Nutrient-based dietary patterns and nasopharyngeal cancer: evidence from an exploratory factor analysis

V Edefonti*[,1], F Nicolussi[2], J Polesel[3], F Bravi[1], C Bosetti[4], W Garavello[5], C La Vecchia[1], E Bidoli[3], A Decarli[1,6], D Serraino[3], S Calza[2] and M Ferraroni[1]

[1]Department of Clinical Sciences and Community Health, Università degli Studi di Milano, via G. Venezian 1, 20122 Milan, Italy; [2]Department of Molecular and Translational Medicine, Università degli Studi di Brescia, viale Europa, 1, 25123 Brescia, Italy; [3]Unit of Epidemiology and Biostatistics, Centro di Riferimento Oncologico Aviano IRCCS, via F. Gallini 2, 33080 Aviano, Italy; [4]Department of Epidemiology, IRCCS - Istituto di Ricerche Farmacologiche 'Mario Negri', via G. La Masa 19, 20156 Milan, Italy; [5]Otolaryngology, Department of Surgery and Translational Medicine, Università degli Studi di Milano-Bicocca, via Cadore 48, 20900 Monza, Italy and [6]Unit of Medical Statistics, Biometry and Bioinformatics, Fondazione IRCSS Istituto Nazionale Tumori di Milano, via G. Venezian 1, 20133 Milan, Italy

Background: To our knowledge, no study assessed the association between dietary patterns and nasopharyngeal carcinoma (NPC) in low-incidence areas.

Methods: We examined this association in a hospital-based case–control study carried out in Italy between 1992 and 2008, including 198 incident NPC cases and 594 controls. *A posteriori* dietary patterns were identified through principal component factor analysis performed on 28 nutrients and minerals derived from a 78-item food-frequency questionnaire. Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated using unconditional multiple logistic regression models on tertiles of factor scores.

Results: We identified five dietary patterns named *Animal products, Starch-rich, Vitamins and fibre, Animal unsaturated fatty acids (AUFAs)*, and *Vegetable unsaturated fatty acids (VUFAs)*. The *Animal product* (OR = 2.62, 95% CI = 1.67–4.13, for the highest *vs* lowest score tertile), *Starch-rich* (OR = 2.05, 95% CI = 1.27–3.33), and *VUFA* (OR = 1.90, 95% CI = 1.22–2.96) patterns were positively associated with NPC. The *AUFA* pattern showed a positive association of borderline significance, whereas the *Vitamins and fibre* pattern was nonsignificantly but inversely associated with NPC.

Conclusions: These findings suggest that diets rich in animal products, starch, and fats are positively related to NPC risk in this low-incidence country.

Nasopharyngeal carcinoma (NPC), here used as a surrogate of nasopharyngeal cancer, is one of the rarer malignancies globally, with an age-adjusted incidence rate of 1.2 per 100 000 persons per year for both sexes combined (Ferlay *et al*, 2010). Its incidence rates worldwide show a noteworthy geographical and ethnic variation (Jemal *et al*, 2011). The southern part of China, other parts of South-Eastern Asia, Northern and Southern Africa, and Alaska (limited to the Inuit populations) define the high-incidence area, whereas populations living elsewhere in the United States of America and Europe have considerably lower NPC incidence rates (Curado *et al*, 2007; Jemal *et al*, 2011). Infection with the Epstein–Barr virus (EBV) and genetic susceptibility appear to have

---

a major role in high-incidence populations, but migrant studies suggest that other environmental factors may also be important (Vokes *et al*, 1997; Chang and Adami, 2006). In low-incidence areas, NPC has been associated with well-known lifestyle risk factors, such as tobacco and alcohol (Polesel *et al*, 2011; Vokes *et al*, 1997; International Agency for Research on Cancer, 2004, 2007).

Among other factors, diet has been associated with NPC in either low- or high-incidence areas (Chang and Adami, 2006; Gallicchio *et al*, 2006; Jia and Qin, 2012). There are suggestions that high intakes of non-preserved fruit and non-starchy vegetables are protective against NPC, whereas Cantonese-style salted fish, characterised by using less salt and a higher degree of fermentation during the drying process, is a cause of NPC (World Cancer Research Fund/American Institute for Cancer Research, 2007). In addition, a few studies considered the role of single nutrients on this cancer site (Lee *et al*, 1994; Farrow *et al*, 1998; Kasum *et al*, 2002; Polesel *et al*, 2012), but only two of them were based on validated food-frequency questionnaires (FFQs) (Farrow *et al*, 1998; Polesel *et al*, 2012). To our knowledge, only one study assessed the association between dietary patterns and NPC risk (Armstrong *et al*, 1998), although dietary patterns address the well-known issues of collinearity of nutrients and interdependencies between foods and nutrients (Hu, 2002). Therefore, the overall evidence on diet and NPC risk is still weak and mainly limited to high-incidence areas (Farrow *et al*, 1998; Kasum *et al*, 2002; Gallicchio *et al*, 2006). This indicated the need for future research.

In the current paper, we will carry out an exploratory principal component factor analysis (PCFA) to identify dietary patterns defined *a posteriori* in a multicentre case–control study on cancer of the nasopharynx conducted in a low-incidence country, like Italy. Previous analyses on single nutrients and NPC on those data showed some positive associations with higher intakes of cholesterol, saturated fatty acids, and phosphorus, and inverse associations with some carotenoids and total fibre (Polesel *et al*, 2012; Bidoli *et al*, 2013). Among foods, higher intakes of eggs were significantly associated to an increased NPC risk, whereas total vegetables, in particular raw and yellow/red vegetables, were inversely related to it (Polesel *et al*, 2013).

The identification of dietary patterns may allow to integrate our knowledge of the association between dietary habits and NPC risk in this low-incidence Mediterranean country.

MATERIALS AND METHODS

A case–control study on NPC was conducted between January 1992 and December 2008 within an established network of centres, including Aviano (Pordenone) and Milan in northern Italy, and Naples and Catania in southern Italy (Polesel *et al*, 2011, 2012; Bidoli *et al*, 2013). Cases were 198 Caucasian subjects (157 men, 41 women; median age 52 years, range 18–76 years) admitted to major teaching and general hospitals in the study areas with incident, histologically confirmed NPC, diagnosed no longer than 1 year before the interview, and with no history of cancer at other sites. Available cases included 137 (68.5%) undifferentiated NPCs (World Health Organization (WHO) type 3), 23 (11.5%) keratinising squamous cell carcinomas (WHO type 1; here referred to as differentiated NPCs), and 38 (20%) not otherwise specified NPCs (Shanmugaratnam, 1991). EBV status was defined based on the detection of EBV nuclear antigen in tissue samples and was available for 61 cases only. All the 57 undifferentiated NPCs tested and 2 out of 4 differentiated NPCs were EBV positive. Controls were 594 Caucasian subjects (471 men, 123 women; median age 52

years, range 19–76 years) admitted to the same hospitals as cases for a wide spectrum of acute, non-neoplastic conditions, unrelated to tobacco smoking or alcohol or to long-term modifications of diet. Three controls were frequency matched to each case according to age (± 2 years), sex, period of interview (± 2 years), and area of residence. Thirty-four percent of the controls were admitted for traumas, 32% for other orthopedic disorders, 22% for acute surgical conditions, and 12% for miscellaneous other illnesses. Less than 3% of both cases and controls originally contacted refused to participate. Cases and controls who agreed to participate in the study signed an informed consent based on recommendations from the individual hospitals' ethics boards.

Centrally trained interviewers administered a structured questionnaire to cases and controls during their hospital stay. The questionnaire included information on socio-demographic characteristics, anthropometric variables, life-style factors, including tobacco and alcohol, a problem-oriented personal medical history, family history of cancer, and, for women, menstrual and reproductive history.

Information on diet referred to the 2 years before diagnosis for cases (or hospital admission, for the controls) and was based on a FFQ that was validated for nutrient intake and tested for reproducibility for specific nutrients and food items (Franceschi *et al*, 1993, 1995; Decarli *et al*, 1996). Subjects were asked to indicate quantity and average weekly frequency of consumption for the period under investigation. The FFQ presented 78 foods and food groups, including some of the most common Italian recipes, and various types of alcoholic beverages. Intakes reported at least once a month but less than once a week were coded as 0.5 per week. Dietary supplements were not considered, given their rare consumption in this population. Other questions on fat-intake pattern, as well as portion size, were used to modulate the composition of recipes.

Italian food composition tables (Gnagnarella *et al*, 2004) were used to calculate intakes of total energy and various nutrients. Losses due to cooking were subtracted from the computation of the vitamin content, when appropriate.

Statistical analysis

Factorability of the original matrix. We conducted the analyses on a selected set of 28 major macro- and micro-nutrients. Nutrients were chosen to provide a comprehensive representation of the Italian diet and to assess their potential role in cancer risk. Moreover, we took into account existing relationships among nutrients to avoid over-representation of single nutrients and consequent artificially higher correlations.

We evaluated the correlation matrix of the nutrients to determine if it was factorable through both visual inspection of the matrix and statistical procedures, including Bartlett's test of sphericity, overall (Kaiser-Meyer-Olkin) and individual measures of sampling adequacy (Pett *et al*, 2003). Given the reassuring results obtained (see Results section), we applied a factor analysis to derive the dietary patterns.

Identification of dietary patterns. Exploratory PCFA (Johnson and Wichern, 2007) was carried out on the correlation matrix of the original nutrients of both cases and controls to describe the variance–covariance structure among nutrients in terms of a few underlying unobservable and randomly varying factors that are generally known as dietary patterns. We chose the number of factors to retain based on the following criteria: factor eigenvalue greater than 1, scree-plot construction and factor interpretability (Johnson and Wichern, 2007). We applied a varimax rotation to the factor loading matrix to achieve a simpler loading pattern. Nutrients with rotated factor loadings greater or equal to 0.63 in absolute value on a given factor were used to name the factors, and are indicated as 'dominant nutrients' hereafter (the

contribution that a factor gives to a nutrient's sample variance is equal to the square of its loading on that factor, so, if we choose a 0.63 cut-off, we expect a minimum contribution of approximately 0.40; Comrey and Lee, 1992). We calculated factor scores following the weighted least squares method. They indicate the degree to which each subject's diet conforms to one of the identified patterns.

To examine the robustness of the identified dietary patterns, we considered the following checks. First, we performed a principal axis factor analysis and a maximum likelihood factor analysis, after a logarithmic transformation of the original nutrients to improve normality in their joint distribution. Second, we confirmed the choice of the number of factors to retain referring to the Velicer's minimum average partial test (Velicer, 1976). Third, we calculated factor scores by applying the multiple regression method and standardising the results (Johnson and Wichern, 2007). Given the robustness in the patterns identified by these complementary checks, we calculated factor scores in all subsequent analyses by using results from the original overall PCFA and applying the weighted least squares method.

To assess reliability and refine the identified factors, we evaluated the internal consistency of those nutrients that load more than 0.40 (in absolute value) on any factor using standardised Cronbach's coefficient α (Cronbach, 1951; Pett *et al*, 2003).

To confirm internal reproducibility of the identified patterns, factor analysis was performed separately in two (randomly selected) subsets of the original data using the same approach of the main analysis. The procedure was repeated several times.

To improve interpretability of the identified dietary patterns, we calculated the values of the Spearman rank correlation coefficient between the continuous factor scores derived from our factor analysis and the weekly number of portions for 29 selected food groups defined on the same data.

Risk estimates. For each factor, participants were grouped into three categories according to tertiles of factor scores among the controls.

We estimated the odds ratios (ORs) and the corresponding 95% confidence intervals (CIs) for tertile categories and a continuous increment in factor scores using unconditional multiple logistic regression models. We fitted separate models for each factor and a composite model including all the factors simultaneously. Tests for linear trend were computed for all these models scoring the tertiles as numbers from 1 to 3. We included in each model the same set of potential confounding and risk factors: age, sex, area of residence, education, year of interview, alcohol drinking, and tobacco smoking. We carried out stratified analyses by age, alcohol, and tobacco.

Calculations were performed using the open-source statistical computing environment R (Ihaka and Gentleman, 1996; R Core Team, 2014), with its libraries psych (Revelle, 2014) and GPArotation (Bernaards and Jennrich, 2005).

RESULTS

The correlation matrix of the original nutrients was amenable to factor analysis. Visual inspection of the correlation matrix revealed that each nutrient showed at least 12 correlation coefficients $\geqslant 0.30$ in absolute value (data not shown), thus allowing to perform the analyses on the entire set of selected nutrients. Table 1 reports results on statistical procedures for checking matrix factorability. Bartlett's test of sphericity allowed to reject the null hypothesis that the correlation matrix is an identity matrix (P-value < 0.001). The Kaiser-Meyer-Olkin statistic was equal to about 0.85, suggesting that the sample size was adequate relative to the number of nutrients. In addition, the individual measures of sampling adequacy gave reassuring results, with 21 nutrients having measures $\geqslant 0.80$, 5 in the 0.70 s, and only 2 nutrients with measures between 0.60 and 0.70.

Table 2 presents the factor loading matrix for the five retained factors, together with the corresponding communality estimates. The retained factors explained $\sim 80\%$ of the total variance in the original data set. The greater was the loading of a given nutrient to a factor, the higher was the contribution of that nutrient to the factor. The first factor, named *Animal products*, had the greatest loadings on calcium, riboflavin, phosphorus, saturated fatty acids, animal protein, and cholesterol (20.90% of the total variance explained); the second factor, named *Starch-rich*, was characterised by the greatest loadings on starch, vegetable protein, and sodium (15.85% of the total variance explained); the third factor, named *Vitamins and fibre*, had the greatest loadings on vitamin C and total fibre, *β*-carotene equivalents, and total folate (15.12% of the total variance explained); the fourth factor, named *Animal unsaturated fatty acids* (AUFAs), had the greatest loadings on vitamin D, other polyunsaturated fatty acids, and niacin (15.06% of the total variance explained); the fifth factor, named *Vegetable unsaturated fatty acids* (VUFAs), had the greatest loadings on linoleic acid, vitamin E, and linolenic acid (12.67% of the total variance explained). Except for lycopene, all the examined nutrients showed at least one loading greater than 0.40 on any factor, thus proving to be good candidates for the original set of nutrients. The percentages of nutrient variance explained by all the retained factors (communality estimates) were generally satisfactory and exceeded the threshold of 80% for most nutrients.

Most of the nutrients contributed to high reliability. Standardised coefficient α for each factor were greater than 0.90, indicating that at least 90% of the variance of the total scores on these subscales for each factor can be attributed to reliable, systematic variance. Standardised *coefficient α, when item deleted* were generally lower than the corresponding overall standardised coefficient α for the same factor, although the differences were small (Appendix Table A1). The internal reproducibility of the sets of patterns identified in the two split-samples was reassuring. Factor loading matrices and communalities obtained from the two

Table 1. Factorability of the correlation matrix of the original nutrients: Bartlett's test of sphericity and measures of sampling adequacy	
Bartlett's test of sphericity: *P*-value < 0.001	
Kaiser-Meyer-Olkin statistic: overall measure of sampling adequacy[a]: 0.85	
Individual measures of sampling adequacy[a]:	
< 0.70	Linoleic acid, retinol
0.70–0.80	Monounsaturated fatty acids, vitamin E, lycopene, calcium, total fibre
0.80–0.90	Starch, iron, vitamin C, vitamin B2, vitamin D, animal protein, niacin, vegetable protein, potassium, cholesterol, saturated fatty acids, vitamin B1
⩾ 0.90	Zinc, total folate, other polyunsaturated fatty acids, sodium, vitamin B6, soluble carbohydrates, *β*-carotene equivalents, phosphorus, linolenic acid
[a]Overall and individual measures of sampling adequacy range between 0 and 1, with values >0.60 indicating a satisfactory size.	

Table 2. Factor loading matrix[a], COMM and explained VAR for the five major dietary patterns identified by factor analysis

Nutrient[b]	Animal products	Starch rich	Vitamins and fibre	AUFAs	VUFAs	COMM
Animal protein	− **0.70**	0.20	—	− 0.55	− 0.26	0.91
Vegetable protein	− 0.17	**0.88**	− 0.29	− 0.15	− 0.23	0.96
Cholesterol	− **0.64**	0.27	—	− 0.52	− 0.23	0.81
Saturated fatty acids	− **0.73**	0.22	− 0.14	− 0.29	− 0.41	0.86
Monounsaturated fatty acids	− 0.25	0.24	− 0.27	− 0.42	− 0.50	0.62
Linoleic acid	− 0.21	0.21	− 0.11	− 0.10	− **0.85**	0.84
Linolenic acid	− 0.39	0.16	—	− 0.19	− **0.78**	0.83
Other polyunsaturated fatty acids	− 0.20	0.22	—	− **0.80**	− 0.39	0.87
Soluble carbohydrates	− 0.54	0.23	− 0.53	—	− 0.10	0.64
Starch	− 0.20	**0.92**	—	− 0.11	− 0.17	0.93
Sodium	− 0.46	**0.76**	—	− 0.21	− 0.16	0.86
Calcium	− **0.90**	—	− 0.23	—	− 0.15	0.90
Potassium	− 0.48	0.39	− 0.60	− 0.32	− 0.23	0.89
Phosphorus	− **0.75**	0.37	− 0.25	− 0.35	− 0.24	0.95
Iron	− 0.33	0.47	− 0.33	− 0.40	− 0.32	0.70
Zinc	− 0.56	0.46	− 0.21	− 0.50	− 0.32	0.93
Thiamin (vitamin B1)	− 0.53	0.48	− 0.39	− 0.35	− 0.22	0.84
Riboflavin (vitamin B2)	− **0.78**	0.19	− 0.37	− 0.32	− 0.11	0.90
Vitamin B6	− 0.47	0.40	− 0.47	− 0.46	− 0.28	0.89
Total folate	− 0.37	0.40	− **0.65**	− 0.33	0.17	0.85
Niacin	− 0.28	0.43	− 0.29	− **0.64**	− 0.23	0.81
Vitamin C	− 0.16	—	− **0.83**	—	—	0.73
Retinol	− 0.19	—	− 0.23	− 0.49	—	0.33
β-Carotene equivalents	− 0.12	—	− **0.70**	—	− 0.26	0.57
Lycopene	—	0.36	− 0.31	− 0.18	− 0.34	0.38
Vitamin D	− 0.19	0.11	—	− **0.83**	− 0.12	0.74
Vitamin E	− 0.16	0.21	− 0.43	− 0.22	− **0.80**	0.93
Total fibre (Englyst)	− 0.16	0.42	− **0.77**	—	− 0.14	0.82
Proportion of VAR explained (%)	20.90	15.85	15.12	15.06	12.67	
Cumulative VAR explained (%)	20.90	36.75	51.87	66.94	79.60	

Abbreviations: AUFAs = animal unsaturated fatty acids; COMM, communality estimates; VAR = variances; VUFAs = vegetable unsaturated fatty acids.
[a]Estimates from a principal component factor analysis performed on 28 nutrients. The magnitude of each loading measures the importance of the corresponding nutrient to the factor.
[b]Loadings greater or equal to 0.63 (in absolute value) defined dominant nutrients for each factor, and were shown in bold typeface; loadings smaller than 0.1 (in absolute value) were suppressed.

factor analyses on the split-samples were almost identical (data not shown).

Table 3 shows the values of the Spearman rank correlation coefficient between the continuous factor scores derived from our factor analysis based on nutrient intakes and the weekly number of portions of 29 food groups on the same data set. For the *Animal product* pattern, the highest values were with milk, cheese, desserts, eggs, sugar and candies, potatoes, butter and margarine, and red meat; for the *Starch-rich* pattern, the highest values of the Spearman coefficient were with bread, desserts, red meat, and pasta and rice; for the *Vitamins and fibre* pattern, the highest values were with other fruits, citrus fruits, fruiting vegetables, root vegetables, other vegetables, olive oil, leafy vegetables, pulses, cruciferous vegetables, and liver; for the *AUFA* pattern, the highest values were with fish, liver, red meat, olive oil, eggs, processed meat, and leafy vegetables; for the *VUFA* pattern, the highest values of the Spearman coefficient were with specified seed oils, unspecified seed oils, red meat, leafy vegetables, butter and margarine.

Table 4 gives the ORs and corresponding CIs for NPC by tertiles of factor scores and continuous factor scores for the retained

dietary patterns. Results refer to the composite model including all the five factors simultaneously, together with the relevant confounding and risk variables. Higher intakes of the dominant nutrients for the *Animal product* dietary pattern were positively associated to NPC risk (OR = 2.62, 95% CI = 1.67–4.13 for the highest *vs* the lowest score tertile, *P* for trend < 0.001). A positive association was also observed between the *Starch-rich* pattern and NPC risk (OR = 2.05, 95% CI = 1.27–3.33, *P* for trend = 0.022), whereas a nonsignificant but inverse association was evident with the *Vitamins and fibre* pattern (OR = 0.68, 95% CI = 0.44–1.05, *P* for trend = 0.223). In addition, there was a borderline significant association between the *AUFA* pattern and NPC risk (OR = 1.55, 95% CI = 1.00–2.39, *P* for trend = 0.038). Finally, the *VUFA* pattern was positively related to NPC risk: the OR was 1.90 (95% CI = 1.22–2.96, *P* for trend = 0.011) for subjects consuming the highest intakes of the dominant nutrients for this pattern. Consistent results were observed for the five models including each factor separately.

Table 5 shows the ORs of NPC and corresponding CIs for the identified dietary patterns in strata of age, tobacco smoking, and alcohol drinking. There is no significant heterogeneity across strata

Table 3. Spearman rank correlation coefficients between continuous factor scores derived from factor analysis on nutrient intakes and weekly number of portions for 29 selected food groups defined on the same data

Food group[a]	Animal products	Starch-rich	Vitamins and fibre	AUFAs	VUFAs
Milk	**0.68**	−0.11	0.13	—	−0.11
Coffee	—	0.12	—	0.18	—
Tea and decaffeinated coffee	—	—	0.11	—	—
Bread	—	**0.85**	—	—	—
Pasta and rice	—	**0.22**	0.11	—	0.16
Soup	—	—	—	—	—
Eggs	**0.33**	—	—	**0.27**	0.16
White meat	—	—	0.12	0.15	—
Red meat	**0.22**	**0.26**	—	**0.42**	**0.32**
Liver	—	—	**0.21**	**0.43**	—
Processed meat	0.17	0.16	—	**0.27**	0.10
Fish	—	—	0.12	**0.62**	—
Cheese	**0.62**	—	—	0.14	0.17
Potatoes	**0.25**	0.17	—	0.19	0.14
Pulses	—	0.10	**0.25**	—	0.11
Leafy vegetables	—	—	**0.29**	**0.22**	**0.24**
Fruiting vegetables	—	—	**0.40**	0.12	0.16
Root vegetables	—	−0.10	**0.37**	—	0.14
Cruciferous vegetables	—	—	**0.22**	0.11	—
Other vegetables	—	—	**0.31**	—	0.19
Citrus fruits	—	—	**0.53**	—	—
Other fruits	0.14	—	**0.68**	—	—
Soft drinks and fruit juice	0.17	—	0.13	—	0.17
Desserts	**0.34**	**0.29**	—	—	—
Sugar and candies	**0.31**	0.15	—	—	0.14
Butter and margarine	**0.23**	—	—	0.10	**0.23**
Specified seed oils	—	—	—	−0.11	**0.55**
Unspecified seed oils	0.18	—	−0.13	—	**0.51**
Olive oil	—	0.13	**0.29**	**0.36**	0.18

Abbreviations: AUFAs = animal unsaturated fatty acids; VUFAs = vegetable unsaturated fatty acids.
[a]Correlations greater or equal to 0.20 (in absolute value) were shown in bold typeface; correlations smaller than 0.1 (in absolute value) were suppressed.

Table 4. ORs[a] of NPC and corresponding 95% CIs according to tertiles of factor scores and continuous factor scores from a principal component factor analysis

Dietary pattern	OR (95% CI), tertile category[b]			P_{trend}^{d}	OR (95% CI) Continuous
	I[c]	II	III		
Animal products	1	1.24 (0.77–2.00)	2.62 (1.67–4.13)	<0.001	1.66 (1.32–2.07)
Starch rich	1	1.53 (0.97–2.41)	2.05 (1.27–3.33)	0.022	1.38 (1.09–1.75)
Vitamins and fibre	1	0.65 (0.42–1.01)	0.68 (0.44–1.05)	0.223	0.82 (0.66–1.02)
AUFA	1	0.78 (0.49–1.25)	1.55 (1.00–2.39)	0.038	1.27 (1.01–1.58)
VUFA	1	1.19 (0.75–1.88)	1.90 (1.22–2.96)	0.011	1.38 (1.10–1.72)

Abbreviations: AUFAs = animal unsaturated fatty acids; CI = confidence interval; OR = odds ratio; VUFAs = vegetable unsaturated fatty acids.
[a]Estimates from a multiple logistic regression model adjusted for age, sex, area of residence, education, year of interview, tobacco smoking, and alcohol drinking. Results refer to the composite model including all the five factors simultaneously.
[b]Cases within tertile categories of factor scores for each dietary pattern were as follows: *Animal products* (I tertile: 47, II tertile: 56; III tertile: 93); *Starch rich* (I tertile: 54; II tertile: 68; III tertile: 74); *Vitamins and fibre* (I tertile: 82; II tertile: 56; III tertile 58); *AUFAs* (I tertile: 5; II tertile: 52; III tertile: 89); *VUFAs* (I tertile: 54; II tertile: 61; III tertile: 81). As tertile categories were calculated on control subjects, the number of controls within categories was approximately of 197 for each pattern.
[c]Reference category.
[d]P-value for linear trend.

Table 5. ORs[a] of NPC and corresponding 95% CIs on tertiles of factor scores in strata of age, tobacco smoking, and alcohol drinking

Dietary pattern	Stratum[b]	OR (95% CI), tertile category			P_{strata}[d]
		I[c]	II	III	
Animal products	Age ⩽ 52 years	1	0.95 (0.47–1.92)	2.87 (1.49–5.55)	0.495
	Age > 52 years	1	1.58 (0.78–3.22)	2.67 (1.36–5.24)	
	Never/former smoker	1	1.26 (0.69–2.29)	2.45 (1.37–4.39)	0.852
	Current smoker	1	1.44 (0.58–3.55)	3.73 (1.60–8.66)	
	Alcohol drinker ⩽ 14.5 drinks per week	1	1.61 (0.81–3.19)	3.03 (1.56–5.91)	0.478
	Alcohol drinker > 14.5 drinks per week	1	0.95 (0.46–1.94)	2.39 (1.24–4.62)	
Starch rich	Age ⩽ 52 years	1	0.10 (0.52–1.93)	1.67 (0.81–3.41)	0.116
	Age > 52 years	1	2.76 (1.35–5.61)	3.32 (1.57–7.02)	
	Never/former smoker	1	1.27 (0.73–2.21)	1.95 (1.08–3.52)	0.447
	Current smoker	1	2.56 (1.02–6.41)	2.78 (1.07–7.24)	
	Alcohol drinker ⩽ 14.5 drinks per week	1	1.31 (0.70–2.45)	2.70 (1.35–5.43)	0.290
	Alcohol drinker > 14.5 drinks per week	1	1.96 (0.92–4.15)	1.70 (0.80–3.63)	
Vitamins and fibre	Age ⩽ 52 years	1	0.82 (0.42–1.58)	0.72 (0.38–1.34)	0.667
	Age > 52 years	1	0.54 (0.29–1.03)	0.61 (0.32–1.17)	
	Never/former smoker	1	0.60 (0.35–1.04)	0.60 (0.35–1.05)	0.952
	Current smoker	1	0.57 (0.26–1.27)	0.62 (0.28–1.38)	
	Alcohol drinker ⩽ 14.5 drinks per week	1	0.38 (0.20–0.73)	0.46 (0.24–0.87)	0.211
	Alcohol drinker > 14.5 drinks per week	1	1.09 (0.57–2.07)	0.96 (0.51–1.80)	
AUFAs	Age ⩽ 52 years	1	0.58 (0.28–1.21)	1.50 (0.79–2.84)	0.395
	Age > 52 years	1	1.15 (0.60–2.22)	1.66 (0.87–3.14)	
	Never/former smoker	1	0.76 (0.42–1.35)	1.34 (0.79–2.30)	0.616
	Current smoker	1	0.91 (0.37–2.23)	1.92 (0.83–4.44)	
	Alcohol drinker ⩽ 14.5 drinks per week	1	0.78 (0.41–1.49)	1.31 (0.70–2.44)	0.365
	Alcohol drinker > 14.5 drinks per week	1	0.70 (0.34–1.46)	1.75 (0.89–3.44)	
VUFAs	Age ⩽ 52 years	1	1.42 (0.73–2.79)	2.01 (1.04–3.91)	0.669
	Age > 52 years	1	0.94 (0.47–1.86)	1.77 (0.94–3.32)	
	Never/former smoker	1	1.31 (0.75–2.31)	1.90 (1.08–3.35)	0.690
	Current smoker	1	0.97 (0.42–2.23)	2.08 (0.95–4.53)	
	Alcohol drinker ⩽ 14.5 drinks per week	1	0.98 (0.53–1.84)	2.23 (1.15–4.30)	0.339
	Alcohol drinker > 14.5 drinks per week	1	1.44 (0.70–2.96)	1.55 (0.81–2.96)	

Abbreviations: AUFAs = animal unsaturated fatty acids; CI = confidence interval; OR = odds ratio; VUFAs = vegetable unsaturated fatty acids.

[a] Estimates from multiple logistic regression models adjusted for age, sex, area of residence, education, year of interview, tobacco smoking, and alcohol drinking, when appropriate. Results refer to the composite models including all the five factors simultaneously.

[b] 52 years was the median age in the overall sample; 14.5 drinks per week was the median alcohol consumption in the overall sample.

[c] Reference category.

[d] P-value for heterogeneity across strata.

for the five patterns. However, subjects in the highest tertile of the *Vitamins and fibre* pattern and drinking less than (or equal to) the median intake of the overall population (14.5 drinks per week) had a significantly lower NPC risk: OR = 0.46 (95% CI = 0.24–0.87) *vs* OR = 0.96 (95% CI = 0.51–1.80) for subjects drinking > 14.5 drinks per week.

DISCUSSION

The present analysis identified five major dietary patterns that explained about 80% of the total variance in the nutrient intakes of this Italian population. After adjustment for several confounders and mutual adjustment for all the other patterns, the *Animal products*, the *Starch-rich,* and the *VUFA* patterns were positively related to NPC, whereas the *AUFA* pattern, still positively related,

showed a borderline significant association with NPC. The *Vitamins and fibre* pattern was inversely, but non-significantly, associated with it.

To our knowledge, there is only one study investigating the role of dietary patterns, identified through a factor analysis, on the risk of NPC (Armstrong *et al*, 1998). This case–control study was carried in the high-incidence Chinese population of the Federal Territory (Kuala Lumpur) and State of Selangor, Malaysia, and assessed the association of NPC with single food items (queried in a 55-item FFQ) and dietary patterns, at two time points (childhood and 5 years before diagnosis). Factor analysis was applied on a selection of 19 foods identified as being associated with NPC risk on the basis of the single-food-item analysis. In accordance with our work, the results showed a significant positive association with a dietary pattern based on meats (Factor 3), including pork/beef, liver and other organ meats (∼60% increase in NPC risk with a change from the 25th to the 75th percentile of each estimated

factor score), which is similar to our *Animal product* and *AUFA* patterns. The decrease in risk associated to their Factor 1 – loading highly on fresh fruit and vegetable consumption, with modest involvement of shrimp and preserved fruit – was replicated in our study, although the association was significant in subjects with at most 52 years only. In addition, the strongest positive association with NPC was found with a pattern based on salted, preserved foods (∼95% increase), which was not typical to the Italian diet. However, the preliminary selection of foods based on significant findings from the single-food-item analysis did not provided a comprehensive picture of the overall diet and therefore prevented from a fair comparison of results.

The composition of the patterns associated with NPC is consistent with findings on nutrients and food groups from our (Polesel *et al*, 2012, 2013; Bidoli *et al*, 2013) and some other studies (Farrow *et al*, 1998; Kasum *et al*, 2002; Chang and Adami, 2006; Gallicchio *et al*, 2006; Turkoz *et al*, 2011; Jia and Qin, 2012) conducted in low-incidence areas, where meat, fresh fruit and vegetables, β-carotene, and vitamin C were associated, to a greater or lesser extent, to NPC. However, apart from the established association with salted fish and salt-preserved foods, typical of high-incidence areas, the evidence on dietary components and NPC is still limited (World Cancer Research Fund/American Institute for Cancer Research, 2007, Jia and Qin, 2012).

Factor analysis is the most common statistical method used today to derive dietary patterns defined *a posteriori* in studies assessing the association between diet and cancer. A proper use of this technique allows to enter in the same regression model different aspects of the overall diet without suffering from severe multicollinearity. As, for instance, people who consume large amounts of red and processed meats tend to consume less poultry, fish, and vegetables, and vice versa, an apparent detrimental effect of these meats could possibly be due, at least in part, to low intakes of these favourable foods (World Cancer Research Fund/American Institute for Cancer Research, 2007). Factor analysis may take care of these issues in a more elaborate way than traditional approaches, and may offer advantages over simultaneous statistical adjustment for food groups used as independent exploratory covariates (Armstrong *et al*, 1998; Imamura *et al*, 2009). However, subjective decisions are involved at each stage of this process (Newby and Tucker, 2004; Edefonti *et al*, 2009). In this paper, we proposed the use of tools for the assessment of the factorability of the correlation matrix. Moreover, we stressed the importance of performing a series of separate robustness analyses to evaluate the impact of all the decisions taken during the analysis. For instance, we checked the solution method using different approaches, we considered two different methods for calculating factor scores and applied factor analysis separately in subgroups of subjects. Only after a complete series of checks, a given data reduction solution is likely to be independent from the specific statistical procedure used to derive it.

Given the strong evidence supporting a causal role of EBV in the development of NPC (Chang and Adami, 2006), the lack of information on EBV status in most of the NPC cases and controls was a potential weakness of the present study. In any case, EBV-status is unlikely to confound the association between dietary habits and NPC. In our case, the role of the identified dietary patterns did not substantially change when analyses were restricted to the undifferentiated NPC patients only, although the *Vitamins and fibre* pattern became significant in the last tertile category (OR = 0.58, 95% CI = 0.34–0.98). The small sample size was an additional weakness, but this is due to the fact that the study was carried out in a low-incidence area, which is actually a strength of the study. Indeed, in low-incidence areas, the choice of an adequate sample size should balance between avoiding statistical power issues, related to an insufficient number of subjects, and limiting the possibility of changes in life-style habits, especially diet, possibly occurring over a long period of observation.

This study has some limitations, but also several strengths of hospital-based case–control investigations. The high participation and the comparable catchment areas of cases and controls have likely avoided substantial selection biases, and comparison of cases with major diagnostic categories of controls led to similar results. Bias in the recall of food intake by cases should be small, given the limited knowledge and attention paid in this population to the possible relationship between diet and NPC. The comparability of the recall between cases and controls was improved by interviewing all the subjects in a hospital setting (D'Avanzo *et al*, 1997). Information on alcohol, tobacco, and diet was satisfactorily reproducible (D'Avanzo *et al*, 1996; Ferraroni *et al*, 1996). Subjects with admission diagnosis related to tobacco smoking, alcohol drinking, and diet modifications were not considered as eligible controls.

In conclusion, the findings of the present study point to the presence of positive associations between NPC and the *Animal product*, the *Starch-rich*, and the *VUFA* dietary patterns, respectively. A nonsignificant, but inverse, association was also evident with the *Vitamins and fibre* pattern in our low-incidence country.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Armstrong RW, Imrey PB, Lye MS, Armstrong MJ, Yu MC, Sani S (1998) Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. *Int J Cancer* **77**: 228–235.

Bernaards CA, Jennrich RI (2005) Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educ Psychol Meas* **65**: 676–696.

Bidoli E, Pelucchi C, Polesel J, Negri E, Barzan L, Franchin G, Franceschi S, Serraino D, De Paoli P, La Vecchia C, Talamini R (2013) Fiber intake and risk of nasopharyngeal carcinoma: a case-control study. *Nutr Cancer* **65**: 1157–1163.

Chang ET, Adami HO (2006) The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* **15**: 1765–1767.

Comrey AL, Lee HB (1992) *A First Course in Factor Analysis*. 2nd edn (Lawrence Erlbaum Associates, Publishers: Hillsdale, New Jersey.

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **18**: 297–334.

Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, Boyle P (eds) (2007) *Cancer Incidence in Five Continents*. Vol IX. IARC Scientific Publications: Lyon, France.

D'Avanzo B, La Vecchia C, Katsouyanni K, Negri E, Trichopoulos D (1996) Reliability of information on cigarette smoking and beverage consumption provided by hospital controls. *Epidemiology* **7**: 312–315.

D'Avanzo B, La Vecchia C, Katsouyanni K, Negri E, Trichopoulos D (1997) An assessment, and reproducibility of food frequency data provided by hospital controls. *Eur J Cancer Prev* **6**: 288–293.

Decarli A, Franceschi S, Ferraroni M, Gnagnarella P, Parpinel MT, La Vecchia C, Negri E, Salvini S, Falcini F, Giacosa A (1996) Validation of a food-frequency questionnaire to assess dietary intakes in cancer studies in Italy. Results for specific nutrients. *Ann Epidemiol* **6**: 110–118.

Edefonti V, Randi G, La Vecchia C, Ferraroni M, Decarli A (2009) Dietary patterns and breast cancer: a review with focus on methodological issues. *Nutr Rev* **67**: 297–314.

Farrow DC, Vaughan TL, Berwick M, Lynch CF, Swanson GM, Lyon JL (1998) Diet and nasopharyngeal cancer in a low-risk population. *Int J Cancer* **78**: 675–679.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**: 2893–2897.

Ferraroni M, Decarli A, Franceschi S, La Vecchia C, Enard L, Negri E, Parpinel M, Salvini S (1996) Validity and reproducibility of alcohol consumption in Italy. *Int J Epidemiol* **25**: 775–782.

Franceschi S, Barbone F, Negri E, Decarli A, Ferraroni M, Filiberti R, Giacosa A, Gnagnarella P, Nanni O, Salvini S, La Vecchia C (1995) Reproducibility of an Italian food frequency questionnaire for cancer studies. Results for specific nutrients. *Ann Epidemiol* **5**: 69–75.

Franceschi S, Negri E, Salvini S, Decarli A, Ferraroni M, Filiberti R, Giacosa A, Talamini R, Nanni O, Panarello G, La Vecchia C (1993) Reproducibility of an Italian food frequency questionnaire for cancer studies: results for specific food items. *Eur J Cancer* **29**: 2298–2305.

Gallicchio L, Matanoski G, Tao XG, Chen L, Lam TK, Boyd K, Robinson KA, Balick L, Mickelson S, Caulfield LE, Herman JG, Guallar E, Alberg AJ (2006) Adulthood consumption of preserved and nonpreserved vegetables and the risk of nasopharyngeal carcinoma: a systematic review. *Int J Cancer* **119**: 1125–1125.

Gnagnarella P, Parpinel M, Salvini S, Franceschi S, Palli D, Boyle P (2004) The update of the Italian Food Composition Database. *J Food Composit Anal* **17**: 509–522.

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

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comp Graph Stat* **5**: 299–314.

Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF (2009) Confounding by dietary patterns of the inverse association between alcohol consumption and type 2 diabetes risk. *Am J Epidemiol* **170**: 37–45.

International Agency for Research on Cancer (2004) Tobacco Smoke and Involuntary Smoking. *IARC Monogr Eval Carcinog Risk Hum* Vol. **83**: IARC: Lyon, France.

International Agency for Research on Cancer (2007) Alcohol Consumption and Ethyl Carbamate. *IARC Monogr Eval Carcinog Risks Hum* Vol. **96**: IARC: Lyon, France.

Jemal A, Bray F, Center M, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA. *Cancer J Clin* **61**: 69–90.

Jia WH, Qin HD (2012) Non-viral environmental risk factors for nasopharyngeal carcinoma: a systematic review. *Semin Cancer Biol* **22**: 117–126.

Johnson RA, Wichern DW (2007) *Applied Multivariate Statistical Analysis.* 6th edn, Prentice-Hall: Upper Saddle River, New Jersey.

Kasum CM, Jacobs Jr DR, Nicodemus K, Folsom AR (2002) Dietary risk factors for upper aerodigestive tract cancers. *Int J Cancer* **99**: 267–272.

Lee HP, Gourley L, Duffy SW, Esteve J, Lee J, Day NE (1994) Preserved foods and nasopharyngeal carcinoma: a case-control study among Singapore Chinese. *Int J Cancer* **59**: 585–590.

Newby PK, Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**: 177–203.

Pett MA, Lackey NR, Sullivan JJ (2003) *Making Sense of Factor Analysis: the Use of Factor Analysis for Instrument Development in Health Care Research.* Sage Publications: Thousand Oaks, California.

Polesel J, Franceschi S, Talamini R, Negri E, Barzan L, Montella M, Libra M, Vaccher E, Franchin G, La Vecchia C, Serraino D (2011) Tobacco smoking, alcohol drinking, and the risk of different histological types of nasopharyngeal cancer in a low-risk population. *Oral Oncol* **47**: 541–545.

Polesel J, Negri E, Serraino D, Parpinel M, Barzan L, Libra M, Bosetti C, Garavello W, Montella M, La Vecchia C, Franceschi S, Talamini R (2012) Dietary intakes of carotenoids and other nutrients in the risk of nasopharyngeal carcinoma: a case-control study in Italy. *Br J Cancer* **107**: 1580–1583.

Polesel J, Serraino D, Negri E, Barzan L, Vaccher E, Montella M, Zucchetto A, Garavello W, Franceschi S, La Vecchia C, Talamini R (2013) Consumption of fruit, vegetables, and other food groups and the risk of nasopharyngeal carcinoma. *Cancer Causes Control* **24**: 1157–1165.

R Core Team (2014) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria, URL http://www.R-project.org, last accessed on: 31 July 2014.

Revelle W (2014) *Psych: Procedures for Psychological, Psychometric, and Personality Research.* Northwestern University: Evanston, Illinois, URL http://CRAN.R-project.org/package=psychR package version 1.4.5, last accessed on: 31 July 2014.

Shanmugaratnam K (1991) Histological classification of tumours of the upper respiratory tract and ear. *In Histological Typing of Tumours of the Upper Respiratory Tract and Ear.* 2nd edn, 3–18. Springer.

Turkoz FP, Celenkoglu G, Dogu GG, Kalender ME, Coskun U, Alkis N, Ozkan M, Turk HM, Arslan UY (2011) Risk factors of nasopharyngeal carcinoma in Turkeyan epidemiological survey of the Anatolian Society of Medical Oncology. *Asian Pac J Cancer Prev* **12**: 3017–3021.

Velicer WF (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika* **41**: 321–327.

Vokes EE, Liebowitz DN, Weichselbaum RR (1997) Nasopharyngeal carcinoma. *Lancet* **350**: 1087–1091.

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.

APPENDIX

Appendix A1. Standardised Cronbach's coefficient α[a] for each factor and standardised Cronbach's *coefficient α when item deleted*[a] for each factor and nutrient

	Animal products	Starch rich	Vitamins and fibre	AUFAs	VUFAs
Standardised α when item deleted					
Animal protein	0.97	—	—	0.92	0.90
Vegetable protein	—	0.95	—	—	—
Cholesterol	0.97	—	—	—	—
Saturated fatty acids	0.97	—	—	—	—
Monounsaturated fatty acids	—	—	—	0.93	0.90
Linoleic acid	—	—	—	—	0.89
Linolenic acid	—	—	—	—	0.88
Other polyunsaturated fatty acids	—	—	—	0.93	—
Soluble carbohydrates	0.97	—	0.92	—	—
Starch	—	0.96	—	—	—
Sodium	0.97	0.95	—	—	—
Calcium	0.97	—	—	—	—
Potassium	0.97	—	0.90	—	—
Phosphorus	0.96	—	—	0.92	—
Iron	—	0.95	—	—	—
Zinc	—	0.95	—	0.92	—
Thiamin (vitamin B1)	0.97	0.95	—	—	—
Riboflavin (vitamin B2)	0.97	—	—	—	—
Vitamin B6	—	—	0.91	0.92	—
Total folate	—	0.95	0.91	—	—
Niacin	—	0.95	—	0.92	—
Vitamin C	—	—	0.92	—	—
Retinol	—	—	—	0.95	—
β-Carotene equivalents	—	—	0.93	—	—
Lycopene	—	—	—	—	—
Vitamin D	—	—	—	0.93	—
Vitamin E	—	—	0.92	—	0.86
Total fibre	—	0.96	0.91	—	—
Standardised α	**0.97**	**0.96**	**0.93**	**0.94**	**0.91**

Abbreviations: AUFAs = animal unsaturated fatty acids; VUFAs = vegetable unsaturated fatty acids.
[a]Cronbach's coefficient α generally range between 0 and 1.