DIET and CANCER the explanatory value of DIETARY PATTERNS

Elleny Balder

TNO VOEDING
BIBLIOTHEEK
26 JAN. 2005
ZEIST

Diet and cancer: the explanatory value of dietary patterns

Elleny Balder

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2004
Zugl.: Maastricht, Univ., Diss., 2004
ISBN 3-86537-279-1

© Balder, Helena Frederika
Diet and cancer: the explanatory value of dietary patterns / HF Balder
Thesis Maastricht University - with summary in Dutch

Subject headings: diet, cancer, dietary patterns, epidemiology, cohort studies, factor analysis, review

Lay-out: Helena F Balder, San Diets
Cover: "The Matrices" - Edwinus B Helsloot, Helena F Balder
Production: Cuvillier, Göttingen, Germany

© CUVILLIER VERLAG, Göttingen 2004
Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.
1. Auflage, 2004
Gedruckt auf säurefreiem Papier

ISBN 3-86537-279-1

Diet and cancer;
the explanatory value of dietary patterns

Proefschrift

ter verkrijging van de graad van doctor
aan de universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op donderdag 23 december 2004 om 14.00 uur

door

Helena Frederika Balder

The work described in and the printing of this thesis was supported by a grant from the Commission of the European Communities (QLK1-CT-1999-00575) and by a grant from the Dutch Cancer Society (TNO V 1999-1963).

Promotor:
Prof.dr.ir. P.A. van den Brandt

Co-promotor
Dr. R. Alexandra Goldbohm, TNO Voeding, Zeist

Beoordelingscommissie:
Prof. dr. P. van Assema
Prof. dr. M. Berger (voorzitter)
Prof. dr. M. von Meyenfeldt
Dr. M. Rookus (Nederlands Kanker Instituut, Amsterdam)
Prof. dr. P. van 't Veer (Universiteit Wageningen)

Voor papa en mama

Table of contents

Chapter		page
1	General Introduction	9
2	Factor analysis of food consumption data: a review of the literature	15
3	Common and country-specific dietary patterns in four European cohort studies	43
4	Dietary patterns associated with sociodemographic and lifestyle characteristics	61
5	Dietary patterns associated with colon and rectal cancers: Results from the DIETSCAN Project	75
6	Dietary patterns associated with male lung cancer risk in the Netherlands Cohort Study	91
7	Dietary patterns and cancer risk in a prospective cohort study	107
8	General discussion	125
	Summary	139
	Samenvatting	143
	Dankwoord	149
	About the author	155

1

General Introduction

Cancer is a major public health problem. In the Netherlands, cancer is -after cardiovascular disease- the second major cause of mortality. Each year approximately 66,000 invasive and 6,000 non-invasive tumors are diagnosed.[1] In men, lung cancer is the most common, followed by prostate and colorectal cancer. In women breast cancer occurred most, followed by colorectal and lung cancer. Cancer is not one single disease, but the collective term for uncontrolled cell proliferation resulting in the invasion of surrounding tissue. Cancer has multiple etiologies. The causes of cancer include genetic predisposition, environmental influences, lifestyle factors (including diet) and infectious agents.

1.1 Epidemiology of diet and cancer

Generally, epidemiological and other studies suggest that diet is an important factor in the etiology of different types of cancer. Since Doll and Peto[2] estimated that diet could be responsible for one third of all cancers in the early 1980s, the epidemiological literature on diet and cancer has grown considerably.[3] Fruits and vegetables and the potentially protective compounds they contain, such as vitamins C, E, beta-carotene and folate have been hypothesized to reduce risk of several cancers although this effect does not appear to be as large as originally thought.[3;4] The types of cancer most likely to be influenced by eating fruit and vegetables are head and neck, esophageal, lung and stomach cancers. The association with bowel cancers remains equivocal. There appears to be no association between fruit and vegetables consumption and breast or prostate cancer risk. Alcohol has been consistently found to increase the risk of several types of cancer, such as mouth, esophageal, liver, breast and colorectal cancer. Meat consumption has been hypothesized to be associated with increased risk for bowel cancer, however the evidence is equivocal. Furthermore, the overall evidence does not confirm the hypothesis that higher meat consumption increases risk of several types of cancer, such as breast, lung, stomach and prostate cancer. High fat intake has been hypothesized to increase risk of breast, colorectal and prostate cancers, but cohort studies did not show a convincing association. A high consumption of milk and dairy products and a high calcium intake have been shown to be associated with a weak but consistent reduction in the risk of bowel cancer. There are indications that the consumption of milk or dairy products does not influence the risk of breast cancer and data on a possible increase of risk for prostate cancer are limited.
Thus, the effects of numerous food items and nutrients have been investigated, but to date, the reported evidence for the relation between dietary habits and a number of cancer types is not as clear-cut and consistent across studies as one would wish for causal inference.

1.2 Rationale for dietary pattern analysis

An explanation for the inconsistent evidence regarding diet and cancer could be that epidemiological studies predominantly analyze a limited number of nutrients or food items. This approach has three major limitations. First, people do not eat isolated nutrients or foods. They eat meals consisting of a variety of foods with complex combinations of numerous constituents. As the number of candidate risk factors is large, many nutrients and foods have to be studied in order to get a comprehensive overview of the possible risk factors. An analysis including a large number of foods also yields statistically significant associations on the basis of chance. Secondly, it is not possible to control for all nutrients and foods that may be confounders, as the intakes of many dietary components are highly correlated. Finally, it is quite possible that different dietary components influence disease

risk interactively. Studies that focus on single nutrients or food items do not allow to take these interactions into account. Examination of dietary patterns that allows to study the impact of many foods and their interactions simultaneously may give more insight into the relation between diet and different types of cancer than analysis of single nutrients. The first studies on dietary patterns and cancer suggested that such an approach might better explain risk of disease than a single nutrient or food approach.[5]

1.3 Statistical analysis of dietary patterns

Overall dietary exposure can be examined by a-posteriori or a-priori approaches. The a-posteriori approach is based on current knowledge about the role of foods and nutrients in the etiology of disease. The diet is scored on whether the intake of certain foods or nutrients are in compliance with nutritional recommendations which in turn are based on the current scientific evidence on the relation between diet and disease. The resulting score can be used as a dietary exposure variable. Examples of this approach are the Diet Quality Index (DQI)[6;7], and Healthy Eating Index (HEI).[8] The a-priori approach is data-driven, using statistical techniques to summarize dietary exposure. Cluster analysis aggregates persons with similar eating habits into clusters. Members of a cluster tend to be relatively similar to each other, but dissimilar to members of other clusters. In this approach a subject usually is part of only one cluster. Factor analysis groups food variables that correlate positively or negatively into a pattern. Each subject will be assigned scores that characterize the extent to which his/her diet conforms to each pattern. The outcomes of all three approaches, dietary indices, clusters and factor-scores, can be used as risk factors in the subsequent analysis to associate overall diet with the outcome of interest. Although the combination of known risk factors are likely to better explain disease than the separate risk factors, the major limitation of the dietary index approach is that it does provide little new insight. As cluster and factor analysis are data driven they are more likely to generate new hypotheses. Neither method is better than the other. Clusters are easier to interpret as they are mutually exclusive. However, intuitively, reducing all dietary information of a person into belonging to only one of a few distinct clusters that can be compared to a reference group seems to be too much loss of potentially relevant information. For the purpose of the research described in this thesis factor analysis was chosen to disentangle a subject's dietary habits into a limited number of independent dimensions that appear to be common in the population(s) under study and to study their relevance for cancer.

1.4 Objectives

The main goal of the research described in this thesis was to gain insight into whether prevailing dietary patterns *consistently* enhance or protect against cancer.

In order to achieve this goal the following questions were formulated:

(1) Is the dietary pattern approach robust?
(2) Are dietary patterns comparable across study populations?
(3) Are dietary patterns part of a broader lifestyle pattern?
(4) Are dietary patterns associated with risk of (several types of) cancer?
(5) What is the value of the dietary pattern approach, compared to the traditional epidemiological approach?

A systematic review of the literature on dietary pattern analysis through factor analysis (including 68 studies) attempted to answer mainly the first two questions. As a literature review, hampered by the huge variation in methodology employed across the studies, has only limited power to produce strong evidence, a multi-center study - the DIETSCAN project - was set up for this purpose. In this project, including four European ongoing cohort studies on diet and cancer, a common methodology was developed to conduct the factor analysis, despite the differences between the methods used to assess diet in the original cohorts. The DIETSCAN study was used in the first place to conduct sensitivity analyses to test the robustness of (subjective) factor analytic decisions on the resulting patterns. In the second place, the study was used to compare the resulting patterns across the participating cohorts. The third purpose of DIETSCAN was to test whether the associations between shared dietary patterns, if any were found, and cancer were consistent across the four cohorts.
In addition, analyses of the complete dietary patterns identified in the Netherlands Cohort Study on Diet and Cancer (NLCS) were conducted to answer the third question, i.e. the association between the patterns and other lifestyle characteristics and also the last two questions.

1.5 The cohort studies

The Netherlands Cohort Study The Netherlands Cohort Study on diet and cancer (NLCS) is a prospective cohort study, which began in September 1986. The cohort included 62,573 women and 58,279 men aged 55 to 69 years who completed a mailed, self-administered questionnaire on dietary habits and other risk factors for cancer. Because of the case-cohort design, a random subcohort of 5000 subjects was sampled from the cohort after the baseline exposure measurement. In a case-cohort approach, follow-up information on vital status of a random subcohort provides an estimate of the person-time experience of the entire cohort. Exposure data of the subcohort is combined with those of incident cases in the entire cohort.

Other cohort studies in DIETSCAN Besides the NLCS, three other cohort studies participated in the DIETSCAN project. These were The Alpha Tocopherol Beta Carotene Cancer Prevention Study (ATBC - Finland), the Swedish Mammography Cohort (SMC), and the ORmoni e Dieta nella Eziologia dei Tumori (ORDET - Italy)
The ATBC Study was a randomized placebo-controlled intervention study conducted among male smokers. The cohort consisted of 29,133 white men aged 50 to 69 years living in Finland. After the intervention ended the follow-up of cancers and deaths has been continued. The SMC study originated from population-based mammography screening in Sweden. A questionnaire that was sent together with an invitation to be screened by mammography, was returned by 66,651 women aged 40 to 76 years. The ORDET Study is a prospective cohort study on hormonal factors and diet in the etiology of breast cancer among 10,788 healthy Italian women, aged 35-69 years.
The main criteria for selecting these studies were that they were prospective cohort studies, with comprehensive dietary assessment using a validated dietary instrument and with follow-up for cancer through record linkage of the cohorts with the national or local cancer registries

1.6 Outline of the thesis

In chapter 2 the literature on dietary patterns identified by factor analysis is reviewed, and subjectivity and comparability of dietary patterns and their associations with personal and lifestyle characteristics and disease are described. Chapter 3 describes the methodology used in the Dietscan project to determine dietary patterns, the robustness of the dietary pattern approach and the major dietary patterns extracted in each of the four cohort studies. Chapter 4 illustrates that major dietary patterns are part of a broader lifestyle pattern. Associations between dietary patterns and incidence of several types of cancer are described in chapter 5 through 7. Across the four cohorts, the common dietary patterns were linked to the risk of breast[9] and colorectal cancer (chapter 5). The choice of these cancers was based on the hypothesized association with diet and a sufficient number of cases available in the cohort studies. In chapter 6 the association of dietary patterns with lung cancer in NLCS men was compared with food-based analyses. In chapter 7, the association of dietary patterns with risk of overall cancer in the NLCS is presented and compared with nutrient-based analyses. Finally, in chapter 8, the results of the research described in this thesis are discussed, the explanatory value of the dietary pattern approach is evaluated and suggestions for further research are given.

REFERENCES

1. Visser O, Coebergh JWW, Dijck JAAM, Siesling S. Incidence of cancer in the Netherlands 1998. Utrecht: Vereniging van Integrale Kankercentra (Netherlands cancer registry), 2002.

2. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J.Natl.Cancer Inst. 1981;66:1191-308.

3. Signaleringscommissie Kanker van KWF Kankerbestrijding. De rol van voeding bij het ontstaan van kanker. Amsterdam: KWF Kankerbestrijding, 2004.

4. IARC Working Group on the Evaluation of Cancer-preventive Strategies. Fruit and vegetables, IARC handbooks of cancer prevention ; 8. Lyon: IARC Press, 2003.

5. Kant AK. Indexes of overall diet quality: a review. J.Am.Diet.Assoc. 1996;96:785-91.

6. Patterson RE, Haines PS, Popkin BM. Diet quality index: capturing a multidimensional behavior. J.Am.Diet.Assoc. 1994;94:57-64.

7. Seymour JD, Calle EE, Flagg EW, Coates RJ, Ford ES, Thun MJ. Diet Quality Index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. Am.J.Epidemiol. 2003;157:980-8.

8. Kennedy ET, Ohls J, Carlson S, Fleming K. The Healthy Eating Index: design and applications. J.Am.Diet.Assoc. 1995;95:1103-8.

9. Mannisto S, Dixon LB, Balder HF et al. Dietary patterns and the risk of breast cancer: Results from three cohort studies in the DIETSCAN project. submitted 2003.

2

Factor analysis of food consumption data: a review of the literature

Balder HF, Van den Brandt PA, Goldbohm RA

Submitted

ABSTRACT

This systematic review summarizes and compares 68 published articles (up to January 1st, 2004) in which dietary patterns were identified by factor analysis. The 65 sets of dietary patterns extracted in the different studies revealed many similarities, despite differences in the data collection methods and subjective factor analytical decisions, such as number and type of input variables and number of factors extracted. Moreover, the association of the dietary patterns with sociodemographic and lifestyle characteristics and chronic disease were comparable. This suggests that factor analysis may be successful in identifying meaningful, common underlying patterns in food consumption in different populations.

INTRODUCTION

Analysis of dietary patterns is a relatively new approach that identifies prevailing combinations of nutrients or foods in a population. Dietary patterns are expected to give more insight into the relation between diet and disease in addition to analysis of single nutrients or foods because they allow studying the impact of many nutrients and/or foods and their interactions simultaneously.[1;2] One of the methods that can be used to study the multidimensionality of diet is factor analysis, an exploratory, data driven technique that identifies independent vectors of variables in a correlation matrix and provides scores that rank individuals in terms of how closely their diet corresponds to the total pattern. Some objections have been brought forward against factor analysis. It has been argued that results of pattern analyses are sample specific and strongly depend on subjective analytic decisions and that differences in eating patterns observed in previous studies might be the result of differences in study methodology, instead of representing true differences in the respective study populations.[3]

Earlier reviews[4-9] have described studies on dietary patterns based on diet indices, cluster analysis and factor analysis. These showed no consistent association between dietary patterns and BMI and/or obesity,[7] found that studies using indices of overall diet quality were related to the risk of disease more strongly than individual nutrients and foods[4] and described the available evidence regarding relationships between dietary patterns and risk of hypertension, type 2 diabetes mellitus, and coronary heart disease.[9] However, these previous reviews did not have a particular focus on the comparability of the identified dietary patterns between studies or the influence of subjective analytical decisions on the resulting dietary patterns. The objective of the present literature survey was to review systematically all studies determining dietary patterns using factor analysis published up to January 1st, 2004. We wanted to study to what extent the identified dietary patterns could be compared, whether the identified dietary patterns were influenced by subjective analytical decisions and whether the dietary patterns showed comparable associations with sociodemographic and lifestyle characteristics and chronic disease.

METHODS

A literature search using the computerized databases PUBMED and Current Contents (WebSPIRS from Silver Platter, version 4.3) was conducted. The following keywords were used to identify relevant publications: diet pattern(s), dietary pattern(s), food pattern(s), eating pattern(s), food intake pattern(s) and factor analysis, principal component(s) analysis, pca. In addition, references in relevant publications were examined further.

The papers were selected for inclusion in this review only if they (a) were published in the English language before January 1st, 2004, (b) reported the use of factor analysis and/or principal components analysis to identify dietary patterns, (c) were based on individual food consumption data in the population at large, and (d) reported comprehensive diet assessment intended to cover the entire diet. In order to focus the review on food intake patterns and to ensure comparability of studies, we did not include papers concerning diet quality scores or cluster analysis, studies based on per capita food consumption, studies based on food preference, attitude or similar behavioral items, or studies reporting dietary patterns of populations consisting of children only.

Table 1: Characteristics of published studies that used factor analysis to determine dietary patterns.

				Dietary measurement			Input variables	
Ref	First author, year	Country	Study period	Type[a]	# items	Time[b] frame	#	type[c]
11	Pelto 1981	Finland	1978	24hr Rec			17	F
12;13	Schwerin 1981	USA	1968-70	24hr DR			15	Q
14	Stellman 1983	USA	1982	FFQ	28		28	F
15-17	Iizumi 1986	Japan	1974-75	3d DR			25	Q
18	Gex-Fabry 1988	Switzerland	1977-82	FFQ + DH	109	2 wk	33	Q
19	Nicklas 1989	USA	1981-83	FFQ	64	1 wk	64	F
20;21	Barker 1990	Northern Ireland		7d W Rd			41	Q
22;23	Hebert 1990	USA	1985-88	FFQ	30	I.d.	30	logF
24	Kristal 1990	USA		FFQ + 2* 4d DR	28		28	F
25-28	Randall 1990	USA	1975-86	FFQ	128	1 yr	110	F
							95	F
29	Gallagher 1993	USA		FFQ	35	1 d	23	F
30	Hebert 1993	USA	1988	FFQ	67	1 yr	67	logF
31	Nolan 1995	Canada		DH		3 m	32	
32	Wolff 1995	USA	1982-84	FFQ	57		18	F
33;34	Whichelow 1996	UK	1984-85	FFQ			39	F
35	Gray-Donald 1997	Canada	1992-93	FFQ	38	3 m	38	F
36	Beaudry 1998	Canada	1990	24 hr Rec			30	Q
37	Gittelsohn 1998	Canada	1992	FFQ	34	3 m	34	F
38	Slattery 1998	USA	1991-94	DH			35	
39	Bree 1999	USA	1989-91	FFQ	99	1 yr	99	F
40-43	Hu 1999	USA	1986-90-94	FFQ+7d Rd	131	1 yr	37-42	F
44	Kumagai 1999	Japan	1988	FFQ	15		25	F
45	Matalas 1999	Greece	1988-89	FFQ+1d Rd	54	1 yr	38	logF
46	Matalas 1999	Greece/USA	1990-92	FFQ	104	1 m	49	logF
47	Maskarinec 2000	Hawaii		FFQ	±200		23	logQ
48	Mullen 2000	UK	1987-88	FFQ	28		28	F
49	Tseng 2000	USA	1988-94	FFQ	63	1 m	63	F
50	Williams 2000	UK		FFQ	35	1 yr	29	F
51;52	Fung 2001	USA	1984-94	FFQ	116	1 yr	36-38	F
53;54	McCann 2001	USA	1986-91	FFQ	190	1 yr	168	F
							56	F
							36	F
55	Ogura 2001	Japan	1996-97	FFQ	33		33	F
56-58	Osler 2001	Denmark	1982-98	FFQ	26		28	F
59	Palli 2001	Italy	1985-87	FFQ	181	1 yr	20	Q
60;78	Schulze 2001	Germany	1994–98	FFQ	148	1 yr	47	Q
61;62	Terry 2001	Sweden	1987-90	FFQ	67	6 m	24	F
63	Tseng 2001	USA	1988-94	FFQ	62	1 m	62	F
64	Handa 2002	Canada	1995-96	FFQ	69		69	F
65	Mishra 2002	Australia	1995	FFQ	100	1 yr	99	F
66	Sichieri 2002	Brazil	1996	FFQ	80	1 yr	16	F
67	Balder 2003	Netherlands	1986	FFQ	150	1 yr	49	Q
		Finland	1985-88	FFQ	276	1 yr	46	Q
		Sweden	1987-90	FFQ	67	6 m	42	Q
		Italy	1987-92	FFQ	107	1 yr	32	Q
68	Bell 2003	USA		FFQ	110		38	F
69	Costacou 2003	Greece		FFQ	150	1 yr	17	Q
70	Chi 2003	Japan	2001	FFQ		e.d.	16	F
71	Kerver 2003	USA	1988-94	FFQ		1 m	35	F
72	Markaki 2003	Greece	1990-93	FFQ	130	a.l.	110	F
73	Masaki 2003	Japan	1988	FFQ	33	2-3 yr	26	F
74	Sánchez-Villegas 2003	Spain	2000-01	FFQ	136		30	F
75	Schulze2003	Germany	1994-98	FFQ	148	1 yr	44	Q
76	Van Dam 2003	Netherlands	1993-97	FFQ	178	1 yr	46	

Table 1 continued

				(Number of) Extracted dietary patterns[e]								
Ref	**First author, year**	**#**	**Gender[d]**	**Total**	**V**	**M**	**Sw**	**Sn**	**T**	**A**	**Other**	**% var expl**
11	Pelto 1981	108	w	2	1				1			48
12;13	Schwerin 1981	11337	m+w	7							7	55
14	Stellman 1983	±5000	m	7	1	1	1				4	51
15-17	Iizumi 1986	101	m+w	8	1	2	1		1	1	2	
18	Gex-Fabry 1988	939	m+w	3	1		1				1	20
19	Nicklas 1989	1275	m+w	17	2	4	3	2			6	57
20;21	Barker 1990	592	m+w	4	1			1	1		1	22
22;23	Hebert 1990	2191	m+w	9	4	1	1	1			2	
24	Kristal 1990	97	w	5							5	
25-28	Randall 1990	1475	m	9	5	1	1	1			1	22
		780	w	9	5		1				3	23
				7	3	1		1	1		1	20
				7	3		1				3	22
29	Gallagher 1993	531	t	4	1		1			1	1	23
		353	m+w	4	1		1				2	20
30	Hebert 1993	1762	m+w	20	2	2	2				14[f]	55
31	Nolan 1995	301	m+w	3	1			1			1	
32	Wolff 1995	549	w	7	1			1	1		4	59
33;34	Whichelow 1996	9003	m+w	4	1		1	1				28
35	Gray-Donald 1997	1432	m+w	5	1			1			3	40
36	Beaudry 1998	2104	m+w	3	1			1	1			18
37	Gittelsohn 1998	721	m+w+c	7	1			1			5	
38	Slattery 1998	1290	m	6	1	1	1			1	2	37
		1120	w	6	1	1	1			1	2	34
39	Bree 1999	4640	m+w	2	1	1						
40-43	Hu 1999	44875	m[g]	2	1	1						20
44	Kumagai 1999	630	m+w	4	1	1					2	44
45	Matalas 1999	356	m+w	11	1	1	1		1	1	6	
46	Matalas 1999	380	m+w	13	1	2		1		1	8	
47	Maskarinec 2000	514	w	4	1	1					2	93
48	Mullen 2000	985	m+w	5	1	1	1	1	1			
49	Tseng 2000	4641	m+w	3	1			1	1			15
				4	2			1	1			20
50	Williams 2000	802	m+w	4	1	1	2					32
51;52	Fung 2001	76399	w[h]	2	1	1						
53;54	McCann 2001	863	w	2	1	1						8
		863	w	2	1	1						13
		863	w	2	1	1						17
55	Ogura 2001	236	m+w	2	1	1						22
56-58	Osler 2001	5871	m+w	2	1	1						20
59	Palli 2001	943	m+w	4	1				1		2	75
60;78	Schulze 2001	8975	m	7	1		1		1	1	3	31
		13379	w	7	1		1		1	1	3	31
61;62	Terry 2001	61463	w	3	1	1				1		25
63	Tseng 2001	5788	m+w	2	1	1						12-14
64	Handa 2002	250	m	13	4	2	2				5	24
		422	w	8	1	1	1				5	16
65	Mishra 2002	3111	m	15	5	3	1	1			5	50
		3569	w	15	4	1	2	1			7	50
66	Sichieri 2002	1198	m	3	1				1		1	
		1528	w	3					1		2	
67	Balder 2003	1525	m	5	2	1	1				1	23
		1598	w	5	1	1	1				2	23
		27111	m	3	1	1				1		20
		61469	w	4	1	1				1	1	22
		9208	w	4	2	1				1		29
68	Bell 2003	135	m+w	7	3	2			1		1	44
69	Costacou 2003	28034	m+w	4	2	1	1					40
70	Chi 2003	1308	m+w	3	1	1					1	
71	Kerver 2003	13130	m+w	6	2	2		1			1	37
72	Markaki 2003	251	m+w	25	4						21[i]	54
73	Masaki 2003	5644	m	4	1	1					2	39
74	Sánchez-Villegas 2003	3847	m+w	2	1			1				15
75	Schulze2003	8552	w[j]	2	1				1			
76	Van Dam 2003	19750	m+w	3	1			1	1			18

Table 1 continued

(a) 24 hr Rec=24 hours record, 24 hr DR=24 hours diet record, FFQ=food frequency questionnaire, DR=diet record, DH=dietary history, 7d W Rd=7 days weighed record; (b) prev=previous, d=day, wk=week, m=month, yr=year; ld="longest duration", e.d.="every day"; a.l.= "average lifetime consumption"; (c) F=frequency, Q=quantity, log=logarithm; (d) w=women, m=men, c=children, t=teenagers; (e) common labels assigned by the authors of this review to the original identified dietary patterns. V=number of patterns characterized by vegetables and/or fruits consumption , M=number of patterns characterized by meat and fat foods , Sw=number of patterns characterized by consumption of sweet foods, desserts and/or baked goods , Sn=number of patterns characterized by consumption of (salty) snacks, fast foods, fried foods, T=number of patterns labeled "Traditional" , A=number of patterns characterized by consumption of alcoholic beverages , Other=number of other dietary patterns; (f) Eight patterns (numbers 9,10,13,15,16,18,19,20) not mentioned in paper; (g) 127 in validation study[40], 466 with bloodsamples[42], 42 504 for Diabetes Mellitus analyses [43], 44 875 for CHD[41]; (h) 69017 women in 1984[51], 76399 women, from 1984-1994[52]; (i) other patterns, that were not associated with disease not mentioned in paper; (j) Exploratory factor analysis on 44 foodgroups from 1994 subjects, Confirmatory factor analysis on 8 foodgroups from 9127 subjects

For inclusion, the articles had to provide sufficient information on the methods used to identify the analysis of dietary patterns and on the resulting patterns. If more than one publication described dietary patterns in the same study population, they were considered as one study. Different publications describing dietary patterns using more than one measurement in time of the diet within the same study population were also treated as one study. If publications described dietary patterns in separate subgroups of the same study population (e.g. different sex or age groups) or the patterns were identified using a different methodology (e.g. different number of input variables or a different number of extracted variables) they were treated as different studies.

For each study, information was collected on characteristics of the study population and the instrument used to measure food consumption. In addition, information was collected on the methods of the dietary pattern analysis and on the labels assigned by the authors to the identified dietary patterns, and the food groups correlated with these patterns. Based on the foods and food groups that loaded highest on the pattern (regardless of aggregation level of the original food variables) we reclassified the original identified dietary patterns that we interpreted to be comparable and assigned common labels to them. All dietary patterns that met one of such descriptive labels were assigned a common pattern label. Therefore, particularly in studies in which a large number of factors were extracted, the same common dietary pattern label could sometimes be assigned to more than one of the dietary patterns within one study. In describing the association between dietary patterns and sociodemographic and lifestyle characteristics and chronic disease these patterns were treated as independent observational units. Only dietary patterns that were considered comparable, i.e. those with a common label, were taken into account in the further analyses in this review.
To study the influence of methodology on the identified dietary patterns, cross-tables were generated of the number of studies that observed one or more of the common dietary patterns, stratified by region (USA, Europe, other), gender (men, women, both), dietary method (FFQ, 24 hr recall, other), timeframe of dietary method (< half year, > half year), number of items of FFQ (<35, 35-100, >35 items), type of input variables (frequency, quantity), number of input variables (<30, 31-50, 51-80, >80), criteria used to determine the number of extracted factors (Eigenvalue, screeplot, interpretation), number of factors extracted (1, 3-4, 5-7, >8), total variance explained (>25%, <25%), type of rotation(Varimax and/or orthogonal, promax, not rotated), and whether or not the association between the patterns and disease were studied (yes, no).
To explore the association of dietary patterns with lifestyle variables the number of studies that observed a significant positive, null or significant negative association between the

dietary patterns and personal or lifestyle characteristics were counted. To describe the association between dietary patterns and cancer, forest plots were generated.[10] Heterogeneity was explored by subgroup analyses according to study design, gender, region, number of input variables and number of extracted variables in the exploratory factor analysis, and tumor site.

RESULTS

In total, 68 published articles describing dietary patterns identified by factor analysis met our criteria. The main characteristics of the populations and the methods described in these articles are listed in table 1.[11-76]. Nine publications described patterns separately for men and women,[25;26;38;49;60;64-67], one publication determined dietary patterns separately for adults and teenagers,[29] Randall *et al*[25-28] described two different sets of identified dietary patterns in the same study population, one publication used three classifications of input variables,[53] and one publication described dietary patterns determined in four separate cohort studies.[67] This resulted in 65 sets of identified dietary patterns that we treated as independent observational units and that we will refer to as "studies" in this review.
The majority of studies originated from the USA (26/65, 40%) and used dietary data from men and women combined (47/65, 72%) (table 2). The sample size of the studies ranged from 97 to 69017 (mean 7445). In most studies dietary information was collected using a food frequency questionnaire (56/65, 86%). In the factor analytical process, most studies used frequency rather than quantity of consumption as input variables (44/65, 68%), reported the use of the Eigenvalue criterium to determine number of extracted factors (41/65, 63%), and used Varimax rotation (46/65, 71%). The number of input variables ranged from 15 to 168 (mean 46) and the number of extracted factors ranged from 2 to 25 (mean: 6), explaining 7.7 to 93% (mean: 32%) of total variance.

Common dietary patterns

Most studies (62/65, 95%) observed at least one dietary pattern characterized by the consumption of vegetables and/or fruits (table 1). The labels assigned to these patterns by the respective authors either described the content of the pattern, such as “fruit and vegetable”, or the interpretation of the underlying behavior, such as “prudent” or “healthy” (appendix table 1). From here on we will refer to this pattern as the vegetable pattern. In most (39/65, 60%) studies also a pattern characterized by consumption of several types of meat (animal protein) and fat foods was identified (table 1). The labels assigned to these patterns by the respective authors described the content, such as “high fat meats and eggs” or “meat and potatoes”, or were based on the interpretation of the underlying behavior such as “western” (appendix table 2). From here on we will refer to this pattern as the meat pattern. A dietary pattern consisting mainly of sweet foods, desserts and/or baked goods was also observed in several (26/65, 40%) studies (table 1; appendix table 3). We will refer to this pattern as the sweet pattern. Another frequently (19/65, 29%) observed pattern was one characterized by consumption of salty snacks such as potato chips, fast foods and/or fried foods (table 1). The labels assigned to this pattern by the respective authors reflected the type of foods consumed, such as “snacks”, and “fast foods” or the nutritional value of the pattern, such as “high fat, low nutrient density”, “nutrient dilute”, “high energy density”, or “high calorie” (appendix table 4). We will refer to this pattern as the snack pattern. The label “Traditional" was also frequently (18/65, 28%)

used for an identified pattern by the investigators, indicating traditionally or culturally determined food choice (table 1). Obviously, the foods of which these patterns consisted differed much between studies/populations (appendix table 5). For instance, in Japan this pattern consisted of pickles, seasonings, and fish and shellfish.[15-17] The traditional dietary pattern in Greece consisted of olives, goat meat, and toasted bread.[45] In Mexican Americans dietary patterns consisting of tortillas, legumes, high fat meats, and sugar[32] and of beans, corn tortillas and chili peppers[49] were considered to be traditional. As this label was not based on a description of the foods correlated to the pattern, in some cases these traditional patterns could also have been classified into one of the previously described common patterns. However, because of the assumed culturally determined food choice we choose to leave them into a separate, although heterogeneous, category, which we will refer to as the traditional pattern. In twelve studies (12/65, 18%) a dietary pattern characterized by consumption of alcoholic beverages (table 1; appendix table 6) was identified.

Influence of methodology on identified dietary patterns

The number of studies in which one (or more) of the above described common dietary patterns were identified, stratified by population and study characteristics and factor analytical decisions are presented in table 2. The three studies that did not extract a vegetable pattern used less than 30 input variables, did not report the use of a screeplot or interpretation to determine the number of factors extracted, did not report varimax rotation and did not relate the dietary patterns to disease. A meat pattern was more likely to be identified in studies using frequency of consumption as input variables (29/39, 74%) or in studies in which either 2 or more than eight patterns were extracted (23% and 28% respectively) and was less likely identified in studies using 24 hr recall data. A sweet pattern was more likely to be observed when five or more factors were extracted (20/26, 76%) and less in studies using a study population of only women (1/26, 4%), studies using 24 hr recall as dietary method (0/26, 0%), and studies that reported interpretation as criterium (4/26, 15%). A snack pattern was more likely identified when eight or more factors were extracted (6/19, 32%) and less likely when dietary data from men only was used (0/19, 0%), or using quantity of consumption as input (2/19, 11%). A pattern was more likely to be labeled "traditional" in studies originating from Europe (9/18, 50%), in studies using quantity of consumption as input (7/18, 39%) and less likely in study populations consisting of men only (0/18, 0%), in studies using an FFQ of less than 35 items (1/18, 6%) or in studies extracting either 2 or 8 or more factors (both 2/18, 11%). An alcohol pattern was more likely to be identified in studies originating from Europe (8/12, 67%) or using quantity of consumption as input (6/12, 50%) and less likely when 24 hr recall was used, when more than 50 input variables were used or when only 2 factors were extracted (0/12, 0%).

Association of dietary patterns with lifestyle

For each subject in a dataset a factor score can be calculated by multiplying the individual consumption of a specific food item by the factor loading of that food item. This score represents the consumption by that subject of all foods that are comprised in the pattern, weighed by their loading, and it can be used to calculate the association with a variable of interest. The number of studies that observed a significant positive, null or significant negative association between dietary pattern scores and personal and lifestyle

Table 2: Number and percentage[a] of studies in which the common dietary patterns[b] were observed, stratified by population and study characteristics and factor analytical decisions, in published studies up to Jan 1st, 2004.

	Total		Vegetables		Meat		Sweet		Snacks		Traditional		Alcohol	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Total	65	100	62	100	39	100	26	100	19	100	18	100	12	100
Country														
USA	26	40	24	39	17	44	12	46	7	37	5	28	3	25
Europe	23	35	23	37	12	31	10	38	6	32	9	50	8	67
other	16	25	15	24	10	26	4	15	6	32	4	22	1	8
Gender														
men	5	8	5	8	5	13	2	8	0	0	0	0	1	8
women	13	20	12	19	9	23	1	4	1	5	3	17	3	25
both	47	72	45	73	25	64	23	88	18	95	15	83	8	67
Dietary method														
FFQ	56	86	54	87	35	90	23	88	16	84	14	78	9	75
24 hr recall method	3	5	2	3	0	0	0	0	1	5	2	11	0	0
other	6	9	5	8	4	10	3	12	2	11	2	11	3	25
Timeframe of dietary method														
< half year	17	26	16	26	6	15	11	42	10	52	5	28	4	33
> half year	31	48	30	48	21	54	8	31	5	26	9	50	6	50
Number of items of FFQ														
<35	12	18	11	18	8	21	6	23	3	16	1	6	1	8
35-100	15	23	14	23	9	23	5	19	5	26	6	33	3	25
>100	27	42	27	44	17	44	11	42	6	32	7	39	5	42
Type of input variables														
frequency	44	68	42	68	29	74	17	65	15	79	10	56	4	33
quantity	17	26	16	26	8	21	7	27	2	11	7	39	6	50
Number of input variables														
< 30	23	35	20	32	12	31	8	31	5	26	8	44	3	25
31-50	25	38	25	40	15	38	9	35	8	42	7	39	9	75
51-80	8	12	8	13	6	15	4	15	3	16	2	11	0	0
>80	9	14	9	15	6	15	5	19	3	16	1	6	0	0
Criteria number of factors														
Eigenvalue														
yes	41	63	39	63	26	67	18	69	10	53	10	56	9	75
no	24	37	23	37	13	33	8	31	9	47	8	44	3	25
Screeplot														
yes	29	45	29	47	17	44	12	46	8	42	8	44	7	58
no	36	55	33	53	22	56	14	54	11	58	10	56	5	42
Interpretation														
yes	21	32	21	34	16	41	4	15	7	37	4	22	3	25
no	44	68	41	66	23	59	22	85	12	63	14	78	9	75
Number of factors extracted														
2	12	18	12	19	9	23	0	0	1	5	2	11	0	0
3-4	23	35	22	35	10	26	6	23	7	37	8	44	5	42
5-7	17	26	15	24	9	23	10	38	5	26	6	33	4	33
>8	13	20	13	21	11	28	10	38	6	32	2	11	3	25
Variance explained														
>25 %	24	37	23	37	15	38	11	42	7	37	6	33	5	42
<25 %	26	40	26	42	16	41	11	42	6	32	6	33	4	33
Rotation														
Varimax and/or orthogonal	46	71	43	69	28	72	15	58	14	74	14	78	10	83
promax	2	3	2	3	1	3	0	0	1	5	1	6	0	0
not rotated	1	2	1	2	0	0	0	0	1	5	0	0	0	0
Patterns related to disease[c]														
yes	26	40	26	42	20	51	9	35	5	26	5	28	5	42
no	39	60	36	58	19	49	17	65	14	74	13	72	7	58

(a) Where percentages do not add up to 100%, information was not available on one or more studies; (b) Common labels assigned by the authors of this review to the original identified dietary patterns; (c) cancer, diabetes, heart disease, mortality, "other disease"

characteristics are presented in table 3. Almost all studies showed a positive association between total energy intake and factor scores of the vegetable pattern (14/17, 82%), the meat pattern (5/6, 83%), the sweet pattern (6/7, 86%), the snack pattern (1/1, 100%) and the traditional pattern (3/3, 100%). However, the alcohol pattern did not consistently show a positive association with total energy intake (2/5, 40%). In the majority of studies the vegetable pattern was found to be positively associated with physical activity (14/15; 93%), use of dietary supplements (7/8; 88%) and education (14/17; 82%) and negatively associated with male gender (10/15, 67%) and smoking (22/36, 61%). The meat pattern was in most studies positively associated with male gender (8/13; 67%), age (12/20; 60%), smoking (10/13; 77%) and consumption of alcohol (7/9; 78%). A negative association was in most studies observed between smoking and the sweet pattern (10/13; 77%).
The snack pattern was in most studies positively associated with consumption of alcohol (5/8; 63%) and negatively with age (9/10; 90%). The "Traditional" pattern was positively associated with age (7/10; 70%). The other associations between dietary patterns and sociodemographic or lifestyle characteristics were either based on too few studies or were not consistently positive or negative.

Table 3: Number of studies that observed a significant positive, null or significant negative association between the common dietary patterns and sociodemographic and lifestyle characteristics or chronic disease, published up to January 1st, 2004

	Vegetables			Meat			Sweet			Snacks			Traditional			Alcohol		
	+	0	-	+	0	-	+	0	-	+	0	-	+	0	-	+	0	-
Personal and lifestyle characteristics																		
Energy intake	14	2	1	5	0	1	6	0	1	1	0	0	3	0	0	2	3	0
Gender (male)	2	3	10	8	4	1	2	5	0	3	1	1	1	1	0	0	0	0
Age	15	11	4	1	7	12	3	3	6	0	1	9	7	2	1	2	0	3
BMI	4	21	3	5	8	3	0	4	6	1	5	2	3	4	3	4	0	3
Education	14	1	2	0	3	4	0	0	3	1	4	1	0	2	3	1	2	0
SES	6	8	3	2	2	2	4	4	1	1	6	0	0	3	0	0	0	0
Physical activity	14	1	0	0	5	4	3	2	2	0	0	3	0	0	3	1	1	2
Supplement use	7	0	1	0	0	4	1	2	0	0	1	1	0	2	1	1	1	0
Smoking	2	12	22	10	2	1	1	3	10	4	4	0	3	3	0	4	0	0
Alcohol	3	13	8	7	1	1	0	6	3	5	2	1	2	1	2	0	0	0
Chronic disease																		
Diabetes	0	2	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CHD	0	1	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Cancer	1	11	9	5	6	0	2	3	0	0	0	0	2	0	0	1	3	0
Total mortality	0	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0

Association with chronic disease

Three authors studied the association of dietary patterns and diabetes mellitus (table 3). The vegetable pattern was found to decrease the risk of previously undiagnosed type 2 diabetes (OR: 0.54 95%CI: 0.32-0.91) by Williams *et al*,[50] but not significantly by Gittelsohn *et al*[37] (OR_{Q4}: 0.86 95%CI: 0.42-1.75) and Van dam *et al*[43] (RR_{Q5}: 0.84 95% CI: 0.70-1.00).Van Dam showed an increased risk of diabetes mellitus type 2 by the meat pattern (RR_{Q5}: 1.59 95%CI 1.32-1.93), and Gittelsohn *et al* by the snack pattern (OR_{Q4}: 2.40 95%CI: 1.13-5.10).
The association between dietary patterns and coronary heart disease was studied in three large cohort studies: the Health Professionals' Follow-up Study[41], the Nurses' Health Study[51] and the Danish Monica Surveys.[57] The vegetable pattern was found to decrease the risk of coronary heart disease by Hu *et al*[41] (RR_{Q5}: 0.75 95%CI: 0.59-0.95) and by Fung *et al*[51] (RR_{Q5}: 0.76 95%CI: 0.60-0.98), but not statistically significant by Osler *et al*[56] ($RR_{per\ score\ unit}$: 0.70 95%CI: 0.47-1.03). Similarly, the meat pattern was found to increase

the risk of CHD by Hu *et al*[41] (RR_{Q5}: 1.43 95%CI: 1.01-2.01) and by Fung *et al*[51] (RR_{Q5}: 1.46 95%CI: 1.07-1.99), but not significantly by Osler *et al*[56] ($HR_{per\ score\ unit}$: 1.21 95%CI: 0.80-1.82).
The association of dietary patterns characterized by consumption of vegetables and/or fruits and colon cancer,[28;38;52] colorectal cancer,[62] stomach cancer,[59;73] endometrial cancer,[53] breast cancer,[61] and kidney cancer[64] are presented in figure 1. Overall, dietary patterns characterized by vegetable consumption appeared to be protective against cancer. Possible sources of heterogeneity were investigated by stratification (figure 2). The protective effect was not observed for kidney cancer, but appeared not to differ significantly by study design, gender, region, number of input variables or number of extracted patterns.
The association of dietary patterns characterized by meat and fat foods and colon cancer, [28;38;52] colorectal cancer,[61] stomach cancer,[73] endometrial cancer,[53] breast cancer,[62] and kidney cancer[64] are presented in figure 3. Overall, dietary patterns characterized by meat and fat foods appeared to increase the risk of cancer. However, when stratified according to possible sources of heterogeneity (figure 4), the meat pattern appeared to be associated with increased risk of colorectal and kidney cancer only, only in case control studies but not in cohort studies, only in studies originating from the USA and in studies in which more than five dietary patterns were extracted. The association was not significantly different for men and women, and for studies using more or less than 68 input variables.
Too few data are available on the association between cancer and dietary patterns characterized by consumption of sweet foods, alcohol, or traditional food items.
The vegetable pattern was associated with a decreased risk of all cause mortality (table 3). Kumagai *et al*[44] observed a decreased risk (RR: 0.77 95%CI: 0.63-0.95) of all cause mortality in 7 years. Osler[56] observed a decreased risk of 15 years all cause mortality both in men ($RR_{per\ 1\ SD}$: 0.84 95%CI: 0.75-0.93) and women ($RR_{per\ 1\ SD}$: 0.74 95%CI: 0.64-0.85). Neither Kumagai[44] nor Osler[56] found the meat pattern to be associated with all cause mortality.

DISCUSSION

In total we reviewed 68 publications in which 65 sets of dietary patterns were identified by factor analysis. We allocated the dietary patterns to six distinct common patterns that we labeled as “Vegetables”, “Meat" “Alcohol”, “Snacks”, “Traditional” and “Sweet".
It has been argued that patterns should be judged as “similar” only when the food groups significantly contributing are similar and their factor loadings are comparable in both direction (positive/negative) and order. [60] However, variations in food groups loading on a dietary pattern can arise because of differences in methodology or simply because of random statistical variations. Furthermore, as the sets of input variables are not the same across studies, the food groups and factor loadings are not expected to be either. Therefore, the reclassification of the original dietary patterns into common patterns was not based on strictly quantitative criteria. Because the behavior underlying the choice of foods does not depend on the exact choice of individual foods, we judged the patterns as "similar" on basis of agreement at the level of food groups. Others could make a different classification of these dietary patterns or identify more patterns in addition to these six patterns, however, we tried to avoid overlapping categories and we consider the common patterns recognizable and meaningful in nutritional epidemiology.

Figure 1 Forest plot of risk ratios and 95% confidence intervals of cancer, comparing the highest vs. the lowest category of scores on a dietary pattern characterized by consumption of ***vegetables and/or fruits***, published up to January 1st 2004. Reference line indicates no association.
Legend: ♦ men, ◊ women. * per 1 sd

Figure 2. Forest plot of crude summary risk ratios and 95% confidence intervals of cancer for dietary patterns characterized by ***vegetables and/or fruits*** consumption, by study design, tumor site, gender, region, number of input variables and number of extracted factors in the exploratory factor analysis, for studies published up to January 1st 2004. Reference line indicates no association.

Figure 3 Forest plot of risk ratios and 95% confidence intervals of cancer, comparing the highest vs. the lowest category of scores on a dietary pattern characterized by consumption of ***meat and fat foods***, published up to January 1st 2004. Reference line indicates no association.
Legend: ♦ men, ◊ women. * per 1 sd

Figure 4. Forest plot of crude summary risk ratios and 95% confidence intervals of cancer for dietary patterns characterized by ***meat and fat foods*** consumption, by study design, tumor site, gender, region, number of input variables and number of extracted factors in the exploratory factor analysis, for studies published up to January 1st 2004. Reference line indicates no association.

We systematically studied whether the identified dietary patterns differed by population and study characteristics and arbitrary factor analytical decisions, as has been argued.[3] Although the number of studies using the 24-hr recall method was low, dietary patterns appeared to differ from FFQ based dietary patterns. An explanation for this is might be that the food choice of 1 specific day is menu-oriented, whereas a food frequency questionnaire refers to the usual diet over a certain period of time and is more driven by food preferences. Furthermore there is also the possibility that items within groups of questions within a questionnaire (such as several types of vegetable items, meat items, alcohol items or sweet items) are artificially highly correlated and hence combine in spurious factors. However, the dietary patterns appear to be broader than just these correlated items. For example, the "vegetables pattern" in many cases, also includes products such as high fiber instead of refined products, and low-fat products instead of high-fat products, fish and poultry instead of red meat, and oil and wine. Moreover, in a subsample from the Health Professionals Follow-up Study qualitatively similar eating patterns were observed across FFQs and diet records.[40] It is likely, therefore, that these dietary patterns are not merely caused by the structure of FFQs.
We did not observe a strong effect of level of aggregation of the input variables on the identified dietary patterns. This is in agreement with McCann *et al*[53] who demonstrated that reducing detailed dietary data obtained from a single FFQ prior to pattern analysis did not affect the character of the "Healthy" and "High fat" dietary patterns.
The criteria used to determine the number of extracted factors did not seem to influence the dietary patterns, whereas the actual number of extracted patterns did. If two factors were extracted, they were most likely a vegetable-based pattern and a meat-based pattern. When extracting more patterns the sweet pattern was more likely to be identified, but this did not influence the previously extracted vegetable or meat pattern. This is in agreement with sensitivity analyses showing that the identified patterns were robust for number of factors extracted.[67] However, extracting too many factors should be avoided to ensure interpretability of the patterns. All patterns are by definition orthogonal and therefore conditional on each other. Thus a subject with a high score on a one pattern is not more likely to have a low score on another pattern than a person with low score on that first pattern. This might not be realistic when food items are not only correlated within but also across patterns.

Dietary patterns are thought to be part of broader lifestyles and are expected to interact with other lifestyle behaviors.[3;14;20;50;60] We performed a qualitative tally (counting the number of studies that observed a significant positive, null or significant negative association) to explore the reported associations of dietary patterns with lifestyle variables. This method can be misleading because the strength of the associations is not taken into account and studies finding no association may have tended on the average towards a positive (or negative) association.[77] Furthermore, reported null findings may have been caused by a lack of power and a tally may obscure the role of bias. However, quantitative pooling of the association between dietary patterns and lifestyle characteristics is beyond the scope of this paper and therefore this review should be considered only as pattern recognition, inviting further research.
We found that sociodemographic and lifestyle characteristics that are considered health-related such as (not) smoking, physical activity, use of dietary supplements and education were positively associated with the dietary pattern characterized by consumption of vegetables and/or fruits. Male gender, younger age, smoking and consumption of alcohol

were positively associated with the meat pattern. Smokers appeared to have higher scores on the alcohol pattern and lower scores on the sweet pattern. The traditional pattern was mostly consumed by older people and consumption of alcohol was positively associated with the (salty) snack pattern. These observed relationships increase the confidence that the statistical solution of factor analysis is meaningful, despite the fact that it might not be possible to completely separate the effects of the dietary patterns from the effects of the lifestyle behaviors.[50]

We used forest plots to present the published associations between cancer and a dietary pattern characterized by consumption of vegetables and or fruits and a dietary pattern characterized by meat and fat foods. Because of the small number of studies, the different study populations, the different study designs and using all types of cancer as outcome, we thought it not appropriate to calculate a summary risk estimate. The dietary pattern characterized by vegetable consumption nevertheless consistently appeared to decrease the risk of diabetes mellitus type II, coronary heart disease, several types of cancer and all-cause mortality. The pattern characterized by meat and fat foods appeared to increase the risk of diabetes mellitus type II, coronary heart disease and colon cancer.
In conclusion, the dietary patterns extracted in the different studies revealed many similarities, despite differences in the data collection methods and subjective factor analytical decisions, such as number and type of food variables included in the analyses and number of factors extracted. Moreover, the association of the dietary patterns with sociodemographic and lifestyle characteristics and chronic disease and mortality were comparable. This suggests that factor analysis is successful in identifying meaningful, common underlying patterns in food consumption in different populations.

REFERENCES

1. Jacobson HN, Stanton JL. Pattern analysis in nutrition. Clin.Nutr. 1986;5:249-53.

2. Nestle M. Fruits and vegetables: protective or just fellow travelers? Nutr.Rev. 1996;54:255-7.

3. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity. Am.J.Epidemiol. 1998;148:17-9.

4. Kant AK. Indexes of overall diet quality: a review. J.Am.Diet.Assoc. 1996;96:785-91.

5. Blackburn GL. Treatment approaches: food first for weight management and health. Obes.Res. 2001;9 Suppl 4:223S-7S.

6. Gerber M. The comprehensive approach to diet: a critical review. J.Nutr. 2001;131:3051S-5S.

7. Togo P, Osler M, Sorensen T, Heitmann B. Food intake patterns and body mass index in observational studies. Int.J.Obes. 2001;25:1741-51.

8. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr.Opin.Lipidol. 2002;13:3-9.

9. Schulze MB, Hu FB. Dietary patterns and risk of hypertension, type 2 diabetes mellitus, and coronary heart disease. Curr.Atheroscler.Rep. 2002;4:462-7.

10. Clark, O. and Djulbegovic, B. Forest plots in excel software (Data sheet). Available at www.evidencias.com. 2001.

11. Pelto, G. H., Pelto, P. J., and Lung'aho, M. S. "Modern" and "traditional" food use in West Finland: an example of quantitative pattern analysis. Nutrition Research 1, 63-71. 1981.

12. Schwerin HS, Stanton JL, Riley AM, Jr. et al. Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. Am.J.Clin.Nutr. 1981;34:568-80.

13. Schwerin HS, Stanton JL, Smith JL, Riley AM, Jr., Brett BE. Food, eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. Am.J.Clin.Nutr. 1982;35:1319-25.

14. Stellman SD. Chairman's remarks. Natl.Cancer Inst.Monogr. 1983;67:145-7.

15. Iizumi H, Amemiya T. Effect of food consumption pattern on total serum cholesterol level: a methodological approach. J.Nutr.Sci.Vitaminol.(Tokyo) 1986;32:205-18.

16. Iizumi H, Amemiya T. Physical conditions affecting food consumption patterns. Int.J.Vitam.Nutr.Res. 1986;56:411-7.

17. Iizumi H, Amemiya T. Eleven-year follow-up of changes in individuals' food consumption patterns. Int.J.Vitam.Nutr.Res. 1986;56:399-409.

18. Gex-Fabry M, Raymond L, Jeanneret O. Multivariate analysis of dietary patterns in 939 Swiss adults: sociodemographic parameters and alcohol consumption profiles. Int.J.Epidemiol. 1988;17:548-55.

19. Nicklas TA, Webber LS, Thompson B, Berenson GS. A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors: the Bogalusa Heart Study. Am.J.Clin.Nutr. 1989;49:1320-7.

20. Barker ME, McClean SI, Thompson KA, Reid NG. Dietary behaviours and sociocultural demographics in Northern Ireland. Br.J.Nutr. 1990;64:319-29.

21. Barker ME, McClean SI, Strain JJ, Thompson KA. Dietary behaviour and health in Northern Ireland: an exploration of biochemical and haematological associations. J.Epidemiol.Community Health 1992;46:151-6.

22. Hebert JR, Kabat GC. Differences in dietary intake associated with smoking status. Eur.J.Clin.Nutr. 1990;44:185-93.

23. Hebert JR, Kabat GC. Implications for cancer epidemiology of differences in dietary intake associated with alcohol consumption. Nutr.Cancer 1991;15:107-19.

24. Kristal AR, Shattuck AL, Henry HJ. Patterns of dietary behavior associated with selecting diets low in fat: reliability and validity of a behavioral approach to dietary assessment. J Am Diet Assoc 1990;90:214-20.

25. Randall E, Marshall JR, Graham S, Brasure J. Patterns in food use and their associations with nutrient intakes. Am.J.Clin.Nutr. 1990;52:739-45.

26. Randall DE, Marshall JR, Brasure J, Graham S. Patterns in food use and compliance with NCI dietary guidelines. Nutr.Cancer 1991;15:141-58.

27. Randall E, Marshall JR, Graham S, Brasure J. High-risk health behaviors associated with various dietary patterns. Nutr.Cancer 1991;16:135-51.

28. Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in western New York. Nutr.Cancer 1992;18:265-76.

29. Gallagher ML, Farrior E, Broadhead L et al. Development and testing of food frequency recall instrument for describing dietary patterns in adults and teenagers. Nutr.Res. 1993;13:177-88.

30. Hebert JR, Stoddard AM, Harris DR et al. Measuring the effect of a worksite-based nutrition intervention on food consumption. Ann.Epidemiol. 1993;3:629-35.

31. Nolan CC, Gray-Donald K, Shatenstein B, O'Loughlin J. Dietary patterns leading to high fat intake. Can-J-Public-Health 1995;86:389-91.

32. Wolff CB, Wolff HK. Maternal eating patterns and birth weight of Mexican American infants. Nutr-Health 1995;10:121-34.

33. Whichelow MJ, Prevost AT. Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. Br-J-Nutr 1996;76:17-30.

34. Prevost AT, Whichelow MJ, Cox BD. Longitudinal dietary changes between 1984-5 and 1991-2 in British adults: association with socio-demographic, lifestyle and health factors. Br.J.Nutr. 1997;78:873-88.

35. Gray-Donald K, O'Loughlin J, Richard L, Paradis G. Validation of a short telephone administered questionnaire to evaluate dietary interventions in low income communities in Montreal, Canada. J.Epidemiol.Community Health 1997;51:326-31.

36. Beaudry M, Galibois I, Chaumette P. Dietary patterns of adults in Quebec and their nutritional adequacy. Can-J-Public-Health 1998;89:347-51.

37. Gittelsohn J, Wolever TM, Harris SB, Harris-Giraldo R, Hanley AJ, Zinman B. Specific patterns of food consumption and preparation are associated with diabetes and obesity in a Native Canadian community. J.Nutr. 1998;128:541-7.

38. Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. Am.J.Epidemiol. 1998;148:4-16.

39. Bree MB, Eaves LJ, Dwyer JT. Genetic and environmental influences on eating patterns of twins aged >/=50 y. Am-J-Clin-Nutr 1999;70:456-65.

40. Hu FB, Rimm E, Smith-warner SA et al. Reproducibility and validity of dietary patterns assessed with a food- frequency questionnaire. Am.J.Clin.Nutr. 1999;69:243-9.

41. Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. Am.J.Clin.Nutr. 2000;72:912-21.

42. Fung TT, Rimm EB, Spiegelman D et al. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. Am.J.Clin.Nutr. 2001;73:61-7.

43. van Dam RM, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Dietary Patterns and Risk for Type 2 Diabetes Mellitus in U.S. Men. Ann.Intern.Med. 2002;136:201-9.

44. Kumagai S, Shibata H, Watanabe S, Suzuki T, Haga H. Effect of food intake pattern on all-cause mortality in the community elderly: a 7-year longitudinal study. J.Nutr.Health Aging 1999;3:29-33.

45. Matalas AL, Franti CE, Grivetti LE. Comparative study of diet and disease prevalence in Greek Chians - Part I rural and urban residents of Chios. Ecology of food and nutrition 1999;38:351-80.

46. Matalas, A. L., Franti, C. E., and Grivetti, L. E. Comparative study of diets and disease prevalence in Greek Chians - Part II Chian immigrants to Athens and to the United States. Ecology of Food and Nutrition 38, 381-414. 1-1-0099.

47. Maskarinec G, Novotny R, Tasaki K. Dietary patterns are associated with body mass index in multiethnic women. J.Nutr. 2000;130:3068-72.

48. Mullen K, Williams R, Hunt K. Irish descent, religion and food consumption in the west of Scotland. Appetite 2000;34:47-54.

49. Tseng M, DeVellis RF, Maurer KR et al. Food intake patterns and gallbladder disease in Mexican Americans. Public Health Nutr. 2000;3:233-43.

50. Williams DE, Prevost AT, Whichelow MJ, Cox BD, Day NE, Wareham NJ. A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. Br-J-Nutr 2000;83:257-66.

51. Fung TT, Willett WC, Stampfer MJ, Manson JE, Hu FB. Dietary patterns and the risk of coronary heart disease in women. Arch.Intern.Med. 2001;161:1857-62.

52. Fung T, Hu FB, Fuchs C et al. Major dietary patterns and the risk of colorectal cancer in women. Arch.Intern.Med. 2003;163:309-14.

53. McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr. 2001;4:989-97.

54. McCann SE, Weiner J, Graham S, Freudenheim JL. Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? Public Health Nutr. 2001;4:903-8.

55. Ogura M, Yamamoto T, Morita M, Watanabe T. A case-control study on food intake of patients with recurrent aphthous stomatitis. Oral Surg.Oral Med.Oral Pathol.Oral Radiol.Endod. 2001;91:45-9.

56. Osler M, Heitmann BL, Hoidrup S, Jorgensen LM, Schroll M. Food intake patterns, self rated health and mortality in Danish men and women. A prospective observational study. J.Epidemiol.Community Health 2001;55:399-403.

57. Osler M, Heitmann BL, Gerdes LU, Jorgensen LM, Schroll M. Dietary patterns and mortality in Danish men and women: a prospective observational study. Br.J.Nutr. 2001;85:219-25.

58. Osler M, Helms AA, Heitmann B et al. Food intake patterns and risk of coronary heart disease: a prospective cohort study examining the use of traditional scoring techniques. Eur.J.Clin.Nutr. 2002;56:568-74.

59. Palli D, Russo A, Decarli A. Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control 2001;12:163-72.

60. Schulze MB, Hoffmann K, Kroke A, Boeing H. Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. Br.J.Nutr. 2001;85:363-73.

61. Terry P, Suzuki R, Hu FB, Wolk A. A prospective study of major dietary patterns and the risk of breast cancer. Cancer Epidemiol.Biomarkers Prev. 2001;10:1281-5.

62. Terry P, Hu FB, Hansen H, Wolk A. Prospective study of major dietary patterns and colorectal cancer risk in women. Am.J.Epidemiol. 2001;154:1143-9.

63. Tseng M, DeVellis RF. Fundamental dietary patterns and their correlates among US whites. JADA 2001;101:929-32.

64. Handa K, Kreiger N. Diet patterns and the risk of renal cell carcinoma. Public Health Nutrition 2002;5:757-67.

65. Mishra G, Ball K, Arbuckle J, Crawford D. Dietary patterns of Australian adults and their association with socioeconomic status: results from the 1995 National Nutrition Survey. Eur.J.Clin.Nutr. 2002;56:687-93.

66. Sichieri R. Dietary patterns and their associations with obesity in the brazilian city of rio de janeiro. Obes.Res. 2002;10:42-8.

67. Balder HF, Virtanen M, Brants HAM et al. Common and country-specific dietary patterns in four European cohort studies. J.Nutr. 2003;133:4246-51.

68. Bell RA, Quandt SA, Vitolins MZ, Arcury TA. Dietary patterns of older adults in a rural, tri-ethnic community: a factor analysis approach. Nutrition Research 2003;23:1379-90.

69. Costacou T, Bamia C, Ferrari P, Riboli E, Trichopoulos D, Trichopoulou A. Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population. Eur.J.Clin.Nutr. 2003;57:1378-85.

70. Chi, D., Nakano M., and Yamamoto, K. Correlates of serum high-density lipoprotein cholesterol: a community-based study of middle-aged and older men and women in Japan. Asia-Pacific Journal of Public Health 15(1), 17-22. 2003.

71. Kerver, J. M., Yang, E. J., Bianchi, L., and Song, W. O. Dietary patterns associated with risk factors for cardiovascular disease in healthy US adults. Am J Clin Nutr 78, 1103-1110. 2003.

72. Markaki I, Linos D, Linos A. The influence of dietary patterns on the development of thyroid cancer. Eur.J.Cancer 2003;39:1912-9.

73. Masaki M, Sugimori H, Nakamura K, Tadera M. Dietary patterns and stomach cancer among middle-aged male workers in Tokyo. Asian Pac.J.Cancer Prev. 2003;4:61-6.

74. Sanchez-Villegas A, Delgado-Rodriguez M, Martinez-Gonzalez MA, Irala-Estevez J. Gender, age, socio-demographic and lifestyle factors associated with major dietary patterns in the Spanish Project SUN (Seguimiento Universidad de Navarra). Eur.J.Clin.Nutr. 2003;57:285-92.

75. Schulze, M. B., Hoffmann, K., Kroke, A., and Boeing, H. Risk of hypertension among women in the EPIC-Potsdam Study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. Am J Epidemiol 158, 365-373. 2003.

76. Dam, R. M. v., Grievink, L., Ocke, M. C., and Feskens, E. J. Patterns of food consumption and risk factors for cardiovascular disease in the general Dutch population. Am J Clin Nutr 77(5), 1156-63. 2003.

77. Greenland S. Meta-analysis. In: Rothman KJ, Greenland S, eds. Modern Epidemiology. NewYork: Lippincott Williams & Wilkins Publishers 1998.

78. Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. Br.J.Nutr. 2003;89:409-19.

Appendix Table 1: Comparison of dietary patterns characterized by consumption of vegetables and/or fruits, identified by factor analysis in published studies up to January 1st, 2004

Ref	Label	food items
11	1 Modern food use	variety of vegetables, servings of vegetables, salad servings, citrus fruits, variety of fruits, sausage servings, high protein foods
14	1 Vitamin-rich foods	green leafy vegetables, raw vegetables, tomatoes, carrots, cabbage, citrus fruit and juices, cheese, squash
15-17	8 Nutrition-consciousness	green & yellow vegetable, soybean products, dairy products, *-dining out*
18	2 Healthfulness	vegetables, fish, poultry, grapes and bananas, fruit, citrus fruit, fruit juice, seafood, rice, *-sugar, -jam*
19	VI Fruit drinks	fruit-flavored soft drinks; fruit or vegetable juice
	VII Fruits, vegetables	canned, frozen, fresh vegetables (prepared without meat); canned, frozen, fresh vegetables (prepared with meat); fresh fruit/dried fruit; beans (red, baked, butter)
20	Component 2 (Cosmopolitan)	fruit, vegetables, rice and pasta, egg and cheese dishes, *- chips, - sausages, peas and beans, savoury pies*
22;23	Factor 1	cauliflower, brussel sprouts, broccoli, spinach, vegetables, carrots, cabbage
	Factor 2	salad, tomatoes, cheese
	Factor 3	fruit, cantaloupe
	Factor 8	vegetable soup, tomatoes, tomato sauce, orange juice
25 men	1 Salad	green pepper, lettuce, celery, cucumber, tomatoes, radishes, onions, carrots
	2 Fruit	peaches, pears, plums, apples, orange and tangerine, cherries, grapes, melon, banana, berries, apricots, dried fruit, *- beer*
	3 Staple vegetables	green beans, wax beans, peas, lima beans, turnips, beets, corn, potatoes, hot dogs, cabbage
	4 Healthful	broccoli, cauliflower, cottage cheese, spinach, poultry, asparagus, fin fish, summer squash, brussels sprouts, unsweetened cereal, *- white bread*
	9 Health foods	oat products, cooked cereal, crackers, *- coffee, - pork chop*
25 women	1 Salad	lettuce, onions, celery, cucumber, tomatoes, green pepper, radishes, carrots, apples
	2 Southern European / healthful	mushrooms, spinach, cauliflower, summer squash, veal, eggplant, broccoli, lamb, poultry, liver, fin fish, green beans, pineapple
	3 Fruit	peaches, pears, plums, grapes, orange and tangerine, apples, grapefruit, cherries, melon, *- shellfish, - distilled alcohol*
	6 Staple vegetables	lima beans, peas, corn, green beans, oat products, potatoes, beets, cooked cereal, wax beans, turnips
	8 Health foods	cooked cereal, dried fruit, apricots, yogurt, oat products, whole-wheat bread, peanut butter, crackers, *- white bread, - coffee*
26 27;28men	1 Salad	lettuce, celery, green pepper, cucumber, tomatoes, radishes, onions, carrots
	2 Fruit	peaches, plums, pears, apples, oranges or tangerines, cherries, grapes, melons, bananas, berries, *- beer*
	3 Healthful	broccoli, spinach, cauliflower, rice, poultry, mushrooms, green beans, summer squash, asparages *- potatoes,- cookies, -white bread*
26 27;28women	1 Salad	lettuce, onion, cucumber, tomato, celery, green pepper, radishes, carrot, apples, broccoli
	2 Healthful	green beans, cauliflower, spinach, beets, summer squash, pineapple, liver, wax beans, broccoli, winter squash, cottage cheese, mushrooms, peas, cabbage
	4 Fruit	pears, peaches, plums, oranges or tangerines, apples, bananas, grapes, grapefruit *- distilled alcohol, -beer, -wine*
29 teenagers	2	other vegetables, yellow/orange vegetables, non-diary milkproducts, red meats, organ meats
29 adults	2	yellow/orange vegetables, fish, fruits and fruit juice, other vegetables, whole milk
30	1 Fruit	apple/pears, oranges, bananas, peaches/apricots, other fruit
	2 Vegetables	cabbage, broccoli, string beans, spinach, carrots

Appendix table 1 continued

31	2 "prudent diet"	whole grain bread, vegetables, fruits, low fat dressing, low fat milk, soup, yoghurt, - potatoes, - white bread
32	1 Nutrient dense	vitamin A & C rich fruits and vegetables, other fruits, low fat dairy products, other vegetables
33	Component 1 ("fresh foods, high fibre foods, and low-fat spread and milk")	summer fruit, winter fruit, summer salad, winter salad, green vegetables, root vegetables, other vegetables, - *chips,* brown bread, low fat spread, semi-skimmed/skimmed milk, - *processed meat, - fried food,* fruit juice
35	1 Avoid fat	fresh fruit as snacks, fruit for dessert, raw vegetables for snacks, a vegetarian dinner, cooked vegetables without butter or margarine, bread rolls or muffins without butter or margarine, green salad with no dressing
36	3 Health conscious	whole grain breads and cereals, low fat dairy products, fresh/baked/frozen vegetables, fresh/baked/frozen fruits, fruit juices, - *white breads*, - *butter*
37;37	Vegetable	peas, carrots, corn, other vegetables
38 men	2 Prudent diet	fresh fruit, legumes, cruciferous vegetables, salad, carrots/tomatoes, other vegetables
38 women	2 Prudent diet	fresh fruit, cruciferous vegetables, salad, carrots/tomatoes, other vegetables
39	Factor 2 ("variety of vegetables, fruit, rice, yogurt, skim milk and dark bread – a healthful eating pattern")	grapefruit, raw spinach, fresh peaches and apricots, fish broiled or baked (not tuna), cauliflower or brussels sprouts, broccoli, frozen yogurt, oranges, other vegetables, winter squash and baked squash, other fruit, fresh strawberries, high-fiber cereals, rice, flavored yogurt, cottage cheese, dark bread, cantaloupe, cooked spinach, skim milk or butter milk, canned/frozen/dried peaches, cooked cereals, orange or grapefruit juice, tuna fish salad and casserole, sweet potatoes and yams, apples/applesauce/pears, corn breads, watermelon, shell fish, tea, wine, lemon in tea, beans (ecxl string beans, green beans and chilli), tomatoes and tomato juice, soups (excl vegetable and tomato soup), carrots
40-43	1 Prudent pattern	other vegetables, tomatoes, green leafy vegetables, oil and vinegar salad dressings, cruciferous vegetables, legumes, fish and other seafood, other soup, dark yellow vegetables, whole grains, condiments, fruit, garlic, poultry
44	factor 1 "plant food"	potatoes, seaweeds, green yellow vegetables, soybean products, fruit
45	8 Fruits and vegetables	fruits, fresh vegetables, bread
46	6 Vegetables, fruits	salad vegetables, fruits and fruit juice, steamed vegetables
47	2 Vegetable	dark yellow vegetables, green leaf vegetables, other vegetables, legumes, fruit
48	1 Fruit, yoghurt, vegetables	fruit in summer, fruit in winter, raw vegetables in summer, raw vegetables in winter, yoghurt, fruit juice
49 men	Vegetable	carrots, salad, broccoli, greens, other vegetables, cauliflower, other fruits, melon, citrus fruit, yellow fruit, soups and stews, cabbage, other juice, citrus juice, fish, chicken
49 women	Vegetable	carrots, salad, broccoli, greens, other vegetables, cauliflower
	Fruit	yellow fruits, citrus juice, other juices, ice cream, melon
50	Component 1	summer fruit, winter fruit, summer salad, winter salad, fish not fried, other vegetables, poultry, green vegetables, pasta/rice, - *crisps), - cakes, - potatoes, - fried fish, - sausage/processed meat, - fried food*
51	1 Prudent	other vegetables, leafy vegetables, cruciferous vegetables, yellow vegetables, fruits, fish, legumes, tomatoes, poultry, garlic, salad dressings, whole grains, low-fat dairy
53;54	1 Healthy	carrots, pineapple, broccoli, spinach, green peppers, cottage cheese, summer squash, grapefruit, peaches, pears, cauliflower, prunes or plums, mushrooms, beets, grapes, asparagus, berries, melon, fish not fried, apples/apple sauce, red peppers, sweet potatoes, cherries, white potatoes
55	factor I	japanes radishes, seaweed, spinach, mushrooms, potatoes, tofu, fish, seeds, miso, fermented soybeans
56;57	1 Prudent	wholemeal bread ; wholemeal ryebread ; porride and groats ; pasta ; rice ; vegetables raw ; vegetables boiled ; fruit ; juice ; jam, honey ; cakes, cookies ; candy, chocolate ; tea ; - *white bread; - white ryebread*
59	1 Vitamin-rich	sugar, fibre, vitamin C, beta-carotene, vitamin E, nitrates
60 men	Fruit and vegetables	olive oil, fresh fruit, raw vegetables, other vegetable oils and fats
60 women	Fruit and vegetables	olive oil, fresh fruit, raw vegetables, other vegetable oils and fats
61;62	1 Healthy	vegetables, fruits, fish, whole grains, low-fat dairy foods, poultry, cereal, eggs, meats
63	1 Vegetable-fruits	tossed salad, broccoli, carrots, other fruits (besides citrus, melon and peaches), brussels sprouts and cauliflower, spinach/greens/collards/kale, peppers, chicken, cabbages, dark breads and rolls, rice, soup/stew, fish, peaches/nectarines/apricots/guava/mango/papaya
64 men	1 Fruits/vegetables	carrots, cabbage, broccoli, cantaloupe, spinach or other greens, yellow squash, other vegetables, other fruit (fresh, canned), tomatoes, baked beans/lentils, apples/pears, soups with vegetables, fish (smoked, salted, dried), oranges, chicken/turkey, bananas
	5 Juice	frozen juice/drink, powdered drinks, other fresh juice, tomato/vegetable juice, fresh citrus juice

Appendix table 1 continued

64 men	6 Healthful	apples/pears, bran or granola cereals, dark bread, bananas, *- french fries, - yellow squash, - cooked cereals, - white bread*
	11 Fruit	oranges, bananas, apples/pears, white bread, *- wine, - potatoes, - -skimmed milk, - liquor*
64 women	1 Fruits/vegetables	other vegetables, carrots, spinach or other greens, other fruit (fresh, canned), cabbage, apples/pears, broccoli, potatoes, oranges, tomatoes, bran or granola cereals, bananas, fish (fresh, frozen, canned)
65 men	1 Salad vegetables	lettuce, tomato, celery, onion, side salad, zucchini, vegetable sandwich
	3Tropical fruits	grapes, pineapple, mango, melon, peach
	4 Other vegetables	brussel sprouts, spinach, broccoli, sweet potato
	5 Traditional vegetables	peas, potato, green beans, pumpkin
	10 Traditional fruit	orange, apple, banana
65 women	1 Traditional vegetables	cauliflower, green beans, brussel sprouts, pumpkin, broccoli, peas, spinach, carrots, potato
	2 Salad	lettuce, celery, tomato, side salad, vegetable sandwich, onion
	3 Fruit	grapes, peach, melon, mango, orange, apple, banana
	9 Fruit and vegetable juice	fruit drink, fruit juice, vegetable juice
66 men	1 Mixed	cereals, fruits, vegetables
67 NL men	1 Salad vegetables	oil, pasta, mushrooms, allium, tomatoes, rice, wine, leaf veg raw
	2 Cooked vegetables	cabbages, leaf veg cooked, legumes, carrots, allium
67 NL women	1 Salad vegetables	cabbages, allium, legumes, leaf veg raw, carrots, leaf veg cooked, tomatoes, oil, rice, pasta
67 FI	1 Salad vegetables	dressings, leaf veg raw, tomatoes, cabbages, carrots, rice, citrus fruit, poultry, berries, *-nonferm fat milk*
67 SE	1 Salad vegetables	leaf veg raw, carrots, tomatoes, apples/pears, cabbages, citrus fruit, fish
67 IT	1 Salad vegetables	leaf veg raw, dressings, tomatoes, oil, carrots
	3 Cooked vegetables	legumes, cabbages, leaf veg cooked, fish, carrots, rice, potatoes, poultry
68	3 Fall/winter healthy foods	leafy greens, carrots/winter squash/other vegetables, apples, *-sweets*
	5 Vitamin A foods	sweet potatoes, tomatoes, cantaloupe
	6 Spring/summer healthy foods	green beans/peas, peaches, broccoli/cauliflower, high fiber grain products
69	1 Mediterranean type diet	olive oil, vegetables, legumes, fish, fruit, *-vegetable seed oils, - margarine, - sugar, - cereal*
	2 Vegetarian diet	vegetable seeds oils, vegetables, other dairy, fruit, margarine, *-potatoes, -meat, - butter, -olive oil*
70	1 Prudent	Fruits, bean foods, vegetables, fishes, red meat, milk and milk products
72	1 Fruits	peaches, cherries, melon, apricots, pears, apples, tangerine, water melon, grapes, oranges, strawberries, lemons, figs
	2 Raw vegetables	cucumbers, raw fresh onions, raw peppers, raw tomatoes, raw lettuce, figs
	4 Fish and cooked vegetables	cauliflower, pineapple, leeks, grapefruit, dried fruits
	17 Mixed vegetables and fruits	okra, aubergines, green beans, courgettes, fish, artichokes, cooked peppers
71	2 American Healthy	green leafy vegetables, salad dressings, tomatoes, other vegetables, cruciferous vegetables, tea, dark-yellow vegetables, potatoes
	5 Southwestern	legumes, refined grains, soup, other vegetables, tomatoes
73	1 Vegetable and fruit	cabbage/lettuce, green leafy vegetables, carrots, tomatoes, chinese cabbage, tomatoes, other fruit, seaweeds, oranges, bean curd (raw or fried)
74	2 "Spanish-Mediterranean"	vegetables, fish and other seafood, fruit, meat products, poultry, olive oil, legumes, nuts, potatoes
75	2 Fruits and vegetables	raw vegetables, vegetable oil, fruits
76	1 Cosmopolitan pattern	vegetable oils, garlic, fried vegetables, white rice, salad vegetables, brown rice, chicken, fish, wine, pasta, warm sauces, organ meat, pizza, soy products, water, mayonnaise, *- potatoes, - added fat (<.35 SFA/g fat), - added sugar, - dairy (>0.02 g fat/g), - coffee, - pastries*

- negative factor loadings are represented by a minus sign preceding the food item, and the food item in *italic* font

Appendix Table 2: Comparison of dietary patterns characterized by consumption of meat and fat foods, identified by factor analysis in published studies up to January 1st, 2004

Ref	Label	food items
14	2 High fat meat and eggs	pork, franks, sausage, eggs, ham, smoked meats, beef
15-17	4 Animal foods	beef and pork, *-fruits*
	5 Prepared foods	processed foods, other meats, eggs
19	I Seafood, meats	tuna fish, salmon, fish sticks; ribs, liver; fish (flounder, trout etc); wild game; hotdogs, corndogs; bacon, salt pork, fatback; crawfish, crabmeat, oysters, shrimp, butter (all brands); ham; luncheon meat; nuts (peanuts, walnuts etc); pork roast, chops, pickled pork
	III Fats, pasta	lard, vegetable fat; margarine; bacon, beef, pork drippings; gravy (pork, chicken, beef, tomato etc); rice, noodles, macaroni, lasagna, spaghetti
	IV Beef, chicken	beef roast, steaks, stew; ground beef (as in meatloaf), casseroles, patties, tacos; chicken, turkey; potato
	X Eggs, sausage	eggs; sausage (smoked, patties, links)
22;23	Factor 7	liver, eggs, sweet potato, cabbage
25 men	5 Eastern European	sausage, bacon, pork roast, sauerkraut, pork chops, distilled alcohol, corned beef, beer, ham, eggs, *- unsweetened cereal*
26 27;28men	6 High fat	eggs, bacon, sausage, steak, beer, salami or pepperoni, distilled alcohol, *- unsweetened cereal, -winter squash, -summer squash,- cookies*
30	3 Ground and processed meat	hamburgers, processed meats, hot dogs
	5 Meat and potatoes	beef/pork/lamb sandwich or mixed dish, beef/pork/lamb main dish, potatoes
38 men	1 Western diet	processed meats, red meat, eggs, refined grain, added sugar
38 women	1 Western diet	processed meats, red meat, fast-food meats, eggs, margarine, potatoes, refined grain
39	Factor 1 ("foods high in fat, salt and sugar – a less healthful eating pattern")	bacon, gravies, ham and lunchmeats, french fries and fried potatoes, sausage, pork incl chops and roast, fried chicken, salty snacks, hot dogs, pies (not pumpkin or sweet potato), mixed dishes with cheese, beef steaks and roasts, hamburgers/cheeseburgers/meatloaf, eggs, fried fish or fish sandwich, biscuits/muffins/burger rolls, doughnuts/cookies/cakes/pastries, cheese and cheese spreads (excl cottage cheese), regular soft drinks, ice cream, beef stew and pot pie with vegetables, white bread/bagels/crackers, chocolate candy, peanuts and peanut butter, butter or margarine on vegetables or potatoes, pizza, candy (nonchocolate)/jelly/honey/brown sugar, pumpkin pie and sweet potato pie, butter on bread or rolls, sugar added to cereal, whole milk and drinks with whole milk, cold cereals (not high in fiber or highly fortified), liverwurst, corn breads
40 41 42;43	2 Western pattern	butter, processed meat, red meat, potatoes, high-fat dairy products, eggs, refined grains, high-energy drinks, sweets and desserts, mayonnaise, french fries, *-fruit*
44	factor 2 "meat, fat and oil"	meat, fat and oil, eggs, noodles
45	7 Non traditional meats	chicken, beef, souvlaki/gyros, delicatessen type meats
46	1 Meats and animal fats	pork, gyros, beef, chips, delicatessen meats, phyllo dough desserts, cakes and candies, dairy spreads and toppings, lamb
	10 Lean meats	fish (fresh/frozen), combined "Greek" dishes, chicken
47	1 Meat	processed meats, red meats, fish, poultry, eggs, fats and oils, condiments
48	5 non-processed cuts of meat and fish	poultry, fish, meat
50	Component 4	cheese, eggs, *- biscuits, - cakes*, fried fish, sausage/processed meat, fried food
51	2 Western	refined grains, processed meats, red meat, french fries, condiments, desserts and sweets, potatoes, high fat dairy, pizza, sugar beverages, margarine, mayonnaise
53;54	2 High fat	white rolls, fast food fries, potato chips, pizza, spaghetti or lasagne, fried potatoes, hot dogs, pepperoni, gravy, ground beef, other beef, doughnuts/pastries, hamburger, chicken wings, pork chops, bacon, bologna, cake, macaroni and cheese, brownies, fast food cheeseburger, ham, sausage, mashed potatoes, cold cuts, corn, candy bars
55	factor II	beef, chicken, pork, sauce (seasoning)
56;57	2 Western (= traditional Danish main meal)	white bread ; white ryebread ; jam, honey ; cakes, cookies ; candy, chocolate ; ice cream, soft drinks ; meat ; sausages ; liver paste, cold meat ; butter, lard, margarine
61;62	2 Western	processed meats, sweets, refined grains, high-fat dairy foods, meats, soda, potatoes, pea soup
63	2 Red meat, starch	beef, processed meats, white potatoes, pork/ham, salted snacks, white bread, cheese, cheese dishes, egg
64 men	3 Meat	hot dogs, eggs, bacon, sausage, hamburger, liver, smoked meat/corned beef
	7 Beef	beef/pork/lamb mixed dish, beef/pork/lamb main dish
64 women	4 Meat	hot dogs, hamburger, luncheon meats, potato chips
65 men	6 Protein foods	bacon, sausage, fried fish, eggs, steamed/grilled/baked fish, chopped lamb

Appendix table 2 continued

65 men	2 Pork and poultry	chopped pork, mixed dishes with pork, mixed dishes with bird, roast poultry, mince, oil/dressing, ham
	13 Mixed meat dishes	mixed beef, mixed lamb, steak beef
65 women	4 Meat dishes	mixed dishes with pork, chopped pork, ham, mince, oil/dressing, soy beans
67 NL men	3 Pork, processed meat, potatoes	coffee, pork, processed meat, potatoes, lowfat margarine
67 NL women	4 Pork, processed meat, potatoes	lowfat margarine, processed meat, coffee, pork, *-butter*
67 FI	3 Pork, processed meat, potatoes	potatoes, pork, spirits, wine, eggs, butter, processed meat
67 SE	2 Pork, processed meat, potatoes	nonferm whole milk, potatoes, processed meat, beef/veal, added sugar, soft drinks, pork, white bread, butter
67 IT	2 Pork, processed meat, potatoes	butter, nonferm whole milk, pasta, beef, potatoes, processed meat, cakes, eggs
68	1 American diet	irish potatoes, spaghetti/pizza, red meats, added fat high in saturated fat, low-fiber grain products
	4 Meats	liver, chicken, mixed dishes, seafood
69	4 Western type diet	Meat, eggs, potatoes, margarine, vegetables, nonalcoholic beverages, vegetable seed oil, fruit, *-cereal, -ethanol*
70	2 Western	Minced meat, cake or candy, mayonnaise, fatty foods, juice
71	1 Western	processed meat, eggs, red meats, high-fat dairy products, potatoes, refined grains
	3 Californian	fish and other seafood, poultry, organ meat, fruit, cruciferous vegetables, dark-yellow vegetables, soup, green leafy vegetables
73	3 meat	pork, beef, chicken, ham/sausage, green leafy vegetables, *- tomatoes, - other fruit*

- negative factor loadings are represented by a minus sign preceding the food item, and the food item in *italic* font

Appendix Table 3: Comparison of dietary patterns characterized by consumption of sweet foods, desserts and/or baked goods, identified by factor analysis in published studies up to January 1st, 2004

Ref	Labels	food items
14	3 Dessert	chocolate, ice cream
15-17	2 Sweets	sugar and sweeteners, tea and coffee, *-soybean products*
18	3 Culinary complexity	butter, jam, citrus fruit, fruits juice, pudding/ice cream, potatoes, grapes/bananas, bread, sugar, vegetables, fruit, tea, chocolate, cookies, *-eggs, -rice*
19	V Desserts	cake; salad dressings; ice cream, ice milk, sherbert, yogurt
	XI Cereals, pancakes, sweets	hot cereals (oatmeal, grits, farina); pancakes, waffles; flavored gelatin, cobbler, pie, pudding
	XV Sugary foods	honey, molasses, syrup, jam, jelly, cocoa
22;23	Factor 4	ice cream, donuts/pastry
25 men	6 Dessert	cookies, candy, cake, ice cream, donut and pastries, pie, noodles, *- broccoli, - asparagus*
25 women	5 Dessert	cake, candy, pie, cookies, ice cream, donut and pastries, muffin and biscuits, *- broccoli, - cauliflower*
26 27;28men	5 Snacks	cookies, candy, crackers, donuts or pastries, hamburger, ice cream, baked beans
26-28 women	5 High fat	pies, candy, cake, potatoes, roast beef, donut or pastries, bacon, rolls, potato chips *- poultry,- yoghurt*
29 teenagers	1	milk products, low-fat milk, sugar and sugar products, dried and dehydrated fruits, flour based desserts, fruit based desserts
29 adults	1	sugar and sugar products, milk products, fats and oils non-red meat
30	4 Sweets	chocolate, candy w/chocolate, cookies
	6 Baked goods	homemade pie, ready-made pie, cake
33	Component 4 ("sweet food items in preference to all types of vegetables")	confectionery, biscuits, cake, *-green vegetables, - root vegetables, - pulses, - other vegetables, - potatoes (excl chips)*
38 men	3 High fat / sugar dairy	yogurt, high fat dairy foods, high sugar dairy foods
38 women	3 High fat / sugar dairy	yogurt, high fat dairy foods, high sugar dairy foods
45	3 Sweets	fruit preserves, phyllo dough desserts, sugar, ice cream, pasta
48	3 Baking and puddings	cakes, biscuits, puddings, sweets
50	Component 3	chocolate, sweets,*- green vegetables, - root vegetables, - pulses, - potatoes*
	Component 2	root vegetables, cheese, pulses, chocolate, sweets, puddings/pies, biscuits, cakes, potatoes
60 men	Sweets	dessert, cake and cookies, confectionery and ice cream, jam, honey and chocolate spread, canned fruit, *- beer*

Appendix table 3 continued

60 women	Sweets	dessert, Cake and cookies, confectionery and ice cream, jam, honey and chocolate spread, canned fruit
64 men	2 Dessert	chocolate, cookies, ice cream, doughnuts, potato chips, hamburger, nuts, peanut butter
	8 Pies, cake	cake, pies, other fresh juice, cooked cereals, cookies, - *beer, - smoked meat/corned beef*
64 women	2 Dessert	doughnuts, chocolate, cake, liver, ice cream, cookies, french fries, pies, potato chips
65 men	8 Cakes and sweets	cake/muffins, pudding, pie/pasties (sweet), chocolate, ice-cream
65 women	11 Cakes and desserts	pudding, cake/muffins, pie/pastie (sweet), ice-cream
	12 Chocolate and confectionery	chocolate, confectionery, chocolate biscuits
67 NL men	4 Sweet and/or savoury snacks	cakes/cookies, sweet sandw spread, sweets/ candies, berries
67 NL women	5 Sweet and/or savoury snacks	savoury snacks, nuts, sweets/candies, cakes/cookies
69	3 Sweet-based	sugar, butter, other dairy, nonalcoholic beverages, *-ethanol, -vegetable seed oils, meat*

- negative factor loadings are represented by a minus sign preceding the food item, and the food item in *italic* font

Appendix Table 4: Comparison of dietary patterns characterized by consumption of (salty) snacks, fast foods, fried foods, identified by factor analysis in published studies up to January 1st, 2004

Ref	Label	food items
19	II Snacks	candy bars; hard candy; regular gum; potato chips, corn chips, pretzels; cookies
	VIII Salty tidbits	popcorn, caramel corn; vegetable oil; peanut butter; crackers; pickles, olives, relishes
20	Component 3 (Convenience diet)	beers, chips, sauces, soft drinks, nuts, cheese, rice and pasta, miscellaneous, savoury pies
22;23	Factor 5	potato chips, fried food, meat, nuts/peanuts, tomato sauce, donuts/pastry, mayonnaise, cabbage
25m	7 Fast foods	french fries, rolls, hamburger, coffee, white bread, cola, hotdogs, - *whole-wheat bread*
31	1 high fat, low nutrient density foods	french fries, fast foods, potato chips, soft drinks, dressings, smoked meats, nuts, baked goods
32	4 Nutrient dilute	salty snacks, non-dairy, sugar
33	Component 3 ("High fat foods, pasta and/r rice and coffee")	chips, crisps, pasta/rice, - *preserves*, processed meat, fried food, soft drinks, - *tea*, coffee
35	2 Junk food	donuts cookies cakes pastries, chocolate or candy, ice cream, french fries or poutine, Snacks such as chips fritoes doritos, dessert with cream or whipped cream, hot dogs salami bologna and other processed meats, spaghetti or noodles with meat butter or cheese sauce, bacon or sausages
36	1 High energy density	breads and cereals other than white and whole grain [rolls, crackers, etc.], high fat dairy products, eggs, salty snacks including fried and roasted potatoes, soft drinks, sauces/salad dressing etc
37;37	Junk foods	chips or french fries, chocolate/candy, cookies/cake, pop, klik (canned meat), canned fruit (in heavy syrup)
46	7 Snacks	french fries, carbonated beverages, beer, eggs, molluscs (octopus/squid), delicatessen type meats, sugar
48	2 Snacks and processed foods	lrn Bru, soft drinks, crisps, chips, sausages, pies
49 men	High Calorie	processed meats, salty snacks, chocolate, white bread, pastries, beef, cheese dishes, soda, fruit drinks, potato, cheese, ice cream, dressing, pizza
49 women	High Calorie	processed meats, salty snacks, chocolate, white bread, pork and ham, pastries, beef, cheese dishes, soda, flour tortillas
65 men	7 Takeaways	pizza, hamburger, meat pie, chips, soft drink
65 women	6 Takeaways	hamburger, soft drink, hot chips, meat pie, chips
71	6 Convenience	pizza, snacks, high-energy drinks, *-coffee, -whole grains*
74	1 "Western"	fast food, french fries, high-fat dairy products, processed meals, red meats, sauces, meat products, processed pastries, eggs, *-low fat dairy products, - wholeweat bread*
76	3 Refined foods pattern	french fries, high sugar beverages, mayonnaise, salty snacks, white bread, low sugar beverages, candy, pizza, warm sauces, - *whole grain bread, - boiled vegetables, - cheese, - fruit, - potatoes, - dairy (>0.02 g fat/g), - coffee, - added fat (<0.35 SFA/g fat), - legumes, - wine, - dairy (<0.02 g fat/g)*

- negative factor loadings are represented by a minus sign preceding the food item, and the food item in *italic* font

Appendix Table 5: Comparison of dietary patterns labeled "Traditional", identified by factor analysis in published studies up to January 1st, 2004

Ref	Label	food items
11	2 Traditional food use	grains (bread, porridge), foods high in sugar, glasses of fluid milk, coffeebread servings, coffee, butter and margarine, potato servings, fresh fruit
15-17	6 Traditional Japanese foods	pickles, seasonings, fish and shellfish, *-dairy products, -tea and coffee*
20	Component 1 (Traditional)	bread, spreading fat, tea and coffee, cakes and desserts, potatoes, milk, preserves, eggs, bacon
26 27;28men	4 Traditional	wax beans, green beans, potatoes, peas, beets, cabbage, roast beef, cakes, pies, corn
32	2 Traditional	tortillas, legumes, high fat meats, sugar
36	2 Traditional	white breads, desserts and sweets, butter, margarine and oil, shortening and lard, lean meats, fresh/baked/frozen vegetables, baked/boiled/mashed potatoes
45	11 Traditional items	olives, *- soft drinks*, goat meat, toasted bread
48	4 Traditional hot mail accompaniments	potatoes, root vegetables, cooked green vegetables, peas and beans
49 men	Traditional	beans, corn tortillas, chilli peppers, organ meats, *- low-fat milk, - diet drinks, - dark bread*
49 women	Traditional	beans, corn tortillas, chilli peppers
59	2 Traditional	total protein, starch, alcohol, nitrite, NDMA
60 men	Plain cooking	cooked vegetables, cabbage family, legumes, cooked potatoes, mushrooms, sauce, poultry, meat except fish and poultry
60 women	Plain cooking	cooked vegetables, cabbage family, legumes, cooked potatoes, mushrooms, sauce, poultry, meat except fish and poultry, soup
66 men	2 Traditional	beans, rice
66 women	2 Traditional	beans, rice, butter/margarine
68	2 traditional breakfast	pork, eggs, modified added fats
75	1 traditional cooking	meat, sauce, cooked vegetables, potatoes, poultry
76	2 Traditional pattern	red meat (>0.05 g fat/g), processed meat, red meat (<0.05 g fat/g), potatoes, added fat (<0.35 g SFA/g fat), coffee, beer, eggs, french fries, organ meat, chicken, spirits, white bread, boiled vegetables, *- soy products, - dairy (<0.02 g fat/g), - breakfast cereals, - tea, - pastries, -fruit, - juice, -dairy (>0.02 g fat/g),- candy, - pizza, - nuts, - brown rice, - cheese*

- negative factor loadings are represented by a minus sign preceding the food item, and the food item in *italic* font

Appendix Table 6: Comparison of dietary patterns characterized by consumption of alcoholic beverages, identified by factor analysis in published studies up to January 1st, 2004

Ref	Label	food items
15-17	3 Alcoholic beverages	alcohol, fish and shellfish
29 teenagers	4	hard alcohol, beer/wine
38 men	4 Drinker	fish, liquor, wine
38 women	6 Drinker	liquor, beer, wine
45	1 Alcohol and appetizers	beer, distilled liquor, wine, coffee, souvlaki/gyros, delicatessen type meats, soft drinks
46	4 Alcohol	wine, ouzo, cheese pies, beer
60 men	Alcohol	fish, nuts, chips and salt sticks, fried potatoes, beer, spirits, wine, other alcoholic beverages, eggs, soup, confectionery and ice cream
60 women	Alcohol	beer, spirits, wine, other alcoholic beverages
61;62	3 Drinker	wine, liquor, beer, snacks, whole grains, poultry
67 FI	2 Alcohol	wine, spirits, beer, *-dry cereals, -added sugar, -berries, -coffee, -cakes/cookies*
67 SE	3 Alcohol	wine, spirits, beer, savoury snacks, beef/veal, *-dry cereals*
67 IT	4 Alcohol	coffee (with milk), nonferm lowfat milk, cakes/cookies, other fruits, *-wine, -spirits*

- negative factor loadings are represented by a minus sign preceding the food item, and the food item in *italic* font

Appendix Table 7: Labels of other[1] dietary patterns, identified by factor analysis in published studies up to January 1st, 2004

Ref	labels
12;13	I More dairy, soups, less sugary; II More non-sugary beverages, less dairy; III more eggs, legumes, cereals; IV more meats, vegetables and fruis, desserts; V More poultry, less meats; VI More mixed protein, shellfish; VII More fish, fats
14	4 Breakfast; 5 High-protein low-fat meats; 6 Carbohydrates; 7 Spreads
15-17	1 Stample foods; 7 Vegetable oil
18	1 Satiating capacity
19	IX Milk, cereals, breads; XII Sandwich dressings, soft drinks; XIII Diets drinks, diet gum; XIV Coffee, cream, sugar; XVI Fruit, milk; XVII Low-fat milk
20	4 Meat and two veg diet
22;23	Factor6 (margarine, skim milk); Factor9 (whole milk)
24	1 Avoid fat as seasoning; 2 avoid meat; 3 modification; 4 substitution; 5 replacement
25 men	8 Southern European
25 women	4 Low cost; 7 Costly; 9 Nonuse
26-28 men	7 Whole grain
26-28 women	3 Low cost; 6 Light; 7 Whole grain
29 teenagers	3 (refined grain products, fats and oils, whole grain products)
29 adults	3 (organ meats, citrus fruits and juice, non-dairy milkproducts); 4 (low-calorie beverages, low-fat milk)
30	7 Soft dairy; 8 Beans, sweet potatoes and bran; 11 High fat breakfasts; 12 Chicken; 14 Low fat breakfasts; 17 Margarine minus butter; 9; 10; 13; 15; 16; 18; 19; 20
31	3 High fat profile
32	3 Transitional; 5 Protein rich; 6 High fat dairy; 7 Mixed dishes
33	Component 2 ("Starchy foods, carcass meat and vegetables")
35	3 High fat traditional food choices; 4 low fat substitutes; 5 Reduce fat content of meat
37;37	3 Bush foods; 4 Breakfast foods; 5Hot meal foods; 6 Tea foods; 7 Bread and butter
38 men	5 Substituter; 6 Fruit juice
38 women	4 Substituter; 5 Coffee and roll
44	3 Bread and rice negative; 4 Salty pickled vegetables, fish and shellfish
45	2 Added oils and fats; 4 Fish and soups; 5 Dairy products; 6 Villager's menu; 9 Seafood /Laborious food; 10 Honey
46	2 Festival-expensive/non-traditional items; 3 Bread with various fats; 5 Tea drinking habit; 8 Added fats and oils; 9 Carbohydrate-rich foods; 11 Dairy products; 12 Coffee drinking habit; 13 Nuts
47	Bean; Cold foods
59	3 Refined; 4 Fat-rich
60 men	Cereals (+meat); High (low-) -fat dairy; Bread and sausage
60 women	Cereals (+meat); High (low-) -fat dairy; Bread and sausage
64 men	4 Rice, tofu; 9 No fast food; 10 Miscellaneous; 12 Butter; 13 Bottled water
64 women	3 Miscellaneous +; 5 Miscellaneous -; 6 High protein; 7 Beverage; 8 Unhealthy
65 men	9 Offal and canned fish; 11 Breakfast cereals; 12 Crackers, cheese and savory spreads; 14 Wholemeal bread; 15 Coffee and milk
65 women	5 Pasta, rice and other mixed dishes; 7 Ethnic vegetables; 8 Breakfast cereal/muesli; 10 Fish and seafood; 13 Crackers, cheese and savory spreads; 14 Coffee and milk; 15 Low-energy drinks
66 men	3 Western
66 women	1Mixed; 3 Western
67 NL men	5 Brown/white bread substitution
67 NL women	2 Fat dairy, 3 Brown/white bread substitution
67 SE	4 Margarine/butter substitution
68	7 Snack foods
70	3 Salty
71	4 Breakfast
73	2 Western breakfast; 4 rice/snack

[1] Other dietary patterns: not characterized by consumption of vegetables and/or fruits, meat and fat foods, sweet foods, desserts and/or baked goods, salty snacks, fast foods or fried foods, alcoholic beverages or labeled "Traditional"

3

Common and country-specific dietary patterns in four European cohort studies

Helena F Balder, Mikko Virtanen, Henny AM Brants, Vittorio Krogh, L Beth Dixon, Frans Tan, Satu Mannisto, Rino Bellocco, Pirjo Pietinen, Alicja Wolk, Franco Berrino, Piet A Van den Brandt, Anne M Hartman, R Alexandra Goldbohm

Journal of Nutrition 2003; 133: 4246-4251

ABSTRACT

The association between diet and cancer, predominantly investigated univariately, has often been inconsistent, possibly because of the large number of candidate risk factors and their high intercorrelations. Analysis of dietary patterns is expected to give more insight than analysis of single nutrients or foods. This study aimed to develop and apply a common methodological approach to determine dietary patterns in four cohort studies originating from Finland, the Netherlands, Sweden and Italy. Food items on each of the food frequency questionnaires were aggregated into 51 food groups, defined on the basis of their position in the diet pattern and possible relevance to cancer etiology. Exploratory factor analysis was used to analyze dietary patterns. Using a standardized approach, three to five stable dietary patterns were identified, explaining 20% to 29% of total variance in consumption of the food groups. Two dietary patterns, which explained most of the variance, were consistent across the studies. The first pattern was characterized by high consumption of (salad) vegetables, the second by high consumption of pork, processed meat and potatoes. In addition, a few specifically national food patterns were identified. Sensitivity analyses showed that the identified patterns were robust for number of factors extracted, distribution of input variables, and energy-adjustment. Our findings suggest that some important eating patterns are shared by the four populations under study, whereas other eating patterns are population-specific.

INTRODUCTION

Epidemiological and other studies suggest that diet is an important factor in the etiology of different types of cancer. The association with various foods and nutrients, predominantly investigated univariately, has often been inconsistent, however. Apart from the obvious differences in sampled populations, possible explanations for these inconsistent findings could be (a) statistically significant associations by chance due to the large number of candidate foods, (b) inseparable effects due to high intercorrelation among many dietary components, or (c) residual confounding or interaction not specified in the model.[1-6]
Analysis of dietary patterns is a relatively new approach that allows studying the impact of many foods and their combinations simultaneously. Several publications suggest that analysis of dietary patterns gives more insight into the relation between diet and cancer than analysis of single nutrients or foods and requires further investigation.[7-14] One of the methods that can be used to study dietary patterns is factor analysis. Factor analysis is a multivariate, modeling technique to examine underlying patterns in a number of observed variables. The aim can be exploration or confirmation of the structure of associations in a set of variables, or data reduction (i.e., to describe the data with fewer variables). Factor analytic models can result in factors that are uncorrelated, unlike the original variables, or can allow for correlation among factors.
The authors are engaged in a multi-center project including four ongoing European cohort studies on diet and cancer (DIETSCAN project). The purpose of the overall project is to investigate whether dietary pattern analysis reveals patterns that are consistently associated with specific cancers across the four populations differing in dietary habits. Such a finding would strongly support the hypothesis that specific food patterns rather than individual foods contribute to the risk of some cancers.

Table 1. Characteristics of the four European cohort studies (1985-1992) participating in the DIETSCAN project

	ATBC [1]	NLCS [1]	SMC [1]	ORDET [1]
Country	Finland	the Netherlands	Sweden	Italy
Baseline year	1985-1988	1986	1987-1990	1987-1992
Baseline cohort size	29 133	120 852 [2]	66 651	10 788
Sex	men	men and women	women	women
Age range	50-69	55-69	40-74	35-69
FFQ				
total number of items	276	150	67	107
FFQ reference period	12 mo	12 mo	6 mo	12 mo
Frequency	units/times per d/wk/mo	7 categories (never – 6-7x/wk)	8 categories (never/seldom - > 4x/d)	times per wk/mo
Quantification	portion size picture booklet (3-5 per item)	natural or household units (fixed weight per unit)	age-specific standard portion sizes	portion size pictures (less/equal/more than 1 to 3 pictures)

[1] ATBC = Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS = Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC = Swedish Mammography Cohort (Sweden); ORDET = ORmoni e Dieta nella Eziologia dei Tumori (Italy). [2] 62 573 women and 58 279 men. Because of the case-cohort design, the subjects used in the analyses were derived from a random sample from the cohort.

Despite the accumulating number of nutritional epidemiologic studies that conducted factor analysis, not much research has been done to assess the influence of (mostly subjective) decisions taken during the analytic process, such as the grouping of input variables or the number of extracted variables on the resulting patterns. Assessing this influence is of particular importance for the comparison of patterns across studies to be able to attribute any differences to true differences between the populations and not to artifacts resulting from the different instruments and criteria used.
The objective of the present paper is to develop a common approach for the different data sets, to evaluate the influence of arbitrary analytical decisions on the identified patterns and to compare similarities and differences in dietary patterns between four different European populations.

SUBJECTS AND METHODS

Four ongoing European cohort studies on diet and cancer were identified that fulfilled the following criteria: (a) prospective cohort study investigating the effects of diet on the risk of cancer; (b) comprehensive diet assessment which includes items that cover the entire diet; (c) a validated dietary assessment instrument. The Alpha Tocopherol Beta Carotene Cancer Prevention Study (ATBC - Finland), the Netherlands Cohort Study on diet and cancer (NLCS), the Swedish Mammography Cohort (SMC), and the ORmoni e Dieta nella Eziologia dei Tumori (ORDET - Italy) fulfilled these criteria. All four studies were established between 1985 and 1992 with follow-up through record linkage of the cohorts with the national or local cancer registries. Characteristics of the cohort studies and the food frequency questionnaires (FFQ) are summarized in **table 1**.

The ATBC Cancer Prevention Study - Finland The Alpha Tocopherol Beta Carotene (ATBC) Cancer Prevention Study was a randomized placebo-controlled intervention study conducted among male smokers.[15,16] The cohort consisted of 29,133 white men aged 50 to 69 years at baseline that smoked five or more cigarettes per day and were living in southwestern Finland. Subjects were recruited between 1985 and 1988 and the intervention, consisting of supplementation with alpha tocopherol and beta-carotene, ended after 5-8 years in 1993. The follow-up of cancers and deaths has been continued after the intervention period. Prior to randomization, a questionnaire was administered on food consumption and other risk factors for cancer. 27111 participants satisfactorily completed a self-administered 276-item FFQ. Subjects were asked to report frequency of consumption over the past 12 months and to estimate portion sizes using a color picture booklet. This dietary instrument which has been evaluated for reproducibility and validity[17] is used in the analysis.

The Netherlands Cohort Study – the Netherlands The Netherlands Cohort Study on diet and cancer (NLCS) is a prospective cohort study, which began in September 1986.[18] The cohort included 62,573 women and 58,279 men aged 55 to 69 years at the beginning of the study. The study population originated from 204 municipalities with computerized population registries. At baseline, the cohort members completed a mailed, self-administered questionnaire on dietary habits and other risk factors for cancer. The nutrition part of the questionnaire was a semi-quantitative FFQ that asked about the habitual consumption of 150 food items in the previous year. The FFQ was validated by 9-day

dietary records in a sub-sample of the cohort.[19] Because of the case-cohort design, a random subcohort of 3500 subjects (1688 men, 1812 women) was sampled from the cohort after the baseline exposure measurement. Because of missing or inconsistent dietary data, the analyses are based on data of 3123 subcohort members (1525 men, 1598 women).

The Swedish Mammography Study – Sweden The Swedish Mammography Cohort originated from population-based mammography screening and started in 1987 – 1990 in the counties of Uppsala and Västmanland in Sweden. A questionnaire, together with a mailed invitation to be screened by mammography was sent to all women aged 40 to 76 years, and was returned by 66,651 (73,8%) women in the source population. The self-administered FFQ asked how often, on average, during the past six months, participants consumed 67 commonly eaten foods. For calculation of daily consumption, age-specific standardized portion sizes were used, based on mean values from 5922 days of weighed food records among 213 women. In a sub-sample of the cohort (n=129), the food frequency data were validated. The design and methods are described in more detail elsewhere.[20]

The ORDET Study – Italy The ORDET Study (Ormoni e Dieta nella Eziologia dei Tumori) is a prospective cohort study among Italian women on hormonal factors and diet in the etiology of breast cancer. The study population was recruited in 1987-1992 and included 10,788 healthy volunteers. All participants were female residents of the Province of Varese in Northern Italy, aged 35-69 years. A life style questionnaire included a 107-item semi-quantitative FFQ (26 food items quantified by photos, 17 items on fat added to dishes). The validity and reproducibility of this questionnaire was described elsewhere.[21,22] The analyses are based on 9208 participants with complete dietary data.

Table 2. Characteristics of the subjects with complete dietary data, included in the dietary pattern analysis in the five datasets

	ATBC [1]	NLCS [1]	NLCS [1]	SMC [1]	ORDET [1]
Sex	males	males	females	females	females
N	27 111	1525 [2]	1598 [2]	61 469	9208
Age (mean, y)	57.7 (5.1)	61.4 (4.2)	61.4 (4.3)	53.7 (9.7)	48.6 (8.6)
BMI (mean, kg/m^2)	26.3 (3.8)	25.0 (2.6)	25.1 (3.5)	24.8 (4.4)	25.4 (4.3)
Energy intake (mean, KJ)	11192 (3138)	9022 (2147)	7055 (1715)	5563 (1573)	7430 (2170)
excl alcohol	10670 (3091)	8597 (2133)	6892 (1699)	5473 (1567)	7137 (2121)
Smoking (%)					
Non	0	13	59	54 [3]	64
Ex	0	51	20	27	16
Current	100	36	21	19	20

[1] ATBC = Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS = Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC = Swedish Mammography Cohort (Sweden); ORDET = ORmoni e Dieta nella Eziologia dei Tumori (Italy). [2] Because of the case-cohort design, the subjects used in the analyses were derived from a random sample from the cohort. [3] Smoking data for 1997 (not available for baseline 1987-90)

Human subjects ATBC was approved by the institutional review boards of the National Public Health Institute, Finland and the US National Cancer Institute. NLCS was approved by the institutional review boards of TNO Toxicology and Nutrition Institute and the

University of Maastricht, the Netherlands. ORDET was approved by the ethical review board of the Italian National Cancer Institute, Italy. SMC was approved by the ethics committee at Uppsala University Hospital and by the regional ethics committee of the Karolinska Institute, Sweden.

Standardization of dietary data and food grouping The dietary data in the four studies were derived from FFQs with different degrees of detail and different numbers of items. Because the number and detail of food variables used as input in the factor analysis is likely to influence the resulting patterns, a common food grouping was developed. Food items of the four food frequency questionnaires were aggregated into 51 food groups (**appendix table 1**), which were defined on the basis of their role in the diet and their possible relevance to cancer etiology. In addition to the standardized food groups, some country specific foods were included to prevent loss of possibly relevant details. These food groups were an important part of the corresponding cohort's diet while relatively unimportant in the other cohorts; thus not so much a case of missing data but rare consumption among the cohorts not including these items.

Data analyses All statistical analyses were performed separately for the four cohort studies and for men and women. Since only the NLCS included both men and women, this resulted in five study data sets. Subjects with complete dietary data were used in the analysis. All data sets had enough observations according to the criteria published by Hatcher,[23] who recommended at least five subjects per input variable in factor analysis.
To identify dietary patterns, exploratory factor analysis (EFA) was performed on the correlation matrix of the 51 food groups. The analyses were performed using the SAS System (SAS Institute Inc., Cary, NC, USA), with exception of the Italian data which were analyzed using Stata Software (Stata Corporation., College Station, TX, USA). The results of both statistical packages were compared for the ORDET and NLCS data sets and the factor solutions yielded by SAS (Proc Factor procedure) and Stata (factor pcf) were very similar.

To examine the influence of analytic decisions on the stability of the dietary patterns, sensitivity analyses were carried out with the number of factors extracted, dichotomization of extremely skewed variables and energy adjustment of the food group variables by the residual method,[24] as explained below. One of the most important decisions in factor analysis is the choice of the number of factors to be extracted. A factor is usually extracted by default if its Eigenvalue (a measure of the amount of variance that is accounted for) is greater than one. A useful additional criterion is the point where the screeplot (plot of the Eigenvalues against the number of factors) levels off. However, this plot does not always show a clear break and may be subject to sampling variation. To assess the impact of extracting an additional factor on the content and interpretation of the previous factors, an increasing number (i.e. 2-6) of factors were extracted and compared for each data set.

Dietary variables often have a large number of one value (zero) and because this affects the correlation between the variables it might lead to spurious factors. Therefore, it was decided to dichotomize variables with more than 75 percent of non-users (non-users=0 *vs.* users=1) because more contrast was observed in whether or not these food groups were consumed than in the quantity in which they were consumed. Other transformations to enhance normality and linearity (such as log or square root) would hamper interpretation of

the factor scores calculated for dietary patterns. Furthermore, as extensive data cleaning in all data sets had been carried out, potential outliers were considered not to be the result of incorrect data and were not dropped.

Our primary interest was in dietary patterns based on the relative composition of the diet, and not in those based on total amount consumed. Therefore we performed pattern analysis with the food-group variables both unadjusted and adjusted for energy by the residual method of Willett and Stampfer,[25] to assess the influence of energy adjustment.

Stability was assessed by comparing the factor solutions between two random halves of each of the five data sets and by comparing the factor solutions across the sensitivity analyses. A factor was considered stable if the food groups significantly contributing were similar and their factor loadings were comparable in both direction (positive/negative) and magnitude.

FIGURE 1. Screeplot of the eigenvalues of the first twenty factors identified by exploratory factor analysis, for the four European cohort studies (1985-1992).

The final factor solution was determined for each study separately on the basis of stability and interpretability of the patterns resulting from the factor solutions. The resulting number of extracted factors may therefore be different for each data set.
Summarizing, the factor loading matrices of (1) untransformed and dichotomized variables, (2) unadjusted variables and variables adjusted for energy intake, (3) extracting an increasing number of factors, and (4) two random halves of the data set, were compared both visually and with procrustes rotation, within each study. With Procrustes rotation, two different factor solutions are compared by rotating the loadings matrix of the other so that it is as similar to the other as possible, while retaining the orthogonality (independence). If comparing, for example, a 3 factor solution to a 4 factor solution, the smaller loading matrix was appended with zeroes and then rotated. The resulting transformation matrix can be regarded as correlation matrix between the factors from different solutions.[26)]
After orthogonal varimax rotation of the factors, food groups with absolute factor loadings >0.35 were considered in the interpretation of the factors. The larger the factor loading of a given food group, the greater the correlation of that food group to that dietary pattern. A negative factor loading indicates that food-groups were inversely associated with the pattern.

RESULTS

Characteristics of the cohort data sets are presented in **table 2**. A listing of the 51 food groups used in the pattern analyses and their mean consumption (grams/day) can be found in the appendix (appendix table 1). Foods not included in the study-specific FFQs as well as country-specific foods included in study-specific FFQs only (e.g. allium vegetables for the Netherlands, other fruits for Italy, pizza for Italy and Finland, and light beer for Sweden) resulted in several missing entries in the appendix table.
Overall, men showed a much higher consumption of potatoes, bread, beer and spirits than women. The Finnish men had a high consumption of dairy products, processed meat, fish, eggs, butter and beer. Among both men and women in the Netherlands, consumption of legumes, cabbages and tea was higher than in the other countries. Among the women in Sweden, the consumption of dry cereals was much higher than in the other countries. In Italy, women's consumption of raw leaf vegetables, tomatoes, pasta, oil, full cream cheese, beef and veal, and wine was relatively high. Italy also showed a high consumption of white bread, whereas in the other countries mostly brown bread types were consumed.

Model selection For all datasets, the scree plot of the Eigenvalues of the first 20 factors was generated by the exploratory factor analysis (**figure 1**). On basis of the scree plots 2–6 dietary patterns could be identified, explaining up to 29 percent of total variance. Because the scree plots did not indicate a single clear break which could be used as an objective criterion to choose the number of factors for any of the studies, for each data set the 2-, 3-, 4-, 5-, and 6-factor solutions were extracted and compared (**table 3**). In general, increasing the number of extracted factors did not materially change the previously extracted factors. For example, comparing the 4-factor solution with the 5-factor solutions of NLCS men, the procrustes rotation indicated that the first factor ("Salad vegetables, oil, pasta") in the 4-factor solution and the first extracted factor in the 5-factor solution showed a correlation of 0.97. Hence this factor was interpreted to be the same. Similar factor 2 ("Cooked vegetables") and factor 4 ("Pork, processed meat, coffee") were interpreted to

remain the same, based on correlations of 0.84 and 0.96, respectively. Occasionally, if extracting an additional factor some food groups shifted between the factors. For example most variation of factor 3 ("Sweet foods") came from the old factor 3, but also from 2 and 4 with respective correlations of 0.42 and -0.22. Some of the variation from factor 5 ("Brown/white bread substitution") came from factor 3 and 2, but a lot was newly accounted for (table 3).

Table 3. Extraction order and similarity of dietary patterns (determined by Procrustes rotation) identified in exploratory factor analysis, when extracting two to six factors and using different sensitivity analyses. NLCS (1986), 1525 men

Factor label	Extracted number of factors					Transformations		
	2	3	4	5	6	UU[1]	DU[1,2]	UE[1]
Salad vegetables, oil, pasta	1	1	1	1	1	1	1	1
Cooked vegetables	2	2/3	2	2	2	2	2	2
Pork, processed meat, potatoes		2/3	3/4	4	3	3	3	3
Sweet foods			3/4	3	4	4	4	3
Brown/white bread substitution				5	5	5	5	4
Fat dairy					6	6	6	5
Low/full fat margarine substitution								6

[1] UU: untransformed-unadjusted; DU dichotomized-unadjusted; UE untransformed-energy adjusted.
[2] Dichotomized variables: fermented whole milk and milk products, low fat cheese and cheese spreads, low fat margarine.

Dichotomizing variables with a high percentage (>75 percent) on non-users did not affect the food groups with significant factor loadings, the magnitude of the factor loadings or the explained variance and thus the order of the extracted patterns (table 3). This resulted in correlations of 0.98-1.00 on the diagonal of the procrustes rotation matrix, and low (<0.10) mutual correlations between the factors.

When using the energy-adjusted food groups in the pattern analysis, the factor solutions were mostly comparable with the unadjusted factor solutions (table 3). As could be expected, mainly the factors with high loadings on energy contributing food groups changed. By using energy-adjusted food variables, substitution of foods such as brown versus white bread and low fat versus medium and full fat dairy products became more important. The energy adjustment forces substitution patterns as it requires each food group to be uncorrelated with energy. Except for some minor changes the other factors were unaffected by energy adjustment. For example, for NLCS males, the factors 1, 2, 4, and 5 showed correlations of 0.95 to 0.98 in the procrustes rotation matrix. Based on these results it was decided to use the variables unadjusted for energy for the final factor solution. The analyses based on split samples were very similar, only the ordering of the factors changed to some extent (results not shown). Based on the results of the sensitivity analyses, the final number of extracted factors was determined to be 3 for the ATBC study, 5 for both males and females of the NLCS study, 4 for SMC, and 4 for the ORDET study.

Interpretation and comparison The dietary pattern labels and percentage variance of the final factor solutions in the five data sets are presented in **table 4**. The factor loadings (>0.35) of the predefined food groups of the identified dietary patterns can be found in the appendix (**appendix table 2**).

Table 4 Percentage variance explained (%) by the varimax rotated dietary patterns, identified in the exploratory factor analysis in the four European cohort studies (1985-1992).

Pattern	ATBC [1]	NLCS [1]	NLCS [1]	SMC [1]	ORDET [1]
(Salad) vegetables	9.6	5.6	6.3	6.9	11.6
Pork, processed meat, potatoes	5.3	4.2	4.3	5.4	7.4
Cooked vegetables		4.8			4.8
Alcohol	5.4			5.3	4.7
Sweet and/or savory snacks		4.3	3.9		
Brown/white bread substitution		4.1	4.3		
Other			4.4[2]	4.2[3]	

[1] ATBC = Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS = Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC = Swedish Mammography Cohort (Sweden); ORDET = ORmoni e Dieta nella Eziologia dei Tumori (Italy). [2] Pattern labeled "Fat dairy". [3] Pattern labeled "Margarine/butter substitution".

Two factors were qualitatively similar across studies and between men and women, although in some instances specific food groups did not load equally and the amount of variance explained varied. The first comparable factor had high factor loadings on raw leaf vegetables and tomatoes and food groups such as carrots and cabbages and could therefore be interpreted as a "(Salad) Vegetable" pattern. This first pattern also included oil, poultry, rice, pasta and fish, although the factor loadings of these food groups were not consistently >0.35 in all data sets. For NLCS males and ORDET, a separate "Cooked vegetables" pattern was also identified, with high loadings on cooked leaf vegetables, cabbages, legumes and carrots. The second similar, although somewhat less consistent, factor had high loadings on pork, processed meat and potatoes and was hence labeled as such. Other food groups with high loadings in this pattern were eggs, butter and coffee.
In addition to these two comparable factors other study-specific patterns emerged. For ATBC, SMC and ORDET an "Alcohol" pattern (high loadings on wine, beer and spirits) was identified. In the NLCS, both in males and females a "White/brown bread substitution" pattern was identified. In NLCS females a "Sweet and/or savory snacks" pattern (savory snacks, nuts, sweets/candies and cakes/cookies) was identified. In NLCS males this pattern consisted of cakes and cookies, sweet sandwich spread, sweets and candies. In NLCS females also a "Fat dairy" pattern (potatoes, non-fermented whole milk, margarine, sweet sandwich spread) was observed and in SMC a "Margarine/butter substitution" pattern (high positive loading for margarine, negative loading for butter).
The total percentage variance explained by the extracted factors reflects the number of input variables in the five data sets. When using a smaller number of food groups (e.g. ORDET had 32 food groups vs. NLCS had 49), there is potentially less unique variance than among a larger number of food groups, so that more variance is explained by a similar number of extracted factors.

DISCUSSION

Using a standardized approach, 3 to 5 stable dietary patterns were identified in cohort studies from Finland, the Netherlands, Sweden, and Italy. Two of the identified patterns, i.e. a "(Salad) Vegetable" and a "Pork, processed meat and potatoes" pattern, were relatively consistent across the studies. In addition to these common dietary patterns, some study-specific food patterns were identified. The identified dietary patterns explained 20% to 29% of the total variance in consumption of the foodgroups. In contrast to e.g., psychometric applications of factor analysis, the items of food consumption questionnaires are not specifically constructed to be highly correlated in order to characterize underlying traits. Other studies that applied factor analysis on dietary data usually find a comparable percentage of total variance explained. Thus, 30% is reasonable for only a few factors. Even more important is the finding that the same or similar factors among the 4 cohorts explained this variance. The importance of these findings will be illustrated through examination of relationships between these factors and sociodemographic factors and specific cancer sites.

The dietary patterns that emerged were easily interpretable and the patterns labeled as "(Salad) Vegetable" and "Pork, processed meat and potatoes" showed a consistent list of the same food groups across the populations. However, the labels assigned are still somewhat arbitrary and represent our interpretation of the data; others may label these dietary patterns differently. Comparing models with a different number of extracted factors proved to be very useful in gaining insight into the data. Studying the dynamics of models helped in restricting to a smaller number, meaningful factors.
A potential objection against the use of factor analysis to define dietary patterns is the influence of subjective decisions on the identified eating patterns.[27] Differences between the FFQs and differences in the detail and number of input variables were minimized by aggregating the food items into a predefined, common food grouping. To make optimal use of the available dietary data in each study, not all studies performed the EFA on the same number of variables. McCann et al[28] described that three methods of reducing detailed dietary data obtained from a single FFQ prior to pattern analysis did not affect the number or the character of the patterns identified. In our study, the differences between the FFQs were reduced by allocating the original FFQ items to the same food groups and we showed that the small differences in number of variables and inclusion of country-specific foods did not hamper the comparability of the results.

Furthermore, we observed that several analytic decisions, such as dichotomization and energy-adjustment of the input variables and increasing the number of extracted factors did not materially change the interpretation of the previously extracted factors. Therefore, methodological differences are not likely to hamper comparison with dietary patterns identified in other studies to a large extent.
As in our analysis, other studies have observed a vegetable-rich dietary pattern (labeled as "Fruits and vegetables",[29,30] "Salad",[4] or "Prudent",[31] and a pattern with high consumption of pork, processed meat and potatoes (labeled as "Western"[31] or "Traditional".[32]
The other study-specific patterns identified in the present study were also similar to patterns found in previous studies. For example a pattern consisting mainly of "Cooked vegetables" was also observed by Hebert[33] in a US population and by Maskarinec[34] in

Hawaiian women. A "Snack" pattern, consisting mainly of sweet foods and/or junk food was observed in several studies,[29,30,33,35-37] and an "Alcohol" or "Drinker" pattern was also identified in several studies.[30,36,38]
In conclusion, the sensitivity analyses suggest that the dietary pattern approach is robust for energy-adjustment, distribution of input variables and number of factors extracted. Our findings suggest that some eating patterns are common to the four European populations under study, but that other eating patterns are country-specific. The finding that shared patterns exist across populations living under different social and cultural circumstances, suggests that to some extent food choices must be driven by endogenous factors (e.g. biological, psychological), whereas other patterns are more clearly the result of sociocultural circumstances and food availability.

REFERENCES

1. Farchi, G., Mariotti, S., Menotti, A., Seccareccia, F., Torsello, S., & Fidanza, F. (1989) Diet and 20-y mortality in two rural population groups of middle-aged men in Italy. Am.J.Clin.Nutr. 50: 1095-1103.

2. Randall, E., Marshall, J., Graham, S., & Brasure, J. (1989) Frequency of food use data and the multidimensionality of diet. J Am Diet Assoc 89: 1070-1075.

3. Nestle, M. (1996) Fruits and vegetables: protective or just fellow travelers? Nutr.Rev. 54: 255-257.

4. Randall, E., Marshall, J. R., Graham, S., & Brasure, J. (1990) Patterns in food use and their associations with nutrient intakes. Am.J.Clin.Nutr. 52: 739-745.

5. Wirfalt, A. K. E. & Jeffery, R. W. (1997) Using cluster analysis to examine dietary patterns: Nutrient intakes, gender, and weight status differ across food pattern clusters. J Am Diet Assoc 97: 272-279.

6. Hu, F. B. (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. Curr.Opin.Lipidol. 13: 3-9.

7. Randall, E., Marshall, J. R., Brasure, J., & Graham, S. (1992) Dietary patterns and colon cancer in western New York. Nutr.Cancer 18: 265-276.

8. Slattery, M. L. & Boucher, K. M. (1998) The senior authors' response: Factor analysis as a tool for evaluating eating patterns. Am.J.Epidemiol. 148: 20-21.

9. Palli, D., Russo, A., & Decarli, A. (2001) Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control 12: 163-172.

10. Terry, P., Suzuki, R., Hu, F. B., & Wolk, A. (2001) A prospective study of major dietary patterns and the risk of breast cancer. Cancer Epidemiol.Biomarkers Prev. 10: 1281-1285.

11. Terry, P., Hu, F. B., Hansen, H., & Wolk, A. (2001) Prospective study of major dietary patterns and colorectal cancer risk in women. Am.J.Epidemiol. 154: 1143-1149.

12. Fung, T., Hu, F. B., Fuchs, C., Giovannucci, E., Hunter, D. J., Stampfer, M. J., Colditz, G. A., & Willett, W. C. (2003) Major dietary patterns and the risk of colorectal cancer in women. Arch.Intern.Med. 163: 309-314.

13. Handa, K. & Kreiger, N. (2002) Diet patterns and the risk of renal cell carcinoma. Public Health Nutrition 5: 757-767.

14. Masaki, M., Sugimori, H., Nakamura, K., & Tadera, M. (2003) Dietary patterns and stomach cancer among middle-aged male workers in Tokyo. Asian Pac J Cancer Prev. 4: 61-66.

15. The ATBC Cancer Prevention Study Group. (1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. Ann.Epidemiol. 4: 1-10.

16. ATBC Cancer Prevention Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N.Engl.J.Med. 330: 1029-1035.

17. Pietinen, P., Hartman, A. M., Haapa, E., Rasanen, L., Haapakoski, J., Palmgren, J., Albanes, D., Virtamo, J., & Huttunen, J. K. (1988) Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am.J.Epidemiol. 128: 655-666.

18. Van den Brandt, P. A., Goldbohm, R. A., Van 't Veer, P., Volovics, A., Hermus, R. J., & Sturmans, F. (1990) A large-scale prospective cohort study on diet and cancer in The Netherlands. J Clin Epidemiol 43: 285-295.

19. Goldbohm, R. A., Van den Brandt, P. A., Brants, H. A., Van 't Veer, P., Al, M., Sturmans, F., & Hermus, R. J. (1994) Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur.J.Clin.Nutr. 48: 253-265.

20. Wolk, A., Bergstrom, R., Hunter, D., Willett, W., Ljung, H., Holmberg, L., Bergkvist, L., Bruce, A., & Adami, H. O. (1998) A prospective study of association of monounsaturated fat and other types of fat with risk of breast cancer. Arch.Intern.Med. 158: 41-45.

21. Faggiano, F., Vineis, P., Cravanzola, D., Pisani, P., Xompero, G., Riboli, E., & Kaaks, R. (1992) Validation of a method for the estimation of food portion size. Epidemiology 3: 379-382.

22. Pisani, P., Faggiano, F., Krogh, V., Palli, D., Vineis, P., & Berrino, F. (1997) Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. Int.J.Epidemiol. 26 Suppl 1: S152-S160.

23. Hatcher, L. (1994) A Step-By-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling SAS Institute, Cary, NC.

24. Willett, W. (1990) Food Frequency Methods ; Nutritional Epidemiology Oxford University Press, Oxford.

25. Willett, W. C. & Stampfer, M. J. (1986) Total energy intake: implications for epidemiologic analyses. Am.J.Epidemiol. 124: 17-27.

26. Mardia, K. V., Kent, J. T., & Bibby, M. C. (1979) In: Multivariate Analysis Academic Press, London.

27. Martinez, M. E., Marshall, J. R., & Sechrest, L. (1998) Invited commentary: Factor analysis and the search for objectivity. Am.J.Epidemiol. 148: 17-19.

28. McCann, S. E., Marshall, J. R., Brasure, J. R., Graham, S., & Freudenheim, J. L. (2001) Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr. 4: 989-997.

29. Nicklas, T. A., Webber, L. S., Thompson, B., & Berenson, G. S. (1989) A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors: the Bogalusa Heart Study. Am.J.Clin.Nutr. 49: 1320-1327.

30. Schulze, M. B., Hoffmann, K., Kroke, A., & Boeing, H. (2001) Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. Br.J.Nutr. 85: 363-373.

31. Hu, F. B., Rimm, E., Smith-warner, S. A., Feskanich, D., Stampfer, M. J., Ascherio, A., Sampson, L., & Willett, W. C. (1999) Reproducibility and validity of dietary patterns assessed with a food- frequency questionnaire. Am.J.Clin.Nutr. 69: 243-249.

32. Barker, M. E., McClean, S. I., Thompson, K. A., & Reid, N. G. (1990) Dietary behaviours and sociocultural demographics in Northern Ireland. Br.J.Nutr. 64: 319-329.

33. Hebert, J. R., Stoddard, A. M., Harris, D. R., Sorensen, G., Hunt, M. K., Morris, D. H., & Ockene, J. K. (1993) Measuring the effect of a worksite-based nutrition intervention on food consumption. Ann.Epidemiol. 3: 629-635.

34. Maskarinec, G., Novotny, R., & Tasaki, K. (2000) Dietary patterns are associated with body mass index in multiethnic women. J.Nutr. 130: 3068-3072.

35. French, S. A., Perry, C. L., Leon, G. R., & Fulkerson, J. A. (1994) Food preferences, eating patterns, and physical activity among adolescents: correlates of eating disorders symptoms. J.Adolesc.Health 15: 286-294.

36. Matalas, A. L., Franti, C. E., & Grivetti, L. E. (1999) Comparative study of diet and disease prevalence in Greek Chians - Part I rural and urban residents of Chios. Ecol Food Nutr. 38: 351-380.

37. North, K. & Emmett, P. (2000) Multivariate analysis of diet among three-year-old children and associations with socio-demographic characteristics. The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) Study Team. Eur.J.Clin.Nutr 54: 73-80.

38. Slattery, M. L., Boucher, K. M., Caan, B. J., Potter, J. D., & Ma, K. N. (1998) Eating patterns and risk of colon cancer. Am.J.Epidemiol. 148: 4-16.

Appendix Table 1 Mean daily consumption (g/d) of 51 food groups used in the exploratory factor analysis.

	ATBC		NLCS				SMC		ORDET	
	men		men		women		women		women	
Number of subjects	27111		1525		1598		61469		9208	
Number of items in EFA	46		49		49		42		32	
	mean	**SD**	**mean**	**SD**	**mean**	**SD**	**mean**	**SD**	**mean**	**SD**
Legumes	4.9	5.1	40	26	35	23	11	15	23	16
Cabbages	14	16	40	25	39	24	10	18	7.2	8.2
Leaf vegetables – cooked	0.2 [1]	1.0	22	18	22	16	8.4	16	19	24
Leaf vegetables – raw	1.4	2.0	10	9.6	10	9.0	17	16	38	34
Allium vegetables			31	27	31	25				
Carrots	14	17	12	14	13	16	24	26	23	25
Tomatoes	24	22	20	21	24	21	27	26	89	55
Mushrooms	2.9	5.1	3.4	4.2	4.0	4.4				
Citrus fruit	59	79	51	61	73	67	52	57	90	74
Other fruit									226	162
Bananas	21	36	13	25	13	27	33	42		
Apples and pears	30	42	67	75	84	82	67	67		
Berries – all types	36	36	6.9	8.3	8.1	8.6				
Potatoes and potato products	180	79	150	85	101	60	95	52	49	39
Rice	5.6	7.4	12	30	9.4	20	2.3	6.2	20	15
Pasta	8.1	12	6.9	9.1	5.3	7.6	15	14	61	34
Dry cereals	15	14	3.3	8.8	3.3	7.3	52	63		
Bread, crackers etc. –white types	19	36	37	57	22	35	21	29	102	61
Bread, crackers etc. – brown/wholemeal types	167	93	144	86	118	59	75	49		
Fermented whole milk and milk products	66	133	16 [1]	40	19	44	58	86	26	40
Fermented medium, lowfat and skimmed milk prod	92	173	69	108	100	122	39	79		
Nonfermented whole milk and milk products	218	321	118	152	79	103	43	89	53	66
Nonfermented medium, lowfat and skimmed milk prod	346	302	103	150	97	129	121	148	43	63
Full cream cheese and cheese spreads	25	28	21	19	21	19	26	20	43	37
Lowfat cheese and cheese spreads	0.6 [1]	5.3	1.7 [1]	7.5	2.0 [1]	6.8				
Beef and veal – fresh	26	21	44	29	37	26	26	18	74	52
Pork – fresh	41	21	45	33	39	31	20	17	5.4 [1]	10
Liver – fresh	3.3	4.2	2.2	4.7	1.8	4.0	4.3	5.1	3.5 [1]	6.5
Poultry – fresh	12	15	13	15	14	16	7.2	7.0	27	21
Processed meat	75	59	22	19	15	14	25	17	22	22
Fish (shellfish and crustacean included)	39	30	14	16	12	14	25	17	16	15
Eggs	53	38	17	13	15	9.9	7.4	8.2	15	11
Butter	39	30	8.9	16	7.3	12	4.4	7.9	3.9	4.3
Margarine	21	22	29	24	19	18	7.7	7.9	0.7 [1]	2.1
Lowfat margarine			8.0*	16	6.5	13				
Oil	1.5	2.9	2.3	3.1	1.9	2.6			24	15
Dressings and other similar sauces	2.2	2.8	4.4	5.5	4.6	5.7			5.3	4.3
Pizza	1.8 [1]	6.1							19	21
Peanuts and other nuts	0.5 [1]	2.7	8.9	15	5.8	11				
Savory snacks			1.4	4.0	1.0	3.9	0.7	1.3		
Cakes, sweet breads and cookies	50	56	25	19	27	19	14	16	38	33
Candies	13	17	4.9	7.4	5.6	7.7	6.3	11		
Added sugar	20	18	29	32	11	20	3.4	6.4		
Sweet sandwich spread			9.6	13	7.7	9.4	5.8	9.2		
Coffee	607	351	574	292	495	242	427	202	14 [2]	13
Tea	73	163	321	244	383	257	132	181		
Fruit juices	24	61	21	50	28	52	38	64		
Soft drinks and syrups	40	91	44	83	32	93	54	102		
Light beer	8.1 [3]	77					39	66		
Beer	175	337	89	203	6.4 [1]	42	4.8	8.7		
Wine and fortified wine	30	42	33	70	35	64	15	21	92	123
Spirits	31	43	25	40	3.8 [1]	16	1.3	2.2	1.5 [1]	4.7

Appendix Table 2 Labels, percentage variance explained and factor loadings (>0.35) of the varimax rotated dietary patterns, identified by the exploratory factor analysis in the four European cohort studies (1985-1992).

ATBC		NLCS men		NLCS women		SMC women		ORDET women	
				"Salad vegetables" pattern (SV)					
dressings	.76	oil	.58	cabbages	.59	leaf veg raw	.64	leaf veg raw	.79
leaf veg raw	.69	pasta	.54	allium	.58	carrots	.59	dressings	.72
tomatoes	.66	mushrooms	.51	legumes	.55	tomatoes	.57	tomatoes	.71
cabbages	.61	allium	.50	leaf veg raw	.51	apples, pears	.47	oil	.59
carrots	.51	tomatoes	.44	carrots	.49	cabbages	.46	carrots	.41
rice	.47	rice	.44	leaf veg cooked	.47	citrus fruit	.46		
citrus fruit	.40	wine	.40	tomatoes	.44	fish	.45		
poultry	.40	leaf veg raw	.40	oil	.40				
berries	.35			rice	.35				
				pasta	.35				
nonferm fat milk	-.38								
				"Pork, processed meat, potatoes" pattern (PPP)					
potatoes	.55	coffee	.52	lowfat margarine	.50	nonferm whole milk	.47	butter	.58
pork	.54	pork	.51	processed meat	.49	potatoes	.43	nonferm whole milk	.53
spirits	.49	processed meat	.51	coffee	.44	processed meat	.43	pasta	.49
wine	.48	potatoes	.37	pork	.43	beef, veal	.41	beef	.45
eggs	.46	lowfat margarine	.35			added sugar	.41	potatoes	.44
butter	.45			butter	-.44	soft drinks	.39	processed meat	.43
processed meat	.41					pork	.38	cakes	.43
						white bread	.37	eggs	.37
						butter	.37		
				"Cooked vegetables" pattern (CV)					
		cabbages	.73					legumes	.59
		leaf veg cooked	.70					cabbages	.54
		legumes	.58					leaf veg cooked	.50
		carrots	.47					fish	.47
		allium	.39					carrots	.45
								rice	.44
								potatoes	.43
								poultry	.38
				"Alcohol" pattern (ALC)					
wine	.73					wine	.60	coffee (with milk)	.60
spirits	.73					spirits	.51	nonferm lowfat milk	.42
beer	.40					beer	.44	cakes	.37
						savoury snacks	.44	other fruits	.35
dry cereals	-.35					beef, veal	.35		
added sugar	-.37							wine	-.55
berries	-.38					dry cereals	-.45	spirits	-.41
coffee	-.38								
cakes, cookies	-.45								

Appendix table 2 continued

ATBC	NLCS men		NLCS women		SMC women		ORDET wom
	"Sweet and/or savory snacks" pattern (SW/SN)						
	cakes, cookies	.53	savoury snacks	.69			
	sweet spread	.46	nuts	.67			
	sweets, candies	.38	sweets, candies	.53			
	berries	.35	cakes, cookies	.38			
	"Brown/white bread substitution" pattern (B/W)						
	white bread	.63	apples, pears	.40			
			brown bread	.63			
	apples,pears	35					
	brown read	-.66	white bread	-.48			
			Other patterns				
			potatoes[1]	.52	margarine[2]	.68	
			nonferm whole milk	.48	brown bread	.39	
			margarine	.46	full cream cheese	.48	
			sweet sandw spread	.46	cakes, cookies	.37	
					butter	-.36	

[1] Pattern labeled "Fat dairy". [2] Pattern labeled "Margarine/butter substitution".

4

Dietary patterns associated with sociodemographic and lifestyle characteristics

Balder HF, Van den Brandt PA, Goldbohm RA

ABSTRACT

The aim of this paper was to describe the association of sociodemographic and lifestyle variables with the adherence to dietary patterns in the Netherlands Cohort Study. A cross-sectional analysis was performed within a random sample (the subcohort) of the NLCS, for men and women separately.

Exploratory factor analysis identified five stable dietary patterns both among male and female subcohort members of the NLCS. The "Salad Vegetable" pattern was characterized by consumption of vegetables, pasta and rice, poultry and fish, and oil. Higher scores on this pattern were associated higher education and with health conscious behavior, such as physical activity, not smoking (never or having quit), and the use of dietary supplements. The dietary pattern characterized by consumption of Pork, processed meat, potatoes, coffee and low fat margarine instead of butter was associated with having a lower education and with less health conscious and less favorable characteristics such as having a higher BMI, smoking and not using dietary supplements. Dietary patterns characterized by consumption of brown instead of white bread (men and women), by sweet foods (men), or by sweet and savory snacks (women) were inversely associated with smoking behavior. In conclusion, the observed associations between these dietary patterns and sociodemographic and lifestyle factors were generally in the expected direction. This increases the confidence that the statistical solution of the dietary pattern analysis is meaningful. We consider it likely that dietary patterns are not isolated, but are part of a broader lifestyle encompassing a combination of health conscious behaviors.

INTRODUCTION

Analysis of dietary patterns has received much attention in nutritional epidemiology, lately.[1] Dietary patterns are expected to be part of broader lifestyles.[2-4] If sociodemographic and lifestyle characteristics such as education, physical activity and (not) smoking are associated with dietary patterns, this increases the confidence that the statistical solution of the pattern analysis is meaningful. Several studies successfully used factor analysis to identify interpretable patterns of food consumption and described their association with sociodemographic or lifestyle variables[5-30] The aim of this paper was to describe the association of sociodemographic and lifestyle variables with the adherence to dietary patterns in the Netherlands Cohort Study on diet and cancer (NLCS). For this purpose we performed a cross-sectional analysis within the healthy subcohort of the NLCS.

SUBJECTS AND METHODS

The Netherlands Cohort Study The *Netherlands Cohort Study on diet and cancer (NLCS)* is a prospective cohort study, which began in September 1986.[31] The cohort included 62,573 women and 58,279 men aged 55 to 69 years at the beginning of the study. The study population originated from 204 municipalities with computerized population registries. At baseline, all participants completed a mailed, self-administered questionnaire on socio-economic background, lifestyle, and dietary habits, including a 150-item semi-quantitative food frequency questionnaire (FFQ) that asked about usual intake of foods and beverages in the previous year. The FFQ was validated by 9-day dietary records in a sub-sample of the cohort.[32] Because of the case-cohort design, a random subcohort of 5000 men and women was sampled from the cohort after the baseline exposure measurement. Because of missing or inconsistent dietary data, the analyses are based on data of 2190 men, and 2251 women. NLCS was approved by the institutional review boards of TNO Toxicology and Nutrition Institute and the University of Limburg, the Netherlands.

Sociodemographic and lifestyle data The sociodemographic and lifestyle characteristics used in the analysis were derived from the self-administered questionnaire at baseline. Age (yr) was calculated from date of birth. Body mass index (BMI) was calculated by dividing self-reported weight (kg) by squared height (m^2). Assessment of baseline non-occupational physical activity was based on two questions in the questionnaire. The first question was "How many minutes do you spend on average per day walking or cycling to your work, to go shopping, or walk your dog?" The subjects could fill in the number of minutes spent per day on these activities. The second question was: "How many hours of your leisure time do you spend on average per week on the following activities: gardening/doing odd jobs, cycling/walking (other than in the first question), and sports/gymnastics?" Possible answers were never, < 1 hour, 1-2 hours or >2 hours per week. We added up the time spent on these activities to an overall measure (minutes per day). Smoking behavior was defined by smoking status (never/ex/current), total years of smoking and the number of cigarettes smoked per day. The highest attained level of education was classified as primary school, lower vocational school (e.g. technical,

Table 1: Food grouping used in the dietary pattern analysis

Food groups	Food items from the NLCS food frequency questionnaire
Legumes	pulses (e.g. white or kidney beans, lentils, marrow fats); cooked sliced beans and string beans; cooked broad beans; legumes in soup as main course
Cabbages	cooked brussels sprouts; cooked sauerkraut; cooked cauliflower; cooked cabbage (savoy, osheart, white ,green) cooked curly kale
Leaf vegetables – cooked	cooked spinach, cooked endive
Leaf vegetables – raw	raw endive, lettuce
Allium vegetables	cooked leek, onions, chinese or indonesian foods, italian food, soup as main course
Carrots	cooked carrots, carrot salad, raw carrots between meals
Tomatoes	tomatoes; tomatoes in italian food
Mushrooms	mushrooms; mushrooms in italian food
Citrus fruit	mandarins; oranges; fresh orange juice; grapefruits; fresh grapefruit juice
Bananas	bananas
Apples and pears	apples; pears
Berries – all types	strawberries
Potatoes and potato products	potatoes (boiled, fried, mashed or in a mixed dish); french fried potatoes (hot meal)
Rice	white rice; brown rice; rice in chinese or indonesian food
Pasta	pasta in chinese or indonesian food; pasta in italian food
Dry cereals	millet, buckwheat, oats and other cereals (hot meal); (dry cereal part of) porridge and cooked oatmeal; muesli; bran; wheat germs
Bread, crackers etc. –white types	white bread; bread instead of hot meal; currant bread; rusks, knackebrod, crackers
Bread, crackers etc. – brown/wholemeal types	brown bread; whole wheat bread; bread instead of hot meal; rye bread
Fermented whole milk and milk products	whole yogurt
Fermented medium, lowfat and skimmed milk prod	skimmed and semi-skimmed yogurt; dry curd; yogurt or dry curd as salad dressing; buttermilk
Nonfermented whole milk and milk products	whole milk; whole milk in coffee (whole milk, whole condensed milk, cream, whole milk powder); milk in porridge, milk in cooked oatmeal; milk in custard pudding; milk in whole chocolate drink
Nonfermented medium, lowfat and skimmed milk prod	low-fat and skimmed milk; semi-skimmed condensed milk (4% fat); low-fat and skimmed chocolate drink, milk in coffee; (semi skimmed milk in) porridge, and cooked oatmeal
Full cream cheese and cheese spreads	40+ and 48+cheese; cheese instead of meat
Lowfat cheese and cheese spreads	low-fat (20+) cheese and cheese spread
Beef and veal – fresh	beef steak, roast beef; other beef; veal; minced beef; half beef/half pork minced meat; beef in italian food
Pork – fresh	pork chops, pork steaks, rolled pork; fricassee (and other lean pork); other pork; half beef/half pork minced meat; pork in chinese or indonesian food; pork in soup as main course
Liver – fresh	fresh pork/veal/chicken liver; cooked liver on sandwiches
Poultry – fresh	chicken, turkey and other poultry
Processed meat	minced meat hot dogs, croquettes, frankfurters (hot meal); fat bacon, smoked bacon (hot meal); smoked sausage in winter (hot meal); boiled ham; rashers, bacon; smoked beef, pork loin roll; other sliced cold meats; processed meat in chinese or indonesian food; processed meat in italian food
Fish (shellfish and crustacean included)	fish (hot meal); fish with or on bread; fish between meals
Eggs	eggs
Butter	butter used for frying meat; butter used for frying/stewing fish; butter on bread
Margarine	margarine for frying meat; margarine for frying/stewing fish; cooking/frying fat for preparing meat; cooking/frying fat for preparing fish; frying fat for french fried potatoes; margarine on bread
Lowfat margarine	low-fat margarine on bread
Oil	oil in a salad dressing; oil for frying meat; oil for frying/stewing fish; oil for french fried potatoes; oil in chinese or indonesian food
Dressings and other similar sauces	mayonnaise (hot meal); salad dressing (creamy salad dressing, dressings with linoleic acid, mayonnaise, dressing without oil)
Peanuts and other nuts	peanut butter; peanuts; other nuts; mixed nuts
Savory snacks	potato chips; salty biscuits; minced meat hot dog, croquettes etc (between meals)
Cakes, sweet breads and cookies	dutch honey cake; pastry, pie or cake; cookies
Candies	liquorice; chocolate
Added sugar	sugar in coffee; sugar in tea
Sweet sandwich spread	sweet sandwich spreads
Coffee	coffee
Tea	tea
Fruit juices	orange juice, grapefruit juice (bottled or canned); other fruit juices
Soft drinks and syrups	coke; other soft drinks; rosehip syrup
Beer	beer
Wine and fortified wine	red wine; white wine; sherry, vermouth, port, campari; blackcurrant gin, sweet liquor, egg-nog
Spirits	liquor (e.g. gin, brandy, whiskey)

domestic science school), junior high school, senior high school, higher vocational school , university and other education. The use of dietary supplements was asked for in the questionnaire as: "During the past five years did you use vitamin tablets, drops or other preparations (for example tonics, vitamins, garlic pills, brewers' yeast, calcium)?"

Assessment of dietary patterns The dietary pattern assessment and related sensitivity analyses have been described in detail elsewhere.[33] In short, to identify dietary patterns for men and women separately, exploratory factor analysis using principal components analysis was conducted on intake (g/day) of 49 food groups (**Table 1**). The Eigenvalue > 1 and the Scree test were used to determine the number of factors to be extracted. Factors were rotated by orthogonal Varimax transformation. Factor loadings represent correlation coefficients between food groups and dietary patterns and food groups with absolute factor loadings > 0.35 were considered as significantly contributing to a dietary pattern. Labels of the dietary patterns were assigned according to the foods with high loadings. Prior to subsequent analysis, factor scores were determined by summing the standardized intakes from each food group weighted by the factor loadings, creating for each participant a unique score for each factor. High scores represented high intake of foods within the corresponding dietary pattern; low scores represented low intake of those foods. Because of our interest in relative quantities, independent of the total energy requirement and intake, we also adjusted the dietary pattern scores for energy intake by the residual method. These energy-adjusted pattern scores were computed as the residuals from the (univariate) regression model with total energy intake as the independent variable and the factor score as the dependent variable. The adjusted factor scores were divided into quintiles.

Statistical analysis The analyses were performed in data of men and women of the subcohort separately. To describe the association between dietary patterns and lifestyle characteristics, the mean age, BMI, energy intake, time spent on non-occupational physical activity, the number of cigarettes smoked per day and the number of years of cigarette smoking were calculated according to quintiles of dietary pattern scores. The association of categorical variables was assessed by calculating the percentage of current, former and never smoking status, higher vocational or university education and use of dietary supplements according to quintiles of dietary pattern scores.

RESULTS

Exploratory factor analysis identified five stable dietary patterns both among male and female subcohort members of the NLCS (**Table 2**).[33] Three of these patterns were observed both in men and women: the "Salad vegetables" (SV: several vegetable items, mainly the raw items in men, but comprising all the vegetable items in women; pasta and rice; poultry; fish; and oil), the "Pork, processed meat, potatoes" (PPP: pork; processed meat and potatoes; coffee and low-fat margarine and negatively on butter) and the "Brown/white bread substitution" (B/W: brown/wholemeal bread types and apples and pears and negatively with white types of bread) pattern. In addition, in men a "Sweet" pattern (SW: with positive loads on cakes and cookies; sweet sandwich spread; candies; tea and strawberries and negative loads on beer and spirits) and in women a "Sweet and savory snacks" pattern (SN: with high loads on savory snacks; nuts; candies and cakes

Table 2 Factor loading matrix

Food groups	Men					Women				
	SV	CV	PPP	SW	B/W	SV	FD	B/W	PPP	SN
Legumes	0.11	**0.58**	0.08	0.04	-0.02	**0.55**	0.23	-0.01	0.12	-0.04
Cabbages	0.13	**0.73**	0.12	-0.05	0.04	**0.59**	0.15	0.03	0.21	-0.11
Leaf vegetables – cooked	0.01	**0.70**	-0.03	0.01	-0.01	**0.48**	0.15	0.03	0.08	-0.15
Leaf vegetables – raw	**0.40**	0.13	0.00	0.28	0.14	**0.52**	-0.13	0.18	0.00	-0.03
Allium vegetables	**0.51**	**0.37**	0.01	-0.14	0.08	**0.59**	-0.13	-0.03	-0.02	-0.04
Carrots	0.20	**0.46**	-0.06	0.26	0.23	**0.50**	0.01	0.25	-0.17	-0.07
Tomatoes	**0.45**	0.26	0.02	0.00	0.14	**0.45**	-0.17	0.16	0.02	-0.07
Mushrooms	**0.52**	0.08	-0.04	-0.11	0.06	**0.39**	-0.34	0.01	0.02	0.09
Citrus fruit	0.21	0.18	-0.30	0.13	0.10	0.18	-0.25	0.20	-0.25	0.07
Bananas	0.14	0.05	0.07	0.25	-0.04	0.11	0.01	0.02	-0.04	0.04
Apples and pears	-0.01	0.08	-0.09	0.26	**0.37**	0.17	0.03	**0.40**	0.04	0.03
Berries – all types	0.14	0.06	-0.05	**0.36**	-0.03	0.12	-0.05	0.10	-0.09	0.26
Potatoes and potato products	-0.18	0.31	**0.38**	0.25	0.02	0.08	**0.52**	0.04	0.29	-0.00
Rice	**0.43**	-0.04	-0.18	0.09	0.02	**0.35**	-0.05	-0.08	-0.31	0.08
Pasta	**0.53**	-0.10	0.20	0.06	-0.04	**0.36**	-0.02	-0.10	-0.05	0.21
Dry cereals	0.15	-0.04	-0.28	0.19	0.13	0.14	0.03	0.26	-0.28	0.09
Bread, crackers etc. –white types	-0.16	0.06	0.04	0.13	**-0.62**	-0.03	0.19	**-0.49**	0.05	0.15
Bread, crackers etc. – brown/wholemeal types	-0.01	-0.03	0.33	0.18	**0.67**	0.07	0.17	**0.64**	0.19	-0.03
Fermented whole milk and milk products	-0.01	-0.01	-0.06	0.28	-0.01	0.09	0.28	-0.03	-0.15	0.05
Fermented medium, lowfat and skimmed milk prod	0.04	0.13	-0.21	0.02	0.33	0.14	-0.26	0.34	-0.10	-0.04
Nonfermented whole milk and milk products	-0.24	0.07	0.15	0.29	-0.29	-0.02	**0.49**	-0.21	-0.09	0.00
Nonfermented medium, lowfat and skimmed milk prod	0.04	-0.08	-0.03	-0.06	0.19	-0.01	-0.15	0.25	0.04	0.04
Full cream cheese and cheese spreads	-0.01	0.02	0.08	0.23	0.22	0.13	0.18	0.30	-0.12	0.12
Lowfat cheese and cheese spreads	-0.03	0.09	-0.02	-0.00	0.18	0.04	-0.23	0.10	0.10	-0.11
Beef and veal – fresh	-0.09	0.30	-0.11	-0.00	-0.07	0.04	0.16	0.01	-0.07	-0.16
Pork – fresh	0.01	0.06	**0.48**	-0.18	-0.17	-0.06	0.07	-0.19	**0.43**	0.06
Liver – fresh	0.15	0.13	0.10	-0.04	-0.10	0.17	-0.20	-0.06	0.28	0.10
Poultry – fresh	0.30	0.01	-0.01	-0.04	-0.01	0.24	-0.32	-0.11	0.02	0.05
Processed meat	0.02	0.02	**0.53**	0.07	-0.08	0.15	0.07	-0.03	**0.50**	0.20
Fish (shellfish and crustacean included)	0.29	0.14	-0.06	-0.11	-0.11	0.33	-0.26	-0.11	-0.04	0.09
Eggs	0.11	0.07	0.19	-0.04	-0.32	0.21	-0.01	-0.19	0.14	0.03
Butter	0.07	-0.01	-0.30	0.14	-0.32	0.13	0.13	-0.12	**-0.45**	0.07
Margarine	-0.30	0.25	0.25	0.12	-0.01	0.07	**0.47**	0.00	0.17	0.04
Lowfat margarine	0.06	-0.12	**0.46**	0.08	0.28	-0.11	-0.15	0.26	**0.50**	0.02
Oil	**0.58**	0.02	0.06	0.05	0.01	**0.41**	-0.14	-0.00	-0.25	0.09
Dressings and other similar sauces	0.07	-0.00	0.22	0.21	-0.05	0.16	0.03	0.06	0.34	0.05
Peanuts and other nuts	0.29	-0.11	0.06	0.20	-0.13	0.10	0.04	0.01	-0.01	**0.68**
Savory snacks	0.32	-0.22	0.06	0.13	-0.12	-0.02	-0.09	-0.07	0.03	**0.70**
Cakes, sweet breads and cookies	-0.06	-0.04	-0.01	**0.53**	0.02	-0.07	0.19	0.25	-0.01	**0.39**
Candies	0.05	-0.01	-0.04	**0.38**	-0.13	-0.10	0.07	0.11	-0.02	**0.54**
Added sugar	-0.07	-0.03	0.30	0.25	-0.34	-0.11	0.32	-0.28	-0.01	0.06
Sweet sandwich spread	-0.11	-0.02	0.04	**0.45**	0.01	-0.05	**0.47**	0.21	-0.09	0.04
Coffee	-0.02	0.04	**0.50**	-0.13	-0.02	-0.01	-0.00	-0.15	**0.44**	-0.13
Tea	-0.06	0.13	-0.27	0.31	0.01	0.00	0.19	0.31	-0.24	0.05
Fruit juices	0.10	0.07	-0.02	0.12	-0.07	0.15	0.09	-0.02	0.04	0.13
Soft drinks and syrups	0.07	-0.04	0.17	0.17	-0.21	-0.03	0.07	-0.15	0.15	0.25
Beer	0.13	0.02	0.21	-0.34	-0.10	0.06	-0.00	-0.23	-0.00	-0.03
Wine and fortified wine	**0.40**	-0.06	-0.09	-0.02	0.03	0.15	-0.25	-0.16	-0.09	0.04
Spirits	0.16	0.09	-0.03	-0.29	-0.19	0.15	-0.07	-0.23	0.07	0.06

SV="(Salad) Vegetables pattern"; CV="Cooked vegetables pattern"; PPP="Pork, processed meat, potatoes pattern"; SW="Sweet foods patterns"; B/W="Brown/white bread substitution pattern"; FD="Fat dairy pattern"; SN="Sweet and/or savory snacks pattern"
For convenience factor loadings **>0.35** are printed **bold**.

and cookies) was observed. Furthermore, in men a "Cooked vegetables" (CV:with high loadings on cooked leaf vegetables, cabbages, legumes and carrots) was observed, whereas in women, a "Fat dairy" (FD, with high loads on potatoes, nonfermented whole milk, margarine and sweet sandwich spread pattern) pattern was identified.
All dietary patterns, except B/W, positively associated with total energy intake. After adjustment for energy intake the patterns were no longer uncorrelated with each other, with Pearson correlation coefficients ranging from 0.00 to -0.45 for the PPP and SW pattern in men.

Table 3: Mean and percentage of sociodemographic and lifestyle variables within the first, third and fifth quintile of the energy adjusted dietary pattern scores among *men* of the subcohort of the NLCS (1986).

Sociodemographic and lifestyle characteristic	SV[1]			CV[1]			PPP[1]			SW[1]			B/W[1]		
	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5
Age (y)	61.9	61.4	60.5	60.6	61.3	62.0	62.1	61.7	60.2	60.7	61.6	61.7	61.1	61.3	61.0
High education (%)	10	14	32	23	19	12	25	17	10	12	16	25	17	16	18
Physical acitivity (min/day)	79.9	82.9	76.8	69.3	81.4	84.7	70.6	80.6	80.1	81.6	76.4	77.7	75.3	79.8	84.1
BMI (kg/m^2)	24.6	25.1	24.9	24.7	24.8	25.3	24.5	25.2	25.2	25.3	24.8	24.7	25.0	25.0	24.8
Smoking status (%)															
current	40	31	33	38	34	36	27	34	49	51	32	20	51	32	24
former	45	57	54	49	52	52	53	54	42	42	56	58	38	56	59
never	15	12	13	13	13	12	20	11	9	7	11	22	11	13	17
Number of cigarettes/day[2]	15.8	16.9	18.5	16.2	17.7	18.0	14.9	16.9	19.2	19.4	17.0	14.9	18.8	16.8	15.9
Years of smoking[2]	35.0	33.6	32.6	34.6	33.8	33.3	30.7	33.8	36.4	36.3	32.9	30.2	36.5	33.1	31.0
Supplement use (%)	16	21	31	22	23	26	30	24	14	20	22	29	18	23	28

[1] SV="(Salad) Vegetables pattern"; CV="Cooked vegetables pattern"; PPP="Pork, processed meat, potatoes pattern"; SW="Sweet foods patterns"; B/W="Brown/white bread substitution pattern"
[2] among former and current smokers

The mean of continuous and percentage of categorical sociodemographic and lifestyle variables in the first, third and fifth quintile of the dietary pattern scores are presented in **table 3** for men and **table 4** for women. Both men and women with a higher score on the SV pattern tended to be younger and more likely to use dietary supplements. In addition, these men were more likely to be higher educated, were less likely to be current smoker and more likely to be former smoker, smoked more cigarettes per day, for fewer years. Women with higher SV pattern scores spent on average more time on non-occupational physical activity and were more likely to be former smoker and less likely to never have smoked. Men with higher CV pattern scores tended to be older, have a slightly higher BMI, spent more time on non-occupational physical activity and were less likely to be higher educated. Men and women with a higher PPP pattern scores tended to be younger, were less likely to have attained higher vocational or university education, were more likely to be current smoker and less likely to be former or never smoker, smoked more cigarettes per day and for more years, were less likely to use dietary supplements. In addition, men with higher PPP pattern scores were more physically active and had a higher BMI. Men with a higher SW pattern score tended to be older, were more likely to be higher educated, had a lower BMI, were less likely to be a current smoker and more likely to be a former or a never smoker, smoked less cigarettes per day and smoked for fewer years, and were more likely to use dietary supplements. Women with higher SN pattern scores tended to be younger, were more likely to be higher educated, had a lower BMI, were more likely to be former smoker and less likely to never have smoked, smoked less cigarettes per day,

Table 4: Mean and percentage of sociodemographic and lifestyle variables within the first, third and fifth quintile of the energy adjusted dietary pattern scores among *women* of the subcohort of the NLCS (1986).

Sociodemographic and lifestyle characteristic	SV[1]			FD[1]			B/W[1]			PPP[1]			SN[1]		
	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5
Age (y)	62.1	61.1	61.0	60.9	61.7	61.7	61.3	61.5	61.4	61.9	61.6	60.7	61.5	61.4	61.0
High education (%)	7.3	9.7	11.0	13.8	10.0	5.9	5.5	8.7	13.3	16.3	7.0	4.41	5.6	8.3	16.1
Physical acitivity (min/day)	55.2	61.7	75.5	66.6	64.5	58.9	61.1	63.6	66.6	66.6	66.4	67.7	68.9	63.1	64.5
BMI (kg/m2)	24.7	25.3	25.1	25.4	25.1	24.6	24.8	25.1	24.8	23.9	24.9	26.3	25.3	24.9	24.5
Smoking status (%)															
current	21	21	23	22	21	22	36	18	12	15	22	27	26	19	22
former	18	18	27	28	18	14	21	24	19	26	22	18	13	20	30
never	61	61	50	50	61	64	43	58	69	59	57	55	61	61	48
Number of cigarettes/day[2]	11.5	10.2	12.0	11.2	11.6	12.1	14.2	10.7	8.6	9.3	11.6	13.9	13.1	11.1	10.3
Years of smoking[2]	29.6	28.0	26.7	27.2	28.0	29.5	31.2	26.7	24.2	25.8	28.7	28.8	29.7	27.9	28.0
Supplement use (%)	33	40	42	45	40	32	30	36	44	49	35	27	32	37	46

[1] SV="(Salad) Vegetables pattern"; "; FD="Fat dairy pattern"; B/W="Brown/white bread substitution pattern; PPP="Pork, processed meat, potatoes pattern"; SN="Sweet and savory snacks pattern"
[2] among former and current smokers

and were more likely to use dietary supplements. Men and women with higher B/W pattern scores were more physically active, were less likely to be current smoker, tended to smoke less and for fewer years and were more likely to use dietary supplements. However, men were most likely to be former smoker, whereas women were more likely to never have smoked. Furthermore, women with higher B/W pattern scores were more likely to be higher educated and to use dietary supplements. Women with high scores on the "fat dairy" pattern tended to be older, were less likely to be higher educated, were less physically active, had a lower BMI, were more likely to be a never smoker and less likely to be a former smoker and less likely to use dietary supplements.

To determine whether the associations between the factor scores were influenced by the energy adjustment of the pattern scores, we repeated the analyses using the original energy-unadjusted factor scores. In **figures 1 to 6** the mean and percentage of the sociodemographic and lifestyle variables are presented across the quintiles of the energy adjusted and the original (not for energy adjusted) PPP pattern scores, both among men and women. The (unadjusted) PPP pattern score was strongly correlated with total energy intake among men ($R_{pearson}$ 0.50) but not so strongly among women ($R_{pearson}$ 0.16).
Most associations were similar for the unadjusted and the adjusted factor scores.

DISCUSSION

Although the NLCS is a prospective cohort study, the present analysis is a cross sectional measurement of diet and lifestyle and does therefore not allow conclusions suggesting causal directions. Furthermore, we can not exclude the possibility that the observed dietary patterns or their association with lifestyle in the present study have been distorted by selective under- or over-reporting of particular food items or selective participation. For example, the *higher* BMI in men and women with *lower* scores on the Sweet foods, sweet and savory snacks or fat dairy patterns might demonstrate that these subjects avoided these foods because they restrained their food intake. However, another possibility would

Fig 1. Mean age (yr) in quintiles of PPP pattern score

Fig 2. Percentage with higher vocational or university education (%) in quintiles of PPP pattern score

Fig 3. Mean physical activity outside profession (min/day) in quintiles of PPP pattern score

Fig 4. Mean BMI (kg/m^2) in quintiles of PPP pattern score

Fig 5. Percentage of current smokers (%) in quintiles of PPP pattern score

Fig 6. Use of dietary supplements (%) in quintiles of PPP pattern score

PPP factor score, men
PPP factor score, women

energy adjusted PPP factor score, men
energy adjusted PPP factor score, women

be that persons with higher BMI were systematically under reporting their consumption of the sweet foods, snacks or fat dairy products that characterized these dietary patterns.
We selected the input food variables on the basis of their possible relevance in the etiology of cancer or a suspected important role in the diet, e.g. foods for which individuals make a deliberate choice, such as brown or white bread and high fat or low fat dairy products were included in the factor analysis separately. Furthermore, we performed sensitivity analyses,[33] which suggested that the dietary pattern approach is robust for the extracted number of patterns, the distribution of input variables and energy adjustment of the input variables.[33] The extracted factors were qualitatively similar, regardless of whether the input food variables were adjusted for energy. As the factor scores are calculated from the product of loading and amount eaten, most of the patterns correlated with energy. Therefore, in the analyses described in the current paper, we adjusted the factor scores for total energy intake to prevent that the associations with lifestyle were determined by the total amount eaten, instead of the composition of the diet. For most food groups the correlation with the factor scores (i.e. the factor loadings) did not materially change. In addition, the associations of the dietary pattern scores and the sociodemographic or lifestyle variables were similar for the unadjusted and the adjusted factor scores. The rather logical explanation of this observation is the orthogonality of the dietary patterns: a person with a high score on for example the SV pattern is not more likely to have a low score on the PPP pattern, than a person with a low score on the SV pattern. This holds by definition for all patterns, because of the method of extracting the patterns and by using an orthogonal rotation of the factor solution. This automatically reduces the influence of total energy intake, by preventing that large eaters will have higher scores on all patterns and small eaters will have low scores on all patterns.
Nevertheless, without energy adjustment, almost all of the dietary patterns were associated positively with total energy intake. This is consistent with the finding that energy intake was positively associated with all patterns in most published studies.[1]
The "Salad Vegetable" pattern and the "Brown/white bread substitution" pattern, dietary patterns that one intuitively would associate with "health consciousness" were indeed associated with higher education and with health conscious behavior such as physical activity, not smoking (never or having quit), and the use of dietary supplements. This is in agreement with the finding that in most published studies a dietary pattern characterized by consumption of fruits and vegetables was positively associated with education, physical activity, use of dietary supplements and inversely associated with smoking.[1]
However, most published studies found this pattern to be positively associated with age, whereas we found the "Salad vegetables" pattern to be inversely associated with age, both in men and women.[1] Our interpretation of the inverse association with age is that this pattern reflects a non-traditional choice of foods, such as salad vegetables, but also pasta, rice and oil, that were not commonly eaten by older subjects at that time. This would be in agreement with our finding that the cooked vegetables pattern was positively associated with older age, as this dietary pattern reflects the traditional hot meal in the Netherlands consisting of cooked vegetables, potatoes and meat.
The dietary pattern characterized by consumption of Pork, processed meat and potatoes, was associated with lower education and with less health conscious and less favorable characteristics, such as having a higher BMI, smoking and not using dietary supplements. This is in agreement with the finding that in the published literature a dietary pattern characterized by consumption of meat and fatty foods mostly was inversely associated with age and positively with smoking.[1] The published literature on the association with

education and the use of dietary supplements tended in the expected direction, however, too few studies reported results on these associations.
Although we observed a clear positive association between BMI and PPP pattern score, independent of total energy intake, in a review Togo *et al*[19], showed no consistent associations between BMI or obesity and food intake patterns. In our review, we also showed that most published studies did not observe a significant association between BMI and dietary patterns.[1]
The dietary patterns characterized by consumption of sweet foods among men and -to a lesser extent- both sweet and savory snacks among women, were associated with higher education, lower BMI, former smoking and the use of dietary supplements. This is in agreement with our finding that most published studies described an inverse association between a pattern characterized by sweet foods and smoking. The first explanation that comes to mind is likely that subjects who quit smoking will start eating more candy. On the other hand, we consider it likely that subjects with a higher education are less likely to have started smoking, because of lower peer pressure, and that higher educated personas are also more likely be health conscious.
The dietary pattern characterized by consumption of full fat dairy products among women appeared to reflect a more traditional food choice. In addition, these women were older, not highly educated, and more likely to never have smoked, all more "traditional" characteristics among women. Furthermore, these women were less physically active and did not use dietary supplements, not indicating weight- or health conscious behavior.
Characteristics that were considered to fit into a health conscious behavior such as higher education, and not smoking and to a lesser extent non-occupational physical activity were positively associated with the dietary patterns characterized by consumption of salad vegetables and by consumption of brown instead of white bread and inversely with the pattern characterized by consumption of pork, processed meat and potatoes. Characteristics reflecting a more traditional personality, such as older age, lower education and in women smoking less, were associated with the "cooked vegetables" pattern among men and the "fat dairy" pattern among women. In conclusion, the observed associations between these dietary patterns and sociodemographic and lifestyle factors were generally in the expected direction and in agreement with the majority of the published literature. This increases the confidence that the statistical solution of the dietary pattern analysis is meaningful. We consider it likely that dietary patterns are not isolated, but are part of a broader lifestyle encompassing a combination of two ore more dimensions, including health consciousness and traditionalism.

REFERENCES

1. Balder HF, van den Brandt PA, Goldbohm RA. Factor analysis for food consumption data: A review of the literature. Submitted for publication 2004.

2. Stellman SD. Chairman's remarks. Natl.Cancer Inst.Monogr. 1983;67:145-7.

3. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity [comment] [see comments]. Am.J.Epidemiol. 1998;148:17-9.

4. Schulze MB, Hoffmann K, Kroke A, Boeing H. Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. Br.J.Nutr. 2001;85:363-73.

5. Gex-Fabry M, Raymond L, Jeanneret O. Multivariate analysis of dietary patterns in 939 Swiss adults: sociodemographic parameters and alcohol consumption profiles. Int.J.Epidemiol. 1988;17:548-55.

6. Nicklas TA, Webber LS, Thompson B, Berenson GS. A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors: the Bogalusa Heart Study. Am.J.Clin.Nutr. 1989;49:1320-7.

7. Barker ME, McClean SI, Thompson KA, Reid NG. Dietary behaviours and sociocultural demographics in Northern Ireland. Br.J.Nutr. 1990;64:319-29.

8. Hebert JR, Kabat GC. Differences in dietary intake associated with smoking status. Eur.J.Clin.Nutr. 1990;44:185-93.

9. Hebert JR, Kabat GC. Implications for cancer epidemiology of differences in dietary intake associated with alcohol consumption. Nutr.Cancer 1991;15:107-19.

10. Randall E, Marshall JR, Graham S, Brasure J. High-risk health behaviors associated with various dietary patterns. Nutr.Cancer 1991;16:135-51.

11. Whichelow MJ, Prevost AT. Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. Br-J-Nutr 1996;76:17-30.

12. Prevost AT, Whichelow MJ, Cox BD. Longitudinal dietary changes between 1984-5 and 1991-2 in British adults: association with socio-demographic, lifestyle and health factors. Br.J.Nutr. 1997;78:873-88.

13. Slattery ML, Edwards SL, Boucher KM, Anderson K, Caan BJ. Lifestyle and colon cancer: An assessment of factors associated with risk. Am.J.Epidemiol. 1999;150:869-77.

14. Maskarinec G, Novotny R, Tasaki K. Dietary patterns are associated with body mass index in multiethnic women. J.Nutr. 2000;130:3068-72.

15. Mullen K, Williams R, Hunt K. Irish descent, religion and food consumption in the west of Scotland. Appetite 2000;34:47-54.

16. North K, Emmett P. Multivariate analysis of diet among three-year-old children and associations with socio-demographic characteristics. The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) Study Team. Eur-J-Clin-Nutr 2000;54:73-80.

17. Williams DE, Prevost AT, Whichelow MJ, Cox BD, Day NE, Wareham NJ. A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. Br-J-Nutr 2000;83:257-66.

18. Fung TT, Rimm EB, Spiegelman D et al. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. Am.J.Clin.Nutr. 2001;73:61-7.

19. Togo P, Osler M, Sorensen T, Heitmann B. Food intake patterns and body mass index in observational studies. Int.J.Obes. 2001;25:1741-51.

20. Tseng M, DeVellis RF. Fundamental dietary patterns and their correlates among US whites. JADA 2001;101:929-32.

21. Mishra G, Ball K, Arbuckle J, Crawford D. Dietary patterns of Australian adults and their association with socioeconomic status: results from the 1995 National Nutrition Survey. Eur.J.Clin.Nutr. 2002;56:687-93.

22. Sichieri R. Dietary patterns and their associations with obesity in the brazilian city of rio de janeiro. Obes.Res. 2002;10:42-8.

23. Sanchez-Villegas A, Delgado-Rodriguez M, Martinez-Gonzalez MA, Irala-Estevez J. Gender, age, socio-demographic and lifestyle factors associated with major dietary patterns in the Spanish Project SUN (Seguimiento Universidad de Navarra). Eur.J.Clin.Nutr. 2003;57:285-92.

24. Dam, R. M. v., Grievink, L., Ocke, M. C., and Feskens, E. J. Patterns of food consumption and risk factors for cardiovascular disease in the general Dutch population. Am J Clin Nutr 77(5), 1156-63. 2003.

25. Costacou T, Bamia C, Ferrari P, Riboli E, Trichopoulos D, Trichopoulou A. Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population. Eur.J.Clin.Nutr. 2003;57:1378-85.

26. Bell RA, Quandt SA, Vitolins MZ, Arcury TA. Dietary patterns of older adults in a rural, tri-ethnic community: a factor analysis approach. Nutrition Research 2003;23:1379-90.

27. Tseng M, Breslow RA, DeVellis RF, Ziegler RG. Dietary patterns and prostate cancer risk in the National Health and Nutrition Examination Survey Epidemiological Follow-up Study cohort. Cancer Epidemiol.Biomarkers Prev. 2004;13:71-7.

28. Sieri S, Krogh V, Pala V et al. Dietary patterns and risk of breast cancer in the ORDET cohort. Cancer Epidemiol.Biomarkers Prev. 2004;13:567-72.

29. Kerver, J. M., Yang, E. J., Bianchi, L., and Song, W. O. Dietary patterns associated with risk factors for cardiovascular disease in healthy US adults. Am J Clin Nutr 78, 1103-1110. 2003.

30. Kim DH, Smith-Warner SA, Spiegelman D et al. Pooled analysis of prospective studies on folate and colorectal cancer. Submitted for publication 2004.

31. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. J Clin Epidemiol 1990;43:285-95.

32. Goldbohm RA, van den Brandt PA, Brants HA et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur.J.Clin.Nutr. 1994;48:253-65.

33. Balder HF, Virtanen M, Brants HA et al. Common and Country-Specific Dietary Patterns in Four European Cohort Studies. J.Nutr. 2003;133:4246-51.

5

Dietary patterns associated with colon and rectal cancers: Results from the DIETSCAN Project

L Beth Dixon, Helena F Balder, Mikko Virtanen, Bahram Rashidkhani, Satu Mannisto, Vittorio Krogh, Franco Berrino, Piet A van den Brandt, Anne Hartman, Pirjo Pietinen, Frans Tan, Jarmo Virtamo, Alicja Wolk, R Alexandra Bausch-Goldbohm.

American Journal of Clinical Nutrition 2004;80:1003-11.

ABSTRACT

Background: Analysis of dietary *patterns*, versus single nutrients or foods, may provide insight regarding diet and colon and rectal cancer.
Objective: A primary aim of the DIETSCAN Project was to develop and apply a common methodological approach to study dietary patterns and cancer in four European cohorts: the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (Finland-ATBC), the Netherlands Cohort Study on diet and cancer (NLCS), the Swedish Mammography Cohort (SMC), and the Ormoni e Dieta nella Eziologia dei Tumori (Italy-ORDET). Three cohorts (ATBC, NLCS, SMC) provided data on colon and rectal cancer for this study.
Design: The cohorts were established between 1985 and 1992 with follow-up through record linkage of the cohorts with national cancer registries, and used validated semi-quantitative food frequency questionnaires (FFQs) to collect dietary data at baseline.
Results: Exploratory factor analysis, conducted within each cohort, identified three to five stable dietary patterns. Two dietary patterns, 'Vegetables' (VEG) and 'Pork, Processed Meat, Potatoes' (PPP), were common across all cohorts. After adjustment for potential confounders, PPP was associated with increased risk of colon cancer in the SMC women ($Q4_{multivariate}$ RR=1.62, 95% CI=1.12-2.34, p for trend=0.01). PPP was also associated with increased risk of rectal cancer in the ATBC men ($Q4_{multivariate}$ RR=2.21, 95% CI=1.07-4.57, p for trend=0.05). Neither pattern was associated with risk of colon or rectal cancer in the NLCS women and men.
Conclusion: Although certain dietary patterns may be consistent across European countries, association of those dietary patterns with risk of colon and rectal cancer is not conclusive.

INTRODUCTION

Colorectal cancer is the second most common cancer, next to lung cancer for men and breast cancer for women, in Europe and North America.[1] In Europe alone, 363,000 incident cases of colorectal cancer were reported in 2000. Approximately 6% of men and women by age 75 are affected. In a recent report on Diet, Nutrition, and the Prevention of Chronic Disease,[2] it is suggested that up to 80% of differences in colorectal cancer rates between countries may be attributable to diet-related factors, particularly overweight/obesity and physical activity. The expert panel also noted that "there is almost universal agreement that some aspects of the westernized diet are a major determinant of risk" and that "there is some evidence that risk is increased by high intakes of meat and fat, and that risk is decreased by high intakes of fruits and vegetables, dietary fiber, folate, and calcium, but none of these hypotheses has been firmly established." Previous studies have focused on the association of those single nutrients or food groups with colorectal cancer outcomes. However, this reductionist approach does not take into account that foods and food constituents likely act synergistically.[3] More recently, nutrition epidemiologists have used factor analysis to examine whether *dietary patterns*, or combinations of foods, are associated with colorectal cancer.[4-7] Factor analysis disentangles complex dietary data into a small number of dietary characteristics (labeled as 'patterns' according to their dominant foods) which are relatively independent from one another and which can be treated as risk factors in subsequent analysis of cancer risk. Results from such analyses[4;5] have shown more consistent and stronger associations between certain dietary patterns and colorectal cancer than those identified for individual nutrients or foods. Such patterns analysis is promising but requires further investigation, particularly across cultures. A primary aim of the DIETSCAN project was to develop and apply a common methodological approach to study dietary patterns and cancer in four European cohorts: the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (Finland-ATBC), the Netherlands Cohort Study on diet and cancer (NLCS), the Swedish Mammography Cohort (SMC), and the Ormoni e Dieta nella Eziologia dei Tumori (Italy-ORDET), and to determine whether the associations of those dietary patterns with cancer are consistent across cultures. Factor analysis was selected as the common methodology and applied 'a posteriori' to the dietary data collected in each cohort. This study reports findings from the factor analysis of dietary patterns and their association with colon and rectal cancer in three of the cohorts (ATBC, NLCS, SMC) that provided colorectal cancer data.

SUBJECTS AND METHODS

Study population Four European cohorts were selected for DIETSCAN because they met the following criteria: a) prospective cohort study designed to investigate the effects of diet on risk of various cancers; b) dietary assessment of common foods and beverages in the entire diet; c) validated dietary assessment instrument. Three of the DIETSCAN cohorts (ATBC, NLCS, SMC) provided follow-up data on colon and rectal cancer from record linkage of the cohorts with national cancer registries (**Table 1**). Colon cancers were defined as those occurring above the peritoneal delineation of the abdominal cavity (ICD-9 code 153-), and rectal cancers were those occurring below this delineation (ICD-9 code 154-). A brief description of each cohort follows. The ATBC Cancer Prevention Study was

Table 1.Characteristics of DIETSCAN cohort studies with colon and rectal cases

	ATBC Men [1]	NLCS Men [2]	NLCS Women [2]	SMC Women [3]
Country	Finland	the Netherlands	the Netherlands	Sweden
Baseline year(s)	1985-1988	1986	1986	1987-1990
Follow-up year(s)	1999	1992	1992	2001
Baseline cohort size	29 133	58,279 [4]	62,573 [4]	66 651
Age range	50-69	55-69	55-69	40-74
Age (y)	57.2±5.1	61.4±4.2	61.4±4.3	53.7±9.7
BMI	26.3±3.8	25.0±2.6	25.1±3.5	24.8±4.4
Education % with high school, vocational or university training	35.5	38.1	31.3	12.4
Smoking				
Number of cigarettes/day	20.4±8.8	14.5±11.5	4.5±7.5	[5]
Years smoked	35.9±8.4	28.9±15.9	11.3±15.8	[5]
Physical activity		81.1±68.3 [6]	62.9±51.6 [6]	[5]
% not working	41.8 [7]			
% < 30 min/day		18.3 [6]	24.7 [6]	
% who do mainly office work	14.0 [7]			
% 30-60 min/day		30.3 [6]	31.9 [6]	
% who walk a lot	18.5 [7]			
% 60-90 min/day		19.1 [6]	21.8 [6]	
% who walk and lift a lot	16.6 [7]			
% > 90 min/day		32.3 [6]	21.6 [6]	
% who do heavy physical work	9.0 [7]			

[1] ATBC = Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); [2] NLCS = Netherlands Cohort Study on Diet and Cancer (Netherlands); [3] SMC = Swedish Mammography Cohort (Sweden); ORDET = ORmoni e Dieta nella Eziologia dei Tumori (Italy). [4] A subcohort of 3500 participants (1688 men, 1812 women) were followed biennially for vital status information. [5] Data on smoking and physical activity were not collected in the SMC; [6] Physical activity outside profession (minutes/day); [7] Type of occupational physical activity in the past year.

a randomized, double-blind, placebo-controlled chemoprevention trial of male smokers in southwestern Finland.[8] Between 1985-1988, 29,133 men between the ages of 50-69 who smoked five or more cigarettes per day were randomized to receive a daily dose of *dl*-alpha-tocopherol (50 mg), B-carotene (20 mg), both, or a placebo. Prior to randomization, 27,111 men (93%) completed a 276-item FFQ that asked about frequency and portion sizes of common foods and beverages in the previous year. Validity of the ATBC FFQ was determined by 12 2-day food records collected from 190 men over a 6-month period.[9] The NLCS on diet and cancer is a population-based prospective cohort study of 58,279 men and 62,573 women, selected from 204 Dutch municipalities with computerized population registries, who were between the ages of 55-69 when the study began in 1986.[10] A case-cohort approach was used for data processing and analysis. In this design, a subcohort is randomly selected from the entire cohort, which provides a comparison group for each occurrence time, i.e., those subjects in the subcohort who are still at risk for the disease under study at a given failure time serve as controls for the occurring failure whether that

failure occurs inside or outside the subcohort. A subcohort of 3500 participants (1688 men and 1812 women) was followed biennially for vital status information. At baseline, 1525 men and 1598 of the subcohort completed a mailed, self-administered questionnaire on dietary habits, including a 150-item FFQ that asked about usual intake of foods and beverages in the previous year. The NLCS FFQ has been validated against 9-day dietary records from a subgroup of 109 participants.[11] The SMC is comprised of 61,463 women from Uppsala and Vastmanland counties in Sweden who were born between 1914 and 1948 and who were invited to participate in a population-based mammography-screening program in 1987-1990.[12] The women completed a 67-item FFQ that asked about frequencies of intake of commonly consumed foods and beverages in the previous 6 months. The SMC FFQ was validated against four sets of weighed 7-day food records from a subgroup of 129 women.[6]

Human Subjects The ATBC Cancer Prevention Study was approved by the institutional review boards of the National Public Health Institute of Finland and the U.S. National Cancer Institute. The NLCS on diet and cancer was approved by the institutional review boards of the TNO Toxicology and Nutrition Institute (Zeist) and the University of Limburg (Maastricht) in the Netherlands. The SMC was approved by the Institutional Ethics Committee at Uppsala University and at Karolinska Institute, Stockholm, Sweden.

Food grouping and assessment of dietary patterns The validated semi-quantitative FFQs used to collect dietary data at baseline differed in number of items, reference period, units of frequency and quantification of portion size, and overall level of detail (**Table 2**). To achieve the objectives of the DIETSCAN Project, a common food grouping was developed.[13] Food items from the FFQs were aggregated into 51 food groups that included foods common to all countries as well as specific foods included in each FFQ. These food groups were selected because of their role in the diet and possible relevance to cancer etiology. To identify dietary patterns within each cohort, exploratory factor analysis using principal components analysis was conducted using the grams per day of the total intake of the FFQ- derived food groups for each country. Food groups with >75% of 0, values representing non- users, were dichotomized. Factors were rotated by orthogonal Varimax transformation. Eigenvalues > 1 and the point at which the scree plot levels off, traditional criteria in factor analysis, were used to determine the number of factors to extract within each cohort (i.e., 3 for ATBC, 5 for NLCS, 4 for SMC). Food groups with absolute factor loadings > 0.35 were considered as contributing to a dietary pattern (**Appendix Table 1**). Factor loadings represent correlation coefficients between food groups and dietary patterns. Food groups with positive loadings are positively associated with a dietary pattern; food groups with negative loadings are inversely associated with a dietary pattern. The proportion of variance explained by each factor was calculated by dividing the sum of the squares of the respective factor loadings by the number of variables (i.e., food groups). Names of dietary patterns within each cohort were determined according to the dominant foods (i.e., foods with high loadings) in the respective patterns. Commonality of dietary patterns across cohorts was determined descriptively and named according to common dominant foods within the respective patterns. Further detail regarding the food grouping and dietary pattern assessment, and related sensitivity analyses conducted in DIETSCAN, is provided by Balder *et al.*[13]

Table 2. Characteristics of the FFQ and factor analysis-derived dietary patterns in the DIETSCAN cohort studies

	ATBC Men	NLCS Men	NLCS Women	SMC Women
Total number of items	276	150	150	67
FFQ reference period	12 months	12 months	12 months	6 months
Frequency	units/times per day/week/month	7 categories (never - 6-7x/wk)	7 categories (never - 6-7x/wk)	8 categories (never/seldom - >4 x/d)
Quantification	portion size picture booklet (3-5 per item)	natural or household units (fixed weight per unit)	natural or household units (fixed weight per unit)	age-specific standard portion sizes
# Food groups used to determine factors	46	49	49	42
# Factors extracted	3	5	5	4
Factor (variance)				
Vegetables	**Salad vegetables** (9.6%)	**Salad vegetables** (5.6%)	**Salad vegetables** (6.3%)	**Salad vegetables** (6.9%)
Pork, processed meat, potatoes	**PPP** (5.3%)	**PPP** (4.2%)	**PPP** (4.3%)	**PPP** (5.4%)
Other factors	Alcohol (5.4%)	Cooked vegetables (4.8%) Sweet & savory snacks (4.3%) Brown/white bread substitute (4.1%)	Full fat dairy (4.4%) Brown/white bread substitute (4.3%) Sweet & savory snacks (3.9%)	Alcohol (5.3%) Bread cheese cookies (4.2%)
Mean daily total energy (mJ) from FFQ	6.73	9.04	7.07	5.56

Statistical analyses Within ATBC and SMC, Cox proportional hazards models were constructed to estimate hazard ratios and 95% confidence intervals relating the common factors to the incidence of colorectal, colon, and rectal cancers. NLCS used survival analysis with exponential distribution to estimate the standard errors using the robust option to account for additional variance introduced by the case-cohort design. Prior to statistical analysis, factor scores were determined by summing the standardized intakes from each food group weighted by the factor loadings. Thus, within each cohort, each participant had a unique score for each factor. High scores represented high intake of foods within the corresponding dietary pattern; low scores represented low intake of those foods. Because factor scores represent standardized variables, each score had a mean of zero and a standard deviation of one. The statistical models included the consistent factor scores (VEG and PPP) and the study- specific factor scores entered as linear (i.e., one estimate per factor), as grouped (i.e., quartiles), or as grouped continuous with medians (i.e., each group assigned the median value within each group) to test for linear trend. The models included all factor scores because dietary patterns are conditional on each other. The models were also adjusted for potential confounding variables determined by previous analyses of nutrients or food groups and incidence of colorectal cancer conducted within each cohort.[14-16] In ATBC, the models included age (<55 y, 55-59.9 y, 60- 64.9 y, >65 y), treatment group (a-tocopherol, b-carotene, both, placebo), body mass index (BMI) (<25 kg/m2, 25-27.4 kg/m2, >27.5 kg/m2), level of education (primary, high school, vocational, university), number of cigarettes smoked per day (<17, 17-20, >20), number of years of smoking regularly (<33, 34-40, >40), and occupational physical activity in the past year

(not working, mainly office work, quite a lot of walking but no lifting or carrying, lot of walking and lifting, heavy physical work). In NLCS, the models included age, BMI (<23 kg/m2, 23-25 kg/m2, 25-27 kg/m2, >27 kg/m2), education (primary, junior high school, vocational, high school/higher vocational/university), number of cigarettes smoked per day, number of years of smoking regularly, minutes per day of physical activity outside of the profession (<30 min/d, 30-60 min/d, 60-90 min/d, >90 min/d), and family history of colorectal cancer (yes, no). In SMC, the models included age, BMI (quartiles), and education (less than high school, high school, university). All models also included energy intake as a continuous variable. Because energy-contributing food groups tend to group together, pattern analysis was performed with the food group variables both unadjusted and adjusted for energy using the residual method.[17] The factor scores remained unadjusted in this analysis because energy adjustment of the food group variables yielded comparable factor solutions[13]. All reported p values are two-sided. PROC FACTOR and PROC PHREG in SAS version 8.2 (SAS Institute Inc., Cary, North Carolina) were used to analyze ATBC and SMC data. 'Streg, robust' in Stata version 6 (Stata Corporation, College Station, Texas) was used to analyze NLCS data.

RESULTS

Using a standardized approach, three to five stable dietary patterns were identified in cohort studies from Finland, the Netherlands, and Sweden (**Table 2**). Two dietary patterns, 'Vegetables' (VEG) and 'Pork, Processed Meat, Potatoes' (PPP) were observed in all three cohorts. The VEG pattern was characterized by intakes of vegetables and legumes, citrus fruit and berries, pasta and rice, poultry and fish, and oil and salad dressings (**Appendix Table 1**), and was correlated with intakes of Vitamins A, C, E, and folate, and polyunsaturated fatty acids. This pattern accounted for the most variance in dietary intake, ranging from 5.6% in NLCS men to 9.6% in ATBC men. The PPP pattern was characterized by intakes of pork, processed meats, potatoes, and coffee (**Appendix Table 1**), and was correlated with intakes of energy, protein, carbohydrate, fat, saturated and monounsaturated fatty acids, cholesterol, B vitamins, and minerals. This pattern accounted for 4-5% of the variance in dietary intake across the cohorts. In general, both patterns were inversely correlated with age and positively but very weakly correlated with BMI, smoking, and physical activity in the cohorts. The VEG pattern was positively correlated with education, but only in ATBC men.

In general, the VEG pattern was not associated with risk of colorectal, colon, or rectal cancer in the cohorts (**Table 3**). In NLCS women, an inverse association between the VEG pattern and risk of colon cancer was suggestive in the age- and energy-adjusted models when modeled as a continuous variable, but did not achieve statistical significance in the multivariate- adjusted models. In ATBC men, a positive association between the VEG pattern and colorectal cancer was suggestive but only in the age- and energy-adjusted models. A positive association between the VEG pattern and rectal cancer was also suggestive but quartile risk estimates in all models were not statistically significant. The PPP pattern was associated with increased risk of colorectal, colon, and rectal cancers in two of the cohorts (**Table 4**). In SMC women, after adjustment for potential confounders, the PPP pattern was associated with 18% increased risk of colorectal cancer when modeled as a continuous variable and 62% increased risk in the fourth quartile of the categorical analyses with significant tests for trend. This pattern was associated with

Table 3: Relative risks of colorectal cancer by linear and quartile models of the Vegetables (VEG) pattern in the Dietary Patterns and Cancer (DIETSCAN) Project cohort studies[1]

	Linear[2]		Q1	Q2		Q3		Q4		P value
	RR	95%CI	Ref	RR	95%CI	RR	95%CI	RR	95%CI	for Trend
Colorectal cancer										
ATBC men										
cases	322		68	64		101		89		
person years	286,967		69360	70433		72421		74753		
RR (age, kjoules)	1.10	0.99-1.23	1.0	0.94	0.66-1.32	1.51	1.10-2.05	1.34	0.97-1.85	0.02
RR (multivariate[5])	1.07	0.95-1.20	1.0	0.90	0.64-1.27	1.41	1.03-1.93	1.22	0.87-1.73	0.09
NLCS men										
cases	660		162	168		167		163		
person years[4]	10,496		2638	2604		2614		2640		
RR (age, kjoules)	1.06	0.96-1.17	1.0	1.08	0.83-1.42	1.14	0.87-1.49	1.16	0.88-1.53	0.27
RR (multivariate[5])	1.03	0.93-1.15	1.0	1.03	0.78-1.36	1.04	0.79-1.38	1.04	0.78-1.39	0.41
NLCS women										
cases	512		134	131		132		115		
person years	11328		2778	2810		2804		2936		
RR (age, kjoules)	0.90	0.81-1.01	1.0	1.02	0.76-1.36	1.05	0.78-1.42	0.87	0.63-1.20	0.42
RR (multivariate[c])	0.92	0.81-1.03	1.0	1.04	0.77-1.39	1.08	0.79-1.48	0.91	0.65-1.27	0.78
SMC women										
cases	586		179	147		130		130		
person years	749282		190157	188548		186562		184016		
RR (age, kjoules)	1.03	0.93-1.14	1.0	0.85	0.65-1.12	0.90	0.68-1.19	0.97	0.72-1.30	0.95
RR (multivariate[6])	1.03	0.93-1.14	1.0	0.91	0.73-1.13	0.89	0.70-1.13	0.99	0.77-1.27	0.90
Colon cancer										
ATBC men										
cases	191		39	42		62		48		
person years	287375		69459	70504		72552		74860		
RR (age, kjoules)	1.06	0.92-1.23	1.0	1.05	0.68-1.63	1.57	1.05-2.36	1.22	0.79-1.88	0.24
RR (multivariate[3])	1.02	0.87-1.19	1.0	0.99	0.64-1.54	1.43	0.95-2.16	1.05	0.66-1.67	0.66
NLCS men										
cases	400		100	105		103		92		
person years[4]	10509		2638	2610		2615		2645		
RR (age, kjoules)	1.07	0.95-1.21	1.0	1.11	0.80-1.53	1.17	0.85-1.60	1.11	0.80-1.55	0.52
RR (multivariate[5])	1.02	0.89-1.16	1.0	1.04	0.75-1.46	1.03	0.74-1.43	0.93	0.65-1.32	0.93
NLCS women										
cases	360		105	92		89		74		
person years	11334		2778	2810		2810		2936		
RR (age, kjoules)	0.86	0.75-0.98	1.0	0.92	0.66-1.27	0.92	0.66-1.30	0.74	0.51-1.07	0.12
RR (multivariate[c])	0.87	0.76-1.01	1.0	0.94	0.68-1.31	0.96	0.67-1.37	0.78	0.54-1.15	0.29
SMC women										
cases	396		122	94		91		89		
person years	749964		190356	188763		186681		184165		
RR (age, kjoules)	1.03	0.91-1.16	1.0	0.85	0.65-1.12	0.90	0.68-1.19	0.97	0.72-1.30	0.89
RR (multivariate[6])	1.03	0.91-1.16	1.0	0.85	0.65-1.12	0.90	0.68-1.19	0.96	0.71-1.30	0.87

Table 3 continued

	Linear[2]		Q1	Q2		Q3		Q4		P value
	RR	95%CI	Ref	RR	95%CI	RR	95%CI	RR	95%CI	for Trend
Rectal cancer										
ATBC men										
cases	133		29	23		40		41		
person years	287486		69475	70515		72597		74899		
RR (age, kjoules)	1.15	0.98-1.36	1.0	0.81	0.47-1.40	1.44	0.89-2.33	1.51	0.93-2.46	0.02
RR (multivariate[3])	1.14	0.95-1.36	1.0	0.80	0.46-1.38	1.40	0.86-2.30	1.48	0.88-2.49	0.04
NLCS men										
cases	260		62	63		64		71		
person years[4]	10525		2645	2612		2621		2647		
RR (age, kjoules)	1.04	0.91-1.19	1.0	1.04	0.70-1.54	1.09	0.74-1.62	1.23	0.84-1.81	0.26
RR (multivariate[5])	1.05	0.91-1.21	1.0	0.99	0.66-1.49	1.08	0.73-1.61	1.23	0.83-1.83	0.16
NLCS women										
cases	152		29	39		43		41		
person years	11355		2782	2815		2812		2946		
RR (age, kjoules)	1.01	0.85-1.20	1.0	1.39	0.84-2.30	1.53	0.91-2.56	1.34	0.76-2.36	0.37
RR (multivariate[6])	1.01	0.85-1.21	1.0	1.40	0.83-2.34	1.51	0.89-2.58	1.33	0.76-2.35	0.24
SMC women										
cases	193		57	56		39		41		
person years	750318		190477	188768		186804		184270		
RR (age, kjoules)	1.07	0.88-1.31	1.0	1.18	0.80-1.72	0.93	0.60-1.43	1.15	0.72-1.83	0.75
RR (multivariate[6])	1.06	0.87-1.29	1.0	1.17	0.80-1.71	0.92	0.60-1.41	1.12	0.70-1.79	0.84

1 ATBC = Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS = Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC = Swedish Mammography Cohort (Sweden); *2* The factor scores in the linear models are standardized variables with a mean of zero and a standard deviation of one; *3* includes age, ATBC treatment group, BMI, education, smoking (number of cigarettes/ day, years smoked), occupational activity, and energy intake; *4* for the subcohort only; *5* includes age, BMI, education, smoking (number of cigarettes/ day, years smoked), physical activity, family history of colorectal cancer, and energy intake; *6* includes age, BMI, education, and energy intake.

similar increased risk of colon cancer in this cohort. In ATBC men, the PPP pattern was associated with 121% increased risk of rectal cancer (Q4) with a significant test for trend. The PPP pattern was not associated with risk of colorectal, colon, or rectal cancers in NLCS men and women.

DISCUSSION

Findings from DIETSCAN revealed two consistent dietary patterns across the cohorts: one consisting mostly of vegetables, and one of pork, processed meats, and potatoes. A review of the literature on factor analysis of food consumption data shows these two patterns to be common across many different cultures despite different methods of data collection and analysis and different numbers and types of foods consumed.[18] In DIETSCAN, however, the associations of these two patterns with colorectal, colon, and rectal cancer were not as consistent. The VEG pattern was generally not associated with colorectal cancer in any cohort. The PPP pattern was associated with increased risk of colorectal and colon cancer in SMC women and with increased risk of rectal cancer in ATBC men. Our findings regarding the PPP pattern generally agree with those of other diet and cancer studies that have used factor analysis to characterize dietary patterns.

Table 4: Relative risks of colorectal cancer by linear and quartile models of the Pork, processed meat, potatoes (PPP) pattern in the Dietary Patterns and Cancer (DIETSCAN) Project cohort studies[1]

	Linear[2]		Q1	Q2		Q3		Q4		P value
	RR	95%CI	Ref	RR	95%CI	RR	95%CI	RR	95%CI	for Trend
Colorectal cancer										
ATBC men										
cases	322		70	84		91		71		
person years	286967		69852	71593		72836		72685		
RR (age, kjoules)	0.94	0.77-1.14	1.0	1.30	0.93-1.82	1.48	1.03-2.13	1.36	0.85-2.18	0.21
RR (multivariate[5])	0.97	0.79-1.18	1.0	1.34	0.96-1.88	1.59	1.10-2.29	1.49	0.93-2.39	0.10
NLCS men										
cases	660		172	169		177		142		
person years[4]	10496		2551	2579		2573		2793		
RR (age, kjoules)	0.99	0.88-1.12	1.0	1.02	0.78-1.34	1.13	0.85-1.50	0.90	0.65-1.24	0.67
RR (multivariate[5])	0.98	0.86-1.11	1.0	0.99	0.75-1.30	1.11	0.83-1.49	0.90	0.65-1.26	0.75
NLCS women										
cases	512		122	131		149		110		
person years	11328		2869	2803		2685		2972		
RR (age, kjoules)	0.98	0.89-1.09	1.0	1.08	0.81-1.45	1.35	1.02-1.80	0.92	0.68-1.25	0.98
RR (multivariate[c])	0.96	0.86-1.08	1.0	1.08	0.80-1.46	1.31	0.97-1.77	0.89	0.64-1.23	0.88
SMC women										
cases	586		193	150		143		100		
person years	749282		188421	182974		189625		188263		
RR (age, kjoules)	1.18	1.02-1.37	1.0	1.05	0.79-1.38	1.21	0.90-1.64	1.61	1.11-2.33	0.03
RR (multivariate[6])	1.18	1.02-1.37	1.0	1.04	0.83-1.29	1.26	0.99-1.61	1.37	1.00-1.89	0.03
Colon cancer										
ATBC men										
cases	191		43	51		56		41		
person years	287375		69943	71693		72931		72808		
RR (age, kjoules)	0.81	0.63-1.05	1.0	1.21	0.79-1.86	1.34	0.84-2.14	0.99	0.54-1.85	0.95
RR (multivariate[3])	0.85	0.66-1.10	1.0	1.26	0.82-1.94	1.47	0.92-2.35	1.12	0.60-2.09	0.64
NLCS men										
cases	400		112	98		105		85		
person years[4]	10509		2551	2579		2579		2800		
RR (age, kjoules)	0.97	0.84-1.13	1.0	0.93	0.68-1.29	1.09	0.78-1.52	0.91	0.62-1.33	0.82
RR (multivariate[5])	0.98	0.84-1.15	1.0	0.90	0.65-1.27	1.10	0.77-1.57	0.96	0.65-1.44	0.89
NLCS women										
cases	360		81	98		105		76		
person years	11334		2869	2803		2685		2978		
RR (age, kjoules)	0.99	0.88-1.11	1.0	1.22	0.87-1.70	1.43	1.03-1.99	0.98	0.68-1.39	0.85
RR (multivariate[c])	0.96	0.84-1.09	1.0	1.22	0.86-1.72	1.37	0.97-1.95	0.92	0.63-1.35	0.84
SMC women										
cases	396		118	98		93		87		
person years	749964		188729	183164		189780		188292		
RR (age, kjoules)	1.25	1.05-1.49	1.0	1.05	0.79-1.38	1.21	0.90-1.64	1.61	1.11-2.33	0.01
RR (multivariate[6])	1.25	1.05-1.49	1.0	1.05	0.80-1.39	1.22	0.90-1.65	1.62	1.12-2.34	0.01

Table 4 continued

	Linear[2]		Q1	Q2		Q3		Q4		P value
	RR	95%CI	Ref	RR	95%CI	RR	95%CI	RR	95%CI	for Trend
Rectal cancer										
ATBC men										
cases	133		27	34		35		37		
person years	287486		69962	71747		72973		72804		
RR (age, kjoules)	1.14	0.84-1.54	1.0	1.47	0.86-2.49	1.68	0.94-3.00	2.11	1.03-4.33	0.06
RR (multivariate[3])	1.15	0.85-1.57	1.0	1.50	0.88-2.55	1.75	0.98-3.15	2.21	1.07-4.57	0.05
NLCS men										
cases	260		60	71		72		57		
person years[4]	10525		2557	2588		2585		2795		
RR (age, kjoules)	1.02	0.86-1.20	1.0	1.18	0.80-1.74	1.22	0.82-1.83	0.89	0.56-1.42	0.66
RR (multivariate[5])	0.98	0.82-1.17	1.0	1.15	0.78-1.70	1.14	0.75-1.73	0.83	0.51-1.34	0.47
NLCS women										
cases	152		41	33		44		34		
person years	11355		2880	2808		2690		2977		
RR (age, kjoules)	0.98	0.83-1.15	1.0	0.81	0.49-1.32	1.21	0.76-1.92	0.82	0.50-1.35	0.76
RR (multivariate[c])	0.97	0.80-1.16	1.0	0.80	0.47-1.34	1.19	0.74-1.93	0.82	0.48-1.39	0.99
SMC women										
cases	193		76	52		52		13		
person years	750318		188765	183225		189862		188466		
RR (age, kjoules)	0.95	0.71-1.26	1.0	0.97	0.66-1.41	1.26	0.83-1.90	0.54	0.26-1.12	0.52
RR (multivariate[6])	0.95	0.72-1.27	1.0	0.98	0.67-1.44	1.28	0.85-1.93	0.56	0.27-1.17	0.60

1 ATBC = Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS = Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC = Swedish Mammography Cohort (Sweden); *2* The factor scores in the linear models are standardized variables with a mean of zero and a standard deviation of one; *3* includes age, ATBC treatment group, BMI, education, smoking (number of cigarettes/ day, years smoked), occupational activity, and energy intake; *4* for the subcohort only; *5* includes age, BMI, education, smoking (number of cigarettes/ day, years smoked), physical activity, family history of colorectal cancer, and energy intake; *6* includes age, BMI, education, and energy intake.

Slattery *et al*[5] identified a Western pattern of red meat, processed meat, fast food, refined grains, and sugar- containing foods that was associated with increased risk of colon cancer in both men (OR=1.80, 95% CI =1.28-2.53, p for trend<0.01) and women (OR=1.49, 95% CI=1.05-2.12, p for trend=0.02) in a multicenter U.S. population-based case-control study. Fung *et al*[7] identified a similar Western pattern of red and processed meats, French fries, refined grains, and sweets and desserts that was suggestive of increased risk for colorectal cancer (RR=1.46, 95% CI=0.97-2.19, p for trend=0.02) in a large prospective study of nurses in the U.S. Randall *et al*[4] identified a Traditional pattern of meats, common vegetables (e.g., green beans, corn), and baked goods that was associated with increased risk of colon cancer (OR=1.28, 95% CI=1.04-1.57) in a study of 205 men and 1,475 neighborhood controls in Western New York. In a previous analysis of SMC data using only 24 food groups compared with 42 food groups in DIETSCAN, Terry *et al*[6] identified a Western dietary pattern of processed and red meats, soda and sweets, refined breads and potatoes, and high-fat dairy products but this pattern was not associated with increased risk of colorectal cancer.

The previous studies also identified a pattern comprised of vegetables and other "healthful" foods that was generally associated with reduced risk of colorectal cancer. In

the U.S. population-based case-control study by Slattery *et al*[5], a Prudent pattern of vegetables and fruits was associated with reduced risk of colon cancer in both men (OR=0.66, 95% CI=0.50- 0.86, p for trend=0.02) and women (OR=0.73, 95% CI=0.55-0.97, p for trend=0.02). Fung *et al*[7] also identified a Prudent pattern of vegetables, fruits, legumes, fish, poultry, and whole grains that was inversely but not significantly associated with colon cancer (RR=0.71, 95% CI=0.50- 1.00, p for trend=0.31) in U.S. nurses. In the case-control study in Western New York, Randall *et al*[4] showed a pattern of Salad vegetables to be associated with reduced risk of colon cancer in women (OR=0.73, 95% CI=0.60-089) and to approach significance in men (OR=0.84, 95% CI=0.69-1.02). We hypothesized that the VEG pattern would be associated with reduced risk of colon and rectal cancers, but there was only slight evidence of this in NLCS women. It is possible that our findings may differ from those of the previous U.S. studies because of the types of foods included in the VEG factor. Our "healthful" pattern (VEG) included salad vegetables and fruits but also included high-fat foods like salad dressings and high-starch foods like pasta and rice. It is also possible that the dietary patterns common to different European cohorts may reflect differences across countries in consumption of those foods (**Appendix Table 1**) and may differ from amounts typically consumed in the U.S.
In addition, findings from analyses of dietary patterns may not always agree with findings from single foods or nutrients with cancer within the same study population. For example, results from previous analyses of individual nutrients and foods with colorectal cancer in ATBC and SMC tend to differ from our findings in the DIETSCAN analyses. In ATBC men, high calcium, milk protein, and milk products were inversely associated with risk of colorectal cancer but meat, vegetables, and fruits were not associated.[14] In contrast, our results show the PPP pattern, which reflects high meat intake and is more strongly associated with dietary calcium and milk products than the other patterns, to be associated with increased risk. One possible explanation for this difference is that the factor analysis did not produce an independent dietary factor representing milk products as the dominant foods in ATBC men. In SMC women, total vegetable and fruit intake was associated with reduced risk of colorectal cancer,[16] whereas our results and those of a previous factor analysis[6] did not show a protective effect of the VEG pattern. However, the VEG pattern included foods other than vegetables and fruits, reflecting more complex dietary behavior rather than individual foods. Given that it is unlikely that any single food causes cancer, the strength of factor analysis of complex dietary data is the ability to assess many combinations of foods rather than single food groups that are rarely eaten alone. Thus, results from such analysis may lead to new hypotheses for dietary factors and cancer.

Limitations of factor analysis include the following. Although this statistical method is data-driven and may be considered *objective* because it is conducted *a posteriori*, the analytical process is filled with *subjective* decisions such as the number and type of food groups determined from a FFQ, estimation of grams per day versus frequencies of food groups, transformation of food groups prior to principal components analysis, the criteria used to determine how many factors to extract (e.g., Eigenvalue cut-off, interpretation of the scree plot), the rotation employed, the interpretation of the resulting dietary patterns, and the use of factor or sum scores in analyses of cancer outcomes.[19] To address some of these limitations, we minimized the differences in FFQ and the detail and number of input variables in our factor analysis by aggregating the food items into a predefined, common food grouping. We also conducted sensitivity analyses with dichotomization of extremely skewed variables, number of factors extracted, and energy adjustment of the food group

variables by the residual method. Results from these sensitivity analyses showed the VEG and PPP patterns to be robust and the associations with colorectal cancer to be consistent in each cohort.[13] It is possible that our findings are due, in part, to measurement error in the dietary data and to differences among the cohorts. Recent research shows measurement error of FFQs to be larger than previously thought, which likely reduces the magnitude of risk estimates for the association of dietary patterns with cancer.[20;21] Also, although the simple age- and energy- adjusted models were uniform across the cohorts, the multivariate models reflect differences in sociodemographic and behavioral data available from each cohort. For example, SMC did not collect data on family history of colorectal cancer, physical activity, and smoking. Lack of data on other risk factors of colorectal cancer (e.g., aspirin use) limits analysis of interactions between the dietary patterns and these potential confounders across the cohorts. In addition, the participants in ATBC (i.e., male smokers) and SMC (i.e., women who responded to a mammography screening) may not be representative of the general population in their respective countries. Although DIETSCAN is not exempt from these limitations, it is important to emphasize that each study has the merits of being a large prospective cohort with validated semi- quantitative FFQ, pathologically confirmed cases of incident cancers, and extensive follow-up data. Given that the dietary patterns were qualitatively robust across the different cohorts, analysis of these patterns with other cancers (e.g., breast, lung) may be more informative.

In conclusion, dietary patterns represent combinations of foods, and may also represent biological interactions and non-nutrient substances that are difficult to assess. While such analysis has great potential to improve dietary guidance and public health and to create more successful dietary interventions, our opinion is that such analysis complements, rather than replaces, findings from more traditional analyses of single nutrients or food groups with cancer. Results from DIETSCAN show two dietary patterns to be consistent across three European countries but the association of those dietary patterns with risk of colon and rectal cancer is not conclusive.

REFERENCES

1Boyle P, Langman JS. ABC of colorectal cancer – epidemiology. Br Med J 2000;321:805-8.

2 Joint WHO/FAO Expert Report on Diet, Nutrition, and the Prevention of Chronic Disease. Geneva, Switzerland: World Health Organization, 2003.

3. Jacobs DR, Steffen LM. Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. Am J Clin Nutr 2003;78(suppl):508S-13S.

4 Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in Western New York. Nutr Cancer 1992;18:265-76.

5 .Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. Am J Epidemiol 1998;148:4-16.

6. Terry P, Hu FB, Hansen J, Wolk A. Prospective study of major dietary patterns and colorectal cancer risk in women. Am J Epidemiol 2001;154:1143-9.

7. Fung T, Hu FB, Fuchs C, et al. Major dietary patterns and the risk of colorectal cancer in women. Arch Intern Med 2003;163:309-14.

8. The ATBC Cancer Prevention Study Group. The Alpha-Tocopherol Beta-Carotene Lung Cancer Prevention Study: design, methods, participant characteristics, and compliance. Ann Epidemiol 1994;4:1-10.

9. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments, I: a self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol 1988;128:655-66.

10. van den Brandt PA, Goldbohm RA, van't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. J Clin Epidemiol 1990;43:285-95.

11. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994;48:253-65.

12. Wolk A, Bergstrom R, Hunter D, et al. A prospective study of association of monounsaturated fat and other types of fat with risk of breast cancer. Arch Intern Med 1998;158:41-5.

13. Balder HF, Brants HM, Virtanen M, et al. Common and country-specific dietary patterns in four European cohort studies. J Nutr 2003;133:4246-51.

14. Pietinen P, Malila N, Virtanen M, et al. Diet and risk of colorectal cancer in a cohort of Finnish men. Cancer Causes Control 1999;10:387-96.

15. Voorrips LE, Goldbohm RA, van Poppel G, Sturmans F, Hermus RJJ, van den Brandt PA. Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study. Am J Epidemiol 2000;152:1081-92.

16. Terry P, Giovannucci E, Michels KB, et al. Fruit, vegetables, dietary fiber, and risk of colorectal cancer. J Natl Cancer Inst 2001;93:525-33.

17. Willett WC, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124:17-27.

18. Balder HF, van den Brandt PA, Goldbohm RA. Factor analysis of food consumption data: a review of the literature. submitted.

19. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity. Am J Epidemiol 1998;148:17-19.

20. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: The OPEN Study. Am J Epidemiol 2003;158:1-13.

21. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. Am J Epidemiol 2001;153:394-403.

Appendix Table 1 Mean daily consumption (g/d) of 51 food groups used in the exploratory factor analysis. [1]

	ATBC men		NLCS men		NLCS women		SMC women	
Factors and representative food groups	load	intake[2]	load	intake[2]	load	intake[2]	load	intake[2]
Vegetables		*g/d*		*g/d*		*g/d*		*g/d*
Legumes	0.34	5 ± 5	0.11	40 ± 26	0.55	35 ± 23	-0.01	11 ± 15
Cabbages	0.61	14 ± 16	0.13	40 ± 25	0.59	39 ± 24	0.43	10 ± 18
Leaf vegetables, cooked	0.23	0.2 ± 1[3]	0.01	22 ± 18	0.48	22 ± 16	0.25	8 ± 16
Leaf vegetables, raw	0.69	1 ± 2	0.40	10 ± 10	0.52	10 ± 9	0.63	17 ± 16
Allium vegetables	-	-[4]	0.51	31 ± 27	0.59	31 ± 25	-	-[4]
Carrots	0.51	14 ± 17	0.20	12 ± 14	0.50	13 ± 16	0.59	24 ± 26
Tomatoes	0.67	24 ± 22	0.45	20 ± 21	0.45	24 ± 21	0.56	27 ± 26
Mushrooms	0.31	3 ± 5	0.52	3 ± 4	0.39	4 ± 4	-	-[4]
Citrus fruit	0.41	59 ± 79	0.21	51 ± 61	0.18	73 ± 67	0.50	52 ± 57
Apples and pears	0.27	30 ± 42	-0.01	67 ± 75	0.17	84 ± 82	0.51	67 ± 67
Berries, all types	0.35	36 ± 36	0.14	7 ± 8	0.12	8 ± 9	-	-[4]
Rice	0.48	6 ± 7	0.43	12 ± 30	0.35	9 ± 20	0.18	2 ± 6
Pasta	0.33	8 ± 12	0.53	7 ± 9	0.36	5 ± 8	0.02	15 ± 14
Poultry, fresh	0.41	12 ± 15	0.30	13 ± 15	0.24	14 ± 16	0.22	7 ± 7
Fish, incl shellfish and crustaceans	0.26	39 ± 30	0.29	14 ± 16	0.33	12 ± 14	0.41	25 ± 17
Oil	0.27	2 ± 3	0.58	2 ± 3	0.41	2 ± 3	-	-[4]
Dressings and other similar sauces	0.77	2 ± 3	0.07	4 ± 6	0.16	5 ± 6	-	-[4]
Wine and fortified wine	-0.01	30 ± 42	0.40	33 ± 70	0.15	35 ± 64	0.09	15 ± 21
Pork, processed meats, potatoes								
Pork, fresh	0.54	41± 21	0.48	45 ± 33	0.43	39 ± 31	0.57	20 ± 17
Processed meats	0.42	75 ± 59	0.53	22 ± 19	0.50	15 ± 14	0.42	25 ± 17
Potatoes and potato products	0.55	180 ± 79	0.38	150 ± 85	0.29	101 ± 60	0.15	95 ± 52
Coffee	0.23	607 ± 351	0.50	574 ± 292	0.44	495 ± 242	0.08	427 ± 202
Beef and veal	0.32	26 ± 21	-0.11	44 ± 29	-0.07	37 ± 26	0.60	26 ± 18
Pasta	0.10	8 ± 12	0.20	7 ± 9	-0.05	5 ± 8	0.55	15 ± 14
Rice	0.04	6 ± 7	-0.18	12 ± 30	-0.31	9 ± 20	0.47	2 ± 6
Eggs	0.46	53 ± 38	0.19	17 ± 13	0.14	15 ± 10	0.18	7 ± 8
Poultry	0.08	12 ± 15	-0.01	13 ± 15	0.02	14 ± 16	0.40	7 ± 7
Liver	0.19	3 ± 4	0.10	2 ± 5	0.28	2 ± 4	0.40	4 ± 5
Butter	0.46	39 ± 30	-0.30	9 ± 16	-0.45	7 ± 12	0.02	4 ± 8
Low-fat margarine	-	-[4]	0.46	8 ± 16[3]	0.50	7 ± 13	-[4]	-[4]
Wine	0.49	30 ± 42	-0.09	33 ± 70	-0.09	35 ± 64	0.00	15 ± 21
Spirits	0.49	31 ± 43	-0.03	25 ± 40	0.07	4 ± 16[3]	0.01	1 ± 2

[1] Food groups that are listed for each pattern have factor loads ≥ 0.35 in at least one of the DIETSCAN cohorts. [2] mean ± SD. [3] Foodgroups with >75% of participants reporting no intake. [4] Food groups missing from the food frequency questionnaire and thus not entered into the principal components analysis.

6

Dietary patterns associated with male lung cancer risk in the Netherlands Cohort Study

Balder HF, Goldbohm RA, Van den Brandt PA

Accepted Cancer Epidemiology, Biomarkers & Prevention

ABSTRACT

The objective was to study the association between dietary patterns and lung cancer incidence in the Netherlands Cohort Study on Diet and Cancer (NLCS). The baseline measurement of this prospective case cohort study that was completed by 58,279 men in 1986 included a self-administered questionnaire on dietary intake, smoking habits and other covariates. Follow-up was established by computerized record linkage to cancer registries and a pathology register. After 9.3 years of follow-up 1426 confirmed cases of incident male lung cancer were detected. Five dietary patterns were identified by exploratory factor analysis in a randomly sampled subcohort (n=2190).

The dietary pattern labeled "salad vegetables" was associated with decreased risk of lung cancer (RR_{Q5}: 0.71, 95%CI: 0.53-0.95), after multivariate adjustment. This inverse association was most evident among current and former smokers. A dietary pattern labeled "sweet foods" was also inversely associated with lung cancer risk (RR_{Q5}: 0.60, 95%CI: 0.42-0.85). However, the higher intake of mono- and disaccharides and fruits and lower consumption of alcohol associated with this pattern could not account for its full protective effect. The "pork, processed meat, potatoes" pattern was non-significantly associated with increased risk (RR_{Q5}: 1.40, 95%CI: 0.97-2.01) and this positive association was most evident among current smokers. The other dietary patterns characterized by brown/white bread substitution and by consumption of cooked vegetables were not associated with lung cancer risk. These results demonstrate how studying both single factors and dietary patterns gives more insight into the complex, and often seemingly inconsistent, associations between diet and cancer.

INTRODUCTION

Lung cancer is among the most frequently occurring cancers in developed countries, particularly in men. Smoking is by far the most important cause of lung cancer, but lung cancer risk may also be influenced by diet. The most consistent associations with lung cancer have been shown with the consumption of fruits and vegetables (inverse)[1;2] and total and saturated fat (positive).[3]

Traditional analyses in nutritional epidemiology have handled nutritional data either as isolated items or a small number of items simultaneously. Relatively few studies have addressed the broader eating patterns that reflect many dietary exposures together. These dietary patterns may better explain disease occurrence than individual dietary exposures, in particular when not just one dietary risk factor is responsible. One method of identifying and examining broader dietary patterns that may be associated with disease is factor analysis. To date, this approach has been used before in studies of colon or colorectal cancer,[4-8] stomach cancer,[9;10] breast cancer,[11] endometrial cancer,[12] renal cell cancer,[13] and thyroid cancer.[14]

The only study relating lung cancer to dietary patterns used cluster analysis to identify two groups of individuals with distinct dietary patterns among 254 lung cancer cases and 184 healthy controls.[15] A dietary pattern labeled "healthy", characterized by high intake of fiber and carbohydrate and low intake of protein and animal fat, had a protective effect against lung cancer, however, after adjustment for smoking the association was no longer statistically significant (OR: 0.93, 95%CI: 0.59-1.44).

To our knowledge, factor analysis has not previously been applied to dietary associations with lung cancer. As part of the DIETSCAN project, a collaboration with large prospective cohort studies from Sweden, Finland and Italy, we recently identified dietary patterns using factor analysis.[16] This gave us the opportunity to study prospectively the association between detailed dietary patterns and lung cancer incidence in the Netherlands Cohort Study on Diet and Cancer (NLCS). Because of the relatively small number of lung cancer cases among the mainly non-smoking women, only the results for men are presented. In addition, and to compare these results, where appropriate, with food based and nutrients based analyses. With 9.3 years of follow-up and more than a thousand incident lung cancer cases, the analyses could be performed also for separate strata of smoking status and tumor histology.

SUBJECTS AND METHODS

Cohort The Netherlands Cohort Study on diet and cancer (NLCS) was approved by the institutional review boards of the TNO Nutrition and Food Research (Zeist) and Maastricht University (Maastricht). The NLCS started in September 1986 when men and women aged 55 to 69 years from 204 municipalities were enrolled in the cohort using computerized population registries. The case-cohort study design has been reported in detail elsewhere.[17] In total 58,279 men completed a self-administered mailed questionnaire on habitual dietary intake, smoking, lifestyle characteristics, medical history and other potential risk factors for cancer. The dietary part of this questionnaire consisted of a 150-item semi-quantitative food frequency questionnaire on the usual intake of food and beverages in the year preceding the start of the study, which was validated against a 9-day diet record.[18] Questionnaire data of all cases and subcohort members have been key-

Table 1 Factor loadings and mean daily consumption (g/d) of 49 food groups used in the exploratory factor analysis within the first and fifth quintile of dietary pattern scores

Food groups	"Salad Vegetables" pattern (SV)			"Cooked Vegetables" pattern (CV)			"Pork, Processed meat, Potatoes" pattern (PPP)		
	load	Q1	Q5	load	Q1	Q5	load	Q1	Q5
Legumes	0.11	36	46	**0.58**	**22**	**63**	0.08	36	44
Cabbages	0.13	36	44	**0.73**	**21**	**67**	0.12	36	45
Leaf vegetables – cooked	0.01	22	22	**0.70**	**9.5**	**41**	-0.03	22	22
Leaf vegetables – raw	**0.40**	**5.2**	**16**	0.13	8.5	13	0.00	10	11
Allium vegetables	**0.51**	**14**	**53**	**0.37**	**20**	**47**	0.01	31	33
Carrots	0.20	9.3	17	**0.46**	**5.9**	**21**	-0.06	15	11
Tomatoes	**0.45**	**9.1**	**34**	0.26	15	28	0.02	22	21
Mushrooms	**0.52**	**1.1**	**6.9**	0.08	3.0	4.2	-0.04	3.7	3.1
Citrus fruit	0.21	35	67	0.18	37	64	-0.30	85	31
Bananas	0.14	8.4	17	0.05	11	15	0.07	12	15
Apples and pears	-0.01	70	68	0.08	54	74	-0.09	82	60
Berries – all types	0.14	5.0	8.4	0.06	5.4	7.5	-0.05	7.6	6.3
Potatoes and potato products	-0.18	167	128	0.31	116	192	**0.38**	**111**	**199**
Rice	**0.43**	**3.3**	**33**	-0.04	16	14	-0.18	24	9.2
Pasta	**0.53**	**2.3**	**14**	-0.10	8.8	6.1	0.20	4.8	10
Dry cereals	0.15	2.2	5.4	-0.04	4.0	3.3	-0.28	7.7	1.3
Bread, crackers etc. –white types	-0.16	50	24	0.06	31	41	0.04	37	41
Bread, crackers etc. – brown/wholemeal types	-0.01	146	145	-0.03	151	144	0.33	113	189
Fermented whole milk and milk products	-0.01	17	15	-0.01	15	16	-0.06	21	14
Fermented medium, lowfat and skimmed milk prod	0.04	60	72	0.13	55	87	-0.21	102	39
Nonfermented whole milk and milk products	-0.24	180	79	0.07	105	136	0.15	95	151
Nonfermented medium, lowfat and skimmed milk prod	0.04	92	112	-0.08	125	92	-0.03	108	99
Full cream cheese and cheese spreads	-0.01	22	24	0.02	21	23	0.08	21	25
Lowfat cheese and cheese spreads	-0.03	2.1	1.4	0.09	0.73	2.4	-0.02	2.1	1.7
Beef and veal – fresh	-0.09	49	42	0.30	33	56	-0.11	50	37
Pork – fresh	0.01	42	43	0.06	42	48	**0.48**	**23**	**65**
Liver – fresh	0.15	1.1	3.6	0.13	1.7	3.6	0.10	1.8	3.4
Poultry – fresh	0.30	6.8	21	0.01	13	14	-0.01	14	13
Processed meat	0.02	21	23	0.02	22	24	**0.53**	**11**	**39**
Fish (shellfish and crustacean included)	0.29	6.4	21	0.14	11	18	-0.06	15	14
Eggs	0.11	14	19	0.07	16	18	0.19	15	21
Butter	0.07	7.8	10	-0.01	8.6	9.6	-0.30	18	3.3
Margarine	-0.30	41	20	0.25	19	37	0.25	17	37
Lowfat margarine	0.06	5.4	8.3	-0.12	13	5.8	**0.46**	**1.4**	**19**
Oil	**0.58**	**0.6**	**5.0**	0.02	2.3	2.6	0.06	2.3	2.6
Dressings and other similar sauces	0.07	3.3	4.9	-0.00	4.2	4.7	0.22	2.7	6.3
Peanuts and other nuts	0.29	4.9	16	-0.11	12	7.8	0.06	8.2	11
Savory snacks	0.32	0.4	3.3	-0.22	2.6	0.7	0.06	1.1	1.6
Cakes, sweet breads and cookies	-0.06	27	24	-0.04	27	24	-0.01	26	25
Candies	0.05	4.9	5.9	-0.01	5.5	4.9	-0.04	5.7	4.3
Added sugar	-0.07	33	25	-0.03	33	30	0.30	20	45
Sweet sandwich spread	-0.11	12	7.8	-0.02	11	8.9	0.04	9.0	9.7
Coffee	-0.02	574	559	0.04	587	589	**0.50**	**397**	**796**
Tea	-0.06	361	297	0.13	261	361	-0.27	424	246
Fruit juices	0.10	15	30	0.07	19	27	-0.02	23	16
Soft drinks and syrups	0.07	32	49	-0.04	50	40	0.17	25	64
Beer	0.13	48	127	0.02	95	110	0.21	39	164
Wine and fortified wine	**0.40**	**7.8**	**87**	-0.06	48	31	-0.09	46	23
Spirits	0.16	14	34	0.09	19	30	-0.03	26	23

SV="(Salad) Vegetables pattern"; CV="Cooked vegetables pattern"; PPP="Pork, processed meat, potatoes pattern"; SW="Sweet foods patterns"; B/W="Brown/white bread substitution pattern"; FD="Fat dairy pattern"; SN="Sweet and/or savory snacks pattern"
For convenience factor loadings **>0.35** are printed **bold**.

Table 1 continued

Food groups	"Sweet" pattern (SW)			"Brown/White bread substitution" pattern (B/W)		
	load	Q1	Q5	load	Q1	Q5
Legumes	0.04	41	42	-0.02	42	41
Cabbages	-0.05	42	40	0.04	38	42
Leaf vegetables – cooked	0.01	23	23	-0.01	23	22
Leaf vegetables – raw	0.28	7.3	14	0.14	8.1	13
Allium vegetables	-0.14	40	28	0.08	28	36
Carrots	0.26	8.4	17	0.23	9.1	17
Tomatoes	0.00	21	21	0.14	16	25
Mushrooms	-0.11	4.4	3.0	0.06	2.8	3.6
Citrus fruit	0.13	39	60	0.10	42	64
Bananas	0.25	7.2	21	-0.04	14	10
Apples and pears	0.26	40	96	**0.37**	**40**	**115**
Berries – all types	**0.36**	**3.5**	**11**	-0.03	7.5	6.5
Potatoes and potato products	0.25	123	180	0.02	152	157
Rice	0.09	9.4	17	0.02	13	12
Pasta	0.06	6.3	7.9	-0.04	6.7	6.3
Dry cereals	0.19	1.6	5.5	0.13	2.1	5.3
Bread, crackers etc. –white types	0.13	29	48	**-0.62**	**108**	**11**
Bread, crackers etc. – brown/wholemeal types	0.18	117	164	**0.67**	**54**	**221**
Fermented whole milk and milk products	0.28	3.6	35	-0.01	14	13
Fermented medium, lowfat and skimmed milk prod	0.02	63	68	0.33	31	139
Nonfermented whole milk and milk products	0.29	66	195	-0.29	191	61
Nonfermented medium, lowfat and skimmed milk prod	-0.06	112	100	0.19	57	145
Full cream cheese and cheese spreads	0.23	15	27	0.22	18	29
Lowfat cheese and cheese spreads	-0.00	1.6	1.9	0.18	0.7	3.8
Beef and veal – fresh	-0.00	45	44	-0.07	43	41
Pork – fresh	-0.18	55	37	-0.17	52	36
Liver – fresh	-0.04	2.9	2.3	-0.10	3.2	1.4
Poultry – fresh	-0.04	15	13	-0.01	13	12
Processed meat	0.07	20	24	-0.08	25	21
Fish (shellfish and crustacean included)	-0.11	17	12	-0.11	17	12
Eggs	-0.04	19	17	-0.32	24	13
Butter	0.14	6.5	12	-0.32	16	3.2
Margarine	0.12	25	34	-0.01	29	29
Lowfat margarine	0.08	5.7	8.9	0.28	4.1	17
Oil	0.05	2.3	2.5	0.01	2.2	2.4
Dressings and other similar sauces	0.21	2.9	6.1	-0.05	4.7	4.5
Peanuts and other nuts	0.20	5.3	13	-0.13	12	7.0
Savory snacks	0.13	0.7	2.2	-0.12	1.9	0.8
Cakes, sweet breads and cookies	**0.53**	**12**	**39**	0.02	25	25
Candies	**0.38**	**1.7**	**9.6**	-0.13	6.1	3.7
Added sugar	0.25	19	42	-0.34	48	15
Sweet sandwich spread	**0.45**	**3.2**	**19**	0.01	9.4	11
Coffee	-0.13	636	525	-0.02	598	555
Tea	0.31	199	404	0.01	320	333
Fruit juices	0.12	16	34	-0.07	24	17
Soft drinks and syrups	0.17	28	61	-0.21	63	22
Beer	-0.34	218	35	-0.10	120	51
Wine and fortified wine	-0.02	35	33	0.03	26	37
Spirits	-0.29	47	10	-0.19	34	13

entered twice and blinded with respect to case/subcohort status to avoid random and systematic coding errors. Mean daily nutrient intakes were calculated using the computerized Dutch food composition table. After baseline exposure measurement, a subcohort of 2335 male subjects was randomly sampled from the large cohort. Following the case-cohort approach, this subcohort was followed for migration and vital status to calculate person time at risk. The entire cohort has been followed-up for incidence of cancer. The method of record linkage to obtain information on cancer incidence in the entire cohort has been described previously.[19] In short, follow-up for incident cancer has been established by record linkage to the Netherlands Cancer Registry and the Netherlands Pathology Registry (PALGA). After 9.3 years of follow-up, i.e. from September 1986 to December 1995, 1583 male lung cancer cases were identified. Subjects with incomplete or inconsistent dietary information were excluded from the analyses according to criteria described elsewhere.[18] After excluding subjects who reported prevalent cancer other than skin cancer, and subjects with incident in situ lung cancer other than carcinoma (sarcoma, lymphoma, unspecified morphology), or without at least a microscopically confirmed diagnosis, 1426 male incident primary lung carcinoma cases (ICD-0 codes T162.2-T162.9) were available for analysis. From the subcohort, prevalent cancer cases other than nonmelanoma skin cancer were excluded as well, leaving 2190 men for analysis.

Assessment of dietary patterns As part of the DIETSCAN project, the food items from the FFQ were aggregated into 51 food groups based on their role in the diet and possible relevance to cancer etiology.[16] To identify dietary patterns exploratory factor analysis using principal components analysis was conducted in the subcohort, using the FFQ-derived food groups (PROC FACTOR in SAS version 8; SAS Institute Inc., Cary, North Carolina). Factors were rotated by orthogonal Varimax transformation. Eigenvalues > 1 and the Scree test, traditional criteria in factor analysis, were used to determine the number of factors to be extracted. Labels of the dietary patterns were determined according to the dominant foods (i.e., foods with high factor loadings, representing correlation coefficients between food groups and dietary patterns). Further detail regarding the food grouping, dietary pattern assessment, and related sensitivity analyses conducted in DIETSCAN has been described elsewhere.[16]

Statistical analyses Cox proportional hazards models were constructed to estimate hazard ratios and 95% confidence intervals relating the dietary patterns to the incidence of lung cancer (Stata version 8; Stata Corporation, College Station, Texas). The proportional hazards assumption was tested using Schoenfeld residuals. Because of the case-cohort design, the 95% confidence intervals were corrected for the additional variance introduced by using a randomly sampled subcohort instead of the complete cohort, by using the robust option.

Prior to etiological analysis, factor scores for both lung cancer cases and members of the subcohort were determined by summing the standardized intakes from each food group, weighted by the factor loadings (PROC SCORE in SAS). Thus each participant had a unique score for each factor. High scores represented high intake of foods loading on the corresponding dietary pattern; low scores represented low intake of those foods. Because the factor scores represent standardized variables, each score had a mean of zero and a standard deviation of one. Rate ratio estimates for the dietary patterns were calculated for quintiles based on the distribution in the subcohort and for continuous variables with an

increment of 1 SD. Tests for trend in the rate ratios were assessed by fitting ordinal exposure categories as continuous variables. The models were adjusted for potential confounding variables determined by previous analyses conducted within the NLCS[1] and additionally for health related lifestyle factors that we expected to be associated with one or more dietary patterns: age at baseline (y), total energy intake (kJ), current cigarette smoking (yes/no), number of cigarettes smoked per day, number of years of smoking cigarettes, highest attained education (primary school; lower vocational school; junior high school; senior high school, higher vocational school; university), physical activity outside of the profession (<30 min/day, 30-60 min/day, 60-90 min/day, >90 min/day) and family history of lung cancer in first or second degree relatives (yes/no). The models included all factor scores simultaneously because dietary patterns are conditional on each other. Subgroup analyses were performed after stratification into never smokers, former smokers and current smokers, and into Kreyberg group I tumors (squamous cell, large cell, and small cell carcinoma) and Kreyberg group II tumors (adenocarcinoma). To compare the dietary pattern approach with traditional food and nutrient based analyses, Cox regression models were constructed for food groups or nutrients that corresponded closely to several of the dietary patterns.

Table 2 Non-dietary baseline characteristics among male lung carcinoma cases and subcohort members with complete dietary data

	subcohort		cases	
Characteristics	**n**	**mean or %**	**n**	**mean or %**
Age (y)	2190	61.3	1426	62.6
Family history of lung cancer	2190	10 %	1426	12 %
BMI (kg/m^2)	2115	24.9	1380	24.7
Total energy intake (MJ)	2190	9.1	1426	9.0
Smoking habits				
never smoker	277	13 %	52	3.7%
former smoker	1128	52 %	488	34 %
current smoker	785	36%	886	62 %
Number of cigarettes/day	2065	14.8	1282	19.1
Duration of smoking cigarettes (y)	2156	29.4	1398	39.6
Highest level of education				
Primary school	494	25 %	388	32 %
Lower vocational	414	21 %	305	25 %
High school	398	35 %	392	32 %
Higher vocational/university	356	18 %	141	11 %
Physical activity outside profession				
< 30 min/day	398	18 %	319	23 %
30-60 min/day	669	31 %	408	29 %
60-90 min/day	406	19 %	259	18 %
>90 min/day	691	32 %	423	30 %

RESULTS

Exploratory factor analysis identified five stable dietary patterns that explained 23% of the total variance in the dietary input variables.[16] In **table 1** the factor loadings of all food groups and their mean daily consumption within the first and the fifth quintile of dietary pattern scores are presented. The first dietary pattern was one we labeled "(Salad) Vegetables". This pattern was characterized by high factor loadings on several vegetable items, several fruit items, pasta, rice, poultry, fish, and oil and explained 5.6% of total variance. A separate "Cooked vegetables" pattern was also identified, with high loadings on cooked leaf vegetables, cabbages, legumes and carrots, describing 4.8% of total variance. A pattern labeled "Pork, processed meat, potatoes" was also identified. This pattern also loaded positively on coffee and low-fat margarine and accounted for 4.3% of the variance in dietary intake. A pattern with high loads on "Sweet foods", such as cakes and cookies, sweet sandwich spread, sweets and candies and (straw)berries accounted for 4.3% of total variance. In addition a "White/brown bread substitution" pattern was identified, which correlated positively with brown/wholemeal bread types and apples and pears and negatively with white types of bread, it explained 4.1% of total variance.
A description of the 1426 male lung cancer cases and 2190 male subcohort members with complete dietary data is presented in **Table 2**. On average, cases were older, were more likely to smoke cigarettes and had a lower educational level than members of the subcohort.

Table 3 Pearson correlation coefficients of potential covariates with dietary patterns[a] among male subcohort member of the Netherlands Cohort Study

Characteristics	n	SV[a]	CV[a]	PPP[a]	SW[a]	B/W[a]
Age (y)	2190	-0.12	0.06	-0.20	-0.00	0.02
Family history of lung cancer	2190	0.03	0.02	-0.01	-0.02	-0.00
BMI (kg/m^2)	2115	0.02	0.08	0.07	-0.10	-0.03
Total energy intake (MJ)	2190	0.13	0.21	0.50	0.50	-0.15
Current smoking status	2190	-0.03	0.02	0.19	-0.14	-0.21
Number of cigarettes/day	2065	0.07	0.05	0.12	-0.20	-0.12
Years of smoking cigarettes	2156	-0.07	0.01	0.16	-0.19	-0.14
Higher vocational/university education	1973	0.21	-0.09	-0.14	0.07	0.02
Physical activity (min/day)	2164	-0.01	0.08	0.06	0.05	0.03

[a] SV="Salad vegetables" pattern; CV="Cooked vegetables" pattern; PPP="Pork, processed meat, potatoes" pattern; SW="Sweet foods" pattern; B/W="Brown/white bread substitution" pattern.

The association between potential covariates and dietary patterns was studied in the subcohort by calculating Pearson correlation coefficients (**Table 3**). Men with a higher score on the salad vegetables pattern tended to be younger and higher educated. Subjects with higher pork, processed meat, potatoes pattern scores were younger and less educated, were more likely to smoke and tended to smoke more cigarettes per day and smoke for more years. Men with higher scores on the sweet pattern had lower BMI and were less likely to be current smokers and tended to smoke fewer cigarettes per day and smoke for fewer years. Consumption of brown/wholemeal types of bread instead of

Table 4 Rate ratios (RRs) and 95 percent confidence intervals (95% CI) of lung cancer according to dietary patterns among men in the Netherlands Cohort Study (9.3 years of Follow-up).

	Quintile of pattern score					P[b] trend	Continuous, linear (increment=1 SD)
Dietary pattern	Q1[a]	Q2	Q3	Q4	Q5		
"Salad vegetables"							
cases of lung cancer	329	336	300	245	216		1426
person years	3705	3736	3768	3739	3817		18766
RR (age, kjoules)	1	1.06	0.95	0.80	0.72	0.000	0.86
95% CI	ref.	0.86-1.31	0.77-1.17	0.65-1.00	0.57-0.90		0.80-0.93
RR (multivariate[c])	1	1.07	1.02	0.75	0.75	0.008	0.87
95% CI	ref.	0.81-1.40	0.77-1.35	0.56-1.01	0.55-1.01		0.78-0.96
"Cooked vegetables"							
cases of lung cancer	251	296	327	280	272		1426
person years	3736	3763	3767	3810	3688		18766
RR (age, kjoules)	1	1.12	1.28	1.05	1.02	0.88	1.00
95% CI	ref.	0.90-1.40	1.03-1.59	0.84-1.31	0.81-1.28		0.93-1.07
RR (multivariate[c])	1	1.05	1.10	0.90	0.86	0.18	0.95
95% CI	ref.	0.78-1.41	0.83-1.47	0.67-1.22	0.63-1.16		0.86-1.05
"Pork, processed meat, potatoes"							
cases of lung cancer	192	269	316	303	346		1426
person years	3690	3736	3759	3779	3802		18766
RR (age, kjoules)	1	1.52	1.90	1.94	2.67	0.000	1.38
95% CI	ref.	1.20-1.92	1.50-2.40	1.53-2.47	2.06-3.47		1.27-1.50
RR (multivariate[c])	1	1.18	1.32	1.24	1.44	0.08	1.10
95% CI	ref.	0.87-1.61	0.96-1.80	0.90-1.71	0.99-2.09		0.98-1.25
"Sweet foods"							
cases of lung cancer	437	326	268	209	186		1426
person years	3683	3707	3706	3786	3883		18766
RR (age, kjoules)	1	0.68	0.54	0.37	0.31	0.000	0.64
95% CI	ref.	0.55-0.83	0.43-0.66	0.30-0.47	0.24-0.40		0.59-0.70
RR (multivariate[c])	1	0.73	0.70	0.57	0.62	0.002	0.80
95% CI	ref.	0.56-0.95	0.52-0.93	0.41-0.78	0.43-0.89		0.70-0.90
"Brown/white bread substitution"							
cases of lung cancer	340	332	265	282	207		1426
person years	3760	3681	3786	3770	3768		18766
RR (age, kjoules)	1	0.95	0.76	0.76	0.59	0.000	0.85
95% CI	ref.	0.77-1.17	0.61-0.94	0.61-0.95	0.47-0.74		0.80-0.91
RR (multivariate[c,e])	1	1.14	0.92	0.90	0.89	0.18	0.99
95% CI	ref.	0.87-1.50	0.69-1.23	0.67-1.22	0.65-1.20		0.89-1.09

[a] Q, quintile
[b] All P-values are from two-sided tests
[c] adjusted for all other dietary patterns and age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked/day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (<30;30-60;60-90;>90 min/day).;
[d]The hazard ratio of the brown/white bread substitution pattern and education, history of lung cancer in family and physical activity appeared not to be constant over time. When this was adjusted for in the model (using time-dependent variables), the hazard ratio was exp(-0.2027 + 0.0454*t).

white types of bread was negatively associated with current smoker status, number of cigarettes smoked per day and duration of smoking. None of the dietary patterns were associated with family history of lung cancer.

Rate ratios (RR) of lung cancer and their 95% confidence intervals (95%CI) according to quintiles of scores on the dietary patterns are provided in **Table 4**. After adjustment for potential confounders and the other identified dietary patterns, the salad vegetables pattern was associated with decreased risk of lung cancer. (RR_{Q5}: 0.75 95%CI: 0.55-1.01, P for trend: 0.008). High consumption of the pork, processed meat, potatoes pattern appeared to be associated with an increased risk of lung cancer. However, after adjustment for smoking, the other potential confounders and the other four dietary

patterns, the positive association was less strong and no longer statistically significant. After adjustment for all potential confounders, the Sweet pattern was associated with decreased risk of lung cancer (RR_{Q5}: 0.62, 95%CI: 0.43-0.89, P for trend: 0.002). The pattern characterized by brown/white bread appeared to be associated with a decreased risk of lung cancer. However, after adjustment for smoking, the inverse association was less strong and no longer statistically significant. However, the hazard ratio of this pattern appeared not to be constant over time, and neither were the hazard ratios for education, family history of lung cancer and physical activity. When this was adjusted for in the model (using time-dependent variables), the hazard ratio for brown/white bread substitution pattern increased from 0.87 after 1.3 years of follow-up to 1.25 after 9.3 years of follow-up. The pattern characterized by consumption of cooked vegetables was not associated with lung cancer risk.

Table 5 Relative rate (RR) and 95% confidence interval (9%%CI) of lung cancer with an increment in 1 SD in factor score for dietary patterns, stratified by smoking status and histologic type of cancer.

Subgroup	number of cases	person time	SV[a]	CV[a]	PPP[a]	SW[a]	B/W[a]
Total[b]	1087	15536	0.89	0.97	1.17	0.81	1.00
			0.81-0.97	0.88-1.06	1.05-1.30	0.72-0.90	0.91-1.09
Smoking status							
never smoker[c]	52	2452	1.21	0.84	1.20	0.51	0.83
			0.94-1.56	0.63-1.12	0.85-1.67	0.33-0.77	0.59-1.17
former smoker[d]	446	9233	0.77	1.08	1.04	0.88	0.96
			0.65-0.90	0.95-1.24	0.88-1.22	0.75-1.04	0.84-1.10
current smoker[d]	771	5852	0.90	0.97	1.21	0.79	1.04
			0.80-1.01	0.87-1.08	1.06-1.39	0.70-0.91	0.93-1.16
Histologic type of cancer							
Kreyberg I[e]	915	17454	0.85	0.98	1.25	0.77	0.98
			0.77-0.94	0.89-1.07	1.12-1.39	0.69-0.86	0.89-1.07
Kreyberg II[e]	254	17381	0.95	0.97	0.98	0.83	1.00
			0.83-1.10	0.86-1.10	0.83-1.15	0.71-0.96	0.88-1.15

[a] SV="Salad vegetables" pattern; CV="Cooked vegetables" pattern; PPP="Pork, processed meat, potatoes" pattern; SW="Sweet foods" pattern; B/W="Brown/white bread substitution" pattern.
[b] adjusted for age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked/day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderate active, active, very active)
[c] adjusted for age, energy
[d] adjusted for age, energy, number of cigarettes smoked, years of smoking
[e] adjusted for age, energy, current smoking, number of cigarettes, years of smoking

In **Table 5**, the relative rates for dietary patterns stratified according to smoking status are presented. The salad vegetables pattern appeared to be associated with decreased risk of lung cancer in former smokers ($RR_{per\ 1\ SD}$: 0.77, 95%CI: 0.65-0.90; P for interaction: 0.02), and marginally significant in current smokers ($RR_{per\ 1\ SD}$: 0.90, 95%CI: 0.80-1.01; P for interaction: 0.22). The increased risk of lung cancer of the dietary pattern characterized by consumption of pork, processed meat, potatoes pattern was associated with risk of lung cancer among current smokers ($RR_{per\ 1\ SD}$: 1.21, 95%CI: 1.06-1.39; P for interaction 0.36). In former smokers, the pork, processed meat, potatoes pattern was not associated with risk of lung cancer (P for interaction: 0.66), whereas in never smokers the association was not statistically significant ($RR_{per\ 1\ SD}$: 1.20, 95%CI: 0.85-1.67). The sweet pattern was most strongly associated with decreased risk of lung cancer in never smokers ($RR_{per\ 1\ SD}$: 0.51, 95%CI 0.33-0.77) and current smokers ($RR_{per\ 1\ SD}$: 0.79, 95%CI: 0.70-0.91; P for interaction: 0.05). In former smokers the inverse association was not statistically significant (P for interaction: 0.01). The cooked vegetables pattern and the brown/white bread

Table 6 Rate ratios (RRs) and 95 percent confidence intervals (95% CI) of lung cancer according to quintiles of food groups and nutrients that corresponded closely to several of the dietary patterns (Netherlands Cohort Study, 9.3 years of Follow-up)

Food group or nutrient	Quintile of consumption					P[b] trend
	Q1[a]	Q2	Q3	Q4	Q5	
"Salad vegetables" and "Cooked vegetables" patterns						
Total vegetables						
median intake (g/d)	102	144	177	218	294	
cases of lung cancer	339	293	304	254	236	
person years	3684	3760	3786	3812	3724	
RR (multivariate[c])	1	0.97	0.99	0.92	0.66	
95% CI	ref.	0.74-1.26	0.76-1.30	0.70-1.21	0.50-0.87	0.008
Raw vegetables						
median intake (g/d)	6	19	31	46	73	
cases of lung cancer	374	279	290	260	223	
person years	3751	3706	3727	3824	3758	
RR (multivariate[c])	1	0.82	0.97	0.84	0.74	
95% CI	ref.	0.63-1.07	0.75-1.26	0.64-1.10	0.57-0.98	0.071
Cooked vegetables						
median intake (g/d)	79	115	143	177	239	
cases of lung cancer	322	288	280	282	254	
person years	3745	3700	3790	3793	3737	
RR (multivariate[c])	1	0.89	0.91	0.94	0.73	
95% CI	ref.	0.68-1.16	0.69-1.20	0.72-1.23	0.55-0.96	0.074
"Sweet foods" pattern						
Total fruits						
median intake (g/d)	31	88	135	186	290	
cases of lung cancer	408	305	235	235	243	
person years	3717	3746	3737	3799	3766	
RR (multivariate[c])	1	0.77	0.67	0.63	0.69	
95% CI	ref.	0.59-0.99	0.52-0.87	0.48-0.83	0.53-0.91	0.001
Alcohol						
median intake (g/d)	0	2.2	9.3	23	42	
cases of lung cancer	183	241	337	333	311	
person years	2592	3880	5056	4265	2973	
RR (multivariate[c])	1	1.11	1.23	1.08	1.56	
95% CI	ref.	0.80-1.54	0.91-1.67	0.80-1.47	1.11-2.18	0.03
Mono-/disaccharides						
median intake (g/d)	53	78	99	122	163	
cases of lung cancer	306	302	238	291	289	
person years	3669	3808	3736	3731	3822	
RR (multivariate[c])	1	0.95	0.80	0.96	0.88	
95% CI	ref.	0.72-1.25	0.59-1.06	0.71-1.30	0.63-1.23	0.53

[a] Q, quintile
[b] All P-values are from two-sided tests
[c] adjusted for age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked/day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderate active, active, very active)
[c] adjusted for age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked/day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderate active, active, very active), body mass index (kg/m^2)

substitution pattern were not associated with risk of lung cancer, in any of the smoking strata (P for interaction ranging from 0.21-0.63).
As different histologic groups of carcinomas are suspected to have different etiology, rate ratios of dietary patterns are shown for Kreyberg I and Kreyberg II carcinomas separately (Table 5). The salad vegetables pattern was associated with decreased risk of Kreyberg I carcinomas, but not with the less frequent Kreyberg II carcinomas. Similarly, the pork, processed meat, potatoes pattern increased risk of Kreyberg I carcinomas, but not with Kreyberg II carcinomas. The sweet pattern decreased the risk of both Kreyberg I and Kreyberg II carcinomas. The cooked vegetables or the brown/white bread substitution patterns were not associated with either type of carcinomas.

To compare the "salad vegetables" pattern and "cooked vegetables" pattern with traditional food based analyses we also calculated the risk estimates for the groups of total vegetables, raw vegetables and cooked vegetables (**Table 6**). High consumption of total vegetables was associated with decreased risk of lung cancer (RR_{Q5}: 0.66, 95%CI: 0.50-0.87, Test for trend: 0.008). Similarly, raw vegetables (RR_{Q5} 0.74 95%CI: 0.57-0.98, Test for trend: 0.07) and cooked vegetables (RR_{Q5} 0.73 95%CI: 0.55-0.96, Test for trend: 0.07) were associated with decreased risk of lung cancer. The risk reduction for the “salad vegetables” dietary pattern appeared not to be greater than found when studying the groups of total or raw vegetables. Therefore, we repeated the multivariate analyses, additionally adjusting for either total vegetables or raw vegetables. After adjustment for total vegetables consumption there was still a suggestion of a negative association for the “salad vegetables” pattern, however the point estimate and the test for trend no longer reached statistical significance (RR_{Q5}: 0.91, 95%CI: 0.82-1.01, Test for trend: 0.09). Adjustment for raw vegetables yielded similar results.
To compare the "sweet foods" pattern with the traditional reductionist approach we also calculated the risk estimates for total fruits consumption and for intake of mono- and disaccharides and alcohol (Pearson correlation coefficients with the Sweet pattern of 0.37, 0.58 and -0.40 respectively). High consumption of mono- and disaccharides was not associated with decreased risk of lung cancer (RR_{Q5}: 0.88, 95%CI: 0.63-1.23, Test for trend: 0.53). Consumption of alcohol (RR_{Q5} 1.56 95%CI: 1.11-2.18, Test for trend: 0.03) was associated with increased risk of lung cancer and total fruit (RR_{Q5} 0.69, 95%CI: 0.53-0.91, Test for trend: 0.001) with decreased risk. After additional adjustment for total fruits consumption and for total alcohol consumption the sweet pattern was still inversely associated with risk of lung cancer (respectively RR_{Q5}: 0.83, 95%CI: 0.74-0.93, Test for trend: 0.004 and RR_{Q5}: 0.64, 95%CI: 0.44-0.93, Test for trend: 0.006).

DISCUSSION

Five dietary patterns were identified in male participants of the Netherlands Cohort Study. The “salad vegetables” pattern and the “sweet” pattern were inversely associated with lung cancer risk, after adjustment for confounders, whereas the “pork, processed meat, potatoes” pattern was associated with increased risk of lung cancer. The other dietary patterns were not associated with lung cancer risk.
One of the strengths of our study is the completeness of follow-up of both person-years and cancer cases, ensuring that selection bias due to loss of follow-up is unlikely.[20] Another strength is the prospective design in which diet was measured before disease so

that information bias due to change in (recall of) diet due to disease was avoided. As pre-clinical symptoms might have influenced diet, the analyses were repeated excluding cases diagnosed during the first two years of follow-up (n=250). However, this did not affect the results. Even though we used a validated semi-quantitative FFQ to assess the participants' diet, we can not exclude the possibility that the dietary pattern approach combined artificial correlations due to grouping of items within the FFQ. However, spurious factors due to correlated questionnaire items are unlikely as Hu *et al*[21] and Togo *et al*[22] demonstrated that qualitative similar eating patterns were observed across FFQs and diet records. Furthermore, as the identified dietary patterns were very comparable with dietary patterns identified in other populations[16] and as the patterns correlated as expected with health-related lifestyle variables, it is unlikely that these dietary patterns are merely caused by correlated measurement error.

Another potential limitation that should be considered is whether the FFQ at baseline was a reliable estimate of past and future diet. The stability of dietary habits over time was evaluated from five annually repeated questionnaire administrations in independent random samples of the cohort. The mean intakes changed very little and the correlation between two measurements decreased only slightly over an increasing time interval. It was concluded that the single FFQ measurement characterizes dietary habits for a period of at least five to ten years.[23]

Despite the strengths of factor analysis, the technique has been criticized for its subjective decisions.[24] We performed extensive sensitivity analyses and our findings showed that the number of extracted factors and other subjective factor analytical decisions did not affect the identified dietary patterns.[16] However, as McCann *et al*[12] observed that reducing the number of input variables from 168 to 56 to 36 did not affect the derived factors, but did attenuate the odds ratios of endometrial cancer for the Healthy pattern, the classification of the FFQ items into 51 food groups may have attenuated the RRs.

The inverse association we observed between the “salad vegetables” pattern and lung cancer risk is in agreement with both case-control and prospective cohort studies, that have consistently shown an inverse association between fruit and vegetable intake and lung cancer risk.[2;25-27] Both in previous analyses of the Netherlands Cohort Study[1] and the current re-analyses after 9.3 years of follow-up, we found reduced risks of lung cancer for vegetable and fruit consumption. The risk reduction for the “salad vegetables” dietary pattern appeared to be not greater than found when studying total or raw vegetables consumption, and additional adjustment for these food groups diminished the association with the pattern. Therefore, although other foods such as oil and wholemeal products also present in the “salad vegetables” dietary pattern may act synergistically with vegetables to decrease risk of lung cancer, our results favor a major role for (raw) vegetables. The "cooked vegetables" pattern was not significantly associated with lung cancer risk, whereas the sum of total cooked vegetables was. Therefore, the other foods associated with the "cooked vegetables" pattern (such as spirits, beef and veal, margarine, potatoes and potato products, etc.) may counteract the protective effect of the cooked vegetables.

The inverse association between the “sweet” pattern and lung cancer risk was unexpected and not in agreement with a case-control study that showed a positive dose-response relation between consumption of particular desserts (cakes, custard/cream pies) and the risk of lung cancer in men and women.[28] Although adjustment for smoking behavior changed the RR estimate, it is unlikely that the observed association is entirely due to residual confounding, because the association still existed after stratification according to smoking status and appeared to be strongest in never smokers. As additional adjustment

for mono- and disaccharides, total fruit consumption and total alcohol consumption diminished the inverse association only slightly, the higher consumption of mono- and disaccharides and fruits and the lower consumption of alcohol associated with this dietary pattern could not account for its full protective effect. Therefore, either the combination of effects of all food groups adds up to the protective effect of this dietary pattern, or other factors associated to this pattern affect lung cancer risk.
The "pork, processed meat, potatoes" pattern was associated with increased risk of lung cancer in men and this association was most evident among current smokers. So far, the evidence that higher meat and fat intake are associated with higher risk of lung cancer is weakly consistent.[3;26] All constituents together in the "pork, processed meat, potatoes" dietary pattern might act synergistically to increase risk of lung cancer. However, as adjustment for smoking behavior diminished the association, and the association was strongest in current smokers, we cannot exclude the possibility of residual confounding due to smoking.

Dietary patterns represent combinations of foods describing as much variance in the original dietary variables as possible, but not necessarily those combinations of foods that show the strongest association with disease. The dietary pattern approach complements the traditional reductionist approach, by enabling to study not only the effect of single factors, but also of complex combinations of dietary exposures.
In conclusion, although we cannot entirely exclude the possibility of residual confounding, the dietary patterns which we labeled "salad vegetables" and "sweet foods" were associated with a decreased risk of lung cancer, whereas the "pork, processed meat, potatoes" pattern might be associated with an increased risk. The protective effect of cooked vegetables appeared to be counter acted by other foods or factors associated with the "cooked vegetables" pattern. These results demonstrate how studying both single factors and dietary patterns gives more insight into the complex, and often seemingly inconsistent, associations between diet and cancer.

REFERENCES

1. Voorrips, L. E., Goldbohm, R. A., Verhoeven, D. T. H., Poppel, G. A. F. C. v., Sturmans, F., Hermus, R. J. J., et al. Vegetable and fruit consumption and lung cancer risk in the Netherlands Cohort Study on diet and cancer. Cancer Causes Control, 2000; *11:* 101-115.

2. Smith-warner, S. A., Spiegelman, D., Yaun, S. S., Albanes, D., Beeson, W. L., van den Brandt, P. A., et al. Fruits and vegetables and lung cancer: A pooled analysis of cohort studies. Int. J. Cancer, 2003.

3. Smith-warner, S. A., Ritz, J., Hunter, D. J., Albanes, D., Beeson, W. L., van den Brandt, P. A., et al. Dietary fat and risk of lung cancer in a pooled analysis of prospective studies. Cancer Epidemiol., Biomark. Prev., 2002; *11:* 987-992.

4. Randall, E., Marshall, J. R., Brasure, J., Graham, S. Dietary patterns and colon cancer in western New York. Nutr. Cancer, 1992; *18:* 265-276.

5. Slattery, M. L., Boucher, K. M., Caan, B. J., Potter, J. D., Ma, K. N. Eating patterns and risk of colon cancer. Am. J. Epidemiol., 1998; *148:* 4-16.

6. Terry, P., Hu, F. B., Hansen, H., Wolk, A. Prospective study of major dietary patterns and colorectal cancer risk in women. Am. J. Epidemiol., 2001; *154:* 1143-1149.

7. Fung, T., Hu, F. B., Fuchs, C., Giovannucci, E., Hunter, D. J., Stampfer, M. J., et al. Major dietary patterns and the risk of colorectal cancer in women. Arch. Intern. Med., 2003; *163:* 309-314.

8. Dixon, L. B., Balder, H. F., Virtanen, M., Rashidkhani, B., Mannisto, S., Krogh, V., et al. Dietary patterns associated with colon and rectal cancers: Results from the DIETSCAN project. Accepted for publication, Am. J. Clin. Nutr, 2004.

9. Palli, D., Russo, A., Decarli, A. Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control, 2001; *12:* 163-172.

10. Masaki, M., Sugimori, H., Nakamura, K., Tadera, M. Dietary patterns and stomach cancer among middle-aged male workers in Tokyo. Asian Pac. J. Cancer Prev., 2003; *4:* 61-66.

11. Terry, P., Suzuki, R., Hu, F. B., Wolk, A. A prospective study of major dietary patterns and the risk of breast cancer. Cancer Epidemiol. Biomarkers Prev., 2001; *10:* 1281-1285.

12. McCann, S. E., Marshall, J. R., Brasure, J. R., Graham, S., Freudenheim, J. L. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr., 2001; *4:* 989-997.

13. Handa, K., Kreiger, N. Diet patterns and the risk of renal cell carcinoma. Public Health Nutrition, 2002; *5:* 757-767.

14. Markaki, I., Linos, D., Linos, A. The influence of dietary patterns on the development of thyroid cancer. Eur. J. Cancer, 2003; *39:* 1912-1919.

15. Tsai, Y. Y., McGlynn, K. A., Hu, Y., Cassidy, A. B., Arnold, J., Engstrom, P. F., et al. Genetic susceptibility and dietary patterns in lung cancer. Lung Canc., 2003; *41:* 269-281.

16. Balder, H. F., Virtanen, M., Brants, H. A. M., Krogh, V., Dixon, L. B., Frans, T., et al. Common and country-specific dietary patterns in four European cohort studies. J. Nutr., 2003; *133:* 4246-4251.

17. van den Brandt, P. A., Goldbohm, R. A., Van 't Veer, P., Volovics, A., Hermus, R. J., Sturmans, F. A large-scale prospective cohort study on diet and cancer in The Netherlands. J. Clin. Epidemiol., 1990; *43:* 285-295.

18. Goldbohm, R. A., van den Brandt, P. A., Brants, H. A., Van 't Veer, P., Al, M., Sturmans, F., Hermus, R. J. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur. J. Clin. Nutr., 1994; *48:* 253-265.

19. van den Brandt, P. A., Schouten, L. J., Goldbohm, R. A., Dorant, E., Hunen, P. M. Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. Int. J. Epidemiol., 1990; *19:* 553-558.

20. Goldbohm, R. A., van den Brandt, P. A., Dorant, E. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. Tijdschr Soc Gezondheidsz, 1994; *72:* 80-84.

21. Hu, F. B., Rimm, E., Smith-warner, S. A., Feskanich, D., Stampfer, M. J., Ascherio, A., et al. Reproducibility and validity of dietary patterns assessed with a food- frequency questionnaire. Am. J. Clin. Nutr., 1999; *69:* 243-249.

22. Togo, P., Heitmann, B. L., Sorensen, T. I. A., Osler, M. Consistency of food intake factors by different dietary assessment methods and population groups. Br. J. Nutr., 2003; *90:* 667-678.

23. Goldbohm, R. A., Van 't Veer, P., van den Brandt, P. A., van-'t-Hof, M. A., Brants, H. A., Sturmans, F., et al. Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. Eur. J. Clin. Nutr., 1995; *49:* 420-429.

24. Martinez, M. E., Marshall, J. R., Sechrest, L. Invited commentary: Factor analysis and the search for objectivity [comment] [see comments]. Am. J. Epidemiol., 1998; *148:* 17-19.

25. Ziegler, R. G., Mayne, S. T., Swanson, C. A. Nutrition and lung cancer. Cancer Causes Control, 1996; *7:* 157-177.

26. World Cancer Research Fund Food, nutrition and the prevention of cancer: a global perspective. Washington: World Cancer Research Fund / American Institute for Cancer Research, 1997.

27. Goldbohm, R. A., Voorrips, L. E. Epidemiology of nutrition and lung cancer. Nestle. Nutr. Workshop Ser. Clin. Perform. Programme., 2000; *4:* 23-35.

28. Goodman, M. T., Hankin, J. H., Wilkens, L. R., Kolonel, L. N. High-fat foods and the risk of lung cancer. Epidemiology, 1992; *3:* 288-299.

7

Dietary patterns and cancer risk in a prospective cohort study

Balder HF, Van den Brandt PA, Goldbohm RA

ABSTRACT

The objective was to study the association of dietary patterns with overall and some of the most frequently occurring cancers within the Netherlands Cohort Study on Diet and Cancer and compare these results with nutrient based analyses. The baseline measurement of this prospective cohort study in 1986 included information on dietary intake, smoking habits and other covariates by means of a self-administered questionnaire. Follow-up was established by computerized record linkage to cancer registries and a pathology register. Using exploratory factor analysis five dietary patterns were identified in men and women separately. Most of these patterns were comparable between men and women. After 9.3 years of follow-up, in all, 10731 (6507 male and 4224 female) incident primary cancer cases were available for analyses. The dietary patterns characterized by consumption of vegetables appeared to be overall protective of cancer, although the results were not statistically significant. A dietary pattern characterized by consumption of sweet foods in men appeared to be inversely associated with overall cancer risk (RR_{Q5}: 0.77, 95%CI:0.60-0.99, Test for trend: 0.03). Although not statistically significant, among women a similar inverse association was observed for a dietary pattern characterized by sweet and savory snacks. This could not be attributed to the nutritional value of these patterns and is therefore most likely to be due to other factors associated with this pattern. A dietary pattern characterized by consumption of pork, processed meat, potatoes, coffee and low fat margarine was found to increase cancer risk in men (RR_{Q5}: 1.35, 95%CI:1.04-1.74, Test for trend: 0.13) and decrease cancer risk in women (RR_{Q5}: 0.69, 95%CI:0.56-0.85, Test for trend < 0.001). This decrease in cancer risk among women is consistent with a decreased risk associated with higher intake of mono- and polyunsaturated fat and lower intake. Overall risk of cancer was not associated with dietary patterns characterized by consumption brown instead of white types of bread or by fat dairy (in women).

INTRODUCTION

Approximately 66,000 invasive tumors and 6,000 non-invasive tumors were diagnosed in the Netherlands in 1998.[1] Of the invasive tumors lung cancer and breast cancer were the most common, followed by colorectal cancer, prostate cancer and cancer of the urinary tract.

Traditionally, epidemiological studies have investigated the relationship between diet and cancer either by studying consumption of whole foods, such as fruits and vegetables, or by studying specific dietary components. Several review articles have been published on these associations.[2-5] Relatively few studies have addressed the association between cancer and broader eating patterns that reflect many dietary exposures together. These dietary patterns were thought to better explain disease occurrence than individual dietary exposures.[6] Factor analysis is one method of identifying and examining broader dietary patterns that may be associated with disease. To date, this approach has been used before in studies of colon or colorectal cancer,[7-11] stomach cancer,[12;13] breast cancer,[14-16] endometrial cancer,[17] prostate cancer,[18] renal cell cancer,[19] and thyroid cancer.[20] In a review of dietary patterns identified by factor analysis, we showed that dietary patterns characterized by vegetables consumption appeared to be protective against several types of cancer.[21] Dietary patterns characterized by meat and fat or processed foods were found to increase the risk of cancer mainly in case-control studies, but not in cohort studies.

As part of the DIETSCAN project, a collaboration with large prospective cohort studies from Sweden, Finland and Italy, we recently identified dietary patterns using factor analysis among participants of the Netherlands Cohort Study on Diet and Cancer (NLCS).[22] This gave us the opportunity to study prospectively the association between major dietary patterns common between the cohorts and incidence of colorectal[11] and breast cancer[23] within the DIETSCAN cohorts and with lung cancer in the NLCS[24]. The aim of the current paper was to examine the association of dietary patterns with overall cancer and with some of the most frequently occurring cancers within the NLCS after 9.3 years of follow-up. To evaluate the value of the dietary pattern approach, these results were compared with analyses for macro and micronutrients.

SUBJECTS AND METHODS

Study population The Netherlands Cohort Study on diet and cancer (NLCS) was approved by the institutional review boards of the TNO Nutrition and Food Research Institute (Zeist) and Maastricht University (Maastricht). The NLCS started in September 1986 when participants aged 55 to 69 years from 204 municipalities were enrolled in the cohort using computerized population registries. The study design of the NLCS has been reported in detail elsewhere.[25] In total 58,279 men and 62,573 women completed a self-administered mailed questionnaire on habitual dietary intake, sociodemographic and lifestyle characteristics, medical history and other potential risk factors for cancer. The dietary part of this questionnaire consisted of a 150-item semi-quantitative food frequency questionnaire on the usual intake of food and beverages in the year preceding the start of the study, which was validated against a 9-day diet record.[26] Mean daily nutrient intakes were calculated using the computerized Dutch food composition table [27]. Questionnaire data of all cases and subcohort members have been key-entered twice and blinded with respect to case/subcohort status to avoid random and systematic coding errors. After

Table 1 Factor loading matrix

Food groups	Men					Women				
	SV	CV	PPP	SW	B/W	SV	FD	B/W	PPP	SN
Legumes	0.11	**0.58**	0.08	0.04	-0.02	**0.55**	0.23	-0.01	0.12	-0.04
Cabbages	0.13	**0.73**	0.12	-0.05	0.04	**0.59**	0.15	0.03	0.21	-0.11
Leaf vegetables – cooked	0.01	**0.70**	-0.03	0.01	-0.01	**0.48**	0.15	0.03	0.08	-0.15
Leaf vegetables – raw	**0.40**	0.13	0.00	0.28	0.14	**0.52**	-0.13	0.18	0.00	-0.03
Allium vegetables	**0.51**	**0.37**	0.01	-0.14	0.08	**0.59**	-0.13	-0.03	-0.02	-0.04
Carrots	0.20	**0.46**	-0.06	0.26	0.23	**0.50**	0.01	0.25	-0.17	-0.07
Tomatoes	**0.45**	0.26	0.02	0.00	0.14	**0.45**	-0.17	0.16	0.02	-0.07
Mushrooms	**0.52**	0.08	-0.04	-0.11	0.06	**0.39**	-0.34	0.01	0.02	0.09
Citrus fruit	0.21	0.18	-0.30	0.13	0.10	0.18	-0.25	0.20	-0.25	0.07
Bananas	0.14	0.05	0.07	0.25	-0.04	0.11	0.01	0.02	-0.04	0.04
Apples and pears	-0.01	0.08	-0.09	0.26	**0.37**	0.17	0.03	**0.40**	0.04	0.03
Berries – all types	0.14	0.06	-0.05	**0.36**	-0.03	0.12	-0.05	0.10	-0.09	0.26
Potatoes and potato products	-0.18	0.31	**0.38**	0.25	0.02	0.08	**0.52**	0.04	0.29	-0.00
Rice	**0.43**	-0.04	-0.18	0.09	0.02	**0.35**	-0.05	-0.08	-0.31	0.08
Pasta	**0.53**	-0.10	0.20	0.06	-0.04	**0.36**	-0.02	-0.10	-0.05	0.21
Dry cereals	0.15	-0.04	-0.28	0.19	0.13	0.14	0.03	0.26	-0.28	0.09
Bread, crackers etc. –white types	-0.16	0.06	0.04	0.13	**-0.62**	-0.03	0.19	**-0.49**	0.05	0.15
Bread, crackers etc. – brown/wholemeal types	-0.01	-0.03	0.33	0.18	**0.67**	0.07	0.17	**0.64**	0.19	-0.03
Fermented whole milk and milk products	-0.01	-0.01	-0.06	0.28	-0.01	0.09	0.28	-0.03	-0.15	0.05
Fermented medium, lowfat and skimmed milk prod	0.04	0.13	-0.21	0.02	0.33	0.14	-0.26	0.34	-0.10	-0.04
Nonfermented whole milk and milk products	-0.24	0.07	0.15	0.29	-0.29	-0.02	**0.49**	-0.21	-0.09	0.00
Nonfermented medium, lowfat and skimmed milk prod	0.04	-0.08	-0.03	-0.06	0.19	-0.01	-0.15	0.25	0.04	0.04
Full cream cheese and cheese spreads	-0.01	0.02	0.08	0.23	0.22	0.13	0.18	0.30	-0.12	0.12
Lowfat cheese and cheese spreads	-0.03	0.09	-0.02	-0.00	0.18	0.04	-0.23	0.10	0.10	-0.11
Beef and veal – fresh	-0.09	0.30	-0.11	-0.00	-0.07	0.04	0.16	0.01	-0.07	-0.16
Pork – fresh	0.01	0.06	**0.48**	-0.18	-0.17	-0.06	0.07	-0.19	**0.43**	0.06
Liver – fresh	0.15	0.13	0.10	-0.04	-0.10	0.17	-0.20	-0.06	0.28	0.10
Poultry – fresh	0.30	0.01	-0.01	-0.04	-0.01	0.24	-0.32	-0.11	0.02	0.05
Processed meat	0.02	0.02	**0.53**	0.07	-0.08	0.15	0.07	-0.03	**0.50**	0.20
Fish (shellfish and crustacean included)	0.29	0.14	-0.06	-0.11	-0.11	0.33	-0.26	-0.11	-0.04	0.09
Eggs	0.11	0.07	0.19	-0.04	-0.32	0.21	-0.01	-0.19	0.14	0.03
Butter	0.07	-0.01	-0.30	0.14	-0.32	0.13	0.13	-0.12	**-0.45**	0.07
Margarine	-0.30	0.25	0.25	0.12	-0.01	0.07	**0.47**	0.00	0.17	0.04
Lowfat margarine	0.06	-0.12	**0.46**	0.08	0.28	-0.11	-0.15	0.26	**0.50**	0.02
Oil	**0.58**	0.02	0.06	0.05	0.01	**0.41**	-0.14	-0.00	-0.25	0.09
Dressings and other similar sauces	0.07	-0.00	0.22	0.21	-0.05	0.16	0.03	0.06	0.34	0.05
Peanuts and other nuts	0.29	-0.11	0.06	0.20	-0.13	0.10	0.04	0.01	-0.01	**0.68**
Savory snacks	0.32	-0.22	0.06	0.13	-0.12	-0.02	-0.09	-0.07	0.03	**0.70**
Cakes, sweet breads and cookies	-0.06	-0.04	-0.01	**0.53**	0.02	-0.07	0.19	0.25	-0.01	**0.39**
Candies	0.05	-0.01	-0.04	**0.38**	-0.13	-0.10	0.07	0.11	-0.02	**0.54**
Added sugar	-0.07	-0.03	0.30	0.25	-0.34	-0.11	0.32	-0.28	-0.01	0.06
Sweet sandwich spread	-0.11	-0.02	0.04	**0.45**	0.01	-0.05	**0.47**	0.21	-0.09	0.04
Coffee	-0.02	0.04	**0.50**	-0.13	-0.02	-0.01	-0.00	-0.15	**0.44**	-0.13
Tea	-0.06	0.13	-0.27	0.31	0.01	0.00	0.19	0.31	-0.24	0.05
Fruit juices	0.10	0.07	-0.02	0.12	-0.07	0.15	0.09	-0.02	0.04	0.13
Soft drinks and syrups	0.07	-0.04	0.17	0.17	-0.21	-0.03	0.07	-0.15	0.15	0.25
Beer	0.13	0.02	0.21	-0.34	-0.10	0.06	-0.00	-0.23	-0.00	-0.03
Wine and fortified wine	**0.40**	-0.06	-0.09	-0.02	0.03	0.15	-0.25	-0.16	-0.09	0.04
Spirits	0.16	0.09	-0.03	-0.29	-0.19	0.15	-0.07	-0.23	0.07	0.06

SV="(Salad) Vegetables pattern"; CV="Cooked vegetables pattern"; PPP="Pork, processed meat, potatoes pattern"; SW="Sweet foods patterns"; B/W="Brown/white bread substitution pattern"; FD="Fat dairy pattern"; SN="Sweet and/or savory snacks pattern"
For convenience factor loadings **>0.35** are printed **bold**.

baseline exposure measurement, a subcohort of 5000 men and women was randomly sampled from the large cohort. Following the case-cohort approach, this subcohort was followed for migration and vital status to calculate person time. The entire cohort has been followed-up for incidence of cancer. Follow-up for incident cancer in the entire cohort has been established by record linkage to the Netherlands Cancer Registry and the Netherlands Pathology Registry (PALGA)[28]. After 9.3 years of follow-up, i.e. from September 1986 to December 1995, 11,631 incident cases of cancer were identified. In 408 cases the subject was diagnosed with cancer more than once, either because of prevalent cancer at baseline or recurring tumors or a tumor of a different site. Only those cases were included in which the incident cancer was the first tumor that was diagnosed in a person (n=11,223). After excluding subjects with incomplete or inconsistent dietary information and subjects from the case group who reported prevalent cancer other than skin cancer, 10,731 (6507 men and 4224 women) incident primary cancer cases (ICD-0 codes T140.0-T199.9) were available for analyses. From the subcohort, after exclusion of 377 subjects with incomplete or inconsistent dietary information and 482 subjects with prevalent cancer, 2190 men and 2251women remained for analysis.

Assessment of dietary patterns The dietary pattern assessment and related sensitivity analyses as part of the DIETSCAN project have been described in detail elsewhere.[22] To identify dietary patterns in the subcohort, exploratory factor analysis using principal components analysis was conducted on intake (g/day) of 49 food groups. The Eigenvalue > 1 and the Scree test, traditional criteria in factor analysis, were used to determine the number of factors to be extracted. Factors were rotated by orthogonal Varimax transformation. Factor loadings represent correlation coefficients between food groups and dietary patterns and food groups with absolute factor loadings > 0.35 were considered as significantly contributing to a dietary pattern. Labels of the dietary patterns were assigned according to the foods with high loadings. Prior to etiological analysis, factor scores were determined by summing the standardized intakes from each food group weighted by the factor loadings, creating for each participant a unique score for each factor. High scores represented high intake of foods within the corresponding dietary pattern; low scores represented low intake of those foods. Because the factor scores represent standardized variables, the distribution of each factor score had a mean of zero and a standard deviation of one.

Etiological analyses Cox proportional hazards models were constructed to estimate hazard ratios and 95% confidence intervals relating the dietary patterns to the incidence of lung cancer (stcox in Stata version 8; Stata Corporation, College Station, Texas). The proportional hazards assumption was tested using Schoenfeld residuals (stphtest in Stata). Because of the case-cohort design, the 95% confidence intervals were corrected for the additional variance introduced by using a randomly sampled subcohort instead of the complete cohort (robust option in Stata). Rate ratio estimates for the dietary pattern scores were calculated for quintiles based on the distribution in the subcohort and for continuous variables with an increment of 1 SD. Tests for trend in the rate ratios were assessed by fitting ordinal exposure categories as continuous variables. Age at baseline (y), total energy intake (kJ), body mass index (kg/m^2), current cigarette smoking (yes/no), number of cigarettes smoked per day, number of years of smoking regularly, highest attained education (primary school; lower vocational school; junior high school; senior high school, higher vocational school; university), non-occupational physical activity (<30

min/day, 30-60 min/day, 60-90 min/day, >90 min/day), family history of cancer in first or second degree relatives (yes/no) and use of dietary supplements (yes/no) were considered as potential confounders for which the rate ratios were adjusted. Breast cancer analyses were additionally adjusted for reproductive factors (previous benign breast disease, family history of breast cancer, age at menarche, age at menopause, ever use of oral contraceptives, parity and age at birth of first child). We also calculated the risk of overall cancer for intake of total energy, protein (total, animal and plant), fat (total, saturated, mono- and polyunsaturated), carbohydrates (total, poly- and mono/di-saccharides), dietary fiber, alcohol, vitamin C, beta-carotene and folate. Intake of macro nutrients were adjusted for energy by the residual method.[29] The increment used for the macro- and micronutrients corresponded to 1 standard deviation of the distribution in the subcohort.

RESULTS

Exploratory factor analysis identified five stable dietary patterns both among male and female subcohort members of the NLCS (**Table 1**).[22] Three of these patterns were observed both in men and women: the "Salad vegetables" (SV: several vegetable items, fruit, pasta and rice, poultry, fish, and oil), the "Pork, processed meat, potatoes" (PPP: pork, processed meat and potatoes, coffee and low-fat margarine and negatively on butter) and the "Brown/white bread substitution" (B/W: brown/wholemeal bread types and apples and pears and negatively with white types of bread) pattern. In addition, in men a "Sweet" pattern (SW: with positive loads on cakes and cookies, sweet sandwich spread, candies and strawberries and negative loads on beer and spirits) and in women a "Sweet and savory snacks" pattern (SN: with high loads on savory snacks, nuts, candies and cakes and cookies) was observed. Furthermore, in men a "Cooked vegetables" (CV:with high loadings on cooked leafy vegetables, cabbages, legumes and carrots) was observed, whereas in women, a "Fat dairy" (FD, with high loads on potatoes, nonfermented whole milk, margarine and sweet sandwich spread pattern) pattern was identified.

Table 2 Pearson correlation coefficients for dietary pattern scores and energy-adjusted macro- and micro-nutrients among members of the subcohort of the NLCS.

	Men					Women				
	SV	CV	PPP	SW	B/W	SV	FD	B/W	PPP	SN
Total energy	0.13	0.21	0.50	0.50	-0.15	0.32	0.50	0.12	0.16	0.49
Total protein	0.17	0.24	0.05	-0.13	0.27	0.27	-0.33	0.23	0.17	-0.12
Animal protein	0.06	0.18	0.05	-0.22	0.09	0.15	-0.32	0.11	0.16	-0.16
Plant protein	0.22	0.09	0.01	0.16	0.33	0.28	-0.01	0.24	0.05	0.09
Total fat	-0.21	0.04	0.01	-0.01	-0.12	-0.09	0.16	-0.15	0.03	0.01
saturated fat	-0.20	0.01	-0.18	0.07	-0.24	-0.09	0.16	-0.14	-0.27	-0.06
mono-unsaturated fat	-0.11	-0.01	0.13	-0.10	-0.17	-0.12	0.06	-0.18	0.17	0.11
poly-unsaturated fat	-0.05	0.06	0.13	-0.04	0.23	0.03	-0.01	0.08	0.25	-0.00
Total carbohydrates	-0.13	-0.11	-0.00	0.41	0.15	-0.12	0.19	0.23	-0.05	-0.03
mono-/disaccharides	-0.10	-0.10	-0.10	0.35	-0.17	-0.11	0.09	-0.01	-0.14	0.06
polysaccharides	-0.08	-0.02	0.13	0.16	0.44	-0.04	0.15	0.34	0.13	-0.14
Dietary fiber	0.14	0.34	-0.03	0.21	0.68	0.43	-0.03	0.58	0.02	-0.18
Alcohol	0.36	0.03	0.05	-0.40	-0.16	0.21	-0.28	-0.27	-0.06	0.17
Vitamin C	0.30	0.52	-0.16	0.30	0.18	0.52	-0.08	0.30	-0.11	0.08
Beta-carotene	0.27	0.64	0.07	0.24	0.15	0.64	0.08	0.22	0.00	-0.05
Folate	0.25	0.67	0.21	0.35	0.26	0.71	0.15	0.39	0.09	0.08

SV="(Salad) Vegetables pattern"; CV="Cooked vegetables pattern"; PPP="Pork, processed meat, potatoes pattern"; SW="Sweet foods patterns"; B/W="Brown/white bread substitution pattern"; FD="Fat dairy pattern"; SN="Sweet and savory snacks pattern"

All dietary patterns, except B/W among men, were positively associated with total energy intake (**Table 2**). Higher SV and CV pattern scores were associated with higher intake of vitamins and dietary fiber and with lower intakes of mono- and disaccharide's. Furthermore, the SV pattern scores were positively associated with intake of plant protein and alcohol and inversely with intake of saturated and mono-unsaturated fat. The PPP pattern score was associated with higher intakes of mono- and ploy-unsaturated fat and polysaccharides and lower intakes of saturated fat and mono- and disaccharide's. The B/W pattern score was associated with higher intakes of plant protein, polysaccharides, dietary fiber and vitamins and lower intakes of fat and alcohol. Both the SW and SN pattern scores were positively associated with intake of plant protein and inversely with animal protein. However, higher SW pattern scores were associated with higher carbohydrate, fiber and vitamin intakes and with lower alcohol and mono-unsaturated fat intake, whereas higher SN pattern scores were associated with higher alcohol and mono-unsaturated fat intake and lower intake of polysaccharides and dietary fiber. Among women higher FD pattern scores were associated with higher saturated fat and carbohydrate intake, and with lower total and animal protein and alcohol.

Table 3 Mean and distribution (%) of non-dietary baseline characteristics among cancer cases and subcohort members

	Men		Women	
Characteristics	subcohort	cases	subcohort	cases
Total	2190	6507	2251	4224
Age (y)	61.3	62.5	61.4	62.0
Total energy intake (MJ)	9.1	9.0	7.1	7.0
BMI (kg/m^2)	24.9	25.0	25.1	25.3
Smoking habits				
never smoker (%)	12.7	9.5	58.2	56.2
former smoker (%)	51.5	47.3	20.7	20.1
current smoker (%)	35.8	43.2	21.1	23.5
Number of cigarettes/day[1]	17.0	18.0	11.4	12.6
Duration of smoking cigarettes (y)[1]	33.3	36.7	27.8	29.2
Highest level of education (%)				
Primary school	25.0	27.6	33.1	33.2
Lower vocational	21.0	20.8	22.8	22.2
High school	35.4	34.0	34.8	35.6
Higher vocational/university	18.0	17.0	8.8	8.1
Non-occupational physical activity (%)				
< 30 min/day	18.4	19.3	24.9	29.3
30-60 min/day	30.9	29.4	31.1	29.3
60-90 min/day	18.8	19.6	22.5	21.5
> 90 min/day	31.9	31.7	21.4	20.0
Use of dietary supplements (%)	23.0	22.9	36.6	35.2
Family history of cancer (%)	43.4	47.4	47.2	51.1

[1] among former and current smokers

The non-dietary baseline characteristics among 6507 male and 4224 female cancer cases and 2190 male and 2251 female subcohort members with complete dietary data are presented in **Table 3**. On average, cases were older, were more likely to smoke cigarettes, and more often reported family history of cancer. Compared to members of the subcohort, male cancer cases were more likely to have a lower educational level, whereas female cancer cases were less physically active.

Table 4 Rate ratios (RRs) and 95 percent confidence intervals (95% CI) of all cancers combined according to dietary pattern score among men and women in the Netherlands Cohort Study

Dietary pattern	*Men*					
	Q1	Q2	Q3	Q4	Q5	P trend
Salad Vegetables pattern (SV)						
cases	1377	1300	1368	1274	1188	
person years	3577	3634	3662	3614	3684	
RR(age, energy)[2]	1	0.98	1.03	1.01	0.96	
95% CI		0.83-1.15	0.88-1.21	0.86-1.18	0.81-1.13	0.77
RR(multivariate)[3]	1	1.00	1.05	0.97	0.94	
95% CI		0.82-1.23	0.86-1.28	0.79-1.19	0.76-1.15	0.53
RR(cases >2 yrs)[4]	1	0.99	1.08	0.97	0.95	
95% CI		0.80-1.22	0.88-1.33	0.79-1.20	0.76-1.17	0.64
Cooked Vegetables pattern (CV)						
cases	1164	1301	1421	1325	1296	
person years	3614	3651	3646	3726	3534	
RR(age, energy)[2]	1	1.07	1.20	1.06	1.08	
95% CI		0.91-1.26	1.02-1.42	0.90-1.24	0.91-1.28	0.45
RR(multivariate)[3]	1	0.98	1.03	0.91	0.95	
95% CI		0.80-1.21	0.84-1.26	0.74-1.12	0.76-1.17	0.41
RR(cases >2 yrs)[4]	1	0.96	1.02	0.89	0.96	
95% CI		0.77-1.19	0.82-1.26	0.72-1.10	0.77-1.21	0.54
Pork, Processed meat, Potatoes pattern (PPP)						
cases	1196	1301	1484	1250	1276	
person years	3570	3626	3605	3670	3701	
RR(age, energy)[2]	1	1.16	1.40	1.22	1.41	
95% CI		0.98-1.36	1.19-1.66	1.03-1.44	1.17-1.69	0.001
RR(multivariate)[3]	1	1.17	1.27	1.06	1.35	
95% CI		0.95-1.43	1.03-1.56	0.86-1.32	1.04-1.74	0.13
RR(cases >2 yrs)[4]	1	1.16	1.28	1.09	1.30	
95% CI		0.94-1.44	1.03-1.58	0.87-1.36	1.00-1.69	0.17
Brown/White bread substitution pattern (B/W)						
cases	1315	1364	1266	1417	1145	
person years	3630	3569	3690	3663	3620	
RR(age, energy)[2]	1	1.00	0.91	0.98	0.84	
95% CI		0.85-1.17	0.78-1.08	0.83-1.15	0.72-0.99	0.06
RR(multivariate)[3]	1	1.10	0.96	1.05	0.99	
95% CI		0.90-1.35	0.78-1.18	0.85-1.29	0.80-1.22	0.75
RR(cases >2 yrs)[4]	1	1.12	0.96	1.06	1.03	
95% CI		0.91-1.38	0.78-1.19	0.85-1.32	0.83-1.27	0.96
Sweet foods pattern (SW)						
cases	1579	1368	1293	1157	1110	
person years	3547	3623	3554	3651	3797	
RR(age, energy)[2]	1	0.79	0.76	0.62	0.60	
95% CI		0.68-0.93	0.64-0.89	0.53-0.74	0.50-0.73	0.000
RR(multivariate)[3]	1	0.81	0.83	0.75	0.77	
95% CI		0.67-0.99	0.68-1.03	0.61-0.94	0.60-0.99	0.03
RR(cases >2 yrs)[4]	1	0.80	0.87	0.76	0.79	
95% CI		0.65-0.98	0.70-1.07	0.60-0.95	0.61-1.02	0.06

[2] adjusted for age at baseline (y) and total energy intake (kJ) [3] adjusted for age at baseline (y), total energy intake (kJ), body mass index (kg/m^2), current cigarette smoking (yes/no), number of cigarettes smoked per day, number of years of smoking regularly, highest attained education (primary school; lower vocational school; junior high school; senior high school, higher vocational school; university),

Table 4 continued

Dietary pattern	***Women***					
	Q1	Q2	Q3	Q4	Q5	P trend
Salad Vegetables pattern (SV)						
cases	873	865	848	851	787	
person years	3890	3967	3918	3917	3934	
RR(age, energy)[2]	1	1.01	1.01	1.04	0.95	
95% CI		0.86-1.19	0.86-1.19	0.88-1.23	0.81-1.13	0.74
RR(multivariate)[3]	1	0.98	0.93	0.98	0.87	
95% CI		0.80-1.19	0.76-1.13	0.79-1.22	0.70-1.09	0.30
RR(cases >2 yrs)[4]	1	0.98	0.89	0.98	0.86	
95% CI		0.79-1.21	0.72-1.11	0.79-1.23	0.69-1.08	0.26
Fat Dairy pattern (FD)						
cases	816	887	859	855	807	
person years	3934	3872	3905	3983	3931	
RR(age, energy)[2]	1	1.08	1.05	1.04	1.01	
95% CI		0.92-1.27	0.89-1.23	0.88-1.23	0.84-1.21	0.93
RR(multivariate)[3]	1	1.10	0.99	0.95	0.91	
95% CI		0.90-1.35	0.80-1.21	0.76-1.18	0.70-1.17	0.25
RR(cases >2 yrs)[4]	1	1.05	0.99	0.97	0.91	
95% CI		0.85-1.29	0.79-1.23	0.77-1.22	0.70-1.19	0.43
Pork, Processed meat, Potatoes pattern (PPP)						
cases	894	880	883	839	728	
person years	3886	3910	3879	3963	3987	
RR(age, energy)[2]	1	0.98	1.01	0.95	0.84	
95% CI		0.84-1.16	0.86-1.19	0.81-1.12	0.71-0.99	0.05
RR(multivariate)[3]	1	0.84	0.87	0.75	0.69	
95% CI		0.69-1.03	0.71-1.06	0.61-0.92	0.56-0.85	0.000
RR(cases >2 yrs)[4]	1	0.84	0.87	0.76	0.69	
95% CI		0.69-1.04	0.71-1.07	0.61-0.94	0.55-0.86	0.001
Brown/White bread substitution pattern (B/W)						
cases	911	865	805	842	801	
person years	3891	3938	3933	3960	3904	
RR(age, energy)[2]	1	0.94	0.87	0.91	0.89	
95% CI		0.80-1.10	0.74-1.02	0.77-1.07	0.76-1.05	0.14
RR(multivariate)[3]	1	0.97	0.90	0.98	0.98	
95% CI		0.79-1.18	0.73-1.10	0.80-1.21	0.80-1.21	0.93
RR(cases >2 yrs)[4]	1	0.99	0.90	0.97	1.00	
95% CI		0.81-1.22	0.73-1.12	0.78-1.20	0.80-1.25	0.94
Sweet and savoury snacks pattern (SN)						
cases	883	894	871	767	809	
person years	3934	3918	3908	3941	3924	
RR(age, energy)[2]	1	1.01	1.01	0.89	0.96	
95% CI		0.86-1.19	0.86-1.20	0.75-1.05	0.80-1.16	0.32
RR(multivariate)[3]	1	1.02	0.92	0.82	0.87	
95% CI		0.83-1.25	0.74-1.13	0.66-1.03	0.67-1.13	0.08
RR(cases >2 yrs)[4]	1	1.03	0.95	0.86	0.87	
95% CI		0.83-1.28	0.76-1.18	0.68-1.09	0.66-1.15	0.13

non-occupational physical activity (<30, 30-60, 60-90, >90 min/day), family history of cancer in first or second degree relatives (yes/no), use of dietary supplements (yes/no) and mutually adjusted for the other dietary patterns; based on 4320 male and 2872 female cases.
[4] adjusted as above, excluding 827 male and 563 female cases diagnosed within the first two years of follow-up.

As shown in **Table 4**, the SV or the CV pattern were not significantly associated with decreased risk of cancer. The PPP pattern showed an increased risk of male cancer, in the model adjusted for age and total energy intake. However, after additional adjustment for smoking behavior and other potential confounders, the highest quintile was associated with increased risk (RR=1.35, 95%CI: 1.04-1.74) but the test for trend was no longer statistically significant (p=0.13). In contrast, in women the PPP pattern was after adjustment for smoking and other potential confounders associated with decreased risk of cancer (RR_{Q5}=0.69, 95%CI:0.56-0.85 p for trend=<0.001). The SW pattern was associated with a decreased risk of male cancer. Adjustment for smoking attenuated the association somewhat, but after multivariate adjustment cancer risk was still decreased (RR_{Q5}=0.77, 95%CI:0.60-0.99; p for trend=0.03). Among women, there was an indication of an inverse association of the SN pattern with cancer risk (RR_{Q5}=0.87, 95%CI:0.67-1.13; p for trend=0.08), however the association was not statistically significant. The B/W or FD patterns were not associated with risk of all cancers combined. Exclusion of cases that were diagnosed within the first two years of follow up did not materially change the risk estimates.

Multivariate analyses for separate cancer sites are displayed in **Table 5**. Among men, the SV pattern was inversely associated with risk of lung cancer ($RR_{per\ 1\ SD}$:0.86; 95%CI:0.77-0.95) and an inverse association was suggested for stomach cancer. Inverse associations were suggested for the CV pattern and risk of lung, colon, rectum and stomach cancer, although the risk estimates were not statistically significant. The PPP pattern was positively associated with risk of gastric cancer ($RR_{per\ 1\ SD}$:1.25; 95%CI:1.04-1.51). However, excluding cases diagnosed within the first and second year of follow-up did attenuate this association. The SW pattern was inversely associated with risk of lung cancer ($RR_{per\ 1\ SD}$:0.77; 95%CI:0.67-0.87) and bladder cancer ($RR_{per\ 1\ SD}$:0.80; 95%CI:0.66-0.95) and inverse associations were suggested for rectum and stomach cancer. The B/W pattern was positively associated with risk of rectal cancer ($RR_{per\ 1\ SD}$:1.18; 95%CI:1.00-1.39) and marginally significant with increased risk of kidney cancer and decreased risk of gastric and bladder cancer. After excluding cases diagnosed within the first and second year, the B/W pattern was no longer associated with risk of rectal and kidney cancer, but still was inversely associated with risk of stomach cancer ($RR_{per\ 1\ SD}$:0.84 95%CI: 0.71-1.00) and bladder cancer ($RR_{per\ 1\ SD}$:0.90; 95%CI: 0.79-1.04).

Among women, although not statistically significant, the relative risks for the SV pattern were below 1 for most types of cancer. The FD pattern was inversely associated with risk of lung cancer ($RR_{per\ 1\ SD}$:0.71; 95%CI:0.52-0.97) and an inverse association was suggested for rectum and breast cancer. An inverse association was suggested for the B/W pattern and risk of lung cancer, whereas a positive association was suggested with risk of stomach cancer. The PPP pattern was associated with a decreased risk of breast cancer ($RR_{per\ 1\ SD}$:0.87; 95%CI:0.80-0.95) and an inverse association was suggested for stomach cancer. The SN pattern was associated with a decreased risk of rectal cancer ($RR_{per\ 1\ SD}$:0.66; 95%CI:0.47-0.92) and although not statistically significant, the relative risk for the SN pattern was below 1 for most of the other types of cancer. Additional adjustment for reproductive factors did not change the risk estimates for women much (results not shown).

Table 5 Relative rate (RR)[1] and 95% confidence interval (95%CI) of all cancers combined and the ten most frequently occurring[2] site specific cancers among men, with an increment in 1 SD in factor score for dietary patterns

Cancer site	ICD-O-1-(*4)	number of cases	dietary pattern				
			SV	CV	PPP	SW	B/W
Men							
All cancers combined	140.0-199.9	6507[3]	0.97 0.91-1.03	0.97 0.90-1.04	1.04 0.96-1.14	0.89 0.82-0.97	1.02 0.95-1.09
Lung cancer	162.2-162.9	1426	0.86 0.77-0.95	0.93 0.84-1.02	1.05 0.92-1.19	0.77 0.67-0.87	1.02 0.93-1.12
Breast cancer	174.0-175.9						
Colorectal cancer							
of which colon cancer	153.0-153.9	539	0.96 0.85-1.08	0.93 0.81-1.06	1.03 0.89-1.20	0.99 0.86-1.15	1.03 0.92-1.16
of which rectum cancer	154.1	231	1.09 0.92-1.29	0.94 0.80-1.11	1.10 0.90-1.34	0.89 0.72-1.10	1.18 1.00-1.39
Prostate cancer	185.9	1282	1.02 0.94-1.11	0.99 0.91-1.08	1.02 0.91-1.14	1.01 0.90-1.12	1.02 0.94-1.11
Urinary tract cancer							
of which bladder cancer	188.0-188.9	369	1.06 0.93-1.20	1.02 0.89-1.17	1.00 0.82-1.23	0.80 0.66-0.95	0.90 0.79-1.02
of which kidney cancer	189.0	175	0.97 0.79-1.20	1.09 0.92-1.29	1.11 0.86-1.43	1.00 0.82-1.22	1.17 0.96-1.43
Stomach cancer	151.0-151.9	323	0.93 0.79-1.09	0.89 0.76-1.04	1.25 1.04-1.51	0.91 0.75-1.10	0.86 0.73-1.00
Women							
All cancers combined	140.0-199.9	4224[3]	0.94 0.88-1.02	0.95 0.87-1.04	0.98 0.92-1.05	0.90 0.84-0.97	0.92 0.83-1.01
Lung cancer	162.2-162.9	210	0.88 0.72-1.09	0.71 0.52-0.97	0.88 0.73-1.06	1.04 0.84-1.29	0.82 0.59-1.15
Breast cancer	174.0-175.9	1416	0.94 0.86-1.03	0.93 0.84-1.03	0.96 0.88-1.04	0.87 0.80-0.95	0.92 0.81-1.04
Colorectal cancer							
of which colon cancer	153.0-153.9	486	0.94 0.82-1.07	0.98 0.83-1.14	0.96 0.85-1.09	1.02 0.90-1.16	0.95 0.78-1.15
of which rectum cancer	154.1	128	0.90 0.71-1.15	0.80 0.59-1.08	0.95 0.76-1.20	1.02 0.81-1.30	0.66 0.47-0.92
Prostate cancer	185.9	-					
Urinary tract cancer							
of which bladder cancer	188.0-188.9						
of which kidney cancer	189.0	90	0.75 0.54-1.05	0.98 0.66-1.45	0.95 0.71-1.26	1.18 0.91-1.53	0.99 0.70-1.40
Stomach cancer	151.0-151.9	93	1.07 0.81-1.43	1.02 0.73-1.42	1.09 0.83-1.43	0.83 0.66-1.05	0.90 0.59-1.36

SV="(Salad) Vegetables pattern"; CV="Cooked vegetables pattern"; PPP="Pork, processed meat, potatoes pattern"; SW="Sweet foods patterns"; B/W="Brown/white bread substitution pattern"; FD="Fat dairy pattern"; SN="Sweet and savory snacks pattern"
1 adjusted for age at baseline (y), total energy intake (kJ), body mass index (kg/m^2), current cigarette smoking (yes/no), number of cigarettes smoked per day, number of years of smoking regularly, highest attained education (primary school; lower vocational school; junior high school; senior high school, higher vocational school; university), non-occupational physical activity (<30 min/day, 30-60 min/day, 60-90 min/day, >90 min/day), family history of cancer in first or second degree relatives (yes/no), use of dietary supplements (yes/no) and mutually adjusted for the other dietary patterns; 2 based on number of total cases in the Netherlands Cohort Study, in 9.3 years follow-up; 3: all first events, all tumors add up to 11631

To compare the pattern approach with traditional epidemiological analyses, we also calculated the risk of overall cancer for macro and micronutrients (**Table 6**). Among men, (energy adjusted) intake of alcohol was associated with an increased risk of cancer (RR_{per1SD}:1.10; 95%CI:1.03-1.18). Furthermore, there was an indication that intake of total energy, saturated fat, total carbohydrates, vitamin C and folate were associated with decreased risk of cancer in men, however, these associations were not statistically significant. In women, higher relative intake of mono- and poly-unsaturated fat was associated with decreased risk of cancer ($RR_{per\ 1\ SD}$:0.92; 95%CI:0.87-0.99 and $RR_{per\ 1\ SD}$:0.93; 95%CI:0.87-0.99 respectively), whereas there was in indication that higher intake of saturated fat and alcohol was associated with increased risk of cancer.

Table 6 Rate ratios (RRs) and 95 percent confidence intervals (95% CI) of cancer of an increment in 1 SD of intake of energy-adjusted macro- and micro-nutrients (Netherlands Cohort Study, 9.3 years of Follow-up).

	Men				Women			
	mean	SD	RR	95%CI	mean	SD	RR	95%CI
Energy and macro nutrient intake								
Total energy (MJ) [a]	9.1	2.1	0.95	0.89-1.01	7.1	1.7	0.99	0.93-1.06
Total protein [b]	76	11	1.00	0.94-1.07	66	11	0.99	0.92-1.05
Animal protein [b]	49	12	1.02	0.96-1.09	44	11	0.99	0.93-1.06
Plant protein [b]	28	6.1	0.96	0.90-1.03	22	4.6	0.97	0.91-1.04
Total fat (g/d) [b]	94	14	0.97	0.91-1.03	74	10	0.96	0.90-1.02
saturated fat (g/d) [b]	37	7.6	0.96	0.90-1.02	30	5.8	1.06	0.99-1.13
mono-unsaturated fat (g/d) [b]	35	6.7	0.97	0.91-1.03	28	5.0	0.92	0.87-0.99
poly-unsaturated fat (g/d) [b]	20	8.0	1.01	0.95-1.08	15	6.1	0.93	0.87-0.99
Total carbohydrates [b]	227	37	0.95	0.89-1.02	178	27	1.01	0.95-1.08
mono-/disaccharides (g/d) [b]	105	35	0.97	0.90-1.03	85	25	1.00	0.94-1.07
polysaccharides [b]	121	26	0.97	0.91-1.04	94	18	1.02	0.95-1.09
Dietary fiber [b]	29	7.3	0.98	0.92-1.04	25	5.8	0.97	0.91-1.03
Alcohol (g/d) [b]	15	17	1.10	1.03-1.18	5.9	9.6	1.05	0.98-1.12
Micro nutrient intake								
Vitamin C (mg/d) [b]	98	43	0.96	0.90-1.03	108	44	1.02	0.96-1.09
Beta-carotene (mg/d) [b]	3.0	1.6	1.00	1.00-1.00	2.9	1.6	1.00	1.00-1.00
Folate (μg/d) [b]	303	80	0.96	0.90-1.02	276	73	1.01	0.94-1.07

[a] adjusted for age at baseline, current cigarette smoker (yes/no), number of cigarettes smoked/day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderate active, active, very active), body mass index (kg/m^2) [b] adjusted for age at baseline, current cigarette smoker (yes/no), number of cigarettes smoked/day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderate active, active, very active), body mass index (kg/m^2) and total energy intake (kJ).

DISCUSSION

The prospective design and the completeness of follow-up of our study ensured that information bias due to disease status and selection bias due to loss of follow-up are unlikely. Although a validated semi-quantitative FFQ was used to assess the participants' diet, we have to take into account the possibility of artificial correlations in the factor analysis, e.g. due to grouping of items within the FFQ. However, spurious factors solely

due to correlated questionnaire items are unlikely as Hu *et al*[30] and Togo *et al*[31] demonstrated that qualitatively similar eating patterns were observed across FFQs and diet records. Furthermore, the items loading on a pattern were often from separate parts of the questionnaire and from questions that were formulated differently. Moreover, as the identified dietary patterns were very comparable with dietary patterns identified in other populations[21;22] and as the patterns correlated as expected with health-related sociodemographic and lifestyle variables,[32] it is unlikely that these dietary patterns were merely caused by correlated measurement error. Despite the strengths of factor analysis, the technique has been criticized for its subjective decisions.[33] However, we performed extensive sensitivity analyses and our findings showed that the number of extracted factors and other subjective factor analytical decisions did not affect the identified dietary patterns.[22]

High consumption of vegetables has been related to decreased risk of all types of cancer.[3;34;35] The cooked vegetables pattern among men and the salad vegetables pattern among women appeared to be overall protective against cancer, although the results were often not statistically significant. Our results are in agreement with other studies on dietary patterns and cancer some of which found vegetable based patterns to be associated with decreased risk of several types of cancer,[7;8;12;15;17;20] whereas others did not observe significant associations.[9-11;13;14;16;18;19] In a review, we showed that dietary patterns characterized by consumption of vegetables appeared to be overall protective against cancer, although the results were not always statistically significant [21] An explanation for the generally relatively weak associations of vegetable based dietary patterns and the lack of statistical significance, might lie in the common practice to classify the many FFQ items into less food groups prior to factor analysis. Although in our review we did not observe that studies that used a smaller number of more aggregated input variables observed weaker associations, McCann *et al.*[17] demonstrated that reducing the number of input variables and increasing the aggregation level from 168 to 56 to 36 did not affect the derived factors, but did attenuate the odds ratios of endometrial cancer for the Healthy pattern. However, as the grouping of input variables that we used was based not only on the role of foods in the diet, but also on their possible relevance to cancer etiology we expect that our associations were not diluted to a great extent.
We found that the dietary pattern characterized by consumption of pork, processed meat, potatoes, coffee and low fat margarine instead of butter was associated with increased risk of cancer in men and with a decreased risk in women. The inverse association among women appeared to be driven by breast cancer, whereas among men prostate cancer was not associated with the PPP pattern. The results of lung cancer and colorectal cancer were similar for men and women. In a review we observed that in general, dietary patterns characterized by meat and fat foods appeared to be associated with increased risk of cancer in case control studies but not in cohort studies.[21] This suggests that the association might be influenced by changes in the (recall of) diet caused by the disease. However, when we excluded cases that were diagnosed within two years of follow-up, our results were not materially changed. Even though the observed positive association among men is in agreement with traditional epidemiological studies[2;36], we have to take into account the possibility that the association may be partly due to unmeasured factors that are associated with the PPP pattern or by residual confounding due to smoking. The inverse association among women is in agreement with the positive correlation of the PPP pattern with mono- and polyunsaturated fat, which were inversely associated with cancer

risk and the negative correlation with saturated fat, which was positively associated with cancer risk. Although in pooled analyses of cohort studies on breast cancer and meat consumption[37] and intake of fat[38;39] no associations were observed, previous NLCS results did show an inverse association between mono-unsaturated fat and breast cancer risk[40].
The inverse association in men between the dietary pattern characterized by consumption of several sweet foods and overall cancer risk could not be explained by the higher consumption of mono- and disaccharides or by the lower consumption of alcohol associated with the sweet pattern. Although not statistically significant, among women a similar inverse association was observed for a dietary pattern characterized by sweet and savory snacks. Other studies observed a positive association between a sweet dietary pattern and risk of renal cell[19] and colon[7] cancer. The specific types of cancer that were most strongly associated were lung cancer and bladder cancer, both known to be strongly related to smoking.[42-44] Although the inverse association with lung cancer observed in our study appeared to be strongest in never smokers[24], the inverse association with total cancer was restricted to current smokers. Although we cannot exclude the possibility of residual confounding by smoking, either by inaccuracies in the measurement or by unmeasured aspects of smoking behavior, e.g. smoking other types of cigarettes, attempts to quit smoking, or products (not) consumed when (not) smoking, it is also possible that other unmeasured factors associated to this pattern, might be responsible for (part of) its protective effect.
Overall risk of cancer was not associated with dietary patterns characterized by consumption of brown instead of white types of bread, however among men this pattern was positively associated with risk of rectal cancer and inversely with risk of gastric cancer. Botterweck *et al* previously demonstrated that stomach cancer cases diagnosed in the first or second year after the start of the NLCS study avoided components of the hot meal (vegetables, meat, potatoes and coffee)[45] When we conducted analyses excluding cases diagnosed within the first and second year, the B/W pattern was still inversely associated with risk of gastric cancer. Consumption of "whole grain", "brown" or "wholemeal" bread has been consistently associated with decreased risk of stomach cancer.[46] This association was not observed in women, but this could be due to the small number of cases.
In conclusion the dietary patterns characterized by consumption of vegetables appeared to be overall protective against cancer, although the results were not statistically significant. In men a dietary pattern characterized by consumption of several sweet foods appeared to be inversely associated with overall cancer risk. In women a dietary pattern characterized by consumption of both sweet and savory snacks also tended towards an inverse association. This could not be attributed to the nutritional value of these patterns and is therefore most likely to be due to other dietary or non-dietary (behavioral) factors associated with this pattern. In men, a dietary pattern characterized by consumption of pork, processed meat, potatoes, coffee and low fat margarine instead of butter was associated with increased risk of cancer. In women the same dietary pattern was found to be inversely associated with overall cancer risk. This decrease in cancer risk is consistent with a decreased risk associated with intake of mono- and polyunsaturated fat. Overall risk of cancer was not associated with a dietary pattern characterized by consumption of brown instead of white types of bread or with the fat dairy pattern.
Epidemiologists may regard it unsatisfactory that the dietary pattern approach is not able to pinpoint purely dietary factors as a cause of cancer or to elucidate biological mechanisms underlying observed associations. However, unexpected associations for

dietary patterns, such as our findings of the Sweet foods pattern, or remarkable differences in strength of associations between dietary patterns and their corresponding individual foods or nutrients, demonstrate the relevance to cancer occurrence of those combinations of risk factors that tend to occur together in a free living population. Our results demonstrate that dietary patterns, other than the two most commonly identified "vegetable-based" and "meat-based" patterns are potentially relevant for cancer. Nevertheless, similar to results of the traditional epidemiological approach, these findings need to be confirmed in other studies.

REFERENCES

1. Visser O, Coebergh JWW, Dijck JAAM, Siesling S. Incidence of cancer in the Netherlands 1998. Utrecht: Vereniging van Integrale Kankercentra (Netherlands cancer registry), 2002.

2. Key TJ, Schatzkin A, Willett WC, Allen NE, Spencer EA, Travis RC. Diet, nutrition and the prevention of cancer. Public Health Nutr. 2004;7:187-200.

3. Riboli E, Norat T. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. Am.J.Clin.Nutr. 2003;78:559S-69S.

4. Norat T, Riboli E. Dairy products and colorectal cancer. A review of possible mechanisms and epidemiological evidence. Eur.J.Clin.Nutr. 2003;57:1-17.

5. Kushi L, Giovannucci E. Dietary fat and cancer. Am.J.Med. 2002;113 Suppl. 9B:63-70.

6. Kant AK. Indexes of overall diet quality: a review. J.Am.Diet.Assoc. 1996;96:785-91.

7. Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in western New York. Nutr.Cancer 1992;18:265-76.

8. Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. Am.J.Epidemiol. 1998;148:4-16.

9. Terry P, Hu FB, Hansen H, Wolk A. Prospective study of major dietary patterns and colorectal cancer risk in women. Am.J.Epidemiol. 2001;154:1143-9.

10. Fung T, Hu FB, Fuchs C et al. Major dietary patterns and the risk of colorectal cancer in women. Arch.Intern.Med. 2003;163:309-14.

11. Dixon LB, Balder HF, Virtanen M et al. Dietary patterns associated with colon and rectal cancers: Results from the DIETSCAN project. Submitted for publication 2003.

12. Palli D, Russo A, Decarli A. Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control 2001;12:163-72.

13. Masaki M, Sugimori H, Nakamura K, Tadera M. Dietary patterns and stomach cancer among middle-aged male workers in Tokyo. Asian Pac.J.Cancer Prev. 2003;4:61-6.

14. Terry P, Suzuki R, Hu FB, Wolk A. A prospective study of major dietary patterns and the risk of breast cancer. Cancer Epidemiol.Biomarkers Prev. 2001;10:1281-5.

15. Sieri S, Krogh V, Pala V et al. Dietary patterns and risk of breast cancer in the ORDET cohort. Cancer Epidemiol.Biomarkers Prev. 2004;13:567-72.

16. Mannisto S, Smith-Warner SA, Spiegelman D et al. Dietary carotenoids and risk of lung cancer in a pooled analysis of seven cohort studies. Cancer Epidemiol.Biomarkers Prev. 2004;13:40-8.

17. McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr. 2001;4:989-97.

18. Tseng M, Breslow RA, DeVellis RF, Ziegler RG. Dietary patterns and prostate cancer risk in the National Health and Nutrition Examination Survey Epidemiological Follow-up Study cohort. Cancer Epidemiol.Biomarkers Prev. 2004;13:71-7.

19. Handa K, Kreiger N. Diet patterns and the risk of renal cell carcinoma. Public Health Nutrition 2002;5:757-67.

20. Markaki I, Linos D, Linos A. The influence of dietary patterns on the development of thyroid cancer. Eur.J.Cancer 2003;39:1912-9.

21. Balder HF, van den Brandt PA, Goldbohm RA. Factor analysis for food consumption data: A review of the literature. Submitted for publication 2004.

22. Balder HF, Virtanen M, Brants HA et al. Common and Country-Specific Dietary Patterns in Four European Cohort Studies. J.Nutr. 2003;133:4246-51.

23. Mannisto S, Dixon LB, Balder HF et al. Dietary patterns and the risk of breast cancer: Results from three cohort studies in the DIETSCAN project. submitted 2003.

24. Balder HF, Goldbohm RA, van den Brandt PA. Dietary patterns associated with male lung cancer risk in the Netherlands Cohort Study. Provisionally accepted CEBP 2004.

25. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. J Clin Epidemiol 1990;43:285-95.

26. Goldbohm RA, van den Brandt PA, Brants HA et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur.J.Clin.Nutr. 1994;48:253-65.

27. Voorlichtingsbureau vd, V. NEVO table: Dutch Food Composition Table 1986-1987. The Hague: 1986.

28. van den Brandt PA, Schouten LJ, Goldbohm RA, Dorant E, Hunen PM. Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. Int.J.Epidemiol. 1990;19:553-8.

29. Willett WC, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am.J.Epidemiol. 1986;124:17-27.

30. Hu FB, Rimm E, Smith-Warner SA et al. Reproducibility and validity of dietary patterns assessed with a food- frequency questionnaire. Am.J.Clin.Nutr. 1999;69:243-9.

31. Togo P, Heitmann BL, Sorensen TIA, Osler M. Consistency of food intake factors by different dietary assessment methods and population groups. Br.J.Nutr. 2003;90:667-78.

32. Balder HF, van den Brandt PA, Goldbohm RA. Dietary patterns associated with health conscious behavior in the Netherlands. Submitted for publication 2004.

33. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity [comment] [see comments]. Am.J.Epidemiol. 1998;148:17-9.

34. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer prevention: a review. J.Am.Diet.Assoc. 1996;96:1027-39.

35. Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. Nutr.Cancer 1992;18:1-29.

36. Norat T, Riboli E. Meat consumption and colorectal cancer: A review of epidemiologic evidence. Nutr.Rev. 2001;59:37-47.

37. Missmer SA, Smith-Warner SA, Spiegelman D et al. Meat and dairy food consumption and breast cancer: a pooled analysis of cohort studies. Int.J.Epidemiol. 2002;31:78-85.

38. Hunter DJ, Spiegelman D, Adami HO et al. Cohort studies of fat intake and the risk of breast cancer--a pooled analysis [see comments]. N.Engl.J.Med. 1996;334:356-61.

39. Smith-Warner SA, Spiegelman D, Adami HO et al. Types of dietary fat and breast cancer: A pooled analysis of cohort studies. Int.J.Cancer 2001;92:767-74.

40. Voorrips LE, Brants HA, Kardinaal AF, Hiddink GJ, van den Brandt PA, Goldbohm RA. Intake of conjugated linoleic acid, fat, and other fatty acids in relation to postmenopausal breast cancer: the Netherlands Cohort Study on Diet and Cancer. Am.J.Clin.Nutr. 2002;76:873-82.

41. van den Brandt PA, Spiegelman D, Yaun SS et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am.J.Epidemiol. 2000;152:514-27.

42. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J.Natl.Cancer Inst. 1981;66:1191-308.

43. Kuper H, Boffetta P, Adami HO. Tobacco use and cancer causation: association by tumour type. J.Intern.Med. 2002;252:206-24.

44. Zeegers MPA, Tan FES, Dorant E, van den Brandt PA. The impact of characteristics of cigarette smoking on urinary tract cancer risk - A meta-analysis of epidemiologic studies. Cancer 2000;89:630-9.

45. Botterweck AAM, van den Brandt PA, Goldbohm RA. A prospective cohort study on vegetable and fruit consumption and stomach cancer risk in the Netherlands. Am.J.Epidemiol. 1998;148:842-53.

46. Jacobs DR, Jr., Slavin J, Marquart L. Whole grain intake and cancer: a review of the literature. Nutr.Cancer 1995;24:221-9.

8

General Discussion

In this thesis, the analysis of dietary patterns and their association with sociodemographic and lifestyle variables and the incidence of cancer in a prospective cohort study have been described. In this chapter, the major results are summarized and compared with results of traditional epidemiological, food- and nutrient-based analyses. Furthermore, the subjectivity of the dietary pattern approach is evaluated, the public health implications are discussed and suggestions for further research are given. Finally, I draw conclusions regarding the explanatory value of dietary patterns when studying diet and cancer.

8.1 Main findings

In ***table 1***, the main findings of the research described in this thesis are summarized. The dietary patterns published in the literature revealed many similarities, after systematically reclassifying the patterns based on the foods and food groups that loaded highest.[chapter2] These comparable patterns were identified in different countries and despite differences in data-collection methods and factor analytical decisions. A dietary pattern characterized by consumption of vegetables and/or fruits has frequently been reported to decrease risk of diabetes mellitus type II, coronary heart disease, several types of cancer and all-cause mortality. A pattern characterized by consumption of meat and fat foods appeared to be associated with increased risk of cancer in case-control studies but not in cohort studies.
Using a standardized approach across the four cohorts of the Dietscan project, three to five stable dietary patterns were identified by factor analysis.[chapter3] Two dietary patterns, "Salad vegetables" (SV) and "Pork, Processed Meat, Potatoes" (PPP), were very similar across the cohorts. In addition to these common patterns, in NLCS men also a "Cooked vegetables" (CV), "Sweet foods" (SW) and "Brown/white bread substitution" (B/W) pattern were identified, and in NLCS women a "Fat dairy" (FD), "Brown/white bread substitution" (B/W) and "Sweet and savory snacks" (SN) pattern were identified. Sensitivity analyses in the Dietscan cohorts revealed that the identified dietary patterns were robust for the number of extracted factors, dichotomization of variables, energy adjustment, and random sampling. The common dietary patterns were associated with health-related sociodemographic and lifestyle variables in the expected direction.[chapter4]
The associations between the common SV and PPP dietary patterns identified in the Dietscan project and colon and rectal cancer[chapter5] and breast cancer[1] were not consistent across the cohorts. In the NLCS, the dietary patterns characterized by consumption of vegetables appeared to be overall protective of cancer, although the results were not statistically significant.[chapter7] In NLCS men, the SW pattern appeared to be inversely associated with overall cancer risk.[chapter7] Although not statistically significant, among women a similar inverse association was observed for the SN pattern. In women the PPP dietary pattern was not only inversely associated with breast cancer risk,[1] but also with overall cancer risk,[chapter7] whereas in men this pattern was associated with increased risk of overall cancer. Although some specific types of cancer were, overall risk of cancer was not associated with the B/W and FD dietary patterns.

8.2 Dependence of results on subjective decisions of factor analysis

Factor analysis has been criticized for the subjectivity of analytical decisions made at various steps of the factor analysis.[2] However, the influence of these arbitrary decisions on the identified dietary patterns appeared not to be as great as previously thought. Aggregation of foods prior to the analysis was expected to influence the identified dietary

Table 1 Main findings of the chapters described in this thesis

	Comparability of dietary patterns	Associations with sociodemographic and lifestyle variables
Review	The dietary patterns extracted in the different studies revealed many similarities. Many studies identified a pattern characterized by consumption of:	Associations of the dietary patterns with sociodemographic and lifestyle variables comparable and in expected direction. Association with… *(chapter 2)*

	kJ	age	smok	act	vit	edu	bmi
vegetables and/or fruits (62/65)	+	o	-	+	+	+	
meat and/or fat foods (39/65)	+	+	+				
sweet foods (26/65)	+		-				
snacks (19/65)	+	-					
alcohol (12/65)							
or labeled "Traditional" (18/65)	+	+					

NLCS

Similar dietary patterns in men and women *(Chapter 3 and Chapter 7)*

Association with *(chapter 4)*

		kj	age	smok	act	vit	edu	bmi
"Salad Vegetables"	M	+	-	-	o	+	+	o
	W	+	-	o	+	+	+	o
"Pork, processed meat, potatoes"	M	+	-	+	+	-	-	+
	W	+	-	+	o	-	-	+
"Sweet foods"	M	+	+	-	o	+	+	-
"Sweet and/or savory snacks"	W	+	-	o	o	+	+	-
"Brown/white bread substitution"	M	o	o	-	+	+	o	o
	W	+	o	-	+	+	+	o
"Cooked vegetables"	M	+	+	o	+	+	-	+
"Fat dairy"	W	+	+	o	-	-	-	-

Dietscan

Using a standardized methodology 3 to 5 dietary patterns were extracted of which two were common across all four European cohorts *(Chapter 3)*:

Association with *(Unpublished results)*

			kj	age	smok	edu	bmi
"Salad vegetables"	ATBC	M	+	-			+
	NLCS	M	+	-	-	+	o
	NLCS	W	+	-	o	+	o
	SMC	W	+	-	o	+	+
	ORDET	W	+	-	o	o	+
"Pork, processed meat, potatoes"	ATBC	M	+	-			+
	NLCS	M	+	-	+	-	+
	NLCS	W	+	-	+	-	+
	SMC	W	+	-	o	o	-
	ORDET	W	+	-	+	+	-

+/- **Bold** plus, minus or null symbols indicate a clear and/or significant association, whereas symbols in regular (not bold) and smaller font represent an indication for a (non-significant) positive or negative association.

Table 1 continued

	Comparability of dietary patterns	Associations with disease
Review	dietary patterns characterized by consumption of:	The associations of the vegetable and meat pattern with chronic disease and mortality were comparable. Association with disease *(chapter 2)*

		Diabetes Mellitus-II	CHD	mortality	cancer
Review	vegetables and/or fruits (62/65)	-	-	-	-
	meat and/or fat foods (39/65)	+	+	+	+
	sweet foods (26/65)				
	snacks (19/65)				
	alcohol (12/65)				
	or labeled "Traditional" (18/65)				

	dietary patterns in men and women		Association with cancer *(Chapter 6 and Capter 7)*				
			Total	lung	colon	rectal	breast
NLCS	"Salad Vegetables"	M	o	-	o	+	
		W	-	-	o	o	o
	"Pork, processed meat, potatoes"	M	+	+	o	o	
		W	-	o	o	o	-
	"Sweet foods"	M	-	-	o	-	
	"Sweet and/or savory snacks"	W	-	-	o	-	-
	"Brown/white bread substitution"	M	o	o	o	+	
		W	o	-	o	o	o
	"Cooked vegetables"	M	o	o	o	o	
	"Fat dairy"	W	-	-	o	-	o

	Common dietary patterns:			*(Chapter 5 and Mannisto et al, 2004)*		
				colon	rectal	breast
Dietscan	"Salad vegetables"	ATBC	M	o	+	n/a
		NLCS	M	o	+	n/a
		NLCS	W	o	+	o
		SMC	W	o	o	o
		ORDET	W	n/a	n/a	-
	"Pork, processed meat, potatoes"	ATBC	M	o	+	n/a
		NLCS	M	o	o	n/a
		NLCS	W	o	o	-
		SMC	W	+	o	o
		ORDET	W	n/a	n/a	o

patterns strongly. However, in a review of the literature we found that qualitatively similar patterns were identified, regardless of the number of input variables.[chapter2] This is in agreement with McCann *et al*[3] who demonstrated that three methods of reducing detailed dietary data obtained from a single FFQ prior to pattern analysis did not affect the number or character of the identified "Healthy" and "High fat" patterns. Thus, aggregation of food groups does not appear to influence the identified dietary patterns much, although differences in input variables hampers direct comparison of the resulting patterns between studies. Similar to the number of input variables, the number of factors retained does not appear to influence the patterns much. From a review of the literature we learned that if two factors were extracted, they were most likely a vegetable-based pattern and a meat-based pattern. Extracting more patterns did not influence the previously extracted patterns, only increasing the chance of other patterns to be extracted. This is in agreement with findings of sensitivity analyses in the Dietscan project, showing that increasing the number of extracted factors did not materially change the previously extracted factors.[chapter3] One of the most subjective steps is the interpretation and labeling of dietary patterns. Therefore, it is important to show the actual factor loadings of the food items or groups in the publication. I consider it best to label a pattern according to its major constituents or to a common denominator, such as "vegetables" or "sweet foods", instead of implying more about overall diet quality, lifestyle and personality than is warranted with labels such as "Prudent", or "Western".
In conclusion, from a systematic review of the literature and from sensitivity analyses within the four cohorts of Dietscan it has become clear that the dietary pattern approach is robust for subjective factor analytical decisions.

8.3 Stability of dietary patterns

Qualitatively similar dietary patterns can be identified in different populations.[chapter2 and 3] This relative stability of dietary patterns could be explained by the possibility that emerging dietary patterns are caused by the grouping of items within food frequency questionnaires that are mostly used in epidemiological studies. However, three different studies showed qualitatively similar eating patterns across different dietary measurements (repeated FFQ's and diet records).[4-6] Therefore, we consider it very unlikely that the dietary patterns are solely caused by the structure of the dietary measurement method.
Furthermore, the consistency and stability of the dietary pattern approach has been studied by several authors. Schwerin et al[7] showed that the dietary patterns were comparable in three different nationwide food consumption surveys in the USA over a time period of 10 years. Iizumi *et al*[8] and Togo *et al*[5] also showed relatively stable dietary patterns over more than 10 years in Japan and Denmark. Dietary patterns thus appear to be consistent over time.
The finding that similar patterns (e.g. vegetable-based and meat-based) are identified[chapter2] in populations living under different social and cultural circumstances, in gender and age groups and over time suggests that to some extent food choices must be driven by endogenous factors (e.g. biological or psychological mechanisms, common to all people), whereas other patterns are more clearly the result of sociocultural circumstances and food availability.

8.3 *The role of energy adjustment in dietary pattern analysis*

Although not often specifically addressed, whether or not to adjust dietary patterns for total energy intake is an important issue in dietary pattern analysis.[9;10] Similar to nutrients, the scores on almost all dietary patterns tend to be positively correlated with total energy intake.[chapter2] These positive correlations may be partly because some persons tend to eat more of everything so that even foods (such as vegetables, tea) or nutrients (such as vitamins and minerals) without caloric value are correlated with total energy intake. In addition, persons with a high energy requirement could deliberately choose energy contributing food groups. Therefore, much of the variation in a dietary pattern score might be secondary to total energy intake. As total energy intake associated with risk of cancers[11], failure to appropriately adjust may result in lack of apparent association between dietary or in spurious associations.

Depending on our interest in total or relative quantities and the expectations with regard to the outcome variable, there are several possibilities to adjust dietary patterns for energy. Firstly, one might want to obtain dietary patterns that are unrelated to total energy intake. Several methods can be used to accomplish this, such as expressing the input variables as a percent energy contribution. However, this does not allow for non-caloric foods to be included in the analysis, such as coffee, tea, or water while these beverages are part of the dietary pattern. Another way to produce dietary patterns independent of total energy would be to use food or nutrient intake adjusted for total energy by the residual method[12] as input variables for the factor analysis. These energy-adjusted food or nutrient intakes are computed as the residuals from the (univariate) regression model with total energy intake as the independent variable and absolute nutrient intake as the dependent variable and can be viewed conceptually as measures of diet composition rather than as measures of absolute intake.

To assess the influence of energy adjustment on the derived patterns, we performed factor analysis with the food-group variables both unadjusted and adjusted for energy in the Dietscan project.[chapter3] We found that the factor solution was not dramatically affected, but that the energy adjustment forced "substitution" patterns because it requires that food groups are not correlated with total energy. This "substitution" was also clear in the factor-loading matrix of Van Dam *et al.*, [13] the only other study so far that used energy adjusted intake of food groups to obtain factors uncorrelated to total energy intake. They identified three food consumption patterns using energy adjusted food groups in the Dutch Morgen study, two of which had many negative factor loadings.

However, using the residual method on dietary input variables has some major drawbacks. The adjusted food item might not be entirely uncorrelated to energy, because other food items could be correlated to both the food item ánd to energy intake. Since energy intake is a linear combination of all the food items, a simultaneous residual approach is not possible. Finally, although for determining the association with cancer it does not seem useful to find a pattern arising because of "energy requirement" only, creating an artificial population with reduced variation in total energy intake by performing energy adjustment would ignore energy requirement as one of the dimensions of overall diet.

In so far the factor scores of dietary patterns are associated with energy, there are several ways to adjust for this in the subsequent analyses. One way is to take into account in the interpretation of the dietary pattern that (part of) the pattern may be caused by energy requirement and to adjust for total energy intake as a confounder in the etiological analyses. This approach of including both total energy intake and absolute factor scores as independent variables in multivariate models with disease is analogous (although not

Table 2 Previously published associations (RR and 95%CI, comparing the risk of major cancer sites of the highest vs. the lowest category of intake of selected food groups and nutrients in the Netherlands Cohort Study on Diet and Cancer.

Dietary exposure (NLCS)		Cancer of the						
		Lung	Breast	colon	rectum	Stomach	Prostate	Urinary tract
Dietary patterns								
"Salad vegetables" pattern	m	***0.7 (0.5-1.0)***		0.9 (0.6-1.3)	1.2 (0.7-2.1)	0.7 (0.5-1.2)	1.0 (0.8-1.4)	1.3 (0.9-2.1)
	w	0.8 (0.4-1.4)	0.8 (0.7-1.2)	0.8 (0.5-1.2)	0.6 (0.3-1.5)	1.0 (0.5-2.1)		0.5 (0.1-2.0)
"Cooked vegetables" pattern	m	0.9 (0.7-1.3)		0.8 (0.5-1.1)	0.7 (0.4-1.3)	0.9 (0.6-1.4)	1.0 (0.8-1.3)	1.2 (0.7-1.8)
"Pork, processed meat, potatoes"	m	1.3 (0.9-1.8)		1.1 (0.7-1.7)	1.5 (0.8-3.0)	***1.8 (1.1-3.2)***	1.1 (0.8-1.5)	1.3 (0.7-2.3)
	w	1.1 (0.5-2.1)	***0.7 (0.5-0.8)***	0.9 (0.6-1.3)	1.1 (0.5-2.3)	0.6 (0.2-1.5)		0.5 (0.2-1.3)
"Sweet foods (no alcohol)"	m	***0.6 (0.4-0.9)***		1.0 (0.6-1.5)	0.8 (0.4-1.5)	0.9 (0.5-1.5)	1.1 (0.8-1.5)	***0.6 (0.4-1.0)***
"Sweet and savory snacks"	w	0.5 (0.2-1.1)	0.8 (0.6-1.2)	0.9 (0.5-1.4)	0.6 (0.2-1.4)	0.8 (0.3-2.2)		0.8 (0.2-2.9)
"Brown/white bread substitution"	m	1.0 (0.7-1.3)		1.0 (0.7-1.5)	1.4 (0.8-2.5)	0.7 (0.4-1.1)	1.1 (0.8-1.4)	0.7 (0.5-1.1)
	w	0.8 (0.5-1.5)	0.9 (0.7-1.2)	0.8 (0.6-1.2)	0.8 (0.4-1.7)	1.3 (0.6-2.7)		1.3 (0.6-2.8)
"Fat dairy"	w	***0.3 (0.1-0.8)***	0.9 (0.6-1.2)	1.0 (0.6-1.5)	0.5 (0.2-1.2)	1.2 (0.5-2.9)		1.3 (0.4-4.4)
Foods and food groups								
Total vegetables		***0.7 (0.5-1.0)***	0.9 (0.7-1.3)	m:0.9 (0.6-1.3)	m:0.9 (0.6-1.4)	0.9 (0.6-1.3)	0.8 (0.6-1.1)	0.9 (0.7-1.3)*
Total vegetables				w:0.8 (0.5-1.3)	w:1.8 (0.9-3.4)			
Cooked vegetables		0.8 (0.6-1.1)		m:0.9 (0.6-1.4)	m:1.0 (0.6-1.5)	0.8 (0.6-1.2)	0.9 (0.6-1.2)	1.0 (0.7-1.4)*
				w:0.8 (0.5-1.1)	w:1.3 (0.7-2.4)			
Raw vegetables		***0.7 (0.6-1.0)***		m:0.8 (0.5-1.2)	m:0.9 (0.6-1.5)	1.0 (0.6-1.5)	1.0 (0.7-1.3)	1.1 (0.8-1.5)*
				w:1.0 (0.7-1.5)	w:1.2 (0.7-2.3)			
Total fruits		0.8 (0.6-1.1)	0.8 (0.5-1.1)	m:1.3 (0.9-2.0)	m:0.9 (0.6-1.3)	1.0 (0.6-1.5)	***1.3 (1.0-1.8)***	0.7 (0.5-1.0)*
				w:0.7 (0.5-1.1)	w:0.7 (0.3-1.3)			
Total (fresh) meat			1.0 (0.7-1.3)	0.9 (n/a)			1.1 (0.8-1.5)	
				0.9 (n/a)				
Processed meat			0.9 (0.7-1.3)	1.4 (n/a)		1.4 (0.9-2.1)	***1.4 (1.0-1.9)***	
				1.4 (n/a)				
Energy and Nutrient intake								
Total energy intake			1.0 (0.8-1.4)	0.7 (0.5-1.2)			1.0 (0.7-1.4)	
Total fat			1.2 (0.9-1.6)	1.1 (0.7-1.6)			1.1 (0.8-1.5)	
Saturated fat			***1.4 (1.0-2.0)***	1.1 (0.7-1.6)			1.2 (0.9-1.7)	
Monounsaturated fat			***0.6 (0.4-1.0)***	1.0 (0.6-1.6)			1.2 (0.9-1.7)	
Polyunsaturated fat			0.9 (0.7-1.2)	1.4 (0.9-2.2)			0.8 (0.6-1.1)	
Alcohol			1.7 (0.9-3.3)	1.1 (0.3-3.6)	2.0 (0.4-9.6)		1.1 (0.8-1.6)	**m: 1.6 (1.1-2.5)**
								w: 0.8 (0.4-1.4)
Vitamin C		***0.6 (0.5-0.8)***	0.8 (0.6-1.1)			**0.7 (0.5-1.0)**	1.2 (0.9-1.5)	0.8 (0.6-1.1)
Beta-carotene		0.9 (0.6-1.2)	1.0 (0.7-1.4)			1.2 (0.8-1.8)	1.1 (0.8-1.5)	1.2 (0.8-1.6)
Folate		***0.7 (0.5-1.0)***		m:1.0 (0.7-1.5)	m:0.9 (0.6-1.6)	1.0 (0.6-1.4)		1.0 (0.7-1.4)
				w:0.7 (0.5-1.1)	w:1.3 (0.7-2.5)	1.0 (0.6-1.4)		1.0 (0.7-1.4)

identical) to energy adjustment of the factor scores themselves. A possible disadvantage of the latter method is that the factors might loose their statistical advantage of maximized variance and orthogonality. Although both approaches have advantages and disadvantages, I prefer using energy intake as a covariate in the cancer analyses, because the model can be easily interpreted, e.g. by comparing subjects with the same total energy intake, having lower or higher scores on a certain factor and hence have his/her diet fit more or less to a certain dietary pattern.

8.4 Comparing the dietary pattern approach with food or nutrient based analyses

The association between dietary intake and incidence of cancer in the Netherlands Cohort Study has been subject of numerous publications. This provided an opportunity to evaluate whether the overall dietary patterns are stronger predictors of outcome compared to individual nutrients or foods. In table 2 the associations of cancer with both dietary patterns[chapter7] and selected food groups and nutrients from publications of the Netherlands Cohort Study are presented. [14-34]

In general, dietary patterns were not more strongly associated with risk of cancer than food groups or nutrients. Overall, the SV and CV patterns did not show stronger inverse associations with risk of cancer than the corresponding food groups or vitamin C. In general, the associations of the vegetable patterns with several types of cancer were less clear, despite the maximized amount of total variance described by the patterns. We concluded previously[chapter6] that the total or raw vegetables groups play a major role in the association with (lung) cancer.

For men, the "pork, processed meat, potatoes" pattern did not show a stronger positive association with cancer risk than processed meat alone. Only for stomach cancer a higher risk estimate was observed for the pattern, compared to the food group, however, this association was attenuated after excluding cases that were diagnosed within 2 years. For some specific cancer sites (colon, prostate) processed meat was associated with increased risk, where the corresponding pattern was not. This suggests either that other factors (dietary or non-dietary) associated with the pattern counter act the effect or that the association is diluted in the pattern approach by obscuring the rank ordering of subjects with regard to the food group(s) in the pattern that do have a causal relationship with cancer. In women, the "pork, processed meat, potatoes" pattern appeared to protect against breast cancer. However, the inverse association with breast cancer was stronger for monounsaturated fatty acids than for this pattern.

The only dietary pattern that seemed to be more strongly associated with cancer risk than the food groups loading on the pattern was the sweet foods (and savory snacks) pattern. This finding was unexpected and the inverse association appeared not to be explained by higher intake of mono-/disaccharides, higher consumption of fruit or lower consumption of alcohol. Considering the nutritional value of the pattern, we have no biological explanation for it. To explain the strong association with cancer, this dietary pattern appears to rank subjects with respect to another, most likely non-dietary, factor that appears to be strongly related to cancer. As the specific types of cancer showing the strongest inverse association were smoking-related, this unmeasured dimension could comprise smoking behavior not captured by smoking status, quantity and duration, e.g. smoking other types of cigarettes, attempts to quit smoking, or products (not) consumed when (not) smoking. However, it is also possible that other (unmeasured) factors are associated with both this dietary pattern and with etiology of cancer and are hence responsible for (part of) the

protective effect. One might think of biological factors such as hormonal balance or personal characteristics such as stress handling.
Although the comparison between the dietary pattern approach and the single foods or nutrients approach was based on the same population using the same dietary measurement, the food and nutrient analyses are mostly based on data from incident cases from 6.3 years of follow up or less, whereas the dietary pattern results are based on incident cases from 9.3 years of follow-up. As the pattern analyses were based on more cases, these analyses have more statistical power. However, the associations with cancer were generally modest and not stronger than the food based analyses.

8.5 *Dietary patterns and chronic disease*

The dietary patterns that appear to be universal across populations were not consistently associated with risk of cancer. From a review of the literature,[chapter2] the dietary patterns characterized by consumption of vegetables appeared to be associated overall with decreased risk of several types of cancer, although the risk estimates of individual studies often did not reach significance.[10] A pattern characterized by consumption of "meat and fatty foods", appeared to be associated with increased risk of several types of cancer in case-control studies but not in cohort studies. However, the two common patterns identified in the Dietscan project were not consistently associated with colorectal or breast cancer across the four cohort studies.[chapter5;1] There are several explanations for the inconsistent associations of dietary patterns with cancer. One possible explanation might be variation in the occurrence of unmeasured environmental and genetic component causes between populations, determining the relative role of a dietary factor to complete a sufficient cause for cancer and thereby the strength of its associations. [35] Another explanation might be the limited amount of total variance explained by the factors. Although the factor analytical process maximizes the explained variance in the dietary input variables and the dietary patterns explain by definition more variation in intake than the original food variables, it might be too little to explain disease.[9] However, McCann et al.[3] have demonstrated that increasing the amount of explained variance (by collapsing foods into groups) attenuated the association with endometrial cancer, even though qualitatively similar pattern were extracted.[3] Therefore we consider it unlikely that the inconsistent associations are due to too little variation within the populations.
Several publications suggest that dietary patterns are more consistently associated with risk of diabetes mellitus type II, coronary heart disease and all-cause mortality.[chapter2] However, it is too early to conclude that diet is more important in the etiology of these diseases or that adult diet is not that important in the etiology of cancer, after all. Although the etiology of these diseases might be less complex than that of cancer, the apparently stronger associations could also be confounded by publication bias. After all, the early studies reporting on the association between vegetables and fruits and cancer also showed stronger associations than the ones published later.
Finally, the inconsistent results on diet and cancer might be due to the inability of dietary measurement methods used in nutritional epidemiology to capture enough aspects of the diet. For dietary pattern analysis holds: "garbage in, garbage out". Factor analysis can not make up for what was not measured well enough or not measured at all.

8.6 Public health considerations: potential for cancer prevention

The dietary pattern approach will not elucidate etiological mechanisms through which the diet affects our health. However, for public health purposes, it is not necessary to wait until all etiological mechanisms are elucidated before starting health promoting activities and interventions. A frequently used argument in favor of the dietary pattern approach is that dietary patterns are closer to behaviors that can be modified than nutrient intakes, and that studies on dietary patterns are more constructive for the design of successful nutrition-intervention programs.[7;36;37] However, the common dietary patterns were not consistent and generalisable predictors of risk of cancer. We have to conclude that it is not possible to pinpoint purely dietary factors as cause of cancer that can be translated into specific dietary advise in addition to the current general health recommendations for diet.[11]
However, the observation that dietary patterns appear to be quite universal is also important for policy makers and health educators, as it implies that food choices apparently are not only driven by circumstantial and exogenous factors, but also by apparently universally existing endogenous factors.

8.7 Suggestions for further research

The results presented in this thesis do not support the hypothesis that those patterns contributing most to the variance in dietary intake are strongly related to risk of cancer. Exploratory factor analysis aims at maximizing the variance of the linear combination of the original variables and might therefore be particularly useful for increasing the variation in the exposure variables, which is very important in analytical epidemiology. However, this does not necessarily increase the variation in factors that may cause or prevent cancer. Structural equation modeling would be a useful method to reflect complex relations between nutrition and continuous variables describing diseases and symptoms. Unfortunately, structural equation models are not suitable to involve dichotomous variables since they are based on ordinary regression models. Thus, methods such as logistic or proportional hazards regression cannot be involved in the complex structural equations model context. Currently it is a major challenge for biostatisticians to develop a method that combines maximizing variance in food groups with maximizing the variance in disease occurrence. Hoffman et al[38] applied an elegant new statistical approach: reduced rank regression (RRR), by choosing nutrients as disease-specific response variables and determining combinations of food intake that explain as much response variation as possible. The usefulness of this method, especially for cancer research should be further explored.
Furthermore, a continued comparative evaluation of patterns derived from dietary indices, cluster analysis and factor analysis on the one hand and traditional epidemiological food and nutrient-based analysis on the other hand is needed. Finally, further analyses are needed, within the NLCS and/or other Dietscan cohorts, to relate dietary patterns to mortality, to determine whether dietary patterns indeed are more strongly associated with other major causes of death, such as cardiovascular disease.

8.8 Conclusion: the explanatory value of dietary patterns

In conclusion, factor analysis appears to be successful in identifying meaningful and robust, partly universal underlying patterns in food consumption in different populations. The identified dietary patterns represent combinations of foods that describe as much

variance of the original dietary variables as possible, but not necessarily those combinations of foods that show the strongest associations with the disease under study. The observed associations between dietary patterns and risk of cancer generally were relatively modest and appeared not to be stronger than the associations observed for individual components that were used to construct the patterns.
In its current form the dietary pattern approach complements the traditional reductionist approach of studying single foods or nutrients by enabling to study the overall effect of complex combinations of dietary exposures as they prevail in the population under study. The apparent lack of associations and sometimes seemingly inconsistent results from these and other studies cannot be explained by confounding by other dietary factors. Our results suggest that the common prevailing patterns in adult diet may not be strongly associated to risk of cancer.

REFERENCES

1. Mannisto S, Dixon LB, Balder HF et al. Dietary patterns and the risk of breast cancer: Results from three cohort studies in the DIETSCAN project. submitted 2003.

2. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity. Am.J.Epidemiol. 1998;148:17-9.

3. McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr. 2001;4:989-97.

4. Hu FB, Rimm E, Smith-Warner SA et al. Reproducibility and validity of dietary patterns assessed with a food- frequency questionnaire. Am.J.Clin.Nutr. 1999;69:243-9.

5. Togo P, Heitmann BL, Sorensen TIA, Osler M. Consistency of food intake factors by different dietary assessment methods and population groups. Br.J.Nutr. 2003;90:667-78.

6. Rashidkhani B, Ye M, Terry P, Wolk A. Major dietary patterns: Validity and reproducibility among Swedish women. submitted 2003.

7. Schwerin HS, Stanton JL, Smith JL, Riley AM, Jr., Brett BE. Food, eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. Am.J.Clin.Nutr. 1982;35:1319-25.

8. Iizumi H, Amemiya T. Eleven-year follow-up of changes in individuals' food consumption patterns. Int.J.Vitam.Nutr.Res. 1986;56:399-409.

9. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev. 2004 May;62(5):177-203.

10. Kant AK. Dietary patterns and health outcomes. J Am Diet Assoc. 2004 Apr;104(4):615-35.

11. Signaleringscommissie Kanker van KWF Kankerbestrijding. De rol van voeding bij het ontstaan van kanker. Amsterdam: KWF Kankerbestrijding, 2004.

12. Willett WC, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am.J.Epidemiol. 1986;124:17-27.

13. Dam, R. M. v., Grievink, L., Ocke, M. C., and Feskens, E. J. Patterns of food consumption and risk factors for cardiovascular disease in the general Dutch population. Am J Clin Nutr 77(5), 1156-63. 2003.

14. Goldbohm RA, van den Brandt PA, van 't Veer P et al. A prospective cohort study on the relation between meat consumption and the risk of colon cancer. Cancer Res. 1994;54:718-23.

15. Goldbohm RA, van den Brandt PA, van 't Veer P, Dorant E, Sturmans F, Hermus RJ. Prospective study on alcohol consumption and the risk of cancer of the colon and rectum in the Netherlands. Cancer Causes Control 1994;5:95-104.

16. Bausch-Goldbohm RA, Voorrips LE, Bos G, van den Brandt PA. Wel of geen relatie tussen vlees en dikkedarmkanker? Nieuw licht op tegenstrijdige resultaten. Voeding Nu 2001;3:9-12.

17. Botterweck AAM, van den Brandt PA, Goldbohm RA. A prospective cohort study on vegetable and fruit consumption and stomach cancer risk in the Netherlands. Am.J.Epidemiol. 1998;148:842-53.

18. Botterweck AAM, van den Brandt PA, Goldbohm RA. Vitamins, carotenoids, dietary fiber, and the risk of gastric carcinoma - Results from a prospective study after 6.3 years of follow-up. Cancer 2000;88:737-48.

19. Schuurman AG, Goldbohm RA, Dorant E, van den Brandt PA. Vegetable and fruit consumption and prostate cancer risk: A cohort study in the Netherlands. Cancer Epidemiol.Biomarkers Prev. 1998;7:673-80.

20. Schuurman AG, van den Brandt PA, Dorant E, Brants HAM, Goldbohm RA. Association of energy and fat intake with prostate carcinoma risk - Results from the Netherlands Cohort Study. Cancer 1999;86:1019-27.

21. Schuurman AG, van den Brandt PA, Dorant E, Goldbohm RA. Animal products, calcium and protein and prostate cancer risk in the Netherlands Cohort Study. Br.J.Cancer 1999;80:1107-13.

22. Schuurman AG, Goldbohm RA, van den Brandt PA. A prospective cohort study on consumption of alcoholic beverages in relation to prostate cancer incidence (The Netherlands). Cancer Causes Control 1999;10:597-605.

23. Schuurman AG, Goldbohm RA, Brants HAM, van den Brandt PA. A prospective cohort study on intake of retinol, vitamins C and E, and carotenoids and prostate cancer risk (Netherlands). Cancer Causes Control 2002;13:573-82.

24. van den Brandt PA, van 't Veer P, Goldbohm RA et al. A prospective cohort study on dietary fat and the risk of postmenopausal breast cancer. Cancer Res. 1993;53:75-82.

25. van den Brandt PA, Goldbohm RA, van 't Veer P. Alcohol and breast cancer: results from The Netherlands Cohort Study. Am.J.Epidemiol. 1995;141:907-15.

26. van den Brandt PA, Botterweck AA, Goldbohm RA. Salt intake, cured meat consumption, refrigerator use and stomach cancer incidence: a prospective cohort study (Netherlands). Cancer Causes Control 2003;14:427-38.

27. Verhoeven DT, Assen N, Goldbohm RA et al. Vitamins C and E, retinol, beta-carotene and dietary fibre in relation to breast cancer risk: a prospective cohort study. Br.J.Cancer 1997;75:149-55.

28. Voorrips LE, Goldbohm RA, Verhoeven DTH et al. Vegetable and fruit consumption and lung cancer risk in the Netherlands Cohort Study on diet and cancer. Cancer Causes Control 2000;11:101-15.

29. Voorrips LE, Goldbohm RA, Brants HAM et al. A prospective cohort study on antioxidant and folate intake and male lung cancer risk. Cancer Epidemiol.Biomarkers Prev. 2000;9:357-65.

30. Voorrips LE, Goldbohm RA, van Poppel G, Sturmans F, Hermus RJJ, van den Brandt PA. Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study - The Netherlands Cohort Study on Diet and Cancer. Am.J.Epidemiol. 2000;152:1081-92.

31. Voorrips LE, Brants HA, Kardinaal AF, Hiddink GJ, van den Brandt PA, Goldbohm RA. Intake of conjugated linoleic acid, fat, and other fatty acids in relation to postmenopausal breast cancer: the Netherlands Cohort Study on Diet and Cancer. Am.J.Clin.Nutr. 2002;76:873-82.

32. Zeegers MPA, Volovics A, Dorant E, Goldbohm RA, van den Brandt PA. Alcohol consumption and bladder cancer risk: Results from the Netherlands cohort study. Am.J.Epidemiol. 2001;153:38-41.

33. Zeegers MP, Goldbohm RA, van den Brandt PA. Are retinol, vitamin C, vitamin E, folate and carotenoids intake associated with bladder cancer risk? Results from the Netherlands Cohort Study. Br.J.Cancer 2001;85:977-83.

34. Zeegers MP, Goldbohm RA, van den Brandt PA. Consumption of vegetables and fruits and urothelial cancer incidence: a prospective study. Cancer Epidemiol.Biomarkers Prev. 2001;10:1121-8.

35. Rothman KJ, Greenland S. Causation and Causal Inference. In: Rothman KJ, Greenland S, eds. Modern Epidemiology. NewYork: Lippincott Williams & Wilkins Publishers 1998.

36. Nicklas TA, Webber LS, Thompson B, Berenson GS. A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors: the Bogalusa Heart Study. Am.J.Clin.Nutr. 1989;49:1320-7.

37. Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in western New York. Nutr.Cancer 1992;18:265-76.

38. Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Am J Epidemiol 2004;159:935-44.

Summary

Cancer is a major public health problem. Epidemiological and other studies suggest that diet might be an important factor in the etiology of different types of cancer. The effects of numerous food items and nutrients have been investigated, but to date, the reported evidence for the relation between dietary habits and a number of cancer types is not as clear-cut and consistent across studies as one would wish for causal inference. Examination of dietary patterns that allows to study the impact of many foods and their interactions simultaneously may give more insight into the relation between diet and different types of cancer than analysis of single foods or nutrients. The main goal of the research described in this thesis was to gain insight whether prevailing dietary patterns *consistently* enhance or protect against cancer.

In ***chapter* 2** the literature on dietary patterns identified by factor analysis was systematically reviewed. Despite differences in the data collection methods and subjective factor analytical decisions, the dietary patterns extracted in the various studies revealed many similarities. Moreover, the association of the dietary patterns with sociodemographic and lifestyle characteristics and chronic disease were comparable and in the expected direction.

Within four cohort studies originating from Finland (the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; ATBC), the Netherlands (the Netherlands Cohort Study on diet and cancer; NLCS), Sweden (the Swedish Mammography Cohort; SMC), and Italy (the Ormoni e Dieta nella Eziologia dei Tumori; ORDET) three to five stable dietary patterns were identified using a common (exploratory) factor-analytical approach (***chapter* 3**). Two observed dietary patterns were consistent across the studies: the first pattern was characterized by high consumption of salad vegetables, the second by high consumption of pork, processed meat and potatoes. In addition to the common dietary patterns, a few specifically country-specific food patterns were identified. Sensitivity analyses showed that the identified patterns were robust for the extracted number of factors, distribution of input variables, and energy adjustment.

The aim of ***chapter* 4** was to describe the cross-sectional association of sociodemographic and lifestyle variables with the dietary patterns among men and women in the Netherlands Cohort Study. The adherence to a specific dietary pattern can be indicated by a score. Each participant had a unique score for each dietary pattern. High scores represented high intake of foods within the corresponding dietary pattern; low scores represented low intake of those foods. In general, the observed associations between the dietary patterns and health-conscious or less favorable characteristics were in the expected direction. High scores on the "*Salad Vegetables*" pattern were associated with higher education, physical activity, not smoking, and the use of dietary supplements. Higher scores on the "*Pork, processed meat, potatoes*" pattern were associated with lower education, higher BMI, smoking and not using dietary supplements. Dietary patterns characterized by consumption of brown instead of white bread among men and women, by sweet snacks among men, or by sweet and savory snacks among women were inversely associated with smoking.

In chapters 5 through 7 the associations between dietary patterns and incidence of several types of cancer are described. In the three cohorts (ATBC, NLCS, SMC) that determined the association, the common "*Salad Vegetables*" and "*Pork, processed meat, potatoes*" patterns were not consistently associated with risk of colon and rectal cancer (***chapter* 5**). In ***chapter* 6**, the association of the identified dietary patterns and male lung cancer within the NLCS were evaluated. The "*Salad Vegetables*" pattern was associated with decreased

risk of lung cancer. The "*Sweet*" pattern was also inversely associated with lung cancer risk. However, the higher intake of mono- and disaccharides and fruits and lower consumption of alcohol associated with this pattern could not account for its full protective effect. Furthermore, as the association appeared to be strongest in never smokers, it is unlikely that the observed association is entirely due to residual confounding by smoking. The "*Pork, processed meat, potatoes*" pattern was non-significantly associated with increased risk.

In ***chapter 7***, the association of dietary patterns with overall and some of the most frequently occurring cancers within the NLCS were described and compared with nutrient-based analyses. The dietary patterns characterized by consumption of vegetables appeared to be overall protective for cancer, although the results were not statistically significant. The "*Sweet*" pattern in men and the "*Sweet and/or savory snacks*" pattern in women appeared to be inversely associated with overall cancer risk, which could not be attributed to the nutritional value of these patterns. The "*Pork, processed meat, potatoes*" pattern was found to increase overall cancer risk in men and to decrease overall cancer risk in women. The inverse association among women appeared to be driven by breast cancer, whereas among men prostate cancer was not associated with the "*Pork, processed meat, potatoes*" pattern. The results of lung cancer and colorectal cancer were similar for men and women. The decrease in cancer risk among women is consistent with a decreased risk associated with higher intake of mono- and polyunsaturated fat and lower intake of saturated fat. Overall risk of cancer was not associated with the other dietary patterns.

n ***chapter 8***, the major results were summarized and compared with results of food- and utrient-based analyses. Finally, the explanatory value of dietary patterns when studying t and cancer was evaluated. In conclusion, factor analysis appears to be successful in ntifying meaningful and robust, partly universal underlying patterns in food consumption different populations. The identified dietary patterns represent combinations of foods t describe as much variance of the original dietary variables as possible, but not essarily those combinations of foods that show the strongest associations with the ase under study. The observed associations between dietary patterns and risk of cer generally were relatively modest and appeared not to be stronger than the ciations observed for individual components that were used to construct the patterns. current form, the dietary pattern approach complements the traditional reductionist oach of studying single foods or nutrients by enabling to study the overall effect of lex combinations of dietary exposures as they prevail in the population under study. pparent lack of associations and sometimes seemingly inconsistent results from and other studies cannot be explained by confounding by other dietary factors. Our suggest that the common prevailing patterns in adult diet may not be strongly ted to risk of cancer.

Samenvatting

Kanker is een groot probleem voor de volksgezondheid. Epidemiologische en andere onderzoeken wijzen er op dat voeding een belangrijke factor in het ontstaan (de etiologie) van verschillende soorten kanker zou kunnen zijn. Tot nu toe zijn talrijke voedingsmiddelen, voedingsmiddelengroepen en voedingsstoffen onderzocht, maar de bewijzen uit deze onderzoeken zijn niet zo duidelijk dat we er harde conclusies uit kunnen trekken wat betreft een oorzakelijk verband. Het bestuderen van voedings*patronen* stelt ons in staat om de invloed van vele voedingsmiddelen tegelijkertijd te onderzoeken. Dit zou meer inzicht kunnen geven in het verband tussen voeding en verschillende soorten kanker, dan het bestuderen van enkele afzonderlijke voedingsstoffen of voedingsmiddelen. Het hoofddoel van het in dit proefschrift beschreven onderzoek was het verkrijgen van meer inzicht in de vraag of bepaalde voedings*patronen* consistent kanker kunnen veroorzaken of er juist tegen kunnen beschermen.

In ***hoofdstuk 2*** is de vakliteratuur samengevat over voedingspatronen die zijn bepaald met behulp van een statistische techniek genaamd 'factoranalyse'. De voedingspatronen die in verscheidene studies gevonden werden vertoonden veel overeenkomsten, ondanks de verschillende onderzoeksmethoden die gebruikt waren of bijvoorbeeld subjectieve beslissingen die genomen waren bij het uitvoeren van de factoranalyse. Bovendien bleken de verbanden van de gevonden voedingspatronen met leefstijlfactoren en chronische ziekten in de tot nu toe gepubliceerde vakliteratuur vergelijkbaar en in de verwachte richting te zijn.
Met behulp van factoranalyse hebben we voedingspatronen bepaald in vier grote cohortonderzoeken, uit Finland (de ATBC studie), Nederland (de Nederlandse Cohort Studie naar voeding en kanker, de NLCS), Zweden (de SMC studie) en Italië (de ORDET studie)(***Hoofdstuk 3***). In deze cohortonderzoeken werden drie tot vijf stabiele voedingspatronen gevonden, waarvan er twee vergelijkbaar waren in alle onderzoeken: het eerste patroon werd gekenmerkt door consumptie van vooral salade groenten (*"Salad vegetables"*), het tweede patroon werd gekenmerkt door consumptie van varkensvlees, vleeswaren en aardappelen *("Pork, potatoes, processed meat")*.
Behalve die gemeenschappelijke voedingspatronen werden in elk onderzoek nog een aantal andere, landspecifieke, voedingspatronen gevonden. Met behulp van zogenaamde sensitiviteitsanalyses hebben we aangetoond dat de gevonden voedingspatronen niet afhankelijk waren van bijvoorbeeld het aantal patronen, van de verdeling van de oorspronkelijke voedingsvariabelen of van correctie voor totale energie-inneming.
Het doel van ***hoofdstuk 4*** was om voor mannen en vrouwen in de Nederlandse NLCS studie het verband te beschrijven tussen de voedingspatronen en bepaalde leefstijlfactoren, zoals leeftijd, opleidingsniveau, lichamelijke activiteit, 'body mass index' (een maat voor overgewicht), rookgewoonten en het gebruik van voedingssupplementen. De mate waarin iemands voeding binnen een bepaald patroon past kan worden uitgedrukt in een score. Ieder persoon heeft een score voor elk van de voedingspatronen. Een hoge score geeft een hoge consumptie van de kenmerkende voedingsmiddelen(groepen) aan, een lage score geeft een lage consumptie van die voedingsmiddelen(groepen) aan. Over het algemeen waren de gevonden verbanden tussen de voedingspatronen en gezondheidsbewuste of juist minder gunstige leefstijlfactoren in de verwachte richting. Een hoge score op het "*Salade groenten*" patroon was geassocieerd met een hogere opleiding, meer lichamelijke activiteit, met niet-roken en het gebruik van voedingssupplementen. Een hogere score op het "*Varkensvlees, vleeswaren en aardappelen*" patroon was geassocieerd met een lager opleidingsniveau, een hogere body mass index, roken en het

minder gebruiken van voedingssupplementen. Voedingspatronen die gekenmerkt werden door de consumptie van bruin brood in plaats van wit brood *("Brown/white bread substitution")* bij mannen en vrouwen, door zoete voedingsmiddelen *("Sweet")* bij mannen, of door zowel zoete als hartige snacks *("Sweet and/or savory snacks")* bij vrouwen waren omgekeerd geassocieerd met roken.
In de hoofdstukken 5 tot en met 7 worden de verbanden tussen de voedingspatronen en het voorkomen van diverse soorten kanker beschreven. In de cohort studies uit Finland (ATBC), Nederland (NLCS) en Zweden (SMC) werd geen consistent verband gevonden tussen de "*Salade groenten*" en de "*Varkensvlees, vleeswaren en aardappelen*" patronen en de kans op dikke darm kanker (***hoofdstuk 5***).
In ***hoofdstuk 6*** wordt het verband tussen de voedingspatronen en longkanker bij mannen in de Nederlandse NLCS studie beschreven. Een hogere score op het "*Salade groenten*" patroon hing samen met een verlaagd risico op longkanker. Ook een hogere score op het "*Zoet*" patroon hing samen met een verlaagd risico op longkanker, maar dit verband kon niet verklaard worden door de hogere inneming van suikers of fruit of de lagere inneming van alcohol binnen dit patroon. Daar het verband ook in de groep niet-rokers te zien was, kan het verband ook niet geheel toegeschreven worden aan het feit dat de personen met een hoge score op het "*Zoet*" patroon minder rookten. Een hogere score op het "*Varkensvlees, vleeswaren en aardappelen*" patroon leek de kans op longkanker te verhogen, maar dit verband was niet statistisch significant.
In ***hoofdstuk 7*** werd het verband tussen voedingspatronen en de meest voorkomende soorten kanker en totaal kanker beschreven in de NLCS en vergeleken met de verbanden met een aantal voedingsstoffen. De voedingspatronen die gekenmerkt werden door consumptie van groenten leken te beschermen tegen kanker, al waren deze resultaten niet statistisch significant. Het "*Zoet*" patroon bij mannen en het "*Zoete en hartige snacks*" patroon bij vrouwen bleek ook te beschermen tegen kanker, en deze verbanden konden niet verklaard worden uit de voedingskundige waarde van deze patronen. Het "*Varkensvlees, vleeswaar en aardappelen*" patroon bleek in mannen de kans op kanker te verhogen en in vrouwen de kans op kanker te verlagen. De verschillende resultaten voor mannen en vrouwen kunnen verklaard worden uit het feit dat bij vrouwen voornamelijk borstkanker voorkomt en bij mannen voornamelijk long- en prostaatkanker. Deze verlaagde kans op kanker bij vrouwen komt overeen met het effect van een hogere inneming van enkelvoudig- en meervoudig onverzadigd vet en een lagere inneming van verzadigd vet. De kans op alle soorten kanker samen was niet geassocieerd met de andere voedingspatronen.
In ***hoofdstuk 8*** worden de belangrijkste resultaten samengevat en besproken en vergeleken met de resultaten van analyses gebaseerd op voedingsmiddelen en voedingsstoffen. Tot slot wordt de bruikbaarheid geëvalueerd van voedingspatronen als men het verband tussen voeding en kanker wil bestuderen.
Concluderend blijkt factoranalyse succesvol in het bepalen van betekenisvolle en robuuste, deels universele onderliggende patronen in voedselconsumptie in verschillende populaties. Echter, de gevonden voedingspatronen vertegenwoordigen combinaties van voedingsmiddelen die zo veel mogelijk variatie in de oorspronkelijke voedingsvariabelen beschrijven, maar niet noodzakelijkerwijs die combinaties die het sterkste verband met de bestudeerde aandoening laten zien. De gevonden verbanden tussen de voedingspatronen en de kans op kanker waren over het algemeen relatief zwak en bleken niet sterker te zijn dan de verbanden met de afzonderlijke componenten waaruit de patronen bestonden. In de huidige vorm vormt het bestuderen van voedingspatronen een aanvulling op de

traditionelere aanpak waarbij slechts enkele voedingsmiddelen of voedingsstoffen tegelijk worden bestudeerd, doordat het daarmee mogelijk wordt om het effect te bestuderen van complexe combinaties van blootstellingen uit de voeding, zoals die voorkomen in de populatie. Het ogenschijnlijke gebrek aan verbanden en in sommige gevallen de schijnbaar tegenstrijdige resultaten van deze en andere onderzoeken kunnen niet verklaard worden door andere voedingsfactoren.
De resultaten van dit onderzoek wijzen erop dat de veelvoorkomende patronen in de voeding van volwassenen niet sterk geassocieerd zijn met de kans op kanker.

Dankwoord

- "Would you tell me please, which way I ought to go from here?" said Alice.
- "That depends a great deal on where you want to get to," said the Cat
- "I don't much care where--," said Alice.
- "Then it doesn't much matter which way you go," said the Cat.
- "so long as I get somewhere," added Alice as an explanation.
- "Oh, you are sure to do that," said the Cat, "if you only walk long enough."

Alice in Wonderland, L. Caroll

"Everything that has a beginning, also has an end." (The Matrix)

Zo ook mijn AIO-schap. En dit proefschrift natuurlijk ook.
Zo'n proefschrift eindigt meestal met een dankwoord. De aftiteling, de "credits", zeg maar. Van films worden de credits nauwelijks gelezen. Van de meeste proefschriften is het dankwoord één van de meest gelezen hoofdstukken...

Allereerst wil ik Sandra Bausch, mijn copromotor, bedanken. Sandra, ik heb het gevoel dat we vanaf ons eerste gesprek op dezelfde golflengte zaten. Ondanks dat je het altijd razend druk had was er altijd wel ruimte voor een motiverend gesprek. Jouw "Ik ben er van overtuigd dat je het kunt" op momenten dat ik daar zelf hárd aan twijfelde deed wonderen. Vooral van onze discussies in de eindfase heb ik genoten. Ik zal je missen, maar ik wens je heel veel succes en vooral veel plezier tijdens je sabbatical in Boston! Ook mijn promotor, Piet van den Brandt wil ik bedanken voor zijn bijdrage. Piet, Maastricht is niet naast de deur, en ik vond het moeilijk de afstand te overbruggen. Maar op de achtergrond was je altijd aanwezig en, al was ik er wel eens even door uit het veld geslagen, ik heb jouw kritische kanttekeningen altijd als zeer waardevol ervaren.

Op de derde plaats komen de deelnemers en alle (ex-) medewerkers van de Nederlandse Cohort studie. Ook al ken ik de meesten niet persoonlijk: zonder jullie inspanningen aan dit unieke project was mijn onderzoek niet mogelijk geweest. Mijn dank daarvoor is groot!
In het bijzonder wil ik nog Leo Schouten hartelijk bedanken. Dankzij het NLCS-intranet werd, via de elektronische snelweg, de afstand Maastricht-Zeist stukken kleiner! Ook Nathalie Slangen en Boukje van Dijk waren voor mij vertrouwde aanspreekpunten in Maastricht. Dank jullie wel daarvoor. Boukje, jij ook veel succes met jouw proefschrift!

Al mijn collega's op TNO, door de jaren heen, Eric, Hannie, Cootje, Tineke, Michiel, Wil, Maaike, Jolanda, Cor, Marco, Carina, Astrid, Linda, Brenda, Diane, Marijke, Laura, Lucy, Marie-Agnes, Helma, Louise, Susan, Corrine, Martine, Henny, Petra, Annet, Marieke, Karin, Margje, Paul en natuurlijk Adriaan wil ik bedanken voor alle steun, voor de belangstelling die ze hebben getoond en voor alle gezellige koffie- en lunchpauzes. Ik heb me vanaf mijn eerste dag welkom en thuis gevoeld bij jullie op de afdeling. Dank jullie wel! Wendy, mijn kamergenote: wat radiosmaak betreft zijn we totaal verschillend -jij radio3 en yorin en ik radio2- maar wat de chaotische stapels op onze bureaus betreft zijn we aardig aan elkaar gewaagd. Ik vind het erg gezellig een kamer met je te delen en ik wens je heel veel succes met jouw laatste loodjes!

De Nederlandse Kankerbestrijding, het Koningin Wilhelmina Fonds (KWF) en de Europese Commissie ben ik zeer erkentelijk voor hun financiële bijdrage aan de projecten die de basis vormen van dit proefschrift. Ik hoop met mijn werk een stukje aan de oplossing van De Puzzel te hebben bijgedragen.

The partners of our European collaboration Dietscan: Alicja Wolk, Rino Bellocco, Pirjo Pietinen, Satu Mannistö, Mikko Virtanen, Vittorio Krogh, Frans Tan, Anne Hartman, and Beth Dixon: I enjoyed working with you! Very much!

Mijn wortels liggen in Wageningen. Ik wil Pieter van 't Veer bedanken dat hij me destijds de vacature voor deze AIO-plaats onder mijn neus heeft gehouden! Pieter, ik vind het leuk

dat je in mijn beoordelingscommissie zit en ik ben benieuwd wat voor prikkelende vragen je me over dit proefschrift gaat stellen. Jan Burema, jij bent altijd geïnteresseerd in en altijd trots op de student-assistenten van je SAS practica. Mijn liefde voor epidemiologische analyse en SAS in het bijzonder heeft zeker zijn oorsprong gevonden in een van jouw practica. Dankjewel voor die vonk!

Mijn ouders wil ik bedanken voor het vertrouwen dat ze altijd in mij (gehad) hebben. Pap, mam, de wetenschap dat jullie zo trots op me zijn en omdat ik jullie daarin niet wilde teleurstellen heeft me ervan weerhouden het bijltje erbij neer te gooien. Daarom draag ik dit boekje aan jullie op.
Mam, u wil ik extra bedanken voor alle donderdagen dat u op Lone heeft gepast. Ik weet dat u het graag heeft gedaan, maar dat waren waardevolle uurtjes die ik extra aan mijn proefschrift kon besteden!!! Pap, u was een beetje mijn "derde paranimf". Bedankt voor uw hulp bij de organisatie van deze Grote Dag!
Mijn broers, Co en Machiel: jullie zussie did it! Machiel, ik vond het gezellig dat je op donderdag ook vaak in Leimuiden was en ik vind het jammer dat je er vandaag niet bij kunt zijn. Mijn grote broer Co en As'ka: dank jullie wel voor jullie belangstelling. Zullen we gauw weer eens met zijn allen gaan eten?

Ook mijn schoonfamilie, Bert en Riet Helsloot en Leendert en Gunilda Valk-Helsloot, wil ik bedanken voor hun interesse. Nu deze last van mijn schouders is gevallen, komen we vast wat vaker langs! Gunilda wil ik op deze plaats nog heel erg bedanken voor het oppassen op Lone, de eerste maanden. Ik vind 't nog steeds knap: twee baby's tegelijk…

San, mijn hartsvriendin, jij hebt lief en leed met me gedeeld. We kennen elkaar al zo lang dat we aan een blik of een half woord -of soms zelfs helemaal geen woord- genoeg hebben om te begrijpen wat de ander bedoelt. Nog voor ik daadwerkelijk AIO werd hadden we dit al afgesproken, en ik ben blij dat het eindelijk zo ver is: dat je me op deze grote dag terzijde staat als mijn paranimf... =D

Dungeons-en-dragons-vrienden: Rimko en Ursula, Mireille en Remco en Jasper en Thea… of moet ik zeggen DM, Tong, Meng, Lark, Cenedra en Oriana. Mede namens Jiao: bedankt voor alle gezellige uurtjes! Ik heb genoten. Zowel binnen- als buitenspels…
Rimko, van dit stel verdien jij nog een woordje apart. Jij wist me, als ervaringsdeskundige, pep-talks te geven op de momenten dat ik die het hardste nodig had. Heel, heel, heel erg bedankt! Zonder jouw Wijze Woorden had ik het niet gered! Ik ben blij dat je me ook op de Grote Dag als paranimf wilt bijstaan.

Marla Cilley, aka the Flylady from Flylady.net, you tought me how to combine my job with the care for both myself and my family... You tought me to quit whining and how to overcome my paralyzing perfectionism. Thank you SO very much. Without you I would not be standing here today.

De medemoeders van de zwangerschapsgym, Catharien, Addie, Nicole, Céline, Karin en Saskia, ik vond het reuze gezellig om met jullie te puffen en nog veel gezelliger om gezellig te kletsen en ervaringen uit te wisselen. Dank jullie wel voor jullie belangstelling voor mijn aanstelling, werk en promotie. Het is gelukkig allemaal op zijn pootjes terecht gekomen. Addie, jij veel succes met de afronding van jouw proefschrift.

Els, Vera, Veronica, Elles en Gerrie en alle anderen van kinderdagverblijf De Kleine Wereld: bedankt voor jullie goede zorgen voor onze kleine meid! Omdat ik Lone altijd met een gerust hart bij jullie bracht, kon ik me goed op mijn werk concentreren.

Lone, lieve moeps, ook jij hebt bijgedragen aan de totstandkoming van dit proefschrift. Niet alleen doordat je zo'n makkelijk, lief, rustig en voorbeeldig kind bent, of door alle avonden en nachten die je *wel* hebt doorgeslapen, maar vooral door de drie R'en die je in ons leven hebt gebracht. Ik ben *zo* blij dat je er bent en ik hoop dat je later een beetje trots zult zijn op je mama.

En nu jij, lieve Edwin, last but certainly not least. Jij wist me altijd uit de brand te helpen: van verloren gewaande bestanden tot uitgeverij, van "motion blur" tot kaft en van "Magic Words" tot het éven ergens anders (of juist nergens) aan denken. Ik zou niet weten hoe ik het zonder jou had gered. Dank je wel! Voor alles! Met een af te ronden proefschrift naast een drukke agenda en de zorg voor huishouden en kind was tijd voor ons samen de laatste tijd nog al eens de sluitpost. Ik hoop dat daar nu verandering in komt.
Ik hou van jou. "En we leefden nog lang en gelukkig..."

The End

About the author

ABOUT THE AUTHOR

Helena Frederika Balder, Elleny, was born in Leiderdorp, the Netherlands on May 16th 1973. In 1990 she completed secondary school (HAVO) at the 'Herbert Vissers College' in Nieuw-Vennep. After two years of Higher Vocational Education in laboratory technique at the 'Internationale Agrarische Hoge School Larenstein' in Wageningen, she started her study Human Nutrition at Wageningen University. As part of this study, she started her first research project, survival analysis of the elderly of the SENECA study, at both the division of Human Nutrition and the division of Mathematics of Wageningen University. Her passion for data analysis brought her to conduct her second research project, on nutritional risk factors for p53 mutations in colorectal cancer, at the division of Epidemiology. In September 1997 she obtained her M.Sc. degree in Human Nutrition. From September 1997 until September 1999 she was appointed at the division of Human Nutrition and Epidemiology of Wageningen as a research associate to perform analyses on adipose PCB levels and risk of postmenopausal breast cancer in women of the EURAMIC multi-center study. In September 1999 she was appointed at Maastricht University to start her Ph.D. project at TNO Food and Nutrition Research in Zeist, of which the results are described in this thesis. During her Ph.D.-fellowship she joined several nutritional and epidemiological conferences and attended several courses on nutrition, epidemiology, oncology and statistics. Since May 1st 2004 she works as a junior epidemiologist at the department of Food and Chemical Risk Analysis of TNO Food and Nutrition Research in Zeist, the Netherlands.

		4	26	
		33	36	
		-6	25	
		-21	9	
		15	6	
		-3	19	
11		8	13	-2
13		-2	18	4
1		-11	28	-1
40		48	2	14
51		10	29	8
20		-1	-6	23
45		53	23	14
52		-6	0	6
21		19	0	10
14		-30	-18	-4
-1		25	-4	37
14		46	-4	-3
-18		6	7	2
43		22	-11	2
53	58	0	-4	-4
15	73	6	14	13
-16	70	6	12	-62
-1	13	-1	8	67
-1	37	-4	5	-1
4	46	30	21	33
-24	26	4	0	-29
4	8	50	20	19
-1	18	-27	13	22
-3	5	-2	53	18
-9	8	17	38	-7
1	6	0	25	-17
15	31	21	45	-10
30	-4	-9	-13	-1
2	-10	-3	31	-8
29	-4		12	-11
11	6		17	-32
7	-3		0	-32
-30	-1		-34	-1
6	13		-2	28
58	7		-29	1
7	-8			-5
0	2			0
29	9			-13
32	30			-12
-6	6			2
5	13			-13
-7	1			-34