate at least  $400 \text{ g day}^{-1}$  of vegetables and fruits (taken together); this is the minimum recommended intake, equivalent to five servings per day<sup>12</sup>. Association of being a compliant of the recommendation, adjusted for other factors, was assessed by means of the OR, estimated from a binary logistic regression model<sup>13</sup>.

## Results

All the analyses were carried out in the subset of 39 622 subjects (out of 41 448) with complete information for all the variables studied. We tried to find the best model to explain the intake of vegetables and fruits separately. The final model contained the variables listed in Table 1 plus an indicator of the region. Age and BMI were in the model as continuous variables in their natural units, while hours per week of sport and other leisure activities, energy, fatty acids and cholesterol intake were entered as continuous and log-transformed. All other covariates were modelled as categorical variables. Use of oral contraceptives was also checked in an analysis restricted to women but no differences between users and non-users were observed. The introduction of amount of cigarettes smoked and time since quitting smoking were also considered but they did not increase goodness of fit. Hours of activity doing housework showed a high colinearity with sex and other indicators of physical activity and was dropped from the model. The introduction of separate terms for each of the three different types of fatty acids did not produce a substantial improvement compared with a single variable for total fat; however, since they showed opposite effects we decided to keep them in the model. Some interaction terms were statistically significant but they did not modify appreciably the R-square and

sometimes made interpretation difficult, so they were excluded from the model. Finally, the slope for age was non-significant in the model for vegetables, while BMI was not significant in the model for fruits, but they were retained in order to achieve a common model for both food groups. Overall, the model showed a better fit for vegetables (R-square 0.26) than for fruits (R-square 0.082).

The adjusted geometric means and confidence intervals of vegetable and fruit intake for different levels of all the covariates included in the final model are presented in Table 1, while in Table 2 we present the ORs estimated by the cumulative odds model. A differential level of consumption was observed by sex: men ate more vegetables, while women ate more fruits, although no association was found in the OR by sex. Fruit consumption increased with age while no differences were shown for vegetables. Among indicators of physical activity, the hours per week spent playing sports and other leisure activities showed a positive association, mainly with fruits. Smoking was inversely associated to fruit intake, with a marked low consumption for current smokers. The pattern regarding vegetables was completely different: former smokers had a greater consumption than current and never smokers. Non-drinkers consumed less vegetables but no differences appeared by different levels among drinkers, while for fruits only heavy drinkers showed a lower level of intake. However, a different picture arose from the cumulative odds model: a tendency to decreased intake of vegetables and fruits was observed across both levels of alcohol consumption as compared to non-drinkers. Regarding fat intake, the mean intake of fruits decreased across quartiles of saturated fatty acids (SFA) and cholesterol, but no differences were observed for vegetables. Again, these results were slightly different in the cumulative odds model, where a clear inverse association was evident for both groups of foods. As far as both types of unsaturated acids are concerned, the most striking result was the high intake of vegetables associated with the highest quartile of intake of monounsaturated fatty acids (MUFA), while no differences for fruits were found. Polyunsaturated fatty acids (PUFA) were positively associated with vegetable consumption and negatively with fruits.

In Table 3 the same set of covariates from the final model of linear regression was used to estimate the ORs associated with the prevalence of consuming at least  $400 \text{ g day}^{-1}$  of vegetables or fruits (both groups together). Although these results are not directly comparable to those from the previous tables, where vegetables and fruits were considered separately, they are rather similar. Nevertheless, the pattern tended to be more consistent with those from the cumulative odds model. Briefly, the proportion of

**Table 1** Daily consumption of vegetables and fruits in grams, according to several demographic, anthropometric, lifestyle and other dietary factors

Characteristic	%*	Vegetables †	Fruits †
<b>Sex</b>			
Male	37.3	223.6 (219.3–228.0)	188.7 (177.5–200.6)
Female	62.7	204.4 (201.2–207.6)	235.1 (226.3–244.2)
<b>Age (years)</b>			
< 45	34.8	212.7 (208.4–217.1)	152.9 (143.7–162.8)
45–54	39.3	219.2 (215.0–223.5)	244.7 (232.7–257.3)
≥ 55	25.8	200.3 (195.3–205.5)	287.1 (271.1–304.1)
<b>Educational level</b>			
Primary not complete	34.7	198.3 (194.1–202.7)	210.6 (199.5–222.3)
Primary	38.6	210.6 (206.5–214.8)	214.9 (203.3–227.1)
Secondary	14.9	228.1 (221.4–235.1)	194.4 (177.3–213.1)
University	11.8	242.2 (234.0–250.8)	273.1 (252.0–296.0)
<b>Physical activity</b>			
<b>Work activity</b>			
Sedentary	21.8	217.0 (211.3–222.8)	204.4 (190.3–219.5)
Standing	66.3	208.5 (205.4–211.7)	223.6 (215.2–232.4)
Manual	9.7	219.2 (210.8–227.9)	192.5 (169.0–219.2)
Heavy manual	2.2	237.5 (219.3–257.2)	252.1 (197.8–321.3)
<b>Sports</b>			
0 h week <sup>-1</sup>	75.1	200.3 (197.4–203.3)	198.3 (190.7–206.3)
0–3 h week <sup>-1</sup>	12.8	247.1 (239.5–255.0)	270.4 (249.3–293.3)
> 3 h week <sup>-1</sup>	12.1	262.4 (254.0–271.1)	295.9 (271.9–322.0)
<b>Other leisure</b>			
< 1 h week <sup>-1</sup>	6.4	194.4 (185.1–204.1)	103.5 (87.0–123.3)
1–7 h week <sup>-1</sup>	45.6	200.3 (196.6–204.1)	204.4 (194.8–214.4)
> 7 h week <sup>-1</sup>	48.0	225.9 (221.9–229.9)	252.1 (240.9–263.9)
<b>Body mass index</b>			
Lean	0.7	175.9 (149.7–206.7)	141.2 (89.8–221.9)
Normal	21.7	200.3 (194.8–206.0)	208.5 (194.4–223.6)
Overweight	47.6	214.9 (211.1–218.7)	217.0 (206.6–227.9)
Obese	30.0	217.0 (212.2–221.9)	223.6 (211.0–237.0)
<b>Smoking habit</b>			
Never	55.6	206.4 (203.0–209.9)	292.9 (282.4–303.9)
Former	17.6	252.1 (245.3–259.2)	262.4 (243.2–283.2)
Current	26.9	200.3 (195.5–205.3)	102.5 (94.5–111.2)
<b>Alcohol consumption</b>			
Non-consumer	30.4	196.4 (191.5–201.3)	265.1 (250.3–280.7)
< 15 g day <sup>-1</sup>	40.3	221.4 (217.4–225.4)	267.7 (256.2–279.7)
≥ 15 g day <sup>-1</sup>	29.2	217.0 (212.2–221.9)	130.3 (121.0–140.4)
<b>Energy intake</b>			
Quartile 1	24.9	190.6 (185.3–196.0)	179.5 (168.0–191.7)
Quartile 2	24.9	200.3 (195.5–205.2)	219.2 (205.7–233.5)
Quartile 3	25.0	217.0 (212.0–222.2)	239.8 (225.0–255.6)
Quartile 4	25.2	244.7 (239.2–250.3)	230.4 (214.5–247.5)
<b>Fatty acids</b>			
<b>Saturated</b>			
Quartile 1	24.8	210.6 (204.7–216.7)	249.6 (234.4–265.8)
Quartile 2	24.8	206.4 (201.5–211.4)	228.1 (213.7–243.6)
Quartile 3	25.1	208.5 (203.7–213.4)	212.7 (199.2–227.1)
Quartile 4	25.3	221.4 (216.2–226.7)	181.3 (169.0–194.4)
<b>Monounsaturated</b>			
Quartile 1	24.8	141.2 (136.9–145.6)	210.6 (196.9–225.3)
Quartile 2	24.9	192.5 (187.8–197.2)	219.2 (205.8–233.5)
Quartile 3	25.0	235.1 (229.7–240.6)	230.4 (216.2–245.6)
Quartile 4	25.3	314.2 (307.5–321.0)	204.4 (190.7–219.0)
<b>Polyunsaturated</b>			
Quartile 1	24.9	167.3 (162.5–172.3)	235.1 (220.7–250.4)
Quartile 2	25.0	202.3 (197.6–207.2)	214.9 (201.1–229.6)
Quartile 3	25.0	228.1 (223.0–233.4)	219.2 (205.3–234.0)
Quartile 4	25.1	262.4 (256.5–268.5)	198.3 (185.1–212.5)
<b>Cholesterol</b>			
Quartile 1	24.9	212.7 (206.9–218.7)	242.2 (228.1–257.2)
Quartile 2	24.9	212.7 (207.5–218.1)	237.5 (223.0–252.8)
Quartile 3	25.0	208.5 (203.7–213.4)	230.4 (216.0–245.8)
Quartile 4	25.2	214.9 (209.9–219.9)	165.7 (153.7–178.6)

\*Proportion of subjects in each category.

†Geometric means with 95% confidence intervals in parentheses, adjusted for region and mutually adjusted for all variables in the table.

subjects eating 400 g day<sup>-1</sup> or more increased with age, education, physical activity and MUFA consumption, and it is markedly low among current smokers, drinkers of  $\geq 15$  g day<sup>-1</sup> of alcohol, and high consumers of SFA and cholesterol.

## Discussion

We assessed the relationship of vegetable and fruit intake with several potential risk factors in a large population with a high consumption of both food

**Table 2** ORs of the cumulative odds model for vegetable and fruit intake in three levels\*, according to demographic, anthropometric, lifestyle and other dietary factors

	Vegetables†	Fruits‡
Sex		
Male	0.91 (0.86–0.97)	0.99 (0.93–1.05)
Age (years)		
45–54	1.04 (0.99–1.09)	1.37 (1.31–1.44)
$\geq 55$	0.96 (0.91–1.02)	1.57 (1.49–1.66)
Education level		
Primary	1.15 (1.10–1.21)	1.20 (1.15–1.26)
Secondary	1.27 (1.19–1.36)	1.14 (1.06–1.21)
University	1.37 (1.27–1.48)	1.33 (1.23–1.43)
Work activity		
Standing	1.05 (0.99–1.11)	1.05 (1.00–1.11)
Manual	1.01 (0.94–1.10)	1.18 (1.09–1.28)
Heavy manual	1.18 (1.02–1.35)	1.47 (1.28–1.69)
Sports		
0–3 h week <sup>-1</sup>	1.30 (1.22–1.38)	1.23 (1.16–1.30)
>3 h week <sup>-1</sup>	1.47 (1.38–1.57)	1.32 (1.24–1.41)
Other leisure		
1–7 h week <sup>-1</sup>	1.13 (1.04–1.22)	1.24 (1.14–1.35)
>7 h week <sup>-1</sup>	1.40 (1.28–1.52)	1.47 (1.36–1.60)
Body mass index		
Lean	0.79 (0.63–1.01)	0.84 (0.67–1.07)
Overweight	1.11 (1.05–1.17)	1.05 (1.00–1.11)
Obese	1.21 (1.14–1.28)	1.00 (0.94–1.06)
Smoking habit		
Former	1.23 (1.16–1.30)	0.90 (0.85–0.95)
Current	0.90 (0.85–0.95)	0.54 (0.51–0.56)
Alcohol consumption		
< 15 g day <sup>-1</sup>	0.91 (0.87–0.96)	0.82 (0.78–0.86)
$\geq 15$ g day <sup>-1</sup>	0.81 (0.76–0.87)	0.45 (0.42–0.47)
Energy intake		
Quartile 2	0.90 (0.84–0.96)	1.94 (1.81–2.07)
Quartile 3	0.84 (0.77–0.91)	3.25 (2.98–3.54)
Quartile 4	0.80 (0.72–0.90)	5.51 (4.94–6.15)
Saturated fatty acids		
Quartile 2	0.68 (0.63–0.72)	0.73 (0.69–0.78)
Quartile 3	0.48 (0.45–0.52)	0.57 (0.53–0.62)
Quartile 4	0.36 (0.33–0.40)	0.46 (0.42–0.50)
Monounsaturated fatty acids		
Quartile 2	2.57 (2.40–2.74)	0.99 (0.93–1.05)
Quartile 3	5.65 (5.20–6.14)	1.02 (0.94–1.10)
Quartile 4	14.15 (12.79–15.66)	1.10 (1.00–1.21)
Polyunsaturated fatty acids		
Quartile 2	1.13 (1.06–1.21)	0.88 (0.82–0.94)
Quartile 3	1.35 (1.25–1.45)	0.80 (0.75–0.86)
Quartile 4	2.19 (2.03–2.36)	0.78 (0.72–0.83)
Cholesterol		
Quartile 2	0.77 (0.73–0.82)	0.91 (0.86–0.97)
Quartile 3	0.55 (0.52–0.60)	0.82 (0.76–0.88)
Quartile 4	0.38 (0.35–0.42)	0.62 (0.57–0.68)

Reference categories: female, < 45 years old, primary level not complete, sedentary job, 0 h week<sup>-1</sup> of sport, < 1 h week<sup>-1</sup> of other leisure, normal BMI, never smokers, non-consumers of alcohol, and lowest quartile of energy intake, fatty acids (each type) and cholesterol.

\*Three groups for vegetables: < 150, 150–324,  $\geq 325$  g day<sup>-1</sup>; for fruit: < 200, 200–449,  $\geq 450$  g day<sup>-1</sup>.

†ORs with 95% confidence intervals in parentheses, mutually adjusted for all variables in the table and region.

groups: almost 74% of subjects ate at least 400 g day<sup>-1</sup> of vegetables and fruits taken together<sup>5</sup>. Adjusting by sex, age, BMI and caloric intake, a positive association with education and physical activity, and negative association with smoking, alcohol consumption, saturated fats

**Table 3** ORs and 95% confidence intervals\* of consuming  $\geq 400$  g day<sup>-1</sup> of vegetables and fruits, according to demographic, anthropometric, lifestyle and other dietary factors

	OR	95% CI
Sex		
Male	0.90	(0.83–0.97)
Age (years)		
45–54	1.33	(1.26–1.41)
$\geq 55$	1.41	(1.32–1.52)
Education level		
Primary	1.19	(1.12–1.27)
Secondary	1.25	(1.15–1.36)
University	1.47	(1.33–1.61)
Work activity		
Standing	1.05	(0.98–1.12)
Manual	1.12	(1.01–1.23)
Heavy manual	1.36	(1.14–1.63)
Sports		
0–3 h week <sup>-1</sup>	1.39	(1.29–1.50)
>3 h week <sup>-1</sup>	1.53	(1.41–1.66)
Other leisure		
1–7 h week <sup>-1</sup>	1.36	(1.24–1.50)
>7 h week <sup>-1</sup>	1.73	(1.57–1.91)
Body mass index		
Lean	0.75	(0.57–0.99)
Overweight	1.10	(1.03–1.18)
Obese	1.10	(1.02–1.18)
Smoking habit		
Former	0.98	(0.91–1.05)
Current	0.54	(0.51–0.58)
Alcohol consumption		
< 15 g day <sup>-1</sup>	0.89	(0.84–0.94)
$\geq 15$ g day <sup>-1</sup>	0.54	(0.50–0.59)
Energy intake		
Quartile 2	1.49	(1.37–1.62)
Quartile 3	2.04	(1.84–2.27)
Quartile 4	2.71	(2.36–3.11)
Saturated fatty acids		
Quartile 2	0.69	(0.64–0.75)
Quartile 3	0.49	(0.44–0.54)
Quartile 4	0.33	(0.30–0.37)
Monounsaturated fatty acids		
Quartile 2	1.61	(1.49–1.74)
Quartile 3	2.61	(2.37–2.89)
Quartile 4	4.54	(4.01–5.14)
Polyunsaturated fatty acids		
Quartile 2	0.93	(0.86–1.00)
Quartile 3	0.99	(0.91–1.08)
Quartile 4	1.21	(1.11–1.33)
Cholesterol		
Quartile 2	0.87	(0.81–0.94)
Quartile 3	0.70	(0.64–0.77)
Quartile 4	0.45	(0.41–0.51)

Reference categories are: female, <45 years old, primary level not complete, sedentary job, 0 h week<sup>-1</sup> of sport, <1 h week<sup>-1</sup> of other leisure, normal BMI, never smokers, non-consumers of alcohol, and lowest quartile of energy intake, fatty acids (each type) and cholesterol.

\*ORs and 95% CIs are mutually adjusted for all variables in the table and region.

and cholesterol, were observed. Most of these associations have been reported previously; however, they have rarely been considered altogether, most studies focusing on a subset of them. The only study including a wide variety of factors had results quite consistent with ours: less educated and sedentary people, heavy drinkers and heavy smokers reported lower consumption of vegetables and fruits, while no differences were seen across categories of BMI<sup>14</sup>. Another study in the USA<sup>15</sup> reported that median intakes of vegetables and fruits were higher for those with higher education and income and that non-smokers ate more than double the amount of fruit than smokers.

Some have observed a negative correlation between vegetable and fruit consumption and total fat intake<sup>16,17</sup>. Danish adults eating less than 35% calories from fat consumed more fruit and vegetables<sup>16</sup>, while in the USA<sup>17</sup> only fruit intake across levels of fat consumption had a clear inverse trend. We also found a negative relationship with total fat, but finally we included specific types of fatty acids since they have differential effects in most chronic diseases (mainly cardiovascular): while higher SFA and cholesterol levels increase disease risk, unsaturated fatty acids tend to be protective. We observed a clear negative association with SFA and cholesterol intake, which was much more evident for fruits. On the other hand there was a marked positive association of vegetables with MUFA. In Spain most of the MUFA consumed come from olive oil, which is very often consumed as added fat with vegetables<sup>18</sup>. Among women, a lower intake of carotenoids, mainly provided from vegetables and fruits, has been shown among users of oral contraceptives<sup>19</sup>; we did not find differences in vegetable or fruit intake between users and non-users of oral contraceptives.

The most consistent association of vegetables and fruit consumption has been observed with tobacco smoking. The above mentioned studies on lifestyle factors reported lower vegetable and fruit intake among smokers<sup>14,15</sup>. A consistent pattern of greater intake of fruits and/or vegetable among non-smokers has been also reported in the UK<sup>20</sup>, the Netherlands<sup>21</sup>, Spain<sup>22</sup>, among Norwegian workers<sup>23</sup>, and in the NHANES II study<sup>24</sup> for different age groups. In our study the inverse association with smoking was only evident for fruits, with a remarkably low level of intake for current smokers. As far as vegetables are concerned, the main result is the increased level of consumption among ex-smokers. Most previous studies<sup>14,15,21,22</sup> also showed greater differences for smokers for fruit rather than for vegetable intake. It seems that healthy and unhealthy habits tend to cluster; in particular, heavy smokers have shown the most unhealthy dietary profiles<sup>20,21</sup>. It has been suggested

that some physiological properties of tobacco smoking, affecting taste and smell, could modify food preferences<sup>14</sup>. Alternatively, quitting smoking might go together with favourable changes in other lifestyle factors, including diet, because of an increase in health consciousness. The increased intake of vegetables in former smokers in our study, also observed in the Dutch study<sup>21</sup> among women, points in this direction. However, a prospective study in the UK<sup>25</sup> did not observe substantial changes in vegetable and fruit intake 1 year after cessation, compared with those who continued to smoke.

Most of these studies measured dietary intake by means of food frequency questionnaires (FFQs)<sup>14,15,19,22</sup> or 24-hour diet recalls<sup>17,23,24</sup>. FFQs often do not include quantitative assessment of portion sizes, collecting only frequencies of consumption. They provide good estimators of intakes of some foods or groups of foods but not of the whole diet, thus not calculating energy intake<sup>26</sup>. On the other hand, 24-hour diet recalls are meant to provide suitable estimations of means of intake at the group level but not for individuals<sup>26</sup>. We assessed diet, including alcohol consumption, by means of a previously validated diet history method<sup>6</sup>; the diet history is thought to be a valid and reliable instrument in assessing habitual intake at the individual level<sup>26</sup>, though true intakes are likely to be overestimated. However, since the average error tends to be consistent between groups, it allows for valid comparisons of means. Despite overestimation, it seems to remain valid in ranking subjects according to their habitual food consumption. Furthermore, it allows for estimation of total caloric intake. Adjusting for energy is thought to be necessary<sup>27</sup>, especially when foods or nutrients that provide calories are analysed. Finally, we studied a large population with an appreciable degree of variability regarding fruit and vegetable intakes<sup>5</sup>; this allowed us to analyse a huge number of variables simultaneously and gave us the power to compare the means of several groups. Other information apart from dietary factors, like physical activity, also came from validated questionnaires. Anthropometric data to calculate BMI, which are usually self-reported<sup>14</sup>, were directly measured in our study by trained persons using a standard protocol.

Two methods of analysis were applied to our data. The first one was linear regression; this method assumes linearity of relationships and normality of variables. Although distribution of most food and nutrient intake is rather skewed, normality may be assumed after log-transformation. We also used the cumulative odds<sup>10,11</sup> to produce ORs for the association between increase in vegetable and fruit intake and other factors; this method assumes parallelism, i.e. a common OR across levels of vegetable or fruit intake. These two methods produced inconsistent results regarding

alcohol consumption and intakes of SFA and cholesterol. It could be that assumptions for some of them did not hold. Linear regression may be affected by some influential points or outliers. However, taking into account that most continuous variables were used in log-scale and the large sample size, it is unlikely that very few values produced a great effect on group estimates. Some interaction terms were not included in the model to make interpretation easier, but maybe their inclusion would produce better estimates for some variables. We examined assumptions of parallelism for the cumulative odds model<sup>11</sup> for intakes of SFA, cholesterol and alcohol. Only for alcohol consumption did the OR for moving from low to medium–high intake of vegetables and the one for moving from low–medium to high intake have very different estimates. Beyond the fact of being interpreted in terms of the familiar OR, one of the reasons for using cumulative odds in our analysis was that it takes into account how subjects are ordered according to fruit and vegetable intakes, rather than their absolute values. As mentioned, the diet history method has the advantage of ranking subjects according to their levels of intake.

Concerns about confounding factors are worth further discussion. It seems intuitively obvious that diet and other health behaviours might cluster together, at least to some extent, and that changes toward more health consciousness could then affect several factors at the same time. Even dietary recommendations are not meant to be followed in isolation; they aim to promote healthy dietary patterns as a whole. One of the advantages of our study was the consideration of several potential confounders simultaneously; the association of vegetable and fruit intake has often been analysed regarding smoking, dietary factors and other lifestyle factors separately. Dietary factors, including energy intake, BMI, physical activity, alcohol consumption and tobacco smoking constitute a net of interrelationships difficult to disentangle. All our estimates are adjusted by energy and BMI; smoking, alcohol, physical activity and saturated fats showed independent associations, mainly with fruit. Although the issue of confounding must be considered carefully when assessing vegetables and fruits as protective factors for some chronic diseases, it is worthy of note that a huge number of epidemiological studies, conducted in diverse populations and with different methods, have shown consistent results. Thus, no single factor is likely to explain the consistent protective effects. Furthermore, it may be seen in the opposite way: it could be that the effects of some other factors may be partially explained by differences in fruit or vegetable intake. Another interesting point to be taken into account is that different patterns of vegetable and fruit intake are seen for some factors, like alcohol consumption or dietary fats. A proper consideration of vegetables

and fruits as risk factors or protective factors, as well as potential confounders, therefore requires that separate indicators of intake be included in the analysis.

As has been pointed out, the knowledge of relationships between vegetable and fruit intake and other risk factors is important for both aetiological research and prevention of disease<sup>14</sup>. To date, research has emphasized benefits or harms of single nutrients and single disease outcomes. Recently, more integrative strategies have been put forward. According to such strategies, diet is considered as a complex mixture of components which interact, meaning that research on diet should shift toward identification of patterns<sup>28</sup>. Probably this strategy should be applied not only to dietary factors, but to other lifestyle and biological factors as well.

### EPIC Group of Spain

EPIC is a European study coordinated by the Unit of Nutrition and Cancer of the International Agency for Research on Cancer (IARC) (Agreement AEP/93/02). It receives financial support from the Europe Against Cancer Programme of the EU (Agreement SOC 97 200302 05F02), the Health Research Fund (FIS) of the Spanish Ministry of Health (Exp. 96-0032), the participating Regional Governments and the 'Fundación Científica de la Asociación Española contra el Cáncer'.

Members of EPIC (European Prospective Investigation into Cancer and Nutrition) in Spain are as follows.

Antonio Agudo: *Co-investigator, IREC (Institut de Recerca Epidemiològica i Clínica), Mataró (Barcelona), Spain*

Pilar Amiano: *Fieldwork coordinator, Dirección de Salud de Gipúzkoa, San Sebastián, Spain*

Ana Barcos: *Fieldwork coordinator, Departamento de Salud de Navarra, Pamplona, Spain*

Aurelio Barricarte: *Principal investigator, Departamento de Salud de Navarra, Pamplona, Spain*

José M Beguiristain: *Co-investigator, Dirección de Salud de Gipúzkoa, San Sebastián, Spain*

María Dolores Chirlaque: *Fieldwork coordinator, Consejería de Sanidad y Política Social, Murcia, Spain*

Miren Dorronsoro: *Principal investigator, Dirección de Salud de Gipúzkoa, San Sebastián, Spain*

Carlos A González: *Study coordinator and Principal investigator, IREC (Institut de Recerca Epidemiològica i Clínica), Mataró (Barcelona), Spain*

Cristina Lasheras: *Fieldwork coordinator, Consejería de Sanidad y Servicios Sociales de Asturias, Oviedo, Spain*

Carmen Martínez: *Principal investigator, EASP (Escuela Andaluza de Salud Pública), Granada, Spain*

Carmen Navarro: *Principal investigator, Consejería de Sanidad y Política Social, Murcia, Spain*

Guillem Pera: *Co-investigator, IREC (Institut de Recerca Epidemiològica i Clínica), Mataró (Barcelona), Spain*

José R Quirós: *Principal investigator, Consejería de Sanidad y Servicios Sociales de Asturias, Oviedo, Spain*

Mauricio Rodríguez: *Co-investigator, EASP (Escuela Andaluza de Salud Pública), Granada, Spain*

María J Tormo: *Co-investigator, Consejería de Sanidad y Política Social, Murcia, Spain*

### References

- Steinmetz KA, Potter JD. Vegetables, fruit, and cancer prevention: a review. *J. Am. Diet. Assoc.* 1996; **96**: 1027–39.
- Ness AR, Powles JW. Fruit and vegetable, and cardiovascular diseases: a review. *Int. J. Epidemiol.* 1997; **26**: 1–13.
- Marshall JR. Improving Americans' diet—setting public policy with limited knowledge (Editorial). *Am. J. Public Health* 1995; **85**: 1609–11.
- Nestle M. Fruits and vegetables: protective or just fellow travelers. *Nutr. Rev.* 1996; **54**: 255–7.
- EPIC Group of Spain. Dietary intake of vegetables and fruits among adults in five regions of Spain. *Eur. J. Clin. Nutr.* 1999; **53**: 174–80.
- EPIC Group of Spain. Relative validity and reproducibility of a diet history questionnaire in Spain. I. Foods. *Int. J. Epidemiol.* 1997; **26** (Suppl. 1): S91–9.
- Slimani N, Torrent M, Farriols N, et al. *European Prospective Investigation into Cancer and Nutrition (EPIC): Food Composition Tables—Spain*. Lyon: International Agency for Research on Cancer, 1991.
- Pols MA, Peeters PHM, Ocké MC, et al. Estimation of reproducibility and relative validity of the questions included in the EPIC physical activity questionnaire. *Int. J. Epidemiol.* 1997; **26** (Suppl. 1): S181–9.
- Snedecor GW, Cochran WG. *Statistical Methods*. Ames: Iowa State University Press, 1967.
- Armstrong BG, Sloan M. Ordinal regression models for epidemiologic data. *Am. J. Epidemiol.* 1989; **129**: 191–204.
- Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *Int. J. Epidemiol.* 1997; **26**: 1323–33.
- Williams C. Healthy eating: clarifying advice about fruit and vegetables. *BMJ* 1995; **310**: 1453–5.
- Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley & Sons, 1989.
- Serdula MK, Byers T, Mokdad AH, Simoes E, Mendlein JM, Coates RJ. The association between fruit and vegetable intake and chronic disease risk factors. *Epidemiology* 1996; **7**: 161–5.
- Subar AF, Heimendinger J, Patterson BH, Krebs-Smith SM, Pivonka E, Kessler R. Fruit and vegetable intake in the United States: the baseline survey of the Five a Day for Better Health program. *Am. J. Health Promot.* 1995; **9**: 352–60.
- Osler M, Heutmann BL. Food patterns associated with intakes of fat, carbohydrate and dietary fibre in a cohort of Danish adults followed for six years. *Eur. J. Clin. Nutr.* 1997; **51**: 354–61.
- Ursin G, Ziegler RG, Subar AF, Graubard BI, Haile RW, Hoover R. Dietary patterns associated with low-fat diet in the national health examination follow-up study: identification of potential confounders for epidemiologic analyses. *Am. J. Epidemiol.* 1993; **137**: 916–27.
- EPIC Group of Spain. Patterns and sources of fat and specific fatty acids intake in the participants of the European Prospective Study on Cancer and Nutrition (EPIC) in Spain. *Med. Clin. (Barc.)* 1999; **112**: 125–32.



- 19 Nebeling LC, Forman MR, Graubard BI, Snyder RA. Specific and total carotenoids intakes among oral contraceptive and estrogen hormone users in the United States. *J. Am. Coll. Nutr.* 1996; **15**: 608–13.
- 20 Margetts BM, Jackson AA. Interactions between people's diet and their smoking habits; the dietary and nutritional survey of British adults. *BMJ* 1993; **307**: 1381–4.
- 21 Zondervan KT, Ocké MC, Smit HA, Seidell JC. Do dietary and supplementary intakes of antioxidants differ with smoking status? *Int. J. Epidemiol.* 1996; **25**: 70–9.
- 22 Martínez MA, Pérez R, Martínez J, García M, Bueno A. Dietary intake of some food items in smokers and non-smokers in a Mediterranean population. *Eur. J. Public Health* 1997; **7**: 40–4.
- 23 Oshaugh A, Bjonnes CH, Bugge KH, Trygg KU. Tobacco smoking, an independent determinant for unhealthy diet? *Eur. J. Public Health* 1996; **6**: 196–202.
- 24 Subar AF, Harlan LC, Mattson ME. Food and nutrient intake differences between smokers and non-smokers in the US. *Am. J. Public Health* 1990; **80**: 1323–9.
- 25 Thompson RL, Pyke SDM, Scott EA, Thompson SG, Wood DA. Dietary changes after smoking cessation: a prospective study. *Br. J. Nutr.* 1995; **74**: 27–38.
- 26 Bingham SA, Nelson M. Assessment of food composition and nutrient intake. In: Margetts BM, Nelson M, eds. *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press, 1991; 153–91.
- 27 Willet W. *Nutritional Epidemiology*. New York: Oxford University Press, 1990.
- 28 Krebs-Smith SM, Cleveland LE, Ballard-Barbash R, Cook DA, Kahle LL. Characterizing food intake patterns of American adults. *Am. J. Clin. Nutr.* 1997; **65** (Suppl.): S1264–8.

## Appendix

List of vegetables included:

*Leafy vegetables*: parsley, seaweeds, chard, lucerne, borage, thistle, spinachs, chicory, turnip (top), lettuce,

endive (*Cichorium intybus*), endive (*Cichorium endivia*).

*Fruiting vegetables*: green beans, tomato, artichoke, eggplant, courgette, gherkins, pumpkin, capers, big capers, cucumber, sweet pepper, green chilli, avocado.

*Root vegetables*: turnip, radishes, beetroot, carrot, Yecla's carrot, Charlotte carrots, parsnip, celery.

*Cruciferous*: cabbage, red cabbage, Brussels sprouts, cauliflower.

*Mushrooms*: mushroom, fungi.

*Grain and pod vegetables*: maize, peas.

*Onion, garlic*: garlic, young garlics, pickled onions, onion.

*Stalk vegetables, sprouts*: bamboo, hearts palm, leek, asparagus, seeded soya, celeriac, caper stems.

List of fruits included:

*Fresh fruits*: apricot, cherry, custard apple, plum, strawberry, pomegranate, figs, kiwi, mango, apple, melon, nectarine, medlars, date plum, pear, pineapple, banana, watermelon, grape, peach, dried peaches, prickly pear, peach with syrup, canned pineapple, dried plum, dried figs, raisins, date.

*Citrics*: lemon, mandarin, orange, grapefruit.

*Nuts and seeds*: lupin, chestnut, coconut, almonds, hazelnuts, peanuts, walnuts, seeds (sunflower or pumpkin), pistachio, pine kernels.

*Olives*: olives.

*Juices*: orange juice, lemon juice, grape juice, tomato juice, orgeat.