Original Research Paper

Volume-7 | Issue-12 | December-2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 86.18

Epidemiology

A FRAMEWORK TO ADDRESS POTENTIAL BIAS IN COLORECTAL CANCER: ITS IMPLEMENTATION ON A NUTRITIONAL EPIDEMIOLOGIC STUDY IN ARGENTINA

Julia Becaria Coquet	Instituto de Investigaciones en Ciencias de la Salud (INICSA-UNC-CONICET), Universidad Nacional de Córdoba (UNC). Escuela de Nutrición, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Avenida Enrique Barros s/n, Ciudad Universitaria, CP 5,000, Córdoba, Argentina.
Sonia E. Muñoz	Instituto de Investigaciones en Ciencias de la Salud (INICSA-UNC-CONICET), Universidad Nacional de Córdoba (UNC).
María del Pilar Díaz*	Unidad de Bioestadística. Escuela de Nutrición, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Avenida Enrique Barros s/n, Ciudad Universitaria, CP 5,000, Córdoba, Argentina. *Corresponding Author

ABSTRACT Colorectal cancer (CRC) is the third most incident cancer in Argentina and diet is widely recognized for being associated with CRC. The objective of this work was to construct a methodological framework to quantitatively assess systematic errors in a case-control study, carried out in adult population of Córdoba province (2010-2016).

A CRC case-control study (n=490, 161/329 cases/controls) was conducted. Confounding was analyzed applying regression models approach for observed and unobserved variables. Selection bias was investigated utilizing deterministic scenarios, assigning *a priori* participation probabilities. Information bias, derived from missing data in covariates, was handled applying multiple imputation by chained equations (MICE), considering the missing at random mechanism.

Confounding effects of sex and socioeconomic status were found. In respect of selection bias, differences regarding conventional estimates were negligible. MI approach showed a promoting effect of the South Cone dietary pattern. Significant associations with sex, BMI, family history of CRC and socioeconomic status were observed.

KEYWORDS : sensitivity analysis; case-control study; dietary pattern.

Introduction

Colorectal cancer (CRC) is the third most frequent cancer in men (representing 10% of the total) and the second in women (9.2% of total) worldwide. The mortality of this cancer is low (8.5% of total) and most of these deaths (52%) happened in less developed countries, reflecting a worst survival in these regions (International Agency for Research on Cancer [IARC], 2012). In Argentina, CRC is the third most incident cancer, having around 12% of total cancer incidence 13.2% (IARC, 2012; Sierra & Forman, 2016). Regarding age standardized incidence rates, Córdoba province reported 12.45 and 19.43 (x100,000 inhab/year) for women and men respectively. CRC incidence rates were stated in third place for both groups –(Díaz, Corrente, Osella, & Muñoz, 2010).

In Argentina, there is extensive evidence related to the association between socio-cultural and biological risk factors and most incident cancers, such as CCR –(Pou et al., 2014). Diet is a recognized modifiable factor associated with CRC. Argentinean traditional diet is characterized by a high consumption of animal protein and lipids (obtained mainly from cow meat), low consumptions of fish, fruits and vegetables –(Alicia Navarro et al., 2004; Pou et al., 2014). Associations between diet and CRC risk have also been published —(Becaria Coquet et al., 2014; Pou et al., 2014; Pou, Díaz, & Osella, 2012), by using the dietary pattern approach to summarize one measure of diet.

Several epidemiologic study designs can be adopted to collect information to address the diet-cancer relationship. To identify risk factors associated with long induction period diseases like cancer (Breslow & Day, 1980), case-control studies are one of the most used in Latin America's region and in Argentina. These require an adequate planning to avoid bias and to obtain valid and reliable risk estimates of the effect of the dietary consumption and other exposures (or risk factors), on the occurrence of diseases. Assessing bias is equivalent to estimate the systematic error that remains after implementing the study design and the corresponding analysis (Lash, Fox, & Fink, 2009). This analysis is called a sensitivity analysis and constitutes the method used to determine the robustness of an evaluation assessing if the results are affected by changes in the methods, models, values of unmeasured variables or some assumptions (Schneeweiss, 2006). These types of errors can affect the validity of the results.

Systematic errors can be classified in three groups depending the moment or stage of the study that are originated (Rothman, Kenneth J., Greenland, Sander, & Lash, Timothy L., 1998). Confounding, arising due to the impossibility of randomly assign the exposure in observational studies, is one of them (Szklo, Moyses & Nieto, Javier, 2007). Selection bias is another group of systematic errors and usually happens when the exposure and outcome affect the participation of subjects in the study. Case-control studies are particularly vulnerable to this bias, since cases and controls are frequently selected conditioning on the presence or absence of a health event. The third group refers to information bias, which arises when the study groups are assessed. Errors can arise when exposure, outcome or any covariate are measured. The presence of these systematic errors can lead to bias and/or inefficient estimates of parameters and biased standards errors. The aim of present study was to construct a methodological framework to quantitatively assess systematic errors in a case-control epidemiologic study of CRC, carried out in adult population of Argentina.

Material and Methods

Study Design and Data

Data come from a CRC case-control study conducted in adult population of Córdoba province (center of Argentina). One hundred and sixty-one cases under 85 years old with a histopathologically confirmed incident primary diagnosis of colorectal cancer (ICIE10:C18-20) have been enrolled between 2010 and 2016 (identified by the Córdoba Tumor Registry). In the same time, 329 controls were randomly chosen, matched by age (± 5 years) and place of residence with cases. All of them gave their informed consent and ethics approval was obtained (RePIS 044/10). Data were collected by using a structured questionnaire including auto reported information (sociodemographic, anthropometric characteristics, physical activity, smoking habits, family and personal disease history and dietary habits, for more details see —Pou et al., 2014, 2012). Data on diet 5 years (Ambrosini, Fritschi, de Klerk, Mackerras, & Leavy, 2008) before interview (for controls) or diagnosis (for cases) was obtained using a food frequency questionnaire (FFQ) and a photographic atlas, both validated (Navarro et al., 2001, 2007).The dietary exposure was the assigned score of dietary patterns adherence identified in previous studies –(Pou et al., 2014).

Sensitivity Analysis Approach

Confounding

The presence of confounding can result in an over or underestimation of the real association studied –(Hernández-Avila, Garrido, & Salazar-Martínez, 2000). Some situations can take place when facing confounding. One may be when the confounding variable is known and registered (i.e. observed). Other, when the confounding variable has not been measured (Lash et al., 2009). For the former situation, two methodologies are used, being regression models the most used (Hammer, du Prel, & Blettner, 2009). Another strategy is to propose different models with the possible confounders as predictors and evaluate graphically how the magnitudes of the risk estimates are modified, when confounders included are combined. This allows observing the direction and magnitude of the risk estimate studied and analysis the statistical significance of the model (Draper & Smith, 1998; Wang, 2007). This method can be applied when the exposure is a continuous variable. Age, sex, socioeconomic status, body mass index (BMI), physical activity and family history of CRC were the confounding variables proposed in the present work.

Sometimes, there are situations where the researcher does not have any information of the subjects under study. In these cases, it is feasible to consider these unregistered variables as possible confounders. The methodology for this situation proposes to explore how the association between exposure and outcome would have been if this variable was registered, and adjusted for, in the model (Lash et al., 2009). The basic idea of this method is to propose different *a priori* distributions for the unregistered confounder and to observe the impact that the variable would have had on the obtained risk effects (Buis, 2010). In addition, some *a priori* association coefficients between the unobserved variable and some observed variables must be proposed, assuming that they could have similar distributions. The hypothesized coefficients respects to an observed variable (BMI) in present study were 0.2; 0.3; 0.4; 0.5.

Selection Bias

To address selection bias a deterministic analysis method was carried out, *a priori* plausible distributions of probability were assigned to the different groups (exposed cases, unexposed cases, exposed control and unexposed control). This sensitivity analysis was based on Monte Carlo simulations (See Orsini, et al., 2008 for more details). In our study, four scenarios were proposed for the analysis, one was non-differential and the other three were differential regarding the probabilities distributions of group participation. The *a priori* distributions were trapezoidal (see Table 1).

Table 1. Assigned trapezoidal probabilities of participation for deterministic selection bias analysis for colorectal cancer case-control study, according to groups of study.

	Study groups			
	Exposed cases	Unexposed cases	Exposed control	Unexposed control
Non differential scenario	0.7;0.75; 0.85; 0.9	0.7;0.75; 0.85; 0.9	0.7;0.75; 0.85; 0.9	0.7;0.75; 0.85; 0.9
Differential scenarios				
Scenarios 1	0.7;0.75; 0.85; 0.9	0.35;0.4; 0.45; 0.5	0.35;0.4; 0.45; 0.5	0.35;0.4; 0.45; 0.5
Scenarios 2	0.7;0.75; 0.85; 0.9	0.55;0.6; 0.65; 0.7	0.55;0.6; 0.65; 0.7	0.55;0.6; 0.65; 0.7
Scenarios 3	0.35;0.4; 0.45; 0.5	0.7;0.75; 0.85; 0.9	0.7;0.75; 0.85; 0.9	0.7;0.75; 0.85; 0.9

Information bias

Information bias encompasses a variety of systematic errors. Missing data is one of them and a very frequent one in health sciences research. To address the problem of missing data, specifically in covariates, multivariate imputation using chained equations (MICE) was the applied method in the study –(Acock, 2005; White, Royston, & Wood, 2011). In this case, the MAR (Missing At Random) mechanism of missingness was assumed (Rubin, 1976). The imputation process has been described elsewhere (White et al., 2011). Twenty datasets were generated (Sterne et al., 2009) and the imputation method was performed when variables had more than 10% of missing values –(Bennett, 2001). The final selected model was the most appropriate one based on a set of imputation models and the obtained average relative variance increase (RVI) (Acock, 2014). Socioeconomic status (SES) variable was imputed and then used as a predictor variable in the final risk logistic regression model. This variable had a significant amount of missing data mainly because they were included after the study began. SES variable is build with eight variables from the dataset (Education level, number of economic providers in the house, occupation of main provider, having computer at home, having internet at home, having debit card, having health care and having cars, see –Becaria Coquet et al., 2016). When one of these variables is missing, the SES will have missing values too. Therefore, these observed variables were imputed and then the SES was calculated. Six out of 8 variables were imputed. The imputed variables had an elevated percentage of missing data (40%), hence it was decided to apply the MICE analysis in the dataset with only half of the subjects with missing data. A random sample of 50% of them was obtained in addition to the complete database, resulting in a sample of 391 subjects (125 cases and 266 controls). See supplementary files for the exploratory data analysis.

Models

All models considered the presence/absence of CRC as the outcome. The exposure covariate was *Southern Cone dietary Pattern*, which was previously identified in Cordoba's population through a principal component factor analysis. This pattern was characterized by positive high loadings of red meats, starchy vegetables and wine (Pou et al, 2014). Other recognized risk factors for CRC were included: Age, sex, socioeconomic status, body mass index (BMI), physical activity and family history of CRC. A Logistic multiple regression model was used. Stata 13.0 software (StataCorp LP, USA) was used for analysis.

Results

The exploratory analysis presented below show the proposed methods for addressing confounding. Table 2 presents the stepwise method, where a multiple logistic regression model including all the proposed confounding variables was used. Risk estimations and its percentages of change are shown, including all the variables one at a time in an additive way. The order in which they appear is determined by the magnitude of change in the estimate exerted by each variable. Sex and socioeconomic status were the variables with more impact in risk estimations.

Table 2. Association Measurements (Odds Ratio), confidence intervals (CI) and percentages of change. Crude and adjusted analysis with Stepwise method. Colorectal Cancer Case-Control Study Córdoba, Argentina 2010-2016.

Added variables	Odds Ratio	CI 95%	% of change
Crude	1.72	1.017-2.918	-
Sex	2.30	1.208-4.399	33.78%
SES	2.08	1.077-4.030	-9.61%
Family history of CRC	2.00	1.031-3.911	-3.61%
BMI	1.94	0.993-3.821	-3.01%
Physical Activity	1.90	0.968-3.750	-2.16%
Age	1.89	0.960-3.745	-0.49%

Regarding combine analysis, Figure 1 presents risk estimates resulting of logistic regressions when all confounding variables are included graphically, indicating the corresponding p-value. In this case, 64 set of confounding variables were combined, hence, 64 OR are presented in the graph. All OR are significant and the modifications of the estimates were modest. The punctual value of the OR estimate when all the variables were included in the model is situated in the top left of the graph, near to the risk estimate derived from the model with the best performance according to AIC statistic.

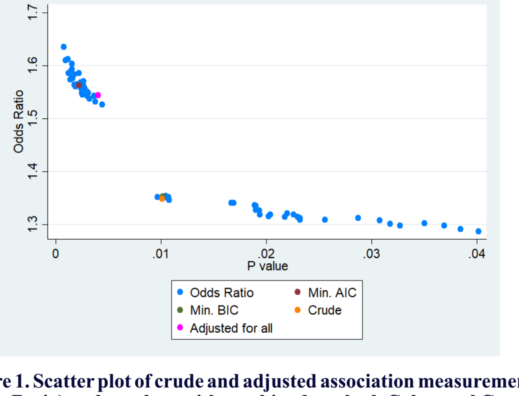

Figure 1. Scatter plot of crude and adjusted association measurements (Odds Ratio) and p-values with combined method. Colorectal Cancer Case-Control Study Córdoba, Argentina 2010-2016.

Table 3 shows the results from the unregistered confounder analysis. Risk estimates and p values were obtained resulting from the different *a priori* proposed scenarios. The OR estimated values increase in magnitude as variability and associations coefficients increase. Maximum modification of effect achieved a 70% of difference comparing to OR without variability imposed (OR=1.44).

Table 3. Association Measurements (Odds Ratio) between adherence to *Southern Cone dietary Pattern* and Colorectal Cancer. Simulation of different risk scenarios with *a priori* assigned variability of the unregistered confounder. Case-Control Study Córdoba, Argentina 2010-2016.

	Observed variable association (BMI)	Odds Ratio	p value
SD=0	0.2	1.44	0.01
	0.3	1.44	0.01
	0.4	1.44	0.01
	0.5	1.44	0.01
SD=1	0.2	1.58	0.006
	0.3	1.60	0.005
	0.4	1.62	0.004
	0.5	1.64	0.003
SD=2	0.2	1.88	0.005
	0.3	1.91	0.003
	0.4	1.62	0.004
	0.5	1.98	0.001
SD=3	0.2	2.30	0.004
	0.3	2.37	0.002
	0.4	2.42	0.001
	0.5	2.47	0.001

Selection bias

Table 4 shows the selection bias analyses, indicating the obtained distribution percentiles of risk estimates and the ratio between the limits of the interquartile range. When non-differential scenarios were imposed, the 50th percentile was the same as the conventional risk estimate. When imposing scenarios where exposed cases had more probabilities of participation (scenarios 1 and 2), the punctual risk estimates where smaller than the conventional one. Imposing scenarios where exposed cases have less participation probabilities, on the other hand, the 50th percentile doubled the one obtained in conventional analysis (scenario 3).

Table 4. Association Measurements (Odds Ratio) between adherence to *Southern Cone dietary Pattern* and Colorectal Cancer. Deterministic selection bias analysis. Case-Control Study Córdoba, Argentina 2010-2016.

	Error	Percentiles			Ratio
		2.5	50	97.5	97.5/2.5
	Conventional	1.17	1.72	2.52	2.16
Deterministic Analysis					
Non- differential Scenario	Systematic Error	1.39	1.72	2.12	1.53
	Random and systematic	1.13	1.72	2.68	2.37
Differential scenarios					
Scenario 1	Systematic Error	0.70	0.91	1.19	1.71
	Random and systematic	0.58	0.91	1.45	2.50
Scenario 2	Systematic Error	1.10	1.34	1.63	1.49
	Random and systematic	0.88	1.34	2.09	2.37
Scenario 3	Systematic Error	2.73	3.43	4.29	1.57
	Random and systematic	2.24	3.44	5.47	2.44

Information bias

Table 5 presents estimations obtained from complete case analysis and multiple imputation by chained equations. Complete case analysis was applied in almost half of study subjects. In general, the estimated ORs are attenuated when Multiple Imputation is performed. Both approaches showed a significant promoting effect of the Southern Cone dietary patter, including significant associations of sex, BMI, family history of CRC and some SES's categories. Overall, more precise confidence intervals were obtained, even considering the uncertainty associated to the imputation method. The final risk estimate showed a value of Relative Variance Increase equal to 0.11, indicating that the variability of the sample estimated was 11% greater than what it would have been observed if covariates had been. Lowest categories of SES were those showing more effect on the % of standard error increase.

Table 5. Association Measurements (Odds Ratio) between adherence to *Southern Cone dietary Pattern* and Colorectal Cancer. Complete and multiple imputation data analyses. Case-Control Study Córdoba, Argentina 2010-2016.

	Complete Case Analysis			Multiple Imputation			% of standard error increase
	(n=252; 51,4%)			(n=346; 70,6%)			
	Odds Ratio	CI 95%	p value	Odds Ratio	CI 95%	p value	
Southern Cone pattern	1.56	1.164-2.116	0.003	1.44	1.115-1.863	0.005	0.46
Age	0.99	0.977-1.020	0.918	0.99	0.976-1.014	0.624	1.25
Sex	2.36	1.158-4.846	0.018	2.17	1.184-3.99	0.012	0.8
SES							
Low-Low	0.54	0.144-2.059	0.372	0.68	0.198-2.347	0.543	10.64
Upper-Low	0.66	0.201-2.217	0.511	0.73	0.234-2.272	0.586	13.12
Middle	0.27	0.077-0.998	0.050	0.31	0.095-1.064	0.063	13.51
Upper Middle	0.30	0.097-0.985	0.047	0.33	0.116-0.977	0.045	9.35
Upper	0.21	0.063-0.722	0.013	0.21	0.069-0.679	0.009	9.27
BMI	1.07	1.000-1.146	0.049	1.05	0.994-1.113	0.075	1.45
Physical Activity	1.00	0.999-1.000	0.583	1.00	0.999-1.000	0.236	0.43
Family history of CCR	3.57	1.251-10.218	0.017	5.12	2.025-12.94	0.001	2.83

Discussion

A methodological approach was built to quantitatively assess possible systematic errors in a colorectal case-control study in Córdoba, Argentina. Main results showed confounding effect of sex and socioeconomic status and a possible confounding effect of the proposed unmeasured confounder. Regarding selection bias, the analysis suggested that this bias probably is not affecting estimations. Concerning to information bias derived from missing data, Multiple Imputation by Chained Equations method showed benefits by considering more information in the analysis and obtaining more precise estimates.

Diet is a recognized factor associated with CRC. In Córdoba (Argentina), the dietary pattern approach previously utilized by Pou *et al* –(2014), based on a sample of controls representing the general population, identified four dietary patterns characterizing usual diet of people from Córdoba. The greater portion of variability was captured by the Southern Cone pattern and so this was adopted as the exposure variable in this study. This pattern was characterized by high consumptions of red meat, wine and starchy vegetables. Evidence suggests that western-like diets similar to this pattern are associated to a greater risk of CRC (Dermadi et al., 2017). Some of the proposed mechanisms are mentioned below. Red meat contains heme iron related to N-nitroso compounds formation and these, like other cytotoxic compounds derived from lipid peroxidation, are potentially carcinogenic (Vieira et al., 2017; Zhou et al., 2016). Regarding ethanol consumption and risk of CRC, various plausible mechanisms have been proposed, among these the carcinogenic effect of acetaldehyde an ethanol metabolite. Besides, alcohol acts like a solvent facilitating the entry of other carcinogenic compounds to the cell. In addition, it has been reported that ethanol participates in retinol metabolism, affecting growth and cellular differentiation and apoptosis (Choi, Myung, & Lee, 2017; Vieira et al., 2017). Higher intake of starchy vegetables (potato, sweet potato, corn) to the detriment of non-starchy vegetables consumption, have also been associated to a greater risk of cancer occurrence ——(Makarem, Lin, Bandera, Jacques, & Parekh, 2015;

Nagle et al., 2015). It may be related mainly to the anti-cancer effects derived from fiber consumption, associated with short chain fatty acid formation by fermentation of colonic flora, reduction of secondary bile acids production, reduction of bowel transit time and insulin resistance (Murphy et al., 2012).

Research works reporting sensitivity analysis applied on observational or experimental studies are still scarce. When it happens, the methodology applied to address bias is not informed in detail nor explained how it was applied, if it was ——(Groenwold, Van Deursen, Hoes, & Hak, 2008; Kahan, Rehal, & Cro, 2015; Lee et al., 2007; Zhang et al., 2017). Even though all studies are susceptible to bias, their effects are often underestimated. While diet and CRC relation is well documented (WCRF/AIC R, 2012), research specifying methodologies and variables considered as possible confounders are scarce (Becaria Coquet et al., 2014; Bingham et al., 2005; Jamshidinaeini, Akbari, Abdollahi, Ajami, & Davoodi, 2016). Some studies included as covariates the same confounder utilized in this work, without specifying the applied methodologies (Go, Chung, & Park, 2016; Klurfeld, 2015; Shin et al., 2017). Regression models, which were used in present work, are often used to address potential confounders (Hammer et al., 2009). Sex, widely reported associated with CCR (Murphy et al., 2011), and socioeconomic status, were adopted as confounders. It is noteworthy that despite of differences in tumor site between men and women, most studies do not consider specificities by sex in design and interpretation of results (Kim et al., 2015). The impact of socioeconomic inequalities in the context of different pathologies, not just cancer, has been studied in the last decades (Salgado-Barreira, Estany-Gestal, & Figueiras, 2014). Specifically regarding CRC, studies suggest an association between socioeconomic status and the disease (Aarts, Lemmens, Louwman, Kunst, & Coebergh, 2010; Doubeni et al., 2012; Manser & Bauerfeind, 2014). Generally, socioeconomic status variable is built from two or more variables and this fact may have an impact in estimates because of its multidimensional nature, and worse, it may present missing data in some of its indicator variables. In fact, our study defined socioeconomic status through 8 variables related with education level, work, health care, having some goods and services regarding the main provider in the house (Becaria Coquet et al., 2014). Aballay et al (2016) reported an association between this construct and other health related factors such as overweight. In Argentina, other authors have considered a set of independent variables to study socioeconomic status and other chronic diseases such as type 2 diabetes (Elgart et al., 2014) and chronic obstructive pulmonary disease (Grigsby et al., 2016). The possible presence of a confounding effect of socioeconomic status should be considered in each case. In addition, the possibility of the information bias occurrence due to missing data should also be considered especially when there are multidimensional variables and or due to conceptual classification error in case the socioeconomic status is defined by independent variables.

However, while confounding variables can be still controlled in the design or analysis, case control studies on diet and disease have many other potential methodological biases. Diet has small variation within populations, intakes are measured with error, and risk estimates are modest but still important due to the high prevalence of the exposure. And these estimates are based mainly on small mean difference of intake among cases and controls. Hence, small systematic errors can heavily affect the relationship.

Selection bias is recognized as a possible threat to estimates validity derived not only from comparative studies such as case-control but cohort studies (for example, lost to follow up, Howe, Cole, Lau, Napravnik, & Eron, 2016), and experimental studies, when treatment allocation bias may be present (La Caze, 2013). We proposed to address this type of bias as deterministic. In case selection bias was considered to be non-differential, the conventional risk effect would be conservative. It is important to note that in this type of studies the concern is related to selection bias being differential between study groups (Gordis, Leon, 2009; Vrijheid, Deltour, Krewski, Sanchez, & Cardis, 2006). This can happen because the selection process may be different for cases and controls; hence, it is not always directly assumed that exchangeability conditions hold (Geneletti, Richardson, & Best, 2009). This selection process may affect the probability of participation of the different groups, i.e. between exposed cases, unexposed cases, exposed controls and unexposed controls. In case control studies, it is possible that differential participation occurs, where exposed cases may have more participation probabilities than the rest of the groups. These situations may arise because exposed cases may be more motivated or have more interest in participating in studies related to modifiable habits, lifestyle and health. If this was the case, estimations of the effect of the exposure on the pathology would be closer to the null the more difference between participation probabilities assigned between exposed cases and the rest of the groups. Evidence suggests that women take more healthy diet choices than men. This may be associated in part they get more involved in weight control and also their beliefs on the effect of diet on health may be stronger than men´s (Ek, 2015; Wardle et al., 2004).The magnitude and direction of the impact of this specific bias could be influenced by the exposure distribution between men and women. This means that the participation probabilities of exposed cases compared with the rest of the groups may differ in a greater extent if in the former group there were more proportion of exposed men than women. In the present work, selection bias seems not to be important. This may be related to the fact that the control group was taken from the general population. Unlike hospital-based controls studies, the former type of studies may be in part avoiding the relationship among the exposure under study and other related diseases. Besides, CRC study and other case-control studies in Córdoba province had high participation rates of eligible controls, –historically around 8-10%- —(Alicia Navarro et al., 2004; Niclis, Román, Osella, Eynard, & Díaz, 2015).

Others methodologies are described in the literature to explore selection bias such as causal diagrams, known as *directed acyclic graphs* o DAGs. These may provide elements to causal models underlying the research problem (M. H. Hernán & Robins, 2018). Several epidemiologic studies have applied this methodology to identify this bias and proposed plausible biological explanations and valid estimates (Geneletti et al., 2009; Hernán, Hernández-Díaz, & Robins, 2004). DAGs are also valuable when proposing less obvious issues, related to apparently surprising results, or inconsistent with scientific evidence. Such are the cases of the birth weight paradox (Hernández-Díaz, Wilcox, Schisterman, & Hernán, 2008) or the obesity paradox (Banack & Kaufman, 2015). This approach is also utilized when studying other types of bias such as confounding (Hernán, Hernández-Díaz, Werler, & Mitchell, 2002).

Missing data in epidemiologic studies is a frequent problem. If observations with missing values are excluded from the analysis, systematic differences between complete and incomplete cases may be ignored and the obtained estimates may be biased (Hernández, Moriña, & Navarro, 2017). Hence, the derived inference may not apply to the reference population, especially when the number of complete cases is small (National Research Council, 2010). In the present work, Multiple Imputation was applied to address this problem in a CRC case-control study with a not large sample size and only half of participants were included in the complete case analysis. Multiple Imputation method implies replacing missing data with plausible values obtained from the predictive distributions, conditioning on observed or complete information. As a result, multiple imputed datasets are obtained that can be analyzed independently with conventional methods and the derived results are then combined by Rubin's rules (1987). When comparing results from complete case analysis and the ones after Multiple Imputation, unreliable p values may be obtained in the first case (Ibrahim, Chu, & Chen, 2012). In some cases, Multiple Imputation can prove to be beneficial when estimating relatively complete covariates coefficients in presence of other incomplete covariates (White & Carlin, 2010). Our study showed interesting results since, in general, estimated ORs were attenuated after MI and more accurate confidence intervals were obtained.

The process of Multiple Imputation is valid when the mechanism of missing data selected is appropriate for the dataset (Molenberghs & Kenward, 2007). Regarding the missing data mechanisms widely described, here the possibility of missing values being missing completely at random (MCAR) was discarded, given the preliminary exploratory analysis. Besides, is not likely a MCAR mechanism to happen in epidemiologic studies, where many variables are registered, that may help to explain missing data patterns in other variables (Soley-Bori, 2013). A missing not at random (MNAR) mechanism neither was considered, because of the impossibility of proving this assumption. The preliminary exploratory analysis (supplementary files) provided valuable insights about the possibility of the missing data mechanism to be missing at random (MAR). Unfortunately, MAR assumption cannot be verified, since missing values are not observed; nevertheless the RVI diagnostic measure, calculated after fitting, indicated good performance of the modeling approach.

In general, one known limitation in cases control studies on diet and disease are overestimation of intakes produced by some instruments like FFQ. In the present CRC study a validated FFQ was used, so small differences in dietary consumption are sometimes not evident. In this specific case-control study, some limitations identified were the study size, making imperative to use as much information as possible, and lack of information regarding tumor site. In addition, the authors recognized as a weakness of the work not including the use of DAGs in the sensitivity analysis. This study has shown that Southern Cone dietary pattern, sex, BMI, family history of CRC and SES are associated with CRC in this population. Additionally, the present study allowed building a methodological framework to quantify systematic errors in a CRC case-control study In Córdoba, Argentina. Due to even large cohort studies do not gather enough information in a reasonable time period, case control studies are central in nutritional epidemiology. Hence, developing and applying methodologies to address systematic error and possible distorted associations become essential.

Acknowledgements

Financial Support

We would like to thank the Science and Technology National Agency, FONCyT grant PICT 2012-1019 for financial support and the National Scientific and Technical Research Council (CONICET) for JBC fellowship.

Supplementary file - Supplementary tables

Table 1. Subjects and missing data: absolute and relative distributions of outcome, exposure and other covariates, colorectal cancer case-control study Córdoba, Argentina 2008-2015.

	n	%	% of missing values				
Total	490	100	**N° of providers**	**Computer**	**Internet**	**Debit Card**	**Cars**
CRC							
No	329	67.1	39.21	40.73	39.51	40.73	40.12
Yes	161	32.9	39.75	39.75	39.75	39.75	39.75
Southern Cone Pattern							
Tertil 1	144	29.5	38.89	38.89	38.89	38.89	38.89
Tertil 2	162	33.1	41.98	44.4	41.98	44.4	43.21
Tertil 3	184	37.4	37.7	38.25	38.25	38.25	38.25
Sex							
Women	230	46.9	39.57	40.43	39.57	40.43	40.0
Men	260	53.1	39.23	40.38	39.62	40.38	40.0
Age							
<45 years	49	10.0	30.61	32.65	30.61	32.65	32.65
45-60 years	117	23.9	41.17	41.03	40.17	41.03	41.03
>60 years	324	66.1	40.43	41.36	40.74	41.36	40.74
BMI							
<25kg/mt^2	180	36.7	33.33	33.89	33.33	33.89	33.89
25-30kg/mt^2	193	39.4	36.27	37.31	36.27	37.31	36.79
>30kg/mt^2	115	23.5	54.78	56.52	55.65	56.52	55.65
Unknown	2	0.4	0.0	0.0	0.0	0.0	0.0
Family history of CRC							
No	456	93.1	39.25	40.13	39.47	40.13	39.91
Yes	34	6.9	41.18	44.12	41.18	44.12	41.18
Physical Activity							
Sedentary	120	24.5	15.83	16.67	15.83	16.67	16.67
Moderate	237	48.4	63.29	64.14	63.29	64.14	63.29
Vigorous	97	19.8	23.71	25.77	24.74	25.77	24.74
Unknown	36	7.3	2.78	2.78	2.78	2.78	5.56

Table 2. Subjects and missing data: absolute and relative distributions of socioeconomic variables, breast cancer case-control study Córdoba, Argentina 2010-2016.

	n	%	% of missing values				
Total	490	100	**N° of providers**	**Computer**	**Internet**	**Debit Card**	**Cars**
Education							
No Studies	4	0.8	75.0	75.0	75.0	75.0	75.0
Incomplete primary	68	13.9	39.71	41.18	41.18	41.18	39.71
Complete primary	128	26.1	46.09	46.88	46.09	46.88	46.09
Incomplete high school	88	18.0	40.91	42.05	40.91	42.05	43.18
Complete high school	92	18.8	40.30	42.39	41.30	42.39	42.39
Higher education	103	21.0	23.30	24.27	23.30	24.27	23.30
Unknown	22	1.4	85.71	85.71	85.71	85.71	85.71
Health Care							
No	56	11.5	51.79	53.57	51.79	53.57	55.36
Yes	424	86.5	36.79	37.74	37.03	37.74	37.03
Unknwown	10	2.0	80.0	80.0	80.0	80.0	80.0
N° of providers							
One	133	27.1	_	2.26	0	2.26	0.75
Two or three	161	32.9	_	1.24	0.62	1.24	1.24
More than three	3	0.6	_	0	0	0	0
Unknwown	193	39.4	_	100	100	100	100
Computer							
No	123	25.1	0	_	0	0	0.81
Yes	169	34.5	0	_	0	0	0
Unknwown	198	40.4	97.47	_	100	100	98.48
Internet							
No	148	30.2	0	0	_	0	0.68
Yes	148	30.2	0	2.70	_	2.7	1.35
Unknwown	194	39.6	99.48	100	_	100	99.48
Debit Card							
No	156	31.8	0	0	0	_	0.64
Yes	136	27.8	0	0	0	_	0
Unknwown	198	40.4	97.47	100	100	_	98.48
Cars							
None	138	42.8	0	0	0	0	_
One	144	39.3	0	2.08	0.69	2.08	_
Two	12	29.4	0	0	0	0	_
Three or more	0	0.0	0	0	0	0	_
Unknwown	196	40.0	98.47	99.49	98.47	99.49	_

*Shaded variables with more than 10% of missing.

Table 3. Subjects and missing data: absolute and relative distributions of occupation of main provider variable, colorectal cancer case-control study Córdoba, Argentina 2010-2016.

	n	%	% of missing values				
Total	490	100					
Occupation of main provider			**Prov**	**Comp**	**Internet**	**DC**	**Cars**
Owner or higher managment with >50 employees	0	0	0	0	0	0	0
Owner or higher managment 6-50 employees	15	3.06	40.0	40.0	40.0	40.0	40.0
Owner or higher managment with 1-5 employees	29	5.92	24.14	24.14	24.14	24.14	24.14
Independet professional	65	13.27	20.0	20.0	20.0	20.0	20.0
Independent technician and dependent relationship	19	3.88	47.37	47.37	47.37	47.37	47.37
trader without employee, craftsman, supervisor	56	11.43	25.0	26.79	25.0	26.79	26.79
Employee without hierarchy	141	28.78	39.72	41.13	39.72	41.13	41.13
Skilled worker	62	12.65	48.39	50	50.0	50.0	48.39
Independent or unskilled worker	90	18.37	55.56	56.67	55.56	56.67	55.56
Unemployed or non-formal/ temporary worker	5	1.02	20.0	20.0	20.0	20.0	20.0
Retiree/pensioner	0	0	0	0	0	0	0
Unknown	8	1.63	87.50	87.50	87.50	87.50	87.5

Prov=N° of providers; Comp=Computer; DC=Debit Card.

References:

1. Aarts, M. J., Lemmens, V. E. P. P., Louwman, M. W. J., Kunst, A. E., & Coebergh, J. W. W. (2010). Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. European Journal of Cancer, 46(15), 2681–2695. https://doi.org/10.1016/j.ejca.2010.04.026
2. Aballay, L. R., Osella, A. R., De La Quintana, A. G., & Diaz, M. del P. (2016). Nutritional profile and obesity: results from a random-sample population-based study in Córdoba, Argentina. European Journal of Nutrition, 55(2), 675–685. https://doi.org/10.1007/s00394-015-0887-0
3. Acock, A. C. (2005). Working with missing values. Journal of Marriage and Family, 67(4), 1012–1028.
4. Acock, Alan C. (2014). Working with missing values-multiple imputation. In A Gentle Introduction to Stata (Fourth edition). College Station, Texas, USA: Stata Press.
5. Ambrosini, G. L., Fritschi, L., de Klerk, N. H., Mackerras, D., & Leavy, J. (2008). Dietary Patterns Identified Using Factor Analysis and Prostate Cancer Risk: A Case Control Study in Western Australia. Annals of Epidemiology, 18(5), 364–370. https://doi.org/10.1016/j.annepidem.2007.11.010
6. Banack, H. R., & Kaufman, J. S. (2015). Does selection bias explain the obesity paradox among individuals with cardiovascular disease? Annals of Epidemiology, 25(5), 342–349. https://doi.org/10.1016/j.annepidem.2015.02.008
7. Becaria Coquet, J., Juárez, S. M., Flores, M. A., Pou, S. A., Aballay, L. R., & Díaz, M. del P. (2014). Identificación de factores de confusión en el estudio de la relación cáncer colorrectal-dieta. Revista Chilena de Salud Pública, 18(2), 161.
8. Becaria Coquet, J., Tumas, N., Osella, A. R., Tanzi, M., Franco, I., & Diaz, M. D. P. (2016). Breast Cancer and Modifiable Lifestyle Factors in Argentinean Women: Addressing Missing Data in a Case-Control Study. Asian Pacific Journal of Cancer Prevention, 17(10), 4567–4575.
9. Bennett, D.A. (2001). How can I deal with missing data in my study? Australian and New Zeland Journal of Public Health, 25(5), 464–469. https://doi.org/10.1111/j.1467-842X.2001.tb00294.x
10. Bingham, S. A., Norat, T., Moskal, A., Ferrari, P., Slimani, N., Clavel-Chapelon, F., … Riboli, E. (2005). Is the association with fiber from foods in colorectal cancer confounded by folate intake? Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology, 14(6), 1552–1556. https://doi.org/10.1158/1055-9965.EPI-04-0891
11. Breslow, N. E., & Day, N. E. (1980). Statistical methods in cancer research. The analysis of case-control studies (Vol. 1). Lyon: International Agency for Research on Cancer.
12. Buis, M. L. (2010). Chapter 7, The consequences of unobserved heterogeneity in a sequential logit model. In: Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands during the 20th Century. (PhD thesis). Retrieved from http://www.maartenbuis.nl/dissertation/chap_7.pdf
13. Choi, Y.-J., Myung, S.-K., & Lee, J.-H. (2017). Light Alcohol Drinking and Risk of Cancer: A Meta-analysis of Cohort Studies. Cancer Research and Treatment. https://doi.org/10.4143/crt.2017.094
14. Dermadi, D., Valo, S., Ollila, S., Soliymani, R., Sipari, N., Pussila, M., … Nyström, M. (2017). Western Diet Deregulates Bile Acid Homeostasis, Cell Proliferation, and Tumorigenesis in Colon. Cancer Research, 77(12), 3352–3363. https://doi.org/10.1158/0008-5472.CAN-16-2860
15. Díaz, M. del P., Corrente, J. E., Osella, A. R., & Muñoz, S. E. (2010). Modeling spatial distribution of cancer incidence in Córdoba, Argentina. Applied Cancer Research, 30(2), 245–252.
16. Doubeni, C. A., Laiyemo, A. O., Major, J. M., Schootman, M., Lian, M., Park, Y., … Sinha, R. (2012). Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study. Cancer, 118(14), 3636–3644. https://doi.org/10.1002/cncr.26677
17. Draper, Norman R., & Smith, Harry. (1998). Applied Regression Analysis (3rd ed). United States: John Wiley & Sons, Inc. Retrieved from http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471170828.html
18. Ek, S. (2015). Gender differences in health information behaviour: a Finnish population-based survey. Health Promotion International, 30(3), 736–745. https://doi.org/10.1093/heapro/dat063
19. Elgart, J. F., Caporale, J. E., Asteazarán, S., De La Fuente, J. L., Camilluci, C., Brown, J. B., … Gagliardino, J. J. (2014). Association between socioeconomic status, type 2 diabetes and its chronic complications in Argentina. Diabetes Research and Clinical Practice, 104(2), 241–247. https://doi.org/10.1016/j.diabres.2014.02.010
20. Geneletti, S., Richardson, S., & Best, N. (2009). Adjusting for selection bias in retrospective, case-control studies. Biostatistics (Oxford, England), 10(1), 17–31. https://doi.org/10.1093/biostatistics/kxn010
21. Go, Y., Chung, M., & Park, Y. (2016). Dietary Patterns for Women With Triple-negative Breast Cancer and Dense Breasts. Nutrition and Cancer, 68(8), 1281–1288. https://doi.org/10.1080/01635581.2016.1225102
22. Gordis, Leon. (2009). Epidemiology - 4th Edition (4th ed). Saunders Elsevier. Retrieved from https://www.elsevier.com/books/epidemiology/gordis/978-1-4160-4002-6
23. Grigsby, M., Siddharthan, T., Chowdhury, M. A., Siddiquee, A., Rubinstein, A., Sobrino, E., … Checkley, W. (2016). Socioeconomic status and COPD among low- and middle-income countries. International Journal of Chronic Obstructive Pulmonary Disease, 11, 2497–2507. https://doi.org/10.2147/COPD.S111145
24. Groenwold, R. H. H., Van Deursen, A. M. M., Hoes, A. W., & Hak, E. (2008). Poor quality of reporting confounding bias in observational intervention studies: a systematic review. Annals of Epidemiology, 18(10), 746–751. https://doi.org/10.1016/j.annepidem.2008.05.007
25. Hammer, G. P., du Prel, J.-B., & Blettner, M. (2009). Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications. Deutsches Arzteblatt International, 106(41), 664–668. https://doi.org/10.3238/arztebl.2009.0664
26. Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. Epidemiology (Cambridge, Mass.), 15(5), 615–625.
27. Hernán, M. A., Hernández-Díaz, S., Werler, M. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. American Journal of Epidemiology, 155(2), 176–184.
28. Hernán, M. H., & Robins, J. M. (2018). Causal Inference Book. United States: Chapman & Hall/CRC. Retrieved from https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
29. Hernández, G., Moriña, D., & Navarro, A. (2017). Imputación de valores ausentes en salud pública: conceptos generales y aplicación en variables dicotómicas. Gaceta Sanitaria, 31(4), 342–345. https://doi.org/10.1016/j.gaceta.2017.01.001
30. Hernández-Avila, M., Garrido, F., & Salazar-Martínez, E. (2000). Sesgos en estudios epidemiológicos. Salud Pública de México, 42(5), 438–446.
31. Hernández-Díaz, S., Wilcox, A. J., Schisterman, E. F., & Hernán, M. A. (2008). From causal diagrams to birth weight-specific curves of infant mortality. European Journal of Epidemiology, 23(3), 163–166. https://doi.org/10.1007/s10654-007-9220-4
32. Howe, C. J., Cole, S. R., Lau, B., Napravnik, S., & Eron, J. J. (2016). Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology (Cambridge, Mass.), 27(1), 91–97. https://doi.org/10.1097/EDE.0000000000000409
33. Ibrahim, J. G., Chu, H., & Chen, M.-H. (2012). Missing Data in Clinical Studies: Issues and Methods. Journal of Clinical Oncology, 30(26), 3297–3303. https://doi.org/10.1200/JCO.2011.38.7589
34. International Agency for Research on Cancer. (2012). GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence worldwide in 2012. Retrieved April 14, 2016, from http://globocan.iarc.fr/Pages/fact_sheets_population.aspx
35. Jamshidinaeini, Y., Akbari, M. E., Abdollahi, M., Ajami, M., & Davoodi, S. H. (2016). Vitamin D Status and Risk of Breast Cancer in Iranian Women: A Case-Control Study. Journal of the American College of Nutrition, 35(7), 639–646. https://doi.org/10.1080/07315724.2015.1127786
36. Kahan, B. C., Rehal, S., & Cro, S. (2015). Risk of selection bias in randomised trials. Trials, 16. https://doi.org/10.1186/s13063-015-0920-x
37. Kim, S.-E., Paik, H. Y., Yoon, H., Lee, J. E., Kim, N., & Sung, M.-K. (2015). Sex- and gender-specific disparities in colorectal cancer risk. World Journal of Gastroenterology, 21(17), 5167–5175. https://doi.org/10.3748/wjg.v21.i17.5167
38. Klurfeld, D. M. (2015). Research gaps in evaluating the relationship of meat and health. Meat Science, 109, 86–95. https://doi.org/10.1016/j.meatsci.2015.05.022
39. La Caze, A. (2013). Why randomized interventional studies. The Journal of Medicine and Philosophy, 38(4), 352–368. https://doi.org/10.1093/jmp/jht028
40. Lash, T. L., Fox, M. P., & Fink, A. K. (2009). Applying Quantitative Bias Analysis to Epidemiologic Data. New York, NY: Springer New York.
41. Lee, W., Bindman, J., Ford, T., Glozier, N., Moran, P., Stewart, R., & Hotopf, M. (2007). Bias in psychiatric case–control studies: Literature survey. The British Journal of Psychiatry, 190(3), 204–209. https://doi.org/10.1192/bjp.bp.106.027250
42. Makarem, N., Lin, Y., Bandera, E. V., Jacques, P. F., & Parekh, N. (2015). Concordance with World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) guidelines for cancer prevention and obesity-related cancer risk in the Framingham Offspring cohort (1991–2008). Cancer Causes & Control : CCC, 26(2), 277–286. https://doi.org/10.1007/s10552-014-0509-9
43. Manser, C. N., & Bauerfeind, P. (2014). Impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer patients: a systematic review. Gastrointestinal Endoscopy, 80(1), 42–60.e9. https://doi.org/10.1016/j.gie.2014.03.011
44. Molenberghs, G., & Kenward, M. G. (2007). Missing Data in Clinical Studies. John Wiley & Sons, Ltd.
45. Murphy, G., Devesa, S. S., Cross, A. J., Inskip, P. D., McGlynn, K. A., & Cook, M. B. (2011). Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. International Journal of Cancer, 128(7), 1668–1675. https://doi.org/10.1002/ijc.25481
46. Murphy, N., Norat, T., Ferrari, P., Jenab, M., Bueno-de-Mesquita, B., Skeie, G., … Riboli, E. (2012). Dietary fibre intake and risks of cancers of the colon and rectum in the European prospective investigation into cancer and nutrition (EPIC). PloS One, 7(6), e39361. https://doi.org/10.1371/journal.pone.0039361
47. Nagle, C. M., Wilson, L. F., Hughes, M. C. B., Ibiebele, T. I., Miura, K., Bain, C. J., … Webb, P. M. (2015). Cancers in Australia in 2010 attributable to inadequate consumption of fruit, non-starchy vegetables and dietary fibre. Australian and New Zealand Journal of Public Health, 39(5), 422–428. https://doi.org/10.1111/1753-6405.12449
48. National Research Council. (2010). The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Retrieved April 14, 2016, from http://www.cytel.com/hs-fs/hub/1670/file-2411099288-pdf/Pdf/MissingDataNationalAcademyof_Medicine.2010.pdf
49. Navarro, A., Cristaldo, P., Andreatta, M. M., Muñoz, S. E., Diaz, M. P., Lantieri, M. J., & Eynard, A. R. (2007). Atlas de alimentos. Córdoba (Argentina): Universidad Nacional de Córdoba, UNC.
50. Navarro, A., Muñoz, S. E., Lantieri, M. J., del Pilar Diaz, M., Cristaldo, P. E., de Fabro, S. P., & Eynard, A. R. (2004). Meat cooking habits and risk of colorectal cancer in Córdoba, Argentina. Nutrition (Burbank, Los Angeles County, Calif.), 20(10), 873–877. https://doi.org/10.1016/j.nut.2004.06.008
51. Navarro, A., Osella, A. R., Guerra, V., Muñoz, S. E., Lantieri, M. J., & Eynard, A. R. (2001). Reproducibility and validity of a food-frequency questionnaire in assessing dietary intakes and food habits in epidemiological cancer studies in Argentina. Journal of Experimental & Clinical Cancer Research: CR, 20(3), 365–370.
52. Niclis, C., Román, M. D., Osella, A. R., Eynard, A. R., & Díaz, M. del P. (2015). Traditional Dietary Pattern Increases Risk of Prostate Cancer in Argentina: Results of a Multilevel Modeling and Bias Analysis from a Case-Control Study. Journal of Cancer Epidemiology, 2015. https://doi.org/10.1155/2015/179562
53. Orsini, Nicola, Belloco, Rino, Bottai, Matteo, Wolk, Alicja, & Greenland, Sander. (2008). A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. The Stata Journal, 8(1), 29–48.
54. Pou, S. A., Díaz, M. del P., & Osella, A. R. (2012). Applying multilevel model to the relationship of dietary patterns and colorectal cancer: an ongoing case–control study in Córdoba, Argentina. European Journal of Nutrition, 51(6), 755–764. https://doi.org/10.1007/s00394-011-0255-7
55. Pou, S. A., Niclis, C., Aballay, L. R., Tumas, N., Román, M. D., Muñoz, S. E., … Díaz, M. del P. (2014). Cáncer y su asociación con patrones alimentarios en Córdoba (Argentina). Nutrición Hospitalaria, 29(3), 618–628. https://doi.org/10.3305/nh.2014.29.3.7192
56. Rothman, Kenneth J., Greenland, Sander, & Lash, Timothy L. (1998). Modern Epidemiology (2nd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins. Retrieved from https://www.readanybook.com/category/books/iscellaneous-problems-in-maritime-navigation-transport-and-shipping-by-adam-weintrit-and-tomasz-neumann
57. Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470316696.fmatter/summary
58. Rubin, Donald B. (1976). Inference and missing data. Biometrika, 63(3), 581–592.
59. Salgado-Barreira, Á., Estany-Gestal, A., & Figueiras, A. (2014). Efecto del nivel socioeconómico sobre la mortalidad en áreas urbanas: revisión crítica y sistemática. Cadernos de Saúde Pública, 30(8), 1609–1621. https://doi.org/10.1590/0102-311X00152513
60. Schneeweiss, S. (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiology and Drug Safety, 15(5), 291–303. https://doi.org/10.1002/pds.1200
61. Shin, S., Saito, E., Sawada, N., Ishihara, J., Takachi, R., Nanri, A., … JPHC Study Group. (2017). Dietary patterns and colorectal cancer risk in middle-aged adults: A large population-based prospective cohort study. Clinical Nutrition (Edinburgh, Scotland). https://doi.org/10.1016/j.clnu.2017.04.015
62. Sierra, M. S., & Forman, D. (2016). Burden of colorectal cancer in Central and South America. Cancer Epidemiology, 44, S74–S81. https://doi.org/10.1016/j.canep.2016.03.010
63. Soley-Bori, M. (2013). Dealing with missing data: Key assumptions and methods for applied analysis (No. Technical Report No. 4). Boston University. Retrieved from http://www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf
64. Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ, 338(jun29 1), b2393–b2393. https://doi.org/10.1136/bmj.b2393
65. Szklo, Moyses, & Nieto, Javier. (2007). Epidemiology Beyond the basics (2nd ed). Estados Unidos: John and Barlett Publishers. Retrieved from https://es.scribd.com/doc/161356288/Epidemiology-Beyond-the-basics-Moyses-Szklo-Javier-Nieto-pdf
66. Vieira, A. R., Abar, L., Chan, D. S. M., Vingeliene, S., Polemiti, E., Stevens, C., … Norat, T. (2017). Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR

Continuous Update Project. Annals of Oncology: Official Journal of the European Society for Medical Oncology, 28(8), 1788–1802. https://doi.org/10.1093/annonc/mdx171

67. Vrijheid, M., Deltour, I., Krewski, D., Sanchez, M., & Cardis, E. (2006). The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. Journal of Exposure Science & Environmental Epidemiology, 16(4), 371–384. https://doi.org/10.1038/sj.jes.7500509
68. Wang, Zhiqiang. (2007). Two postestimation commands for assessing confounding effects in epidemiological studies. Retrieved from http://www.stata-journal.com/article.html?article=st0124
69. Wardle, J., Haase, A. M., Steptoe, A., Nillapun, M., Jonwutiwes, K., & Bellisle, F. (2004). Gender differences in food choice: the contribution of health beliefs and dieting. Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine, 27(2), 107–116.
70. White, I. R., & Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Statistics in Medicine, 29(28), 2920–2931. https://doi.org/10.1002/sim.3944
71. White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine, 30(4), 377–399. https://doi.org/10.1002/sim.4067
72. World Cancer Research Fund International, & American Institute for Cancer Research (Eds.). (2012). Second Expert Report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: AICR.
73. Zhang, Y., Flórez, I. D., Colunga Lozano, L. E., Aloweni, F. A. B., Kennedy, S. A., Li, A., … Guyatt, G. H. (2017). A systematic survey on reporting and methods for handling missing participant data for continuous outcomes in randomized controlled trials. Journal of Clinical Epidemiology, 88, 57–66. https://doi.org/10.1016/j.jclinepi.2017.05.017
74. Zhou, L., Zahid, M., Anwar, M. M., Pennington, K. L., Cohen, S. M., Wisecarver, J. L., … Mirvish, S. S. (2016). Suggestive evidence for the induction of colonic aberrant crypts in mice fed sodium nitrite. Nutrition and Cancer, 68(1), 105–112. https://doi.org/10.1080/01635581.2016.1102298