Main nutrient patterns and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition study

Aurélie Moskal[1], Heinz Freisling*,[1], Graham Byrnes[1], Nada Assi[1], Michael T Fahey[2], Mazda Jenab[1], Pietro Ferrari[1], Anne Tjønneland[3], Kristina EN Petersen[3], Christina C Dahm[4], Camilla Plambeck Hansen[4], Aurélie Affret[5,6], Marie-Christine Boutron-Ruault[5,6], Claire Cadeau[5,6], Tilman Kühn[7], Verena Katzke[7], Khalid Iqbal[8], Heiner Boeing[8], Antonia Trichopoulou[9,10], Christina Bamia[9,10], Androniki Naska[9,10], Giovanna Masala[11], Maria Santucci de Magistris[12], Sabina Sieri[13], Rosario Tumino[14], Carlotta Sacerdote[15,16], Petra H Peeters[17,18], Bas H Bueno-de-Mesquita[19,20,21,22], Dagrun Engeset[23], Idlir Licaj[24], Guri Skeie[24], Eva Ardanaz[25,26,27], Genevieve Buckland[28], José M Huerta Castaño[27,29], José R Quirós[30], Pilar Amiano[27,31], Elena Molina-Portillo[27,32], Anna Winkvist[33,34], Robin Myte[35], Ulrika Ericson[36], Emily Sonestedt[36], Aurora Perez-Cornago[37], Nick Wareham[38], Kay-Tee Khaw[39], Inge Huybrechts[1], Konstantinos K Tsilidis[40,41], Heather Ward[41], Marc J Gunter[1] and Nadia Slimani[1]

Background: Much of the current literature on diet–colorectal cancer (CRC) associations focused on studies of single foods/nutrients, whereas less is known about nutrient patterns. We investigated the association between major nutrient patterns and CRC risk in participants of the European Prospective Investigation into Cancer and Nutrition (EPIC) study.

Methods: Among 477 312 participants, intakes of 23 nutrients were estimated from validated dietary questionnaires. Using results from a previous principal component (PC) analysis, four major nutrient patterns were identified. Hazard ratios (HRs) and 95% confidence intervals (CIs) were computed for the association of each of the four patterns and CRC incidence using multivariate Cox proportional hazards models with adjustment for established CRC risk factors.

Results: During an average of 11 years of follow-up, 4517 incident cases of CRC were documented. A nutrient pattern characterised by high intakes of vitamins and minerals was inversely associated with CRC (HR per 1 s.d. = 0.94, 95% CI: 0.92–0.98) as was a pattern characterised by total protein, riboflavin, phosphorus and calcium (HR (1 s.d.) = 0.96, 95% CI: 0.93–0.99). The remaining two patterns were not significantly associated with CRC risk.

Conclusions: Analysing nutrient patterns may improve our understanding of how groups of nutrients relate to CRC.

In 2012, colorectal cancer (CRC) was the fourth most common cancer worldwide, with ∼1 360 000 cases diagnosed (Ferlay *et al*, 2015). Dietary and other lifestyle choices play a significant role in CRC development (Bingham, 2000; Willett, 2005; World Cancer Research Fund/American Institute for Cancer Research, 2007). The panel of the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) in the Continuous Update Project (CUP) judged the evidence that foods high in dietary fibre protect against CRC, and that consumption of red meat, processed meat and alcohol (especially in men) increase the risk of developing CRC as 'convincing' (World Cancer Research Fund/American Institute for Cancer Research, 2011). Milk and calcium have been related to a probable reduction in CRC risk, whereas evidence is limited regarding a possible protective effect of folate,

selenium and vitamin D (World Cancer Research Fund/American Institute for Cancer Research, 2011).

In order to further advance our understanding of the relationship between diet and cancer development, the CUP panel recommended the use of new methods of investigation, such as patterns of diets in contrast to individual foods or nutrients (World Cancer Research Fund/American Institute for Cancer Research, 2007, 2011). Dietary patterns summarise a large number of correlated items (e.g., foods, 'food patterns', or nutrients, 'nutrient patterns') into fewer independent components capturing a large proportion of the dietary variability in a population (van Dam, 2005; O'Sullivan *et al*, 2011). This approach may be particularly relevant for multifactorial diseases such as CRC, where the aetiology possibly depends on more than a restricted list of dietary items (Hu, 2002; Jacobs and Steffen, 2003).

Most of the observational studies that have investigated the association between dietary patterns and CRC risk have focussed on food patterns, summarised in four literature reviews (Randi *et al*, 2010; Miller *et al*, 2010; Yusof *et al*, 2012; Fung and Brown, 2013). Despite notable differences in population characteristics, study design and methods used across the different studies, a plant-based diet with some dairy and fish was associated with a lower CRC risk, whereas a diet high in meats, refined grains and added sugar appeared to increase risk (Randi *et al*, 2010; Miller *et al*, 2010; Yusof *et al*, 2012; Fung and Brown, 2013). However, no studies published to date have examined associations between dietary patterns at the nutrient level and CRC risk in a large prospective cohort. Although results from pattern analyses conducted on foods are easier to translate into public health recommendations (Jacobs and Tapsell, 2007), nutrient pattern studies have several other advantages. Nutrients are universal exposures, that is, virtually everyone is exposed, and functionally not exchangeable. In contrast to food patterns, nutrient patterns characterise specific dietary habits more easily and in a more comparable way across populations and mirror better a combination of nutrients in complex biological pathways associated with diseases (Bravi *et al*, 2010; De Stefani *et al*, 2012; Moskal *et al*, 2014).

In the European Prospective Investigation into Cancer and Nutrition (EPIC) study, a large prospective cohort study across 23 centres in 10 European countries, we previously identified four main nutrient patterns based on dietary questionnaire data using principal component analysis (PCA) and successfully validated these patterns relative to standardised 24-h dietary recalls (Moskal *et al*, 2014).

Building on this previous methodological work, the main aim of this study was to investigate the relationship between four main nutrient patterns and the risk of CRC in the EPIC study. In addition, we investigated associations by anatomical subsite (colon, rectum) and location of the tumour within the colon (proximal, distal).

MATERIALS AND METHODS

Study population. The EPIC study is a multicentre prospective cohort designed to investigate the associations between diet, cancer and other chronic diseases across 10 European countries: Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom. Participants were recruited between 1992 and 1998 and included 521 330 men and women aged 35–70 years. Details on recruitment and study design have been published previously (Riboli and Kaaks, 1997; Riboli *et al*, 2002). All participants gave written informed consent, and the project was approved by ethical review boards of the International Agency for Research on Cancer and local participating centres.

Dietary data and lifestyle questionnaires. Usual diet was assessed at study baseline using validated country/centre-specific dietary questionnaires (DQs) (Riboli *et al*, 2002). In most centres, DQs were self-administered, with the exception of Greece, Ragusa (Italy), Naples (Italy) and Spain where face-to-face interviews were performed. Extensive quantitative DQs were used in northern Italy, the Netherlands, Germany and Greece that were structured by meals in Spain, France and Ragusa. Semiquantitative food-frequency questionnaires (FFQs) were used in Denmark, Norway, Naples and Umeå (Sweden). In the United Kingdom, both a semiquantitative FFQ and a 7-day record were used, whereas a method combining a short nonquantitative FFQ with a 7-day record on hot meals was used in Malmö (Sweden) (Riboli *et al*, 2002).

Individual intakes of 23 nutrients, alcohol and total energy were estimated from the baseline country-specific DQs using a common harmonised food composition database (EPIC Nutrient Database, ENDB) (Slimani *et al*, 2007; Nicolas *et al*, 2016). Supplement use (vitamins/minerals) was not included in the calculation of nutrient intakes.

Information on physical activity, history of tobacco smoking, alcohol consumption and education was collected at baseline by questionnaires. Weight and height were measured at the baseline examination in all centres except from part of Oxford and France and Norway, where weight and height were self-reported (Riboli *et al*, 2002).

Cohort follow-up and identification of CRC cases. Population cancer registries were used in Denmark, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom to identify participants with incident cancer. In France, Germany and Greece, cancer cases were identified through active follow-up, directly through study participants or next of kin, and confirmed by a combination of methods including health insurance records, cancer and pathology registries. All self-reported CRC cases were systematically verified using clinical and pathological records. Mortality data were obtained from mortality registries at the regional or national level. The participants were followed up from the date of enrollment (1992–1998) until the first date of diagnosis of cancer, death or until end of the follow-up period.

Cancer follow-up censoring dates varied among centres, ranging between 2005 and 2010, but completeness was > 98.5% across all centres. For the current study, the end point of interest was the first occurrence of primary CRC. Cancer incidence data were coded in accordance with the 10th Revision of the International Classification of Diseases (ICD-10) and the second revision of the International Classification of Disease for Oncology (ICDO-2). All incident cases of colon (C18) and rectal cancer (C20) were included. Proximal colon cancer included tumours of the caecum, appendix, ascending colon, hepatic flexure, transverse colon and splenic flexure (C18.0–18.5). Distal colon cancer included those in the descending and sigmoid colon (C18.6–18.7). Overlapping (C18.8) and unspecified (C18.9) lesions of the colon were grouped among colon cancers only. Cancer of the rectum included cancer occurring at the recto sigmoid junction (C19) and rectum (C20). Anal canal tumours were excluded.

Exclusion criteria in analysis. Among the 521 330 EPIC participants, we excluded 11 345 subjects with missing dietary or other lifestyle information and 10 241 subjects in the lowest or highest 1% of the ratio of reported total energy intake to energy requirements (estimated from age, sex and body weight). In addition, 22 432 participants were excluded because they had a prevalent cancer at any site at baseline other than non-melanoma skin cancer or were lost during the follow-up. Statistical pattern analyses were therefore conducted on 477 312 participants.

Nutrient patterns. The identification, validation and interpretation of the patterns have been described previously (Moskal *et al*,

2014). Briefly, nutrient pattern analyses were performed on individual intakes of 23 harmonised nutrients as available in the ENDB estimated from the baseline DQs (Supplementary Table S1). Alcohol consumption was considered as a lifestyle factor and was thus not included in the PCA (Moskal et al, 2014). Independence of scale of the variances and co-variances was achieved by taking the natural log of the input variables. In order to capture variability of nutrient intakes independently from variation in energy intake, nutrients (log variables) were adjusted for alcohol-free energy before applying PCA using the nutrient density method (Willett et al, 1997). Nutrient patterns were then identified by PCA using the covariance matrix. The data were sufficiently similar across centres and hence the first four PCs from the combined data captured 67% of the total variation (Moskal et al, 2014). PCA were conducted on both sexes combined (Moskal et al, 2014). Based on the interpretability of the patterns, the percentage of total variance explained and the scree-plots of eigenvalues, four PCs or nutrient patterns were retained (Johnson and Wichern, 2007). The PC loadings represent how much a variable contribute to a pattern (Supplementary Table S1). Nutrients from plant food sources such as β-carotene, vitamin C, folate and dietary fibre loaded positively on PC1, whereas nutrients typically found in animal foods such as retinol, vitamin D, vitamin B_{12}, cholesterol and saturated fatty acids loaded negatively. A variety of vitamins and minerals contributed to PC2; PC3 was characterised mainly by vitamin D; and PC4 by vitamin B_{12}, riboflavin, calcium, cholesterol, total proteins and phosphorus (Supplementary Table S1). Individual PC scores were computed from each retained PC as the sum of products of the observed variables (nutrient intakes multiplied by its loadings; Moskal et al, 2014).

Statistical analyses. The association between the individual nutrient pattern scores was investigated for CRC, as the primary objective, and then by anatomical subsite (colon, rectum) and location of the tumour within the colon (proximal, distal). Hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated using Cox proportional hazards models. Age was the primary time variable in all models; time at entry was age at recruitment and exit time was the age at whichever of the following came first: CRC diagnosis, diagnosis of a cancer other than CRC, death, emigration or the date at which follow-up was considered complete in each centre. To control for differences in questionnaires, follow-up procedures and other centre effects, all analyses were stratified by study centre. Models were also stratified by sex and age at recruitment (1-year intervals) (Ferrari et al, 2008). Nutrient patterns were modeled as continuous variables (ln (HR) per 1 s.d. increase) and as quintiles defined across the whole cohort. To test for linear trends, the medians of each quintile were entered as continuous terms in the Cox model. Graphs based on Schoenfeld residuals were used to assess proportional hazards assumptions that were satisfied.

All analyses were conducted for men and women combined as no significant interaction by sex was observed. First, models were adjusted for log-transformed (to improve normality) total energy intake (continuous, kcal per day) and mutually for each retained pattern score (minimally adjusted model). Then, in our fully adjusted model, we further adjusted for log-transformed alcohol intake (continuous, g per day, $+1$ was added before log transformation to take zero consumers into account), smoking status at baseline (never, former, smoker, unknown $n = 9733$, 2%), height (sex- and centre-specific tertiles), weight (sex- and centre-specific tertiles) and physical activity level (inactive, moderately inactive, moderately active, active, missing $n = 66\,265$, 13.9%). Further adjustment for education (none/primary school; technical/ secondary school; longer education incl. university degree; unknown $n = 10\,707$, 2.2%) and for BMI (sex- and centre-specific tertiles) instead of weight were also assessed, but led to virtually the same risk estimates (Supplementary Table S2).

Sensitivity analyses were performed by excluding CRC cases diagnosed within 2 or 5 years of follow-up and, second, by excluding underreporters of dietary energy intake according to Goldberg that has been shown to partially account for BMI-related biases (Freisling et al, 2012).

Likelihood ratio tests were used in nested models with and without multiplicative interaction terms to evaluate potential effect modification. A priori selected interaction terms entered into the statistical model were pattern scores as continuous variables with sex, total energy intake (log-continuous), alcohol intake (($+1$)log-continuous), height (continuous), weight (continuous), BMI (continuous), smoking status (never, former, smoker, unknown) and physical activity (inactive, moderately inactive, moderately active, active, missing).

The heterogeneity across countries/centres was explored by using random-effect meta-analysis (Greenland and Longnecker, 1992) and quantified by I^2 scores (Higgins and Thompson, 2002).

Statistical tests were two sided and P-values of < 0.05 were considered statistically significant. Statistical analyses were performed with SAS version 9.4 (Cary, NC, USA) and Stata version 12.1 (StataCorp, College, Station, TX, USA).

RESULTS

After a mean follow-up of 11.3 (s.d. 2.5) years, 4517 incident CRC cases were identified among the 477 312 participants and 3 763 676 person-years. This included 2869 colon cancer cases (1266 distal, 1298 proximal and 305 overlapping or unspecified) and 1648 rectal cancer cases. The distributions of CRC cases and person-years by country/centre are shown in Table 1.

Characteristics of study population. Table 2 shows the main characteristics of the study population by quintiles of the four PCs. A higher proportion of participants with higher scores on PC1 and PC2 were women, had a university degree and were never smokers (all $P < 0.001$). A higher proportion of participants with higher scores on PC3 and PC4 were men and had less than a university degree (all $P < 0.001$). Food consumption across quintiles of the four nutrient patterns differed considerably for some food groups (Table 2). For example, mean fruit and vegetable consumption more than doubled comparing the highest with the lowest quintile of PC1. Subjects with high scores on PC2 showed higher mean consumption of fruits and vegetables, fish, and milk and yoghurt. The PC3 was characterised by a high consumption of fish, whereas PC4 was characterised by a high consumption of fish, and in particular of milk and yoghurt (Table 2).

Nutrient patterns and CRC risk. The PC1 was not associated with the risk of overall CRC or subsites, with the exception of a suggestive inverse association with cancer of the distal colon, when considered as a continuous score (Table 3). The HR for cancer of the distal colon for a 1 s.d. increase in PC1 was 0.93 (95% CI: 0.86–1.00). However, there was no indication of a linear trend when we compared the quintiles of PC1 scores ($P_{trend} = 0.23$). The PC2 was inversely associated with CRC risk (HR $_{per\ 1\ s.d.\ increase}$ 0.94, 95% CI: 0.92–0.98) (Table 4). This association was confirmed when we compared the highest with the lowest quintile of PC2 (HR $_{Q5\ vs\ Q1}$ 0.88, 95% CI: 0.80–0.98, $P_{trend} = 0.02$). The PC2 score was inversely associated with cancer of the colon and proximal colon, but not with distal colon and rectal cancers (Table 4). Significant associations of PC3 with risk of CRC and with subsites were largely absent (Table 5). The PC4 was inversely associated with CRC risk (HR $_{per\ 1\ s.d.\ increase}$ 0.96, 95% CI: 0.93–0.99) (Table 6). When comparing the highest with the lowest quintile of PC4, we observed a risk reduction by 10% (HR $_{Q5\ vs\ Q1}$ 0.90, 95% CI: 0.81–0.99, $P_{trend} = 0.02$). Similar to PC2, the inverse association

Table 1. Characteristics of study subjects and CRC cases (first tumour only) in the EPIC cohort by country

Country	Study subjects		Age in years at enrolment[a]	Follow-up time in years[a]	Total person-years		CRC cases	
	Men	Women			Men	Women	Men	Women
France	0	67 385	51.5	11.8	–	703 730	–	423
Italy	14 029	30 512	50.5	11.8	162 766	351 650	173	245
Spain	15 148	24 854	48.6	12.7	186 950	305 507	185	144
UK-general population	12 346	17 161	56.8	12.3	143 264	209 220	256	223
UK-health conscious	10 506	35 382	42.1	11.4	116 256	393 550	68	181
The Netherlands	9 639	26 866	51.2	12.3	117 177	325 230	82	305
Greece	10 807	15 225	52.9	11.6	100 272	150 640	61	44
Germany	21 172	27 411	50.9	11.1	214 857	279 344	265	172
Sweden	22 309	26 375	50.9	14.1	302 244	365 021	339	313
Denmark	26 294	28 722	56.2	11.6	293 173	329 386	475	353
Norway	0	35 169	48.0	10.1	–	350 397	–	210
Total	142 250	335 062	51.5	11.7	1 636 959	3 763 675	1904	2613

Abbreviations: CRC = colorectal cancer; EPIC = European Prospective Investigation into Cancer and Nutrition.
[a]Median is given for age at enrolment and follow-up time.

for PC4 was significant for colon and proximal colon but not for distal colon and rectum cancers (Table 6).

Sensitivity analyses, effect modification and heterogeneity by country/centre. After excluding cancer cases diagnosed within the first 5 years of follow-up, the association between PC1 and cancer of the distal colon became stronger (HR $_{\text{per 1 s.d. increase}}$ 0.89, 95% CI: 0.81–0.98); results for associations between PC2–PC4 and CRC were largely unchanged (Supplementary Table S2).

Exclusion of participants with implausibly high or low total energy intake ($n = 163\,084$) led to virtually the same risk estimates (Supplementary Table S2).

No significant statistical interactions for the association of pattern scores and CRC risk, and the two subsites colon and rectum, were observed with total energy intake, alcohol intake, height, weight, BMI, smoking status and physical activity (all $P > 0.06$) (data not shown).

Although there were no significant interactions for associations of pattern scores and CRC risk, and the two subsites colon and rectum, with sex (all $P > 0.07$), results from analyses conducted separately for men and women are provided in Supplementary Table S2. In men, we observed a stronger association between PC2 and CRC (HR $_{\text{Q5 vs Q1}}$ 0.84, 95% CI: 0.71–0.99, $P_{\text{trend}} = 0.02$) as compared with the combined estimates, whereas associations were weaker in women and lost statistical significance (HR $_{\text{Q5 vs Q1}}$ 0.95, 95% CI: 0.83–1.11, $P_{\text{trend}} = 0.36$). Results for PC1, PC3 and PC4 remained virtually the same after stratification by sex.

There was an indication for moderate heterogeneity by country/ centres for PC1 ($I^2 = 43\%$, $P_{\text{heterogeneity}}$ 0.07) (Supplementary Figure S1) and PC3 ($I^2 = 42\%$, $P_{\text{heterogeneity}}$ 0.07) (Supplementary Figure S3). The forest plot for PC1 (Supplementary Figure S1) suggested that the HR for Norway differed substantially from that of the other countries. After removing Norway from the analysis, no heterogeneity was observed ($I^2 = 22\%$, $P_{\text{heterogeneity}}$ 0.23) and the estimated HR for a 1s.d. increase in PC1 was 0.97 (95% CI: 0.93–1.00). No heterogeneity by country/centre and CRC risk was observed for PC2 ($I^2 = 0\%$, $P_{\text{heterogeneity}}$ 0.88) (Supplementary Figure S2) and PC4 ($I^2 = 0\%$, $P_{\text{heterogeneity}}$ 0.77) (Supplementary Figure S4).

DISCUSSION

In this prospective analysis, we identified a nutrient pattern characterised by a high variety of vitamins and minerals (PC2) and a nutrient pattern characterised by vitamin B_{12}, riboflavin, calcium, cholesterol, total proteins and phosphorus (PC4) to be independently and inversely associated with CRC risk. Associations for both PC2 and PC4 appear to be more pronounced for cancers of the colon, in particular of the proximal colon, than for distal colon or rectal cancer. The remaining two patterns studied were not significantly associated with the risk of CRC or its anatomical subsites.

To our knowledge, this is the first prospective cohort study relating dietary patterns at the nutrient level to CRC risk. Two case–control studies have identified a pattern labelled as 'vitamins and fibre' (Bravi *et al*, 2010; Turati *et al*, 2011), similar to the PC2 in our analysis that was found to be associated with a reduced risk of rectal cancer (Bravi *et al*, 2010). The results of a recent review of case–control and cohort studies on dietary patterns and CRC risk indicate that a plant-based pattern with moderate amounts of dairy, a pattern that is usually characterised by a high variety of vitamins and minerals, was associated with a lower CRC risk (Fung and Brown, 2013). Possible explanations for the apparent protective effect of PC2 may come from the antioxidant properties of nutrients with high loadings on PC2 including β-carotene and vitamin C that may reduce cell proliferation (World Cancer Research Fund/American Institute for Cancer Research, 2007; Stone *et al*, 2014). One of the most striking differences between PC1 and PC2 is the differential loading of vitamin B_{12}. This would support the relevance of the one-carbon metabolism pathway contributing to colon carcinogenesis potentially through modification of DNA methylation (Kim *et al*, 2004; Davis and Uthus, 2004; Ulrey *et al*, 2005; Ulrich *et al*, 2008; Zhang *et al*, 2013). Interestingly, the patterns associated with lower CRC risk (i.e., PC2 and PC4) are those that are concomitantly associated with folate and vitamin B_{12}. Furthermore, PC4, but also PC2, loaded on calcium (loadings on PC2 and PC4 were 13% and 28% respectively) that may have an antitumourigenic effect in the colon through reductions in cellular proliferation, by binding secondary bile acids and ionised fatty acids (Aune *et al*, 2012), and promotion of differentiation and apoptosis in both normal and tumour colorectal cells (Lamprecht and Lipkin, 2001). Riboflavin

Table 2. Characteristics of study subjects and their food consumption by lowest, middle and highest quintiles of the four main nutrient patterns (PC1–PC4)

	PC1			PC2			PC3			PC4		
	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5	Q1	Q3	Q5
N subjects/CRC cases	95 462/1133	95 462/916	95 462/707	95 462/880	95 462/886	95 462/935	95 462/790	95 462/937	95 462/888	95 462/831	95 462/923	95 462/972
Age (years)[a]	52.0±8.9	51.2±9.4	50.6±11.9	49.4±10.3	51.2±10.0	53.1±9.1	51.8±9.0	51.5±10.2	50.2±10.1	49.7±10.6	51.5±9.6	52±9.9
BMI (kg m^{-2})[a]	25.6±4.1	25.4±4.2	25.4±4.6	25.7±4.3	25.4±2.3	25.3±4.2	25.2±4.4	25.5±4.2	25.5±4.2	24.8±4.2	25.3±4.2	26.2±4.4
Height (cm)[a]	169.4±9.2	165.6±8.6	163.5±8.4	166.9±9.9	166.1±8.8	164.8±7.9	164.1±8.6	166.4±9.1	166.9±8.6	167±8.9	165.9±8.9	165.2±9
University degree (%)[b]	19	24	28	19	26	25	25	25	22	25	25	21
Men (%)[b]	47	26	22	45	29	16	25	32	31	30	29	31
Current smoker (%)[b]	33	21	15	28	21	19	22	23	22	22	22	24
Physically active (%)[b]	7	8	9	9	9	7	8	10	6	8	8	9
Vegetables[a,c]	119±79	196±110	342±184	147±116	206±137	286±159.3	243±161	205±137	194±140	236±174	211±136	188±131
Fruits[a,c]	135±111	222±150	353±231	217±191	229±175	255±181	265±192	221±172	226±188	207±165	238±177	251±204
Cereals/products[a,c]	219±107	221±110	216±116	285±142	211±94	173±85	227±122	215±106	219±109	235±110	226±114	190±102
Cakes and biscuits[a,c]	46±48	44±43	28±29	52±55	41±39	29±29	39±40	42±45	38±33	52±53	42±40	28±31
Red meat[a,c]	51±41	45±34	31±32	44±37	43±36	42±35	51±37	47±37	29±29	34±31	46±35	46±39
Poultry[a,c]	17±18	20±20	18±21	20±21	18±19	21±20	21±19	19±19	19±22	14±15	19±18	24±24
Processed meat[a,c]	48±37	32±27	13±19	35±36	32±31	27±24	31±32	33±31	30±28	34±34	32±30	29±28
Fish and shellfish[a,c]	48±44	37±33	27±29	26±27	35±32	54±45	25±21	32±26	63±53	32±36	36±33	45±40
Milk[a,c]	229±245	189±201	161±176	163±195	192±205	217±219	180±220	196±209	189±185	105±133	171±175	329±266
Yoghurt/fermented milk[a,c]	62±95	64±84	59±77	48±81	65±85	73±87	64±86	67±86	52±77	47±64	65±81	74±107
Cheese[a,c]	40±35	41±35	32±31	41±39	40±34	33±31	60±47	35±28	23±22	36±30	44±37	31±33
Fruit and vegetable juices[a,c]	42±76	70±116	70±118	49±107	71±116	64±95	59±109	69±116	58±97	69±114	65±110	55±99
Carbonated/soft drinks[a,c]	87±172	73±159	76±178	96±194	74±157	64±155	56±149	84±176	83±164	90±178	72±160	75±173
Energy intake[a,c]	2200±670	2100±600	1900±560	2300±690	2100±580	1900±530	2200±640	2100±600	1900±580	2100±640	2100±600	1900±590
Alcohol[a,c]	15±20	12±16	9±14	13±19	12±17	11±16	13±18	13±17	10±16	9±13	12±16	15±22

Abbreviations: BMI = body mass index; CRC = colorectal cancer; PC = principal component.
[a]Means±s.d. are presented for continuous variables, all P<0.001 (analysis of variance (ANOVA) performed across the five quintiles).
[b]Frequencies are presented for categorical variables, all P<0.001 (χ^2 test performed across the five quintiles).
[c]Estimated from the country-specific validated dietary questionnaires in the European Prospective Investigation into Cancer and Nutrition (EPIC) study, in g per day except for energy in kcal per day.

Table 3. HRs (95% CI) for CRC and subsites by quintiles (Q) of PC1 score and for 1 s.d. increase in PC1 score

	Q1	Q2	Q3	Q4	Q5	P-trend	HR (95% CI) per 1 s.d. increase	P-value HR continuous for 1 s.d. increase
Colorectum								
N cases	1133	949	916	812	707			
Basic[a]	1.00 (Ref)	0.97 (0.89–1.07)	1.00 (0.9–1.10)	0.92 (0.83–1.02)	0.95 (0.84–1.06)	0.261	0.96 (0.93–1.00)	0.061
Multivariable[b]	1.00 (Ref)	0.99 (0.90–1.08)	1.02 (0.92–1.12)	0.94 (0.85–1.05)	0.98 (0.88–1.11)	0.686	0.98 (0.94–1.02)	0.264
Colon								
N cases	691	598	581	517	482			
Basic[a]	1.00 (Ref)	0.96 (0.86–1.08)	0.96 (0.85–1.08)	0.87 (0.76–0.99)	0.95 (0.82–1.10)	0.304	0.96 (0.91–1.01)	0.093
Multivariable[b]	1.00 (Ref)	0.97 (0.86–1.09)	0.97 (0.86–1.10)	0.89 (0.78–1.01)	0.98 (0.85–1.13)	0.558	0.97 (0.92–1.02)	0.223
Colon, proximal								
N cases	324	269	253	241	211			
Basic[a]	1.00 (Ref)	0.94 (0.80–1.12)	0.92 (0.76–1.10)	0.90 (0.74–1.09)	1.00 (0.80–1.24)	0.971	1.00 (0.93–1.08)	0.979
Multivariable[b]	1.00 (Ref)	0.96 (0.81–1.14)	0.94 (0.78–1.13)	0.93 (0.77–1.13)	1.04 (0.84–1.29)	0.692	1.02 (0.95–1.10)	0.631
Colon, distal								
N cases	320	265	263	218	200			
Basic[a]	1.00 (Ref)	0.92 (0.78–1.09)	0.95 (0.79–1.13)	0.80 (0.66–0.98)	0.90 (0.72–1.12)	0.168	0.92 (0.86–0.99)	0.0308
Multivariable[b]	1.00 (Ref)	0.92 (0.78–1.09)	0.95 (0.79–1.14)	0.81 (0.66–0.99)	0.91 (0.74–1.14)	0.226	0.93 (0.86–1.00)	0.05
Rectum								
N cases	442	351	335	295	225			
Basic[a]	1.00 (Ref)	1.00 (0.86–1.15)	1.06 (0.91–1.24)	1.01 (0.85–1.20)	0.92 (0.76–1.13)	0.616	0.97 (0.91–1.04)	0.376
Multivariable[b]	1.00 (Ref)	1.02 (0.88–1.18)	1.10 (0.93–1.28)	1.05 (0.89–1.25)	0.98 (0.80–1.19)	0.927	0.99 (0.93–1.06)	0.799

Abbreviations: CI = confidence interval; CRC = colorectal cancer; HR = hazard ratio; PC = principal component; Ref = reference.
[a]Basic model: Cox regression adjusted for total energy intake (log-transformed continuous variable) and for each score, and stratified by age (1-year categories), sex and centre.
[b]Multivariable model: Cox regression adjusted for total energy intake (log-transformed continuous variable), alcohol (log-transformed continuous variable), smoking status (never, former, smoker, unknown), height (tertiles), weight (tertiles), physical activity level (inactive, moderately inactive, moderately active, active, missing) and for each score, and stratified by age (1-year categories), sex and centre.

Table 4. HRs (95% CI) for CRC and subsites by quintiles (Q) of PC2 score and for 1 s.d. increase in PC2 score

	Q1	Q2	Q3	Q4	Q5	P-trend	HR (95% CI) per 1 s.d. increase	P-value HR continuous for 1 s.d. increase
Colorectum								
N cases	880	877	886	939	935			
Basic[a]	1.00 (Ref)	0.97 (0.88–1.07)	0.93 (0.84–1.03)	0.96 (0.87–1.07)	0.92 (0.83–1.02)	0.123	0.96 (0.93–1.00)	0.035
Multivariable[b]	1.00 (Ref)	0.96 (0.87–1.06)	0.92 (0.83–1.01)	0.94 (0.85–1.04)	0.88 (0.80–0.98)	0.023	0.94 (0.92–0.98)	0.004
Colon								
N cases	551	572	564	603	579			
Basic[a]	1.00 (Ref)	1.03 (0.91–1.16)	0.96 (0.85–1.09)	0.99 (0.87–1.13)	0.90 (0.79–1.03)	0.083	0.94 (0.91–0.99)	0.018
Multivariable[b]	1.00 (Ref)	1.02 (0.91–1.16)	0.95 (0.84–1.07)	0.97 (0.86–1.10)	0.87 (0.76–1.00)	0.023	0.94 (0.90–0.98)	0.003
Colon, proximal								
N cases	239	255	270	278	256			
Basic[a]	1.00 (Ref)	1.01 (0.84–1.21)	0.98 (0.81–1.18)	0.97 (0.80–1.17)	0.85 (0.70–1.04)	0.069	0.92 (0.87–0.99)	0.016
Multivariable[b]	1.00 (Ref)	1.01 (0.84–1.21)	0.98 (0.82–1.18)	0.96 (0.80–1.16)	0.84 (0.69–1.03)	0.05	0.92 (0.86–0.98)	0.0106
Colon, distal								
N cases	252	255	236	257	266			
Basic[a]	1.00 (Ref)	1.05 (0.88–1.26)	0.94 (0.78–1.13)	1.00 (0.83–1.22)	0.99 (0.81–1.20)	0.798	0.99 (0.92–1.05)	0.667
Multivariable[b]	1.00 (Ref)	1.03 (0.86–1.24)	0.91 (0.75–1.10)	0.96 (0.79–1.17)	0.92 (0.76–1.13)	0.352	0.96 (0.90–1.03)	0.252
Rectum								
N cases	329	305	322	336	356			
Basic[a]	1.00 (Ref)	0.87 (0.74–1.02)	0.88 (0.75–1.04)	0.91 (0.78–1.08)	0.95 (0.80–1.13)	0.787	0.99 (0.93–1.05)	0.718
Multivariable[b]	1.00 (Ref)	0.86 (0.73–1.01)	0.87 (0.74–1.02)	0.89 (0.75–1.05)	0.91 (0.76–1.08)	0.43	0.97 (0.92–1.03)	0.374

Abbreviations: CI = confidence interval; CRC = colorectal cancer; HR = hazard ratio; PC = principal component; Ref = reference.
[a]Basic model: Cox regression adjusted for total energy intake (log-transformed continuous variable) and for each score, and stratified by age (1-year categories), sex and centre.
[b]Multivariable model: Cox regression adjusted for total energy intake (log-transformed continuous variable), alcohol (log-transformed continuous variable), smoking status (never, former, smoker, unknown), height (tertiles), weight (tertiles), physical activity level (inactive, moderately inactive, moderately active, active, missing) and for each score, and stratified by age (1-year categories), sex and centre.

(vitamin B_2), another nutrient with a high loading on PC4, is a micronutrient that plays a pivotal role in one-carbon metabolism and has already been implicated in a reduced risk of CRC in EPIC (Eussen *et al*, 2010) and in a large cohort of women (Zschabitz *et al*, 2013).

We did not find a significant association between PC1 and overall CRC risk. However, there was a small suggestive inverse association with cancer of the distal colon and also with overall CRC risk after excluding subjects from Norway. In a subanalysis

stratified by country, we found that Norway contributed to most of the heterogeneity observed for associations between PC1 and CRC risk. A previous study within EPIC using a single 24-h recall has shown that intakes of nutrients with high loadings on PC1 such as

dietary fibre, folate and vitamin C were consistently below the overall EPIC mean in Norway (Freisling *et al*, 2010). These nutrients therefore appear to be underrepresented in the Norwegian diet in comparison with the EPIC overall mean. As this was not apparent in

Table 5. HRs (95% CI) for CRC and subsites by quintiles (Q) of PC3 score and for 1s.d. increase in PC3 score

	Q1	Q2	Q3	Q4	Q5	*P*-trend	HR (95% CI) per 1 s.d. increase	*P*-value HR continuous for 1 s.d. increase
Colorectum								
N cases	790	899	937	1003	888			
Basic[a]	1.00 (Ref)	1.03 (0.94–1.14)	1.02 (0.92–1.13)	1.07 (0.97–1.19)	0.99 (0.89–1.12)	0.920	1.00 (0.97–1.04)	0.817
Multivariable[b]	1.00 (Ref)	1.03 (0.93–1.13)	1.01 (0.91–1.12)	1.06 (0.96–1.18)	0.98 (0.88–1.10)	0.935	1.00 (0.96–1.04)	0.99
Colon								
N cases	523	553	606	644	543			
Basic[a]	1.00 (Ref)	0.99 (0.88–1.12)	1.04 (0.92–1.18)	1.09 (0.95–1.24)	0.95 (0.83–1.10)	0.854	0.98 (0.95–1.04)	0.722
Multivariable[b]	1.00 (Ref)	0.98 (0.87–1.12)	1.04 (0.91–1.18)	1.08 (0.95–1.23)	0.94 (0.82–1.09)	0.762	0.99 (0.94–1.04)	0.608
Colon, proximal								
N cases	199	258	271	314	256			
Basic[a]	1.00 (Ref)	1.16 (0.96–1.40)	1.12 (0.92–1.36)	1.23 (1.01–1.50)	1.05 (0.84–1.30)	0.626	0.99 (0.92–1.06)	0.752
Multivariable[b]	1.00 (Ref)	1.15 (0.95–1.40)	1.12 (0.92–1.36)	1.23 (1.00–1.50)	1.05 (0.84–1.31)	0.594	0.99 (0.92–1.06)	0.771
Colon, distal								
N cases	241	234	276	273	242			
Basic[a]	1.00 (Ref)	0.92 (0.77–1.11)	1.04 (0.87–1.26)	1.05 (0.86–1.27)	0.95 (0.77–1.18)	0.941	1.01 (0.94–1.08)	0.849
Multivariable[b]	1.00 (Ref)	0.91 (0.76–1.10)	1.03 (0.86–1.24)	1.02 (0.84–1.25)	0.93 (0.75–1.15)	0.768	1.00 (0.93–1.07)	0.94
Rectum								
N cases	267	346	331	359	345			
Basic[a]	1.00 (Ref)	1.11 (0.94–1.31)	0.98 (0.83–1.17)	1.05 (0.89–1.29)	1.07 (0.89–1.29)	0.681	1.03 (0.96–1.09)	0.40
Multivariable[b]	1.00 (Ref)	1.10 (0.94–1.30)	0.97 (0.82–1.16)	1.04 (0.87–1.24)	1.05 (0.87–1.27)	0.799	1.02 (0.96–1.09)	0.515

Abbreviations: CI = confidence interval; CRC = colorectal cancer; HR = hazard ratio; PC = principal component; Ref = reference.
[a]Basic model: Cox regression adjusted for total energy intake (log-transformed continuous variable) and for each score, and stratified by age (1-year categories), sex and centre.
[b]Multivariable model: Cox regression adjusted for total energy intake (log-transformed continuous variable), alcohol (log-transformed continuous variable), smoking status (never, former, smoker, unknown), height (tertiles), weight (tertiles), physical activity level (inactive, moderately inactive, moderately active, active, missing) and for each score, and stratified by age (1-year categories), sex and centre.

Table 6. HRs (95% CI) for CRC and subsites by quintiles (Q) of PC4 score and for 1 s.d. increase in PC4 score

	Q1	Q2	Q3	Q4	Q5	*P*-trend	HR (95% CI) per 1 s.d. increase	*P*-value HR continuous for 1 s.d. increase
Colorectum								
N cases	831	860	923	931	972			
Basic[a]	1.00 (Ref)	0.97 (0.88–1.07)	0.99 (0.90–1.09)	0.94 (0.85–1.03)	0.94 (0.85–1.04)	0.172	0.98 (0.94–1.01)	0.155
Multivariable[b]	1.00 (Ref)	0.96 (0.87–1.05)	0.96 (0.88–1.06)	0.91 (0.82–1.00)	0.90 (0.81–0.99)	0.0192	0.96 (0.93–0.99)	0.015
Colon								
N cases	517	563	591	587	611			
Basic[a]	1.00 (Ref)	1.01 (0.89–1.14)	0.99 (0.88–1.12)	0.92 (0.82–1.04)	0.92 (0.81–1.04)	0.076	0.97 (0.93–1.01)	0.12
Multivariable[b]	1.00 (Ref)	0.99 (0.88–1.12)	0.97 (0.86–1.10)	0.89 (0.79–1.01)	0.88 (0.77–1.00)	0.013	0.95 (0.91–0.99)	0.022
Colon, proximal								
N cases	263	252	259	254	270			
Basic[a]	1.00 (Ref)	0.91 (0.77–1.09)	0.89 (0.75–1.07)	0.82 (0.68–0.98)	0.84 (0.70–1.01)	0.04	0.95 (0.89–1.01)	0.08
Multivariable[b]	1.00 (Ref)	0.91 (0.76–1.08)	0.88 (0.74–1.05)	0.80 (0.67–0.98)	0.81 (0.67–0.98)	0.014	0.93 (0.88–0.99)	0.03
Colon, distal								
N cases	209	243	275	259	280			
Basic[a]	1.00 (Ref)	1.06 (0.88–1.28)	1.10 (0.91–1.32)	0.95 (0.78–1.15)	0.95 (0.78–1.16)	0.33	0.97 (0.91–1.03)	0.362
Multivariable[b]	1.00 (Ref)	1.03 (0.86–1.25)	1.06 (0.88–1.28)	0.91 (0.75–1.10)	0.90 (0.74–1.09)	0.115	0.95 (0.89–1.01)	0.125
Rectum								
N cases	314	297	332	344	361			
Basic[a]	1.00 (Ref)	0.91 (0.78–1.07)	0.98 (0.84–1.15)	0.96 (0.82–1.13)	0.99 (0.84–1.16)	0.94	0.99 (0.94–1.05)	0.76
Multivariable[b]	1.00 (Ref)	0.89 (0.76–1.05)	0.95 (0.81–1.12)	0.93 (0.79–1.09)	0.93 (0.79–1.10)	0.531	0.97 (0.92–1.03)	0.307

Abbreviations: CI = confidence interval; CRC = colorectal cancer; HR = hazard ratio; PC = principal component; Ref = reference.
[a]Basic model: Cox regression adjusted for total energy intake (log-transformed continuous variable) and for each score, and stratified by age (1-year categories), sex and centre.
[b]Multivariable model: Cox regression adjusted for total energy intake (log-transformed continuous variable), alcohol (log-transformed continuous variable), smoking status (never, former, smoker, unknown), height (tertiles), weight (tertiles), physical activity level (inactive, moderately inactive, moderately active, active, missing) and for each score, and stratified by age (1-year categories), sex and centre.

the centre-wide PCA based on long-term dietary instruments, we decided to keep Norway in the overall PCA.

The PC1 shares some of the nutrient characteristics of PC2, with the exception of being poor in nutrients from animal food sources such as retinol and vitamin B_{12}, and also in calcium, phosphorus and riboflavin. As reported previously, in terms of contributing foods, individuals with a high adherence to PC1, compared with individuals with high adherence to PC2, had slightly higher intakes of vegetables and fruits, cereals and cereal products, but substantially lower intakes of milk and yoghurt, and also of fish (Moskal et al, 2014). A high consumption of fruits and vegetables may not be sufficient to markedly reduce the risk of CRC. A pooled analysis of 14 prospective studies concluded that fruit and vegetable intake is not strongly associated with a decreased risk of colon cancer overall, but may be associated with a lower risk of distal colon cancer (Koushik et al, 2007), and this is in line with our finding of a 7% reduced risk of distal colon cancer per 1 s.d. increase in PC1. These findings deserve further investigation in future studies.

In a case–control study, a nutrient pattern similar to our PC3 was labelled 'unsaturated fats (animal source)', and was positively related to the risk of CRC (Bravi et al, 2010), but in our study no significant associations were found. This is also in line with a meta-analysis of dietary vitamin D, where no significant associations with CRC risk were found (Huncharek et al, 2009).

None of our nutrient patterns reflected a nutrient profile associated with a strikingly high consumption of red and/or processed meat (Table 2), and this is probably a main reason why we did not observe a higher CRC risk associated with any of the four patterns. For example, mean consumption of red meat in the highest quintile of PC4 was < 50 g per day, and this is at or below the current dietary guidelines for cancer prevention, that is, < 70 at population level or < 40 g per day at individual level (Norat et al, 2015).

The main strengths of our study include its prospective design with a long follow-up, and the very large sample size combined with a large number of incident cancer cases. The analysis benefitted from these unique features of the whole EPIC cohort allowing us to explore differences in CRC risk by anatomical subsites and countries. We believe that reverse causation (changes in dietary habits because of early symptoms of an undetected cancer) is unlikely because our risk estimates remained virtually unchanged after excluding cases identified within the first 5 years of follow-up. On the other hand, we were not able to account for potential changes in diet during follow-up, because diet was assessed only at baseline. However, previous studies have demonstrated a reasonable stability of dietary patterns over time (Hu et al, 1999).

We used common nutrient patterns across 23 centres from 10 countries, accounting for a wide heterogeneity in the diet (Slimani et al, 2002; Freisling et al, 2010), allowing generalisation across these diverse populations. In this regard, we would like to highlight the consistency of findings, in particular for PC2 and PC4, across countries/centres.

Finally, by exploring macro- and micronutrients, the present study aimed to provide a holistic representation of individuals' diet. It is likely that the combined intake of multiple dietary factors acts synergistically on a given health outcome and such cumulative effects may be easier to observe in a population setting (Hu, 2002; Jacobs et al, 2009). Individuals also vary considerably in their susceptibility to events that either inhibit or contribute to cancer development in response to the very same dietary component (World Cancer Research Fund/American Institute for Cancer Research, 2007). Therefore, the net 'synergistic' effect of a dietary pattern may also be that a greater proportion of individuals of a population are susceptible to at least one dietary component of a given pattern. One of the main advantages of dietary patterns at the

nutrient level is the potential to identify combinations of nutrients that could reflect underlying biological mechanisms. Nutrient patterns can therefore be seen as complementary to food patterns and ideally confirm observed associations between foods and a health outcome. Furthermore, because different foods contributed to the same nutrient patterns, it is unlikely that results are confounded by other dietary compounds not captured by a given pattern, adding further strength to our findings. This appears to be particularly the case for nutrient patterns, where many nutrients contribute to a given pattern (i.e., PC2 and PC4). In contrast, PC3 seems to be mainly driven by vitamin D and PC1 shows sizeable positive associations only with β-carotene and vitamin C. Indeed, for these two patterns (PC1 and PC3), confounding by 'healthier' foods across countries cannot entirely be excluded and this may be one explanation why heterogeneity in associations with CRC across countries was larger for PC1 and PC3 as compared with PC2 and PC4.

Some caution is warranted with the interpretation of our findings. First, nutrient patterns identified by the exploratory techniques are meant to reflect dietary habits independently from any a priori hypotheses regarding known or unknown dietary effects on health outcomes. Indeed, PCA aims at maximising the fraction of variance explained by a weighted linear combination of the original nutrients. However, the nutritional components that are most variable might not be those that are most strongly associated with disease (McCann et al, 2001). As a consequence, the dietary patterns identified and used in diet–disease association studies are not necessarily relevant for cancer risk, and this could explain, at least in part, the lack of association reported in some studies (McCann et al, 2001; Randi et al, 2010). In future analysis, statistical methods of dimension reduction with respect to outcome such as PLS or latent class analysis could be of interest to link patterns to cancer outcomes.

Second, although several statistical tests were performed in this work, no Bonferroni correction was applied. As individuals do not belong exclusively to one pattern, but rather dietary habits can be spanned into different patterns, there exists an inherent degree of intercorrelation between some of the components of a given pattern (although the patterns itself are uncorrelated with each other). This is likely to make any multiple testing corrections too restrictive.

Third, this study relied on dietary questionnaires to assess nutrient intakes that are prone to measurement errors. In addition, questionnaires used to estimate food intakes were country specific and might have introduced systematic between-country differences in dietary assessment. However, our models were stratified by country/centre to account for such potential differences and adjusted for total energy intakes partly accounting for such differences and measurement errors (Ferrari et al, 2008). In addition, a harmonised nutrient database was used to convert food consumption into nutrient intakes (Slimani et al, 2007) that should improve the comparability of nutrient intakes across the countries participating in EPIC. A specific limitation of the ENDB is that not all essential nutrients are available because of the lack of available and reliable data. For example, in our nutrient patterns we could not separate the essential fatty acids linoleic and linolenic acids from total polyunsaturated fatty acids or account for sodium (salt) intake that are all key nutrients in dietary recommendations or cancer prevention guidelines (World Cancer Research Fund/ American Institute for Cancer Research, 2007). However, we have good reasons to believe that the lack of sodium in our nutrient patterns did not affect our observed associations because it appears that salt intake is not associated with CRC risk (but with stomach cancer risk) (World Cancer Research Fund/American Institute for Cancer Research, 2007). Similarly, the evidence for an association of n-3 and n-6 fatty acids with CRC development is not very strong (Song et al, 2014; Chen et al, 2015) and total PUFAs did not explain a large part of the variation in nutrient intakes in EPIC.

Fourth, although we were able to adjust for most of the established CRC risk factors, we cannot exclude the possibility of residual confounding; for example, we were not able to adjust for use of nonsteroidal anti-inflammatory drugs (NSAIDs) (Rothwell et al, 2010) as this information was not collected in EPIC. Furthermore, genetic factors, which may play a role in CRC development, were not accounted for in this analysis. In future analysis, it would be worthwhile to investigate how genetic factors may have interacted with nutrient patterns on CRC risk.

Finally, a disadvantage of a nutrient-based approach is that nutrients are less directly related to dietary recommendations because, ultimately, nutrient intakes are largely determined by the choice of food sources. As many food sources exist for the same nutrient, it is challenging to make food-based dietary recommendations. However, there is accumulating evidence that health-conscious consumers are increasingly using nutrient-based information to make healthier food choices (Hersey et al, 2013). In addition, future studies may also look to integrating biological marker information with nutrient pattern analyses and further improve our understanding of the role of dietary factors in chronic disease aetiology, where nutrient patterns act as an interface between food patterns and the food metabolome (O'Sullivan et al, 2011). Nonetheless, in previous descriptive analyses (and briefly in Table 2), we were able to identify the related food sources that characterised each of the nutrient patterns in EPIC (Moskal et al, 2014).

If the inverse associations between PC2 and PC4 and CRC risk were causal and under the assumption that study subjects, for example, shift one quintile upwards in the distribution of nutrient patterns, a major effect on cancer incidence is unlikely. However, the results contribute to an improved understanding on how overall dietary patterns affect CRC development.

CONCLUSION

In summary, our study provides the first assessment of the relation between nutrient patterns and CRC risk in a large prospective multicentre cohort setting. Of the four nutrient patterns examined, the results were not statistically significant for PC1, a pattern characterised by nutrients from plant food sources such as vitamin C or β-carotene, and PC3, a pattern mainly driven by dietary vitamin D. Results for PC2, a nutrient pattern characterised by a high variety of vitamins and minerals, and PC4, a pattern driven by vitamin B_{12}, riboflavin, calcium, phosphorus, cholesterol and total proteins, were both significantly and independently associated with decreased CRC risk. These findings suggest that analysing nutrient patterns may improve our understanding of how groups of nutrients consumed together might relate to CRC.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aune D, Lau R, Chan DS, Vieira R, Greenwood DC, Kampman E, Norat T (2012) Dairy products and colorectal cancer risk: a systematic review and meta-analysis of cohort studies. Ann Oncol 23(1): 37–45.

Bingham SA (2000) Diet and colorectal cancer prevention. Biochem Soc Trans 28(2): 12–16.

Bravi F, Edefonti V, Bosetti C, Talamini R, Montella M, Giacosa A, Franceschi S, Negri E, Ferraroni M, La VC, Decarli A (2010) Nutrient dietary patterns and the risk of colorectal cancer: a case-control study from Italy. Cancer Causes Control 21(11): 1911–1918.

Chen GC, Qin LQ, Lu DB, Han TM, Zheng Y, Xu GZ, Wang XH (2015) N-3 polyunsaturated fatty acids intake and risk of colorectal cancer: meta-analysis of prospective studies. Cancer Causes Control 26(1): 133–141.

Davis CD, Uthus EO (2004) DNA methylation, cancer susceptibility, and nutrient interactions. Exp Biol Med (Maywood) 229(10): 988–995.

De Stefani E, Ronco AL, Boffetta P, Deneo-Pellegrini H, Correa P, Acosta G, Mendilaharsu M (2012) Nutrient-derived dietary patterns and risk of colorectal cancer: a factor analysis in Uruguay. Asian Pac J Cancer Prev 13(1): 231–235.

Eussen SJ, Vollset SE, Hustad S, Midttun O, Meyer K, Fredriksen A, Ueland PM, Jenab M, Slimani N, Boffetta P, Overvad K, Thorlacius-Ussing O, Tjonneland A, Olsen A, Clavel-Chapelon F, Boutron-Ruault MC, Morois S, Weikert C, Pischon T, Linseisen J, Kaaks R, Trichopoulou A, Zilis D, Katsoulis M, Palli D, Pala V, Vineis P, Tumino R, Panico S, Peeters PH, Bueno-de-Mesquita HB, van Duijnhoven FJ, Skeie G, Munoz X, Martinez C, Dorronsoro M, Ardanaz E, Navarro C, Rodriguez L, VanGuelpen B, Palmqvist R, Manjer J, Ericson U, Bingham S, Khaw KT, Norat T, Riboli E (2010) Plasma vitamins B2, B6, and B12, and related genetic variants as predictors of colorectal cancer risk. Cancer Epidemiol Biomarkers Prev 19(10): 2549–2561.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136(5): E359–E386.

Ferrari P, Day NE, Boshuizen HC, Roddam A, Hoffmann K, Thiebaut A, Pera G, Overvad K, Lund E, Trichopoulou A, Tumino R, Gullberg B, Norat T, Slimani N, Kaaks R, Riboli E (2008) The evaluation of the diet/disease relation in the EPIC study: considerations for the calibration and the disease models. Int J Epidemiol 37(2): 368–378.

Freisling H, Fahey MT, Moskal A, Ocke MC, Ferrari P, Jenab M, Norat T, Naska A, Welch AA, Navarro C, Schulz M, Wirfalt E, Casagrande C,

Amiano P, Ardanaz E, Parr C, Engeset D, Grioni S, Sera F, Bueno-De-Mesquita B, van der Schouw YT, Touvier M, Boutron-Ruault MC, Halkjaer J, Dahm CC, Khaw KT, Crowe F, Linseisen J, Kroger J, Huybrechts I, Deharveng G, Manjer J, Agren A, Trichopoulou A, Tsiotas K, Riboli E, Bingham S, Slimani N (2010) Region-specific nutrient intake patterns exhibit a geographical gradient within and between European countries. *J Nutr* **140**(7): 1280–1286.

Freisling H, van Bakel MM, Biessy C, May AM, Byrnes G, Norat T, Rinaldi S, Santucci de MM, Grioni S, Bueno-de-Mesquita HB, Ocke MC, Kaaks R, Teucher B, Vergnaud AC, Romaguera D, Sacerdote C, Palli D, Crowe FL, Tumino R, Clavel-Chapelon F, Boutron-Ruault MC, Khaw KT, Wareham NJ, Trichopoulou A, Naska A, Orfanos P, Boeing H, Illner AK, Riboli E, Peeters PH, Slimani N (2012) Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *Br J Nutr* **107**(6): 910–920.

Fung TT, Brown LS (2013) Dietary patterns and the risk of colorectal cancer. *Curr Nutr Rep* **2**(1): 48–55.

Greenland S, Longnecker MP (1992) Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* **135**(11): 1301–1309.

Hersey JC, Wohlgenant KC, Arsenault JE, Kosa KM, Muth MK (2013) Effects of front-of-package and shelf nutrition labeling systems on consumers. *Nutr Rev* **71**(1): 1–14.

Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**(11): 1539–1558.

Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**(1): 3–9.

Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett WC (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* **69**(2): 243–249.

Huncharek M, Muscat J, Kupelnick B (2009) Colorectal cancer risk and dietary intake of calcium, vitamin D, and dairy products: a meta-analysis of 26,335 cases from 60 observational studies. *Nutr Cancer* **61**(1): 47–69.

Jacobs Jr DR, Gross MD, Tapsell LC (2009) Food synergy: an operational concept for understanding nutrition. *Am J Clin Nutr* **89**(5): 1543S–1548S.

Jacobs Jr DR, Steffen LM (2003) Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *Am J Clin Nutr* **78** (3 Suppl): 508S–513S.

Jacobs Jr DR, Tapsell LC (2007) Food, not nutrients, is the fundamental unit in nutrition. *Nutr Rev* **65**(10): 439–450.

Johnson R, Wichern D (2007) *Applied Multivariate Statistical Analysis*. 5th (ed). Prentice-Hall: Upper Saddle River, NJ, USA.

Kim MK, Sasaki S, Sasazuki S, Tsugane S (2004) Prospective study of three major dietary patterns and risk of gastric cancer in Japan. *Int J Cancer* **110**(3): 435–442.

Koushik A, Hunter DJ, Spiegelman D, Beeson WL, van den Brandt PA, Buring JE, Calle EE, Cho E, Fraser GE, Freudenheim JL, Fuchs CS, Giovannucci EL, Goldbohm RA, Harnack L, Jacobs Jr DR, Kato I, Krogh V, Larsson SC, Leitzmann MF, Marshall JR, McCullough ML, Miller AB, Pietinen P, Rohan TE, Schatzkin A, Sieri S, Virtanen MJ, Wolk A, Zeleniuch-Jacquotte A, Zhang SM, Smith-Warner SA (2007) Fruits, vegetables, and colon cancer risk in a pooled analysis of 14 cohort studies. *J Natl Cancer Inst* **99**(19): 1471–1483.

Lamprecht SA, Lipkin M (2001) Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis. *Ann NY Acad Sci* **952**: 73–87.

McCann SE, Weiner J, Graham S, Freudenheim JL (2001) Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? *Public Health Nutr* **4**(4): 903–908.

Miller PE, Lesko SM, Muscat JE, Lazarus P, Hartman TJ (2010) Dietary patterns and colorectal adenoma and cancer risk: a review of the epidemiological evidence. *Nutr Cancer* **62**(4): 413–424.

Moskal A, Pisa PT, Ferrari P, Byrnes G, Freisling H, Boutron-Ruault MC, Cadeau C, Nailler L, Wendt A, Kuhn T, Boeing H, Buijsse B, Tjonneland A, Halkjaer J, Dahm CC, Chiuve SE, Quiros JR, Buckland G, Molina-Montes E, Amiano P, Huerta Castano JM, Gurrea AB, Khaw KT, Lentjes MA, Key TJ, Romaguera D, Vergnaud AC, Trichopoulou A, Bamia C, Orfanos P, Palli D, Pala V, Tumino R, Sacerdote C, de Magistris MS, Bueno-de-Mesquita HB, Ocke MC, Beulens JW, Ericson U, Drake I, Nilsson LM, Winkvist A, Weiderpass E, Hjartaker A, Riboli E, Slimani N (2014) Nutrient patterns and

their food sources in an International Study Setting: report from the EPIC study. *PLoS One* **9**(6): e98647.

Nicolas G, Witthoft CM, Vignat J, Knaze V, Huybrechts I, Roe M, Finglas P, Slimani N (2016) Compilation of a standardised international folate database for EPIC. *Food Chem* **193**: 134–140.

Norat T, Scoccianti C, Boutron-Ruault MC, Anderson A, Berrino F, Cecchini M, Espina C, Key T, Leitzmann M, Powers H, Wiseman M, Romieu I (2015) European Code against Cancer 4th Edition: Diet and Cancer. *Cancer Epidemiol* **39**(Suppl 1): S56–S66.

O'Sullivan A, Gibney MJ, Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* **93**(2): 314–321.

Randi G, Edefonti V, Ferraroni M, La VC, Decarli A (2010) Dietary patterns and the risk of colorectal cancer and adenomas. *Nutr Rev* **68**(7): 389–408.

Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-de-Mesquita HB, Peeters PH, Lund E, Engeset D, Gonzalez CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5**(6B): 1113–1124.

Riboli E, Kaaks R (1997) The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26**(Suppl 1): S6–S14.

Rothwell PM, Wilson M, Elwin CE, Norrving B, Algra A, Warlow CP, Meade TW (2010) Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *Lancet* **376**(9754): 1741–1750.

Slimani N, Deharveng G, Unwin I, Southgate DA, Vignat J, Skeie G, Salvini S, Parpinel M, Moller A, Ireland J, Becker W, Farran A, Westenbrink S, Vasilopoulou E, Unwin J, Borgejordet A, Rohrmann S, Church S, Gnagnarella P, Casagrande C, van BM, Niravong M, Boutron-Ruault MC, Stripp C, Tjonneland A, Trichopoulou A, Georga K, Nilsson S, Mattisson I, Ray J, Boeing H, Ocke M, Peeters PH, Jakszyn P, Amiano P, Engeset D, Lund E, de Magistris MS, Sacerdote C, Welch A, Bingham S, Subar AF, Riboli E (2007) The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr* **61**(9): 1037–1056.

Slimani N, Fahey M, Welch AA, Wirfalt E, Stripp C, Bergstrom E, Linseisen J, Schulze MB, Bamia C, Chloptsios Y, Veglia F, Panico S, Bueno-de-Mesquita HB, Ocke MC, Brustad M, Lund E, Gonzalez CA, Barcos A, Berglund G, Winkvist A, Mulligan A, Appleby P, Overvad K, Tjonneland A, Clavel-Chapelon F, Kesse E, Ferrari P, Van Staveren WA, Riboli E (2002) Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public Health Nutr* **5**(6B): 1311–1328.

Song M, Chan AT, Fuchs CS, Ogino S, Hu FB, Mozaffarian D, Ma J, Willett WC, Giovannucci EL, Wu K (2014) Dietary intake of fish, omega-3 and omega-6 fatty acids and risk of colorectal cancer: a prospective study in U.S. men and women. *Int J Cancer* **135**(10): 2413–2423.

Stone WL, Krishnan K, Campbell SE, Palau VE (2014) The role of antioxidants and pro-oxidants in colon cancer. *World J Gastrointest Oncol* **6**(3): 55–66.

Turati F, Edefonti V, Bravi F, Ferraroni M, Franceschi S, La VC, Montella M, Talamini R, Decarli A (2011) Nutrient-based dietary patterns, family history, and colorectal cancer. *Eur J Cancer Prev* **20**(6): 456–461.

Ulrey CL, Liu L, Andrews LG, Tollefsbol TO (2005) The impact of metabolism on DNA methylation. *Hum Mol Genet* **14**((Spec No 1): R139–R147.

Ulrich CM, Reed MC, Nijhout HF (2008) Modeling folate, one-carbon metabolism, and DNA methylation. *Nutr Rev* **66**(Suppl 1): S27–S30.

van Dam RM (2005) New approaches to the study of dietary patterns. *Br J Nutr* **93**(5): 573–574.

Willett WC (2005) Diet and cancer: an evolving picture. *JAMA* **293**(2): 233–234.

Willett WC, Howe GR, Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65**(4 Suppl): 1220S–1228S.

World Cancer Research Fund/American Institute for Cancer Research (2007) *Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective*. AICR: Washington DC, USA.

World Cancer Research Fund/American Institute for Cancer Research (2011) Continuous Update Project Report Summary. *Food, Nutrition, Physical Activity, and the Prevention of Colorectal Cancer*.

Yusof AS, Isa ZM, Shah SA (2012) Dietary patterns and risk of colorectal cancer: a systematic review of cohort studies (2000-2011). *Asian Pac J Cancer Prev* **13**(9): 4713–4717.

Zhang XH, Ma J, Smith-Warner SA, Lee JE, Giovannucci E (2013) Vitamin B6 and colorectal cancer: current evidence and future directions. *World J Gastroenterol* **19**(7): 1005–1010.

Zschabitz S, Cheng TY, Neuhouser ML, Zheng Y, Ray RM, Miller JW, Song X, Maneval DR, Beresford SA, Lane D, Shikany JM, Ulrich CM (2013) B vitamin intakes and incidence of colorectal cancer: results from the Women's Health Initiative Observational Study cohort. *Am J Clin Nutr* **97**(2): 332–343.

[1]International Agency for Research on Cancer (IARC-WHO), 150 Cours Albert Thomas, 69372 Lyon Cedex 08, France; [2]Department of Statistics, Data Science and Epidemiology, Swinburne University of Technology, Hawthorn, Melbourne, VIC 3122, Australia; [3]Danish Cancer Society Research Center, Unit of Diet, Genes and Environment, Strandboulevarden 49, Copenhagen DK-2100, Denmark; [4]Section for Epidemiology, Department of Public Health, Aarhus University, Bartholins Alle 2, Aarhus C DK-8000, Denmark; [5]Université Paris-Saclay, Université Paris-Sud, UVSQ, CESP, INSERM UMRS 1018, Villejuif F-94805, France; [6]Institut Gustave Roussy, Villejuif F-94805, France; [7]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany; [8]Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, Nuthetal 14558, Germany; [9]Hellenic Health Foundation, 13 Kaisareias Street, Athens GR-115 27, Greece; [10]WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Mikras Asias 75, Goudi, Athens GR-115 27, Greece; [11]Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute–ISPO, Ponte Nuovo Palazzina 28A 'Mario Fiori', Via delle Oblate 4, Florence 50141, Italy; [12]Department of Clinical and Experimental Medicine, Federico II University, via Pansini 5, 80131 Naples, Italy; [13]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian, 1, Milano 20133, Italy; [14]Cancer Registry and Histopathology Unit, 'Civic–M.P.Arezzo' Hospital, ASP, Via Dante No. 109, Ragusa 97100, Italy; [15]Unit of Cancer Epidemiology, University of Turin, Via Santena 7, Turin 10126, Italy; [16]Centre for Cancer Epidemiology and Prevention (CPO Piemonte), Via Santena 7, Turin 10126, Italy; [17]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Huispost Str. 6.131, PO Box 85500, Utrecht 3508 GA The Netherlands; [18]MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK; [19]Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and The Environment (RIVM), PO Box 1, Bilthoven 3720 BA, The Netherlands; [20]Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Heidelberglann 100, Utrecht 3584 CX, The Netherlands; [21]Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK; [22]Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia; [23]The Norwegian Food Safety Authority, Head Office, Postboks 383, 2381 Brumunddal, Norway; [24]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø N-9037, Norway; [25]Navarra Public Health Institute, c/Leyre 15, Pamplona 31003, Pamplona, Spain; [26]IdiSNA, Navarra Institute for Health Research, Recinto de Complejo Hospitalario de Navarra c/Irunlarrea 3, Pamplona, 31008, Spain; [27]CIBER, Epidemiology and Public Health (CIBERESP), Melchor Fernández Almagro, 3-5, Madrid 28029, Spain; [28]Unit of Nutrition and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet de Llobregat, Avda Gran Via 199-203, 08908 Barcelona, Spain; [29]Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Ronda de Levante 11, Murcia 30008, Spain; [30]Public Health Directorate, Asturias, Ciriaco Miguel Vigil St 9, Oviedo 33006, Spain; [31]Public Health Division of Gipuzkoa-BIODonostia Research Institute, Basque Regional Health Department, 20013 San Sebastian, Spain; [32]Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria ibs, Hospitales Universitarios de Granada/Universidad de Granada, Cuesta del Observatorio, 4, Campus Universitario de Cartuja, Granada 18080, Spain; [33]Department of Internal Medicine and Clinical Nutrition, University of Gothenburg, PO Box 459, 40530 Gothenburg, Sweden; [34]Department of Public Health and Clinical Medicine, Nutrition Research, Umeå University, SE-901 87 Umeå, Sweden; [35]Department of Radiation Sciences, Oncology, Umeå University, SE-901 87 Umeå, Sweden; [36]Department of Clinical Sciences Malmö, Lund University, Box 117, SE-221 00 Lund, Sweden; [37]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Richard Doll Building, Oxford OX3 7LF, UK; [38]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Box 285, CB2 0QQ Cambridge, UK; [39]Clinical Gerontology Unit, School of Clinical Medicine, University of Cambridge, Addenbrooke's Hospital, Hills Rd, Cambridge CB2 0SP, UK; [40]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus, Ionnina 45110, Greece and [41]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, Cambridge

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)