# Supplemental Digital Content

## Manuscript: "Shared and study-specific dietary patterns and their association with head and neck cancer risk in an international consortium"

## eAppendix

We included in the main text methods and results of a multi-study factor analysis based on controls only (controls-only analysis from here onwards). We also carried out a parallel analysis based on the combined sample of head and neck cancer cases and controls (cases+controls multi-study factor analysis from here onwards). In this eAppendix, we include additional details on the former analysis (methods and results) and main results from the latter analysis.

## Statistical analysis

In the following, we describe the main steps of the statistical analysis, which includes dietary pattern identification from controls-only multi-study factor analysis and assessment of related head and neck cancer risk.

### *Identification of dietary patterns*

We carried out multi-study factor analysis[1] to describe the variance-covariance structure among nutrients in terms of a few underlying unobservable and randomly varying shared

and study-specific factors, or dietary patterns. Specifically, we considered S=7 studies, each represented by the same set of P=23 nutrients. Study s has $n_s$ control subjects, each represented by a P-dimensional log-transformed and standardized data vector, $x_{is}$, with $i=1,\ldots,n_s$, $s=1,\ldots,S$. The $x_{is}$ are expressed by multi-study factor analysis in terms of K shared factors and $J_s$ additional study-specific factors, giving a total $T_s=K+J_s$ factors. Let $f_{is}$ be the $(K \times 1)$ shared latent factor vector for subject i in study s, and $\Phi$ be the $(P \times K)$ shared factor-loading matrix. Moreover, let $l_{is}$ be the $(J_s \times 1)$ study-specific latent factor vector and $\Lambda_s$ be the $(P \times J_s)$ study-specific factor-loading matrix. Multi-study factor analysis assumes that the P-dimensional vector $x_{is}$ can be written as in the following equation:

$$x_{is} = \Phi f_{is} + \Lambda_s l_{is} + e_{is} \quad i=1,\ldots, n_s \quad s=1,\ldots, S,$$

where the error term $e_{is}$ has a diagonal covariance matrix $\Psi_s = \mathrm{diag}(\psi_{s1},\ldots, \psi_{sp})$.

The corresponding likelihood is a product over studies of the usual product over subjects found in standard maximum likelihood factor analysis. In addition, each product term includes the corresponding study-specific correlation matrix that reflects the simultaneous presence of the shared and study-specific factors: $\Sigma_s = \Phi \Phi^T + \Lambda_s \Lambda_s^T + \Psi_s$.

Multi-study factor analysis is fitted using the Expectation Conditional Maximization algorithm[2], which is a generalization of the Expectation Maximization algorithm[3] used in standard maximum likelihood factor analysis. In multi-study factor analysis, the maximization step of the Expectation Maximization algorithm is replaced by a set of sequential conditional maximization steps for each parameter or groups of parameters.

To choose the number of factors to retain, we first estimated the total number of factors, $T_s$, for each of the S studies, using a combination of standard techniques for factor

analysis, including Horn's parallel analysis, Cattell's scree plot, and the Steiger's root mean square error of approximation index[4]. Next, we used the Akaike Information Criterion[5] on the overall model to select the number of shared factors, K. The number of study-specific factors, $J_s$, for each study s, was then found by difference as $T_s$ - K, s=1,…,S. A global Akaike Information Criterion was also used to identify the optimal pair (K, $J_s$).

We then applied a varimax rotation to the factor-loading matrix of the shared factors to achieve a better-defined loading structure. To name the 'dominant nutrients'[6], we used nutrients with a shared (rotated) factor loading of at least 0.60 or a study-specific (unrotated) factor loading of at least |0.25|. We derived nutrient communalities from the overall factor-loading matrix, [Φ | $\Lambda_s$], obtained juxtaposing the shared and study-specific factor loadings derived from the controls-only analysis.

Factor scores indicate the degree to which each subject's diet conforms to one of the identified patterns. We calculated factor scores in multi-study factor analysis by adapting the standard Bartlett and Thurstone methods for factor analysis[7,8]. In detail, we calculated a factor score for each subject (case or control) and factor within each study by using the study-specific correlation matrix of the log-transformed data (from the overall sample of cases and controls) and the factor-loading matrix [Φ | $\Lambda_s$] from the controls-only factor analysis.

The correlations between the two types of scores were 0.99 for all factors, so we continued with the Bartlett method, since its scores have zero sample mean vector and zero sample covariances[7].

To evaluate the internal consistency of patterns, we calculated standardized Cronbach's alpha coefficients for those nutrients that load more than 0.40 (in absolute value) on the

shared factors and more than 0.20 (in absolute value) on the study-specific factors. We calculated these coefficients for the shared factors using the merged-data correlation matrix and for the study-specific factors using the study-specific correlation matrices. We also calculated Cronbach's *alphas-when-item-deleted* for each factor and for each nutrient loading more than 0.40 and 0.20 (in absolute values) for the shared and study-specific patterns, respectively[6].

To examine the internal stability of patterns, we randomly assigned individuals to one of two equally sized sub-samples and performed multi-study factor analysis separately in both samples using the same approach of the main analysis. We repeated this procedure several times.

We also derived separate sets of dietary patterns by sex (females and males). Finally, we carried out multi-study factor analysis on the subset of five studies analyzed in Edefonti et al.[9], and included also in the current analysis.

***Estimates of the association between the identified dietary patterns and head and neck cancer***

For each shared factor, participants were grouped into five categories according to quintiles of factor scores among the controls. For each study-specific factor, participants were similarly 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), of cancers of the oral cavity and pharynx combined, and larynx, for each quantile category using unconditional multiple logistic regression models[10]. We fitted separate models for each factor, a shared-factor regression model and a composite regression

model including all the shared and one study-specific factor at a time. Study-specific factors were analyzed only for the studies in which they were identified. Models included adjustments for age, sex, race, study center (when appropriate), education, pack-years of cigarette smoking, cigar smoking status, pipe smoking status, and alcohol drinking intensity (number of drinks per day) (see **Table 1** for a complete list of the covariate categories used). Tests for linear trend were computed for all models scoring the quintiles as numbers from 1 to 5 and the tertiles as numbers from 1 to 3.

To accommodate heterogeneity of the shared patterns' associations across studies, we estimated the corresponding ORs and CIs using a random-slope generalized linear mixed model with logit link function and binomial family[11]. The random-effects terms included the quintile categories' effects (except for the reference) as random slopes and study center as common grouping factor (eight levels). No random intercepts or correlations between random effects were included. Overall, each shared pattern potentially contributed to the models with four random-effects terms (one for each quintile category, reference category excluded) and thirty-two random effects (one for each study center and quintile category, reference category excluded). However, given the limited number of levels of the grouping factor, we added random-effects terms for those patterns showing evidence in favor of them from both fixed and mixed models. All statistical tests were two-sided.

## Results

### *Identification of dietary patterns*

Correlations among individual nutrients were strong enough to suggest that the study-specific correlation matrices were factorable. Bartlett's test of sphericity rejected the null hypothesis that the correlation matrix is the identity matrix (p-value < 0.001) for each of the seven studies. The overall Kaiser-Meyer-Olkin statistic ranged from 0.88 [Italy Multicenter,

Memorial Sloan Kettering Cancer Center (MSKCC) and Milan (2006-2009) studies] to 0.93 (Boston study) across the seven studies, indicating that we had an adequate sample size relative to the number of nutrients within each study. The individual measures of sampling adequacy were very high, with no nutrients with measures smaller than 0.70 in any study. We obtained similar results for the correlation matrix of all the studies together: p-value for the Bartlett's test of sphericity < 0.001, overall Kaiser-Meyer-Olkin measure = 0.92, and individual measures of sampling adequacy ranging from 0.87 to 0.96.

Standardized Cronbach's alphas for the shared factors are 0.98, 0.96 and 0.95, for the *Animal products and cereals*, the *Anti-oxidant vitamins and fiber* and the *Fats* patterns, respectively; standardized Cronbach's alphas for the study-specific factors ranged from 0.91 (Boston study) to 0.96 (MSKCC study). Most of the standardized Cronbach's *alphas-when-item-deleted,* were lower than the corresponding standardized alphas for the same factor, although differences were generally small. These findings indicate that most nutrients are contributing to internal consistency and further support the choice of the selected nutrients.

The internal stability of the sets of patterns identified in the two split-samples for each study-specific dataset was good. Multi-study factor analysis identified the same number of shared and study-specific factors as the main analysis in the two sub-samples including half of the controls sampled at random from each study. In addition, the shared pattern factor-loading matrices were similar across the sub-samples and were in agreement with the main analysis. However, although the loading for calcium was still consistently reproduced in the sub-samples, other study-specific factor loadings showed some differences across the sub-samples and compared to the main analysis. This happened mostly with the MSKCC study, where the number of controls was only 123. Communalities were in agreement with the main analysis.

In addition, when we carried out stratified multi-study factor analyses by sex, the best models according to the Akaike Information Criterion differed in terms of number of shared patterns. For males, the best model showed three common factors and one additional factor for each of the US studies, as in the main analysis; for females, it showed four shared patterns and no extra study-specific patterns were assumed. For the former group (4,653 control males), the identified patterns and corresponding percentages of explained variances (36, 24, and 16% for the shared patterns, and 3, 3, 5, and 3%, for the study-specific patterns, with patterns ordered as in Table 2 and eTable3, respectively) were similar to those derived on the combined set of males and females. The few differences in the factor loadings concerned dominant nutrients of the study-specific patterns that did not show up anymore in the male stratum alone. For the latter group (2,166 control females), the *Anti-oxidant vitamins and fiber* and the *Fats* patterns were very similar to the corresponding ones from the main analysis, with percentages of explained variances of 26 and 18%, respectively, However, the original shared *Animal products and cereals* pattern derived on males and females combined was split into two other shared patterns (20 and 14% of explained variance) in females only. The first pattern (20% of variance) included niacin, protein, and zinc as dominant nutrients; the second one (14% of variance) showed the following dominant nutrients: calcium, phosphorus, and riboflavin.

Finally, when we carried out multi-study factor analysis on the subset of five studies analyzed in Edefonti et al.[9], the three previously derived patterns were satisfactorily reproduced in the form of our shared patterns. The boxplots representing the distribution of the shared pattern loadings on the 100 bootstrapped random sets were narrower with multi-study than with standard maximum likelihood factor analysis (e**Figure 2**). Similarly, the standard errors of the shared pattern loadings were smaller under the latter approach. In addition, multi-study factor analysis estimated one extra pattern for each of the US

studies: the American study-specific patterns were similar to the corresponding ones from the more recent analysis on seven studies. Percentages of explained variances were similar for the corresponding patterns in both the analyses (**eFigure 3**).

We observed generally similar results with the cases+controls, compared to controls-only, multi-study factor analysis (**eFigure 4**). The most important differences are related to likely loss of power when subsamples of the smaller control sample are used to derive factors (i.e. split-half analysis and stratified multi-study factor analysis by sex). In the analysis by sex, when the combined sample of cases and controls was used to derive factors, the two best multi-study factor analysis models for both males and females are characterized by a consistent set of three shared patterns and one study-specific pattern for each of the US studies. This is also the case for the main cases+controls multi-study factor analysis on the combined set of males and females.

***Estimates of the association between the identified dietary patterns and head and neck cancer***

Results from the seven models including each dietary pattern separately, together with the selected confounding factors, were in agreement with those from the main analysis for both cancer sites.

Extra adjustment by non-alcohol energy intake in mixed-effects models was carried out with an approximated solution, based on penalized iteratively reweighted least squares, due to convergence issues. After the adjustment, the point estimates of the association between dietary pattern scores and head and neck cancer remained similar to the ones presented in the main analysis. For the *Anti-oxidant vitamins and fiber* patterns, the ORs for the last quintile category were 0.58 (95% CI: 0.42-0.80) for oral and pharyngeal cancer and 0.57 (95% CI: 0.34-0.96) for laryngeal cancer; for the *Fats* pattern, the OR of laryngeal

cancer was 1.59 (95% CI: 1.19-2.12). However, some CIs were wider, with the corresponding ORs being 1.23 (95% CI: 0.82-1.85) for the *Animal products and cereals* pattern and laryngeal cancer and 0.81 (95% CI: 0.66-1.00) for the *Fats* pattern and oral cavity and pharyngeal cancer. This also happened with the *Dairy products and breakfast cereals* patterns identified in the Los Angeles (OR=0.95, 95% CI: 0.57-1.60) and Boston studies (OR=1.25, 95% CI: 0.87-1.79) for oral cavity and pharyngeal cancer, in the models including the three shared dietary patterns too. The point estimates of the extra adjustment by non-alcohol energy intake were close to one across quintile categories for both cancer sites, with the partial exception of the two models for the Boston study, where there was evidence of a trend for increasing quintile categories of non-alcohol energy intake, although the ORs have very wide CIs.

Information on never/ever use of supplements of vitamin C, vitamin E, or beta-carotene was available and usable for the Los Angeles, Boston, and North Carolina (2002-2006) studies. Extra adjustment for supplement use was therefore not possible for the shared patterns regression models, but only for the models including the study-specific dietary patterns, together with the three shared ones. Adjustment by any supplement use was relevant in the North Carolina (2002-2006) study only, for both cancer sites. However, results were similar to the ones from the main analysis: for the last tertile category, the ORs of both cancer sites were around 0.90 and the corresponding CIs include 1. For the Los Angeles and Boston studies, the inclusion of information on one supplement at a time was not relevant and it did not modify the protective/detrimental role of the study-specific pattern on oral and pharyngeal cancers combined, although the CIs were very wide. However, results were based on 38% and 45-61% (depending on the supplement considered) of the original samples, respectively.

We observed similar results when using logistic regression models including quantile categories of factor scores derived from the cases+controls, instead of controls-only, multi-study factor analysis (**eFigure 4**, **eTable 6**, and **eTable 7**). However, given the same OR=0.62 (last quintile score category) for the Anti-oxidant vitamins and fiber pattern and laryngeal cancer, the CI from the cases+controls multi-study factor analysis was narrower (0.40-0.97 eTable 6) compared to the controls-only analysis (0.37-1.04 Table 3 in the main text).

## References


1. de Vito R, Bellio R, Trippa L, Parmigiani G. Multi-study factor analysis. *arXiv preprint arXiv:1611.06350* 2016.
2. Meng X-L, Rubin DB. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 1993;**80**(2):267-278.
3. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)* 1977:1-38.
4. Mulaik SA. *Foundations of factor analysis.* CRC Press, 2009.
5. Akaike H. A new look at the statistical model identification. *IEEE transactions on automatic control* 1974;**19**(6):716-723.
6. Pett MA, Lackey NR, Sullivan JJ. *Making sense of factor analysis: the use of factor analysis for instrument development in health care research.* Thousand Oaks, CA: Sage, 2003.
7. Johnson RA, Wichern DW. *Applied multivariate statistical analysis*. 5th ed. Upper Saddler River, NJ: Prentice Hall, 2002.
8. DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation* 2009;**14**(20):1-11.
9. Edefonti V, Hashibe M, Ambrogi F, Parpinel M, Bravi F, Talamini R, Levi F, Yu G, Morgenstern H, Kelsey K, McClean M, Schantz S, Zhang Z, Chuang S, Boffetta P, La Vecchia C, Decarli A. Nutrient-based dietary patterns and the risk of head and neck cancer: a pooled analysis in the International Head and Neck Cancer Epidemiology consortium. *Ann Oncol* 2012;**23**(7):1869-80.
10. Hosmer DW, Lemeshow S. *Applied logistic regression, 2nd edn*. New York, NY:

John Wiley & Sons, Inc, 2000.

11. Pinheiro JC, Bates DM. *Mixed-effects models in S and S-PLUS*. New York, NY: Springer-Verlag, 2000.

# eTables

**eTable 1. Characteristics of the individual studies from the International Head and Neck Cancer Epidemiology (INHANCE) Consortium used in the current analysis.**

| Study Reference paper | Recruitment period | Source (cases/ controls) | Participation rate, % (cases/controls) | Age eligibility (years) | Number of subjects (cases/ controls) | Questionnaire, administration, reference period for the recall | Frequency | Serving size[a] | # Food items (including non-alcoholic beverages) |
|---|---|---|---|---|---|---|---|---|---|
| **Italy Multicenter** Bosetti et al., 2003[b] | 1990-1999 | Hospital/Hospital-unhealthy | >95/>95 | 18-80 | 1261/2716 | FFQ, interviewer-administered, 2 year before disease | Raw data | Small/ Medium/ Large | 78 (including 6 non-alcoholic beverages) |
| **Switzerland** Levi et al., 1998[b] | 1991-1997 | Hospital/Hospital-unhealthy | >95/>95 | <80 | 516/883 | FFQ, interviewer-administered, 2 year before disease | Raw data | Small/ Medium/ Large | 78 (including 6 non-alcoholic beverages) |
| **Los Angeles**, CA, USA Cui et al., 2006 | 1999-2004 | Cancer registry/ Neighborhood | 49/68 | 18-65 | 417/1005 | FFQ, interviewer-administered, during the past year | Raw data | Medium | 78 (including 11 non-alcoholic beverages) |
| **Boston**, MA, USA Peters et al., 2005 | 1999-2004 | Hospital/ Residential records | 88.7/48.7 | ≥18 | 584/659 | FFQ, self-administered, during the past year | Categories | Medium | 138 (including 12 non-alcoholic bevarges) |
| **New York, MSKCC**, USA Schantz et al., 1997 | 1992-1994 | Hospital/Blood donors | NA | NA | 134/169 | FFQ–diet history, self-administered, during the past year[c] | Raw data | Small/ Medium/ Large | 88 (including 5 non-alcoholic beverages) |

| Study Reference paper | Recruitment period | Source (cases/ controls) | Participation rate, % (cases/controls) | Age eligibility (years) | Number of subjects (cases/ controls) | Questionnaire, administration, reference period for the recall | Frequency | Serving size[a] | # Food items (including non-alcoholic beverages) |
|---|---|---|---|---|---|---|---|---|---|
| **Milan (2006-2009)**, Italy Bravi et al., 2013[b] | 2006-2009 | Hospital/Hospital-unhealthy | >95/>95 | 18-80 | 367/750 | FFQ, interviewer-administered, 2 years before disease | Raw data | Small/ Medium/ Large | 78 (including 6 non-alcoholic beverages) |
| **North Carolina (2002-2006)**, USA Divaris et al., 2010[c] | 2002-2006 | Cancer registry/ Department of Motor Vehicles files | 82/61 | 20-80 | 1368/1396 | FFQ, interviewer-administered, during the past year | Categories | Medium | 72 (including 5 non-alcoholic beverages) |

ABBREVIATIONS: FFQ: food-frequency questionnaire; MSKCC: Memorial Sloan Kettering Cancer Center; NA: not available. [a]A quantification of the medium serving size was provided in all the studies, except for the Japan one. [b]Italy Multicenter, Milan (2006-2009) and Switzerland studies were based on the same food-frequency questionnaire. [c]The food-frequency questionnaire from the North Carolina study provided combined questions concerning consumption of specific food items and corresponding condiment habits or fat content of the food item of interest (i.e. while asking for cooked or raw vegetable consumption, the food frequency questionnaire asked for extra information on fat, sauce, or dressing added after cooking or at the table).

**eTable 2. Akaike Information Criterion - based results from the model selection procedure of the best controls-only multi-study factor analysis model[a]. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

| Number of shared factors | Number of study-specific factors | | Akaike Information Criterion value |
|---|---|---|---|
| | European studies[b] | American studies[b] | |
| 1 | 2 | 3 | 41934.78 |
| 2 | 1 | 2 | 43678.39 |
| **3** | **0** | **1** | **36762.55** |

[a] We started the model selection procedure assuming a fixed value for the study-specific total latent factors, $T_s$, as indicated by a combination of criteria used in standard factor analysis (see details in the Statistical Analysis section in the main manuscript and eAppendix). In detail, we assumed $T_s$=3 for the European studies and $T_s$=4 for the American studies. For each $T_s$, we then explored models showing the available combinations of number of shared and number of study-specific dietary patterns identified in the table rows. We finally selected the model with the minimum Akaike Information Criterion value (**36762.55**). The selected model was indicated in bold typeface. [b] The European studies included Italy Multicenter, Switzerland, and Milan (2006-2009); the American studies included Boston, Los Angeles, Memorial Sloan Kettering Cancer Center, and North Carolina (2002-2006).

**eTable 3. Study-specific factor-loading matrix and explained variances (VAR)[a] for the four study-specific dietary patterns identified by controls-only multi-study factor analysis. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

| Nutrient | *Los Angeles*[b] | *Boston*[b] | *MSKCC*[b] | *North Carolina (2002-2006)*[b] |
|---|---|---|---|---|
| Total protein[c] (g)[d] | - | - | 0.22 | - |
| Cholesterol (mg)[d] | - | - | 0.16 | - |
| Saturated fatty acids (g) | **0.25** | 0.23 | 0.21 | 0.19 |
| Monounsaturated fatty acids (g) | 0.15 | 0.12 | 0.19 | - |
| Polyunsaturated fatty acids (g) | - | - | 0.20 | - |
| Total carbohydrates (g) | **-** | - | - | 0.22 |
| Calcium (mg)[b] | **0.65** | **0.53** | **0.45** | **0.63** |
| Sodium (mg) | 0.17 | - | 0.16 | - |
| Potassium (mg) | - | 0.16 | 0.24 | 0.20 |
| Phosphorus (mg) | **0.28** | 0.24 | **0.34** | **0.25** |
| Iron (mg) | -0.22 | -0.21 | **-0.26** | -0.10 |
| Zinc (mg) | - | - | **-0.29** | - |
| Thiamin (vitamin B1, mg) | -0.14 | - | **-0.32** | - |
| Riboflavin (vitamin B2, mg) | 0.20 | 0.14 | -0.12 | 0.23 |
| Vitamin B6 (mg) | **-0.29** | -0.19 | **-0.30** | - |
| Vitamin C (mg) | -0.10 | - | - | - |
| Total folate (μg) | -0.22 | - | **-0.27** | - |
| Niacin (vitamin B3, mg) | **-0.34** | **-0.30** | **-0.32** | -0.21 |
| Lutein (μg) | - | - | - | - |
| Total carotene (μg) | - | - | - | - |
| Lycopene (μg) | - | - | 0.19 | -0.10 |
| Vitamin E (mg) | -0.10 | - | **-0.33** | - |
| Total fiber (g) | -0.14 | - | - | - |
| Proportion of VAR explained (%) | 5 | 3 | 6 | 3 |

ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

[a]Estimated from a multi-study factor analysis carried out on twenty-three nutrients. The magnitude of each loading measures the importance of the corresponding nutrient to the factor. [b]The table showed additional study-specific dietary patterns for the US studies only, as the best-fitting multi-study factor analysis model did not include additional study-specific patterns for the European studies. [c]Loadings ≥0.25 in absolute value define the dominant nutrients for each factor and were shown in bold typeface; loadings <0.1 in absolute value were suppressed. [d]The units of the nutrients indicated their original scale, but loadings were derived from log-transformed and standardized nutrient intakes entered into the multi-study factor analysis model.

**eTable 4. Nutrient communalities for the shared and study-specific dietary patterns identified within the controls-only multi-study factor analysis[a]. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

| | Shared patterns | Study-specific patterns | | | |
|---|---|---|---|---|---|
| **Nutrient** | **Italy Multicenter Milan (2006-2009) Switzerland** | **Los Angeles** | **Boston** | **MSKCC** | **North Carolina (2002-2006)** |
| Total protein[b] (g) | 0.95 | 0.95 | 0.95 | 0.99 | 0.95 |
| Cholesterol[b] (mg) | 0.74 | 0.74 | 0.74 | 0.75 | 0.74 |
| Saturated fatty acids (g) | 0.80 | 0.85 | 0.84 | 0.83 | 0.83 |
| Monounsaturated fatty acids (g) | 0.89 | 0.90 | 0.89 | 0.89 | 0.90 |
| Polyunsaturated fatty acids (g) | 0.76 | 0.76 | 0.75 | 0.77 | 0.76 |
| Total carbohydrates (g) | 0.64 | 0.65 | 0.66 | 0.65 | 0.70 |
| Calcium (mg) | 0.63 | 1.00 | 0.91 | 0.89 | 1.00 |
| Sodium (mg) | 0.77 | 0.79 | 0.77 | 0.79 | 0.77 |
| Potassium (mg) | 0.87 | 0.88 | 0.91 | 0.93 | 0.91 |
| Phosphorus (mg) | 0.91 | 0.96 | 0.96 | 1.00 | 0.96 |
| Iron (mg) | 0.76 | 0.80 | 0.79 | 0.81 | 0.77 |
| Zinc (mg) | 0.93 | 0.94 | 0.93 | 0.99 | 0.94 |
| Thiamin (vitamin B1, mg) | 0.86 | 0.87 | 0.86 | 0.93 | 0.86 |
| Riboflavin (vitamin B2, mg) | 0.84 | 0.88 | 0.87 | 0.85 | 0.89 |
| Vitamin B6 (mg) | 0.88 | 0.96 | 0.91 | 0.94 | 0.89 |
| Vitamin C (mg) | 0.64 | 0.63 | 0.64 | 0.63 | 0.66 |
| Total folate (μg) | 0.81 | 0.84 | 0.82 | 0.86 | 0.81 |
| Niacin (vitamin B3, mg) | 0.80 | 0.91 | 0.89 | 0.92 | 0.85 |
| Lutein (μg) | 0.40 | 0.40 | 0.41 | 0.42 | 0.41 |
| Total carotene (μg) | 0.49 | 0.49 | 0.50 | 0.50 | 0.50 |
| Lycopene (μg) | 0.21 | 0.21 | 0.20 | 0.22 | 0.22 |
| Vitamin E (mg) | 0.88 | 0.88 | 0.87 | 0.94 | 0.88 |
| Total fiber (g) | 0.77 | 0.77 | 0.77 | 0.77 | 0.77 |

ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

[a]For the European studies, nutrient communalities were based on the three shared dietary patterns only; for the American studies, nutrient communalities were based on all the four (shared and study-specific) derived dietary patterns.

**eTable 5. Odds Ratios (ORs)[a] of oral cavity and pharyngeal cancers combined, and laryngeal cancer and corresponding Confidence Intervals (95% CIs) on study-specific factor scores tertile categories, as derived from the controls-only multi-study factor analysis. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

| Study-specific dietary pattern | Controls | Oral and pharyngeal cases | OR | 95% CI | Laryngeal cases | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| ***Los Angeles-specific*** | | | | | | | |
| I Tertile (-4.66, -0.40] | 268 | 87 | 1[b] | - | 18 | 1[b] | - |
| II Tertile (-0.40, 0.41] | 267 | 84 | 1.0 | 0.68 - 1.48 | 17 | 1.10 | 0.48 - 2.54 |
| III Tertile (0.41, 4.31] | 266 | 68 | 0.66 | 0.44 - 0.99 | 23 | 1.37 | 0.61 - 3.07 |
| $p_{for\ trend}$[c] | | | | 0.048 | | | 0.531 |
| ***Boston-specific*** | | | | | | | |
| I Tertile (-4.68, -0.50] | 203 | 88 | 1[b] | - | 26 | 1[b] | - |
| II Tertile (-0.50, 0.30] | 202 | 94 | 1.18 | 0.80 - 1.75 | 24 | 1.10 | 0.54 - 2.23 |
| III Tertile (0.30, 6.18] | 200 | 127 | 1.55 | 1.06 - 2.28 | 19 | 0.71 | 0.34 - 1.49 |
| $p_{for\ trend}$[c] | | | | 0.034 | | | 0.541 |
| ***MSKCC-specific*** | | | | | | | |
| I Tertile (-5.13, -0.13] | 38 | 21 | 1[b] | - | 13 | 1[b] | - |
| II Tertile (-0.13, 0.36] | 38 | 23 | 1.14 | 0.44 - 2.94 | 4 | 0.16 | 0.03 - 0.88 |
| III Tertile (0.36, 1.76] | 33 | 20 | 0.98 | 0.38 - 2.56 | 12 | 0.76 | 0.22 - 2.63 |
| $p_{for\ trend}$[c] | | | | 0.918 | | | 0.853 |
| ***North Carolina (2002-2006)-specific*** | | | | | | | |
| I Tertile (-3.47, -0.58] | 369 | 218 | 1[b] | - | 96 | 1[b] | - |
| II Tertile (-0.58, 0.31] | 369 | 215 | 0.76 | 0.57 – 1.00 | 147 | 0.96 | 0.66 - 1.39 |
| III Tertile (0.31, 6.18] | 372 | 247 | 0.83 | 0.63 - 1.10 | 128 | 0.83 | 0.57 - 1.22 |
| $p_{for\ trend}$[c] | | | | 0.184 | | | 0.333 |

ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

[a]Estimated from multiple logistic regression models adjusted for age, sex, race, education, pack-years of cigarette smoking, cigar smoking status, pipe smoking status, and alcohol drinking intensity (number of drinks per day). Due to the low number of cases and controls for the MSKCC study, covariates were restricted to age, sex, race, education, pack-years of cigarette smoking, and alcohol drinking intensity and the number of categories used was smaller than in the other study-specific analyses. Results refer to the fixed-effects composite models including one study-specific dietary pattern at a time and the three shared dietary patterns. [b]Reference category. [c]P for linear trend.

**eTable 6. Odds Ratios (ORs)[a] of oral cavity and pharyngeal cancers combined, and laryngeal cancer and corresponding Confidence Intervals (95% CIs) on shared factor scores quintile categories, as derived from the cases+controls multi-study factor analysis. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

| Shared dietary pattern | Controls | Oral and pharyngeal cases | $p_{studies}$ [b] | OR | 95% CI[c] | Laryngeal cases | $p_{studies}$ [b] | OR | 95% CI[c] |
|---|---|---|---|---|---|---|---|---|---|
| ***Animal products and cereals*** | | | | | | | | | |
| I Quintile (-5.57, -0.90] | 1,337 | 378 | | 1[d] | - | 135 | | 1[d] | - |
| II Quintile (-0.90, -0.33] | 1,334 | 376 | | 0.83 | 0.69 - 1.00 | 192 | | 1.10 | 0.84 - 1.43 |
| III Quintile (-0.33, 0.15] | 1,345 | 476 | 0.06 | 0.97 | 0.81 - 1.15 | 219 | <0.001 | 1.17 | 0.90 - 1.51 |
| IV Quintile (0.15, 0.74] | 1,336 | 582 | | 1.15 | 0.96 - 1.36 | 317 | | 1.40 | 0.99 - 1.98 |
| V Quintile (0.74, 4.90] | 1,342 | 676 | | 1.06 | 0.89 - 1.26 | 394 | | 1.52 | 1.12 - 2.07 |
| $p_{for\ trend}$[e] | | | | | 0.03 | | | | <0.001 |
| ***Anti-oxidant vitamins and fiber*** | | | | | | | | | |
| I Quintile (-6.50, -0.74] | 1,336 | 669 | | 1[d] | - | 355 | | 1[d] | - |
| II Quintile (-0.74, -0.15] | 1,336 | 498 | | 0.80 | 0.68 - 0.94 | 266 | | 0.74 | 0.60 - 0.91 |
| III Quintile (-0.15, 0.37] | 1,347 | 532 | <0.001 | 0.86 | 0.73 - 1.01 | 255 | <0.001 | 0.68 | 0.51 - 0.90 |
| IV Quintile (0.37, 0.92] | 1,343 | 435 | | 0.69 | 0.56 - 0.85 | 200 | | 0.55 | 0.43 - 0.71 |
| V Quintile (0.92, 4.65] | 1,332 | 354 | | 0.55 | 0.42 - 0.72 | 181 | | 0.62 | 0.40 - 0.97 |
| $p_{for\ trend}$[e] | | | | | <0.001 | | | | 0.03 |
| ***Fats*** | | | | | | | | | |
| I Quintile (-7.90, -0.83] | 1,336 | 481 | | 1[d] | - | 155 | | 1[d] | - |
| II Quintile (-0.83, -0.23] | 1,343 | 484 | | 0.96 | 0.81 - 1.13 | 205 | | 1.30 | 1.00 - 1.68 |
| III Quintile (-0.23, 0.24] | 1,340 | 488 | 0.310 | 1.00 | 0.84 - 1.19 | 239 | 0.088 | 1.65 | 1.28 - 2.12 |
| IV Quintile (0.24, 0.76] | 1,334 | 506 | | 0.89 | 0.75 - 1.05 | 286 | | 1.63 | 1.27 - 2.09 |
| V Quintile (0.76, 3.69] | 1,341 | 529 | | 0.82 | 0.69 - 0.97 | 372 | | 2.03 | 1.60 - 2.58 |
| $p_{for\ trend}$[e] | | | | | 0.02 | | | | <0.001 |

[a]Estimated from multiple logistic regression models adjusted for age, sex, race, study center, education, pack-years of cigarette smoking, cigar smoking status, pipe smoking status, alcohol drinking intensity (number of drinks per day). Results refer to the composite models including all the three shared factors simultaneously. [b]P for heterogeneity between studies. [c]For both cancer sites, we reported results from a generalized linear mixed model including a random-slope for the *Anti-oxidant vitamins and fiber* pattern with oral and pharyngeal cancers combined and a random slope for the *Animal products and cereals* and the *Anti-oxidant vitamins and fiber* patterns with laryngeal cancer. [d]Reference category. [e]P for linear trend.

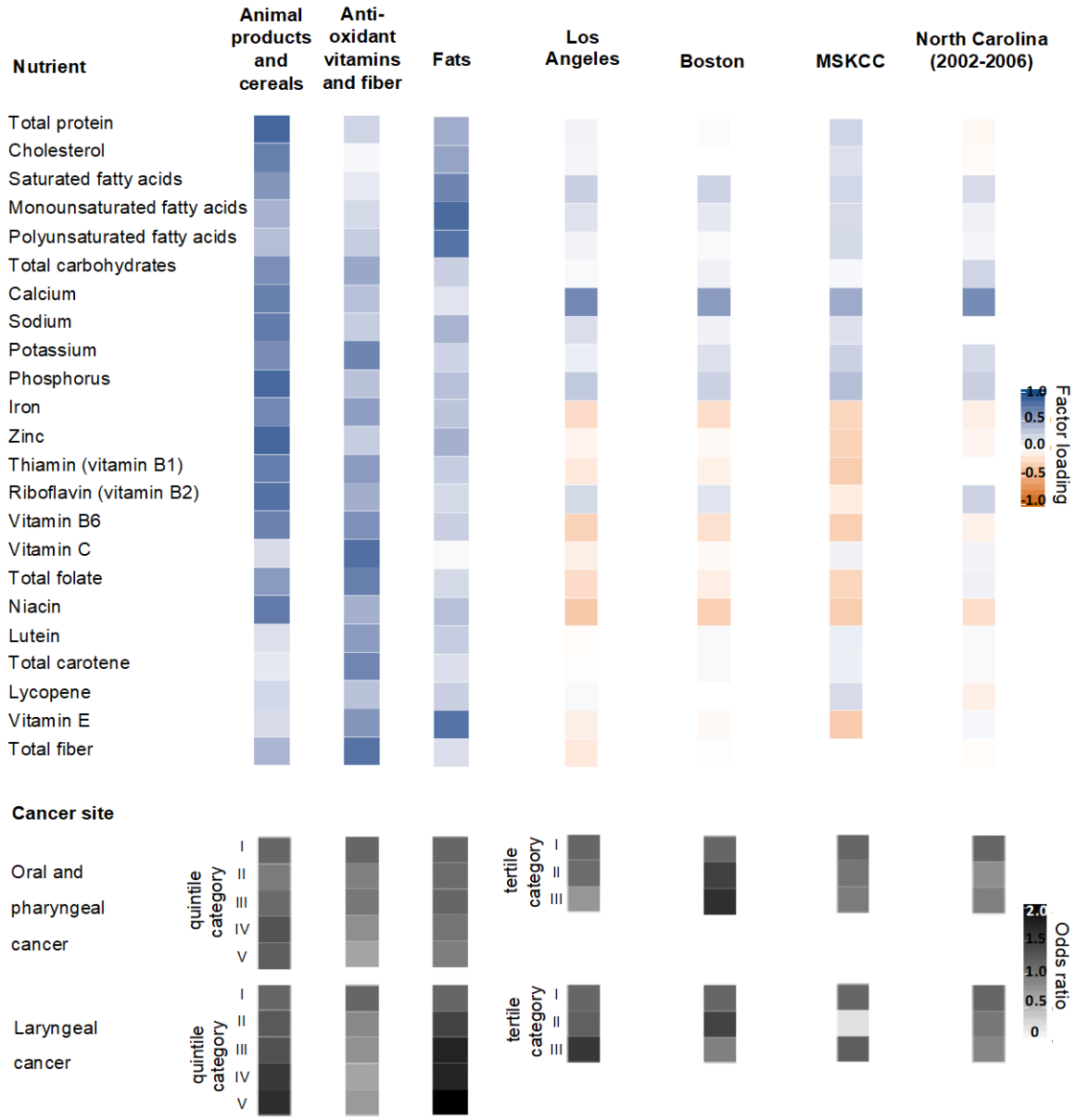
e**Table 7. Odds Ratios (ORs)[a] of oral cavity and pharyngeal cancers combined, and laryngeal cancer and corresponding Confidence Intervals (95% CIs) on study-specific factor scores tertile categories, as derived from the cases+controls multi-study factor analysis. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

| Study-specific dietary pattern | Controls | Oral and pharyngeal cases | OR | 95% CI | Laryngeal cases | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| ***Los Angeles-specific*** | | | | | | | |
| I Tertile (-4.94, -0.40] | 266 | 92 | 1[b] | - | 18 | 1[b] | - |
| II Tertile (-0.40, 0.44] | 270 | 81 | 0.95 | 0.65 - 1.40 | 18 | 1.05 | 0.45 - 2.48 |
| III Tertile (0.44, 4.27] | 265 | 66 | 0.64 | 0.43 - 0.96 | 22 | 1.43 | 0.62 - 3.29 |
| $p_{for\ trend}$[c] | | | | 0.03 | | | 0.46 |
| ***Boston-specific*** | | | | | | | |
| I Tertile (-4.84, -0.53] | 203 | 88 | 1[b] | - | 25 | 1[b] | - |
| II Tertile (-0.53, 0.35] | 202 | 102 | 1.32 | 0.89 - 1.95 | 25 | 1.30 | 0.65 - 2.63 |
| III Tertile (0.35, 6.39] | 200 | 119 | 1.50 | 1.02 - 2.21 | 19 | 0.80 | 0.38 - 1.68 |
| $p_{for\ trend}$[c] | | | | 0.06 | | | 0.72 |
| ***MSKCC-specific*** | | | | | | | |
| III Tertile (-2.16, -0.36] | 38 | 23 | 1[b] | - | 12 | 1[b] | - |
| II Tertile (-0.36, 0.18] | 38 | 23 | 0.88 | 0.35 - 1.22 | 6 | 0.17 | 0.03 - 0.85 |
| I Tertile (0.18, 5.12] | 33 | 18 | 0.83 | 0.32 - 2.17 | 11 | 1.06 | 0.31 - 3.62 |
| $p_{for\ trend}$[c] | | | | 0.69 | | | 0.91 |
| ***North Carolina (2002-2006)-specific*** | | | | | | | |
| I Tertile (-3.42, -0.54] | 370 | 231 | 1[b] | - | 107 | 1[b] | - |
| II Tertile (-0.54, 0.32] | 368 | 205 | 0.70 | 0.53 - 0.93 | 140 | 0.86 | 0.60 - 1.25 |
| III Tertile (0.32, 5.75] | 372 | 244 | 0.81 | 0.61 - 1.07 | 124 | 0.76 | 0.52 - 1.11 |
| $p_{for\ trend}$[c] | | | | 0.09 | | | 0.11 |

ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

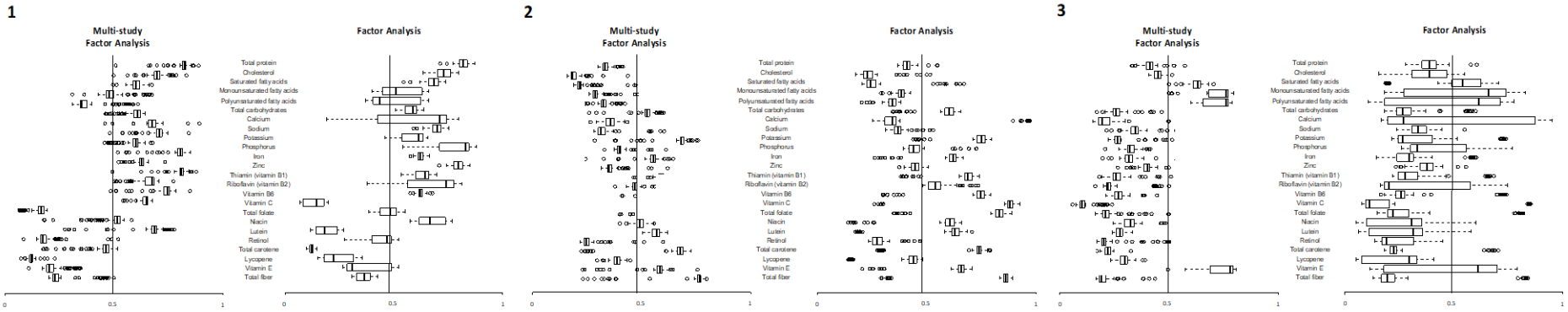
[a]Estimated from multiple logistic regression models adjusted for age, sex, race, education, pack-years of cigarette smoking, cigar smoking status, pipe smoking status, and alcohol drinking intensity (number of drinks per day). Due to the low number of cases and controls for the MSKCC study, covariates were restricted to age, sex, race, education, pack-years of cigarette smoking, and alcohol drinking intensity and the number of categories used was smaller than in the other study-specific analyses. Results refer to the fixed-effects composite models including one study-specific dietary pattern at a time and the three shared dietary patterns. [b]Reference category. [c]P for linear trend.

**eFigure 1. Heatmap of the estimated factor loadings and cancer-specific odds ratios for the shared and study-specific dietary patterns identified with the controls-only multi-study factor analysis. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

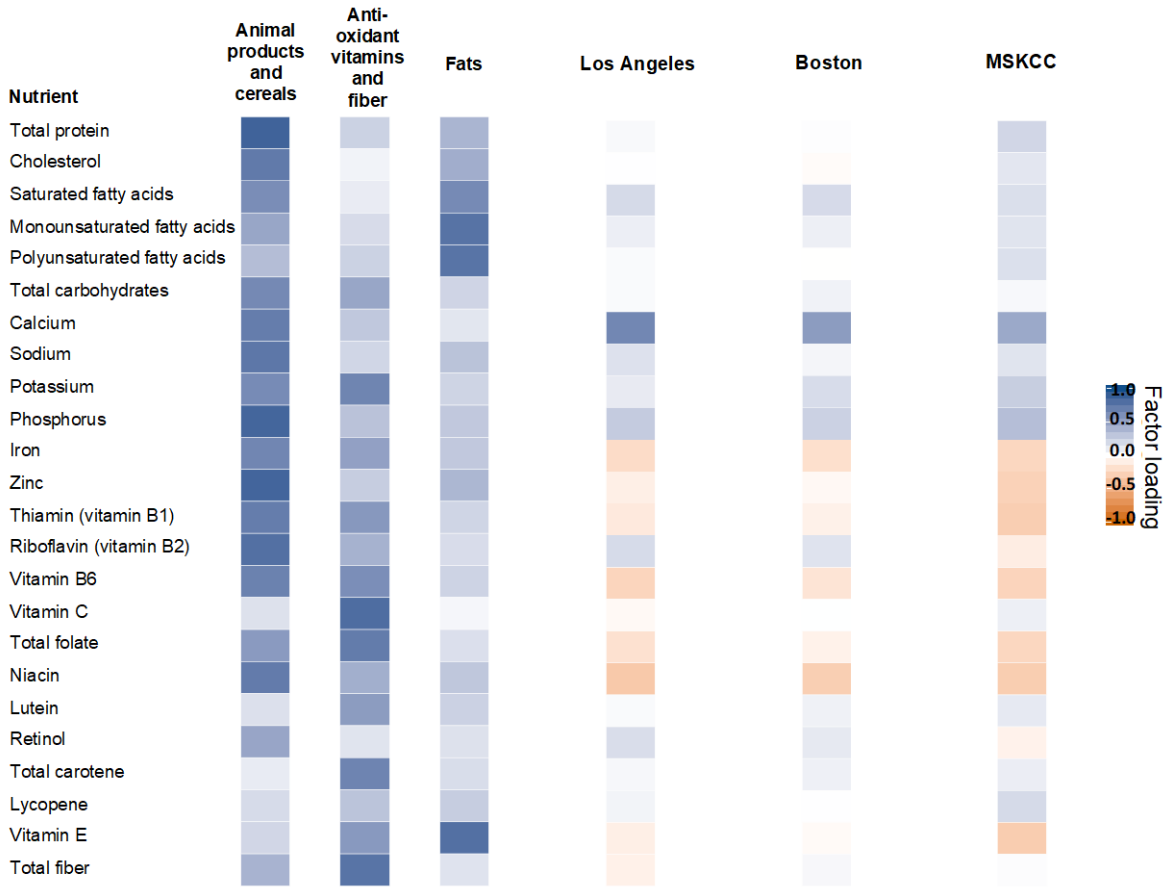



ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

**eFigure 2. Distribution of the estimated factor loadings for the three shared dietary patterns, *Animal products and cereals* (1), *Anti-oxidant vitamins and fiber* (2), and *Fats* (3), as identified by multi-study factor analysis and by standard factor analysis, from 100 bootstraps of the control sample including the subset of five studies analyzed in Edefonti et al.[9]. The distribution of the loadings of each nutrient to each factor is represented through a boxplot. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**

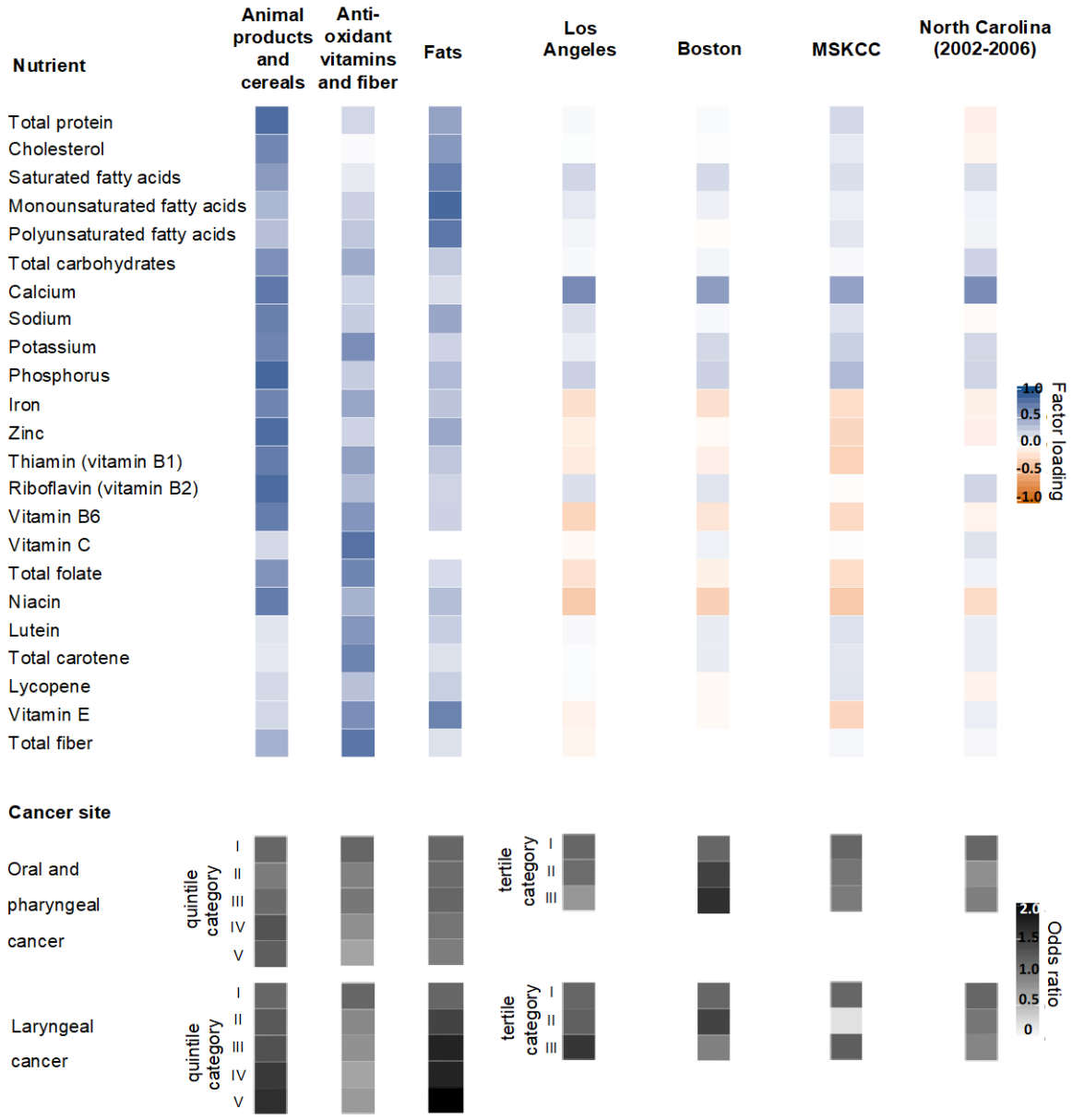


**eFigure 3. Heatmap of the estimated factor loadings for the shared and study-specific dietary patterns as identified with the controls-only multi-study factor analysis on the subset of five studies included in Edefonti et al.[9]. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**




ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

**eFigure 4. Heatmap of the estimated factor loadings and cancer-specific odds ratios for the shared and study-specific dietary patterns identified with the cases+controls multi-study factor analysis. International Head and Neck Cancer Epidemiology (INHANCE) Consortium.**




ABBREVIATIONS: MSKCC: Memorial Sloan Kettering Cancer Center.

## Data collection instruments:

The study-specific food-frequency questionnaires on which data are based are available on the INHANCE website: http://www.inhance.utah.edu/index.php under the login member area. Therefore, they are available upon request to the Corresponding Author at valeria.edefonti@unimi.it.