DIETARY PATTERNS AND ORAL AND PHARYNGEAL CANCER USING LATENT CLASS ANALYSIS

Michela Dalmartello[1], Adriano Decarli[1], Monica Ferraroni[1], Francesca Bravi[1], Diego Serraino[2], Werner Garavello[3], Eva Negri[4], Jeroen Vermunt[5], Carlo La Vecchia[1]

[1]Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy

[2]Unit of Cancer Epidemiology, Centro di Riferimento Oncologico Aviano IRCCS, Aviano, Italy

[3]Department of Otorhinolaryngology, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

[4]Department of Biomedical and Clinical Sciences, Università degli Studi di Milano, Sacco Hospital, Milan, Italy

[5]Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

Short title:
Dietary patterns with latent class analysis and oral and pharyngeal cancer

Correspondence to:
Dr Michela Dalmartello,
Department of Clinical Sciences and Community Health, University of Milan
Via A. Vanzetti, 5, 20133 Milan, Italy
Phone: +39 0250320880
fax: +39 0250320866
email: michela.dalmartello@unimi.it

Keywords:
Case-control study; Dietary patterns; Oral cancer; Pharyngeal cancer; Latent class analysis

Abbreviations:
BIC: Bayesian Information criterion; BVR: bivariate residuals; CA: cluster analysis; CI: confidence interval; FA: factor analysis; LCA: latent class analysis; ML: maximum likelihood; PCA: principal components analysis; PCFA: principal component factor analysis; OPC: oral/pharyngeal cancer; OR: odds ratio.

Article category:
Cancer epidemiology

Novelty and Impact:
Latent class analysis gives relevant insights to dietary pattern research, allowing for the identification of mutually exclusive subgroups in the population characterized by different dietary choices with a solid parametric approach. Subjects with a diet rich in fruits and vegetable products and poor in red meat, or with high diet diversity had a decrease in oral/pharyngeal cancer risk, independently from the major recognised risk factors, compared to the other groups.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.32769

Abstract

The methods traditionally used to identify *a posteriori* dietary patterns are principal components, factor and cluster analysis. The aim of this study is to assess the relation between dietary patterns derived with latent class analysis and oral/pharyngeal cancer risk (OPC), highlighting the strengths of this method compared to traditional ones.

We analysed data from an Italian multicentric case-control study on OPC including 946 cases and 2492 hospital controls. Dietary patterns were derived using latent class analysis on 25 food groups. A multiple logistic regression model was used to derive odds ratios (ORs) and corresponding 95% confidence intervals (CIs) for OPC according to the dietary patterns identified.

We identified four dietary patterns. The first one was characterized by high intake of leafy and fruiting vegetable and fruits ('Prudent pattern'), the second one showed a high intake of red meat and low intake of selected fruits and vegetables ('Western pattern'). The last two patterns showed a combination-type of diet. We labelled 'Lower consumers-combination pattern' the cluster that showed a low intake of the majority of foods, and 'Higher consumers-combination pattern' the one characterized by a high intake of various foods. Compared to the 'Prudent pattern', the 'Western' and the 'Lower consumers-combination' ones were positively related to the risk of OPC (OR=2.56, 95% CI: 1.90 - 3.45 and OR=2.23, 95% CI: 1.64 - 3.02). No difference in risk emerged for the 'Higher consumers-combination pattern' (OR=1.28, 95% CI: 0.92 - 1.77).

Introduction

Oral cavity and pharyngeal cancer (OPC hereafter) collectively ranks seventh for incidence and eighth for cancer mortality [1]. Tobacco smoking and excessive alcohol drinking are recognized as the major risk factors for OPC. Among other factors, diet has been suggested to play an important role. In particular, an inverse association between intake of vegetable and fruits and a possible positive association between meat and OPC risk were reported [2-18]. Most of the evidence came from studies focusing on single foods while the relationship between diet and oral and pharyngeal cancer has been less frequently addressed considering dietary patterns.

The dietary pattern approach is useful to study the effect of the overall diet on health outcome, considering the network of complex interactions between foods or nutrients. The methods traditionally used to identify dietary patterns are principal components analysis (PCA), factor analysis (FA), principal components factor analysis (PCFA) and cluster analysis (CA). With regard to *a posteriori* dietary patterns, association between diet and OPC has been traditionally assessed by PCA and PCFA [6 7 19-25].

Latent Class Analysis (LCA) is a latent variable model, which has advantages in comparison to the previous methods [26]. Unlike PCA/PCFA/FA, it can be used to classify individuals into mutually exclusive groups/dietary patterns and differently from CA, it permits quantification of the uncertainty of class membership, and assessment of goodness of fit. Moreover, it allows for adjustment for covariates directly in the pattern identification.

The aim of this study is to identify dietary patterns through LCA, hence adding a new perspective on the association between dietary habits and OPC in Italy.

Materials and methods

Study population

We used data from a multicentric case-control study on OPC carried out between 1992 and 2009, in the greater Milan area (Northern Italy), the provinces of Pordenone (North-East Italy), Rome and Latina (Central Italy). The study included 946 patients (756 men, and 190 women; median age 58 years, range 22–79 years) admitted to major hospitals in the study areas with incident, histologically confirmed OPC diagnosed within 1 year prior to the interview. Controls were 2492 subjects (1497 men and 995 women; median age 58 years, range 19–82 years) admitted to the same hospital networks in the same period for acute, non-neoplastic conditions, unrelated to alcohol drinking, tobacco smoking or long term dietary modifications. Among the controls, 24% were admitted for traumas, 27% for other orthopedic causes, 22% for surgical conditions, 9% for eye diseases, and 19% for miscellaneous other illnesses. Less than 5% of potential cases and controls contacted refused to participate. Centrally trained interviewers used the same structured questionnaire and coding material in all centers. The questionnaire collected information on socio-demographic characteristics such as education and occupation, tobacco and alcohol consumption, physical activity, anthropometric measures, personal medical history and family history of cancer. The study protocol was approved by the local ethical committees and all participants gave informed consent to participate. Detailed information on the study is given elsewhere[27].

Dietary intake assessment

Dietary intake was assessed through a validated [28] and reproducible [29] food frequency questionnaire(FFQ) including weekly consumption of 78 food items or recipes and five alcoholic beverages. Intake frequencies lower than once in a week, but at least once per month were coded as 0.5. Italian food composition tables were used to calculate energy intake and nutrients [30] .

Food items and recipes were grouped into 25 food groups according to their nutritional characteristics. Daily intake (g/d) was calculated for the food groups (Supplementary Table 1) using standard portion sizes. The major part of food groups' distributions were skewed with a huge spike at zero (nonconsumers). We decided for categorization instead of transformation as we wanted to treat zeros differently from non-zeros. Especially with FFQ [26], zeros are expected to represent habitual non-consumption, therefore, they are likely to correspond to specific population subgroups, e.g. vegetarians. Moreover, original variables were not continuous in nature. Categorization was done as follows. Indicators with a percentage of nonconsumers less than 10% (n=16) were categorized in a 2-level variable: below or above the median. Indicators with a proportion of non consumers between 10-50% (n=6) were categorized in a 3-level variable: nonconsumers and below or above the median among consumers (g/d>0). Indicators with a proportion of nonconsumers (n=3) equal or higher than 50% were dichotomized in consumers and nonconsumers. Categories were considered to be nominal, rather than ordinal due to a higher classification performance.

Statistical methods

We defined dietary patterns as unobserved classes in a population having different food consumption probability distributions. LCA was used to identify a set of mutually exclusive clusters of individuals, based on their responses to the set of observed food groups (indicators).

In our latent class model, total non-alcoholic energy intake was used as a control variable affecting each of the food groups directly, and these effects were evaluated using Wald tests. Thus, we made sure that the encountered latent patterns represent actual diet types rather than energy intake types. We also allowed for total non-alcoholic energy intake to be associated with the latent diet patterns, since it is unlikely that such an association is fully absent. This latter effect was also evaluated using the Wald test.

Given the assumption of conditional independence, any residual association between two indicators after including the latent variable indicates a violation. These can be quantified and tested using the bivariate residuals (BVR) statistic. When the BVR becomes too high, and it is theoretically warranted, the indicators can be allowed to covary to locally relax the assumption. Therefore, we evaluated the within-class residual correlations (local dependencies) among food groups intake by checking the BVR between pairs of food groups, and allowed for correlated errors between food groups that showed high values of the statistic.

Class parity was determined as follows. The trivial 1-class model, where all individuals belong to the same class, was first fitted. The number of classes was successively increased by 1 in each subsequent model until the value of the Bayesian information criterion (BIC) ceased to monotonically decrease or until the number of classes reached 10. This parity was chosen as the maximum to ensure substantial reduction in dimension from 25 food groups.

Names of the clusters were chosen according to the conditional distribution of food groups intake giving the latent classes (class-specific response probabilities).

Subjects were assigned to latent classes based on their posterior class membership probabilities. These were obtained from the estimated parameters of the LC model and their observed responses. Proportional allocation was chosen to permit a 'soft' classification, assigning subjects to each class with a weight equal to posterior membership probability for that class.

LCA was performed on both cases and controls. Analysis on controls only was carried out to check the robustness of the previous solution. As dietary patterns identified on controls were consistent (number and characteristics of the patterns) with the ones obtained on the overall dataset we based all our analysis on the overall dataset. To confirm the internal reproducibility of the chosen solution the analysis was performed separately in two randomly selected subsets of the original data several times.

We examined the distribution of the identified clusters according to the selected nutrients used as dietary variables in a previous publication [7], performing a comparison between the previous nutrient based dietary patterns and the ones from the current analysis. As the previous publication regarded a subsample of the current database (data collected between 1992 and 2005 only), the same analysis of the previous study was repeated and the robustness of the solution was checked and satisfactory proved (data not shown).

We also assessed the characterization of the clusters in terms of selected demographic/ anthropometric characteristics and the known main risk factors for OPC, tobacco and alcohol consumption.

Odds ratios (OR) and related 95% confidence intervals (CI) for OPC risk were derived through a multiple logistic regression model using the class assignment to evaluate the effect of dietary patterns on the risk of OPC. To account for the major risk factors and for differences between cases and controls [7], models were adjusted including terms for age, sex, education, body mass index (BMI), tobacco and alcohol consumption, as confounders. Bolck et al. demonstrated that the classical three-step approach, which first identifies patterns, then assigns subject to each cluster and finally builds the prediction model, underestimates the associations between covariates and class membership [31]. They proposed resolving this problem by means of a specific correction method. In this study, a new maximum likelihood (ML) correction method proposed by Vermunt[32] was applied which is more efficient and incorporates uncertainty about classification in the estimation procedure[31]. As classification errors exist even in proportional assignment, this source of error or uncertainty must be taken into account when estimating effects between the latent variable and the outcome variables.

Stratified analyses were conducted according to smoking and alcohol drinking status. Heterogeneity across strata was tested by loglikelihood ratio test comparing models with and without interaction terms between the strata variable and the clusters variable. We also estimated ORs according to the clusters for sub-types of cancer separately. Heterogeneity between sub-sites was tested by loglikelihood ratio test comparing multinomial models with and without restrictions on the parameters constraining the effect of clusters to be equal between the different sub-types.

Statistical analysis were performed using SAS 9.4 (SAS Institute, Cary, NC, USA) and Latent GOLD 5.1 (Vermunt & Magidson,2016) statistical software.

Data availability

Anonymous data will be made available upon reasonable request.

Results

When fitting the LC model, we chose the solution with 4 classes according to the BIC.

Cluster prevalence and food groups consumption were conditioned on total non-alcoholic intake in the final models as there were significant associations according to Wald tests on the related regression parameters (data not shown).

The BVR statistics showed high correlated errors between sugar and coffee food groups and between pulses and soups food groups. As the FFQ questions on sugar were related to hot beverages and in the construction of food group variables pulses and soup shared an item, we specified correlated errors between coffee and sugar groups and between soup and pulses groups in the final model. The split-half assessment confirmed internal reproducibility of the chosen solution in subsamples.

Table 1 reports the conditional distribution of food groups intake, giving the latent classes for the food groups more relevant in discriminating and labeling the clusters. The complete table with all food items is given in Supplementary Table 2. Cluster 1 labeled 'Prudent pattern', showed higher probability to consume more fruiting vegetables, citrus fruit and all other kinds of fruits, leafy vegetables, tea and decaffeinated coffee and lower probability to consume red meat. Subjects in Cluster 2, 'Western pattern', reported higher consumption of red meat and lower consumption of fruiting and cruciferous vegetables, and fruits. Clusters 3 and 4 were related to similar food groups, but with a difference in the amount of intake. We termed Cluster 3 'Lower consumers-combination pattern' as people in it were less likely to eat leafy and fruiting vegetables, fish, fruits, red meat, bread, pulses, tea/decaffeinated coffee, potatoes and white meat. Cluster 4 had higher probability to eat bread, leafy, fruiting and other vegetables, red and white meat, while showed a lower probability to consume sugary drinks and desserts, fish, processed meat, citrus fruit, tea/decaffeinated coffee, cheese and coffee. We called this cluster 'Higher consumers-combination pattern'. Estimated cluster's sizes were 36.8% (n=1265) for the 'Prudent pattern', 27.0% (n=929) for the 'Western pattern', 21.1% (n=725) for the 'Lower consumers-combination pattern' and 15.1% (n=519) for the 'Higher consumers-combination pattern'.

Descriptions of the clusters for selected variables are given in Table 2. Differences between the clusters in demographics (age, sex, education) emerged. Regarding the two main risk factors for OPC, the 'Prudent pattern' was characterized by a lower consumption of alcohol and tobacco (respectively, 24.5% and 44.8% were non-consumers) with respect to other clusters. The 'Higher consumers-combination pattern' had the highest proportion of heavy drinkers (64.4%), followed by the 'Western pattern' (52.5%). These two pattern showed a similar distribution in terms of tobacco consumption, with the smallest proportion of non-smokers (29.0% and 27.1%, respectively).

Table 3 reports cluster's characteristics in terms of non-alcoholic energy intake and nutrients intake: the 'Higher consumers-combination' pattern showed the highest energy intake, followed by the 'Western', the 'Prudent' and the 'Lower consumers-combination'. The 'Prudent' pattern's diet was characterized by high intake of soluble carbohydrates, calcium, vitamin C, beta-carotene equivalents and total fiber. Subjects in the 'Western' pattern reported a diet rich in animal protein, animal fat, cholesterol, saturated fatty acids, phosphorus, vitamin B2 and retinol. The 'Higher consumers-combination' pattern exhibited a diet with high intake of most nutrients. Subjects in the 'Lower consumers-combination' pattern reported a diet with a low intake of every nutrient.

Table 4 gives the ORs and corresponding CIs for OPC and sub-sites of OPC, by the classification in the four dietary pattern from the composite model including the relevant confounding and risk variables. Compared to the 'Prudent' pattern, the 'Western' and the 'Lower consumers-combination' ones were positively related to the risk of OPC (OR=2.56, 95% CI: 1.90 – 3.45 and OR=2.23, 95% CI: 1.64 – 3.02). The 'Higher consumers-combination pattern' did not differ significantly from the Prudent pattern (OR=1.28, 95% CI: 0.92 – 1.77). Consistent results were also observed in relation to oral (OR=2.47, 95% CI: 1.70 – 3.59 for 'Western', OR=2.35, 95% CI: 1.63 – 3.39 for 'Lower consumers-combination' and OR=1.06, 95% CI: 0.67 – 1.67 for 'Higher consumers-combination' pattern), oropharyngeal (OR=2.61, 95% CI: 1.56 – 4.38 for 'Western', OR=1.92, 95% CI: 1.10 – 3.36 for 'Lower consumers-combination' and OR=1.47, 95% CI: 0.83 – 2.60 for 'Higher consumers-combination' pattern), and hypopharyngeal cancer (OR=3.48, 95% CI: 1.78 – 6.81 for 'Western', OR=2.41, 95% CI: 1.11 – 5.24 for 'Lower consumers-combination' and OR=2.05, 95% CI: 0.97 – 4.32 for 'Higher consumers-combination' pattern). No significant differences emerged between sub-sites ($p_{heterogeneity}$= 0.715).

No relevant differences emerged across strata of smoking status ($p_{heterogeneity}$=0.147) and alcohol drinking habits ($p_{heterogeneity}$=0.542) (Table 5).

Discussion

Our objective was to identify dietary patterns conceived as mutually exclusive groups of people characterized by similar food intake and to compare the resulting patterns in terms of OPC risk.

We identified four dietary patterns. We found a protective effect on OPC of the pattern characterized by high intake of fruits and leafy/fruiting vegetables and low intake of red meat ('Prudent pattern'). Compared to this group, the pattern distinguished by a diet rich in red meat and poor in fruit, cruciferous and fruiting vegetables consumption ('Western pattern'), was associated to an increase in the risk of OCP. The 'Lower consumers-combination' pattern, which showed a diet poor in most of the foods considered, was positively related OPC risk. No difference in risk between the 'Prudent pattern' and the 'Higher consumers-combination' emerged. This pattern exhibited a diet extremely varied in terms of foods.

Our results were consistent with the evidence from the studies addressing influence of *a posteriori* dietary patterns on OPC [6 19 21-25]. A pattern related to high intake of fruit and vegetables, named 'Prudent' or 'Vegetable and Fruit' was found and associated to a decrease of OPC risk in most of these studies. A second pattern characterized by high intake of meat was also often found and related to an increased OCP risk, even though the strength of the association is not consistent throughout all the studies. Apart from meat, this pattern was also distinguished by other foods that varied in the different studies, which is why the label used changed from 'Western' to 'Snacks', for example. Beside these two main patterns, the number and the types of further patterns differed in the various publications. Patterns related to a combination of different food types varied from 'Traditional' country specific diet, 'Combination', 'Modern' 'Monotonous', 'Starchy'. Concerning a potential 'Traditional' pattern, a characteristic of our study was the adjustment for non-alcoholic energy intake directly in pattern identification. This had influence in particular on foods having higher correlation with energy intake, like pasta, bread and sugary types of food. In fact, adjusting for this covariate had the effect of not identifying a pattern strongly related to these foods. This effect was noted also in other studies analysing the effect of energy correction in dietary patterning [33 34]. Studies focusing on *a priori* dietary patterns highlighted a protective effect of the Mediterranean-type of diet [35-40] and of diversity in food consumption [41 42], particularly within vegetables and fruits [41]. In our analysis the pattern characterized with an extremely varied diet in terms of foods and nutrients ('Higher consumers-combination pattern') did not significantly differ from the 'Prudent' one in the risk of OCP.

LCA can provide interesting insights into dietary patterning, allowing to identify prevalent types of eating behaviour in a population and to compare risk for people with different types of diet.

Empirical *a posteriori* dietary patterns are derived predominantly using PCA, exploratory/ confirmatory FA or CA. Despite the same label, dietary patterns derived from different methods are conceptually different [34]. PCA groups dietary variables in combinations that are representative of the original features of the dietary dataset. These combinations are the identified dietary patterns. FA groups dietary items into dimensions with the assumption that if those items highly correlate, they may measure aspects of a common underlying dimension that represents a dietary pattern. Therefore, these techniques help understand which dietary items are eaten in combination and to study the effect of these dietary dimensions/combinations on health outcomes. A disadvantage of these methods is that they do not give rise to mutually exclusive groups. Thus, when the interest is to compare subgroups of subjects, an additional step of cross-classification of the dimensions/combinations is needed. While FA and PCA group foods/nutrients items, CA groups individuals into relatively homogeneous classes. Therefore, CA define dietary patterns as classes of subjects where subjects share similar dietary habits and they are useful to study how these groups differs in terms of risk of an health outcome. However, this method assumes classification uncertainty to be 0 and it mostly relies on non-parametric approaches which lack in assessment of goodness of fit. Moreover, all the above-mentioned techniques do not take into account external covariates (e.g. confounders).

The application of a LC model to the Italian case-control study on OPC overcomes problems inherent in the traditional methods and gives further advantages in dietary patterning, such as a probability-based classification under a general parametric approach, pattern prevalence estimation and covariate adjustment.

When the method applied is devoted to classifying individuals according to their food/nutrient intake and the data came from areas characterized by homogeneous diet, the clustering method likely identifies groups characterized by similar diet but different amount of food intake and consequently, energy intake. Thus, not taking into account energy intake could lead to results that reflect the effect of the energy intake on disease instead of the real effect of food itself. Direct covariate adjustment is not possible with other standard methods. With the classical methods, total non-alcoholic energy intake can be taken into account by working with foods already adjusted, usually with the residual method [43] or by the inclusion of the related variable in dimensions identification together with the dietary indicators. Even though the last approach has been commonly applied, it implies that the energy variable will influence the formation of dimensions and would, in essence become an 'indicator' of the dietary patterns, treated in the same way as the proper dietary variables. Theoretically, it would be preferable to separate proper indicators from confounding variables. In contrast, LCA can be easily extended to include confounders keeping the measurement part of the model (which defines the relation between the latent variables and its indicators) separate from the structural one (which defines the relations with external variables) and permitting both to be estimated using a single ML estimation algorithm. In the current study, with LCA, we could correct dietary patterns including the energy intake as an external covariate, keeping it separated from the proper dietary indicators (the single food group items) to work with items that represented relative (adjusted) rather than absolute food intake. Then as the patterns size also resulted depending on energy intake, it was possible to take this into account in the pattern identification by allowing the distribution of the latent pattern variable itself to depend on the covariate.

Another strength of LCA is the possibility first to inspect and then take into account for possible correlation between some dietary variables within classes. When the same measurement instrument (e.g. FFQ) is used for all foods, correlated errors are expected because of self-reporting bias and similar ways of wording items in the questionnaire. That leads to the importance of considering the effect of correlated errors between food intakes in the dietary pattern identification[26]. We were able to identify and allow correlated errors between some food groups that resulted from the nature of the assessment instrument and coding of food groups.

Further, techniques like FA/PCA can be fairly applied on continuous variables (like nutrients), but when dealing with categorical ones (like foods/food groups) they may result in biased estimation. LCA, instead, can properly deal with categorical indicators and also be extended to ordinal or continuous variables in the framework of finite mixture models.

A previous analysis on dietary patterns and OCP [7] used data from this multicentric case-control study collected between 1992 and 2005, while the current database was updated including additional 142 cases and 412 controls. The previous study aimed to identify dietary patterns conceived as 'combination of dietary components intended to summarize key aspect of the diet for a given population' [7] by performing PCFA on selected 28 nutrients and total non-alcoholic energy.

In contrast, in the current study, we hypothesized the existence of unobservable, underlying groups that we defined as dietary patterns, which accounted for the difference in food consumption between subjects. Hence, the two studies were conceptually and methodologically different. However, together they provide two different perspectives that combined can give a broader insight on the effect of the diet on the risk of OPC cancer. The consistence of our results with the previous findings reinforces the reliability of the identified dietary patterns. The 'Prudent' pattern's diet was associated to all the nutrients that showed highest factor loadings for the pattern with the lowest risk for OPC ('Vitamin and fiber') in the previous study. Instead, the ‘Western’ pattern was associated to many nutrients that showed highest factor loadings for a previous pattern associated to an increased risk of OCP (‘Animal products’). The ‘Higher consumers-combination’ pattern exhibit high consume of nutrients associated to both previous protective (‘Vitamin and fiber’, ‘Starch’, ‘Unsaturated fats’) and risk factors patterns (‘Animal products’).

With reference to possible study limitations, hospital controls may be not representative of the general population for various aspects including dietary habits [44]. To limit this potential bias, controls were included according to a large spectrum of admission diagnoses, excluding the ones related to major know risk factors for OPC, such as tobacco and alcohol habits or long term dietary modifications. The recent diagnosis may affect patient’s recall, but at the time of data collection, as the knowledge and awareness of the potential role of diet on OPC risk was scarce, such a misclassification was likely limited. Moreover, both cases and controls were interviewed in the same settings, by the same interviewers and with the same reproducible [29] and valid [28] FFQ. Among the strengths of this study were the large sample size, the almost complete participation and the comparable catchment areas of cases and controls. Information on diet, alcohol and tobacco consumption was satisfactorily reproducible [28 29 45].

In conclusion, LCA gives further insights to dietary pattern research, allowing for the definition and estimation of the prevalence of different groups of subject characterized by different dietary choice s, and comparing those groups in relation to important health outcomes like OPC. Thus, it adds a new perspective to the classical PCA/FA attempting to explain which foods are eaten in combination and their effect on health outcomes, and it has inferential advantages compared to CA.

Conflict of interest

None declared.

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, Znaor A, Bray F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 2019;**144**(8):1941-53.
2. Boeing H, Dietrich T, Hoffmann K, Pischon T, Ferrari P, Lahmann PH, Boutron-Ruault MC, Clavel-Chapelon F, Allen N, Key T, Skeie G, Lund E, Olsen A, Tjonneland A, Overvad K, Jensen MK, Rohrmann S, Linseisen J, Trichopoulou A, Bamia C, Psaltopoulou T, Weinehall L, Johansson I, Sanchez MJ, Jakszyn P, Ardanaz E, Amiano P, Chirlaque MD, Quiros JR, Wirfalt E, Berglund G, Peeters PH, van Gils CH, Bueno-de-Mesquita HB, Buchner FL, Berrino F, Palli D, Sacerdote C, Tumino R, Panico S, Bingham S, Khaw KT, Slimani N, Norat T, Jenab M, Riboli E. Intake of fruits and vegetables and risk of cancer of the upper aero-digestive tract: the prospective EPIC-study. Cancer Causes Control 2006;**17**(7):957-69.
3. Bravi F, Polesel J, Garavello W, Serraino D, Negri E, Franchin G, La Vecchia C, Bosetti C. Adherence to the World Cancer Research Fund/American Institute for Cancer Research recommendations and head and neck cancers risk. Oral Oncol 2017;**64**:59-64.
4. Chuang SC, Jenab M, Heck JE, Bosetti C, Talamini R, Matsuo K, Castellsague X, Franceschi S, Herrero R, Winn DM, La Vecchia C, Morgenstern H, Zhang ZF, Levi F, Dal Maso L, Kelsey K, McClean MD, Vaughan T, Lazarus P, Muscat J, Ramroth H, Chen C, Schwartz SM, Eluf-Neto J, Hayes RB, Purdue M, Boccia S, Cadoni G, Zaridze D, Koifman S, Curado MP, Ahrens W, Benhamou S, Matos E, Lagiou P, Szeszenia-Dabrowska N, Olshan AF, Fernandez L, Menezes A, Agudo A, Daudt AW, Merletti F, Macfarlane GJ, Kjaerheim K, Mates D, Holcatova I, Schantz S, Yu GP, Simonato L, Brenner H, Mueller H, Conway DI, Thomson P, Fabianova E, Znaor A, Rudnai P, Healy CM, Ferro G, Brennan P, Boffetta P, Hashibe M. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. Cancer Causes Control 2012;**23**(1):69-88.
5. Di Maso M, Talamini R, Bosetti C, Montella M, Zucchetto A, Libra M, Negri E, Levi F, La Vecchia C, Franceschi S, Serraino D, Polesel J. Red meat and cancer risk in a network of case-control studies focusing on cooking practices. Ann Oncol 2013;**24**(12):3107-12.
6. Edefonti V, Bravi F, Garavello W, La Vecchia C, Parpinel M, Franceschi S, Dal Maso L, Bosetti C, Boffetta P, Ferraroni M, Decarli A. Nutrient-based dietary patterns and laryngeal cancer: evidence from an exploratory factor analysis. Cancer Epidemiol Biomarkers Prev 2010;**19**(1):18-27.
7. Edefonti V, Bravi F, La Vecchia C, Randi G, Ferraroni M, Garavello W, Franceschi S, Talamini R, Boffetta P, Decarli A. Nutrient-based dietary patterns and the risk of oral and pharyngeal cancer. Oral Oncol 2010;**46**(5):343-8.
8. Franceschi S, Bidoli E, La Vecchia C, Talamini R, D'Avanzo B, Negri E. Tomatoes and risk of digestive-tract cancers. Int J Cancer 1994;**59**(2):181-4.
9. Franceschi S, Favero A, Conti E, Talamini R, Volpe R, Negri E, Barzan L, La Vecchia C. Food groups, oils and butter, and cancer of the oral cavity and pharynx. Br J Cancer 1999;**80**(3-4):614-20.
10. Freedman ND, Park Y, Subar AF, Hollenbeck AR, Leitzmann MF, Schatzkin A, Abnet CC. Fruit and vegetable intake and head and neck cancer risk in a large United States prospective cohort study. Int J Cancer 2008;**122**(10):2330-6.
11. Fund WCR, Research AIfC. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. 2007;**1**.
12. Garavello W, Lucenteforte E, Bosetti C, La Vecchia C. The role of foods and nutrients on oral and pharyngeal cancer risk. Minerva Stomatol 2009;**58**(1-2):25-34.
13. La Vecchia C, Chatenoud L, Franceschi S, Soler M, Parazzini F, Negri E. Vegetables and fruit and human cancer: update of an Italian study. Int J Cancer 1999;**82**(1):151-2.
14. La Vecchia C, Negri E, D'Avanzo B, Boyle P, Franceschi S. Dietary indicators of oral and pharyngeal cancer. Int J Epidemiol 1991;**20**(1):39-44.

15. Lagiou P, Talamini R, Samoli E, Lagiou A, Ahrens W, Pohlabeln H, Benhamou S, Bouchardy C, Slamova A, Schejbalova M, Merletti F, Richiardi L, Kjaerheim K, Agudo A, Castellsague X, Macfarlane TV, Macfarlane GJ, Biggs AM, Barzan L, Canova C, Simonato L, Lowry RJ, Conway DI, McKinney PA, Znaor A, McCartan BE, Healy CM, Marron M, Hashibe M, Brennan P. Diet and upper-aerodigestive tract cancer in Europe: the ARCAGE study. Int J Cancer 2009;**124**(11):2671-6.
16. Levi F, Pasche C, Lucchini F, Bosetti C, La Vecchia C. Processed meat and the risk of selected digestive tract and laryngeal neoplasms in Switzerland. Ann Oncol 2004;**15**(2):346-9.
17. Tavani A, La Vecchia C, Gallus S, Lagiou P, Trichopoulos D, Levi F, Negri E. Red meat intake and cancer risk: a study in Italy. Int J Cancer 2000;**86**(3):425-8.
18. Turati F, Rossi M, Pelucchi C, Levi F, La Vecchia C. Fruit and vegetables and cancer risk: a review of southern European studies. Br J Nutr 2015;**113 Suppl 2**:S102-10.
19. Amtha R, Zain R, Razak IA, Basuki B, Roeslan BO, Gautama W, Purwanto DJ. Dietary patterns and risk of oral cancer: a factor analysis study of a population in Jakarta, Indonesia. Oral Oncol 2009;**45**(8):e49-53.
20. Bravi F, Edefonti V, Randi G, Ferraroni M, La Vecchia C, Decarli A. Dietary patterns and upper aerodigestive tract cancers: an overview and review. Ann Oncol 2012;**23**(12):3024-39.
21. De Stefani E, Boffetta P, Ronco AL, Correa P, Oreggia F, Deneo-Pellegrini H, Mendilaharsu M, Leiva J. Dietary patterns and risk of cancer of the oral cavity and pharynx in Uruguay. Nutr Cancer 2005;**51**(2):132-9.
22. De Stefani E, Deneo-Pellegrini H, Boffetta P, Ronco AL, Aune D, Acosta G, Mendilaharsu M, Brennan P, Ferro G. Dietary patterns and risk of cancer: a factor analysis in Uruguay. Int J Cancer 2009;**124**(6):1391-7.
23. 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.
24. Marchioni DM, Fisberg RM, Francisco de Gois Filho J, Kowalski LP, Brasilino de Carvalho M, Abrahao M, Latorre Mdo R, Eluf-Neto J, Wunsch Filho V. Dietary patterns and risk of oral cancer: a case-control study in Sao Paulo, Brazil. Rev Saude Publica 2007;41(1):19-26.
25. Toledo AL, Koifman RJ, Koifman S, Marchioni DM. Dietary patterns and risk of oral and pharyngeal cancer: a case-control study in Rio de Janeiro, Brazil. Cad Saude Publica 2010;**26**(1):135-42.
26. Fahey MT, Thane CW, Bramwell GD, Coward WA. Conditional Gaussian mixture modelling for dietary pattern analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2007;**170**(1):149-66.
27. Shivappa N, Hebert JR, Rosato V, Garavello W, Serraino D, La Vecchia C. Inflammatory potential of diet and risk of oral and pharyngeal cancer in a large case-control study from Italy. Int J Cancer 2017;**141**(3):471-79.
28. Decarli A, Franceschi S, Ferraroni M, Gnagnarella P, Parpinel MT, La Vecchia C, Negri E, Salvini S, Falcini F, Giacosa A. Validation of a food-frequency questionnaire to assess dietary intakes in cancer studies in Italy. Results for specific nutrients. Ann Epidemiol 1996;**6**(2):110-8.
29. Franceschi S, Negri E, Salvini S, Decarli A, Ferraroni M, Filiberti R, Giacosa A, Talamini R, Nanni O, Panarello G, et al. Reproducibility of an Italian food frequency questionnaire for cancer studies: results for specific food items. Eur J Cancer 1993;**29A**(16):2298-305.
30. Gnagnarella P, Parpinel M, Salvini S, Franceschi S, Palli D, Boyle P. The update of the Italian food composition database. . Journal of Food Composition and Analysis 2004;**17**(3-4):509-22.
31. Bolck A, Croon M, Hagenaars J. Estimating latent structure models with categorical variables: one-step versus three-step estimators. Political Analysis 2004;**12**(1):3-27.
32. Vermunt JK. Latent class modeling with covariates: two improved three-step approaches. Political Analysis 2010;**18**(4):450-69.
33. Balder HF, Virtanen M, Brants HA, Krogh V, Dixon LB, Tan F, Mannisto S, Bellocco R, Pietinen P, Wolk A, Berrino F, Van den Brandt PA, Hartman AM, Goldbohm RA. Common and country-specific dietary patterns in four European cohort studies. J Nutr 2003;**133**(12):4246-51.

34. Northstone K, Ness AR, Emmett PM, Rogers IS. Adjusting for energy intake in dietary pattern investigations using principal components analysis. Eur J Clin Nutr 2008;**62**(7):931-8.
35. Bosetti C, Gallus S, Trichopoulou A, Talamini R, Franceschi S, Negri E, La Vecchia C. Influence of the Mediterranean diet on the risk of cancers of the upper aerodigestive tract. Cancer Epidemiol Biomarkers Prev 2003;**12**(10):1091-4.
36. Filomeno M, Bosetti C, Garavello W, Levi F, Galeone C, Negri E, La Vecchia C. The role of a Mediterranean diet on the risk of oral and pharyngeal cancer. Br J Cancer 2014;**111**(5):981-6.
37. Giraldi L, Panic N, Cadoni G, Boccia S, Leoncini E. Association between Mediterranean diet and head and neck cancer: results of a large case-control study in Italy. Eur J Cancer Prev 2017;**26**(5):418-23.
38. Grosso G, Buscemi S, Galvano F, Mistretta A, Marventano S, La Vela V, Drago F, Gangi S, Basile F, Biondi A. Mediterranean diet and cancer: epidemiological evidence and mechanism of selected aspects. BMC Surg 2013;**13 Suppl 2**:S14.
39. Li WQ, Park Y, Wu JW, Goldstein AM, Taylor PR, Hollenbeck AR, Freedman ND, Abnet CC. Index-based dietary patterns and risk of head and neck cancer in a large prospective study. Am J Clin Nutr 2014;**99**(3):559-66.
40. Samoli E, Lagiou A, Nikolopoulos E, Lagogiannis G, Barbouni A, Lefantzis D, Trichopoulos D, Brennan P, Lagiou P. Mediterranean diet and upper aerodigestive tract cancer: the Greek segment of the Alcohol-Related Cancers and Genetic Susceptibility in Europe study. Br J Nutr 2010;**104**(9):1369-74.
41. Garavello W, Giordano L, Bosetti C, Talamini R, Negri E, Tavani A, Maisonneuve P, Franceschi S, La Vecchia C. Diet diversity and the risk of oral and pharyngeal cancer. Eur J Nutr 2008;**47**(5):280-4.
42. Levi F, Pasche C, La Vecchia C, Lucchini F, Franceschi S, Monnier P. Food groups and risk of oral and pharyngeal cancer. Int J Cancer 1998;**77**(5):705-9.
43. Willett W. *Nutritional epidemiology.* : (Vol. 40) Oxford university press, 2012.
44. La Vecchia C, Tavani A, Franceschi S, Levi F, Corrao G, Negri E. Epidemiology and prevention of oral cancer. Oral Oncol 1997;**33**(5):302-12.
45. Ferraroni M, Decarli A, Franceschi S, La Vecchia C, Enard L, Negri E, Parpinel M, Salvini S. Validity and reproducibility of alcohol consumption in Italy. Int J Epidemiol 1996;**25**(4):775-82.

Table 1. Probabilities of consumption for selected food items by dietary patterns derived from LCA. Italy, 1992-2009.

		Prudent %	Western %	Lower consumers-combination %	Higher consumers-combination %
Bread	Below median	57,8	44,1	**63,8**	22,9
	Above median	42,2	55,9	36,2	**77,1**
White meat	Below median	45,9	52,2	**61,3**	38,1
	Above median	54,1	47,8	38,7	**61,9**
Red meat	Below median	**61,1**	30,9	**65,0**	35,7
	Above median	38,9	**69,1**	35,0	**64,3**
Processed meat	Below median	50,7	47,3	50,8	**63,2**
	Above median	49,3	52,7	49,2	36,8
Fish	Below median	45,6	49,2	**69,1**	**66,8**
	Above median	54,4	50,8	30,9	33,2
Cheese	Below median	41,1	48,4	58,5	**62,6**
	Above median	58,9	51,6	41,5	37,4
Potatoes	Below median	58,2	45,7	**61,7**	41,0
	Above median	41,8	54,3	38,3	59,0
Pulses	Below median	42,7	57,3	**62,6**	44,0
	Above median	57,3	42,7	37,4	56,0
Leafy vegetables	Below median	35,5	59,8	**74,6**	22,5
	Above median	**64,5**	40,2	25,4	**77,5**
Fruiting vegetables	Below median	24,4	**79,2**	**71,1**	31,0
	Above median	**75,6**	20,8	28,9	**69,0**
Cruciferous vegetables	Not consumed	18,3	14,8	51,5	24,9
	Below median	25,1	**67,3**	19,8	46,6
	Above median	56,6	17,9	28,7	28,5
Other vegetables	Not consumed	6,3	1,5	41,3	6,4
	Below median	36,1	55,9	54,3	22,7
	Above median	57,6	42,5	4,4	**70,9**
Citrus fruit	Not consumed	4,1	7,2	25,3	17,3
	Below median	24,2	59,0	39,9	**62,8**
	Above median	**71,7**	33,8	34,8	19,9
Other fruits	Below median	29,8	**63,6**	**67,8**	50,2
	Above median	**70,2**	36,4	32,3	49,8
Sugary drinks	Not consumed	54,2	41,2	59,9	**70,9**
	Consumed	45,8	58,8	40,1	29,1
Desserts	Below median	44,3	51,8	59,2	**68,0**
	Above median	55,7	48,2	40,8	32,0
Coffee	Below median	52,2	52,8	58,5	**60,6**
	Above median	47,8	47,2	41,5	39,4
Tea/Decaffeinated coffee	Not consumed	39,9	44,6	**62,3**	**62,8**
	Consumed	**60,1**	55,4	37,7	37,2
Cluster's size		36,8	27,0	21,1	15,1

Probabilities greater or equal to 60.0% defined dominant food groups for each cluster, and were shown in bold typeface. The latent class model was adjusted for non-alcoholic energy intake.

Table 2. Dietary patterns' characteristics according to selected sociodemographic variables. Italy, 1992-2009.

		Prudent %	Western %	Lower consumers-combination %	Higher consumers-combination %
	Cases	20,0	35,1	32,6	23,5
	Controls	80,0	64,9	67,4	76,5
Age (years)	<50	22,7	25,3	20,7	19,4
	50-59	30,1	31,4	26,8	34,8
	60-69	33,6	31,6	36,0	35,7
	>69	13,6	11,7	16,5	10,1
Sex	Male	53,1	73,8	65,6	81,1
	female	46,9	26,2	34,4	18,9
Education (years)	<7	46,4	55,7	55,3	64,7
	7-11	30,5	28,6	27,5	26,1
	>11	23,1	15,7	17,2	9,2
Alcoholic intake (weekly units)	0	24,5	12,4	22,0	9,0
	1-6	14,0	8,5	9,4	5,7
	7-13	15,0	10,5	11,1	8,3
	14-20	19,2	16,1	16,5	12,6
	>20	27,3	52,5	41,0	64,4
Smoking Habit (cig/d)	Never smoked	44,8	27,1	34,0	29,0
	Ex smoker	28,9	32,5	27,2	32,5
	<15	10,8	13,7	12,2	12,1
	>14	15,5	26,7	26,6	26,4
BMI	<18.5	2,1	2,1	3,4	1,8
	18.6-25.9	54,3	56,2	58,9	51,8
	26-29.9	32,0	29,2	25,6	32,8
	>29.9	11,6	12,5	12,1	13,6

BMI: body mass index; cig/d: cigarettes/day.

Table 3. Dietary patterns' characteristics according to non-alcoholic energy intake and selected nutrients. Mean intake for each dietary pattern. Italy, 1991-2009.

	Prudent	Western	Lower consumers-combination	Higher consumers-combination
Energy intake (kcal)	2252,4	2305,9	1838,0	2582,6
Animal protein (g)	59,2	63,6	50,7	64,4
Vegetable protein (g)	32,0	32,2	26,8	38,6
Animal fat (g)	41,7	46,0	36,3	46,4
Vegetable fat (g)	45,8	43,5	31,1	56,1
Cholesterol (mg)	301,7	340,0	253,1	336,7
Saturated fatty acids (g)	28,4	29,6	23,2	31,5
Monounsat. fatty acids(g)	41,3	39,5	30,0	46,8
Polyunsat. fatty acids (g)	1,2	1,3	1,0	1,4
Starch (g)	179,1	194,0	162,9	229,2
Soluble carbohydrate(g)	115,5	105,7	84,1	105,7
Sodium (mg)	2177,9	2318,1	1931,6	2624,3
Calcium (mg)	987,8	949,5	777,2	954,9
Potassium (mg)	4084,5	3873,1	3162,9	4357,0
Phosphorus (mg)	1554,0	1628,7	1317,2	1728,1
Iron (mg)	14,3	15,6	12,2	18,0
Zinc (mg)	12,9	13,8	10,9	14,7
Thiamin (vit. B1) (mg)	0,9	0,9	0,7	1,0
Riboflavin (vit. B2) (mg)	1,7	1,7	1,3	1,7
Vitamin B6 (mg)	2,0	2,1	1,6	2,3
Total folate (μg)	301,7	284,0	214,1	316,5
Niacin (mg)	19,1	19,9	15,9	21,3
Vitamin C (mg)	180,5	131,1	99,7	145,9
Retinol (μg)	770,0	1268,4	576,8	873,7
Beta-carotene equivalents (μg)	4807,6	3590,7	2501,7	4600,1
Lycopene (μg)	7172,4	7123,2	6324,0	8782,7
Vitamin D (μg)	3,3	3,3	2,5	3,1
Vitamin E (mg)	15,3	14,2	10,4	18,2
Total fiber (g)	18,1	15,0	12,3	18,2

Table 4. Odds ratios (OR) and the corresponding 95% confidence intervals (CIs) of oral and pharyngeal cancer for each cluster adjusting for known confounders: overall and per cancer subsite. Italy, 1992-2009.

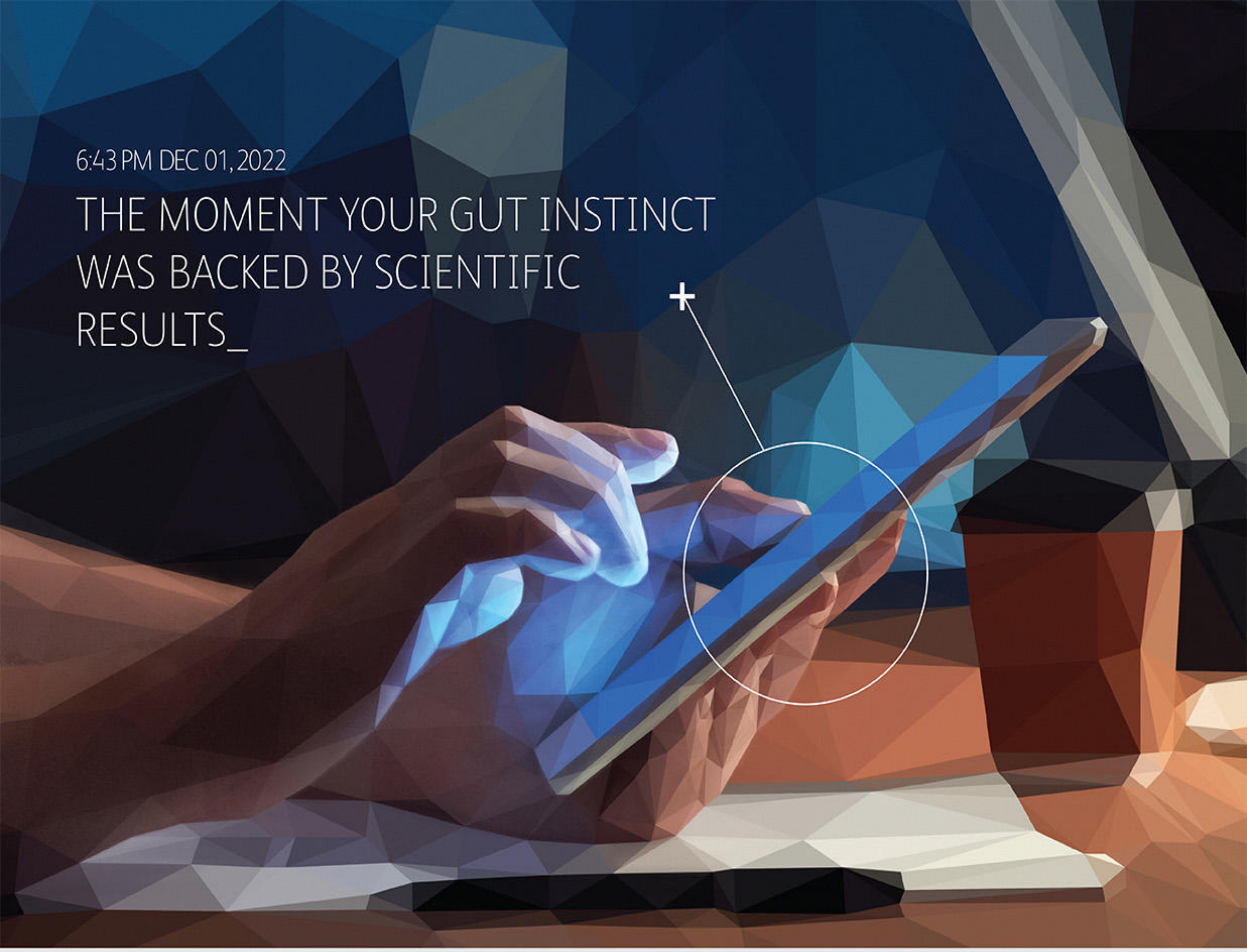
		Adjusted OR (95% CI) [a]			
		Overall	*Cancer subtype* [c]		
			Oral	Oropharynx	Hypopharynx
	cases/controls	946/2492	506/2492	258/2492	160/2492
Dietary Patterns	Prudent [b]	1	1	1	1
	Western	2.56 (1.90 – 3.45)	2.47 (1.70 – 3.59)	2.61 (1.56 – 4.38)	3.48 (1.78 – 6.81)
	Lower consumers-combination	2.23 (1.64 – 3.02)	2.35 (1.63 – 3.39)	1.92 (1.10 – 3.36)	2.41 (1.11 – 5.24)
	Higher consumers-combination	1.28 (0.92 – 1.77)	1.06 (0.67 – 1.67)	1.47 (0.83 – 2.60)	2.05 (0.97 – 4.32)
				$p_{heterogeneity}=0.715$	

[a] Adjusted for sex, age, education, body mass index, tobacco and alcohol intake
[b] Reference category.
[c] The sum of the cancer subsites does not add up to the total number of cases because of 22 cases with 'unspecified' cancer subsite.

Table 5. Odds ratios (OR) and the corresponding 95% confidence intervals (CIs) of oral and pharyngeal cancer for each cluster adjusting for known confounders across strata of alcohol and tobacco consumption. Italy, 1992-2009.

		Adjusted OR (95% CI) [a]			
		Smoking habit		Alcoholic intake	
		Ex/non-smoker	Current smoker	≤20 units/week	>20 units/week
	cases/controls	406/1843	540/649	313/1661	633/831
Dietary Patterns	Prudent [b]	1	1	1	1
	Western	2.67 (1.81 – 3.94)	3.02 (1.82 – 5.00)	2.10 (1.36 – 3.25)	4.00 (2.44 – 6.57)
	Lower consumers-combination	2.03 (1.36 – 3.03)	3.01 (1.78 – 5.12)	2.32 (1.56 – 3.47)	2.96 (1.79 – 4.87)
	Higher consumers-combination	0.97 (0.55 – 1.70)	1.71 (0.98 – 2.98)	1.09 (0.50 – 2.36)	1.76 (1.08 – 2.86)
		$p_{heterogeneity}=0.147$		$p_{heterogeneity}=0.542$	

[a] Adjusted for sex, age, education, body mass index, tobacco and alcohol intake when the covariate was not a stratification variable.
[b] Reference category.



6:43 PM DEC 01, 2022
THE MOMENT YOUR GUT INSTINCT WAS BACKED BY SCIENTIFIC RESULTS_