Eur J Nutr (2012) 51:755–764
DOI 10.1007/s00394-011-0255-7

ORIGINAL CONTRIBUTION

Applying multilevel model to the relationship of dietary patterns and colorectal cancer: an ongoing case–control study in Córdoba, Argentina

Sonia Alejandra Pou · María del Pilar Díaz · Alberto Rubén Osella

Received: 9 August 2011 / Accepted: 26 September 2011 / Published online: 12 October 2011

Abstract
Purposes Scientific literature has consistently shown the effects of certain diets on health but regional variations of dietary habits, and their relationship colorectal cancer (CRC) has been poorly studied in Argentina. Our aims were to identify dietary patterns and estimate their effect on CRC occurrence and to quantify the association between family history of CRC and CRC occurrence by applying multilevel models to estimate and interpret measures of variation.
Methods Principal components factor analysis was performed to identify dietary patterns that were then used in a multilevel logistic regression applied to an ongoing case–control data about dietary exposure and CRC occurrence taking into account familiar clustering.
Results Three dietary patterns were identified: "Southern Cone pattern" (red meat, wine, and starchy vegetables), "High-sugar drinks pattern", and "Prudent pattern". The study considered 41 cases and 95 controls. There was a significant promoting effects on CRC of "Southern Cone" (OR 1.5, 95%CI 1.0–2.2) and "High-sugar drinks" (OR 3.8, 95%CI 2.0–7.1) patterns, whereas "Prudent pattern" (OR 0.3, 95%CI 0.2–0.4) showed a significant protective effect at third tertile level. BMI, use of NSAIDs, and to have medical insurance showed significant effects. Variance of the random effect of family history of CRC was highly significant.
Conclusions This novel approach for Argentina showed that Southern Cone and High-sugar drinks patterns were associated with a higher risk of CRC, whereas the Prudent pattern showed a protective effect. There was a significant clustering effect of family history of CRC.

Keywords Dietary patterns · Argentina · Meat intake · Colorectal cancer · Multilevel

Introduction

Evidence about the promoting or preventing role of several foods or dietary components on cancer occurrence has been documented [1], although the focus on dietary patterns (integration of several dietary exposures into a single exposure) has emerged in cancer epidemiology in relatively recent times [2, 3]. The main advantage of this approach is that dietary patterns analysis allows to address the issue of collinearity of nutrients and interdependencies between foods and nutrients [4], simplifying also the interpretation of a complex and multidimensional phenomenon such as dietary intake. This approach has been recognized as particularly useful for colorectal cancer (CRC), as many dietary components have been associated

S. A. Pou
National Research Council (CONICET). Biostatistics Unit, School of Nutrition, Faculty of Medical Sciences, University of Córdoba, Córdoba, Argentina
e-mail: pousonia@conicet.gov.ar

M. d. P. Díaz (✉)
Biostatistics Unit, School of Nutrition, Faculty of Medical Sciences, University of Córdoba, Avenida Enrique Barros s/n Ciudad Universidad, CP 5000 Córdoba, Argentina
e-mail: pdiaz@fcm.unc.edu.ar

A. R. Osella
Laboratory of Epidemiology and Biostatistics, IRCCS "Saverio de Bellis", Castellana Grotte, Bari, Italy
e-mail: arosella@irccsdebellis.it

with this disease [5]. Previous analyses on single nutrients or foods and CRC risk showed direct relationships with red meat, processed meat, and alcoholic drinks and inverse relationships with foods containing dietary fiber, garlic, milk, and calcium [1].

There is also some evidence about the potential interaction between dietary factors and family history of cancer, including the CRC [5]. Particularly in this cancer type, it is well known the strong effect of family history of CRC on the occurrence of this disease [1, 6].

The concept of contextual phenomenon has a long history in epidemiology, and it is mentioned in different forms in Durkheim's concept of social fact [7], Rose's notion of population diseases rates [8], and John Snow's findings on cholera incidence [9]. This conception supports the idea that the distribution and determinants of population health is epistemologically multilevel and needs to consider people and context. From a statistical point of view, contextual phenomenon corresponds to clustering [10].

In order to study the role of dietary patterns in colorectal cancer risk assessment, we applied multilevel analysis, a methodological approach that considers the hierarchical structure of the data (individual characteristics nested in family characteristics) [11]. This methodological strategy has been recently applied by a growing number of epidemiologic studies for the investigation of causality in chronic diseases. It allows obtaining suitable epidemiological information about the importance of contextual characteristics (family, neighborhoods, etc.) for understanding disparities in health and health-related behavior [12]. It has been highlighted that the measuring of health variation from multilevel analysis is relevant not only for statistical reasons (improved estimation) but also for substantive epidemiologic reasons (quantification of the importance of the contextual factors for understanding individual health) [13].

The Argentinean population traditionally has a high consumption of animal protein and fats obtained mainly from red meat and a low intake of fiber; nevertheless, some comprehensive characterization of dietary pattern is still absence in the country.

Aims of this study were: (1) to identify dietary patterns in Córdoba, Argentina; (2) to estimate the effect of dietary patterns on CRC occurrence, and (3) to quantify the association between family history of CRC and CRC occurrence by studying the propensity CRC to occur from persons belonging to different clusters.

Thus, in this paper, special emphasis is given to the relation between epidemiological and methodological aspects through the construction of dietary patterns and the application of multilevel approach to estimate and interpret measures of variation in nutritional epidemiology.

Subjects and methods

Study design

A project named "Environmental Epidemiology of Cancer in Córdoba" (EECC) started in 2004 and includes several aspects of cancer epidemiology namely descriptive epidemiology from 1986 to 2006 by performing Age-Period-Cohort models [14–17], incidence and spatial distribution [18, 19], and case–control studies about dietary and other environmental exposures [20, 21] relationship with most frequent cancers identified in Argentina: breast, prostate, colon, and bladder.

This case–control study began in 2006 and ended with the last case recruited in 2010. Cases were 41 subjects with an incident, histologically confirmed, primary adenocarcinoma of the colon or rectum (C18-C20) [22], diagnosed in the previous 6 months, and identified by the Córdoba Tumor Registry. In the same time period, all subjects matched by sex, age, and place of residence identified through the census list were considered as possible candidates as controls. Controls were randomly chosen from census list and included only after an accurate verification about the absence of any neoplastic or related condition as well as diseases or other social, religious, or lifestyle conditions that generate a long-term modification of the diet. All participants (cases and controls) completed a structured questionnaire which provides information on personal history of diseases, included cancer. The study base was constituted by 41 cases and 95 controls (1:2 cases–controls ratio).

The area from which cases come from includes Cordoba City (Capital, 1,300,000 inhabitants), as well as some rural and urban counties which are representative from the whole population of Córdoba province [18].

The study was approved by the Ethical Committee of the Faculty of Medical Sciences, University of Córdoba, and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. All participants gave informed consent prior to their inclusion in the study.

Subject Information

All participants completed a structured questionnaire which included information about social-demographic characteristics, smoking, alcohol consumption, self-reported anthropometric characteristics, use of nonsteroidal anti-inflammatory drugs (NSAID), medical insurance, personal medical history, and family history of cancer (at least one first-degree relative with CRC). To assess dietary exposure, a validated FFQ was used [23]. This FFQ contains 127 items and includes physical activity items, which allow

classifying it into low, moderate, and intense categories. The FFQ was coupled with an also validated photographical atlas based on standard portion sizes in Argentina [24]. FFQ focused on the five-year period before the date of interview (diagnosis for cases). An at-home face-to-face interview was carried out by trained interviewers.

Statistical analysis

Dietary pattern identification method

Dietary patterns, as combinations of dietary components (food items, food groups, nutrients, or both), summarize the total diet or the key aspects of the diet for the population under study [5]. In the present work, characterization of dietary patterns was performed by pooling 373 controls from all ongoing case–control studies of the project as they come from the same population-time experience.

A principal component factor analysis (PCFA) was performed. Twelve representative food groups of Argentinean diet were chosen to assess their potential role in CRC risk: dairy foods, processed meat, red meat, fish and poultry, eggs, fruit and nonstarchy vegetables, starchy vegetables, grains products, added sugar and sweets, high-sugar drinks, wine, fats and oils. The factorability of the correlation matrix was evaluated through visual inspection of the matrix and statistical indicators such as Bartlett's test of sphericity and Kaiser–Meyer–Olkin (KMO), an index used to measure the appropriateness of factor analysis. KMO with values between 0.5 and 1.0 indicate factor analysis is appropriate [25]. Also, eigenvalue values, Akaike information (AIC), and Bayesian Information (BIC) were taken into account for parsimony and plausibility of the factors [2]. A varimax rotation was then applied to the factor loading matrix to facilitate interpretability of the factors. After, factor scores were calculated by applying the regression method. These scoring coefficients indicate the degree to which each subject's diet conforms to each one of the identified patterns [2]. Each factor was labeled by its dominants food groups, and only those item with absolute rotated factor loading $\geq$0.60 were considered. Finally, all participants of the CRC case–control study were categorized into tertiles (low, medium, and high) of each factor scores, based on the pooled distribution obtained from the controls.

As a second step, each pattern was correlated with some lifestyle and socio-demographic characteristics as well as nutritional variables, using direct and partial coefficients.

Risk estimates

We proposed a multilevel logistic regression (MLR) model for the binary response (equal to 1 for CRC occurrence, 0 for absence) in order to considerate that the individual probability of an outcome is dependent on both individual-level variables (dietary exposure) as well as contextual or group variables (family history characteristics) of the subjects [12]. Thus, the first-level variables included: dietary patterns, BMI, NSAID use, and medical insurance, whereas family history of CRC was considered as second-level or clustering variable. No covariate was included in this second hierarchy, then only one parameter was estimated: the variance. By means of this model, the presence of significant familiar aggregation and to which extent familiar level differences are explained by the individual characteristics, related to mainly dietary patterns, was probed. Multilevel logistic models were also used, unlike the classical logistic regression model, in order to avoid underestimating the standard error of the regression coefficient of aggregate risk factor, leading to overestimation of the significance of the risk factor [26]. This is an important aspect to consider mainly because of the small sample size of this work. There is agreement about that a small sample size at level two (meaning a sample of 10 clusters or less) leads to biased estimates of the second-level standard errors, but only a variance (family history of CC) and not regression coefficients is estimated in our work. Indeed, Maass and Hox [27] showed that the standard errors of the second-level variances are estimated about 15% smaller when the number of groups is substantially lower than 30 groups, resulting in a noncoverage rate of almost 8.9%, instead of 5%. This is clearly different from the nominal 5%, but in practice acceptable.

Steps of MLR estimation included: (1) an initial empty model that includes only random intercepts to explore familiar aggregation, such that only considers the variance at second level or between groups, without any effect of covariates; (2) an intermediate model which includes dietary patterns effects on CRC occurrence while taking into account the familiar aggregation (Model A) and a full or current model (Model B) which also includes dietary patterns effects and other characteristics, such as BMI (continuous), regular use of NSAID (yes/no), and medical insurance (no/yes). In the three models, the effects of all covariates were considered fixed and transformed to odds ratio (OR) for interpretation, while family history (random) effects were summarized into a variance component. AIC was used to select the appropriate model. This sequential strategy was performed to observe possible changes in the variance estimates (model A or B from empty model) and in the OR estimates for dietary patterns when other covariates are included in the model.

Intraclass correlation coefficient (ICC) of the outcome given the covariates was also calculated. ICC quantifies the second-level (between-group) heterogeneity and is expressed as $\mathrm{ICC} = \frac{\psi}{\psi+\pi^2/3}$, where ψ denotes the cluster variance.

Finally, the residual heterogeneity between cluster groups in terms of odds ratios was computed by the median odds ratio (MOR) [12] as $\text{MOR} = \exp\left(\sqrt{2\psi^2}\phi^{-1}(0.75)\right)$, with $\phi^{-1}(0.75)$ the 75th percentile of the standard normal distribution. By quantifying cluster heterogeneity, MOR quantifies and facilitates a direct comparison between covariate effects and the magnitude of heterogeneity in terms of well-known odds ratios [13]. The MOR considers all possible pairs of subjects with similar covariates (dietary patterns, BMI, use of NSAID, and medical insurance in this case) but with different values for the clustering variable, and so quantify how many times the individual odds of outcome occurrence increases (or decreases) when a subject moves from one clustering variable value to another one.

As cluster differences in CRC occurrence may be attributable not only to contextual influences (as some family characteristic) but also to differences in the individual composition of clusters in terms of BMI, use of NSAIDs, medical insurance and other unmeasured individual characteristics not considered in our study, another diagnostic measure, proportional change in variance (PCV) [28], was calculated. By adjusting for individual characteristics, some portion of the compositional differences that explain some of the cluster-level variance detected in the empty model are taken into account through this measure.

Stata 11.2 software (Statacorp LP. College Station, TX: USA) was used for all analysis.

Results

Identification of dietary patterns

Table 1 shows the results from principal components factor analysis. Both Bartlett's test for sphericity and overall estimate KMO values indicated that factor analysis was suitable for the dataset. As shown in Table 1, three major dietary patterns (factors) were identified and retained. These factors explained 46% of the total variance in the original dataset of food groups. The first factor was characterized by strong factor loadings on red meat, wine, and starchy vegetables (representing mainly potato intake). This factor was labeled "Southern Cone pattern". The second had the strongest factor loading on high-sugar drinks, named in consequence "High-sugar drinks pattern". Finally, the third factor, defined as "Prudent pattern", presented strongest factor loadings for fruit and nonstarchy vegetables group, and dairy foods. KMO measure of sampling adequacy ranged from 0.63 for starchy vegetables to 0.80 for high-sugar drinks.

Correlation between dietary patterns and selected lifestyle characteristics and nutritional variables evidenced statistically significant positive coefficients between *Southern Cone pattern* and intakes of total energy, proteins, carbohydrates, lipids, and vitamin E, as well as with number of cigarettes smoked and years smoked (Table 2). Also *High-sugar drinks pattern* was positive associated with total energy, proteins, carbohydrates, lipids, and vitamins A and C intakes but negatively with age, whereas the *Prudent pattern* was associated negatively with years smoked and positively with proteins, vitamins A, and fiber. Partial correlations between food patterns and the same variables as above, which estimate the correlation with each variable after removing the effect of all other variables, substantially confirmed the correlations (Table 2). *Southern Cone pattern* was positively partially correlated with age, number of cigarettes/day, total energy intake, and vitamin E and negatively correlated with proteins and carbohydrates. *High-sugar drinks pattern* was negatively partially correlated with age, total energy intake, proteins, and fiber but positively carbohydrates and vitamin A. *Prudent pattern* was positively associated with proteins, total fat, and fiber but negatively with total energy intake.

Case–control study

In this exploratory analysis, 136 subjects were considered (mean age 61.8, SE 1.2) of which 41 cases (mean age 61.1, SE 2.3) and 95 controls (mean age 62.0, SE 1.5); 53.7% of the cases were men. Male cases (mean age 66.9, 95%CI 62.6–71.2) were older than female cases (mean age 54.3, 95%CI 46.8–61.8). Table 3 shows the distribution of cases and controls according to variables of interest. Cases tended to be more concentrated in medium to high categories of *Southern Cone pattern* (87.8%) and *High-Sugar Drinks pattern* (87.9%); the same figures for controls were 74.8 and 66.3%, respectively. As regards the *Prudent pattern,* 48.8 and 62.1% of cases and controls, respectively, had a medium or high intake. A higher proportion of cases than controls presented obesity (34.2 vs. 28.4%, respectively), reported no using aspirin or NSAID (75.6 vs. 64.2%), had no medical insurance (41.5 vs. 14.7%), and had a history of CRC in their family (17.1 vs. 3.2%).

Table 4 shows results from MRL. Variance of the random effect of family history of CRC clustering variable was highly significant in all models. Model A showed significant promoting effects of *Southern Cone* and *High-sugar drinks dietary* patterns on CRC, whereas *Prudent* pattern showed a significant protective effect. Model B evidenced the same exposure effects although less strong (except for the III tertile of *Prudent* pattern) as well as significant effects of BMI (promoting) and No Use of NSAIDs (preventing).

Table 1 Rotated factor loading matrix, explained variances, and diagnostic measures for the three major dietary patterns identified by factor analysis among controls ($n = 373$) in the EECC Study, Córdoba, Argentina, 2006–2010

Food groups*	Southern Cone pattern Red meat, wine and potatoes	High-sugar drinks pattern High-sugar drinks	Prudent pattern Fruit and vegetable, dairy
Dairy foods	−0.02	0.01	**0.69***
Eggs	0.53	−0.04	0.26
Red meat	**0.72****	0.12	0.03
Processed meat	0.56	0.28	−0.02
Fish and poultry	0.09	0.39	0.36
Fruit and nonstarchy vegetables	0.06	−0.05	**0.76****
Starchy vegetables	**0.63***	0.16	0.21
Cereal products	0.29	0.42	0.07
Fats and oils	0.54	0.35	0.24
Added sugar and sweets	0.24	0.57	0.10
High-sugar drinks	−0.05	**0.80*****	−0.10
Wine	**0.69***	−0.36	−0.21
Proportion of explained variances (%)	20.23	13.90	11.89
Cumulative explained variances (%)	20.23	34.13	46.02

Food groups with absolute rotated factor loading ≥0.60, considered as dominant food groups for each factor, were shown in bold typeface

* Barlett's test for Sphericity: χ^2 584.7, $p < 0.0001$; Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy: 0.75 (middling)

KMO: * mediocre, ** middling, *** meritorious

Table 2 Pearson and partial correlations between selected variables and dietary pattern ($n = 373$) in the EECC Study, Córdoba, Argentina, 2006–2010

Variable	Correlations			Partial correlations		
	Southern cone	High-sugar drink	Prudent	Southern cone	High-sugar drink	Prudent
Age (years)	0.12	−0.31*	−0.07	0.28*	−0.24*	−0.01
Education	0.00	−0.10	0.01	0.14	−0.12	−0.03
BMI	0.16	0.12	−0.02	−0.01	0.16	0.02
Years Smoked	0.39*	−0.15	−0.34*	0.00	0.01	−0.20
Cigarettes/day	0.37*	0.21	−0.17	0.25*	−0.06	−0.11
Total Energy	0.67*	0.49*	0.13	0.49*	−0.38*	−0.42*
Protein	0.53*	0.46*	0.34*	−0.27*	−0.27*	0.39*
Carbohydrates	0.33*	0.67*	0.25	−0.43*	0.69*	0.09
Total Fat	0.71*	0.37*	0.18	0.12	0.19	0.28*
Vitamin A	0.01	0.28*	0.49*	−0.19	0.45*	0.18
Vitamin C	0.07	0.52*	0.26	−0.12	0.41*	0.06
Vitamin E	0.35*	0.02	0.25	0.27*	−0.17	0.12
Fiber	0.19	0.24	0.61*	0.13	−0.67*	0.46*

* $p < 0.05$

ICC was 0.451, 0.285, and 0.375 for the empty, A, and B models, respectively, meaning that roughly 38% of the variance was attributable to familiar aggregation. MOR was 2.98 and 3.82 for models A and B, respectively, whereas PVC was 51.57 and 26.91% for models A and B, respectively.

Discussion

These preliminary results show that the Southern Cone (red meat, starchy vegetables, wine) and High-sugar drinks dietary patterns have a promoting effect on CRC occurrence but a Prudent pattern (dairy foods, fruit and nonstarchy vegetables) has a preventing one. BMI was positively associated, whereas use of NSAIDS (vs. no use) and to have medical insurance (vs. not having) were negatively associated with CRC. Moreover, there was a significant family history of CRC's clustering effect.

Some methodological issues concerning the study design need to be considered. In order to reduce the amount of possible exposure misclassifications, absence of neoplastic conditions or related diseases, or long-term modifications of diet was incorporated as criteria of eligibility for controls; but the error in measuring the level of dietary exposure may still be present. Given that the most common bias one would expect would be to give socially desirable answers, this would lead to an underestimation of the

Table 3 Cases and control characteristics according to selected variables

Characteristics	Cases Subjects (%)	Controls Subjects (%)
Dietary patterns (tertiles)[a]		
Southern Cone pattern		
I	5 (12.2)	24 (25.3)
II	19 (46.3)	30 (31.6)
III	17 (41.5)	41 (43.2)
High-sugar drinks pattern		
I	5 (12.2)	32 (33.7)
II	14 (34.2)	30 (31.6)
III	22 (53.7)	33 (34.7)
Prudent pattern		
I	21 (51.2)	36 (37.9)
II	14 (34.2)	27 (28.4)
III	6 (14.6)	32 (33.7)
BMI		
<18.5	2 (4.9)	1 (1.1)
18.5–24.9	16 (39.0)	31 (32.6)
25–29.9	9 (22.0)	36 (37.9)
≥30.09	14 (34.2)	27 (28.4)
Regular use of NSAID		
Yes	10 (24.4)	34 (35.8)
No	31 (75.6)	61 (64.2)
Medical insurance		
Yes	24 (58.5)	81 (85.3)
No	17 (41.5)	14 (14.7)
Family history of CRC		
No	34 (82.9)	92 (96.8)
Yes	7 (17.1)	3 (3.2)

$n = 373$ in the EECC Study, Córdoba, Argentina, 2006–2010

NSAID nonsteroidal anti-inflammatory drugs, *CRC* colorectal cancer

[a] Tertiles of each factor scores (pattern) are based on the distribution of healthy subjects that participated of the EECC Study

exposure. This misclassification would be no differential, and the bias probably introduced is toward the null. Our FFQ has a satisfactory level of validity and reproducibility, although reproducibility measurements tended to be better than validity ones.

In this sense, the degree of agreement with respect to classification into categories is not devoid from interpretational difficulties due to that the division into percentiles strongly depends on the distribution of real intake levels in the study population and the ability of the FFQ to correctly classify subjects. Overall, agreement measurements of the FFQ were as good or better for most nutrients considering both validity and reproducibility issues [24].

Recall bias also needs to be considered. As in other case–control studies, this study may suffer from this bias in dietary estimation, probably increased for the time-frame of the FFQ. Nevertheless, the reproducibility of this 5-year FFQ has been probed by Navarro et al. [24], who concluded that this questionnaire can provide an acceptable assessment of long-term dietary intake in the Argentine population and in other Latin-American areas in epidemiological cancer studies. It is possible, however, that error in recall occurred, but in both cases and controls, or that the referent period of the study did not capture the relevant etiologic time of dietary exposure.

Moreover, the reproducibility of FFQs has been examined by Willet [29] who found that, in general, the studies show correlations ranged from 0.5 to 0.7 for nutrients intakes measured at periods of 1–10 years, which indicate a acceptable level of reproducibility. Finally, residual confounding may be present, as we are not sure whether the study base was selected in such a way as to include covariant factors that do not occur in the general population and whether the sample size is sufficient to detect a reliable effect of dietary pattern exposures and of family history of CRC clustering on CRC occurrence.

The research methodology "top-down" as opposed to "down-top" has been proposed to study health effects of the combinations of single nutrients in foods and food patterns [30]. This approach has been applied to the relationship between dietary factors and CRC since the pioneer works of Randall et al. [31] and Slattery et al. [32]. Factor analysis has an intuitive appeal in nutritional epidemiology as it offers a means of factoring intakes of a variety of foods reflecting underlying eating patterns. However, a number of issues must be addressed when considering factor analysis in nutritional epidemiologic studies. The first issue is heterogeneity in the data. A criterion of 0.60 as the lower limit for a meaningful factor loading was used in this study. Similar studies from this [33] and others areas of the world [3, 34] have used similar cut-off values. With low loadings, one variable can flip from one factor to another. Three factors were extracted in this study; these factors accounted for 46.0% of the cumulative proportion of total variance which is a reasonably good proportion of the between-person variance in the diet [32, 33]. The most of exploratory works in this issues report cumulative variance below 30% [4].

Another concern regards reproducibility of results. This is particularly important with CRC as some inconsistencies have been reported [35]. However, our results are like those of studies conducted in Uruguay [33, 36] whose major ethnic streams and diet are similar to those Argentinean, although recently the existence of country specific patterns has been emphasized [37].

By performing principal components factor analysis, three dietary patterns were identified. De Stefani E et al. reported, in an area similar to Córdoba from the geographical and lifestyle point of view, four dietary patterns

Table 4 CRC risk estimates obtained from multilevel logistic modeling on tertiles of the scored dietary patterns and other individual-level covariates

	Empty model	Models with individual-level variables	
	(AIC:167.019)	Model A (AIC: 163.590)	Model B (AIC: 162.996)
Measures of association (OR, 95% CI)			
Dietary patterns (vs. tertile I)			
Southern Cone, tertile II		3.086 (1.094; 8.715)*	1.920 (1.174; 3.139)*
Southern Cone, tertile III		2.451 (1.454; 4.130)**	1.484 (1.006; 2.980)*
High-sugar drinks, tertile II		3.977 (2.972; 5.323)***	4.438 (2.298; 8.571)***
High-sugar drinks, tertile III		4.107 (2.068; 8.155)***	4.008 (2.127; 7.553)***
Prudent, tertile II		0.665 (0.612; 0.722)***	0.474 (0.434; 0.518)***
Prudent, tertile III		0.277 (0.176; 0.436)***	0.159 (0 .153; 0.164)***
BMI (continuous)			1.060 (1.050; 1.070)***
Use of NSAIDs (vs. no using)			0.542 (0.432; 0.679)***
Medical insurance (vs. no)			0.337 (0.320; 0.354)***
Measures of variation or clustering			
Family history of CC variance (SE)	2.705 (0.042)	1.310 (0.288)	1.669 (0.376)
ICC[†]	0.451	0.285	0.337
MOR[‡]	4.801	2.980	3.430
PCV[§]	Reference	51.571%	38.266%

AIC Akaike information criterion, *OR* odds ratio, *NSAID* nonsteroidal anti-inflammatory drugs, *CC* colorectal cancer, *SE* standard error, *ICC* intraclass correlation, *MOR* median odds ratio, *PCV* proportional change in variance

[†] ICC computed by latent variable method. [‡] MOR quantifies cluster heterogeneity in terms of odds ratios. [§] The PCV expresses the change in the cluster variance between the empty model and the model (A or B) with individual-level variables

* $p < 0.05$, ** $p < 0.01$ or *** $p < 0.001$ as levels of significance

[33] in a multisite cancer, hospital-based case–control study.

The *Southern Cone* pattern, basically composed by red meat, wine, and potatoes, explained 20% of the total variance in the analysis of foods. As reported in other studies, this type of diet loaded mainly on red meat has been implicated in colon and rectal cancer [38]. Although our *Southern Cone* pattern is different from *Western pattern* identified by other studies, for which there is no specific definition, it is clear that red meat, processed meat, and eggs are always present. As regards meat intake, its cooking method is an important variable when considering risk for CRC. Barbecued red meat is the most prevalent in the Argentinean population, where there is traditionally a high consumption of animal protein and fats obtained mainly from red meat and a low intake of fiber. In fact, in previous studies in Córdoba, it was reported that consumption of meat and meat products is unusually high (280 g/d) and contributes 65% protein, 71% dietary lipids, and 80% saturated fatty acids [39]. Reported results in this area showed that increased risk of CRC seems to be related to cooking temperature and close food contact with the heating source, as higher risks were observed for heavily browned surfaces when meats were barbecued or iron-pan cooked [39]. Red meat, usually fatty (30–33% of total lipids), is the one preferred by Argentinean population, as is barbecuing or iron-pan cooking methods. Moreover, in Argentina, red meat consumption is frequently accompanied by wine intake about 5–6 times a week. In our results, the fourth pattern (Drinker) identified by De Stefani et al. is perfectly integrated into our first factor, Southern Cone pattern, as wine is a strong loading factor. Alcohol consumption (including wine and beer) is around 181 g/person/day in Argentina (period 2001–2003) [40]. It is well known that wine is the most common alcoholic drink in this population [41, 42]. In fact, this country has traditionally been among those with the highest per capita production of wine [43] and, in consequence, shows high levels of consumption. A causal link has been established between alcohol consumption and several cancers, including colon and rectum [44]. The evidence that consumption of more than 30 g/day of ethanol from alcoholic drinks is a cause of colorectal cancer in men is convincing, and probably also in women [1]. A meta-analysis of published cohort studies between 1990 and 2005 on the relationship between alcohol intake and colorectal cancer reported a 15% increase in risk for an increase of 100 g of alcohol intake per week [45].

A high risk of CRC cancer was evident for increasing tertiles of High-sugar drinks pattern. A similar pattern was

found by Slattery et al. [32], and as in our case, this pattern loaded only drinks rich in sugar. This pattern has not been identified in other areas of South America [47]. Argentinean intake of high-sugar drinks is high as evidenced by FAO [40]. The increased risk of CRC found in this study is consistent with other findings which evidenced low levels of deglycating enzymes in colorectal cancer patients [48] and the increased risk of colorectal adenoma with the level of fructosamine, an indicator of the level of glucose in the blood more sensitive to foods with a high glycemic index [49].

Prudent pattern has emerged repeatedly in studies using factor analysis to study dietary patterns in North America [50], Europe [51, 52], Asia [53], and South America [47]. All findings are consistent across the studies displaying lower risk for higher centiles of the pattern. This may be linked to the strong evidence that suggest that foods containing dietary fiber, such as fruits and nonstarchy vegetables that conform this pattern, probably protect against CRC [1]. This is consistent with the outcome from the correlation analysis carried out, which found positive association between this pattern and fiber. However, we cannot exclude a role for other bioactive compounds in fruit and vegetables highly correlated with dietary fiber that also are related to a lower risk of colon cancer as β-carotene or lycopene [54]. Dietary patterns, along with other characteristics such as physical activity, overweight-obesity, and smoke habits have been long suspected to contribute to higher rates of many chronic diseases [55]. Particularly in CRC, there is abundant and consistent epidemiological evidence showing lower risk of CRC with higher levels of physical activity and increased risk with increased body fatness [1]. In addition, the literature suggests that other individual variables such as fiber intake or smoking habit have a potential effect on CRC risk. The probable protective role of fiber is linked to plausible mechanisms such as the dilution of fecal content, the decrease in transit time, and their effect on gut fermentation [1]. Smoking, in turn, is also an established risk factor for many forms of cancer, and there is evidence that supports the addition of CRC to the list of tobacco-associated malignancies [56]. However, in this anatomic site, the link between smoking and cancer needs additional study, considering, for example, the adjustment of estimates for alcohol, other relevant risk factor for CCR that is also strongly correlated with cigarette smoking [57].

In our study, the Southern Cone pattern was associated with other lifestyle characteristics such as daily number of cigarettes and years smoked as well as a higher caloric intake, whereas Prudent pattern was associated with a younger age, a lower number of cigarettes and years smoked, lower caloric intake and higher intake of dietary fiber, as was mentioned above. As other authors have argued the ability of factor analysis to identify dietary pattern and associate them with other lifestyle characteristics help to feel more confident that meaningful eating patterns have been identified [32]. We focused our attention on dietary patterns and CRC as several studies in this area [39, 46] have yet analyzed the effect of and/or nutrients but no attempt had been made to evaluate dietary patterns, which represent a broader picture of dietary consumption and may be more associated with risk than any individual food or nutrient.

We used a multilevel or hierarchical model in order to decompose the total variance between responses into variance components, specifically the between-cluster variance ψ and the within-cluster variance θ. This was accomplished by specifying a model that included random intercepts and induced correlations among responses for subjects in the same cluster (family history of CRC), presented as the ICC. ICC can be thought of a measure of absolute agreement or reliability [58] of the total variance decomposition, and the significant value here obtained (0.378) implies that the proportion coming from familiar aggregation is essential for obtaining odds ratios accurateness.

Also, we fit MLR models in three steps, from empty to model B and calculated the PCV measure to quantify the change in the cluster variance between the empty model and each model (A or B) due to the inclusion of individual-level variables.

Our results indicated that there was variation between family history of CRC groups. Thus, for two persons with the same individual-level covariates (dietary patterns, BMI, use of NSAID, medical insurance) but belonging to with or without family history of CRC groups, the risk of CRC increases, in median, 3.43 times. In other words, MOR shows us the extent to which the individual probability (not population-average probability) of having CRC is determined by the family history of the disease [13].

Finally, as regards of modeling approach, this work defines subject-specific models which are differentiated from marginal- or population-averaged models by the inclusion of parameters specific to the subject (cluster here) [59]. This allows describing different types of covariates effect on the expectations, for example, for binary responses (our case), the effect of covariates on the probabilities is conditional upon the level of the subject (cluster)-specific effect. In our case, a unit change in the covariate (BMI, for instance) translates into an appropriate change in probability, keeping the level of the family history of CRC-specific effect fixed. Some covariates were included in the model in order to account for residual confounding and to obtain more accurate estimates of dietary patterns which constitute the objective of this paper. BMI in particular was included not only for its known association with the cancers in general and particularly with CRC.

Summarizing, we conducted a factor analysis on foods, identified three recognized dietary patterns, and explored their relationship with CRC in Córdoba, Argentina. In the study, the scores of foods patterns were associated with an increased risk for the *Southern Cone* and *High-sugar drinks* patterns, whereas the *Prudent* showed a strong protective effect. Further studies could reassure the validity of our findings. In particular, confirmatory factor analysis could clarify the validity of the dietary patterns found in the present study.

Acknowledgments We are especially grateful to the Córdoba Cancer Registry, Ministry of Health of Córdoba (Argentina), for allowing us to use the database and for their assistance with all the data-management steps, also to the physicians and all who participated in this study. We would like to thank the National Scientific and Technical Research Council (CONICET) for the SAP's fellowship.

This research was partially supported by the Science and Technology National Agency, FONCyT grant PICT 2008-1814, PICT-O 2005-36035, and Science and Technical Secretary of the University of Córdoba (SECyT-UNC) grant 05/H207.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. World Cancer Research Fund, American Institute for Cancer Research (eds) (2007) Food, nutrition, physical activity, and the prevention of cancer: a global perspective. AICR, Washington
2. Edefonti V, Decarli A, La Vecchia C et al (2008) Nutrient dietary patterns and the risk of breast and ovarian cancers. Int J Cancer 122(3):609–613
3. Edefonti V, Bravi F, Garavello W et al (2010) Nutrient-based dietary patterns and laryngeal cancer: evidence from an exploratory factor analysis. Cancer Epidemiol Biomarkers Prev 19(1):18–27
4. Edefonti V, Randi G, La Vecchia C et al (2009) Dietary patterns and breast cancer: a review with focus on methodological issues. Nutr Rev 67(6):297–314
5. Randi G, Edefonti V, Ferraroni M et al (2010) Dietary patterns and the risk of colorectal cancer and adenomas. Nutr Rev 68(7):389–408
6. Camp NJ, Slattery ML (2002) Classification tree analysis: a statistical tool to investigate risk factor interactions with an example for colon cancer (United States). Cancer Causes Control 13(9):813–823
7. Durkheim E (1964) The rules of sociological method, 8th edn. Free Press of Glencoe, New York
8. Rose GA (1992) The strategy of preventive medicine. Oxford University Press, Oxford
9. Snow J (1936) Snow on cholera. A reprint of two papers by John Snoe, MD, together with a biographical memoir by BW Richardson, MD, and an introduction by Wade Hampton Frost. The Commonwealth Fund, New York
10. Diez-Roux AV (2000) Multilevel analysis in public health research. Annu Rev Public Health 21:171–192
11. Rabe-Hesketh S, Skrondal A, Pickles J (2001) Generalized multilevel structural equation modelling. Psychometrika 69(2): 167–190
12. Merlo J, Chaix B, Ohlsson H et al (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health 60(4):290–297
13. Larsen K, Merlo J (2005) Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. Am J Epidemiol 161(1):81–88
14. Pou SA, Osella AR, Eynard AR et al (2010) Cancer mortality in Córdoba, Argentina, 1986–2006: an age-period-cohort analysis. Tumori 96(2):202–212
15. Pou SA, Osella AR, Eynard AR et al (2009) Colorectal cancer mortality trends in Córdoba, Argentina. Cancer Epidemiol 33:406–412
16. Niclis C, Del Pilar Díaz M, La Vecchia C (2010) Breast cancer mortality trends and patterns in Córdoba, Argentina in the period 1986–2006. Eur J Cancer Prev 19(2):94–99
17. Niclis C, Pou SA, Bengió RH et al (2011) Prostate cancer mortality trends in Argentina 1986–2006: an age-period-cohort and joinpoint analysis. Cad Saude Publica 27(1):123–130
18. MdelP Díaz, Osella AR, Aballay LR et al (2009) Cancer incidence pattern in Córdoba, Argentina. Eur J Cancer Prev 18(4):259–266
19. Díaz MP, Corrente JE, Osella AR et al (2010) Modeling spatial distribution of cancer incidence in Cordoba, Argentina. Appl Cancer Res 30(2):245–252
20. Pou SA, Osella AR, Díaz MP (2011) Bladder cancer mortality trends and patterns in Córdoba, Argentina (1986–2006). Cancer Causes Control 22(3):407–415
21. Francisca FM, Carro Perez ME (2009) Assessment of natural arsenic in groundwater in Cordoba Province, Argentina. Environ Geochem Health 31(6):673–682
22. World Health Organization (1992) International classification of disease: 10th revision, 2nd edn. World Health Organization, Geneva
23. Navarro A, Cristaldo PE, Díaz MP et al (2000) Atlas fotográfico para cuantificar el consumo de alimentos y nutrientes en estudios nutricionales epidemiológicos en Córdoba, Argentina. Rev Fac Cienc Méd Córdoba 57(1):67–74
24. Navarro A, Osella AR, Guerra V et al (2001) Reproducibility and validity of a food-frequency questionnaire in assessing dietary intakes and food habits in epidemiological cancer studies in Argentina. J Exp Clin Cancer Res 20(3):203–208
25. Bertuccio P, Edefonti V, Bravi F et al (2009) Nutrient dietary patterns and gastric cancer risk in Italy. Cancer Epidemiol Biomarkers Prev 18(11):2882–2886
26. Blakely T, Woodward A (2000) Ecological effects in multi-level studies. J Epidemiol Community Health 54:367–374
27. Maas CLM, Hox JJ (2005) Sufficient sample sizes for multilevel modeling. Methodology 1(3):86–92
28. Merlo J, Yang M, Chaix B et al (2005) A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. J Epidemiol Community Health 59(9):729–736
29. Willet W (1998) Nutritional epidemiology, 2nd edn. Oxford University Press, Oxford
30. Jacobs DR, Steffen LM (2003) Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. Am J Clin Nutr 78(suppl):508S–513S
31. Randall E, Marshall JR, Brasure J et al (1992) Dietary patterns and colon cancer in Western New York. Nutr Cancer 18:265–276
32. Slattery ML, Boucher KM, Caan BJ et al (1998) Eating patterns and risk of colon cancer. Am J Epidemiol 148:4–16
33. De Stefani E, Deneo-Pellegrini H, Boffetta P et al (2009) Dietary patterns and risk of cancer: a factor analysis in Uruguay. Int J Cancer 124:1391–1397

34. Bertuccio P, Edefonti V, Bravi F et al (2009) Nutrien dietary patterns and gastric cancer risk in Italy. Cancer Epidemiol Biomarkers Prev 18(11):2882–2886
35. Witte JS, Greenland S, Haile RW et al (1994) Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology 5:612–621
36. De Stefani E, Ronco AL, Deneo-Pellegrini H et al (2010) Dietary patterns and risk of advanced prostate cancer: a principal component analysis in Uruguay. Cancer Causes Control 21:1009–1016
37. Balder HF, Virtanen M, Brants HA et al (2003) Common and country-specific dietary patterns in four European cohort studies. J Nutr 133:4246–4251
38. Donaldson MS (2004) Nutrition and cancer: a review of the evidence for an anti-cancer diet. Nutr J 3:19
39. Navarro A, Muñoz SE, Lantieri MJ et al (2004) Meat cooking habits and risk of colorectal cancer in Cordoba, Argentina. Nutrition 20(10):873–877
40. Food and Agriculture Organization (FAO) Statistics Division (2009) FAO food balance sheets. http://www.fao.org/statistics/faostat/foodsecurity/. Accessed 24 Feb 2009
41. Muñoz SE, Navarro A, Lantieri MJ et al (1998) Alcohol, methylxanthine-containing beverages, and colorectal cancer in Córdoba, Argentina. Eur J Cancer Prev 7:207–213
42. Castelletto R, Castellsague X, Muñoz N et al (1994) Alcohol, tobacco, diet, mate drinking, and esophageal cancer in Argentina. Cancer Epidemiol Biomarkers Prev 3(7):557–564
43. Munné MI (2005) Alcohol and the economic crisis in Argentina: recent findings. Addiction 100(12):1790–1799
44. Boffetta P, Hashibe M (2006) Alcohol and cancer. Lancet Oncol 7(2):149–156
45. Moskal A, Norat T, Ferrari P et al (2007) Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. Int J Cancer 120(3):664–671
46. Navarro A, Diaz MP, Muñoz SE et al (2003) Characterization of meat consumption and risk of colorectal cancer in Cordoba, Argentina. Nutrition 19(1):7–10
47. De Stefani E, Deneo-Pellegrini H, Ronco AL et al (2011) Dietary patterns and risk of colorectal cancer: a factor analysis in Uruguay. Asian Pac J Cancer Prev 12(3):753–759
48. Notarnicola M, Caruso MG, Tutino V et al (2011) Low red blood cell levels of deglycating enzymes in colorectal cancer patients. World J Gastroenterol 17(3):329–333
49. Misciagna G, De Michele G, Guerra V et al (2004) Serum fructosamine and colorectal adenomas. Eur J Epidemiol 19(5):425–432
50. Flood A, Rastogi T, Wirfält E et al (2008) Dietary patterns as identified by factor analysis and colorectal cancer among middle-aged Americans. J Clin Nutr 88:176–184
51. Dixon LB, Balder HF, Virtanen MJ et al (2004) Dietary patterns associated with colon and rectal cancer: results from the dietary patterns and cancer (DIETSCAN) project. Am J Clin Nutr 80:1003–1011
52. Magalhães B, Bastos J, Lunet N (2011) Dietary patterns and colorectal cancer: a case-control study from Portugal. Eur J Cancer Prev [Epub ahead of print]
53. Kim MK, Sasaki S, Otani T et al (2005) Dietary patterns and subsequent colorectal cancer risk by subsite: a prospective cohort study. Int J Cancer 115:790–798
54. Van Duijnhoven FJ, Bueno-De-Mesquita HB, Ferrari P et al (2009) Fruit, vegetables, and colorectal cancer risk: the European prospective Investigation into cancer and nutrition. Am J Clin Nutr 89(5):1441–1452
55. Adlercreutz H (1990) Western diet and western diseases: some hormonal and biochemical mechanisms and associations. Scand J Clin Lab Invest Suppl 201:3–23
56. Giovannucci E (2001) An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 10(7):725–731
57. Gandini S, Botteri E, Iodice S et al (2008) Tobacco smoking and cancer: a meta-analysis. Int J Cancer 122(1):155–164
58. Rabe-Hesketh S, Skrondal A (2008) Multilevel and longitudinal modeling using stata, 2nd edn. Stata Press, College Station
59. Molenberghs G, Verbeke G (2005) Models for discrete longitudinal data. Springer, NY