



Cancer occurrence among women resident in the Gorizia province (North-Eastern Italy). May ecological data be used for adjusted epidemiological measures? Deterministic and probabilistic sensitivity analysis

Luigi Castriotta^{1,2*}, Valentina Rosolen^{2†}, Ettore Bidoli^{3†}, Paolo Collarile^{3†}, Diego Serraino^{3†} and Fabio Barbone^{4†}

*Correspondence: luigi.castriotta@asuiud.sanita.fvg.it



CrossMark

← Click for updates

[†]These authors contributed equally this work.

¹Istituto di Igiene ed Epidemiologia Clinica, Azienda Sanitaria Universitaria Integrata di Udine.

²Dipartimento di Area Medica, Università degli Studi di Udine.

³SOC Epidemiologia Oncologica, IRCCS Centro di Riferimento Oncologico, Aviano.

⁴Istituto di Ricerca e Cura a Carattere Scientifico materno infantile "Burlo Garofolo", Trieste.

Abstract

Background: The evaluation of the confounders is crucial to accurately estimate the association between environmental factors and diseases. The deterministic sensitivity analysis permits an external adjustment of the observed measures of effect. The probabilistic sensitivity analysis allows to define several probability density functions for the bias parameters and to use these prior distributions to calculate limits for the bias-adjusted exposure-disease relative risks. The confounding effect of known risk factors, for lung, breast and bladder cancer, in women residing in the Gorizia Province (GP), was estimated.

Methods: The deterministic sensitivity analysis was performed by calculating confounding risk ratios (CRRs) under different scenarios of confounding. Several parameters were defined, as the population being studied; the prevalence of women residing in the GP and in the southern and northern Gorizia province (SGP and NGP respectively); the prevalence of the different risk factors in the selected areas; and the risk factor-disease risk ratios per specific type of cancer. The CRRs were finally used to adjust the age-standardised relative ratios. Probabilistic sensitivity analysis allowed to specify prior probability distributions for the lognormal risk factors-disease risk ratio with 95% lognormal confidence limits $\ln(RR1)$ and $\ln(RR2)$, mean and standard deviation equal to $\{\ln(RR1)+\ln(RR2)\}/2$ and $\{\ln(RR2)-\ln(RR1)\}/(2*1.96)$ respectively.

Results: In women resident in the GP, the age-standardised IRR for bladder cancer is equal to 1.18 but, adjusting for smoking through deterministic analysis, the new IRR ranges from 1.13 to 1.25. In SGP the same age-standardised IRR is equal to 1.45 while after adjustment for smoking it ranges from 1.31 to 1.67. Adjusting for the deprivation index in the areas of interest, the IRRs for bladder cancer in GP range from 1.17 to 1.18 while in SGP resident from 1.45 to 1.47. Results from the probabilistic analysis show that the RR for bladder cancer in GP women, when adjusted for smoking results to be equal to 1.20 (2.5th – 97.5th: 0.99 – 1.44) when adjusted for deprivation, appears to be equal to 1.19 (2.5th – 97.5th: 0.99 – 1.43). In the SGP the IRR for bladder cancer, when adjusted for smoking, results equal to 1.58 (2.5th – 97.5th: 1.09 – 2.25) when adjusted for deprivation, 1.57 (2.5th – 97.5th: 1.10 – 2.23). Regarding breast and lung cancers, no risk excess arises in the studied areas, even after adjustment in both deterministic and probabilistic analysis.

Conclusions: Data on bladder cancer would seem to suggest a higher incidence of this cancer among the women residing in GP and, in particular, in the SGP. This excess risk is confirmed after adjustment for smoking and deprivation. No risk excesses in the studied areas for lung and breast cancers.

Keywords: Confounding risk ratio, deterministic analysis, probabilistic analysis, cancer

Introduction

When studying the possible association between the exposure to a risk factor, such as an environmental factor, and a certain disease, the evaluation of the effect of potential confounders is crucial to estimate the risk accurately. Even in the case of a descriptive study, the information on confounders contributes to the correct definition of the hypotheses on which analytical studies are after that based. It is an elaborate effort to obtain individual data on confounders and, if the research investigates wide groups of the population (as is usually the case with descriptive studies), individual information may not be available. The challenge, therefore, lies in using ecological data, which are more easily available, to determine the extent of confounding within the estimated measures.

Sensitivity analysis (SA) provides a way to account for systematic errors and to assess the magnitude of biases; several different SA techniques have been proposed and therefore literature is large. This field was initiated by Cornfield and colleagues when, in 1959, they published a SA for the association between cigarette smoking and lung cancer [1]. For their analysis, authors assumed a single binary confounder, no interaction between the effect of the exposure and the confounder on the outcome, and the null hypothesis of no causal effect. Under these three simplifying assumptions, also known as the Cornfield conditions, authors could determine how strong the potential confounder would have been to completely explain the effect of the exposure (cigarette smoking) on the outcome (lung cancer). Although computationally easy, performing quantitative bias analysis under the Cornfield conditions does not permit the estimate of the effect under alternative (weaker) confounding scenarios.

As cited, literature is large and bias analytic strategies range from easy-to-compute approaches to computationally intensive ones. In the simple deterministic SA, the estimate of the association between exposure and outcome is adjusted a single time for a fixed quantity of the bias parameter. An extension of this technique can be achieved by repeating the earlier analysis for different values of the same confounder. This multidimensional SA permits, therefore, an external adjustment of the effect measures, under different scenarios of confounding, even if it provides none or scarce information of the most likely adjusted estimate. Moreover, both simple and multidimensional SA address only one confounder at a time and, due to the absence of hypothesis about the prior distributions of the bias parameters, do not explicitly account for uncertainty [2,3].

Probabilistic bias analysis (PBA) can be viewed as a generalization of the simple SA in which the bias parameters are defined by probability density functions (joint prior distributions) rather than by one or more fixed values within a range. The so-called Monte-Carlo sensitivity analysis (MCSA) represents the easiest technique to implement PBAs: through this approach, bias parameters, extracted from their joint prior distributions, are used in the deterministic SA formula [3]. However, differently

from the deterministic approach, the joint prior distributions capture the uncertainty about the bias parameters, allowing the definition of the limits for the bias-adjusted effect measure that can be interpreted as the frequentist confidence interval. Similarly to the deterministic SA, PBA accounts only for one source of confounding at a time [3,5].

When multiple sources of systematic errors (selection bias, misclassification of the exposure, unmeasured confounding) are concurrently present, multiple bias modeling (MDM) can be performed with adjustment in the reverse order of bias occurrence. Typically, for each bias source, a different joint prior distribution is defined and all these distributions are assumed to be independent. When this assumption does not hold, prior correlations or hierarchical models can be used to address dependency [3].

Another strategy consists in using frequency distributions of the bias parameters as prior distributions in Bayesian bias analysis (BBA). This approach starts from the same assumptions of MCSA and if prior distributions are defined for all parameters in the model then the analysis is known as fully Bayesian, else semi-Bayesian. Performing BBA is computationally demanding nevertheless it gives more reliable posterior probability distributions [3]. This occurs because while in BBA prior distributions are combined with the likelihood functions for the parameters, in MCSA, as previously explained, the bias parameters are repeatedly sampled from the prior distributions and then used to calculate bias-adjusted estimates [6]. However, according to Greenland [4], BBA and MCSA may give very similar results if, in BBA, prior and posterior distributions of bias parameters do not differ.

Friuli Venezia Giulia (FVG) region (North-Eastern Italy) has a cancer registry (RT-FVG) which is part of the national system of cancer registration and is a member of the Italian Association of Cancer Registries (AIRTUM). RT-FVG is internationally accredited by the International Association of Cancer Registries at the International Agency for Research on Cancer in Lyon [7].

Since 1995 the RT-FVG has been monitoring all new incident cancer cases (that were diagnosed in the time period of interest) among the population residing in the region. The starting information system is the regional epidemiological data warehouse, storing all the information gathered by the regional health information system (SISR). The main data bases used by the RT-FVG are the ones referring to hospital discharge forms (SDO), pathology reports (AP), general death forms (MG) and health data repository of the region inhabitants (A). Each disease is encoded according to the International Classification of Diseases (10th Revision - ICD-10).

On 2011 a monograph by Birri S, Bidoli E, Zucchetto A et al., 'I Tumori in Friuli Venezia Giulia. Dati di incidenza, sopravvivenza e prevalenza: aggiornamento al 2007' published by the Region of Friuli Venezia Giulia, together with RT-FVG, describes the frequency of cancer. Data on incidence, survival, and prevalence refer to the 1995-2007 time period and derive from the RT-FVG, while the corresponding data on mortality

were provided by the National Institute of Statistic (ISTAT). The monograph shows that, for that period, women of the Gorizia province (GP) had a higher mortality, incidence and reduced survival for all types of cancer and, in particular, for malignant lung cancer, and malignant breast, and bladder cancer. Some of these excesses seem especially to affect the municipalities of the area covered by the Southern Gorizia province (SGP) [8].

This preliminary evidence triggered the development of hypotheses by public health officers and groups of citizen that exposure to environmental factors may explain the cited cancer excesses. As the first step, we performed a descriptive study to calculate the 1995-2009 age-adjusted cancer incidence and incidence rate ratios (IRRs) for lung, breast, and bladder cancer in women residing in GP. GP rates were compared with the rest of FVG region, and SGP rates were compared with the Northern Gorizia province (NGP). Afterwards, we tried to determine the magnitude of confounding in the observed findings by using aggregate data, which are easily available and easy-to-compute analytic approaches for sensitivity analysis.

Materials and methods

New cancer incident cases in FVG are recorded on the basis of the international methods defined by the International Association of Cancer Registries (IACR) and the Italian Association of Cancer Registries (AIRTUM). After excluding autopsy cases, a data set was obtained for the 1995-2009 period comprising 8,852 incident cases of malignant neoplasms, 7,000 after excluding non-melanoma skin cancer. Age-standardised rates (ASR) were calculated with SEER*Stat software by the direct method using the conventional European population (EU 1960) as the standard population. For the 1995-2009 period, the population at risk used as denominator was the yearly average resident female population (calculated as the average between the resident population on January 1st of the year under consideration and the resident population on January 1st of the following year). All rates were expressed per 100,000 inhabitants. Ratios between age-standardised rates were also calculated (IRRs).

To select among the anatomical sites of cancer and the potential risk factors to be studied, we considered: frequency of cancer, availability of ecological data, evidence in the literature of an association between cancer and risk factors that were both environmental and related to smoking, alcohol consumption, occupational exposure and socioeconomic status. As mentioned, to evaluate the extent of confounding, within these measures, we implemented both deterministic and probabilistic sensitivity analysis, based on available ecological data.

Deterministic sensitivity analysis

The deterministic sensitivity analysis was performed by calculating confounding risk ratios (CRRs) under different scenarios of confounding. CRR is defined as the ratio between the crude

risk ratio and the risk ratio after removing the effect of the risk factor and measures the degree of confounding [9-12].

In order for the CRR to be calculated, it was necessary to define several parameters as the population being studied (women residing in FVG on 31.12.2009); the prevalence of women residing in the GP (exposed to the unknown environmental factor), and in SGP or NGP on 31.12.2009; the prevalence of women residing in the remaining territory of FVG on 31.12.2009; the prevalence of the different risk factors (smoking, alcohol and deprivation index) in the selected areas; and the risk factor-disease risk ratios per specific type of cancer. The CRRs were finally used to adjust the age-standardised IRRs.

Discrete scenarios of confounding were hypothesized by varying the strength of association between risk factors and the selected types of cancer and using both the point measure and the confidence limits of the risk factors' prevalence values, to simulate both the most likely and the extreme scenarios.

The 95% confidence intervals of the prevalence were calculated when data were based on samples (i.e. the PASSI survey on smoking habits and alcohol intake) [13], while for data based on the national census (i.e. employees by economic activity and deprivation index [14,15]) only the point measure was obtained.

Therefore, when available, the confidence limits were used to force the differences between prevalence values in the compared areas, alternating the lower limit in one area and the upper limit in the compared one and vice versa. Consequently, CRR_1 derives from using the lower limit of the confidence intervals for the GP and the SHD and the upper limit for the rest of the region FVG and the NHD, while CRR_2 derive from the upper limit for the GP and the SHD and the lower limit for the rest of the region FVG and the NHD.

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis, through Monte Carlo simulations, allowed to define several probability density functions for the risk factors and to use these prior distributions to calculate limits for the bias-adjusted exposure-disease relative risks (RRs) [2,4,5]. The term RR is here used as a generic term for the risk ratio, the rate ratio or odds ratio. Indeed, the probabilistic sensitivity analysis was conducted through Monte Carlo simulations by firstly drawing a random sample from the specified probability density functions of the bias parameters, and secondly back-calculating the bias-adjusted RR. These two steps were iterated to obtain the distribution of the bias-adjusted RRs [5]. The *a priori* distribution of the bias parameters derived from background information (prevalence values and 95% confidence intervals). The probabilistic sensitivity analysis resulted in probability density functions (distributions) of adjusted measures that took into account the uncertainty about the parameters [4]. We, therefore, specified prior probability distributions for the lognormal risk factors-disease risk ratio with 95% lognormal confidence limits $\ln(RR1)$ and $\ln(RR2)$, mean equals to $\{\ln(RR1)+\ln(RR2)\}/2$

and standard deviation equals to $\{\ln(RR2)-\ln(RR1)\}/(2*1.96)$.

For this study, we ran our sensitivity analysis only under the hypothesis of uncontrolled confounders assuming no exposure misclassification or selection bias.

Risk factors and data sources

As previously mentioned, risk factors data derived from ecological studies and databases. Therefore, for the selected areas, data used for our analysis were the prevalence of consumers at risk from alcohol drinking [13], the prevalence of smokers [13], the prevalence of employees by economic activity [14], from 2011 national census and, lastly, the distribution of the deprivation index (developed on the basis of the information gathered through the 2001 national census) [15].

Risk factors definitions and related diseases

We assumed as definition of consumers at risk from alcohol drinking, women younger than 18 years who have drunk an alcoholic beverage; adult women exceeding 1 standard alcoholic unit and all women – regardless of age - who have binged (binge drinking being defined as the consumption of 6 glasses or more of alcoholic beverages in a single session) at least once a year. (In the PASSI study, the survey on alcohol consumption refers to the last 30 days before the interview) [13]. In our analysis, alcohol was considered a risk factor for breast cancer [16,17].

According to the World Health Organization definition, we assumed women to be smokers, someone who smoked at least 100 cigarettes in her lifetime (5 packets of 20 cigarettes) and was still smoking at the time of the PASSI survey or stopped smoking less than 6 months before [12]. Smoking was considered a risk factor for lung, breast and bladder cancer [18-20]. Prevalence of employees by economic activity (agriculture, industry, commerce, transport, financial activity, and other activity) was calculated on the basis of the 2011 Census data [14]. In our analysis, in order to adjust the IRRs, we used the prevalence of employed in industrial sites as a risk factor for lung cancer [21].

As mentioned above, for each area under analysis, the deprivation index used was developed by Caranci et al. on the basis of the information gathered through the 2001 national census. The deprivation index was subdivided into quintiles, ranking from very affluent to very disadvantaged. For our analysis, we calculate the prevalence of the most disadvantaged class with reference to lung and bladder cancer, while in the case of breast cancer the prevalence of the 'very affluent' category [15,22-26].

For the purpose of this paper, women resident in FVG region were stratified by areas, and comparisons were made among GP versus the rest of the region and the SGP versus the NGP.

The software used for data management and analysis were Microsoft Office Excel (Redmond, Washington: Microsoft, 2013) and STATA (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP).

Results

The reference population was defined by all women resident in FVG on January 1st, 2010; according to the Regional Healthcare Data Warehouse, FVG region had 637,251 inhabitants, while the province of Gorizia had 73,026 inhabitants.

The median age of women residing in FVG is 47 years (in Gorizia 48 years); women over 65 years of age account for 26.7% of the regional population (in Gorizia 28.8%).

Age standardised incidence rates in females by cancer site, study period and area, are reported in **Table 1**. An increasing trend of cancer incidence, particularly for breast cancer in both the entire Region as well as the Gorizia Province, can be noticed.

Although compared with the regional data, the confidence intervals of the estimates referring to the GP and to the SGP and NGP are wider (due to the lower numbers of cases) bladder cancer ASR appears to be higher in the GP (ASR 11.2; 95%CI 9.7-13) and particularly in the SGP (ASR 13.3; 95%CI 10.8-15.8) than in the FVG (ASR 9.5; 95%CI 9-10.1). No relevant differences among lung cancer ASRs arise in the studied areas.

Table 2 shows both incident (1995-2009) and prevalent cancer cases on 01.01.2010 in the GP and in FVG; prevalence values are also indicated for the NGP and SGP.

Table 3 shows prevalence and 95% confidence intervals of smoking and alcohol drinking habits by different areas. There is no evidence of heterogeneity among the different areas of interest.

Concerning the socioeconomic conditions, **Table 4** shows the deprivation index distribution in the study areas. While data belonging to the rest of the Region and the GP seem to be comparable, within the GP, the NGP shows a higher prevalence of the very affluent category, compared to the SGP.

Deterministic analysis results

As already mentioned, firstly a deterministic sensitivity analysis was performed and confounding risk ratios (CRRs) were calculated and used to adjust the age-standardised IRRs; **Tables 5** and **6** show the age-standardised IRRs and the age-standardised adjusted for the CRRs calculated under different scenarios. The IRRs concerning the bladder cancer in women resident in the GP and particularly in the SGP seem to be significant even after adjustment for different hypothetical scenarios of confounding. In women resident in the GP, the age-standardised IRR is equal to 1.18 but, adjusting for smoking, the new IRR ranges from 1.13 to 1.25. In SGP resident, the same age-standardised IRR is equal to 1.45 while, after adjustment for smoking it ranges from 1.31 to 1.67. Adjusting for the different distributions of the deprivation index in the areas of interest, the IRRs for bladder cancer in GP range from 1.17 to 1.18 while in SGP resident from 1.45 to 1.47. As predictable from the age-adjusted incidence rates and the prevalences of the risk factors, for lung and breast cancers, the deterministic analysis shows no relevant results (data not shown).

Table 1. Incident cases of cancer, age standardized incidence rates, ASR (European population, per 100000 women-year), with 95% confidence interval (IC95%), in female by cancer site, study period and area.

Cancers by site (ICD10)	FVG region (Gorizia Province excluded)			Gorizia Province			Northern Gorizia Province			Southern Gorizia Province		
	YEARS	N.	ASR (95%CI)	N.	ASR (95%CI)	N.	ASR (95%CI)	N.	ASR (95%CI)	N.	ASR (95%CI)	
Allsites, skin excluded												
1995-2009	50,267	371.9	(368.3-375.6)	7,000	376.9	(367-387.1)	3,587	373.3	(359.4-387.3)	3,413	382.3	(367.8-396.7)
1995-1999	16,210	370.1	(363.8-376.5)	2,209	364.7	(347.8-382.3)	1,122	357.4	(333.8-381)	1,087	347.1	(349.4-398.9)
2000-2004	16,544	369.2	(362.9-375.6)	2,321	375.2	(358-393.1)	1,195	373.6	(349.4-397.8)	1,126	378.5	(353.5-403.5)
2005-2009	17,513	376.6	(370.4-382.9)	2,470	392	(374.4-410.3)	1,270	390.2	(365.3-415.1)	1,200	395.8	(370.2-421.3)
Lung (C33-34)												
1995-2009	3,104	20.4	(19.6-21.2)	444	20.7	(18.6-23.1)	220	19.6	(16.6-22.5)	224	22.0	(18.9-25.2)
1995-1999	976	20	(18.6-21.4)	138	20.4	(16.8-24.6)	60	17.3	(12.4-22.2)	78	23.9	(18.2-29.6)
2000-2004	970	18.8	(17.5-20.2)	130	19.4	(15.8-23.7)	61	15.8	(11.3-20.3)	69	23.1	(17.1-29.1)
2005-2009	1,158	22.5	(21.1-24)	176	22.4	(18.8-26.7)	99	25.6	(19.8-31.4)	77	19.2	(14.5-24)
Breast (C50)												
1995-2009	15,013	125.3	(123.1-127.5)	2,035	125.5	(119.6-131.7)	1,068	127.5	(119.1-135.9)	967	123.8	(115.3-132.3)
1995-1999	4,488	115.2	(111.6-118.9)	620	114.8	(105.1-125.2)	327	117.5	(103.5-131.6)	293	112.1	(98.2-126)
2000-2004	4,941	124.5	(120.7-128.3)	662	124.4	(114.2-135.3)	351	127.0	(112.4-141.6)	311	122.0	(107.2-136.7)
2005-2009	5,584	135.6	(131.8-139.5)	753	137.5	(127-148.8)	390	137.8	(122.8-152.9)	363	138.3	(122.7-153.9)
Bladder (C67, D9.0, D41.4)												
1995-2009	1,552	9.4	(8.9-10)	246	11.1	(9.6-12.9)	107	9.2	(7.2-11.3)	139	13.3	(10.8-15.8)
1995-1999	544	10.2	(9.3-11.3)	79	10.5	(8.1-13.7)	34	8.2	(5.1-11.3)	45	13.3	(8.9-17.6)
2000-2004	481	8.6	(7.8-9.6)	79	11.4	(8.7-14.8)	41	11.6	(7.5-15.7)	38	11.2	(7.2-15.2)
2005-2009	527	9.5	(8.6-10.5)	88	11.4	(8.8-14.7)	32	7.8	(4.7-11)	56	15.2	(10.6-19.8)

Table 2. Incident and prevalent cases of cancer at 1st January 2010, by site and area infemale. Diagnosis period: 1995-2009.

Cancer site (ICD10)	FVG region (Gorizia Province excluded)		Gorizia Province		Northern Gorizia Province		Southern Gorizia Province	
	Incident	Prevalent	Incident	Prevalent	Incident	Prevalent	Incident	Prevalent
All sites (C00-97) skin excluded	50,267	24,103	7,000	3,262	3,587	1,636	3,413	1,623
Lung (C33-34)	3,104	470	444	58	220	28	224	30
Breast (C50)	15,013	10,540	2,035	1,409	1,068	725	967	684
Bladder (C67, D9.0, D41.4)	1,552	834	246	129	107	51	139	78

Table 3. Prevalence of smoking and alcohol habits.

	Smoking habits			Alcohol drinking habit		
	Prevalence	95%CI		Prevalence	95%CI	
		ll	ul		ll	ul
FVG, Gorizia province excluded (n=3,771)	0.231	0.218	0.245	0.099	0.089	0.108
Gorizia Province (n=681)	0.229	0.198	0.261	0.095	0.073	0.118
Southern Gorizia Province (n=117)	0.214	0.139	0.288	0.086	0.035	0.136
Northern Gorizia Province (n=564)	0.231	0.196	0.265	0.098	0.073	0.122

Table 4. Deprivation index distribution in the study areas.

Deprivationcategory	Area	Women resident*	%
0 missing	FVG	229	0.04
1 Very affluent		107,409	19.68
2 Affluent		110,966	20.33
3 Medium		106,468	19.51
4 Disadvantaged		110,804	20.30
5 Very disadvantaged		109,863	20.13
0 missing	GO prov.	7	0.01
1 Very affluent		15,019	21.76
2 Affluent		10,861	15.74
3 Medium		16,022	23.21
4 Disadvantaged		12,295	17.81
5 Very disadvantaged		14,819	21.47
0 missing	NorthernGorizia Province	2	0.01
1 Very affluent		11,670	32.65
2 Affluent		5,608	15.69
3 Medium		7,184	20.10
4 Disadvantaged		4,239	11.86
5 Verydis advantaged		7,039	19.69
0 missing	Southern Gorizia Province	5	0.02
1 Very affluent		3,349	10.06
2 Affluent		5,253	15.78
3 Medium		8,838	26.56
4 Disadvantaged		8,056	24.21
5 Very disadvantaged		7,780	23.38

* (ISTAT census, 2001)

Probabilistic analysis results

Tables 7 and **8** show the results of the probabilistic analysis in terms of posterior probability distributions of RRs by cancer-type, in women resident in GP compared to FVG and in women resident in the SHD compared to the NHD, both adjusted by risk factors. From the 20,000 draws for each cancer type and bias parameter, the median risk factors-adjusted residence-cancer RRs and 2.5th and 97.5th percentiles are presented, together with the 97.5th/2.5th ratio. As expected, the ratios deriving from systematic and random errors are equal or higher than the ratios of the conventional limits. **Figures 1** and **2** show the distributions of the bias-adjusted RRs, including both systematic and random error. Consistently with the results from the deterministic analysis, among all distributions, the ones concerning the bladder cancer RR in GP women and particularly in SGP resident appear to be relevant even after adjustment for smoking or the deprivation index. Considering both the systematic and the random error, the RR for bladder

cancer in GP women when adjusted for smoking results to be equal to 1.20 (2.5th – 97.5th: 0.99 – 1.44), when adjusted for deprivation appears to be equal to 1.19 (2.5th – 97.5th: 0.99 – 1.43). In the SHD the RR for bladder cancer, when adjusted for smoking, results equal to 1.58 (2.5th – 97.5th: 1.09 – 2.25) when adjusted for deprivation, 1.57 (2.5th – 97.5th: 1.10 – 2.23). Again, analysis concerning lung and breast cancers show no relevant results.

Discussion

As mentioned, the age-standardised incidence rates seem to show an increasing trend over the years for all types of cancer. In particular, this increase appears to be more remarkable for breast cancer (**Table 1**); this trend affects both the province of Gorizia and the rest of FVG, even though caution is necessary when drawing conclusions because of the overlapping confidence intervals. This trend, if real, could be explained in different ways: it may be due to the regional screening program for breast cancer (started in 2005), to an increase in the prevalence of the risk factors over last decades, and to improvements in diagnostic capabilities. The comparison of cancer incidence rates in the province of Gorizia with the rest of FVG, regarding lung and breast cancer, did not show any relevant difference. On the contrary, data on bladder cancer, particularly those referring to the entire period, would seem to suggest a higher incidence of this cancer type among women residing in the province of Gorizia. This is especially true in the Southern Gorizia Province where the ASR reaches the highest value (1995-2009, ASR:13.3. 95%CI: 10.8-15.8). On the contrary, rates in Northern Gorizia Province seem to be comparable with the rest of the region. Although this difference seems worthy of further investigations, it should be noted that bladder cancer classification includes both in situ forms and neoplasms of uncertain behavior. Therefore, it cannot be excluded that a different diagnostic assessment explains this result.

The use of aggregate data from available ecological sources to assess the effect of possible confounders (i.e. smoking, alcohol, deprivation, occupational exposure) on the association between exposure to environmental risk factors and outcome may be relevant and efficient. In fact, it allows an advantageous, in terms of time and costs, estimation of the magnitude of the confounding. However, limitations are equally evident: firstly, ecological data are available only for a limited number of all known risk factors, secondly the use of aggregate data does not provide a measurement of individual exposure but only an indirect calculation. Therefore, despite the adjustment of the effect measures for the above mentioned confounders, a different prevalence of those risk factors that we did not consider and the use of proxies for individual exposures may have left residual confounding in the estimates.

In particular, there are two principal limitations bearing upon the potential use of aggregate data and the reliability of the estimates that derive from these data: a) the limited

Table 5. IRRs by residence and cancer type, adjusted for CRRs calculated under different scenarios of confounding.

Comparison	Risk factor	Cancer type	IRR	RR _{CE+}	CRR	IRR adjusted for CRR	CRR1	IRR adjusted for CRR1	CRR2	IRR adjusted for CRR2
Gorizia province vs Rest of the region	Smoking	Lung Cancer	1.015	15	0.993	1.022	0.851	1.193	1.148	0.884
				20	0.992	1.023	0.841	1.206	1.165	0.871
	Smoking	Breast Cancer	1.002	1.1	1.000	1.002	0.995	1.006	1.004	0.997
				1.5	0.999	1.003	0.979	1.023	1.019	0.983
	Smoking	Bladder Cancer	1.181	2	0.999	1.182	0.962	1.227	1.035	1.141
				2.5	0.998	1.183	0.948	1.245	1.048	1.126
Southern vs Northern Gorizia Province	Alcohol	Breast Cancer	1.002	1.1	1.000	1.002	0.997	1.005	1.003	0.999
				1.5	0.999	1.003	0.984	1.018	1.014	0.988
	Smoking	Lung Cancer	1.122	15	0.944	1.189	0.626	1.792	1.345	0.834
				20	0.941	1.193	0.604	1.858	1.371	0.819
	Smoking	Breast Cancer	0.971	1.1	0.998	0.973	0.988	0.983	1.009	0.962
				1.5	0.992	0.978	0.944	1.028	1.042	0.932
Smoking	Bladder Cancer	1.446	2	0.992	1.457	0.901	1.605	1.077	1.342	
			2.5	0.986	1.466	0.865	1.671	1.107	1.306	
Alcohol	Breast Cancer	1.002	1.1	0.999	1.003	0.991	1.010	1.006	0.995	
			1.5	0.994	1.007	0.959	1.045	1.030	0.972	

Table 6. IRRs by residence and cancer type, adjusted for CRRs calculated under different scenarios of confounding.

Comparison	Risk factor	Cancer type	IRR	RR _{CE+}	CRR	IRR adjusted for CRR
Gorizia province vs Rest of the Region	DI	Lung Cancer	1.015	1.1	1.001	1.013
				1.5	1.006	1.009
				2	1.017	0.985
	DI	Breast Cancer	1.002	1.1	1.002	1.000
				1.5	1.009	0.992
				2	1.017	0.985
	DI	Bladder Cancer	1.181	1.1	1.001	1.179
				1.5	1.006	1.174
				2	1.017	0.985
Occupational	Lung Cancer	1.015	1.5	0.984	1.031	
			2	0.970	1.046	
			2.5	0.958	1.059	
Southern vs Northern Gorizia Province	DI	Lung Cancer	1.122	1.1	0.996	1.127
				1.5	0.984	1.141
				2	0.970	1.046
	DI	Breast Cancer	0.971	1.1	0.978	0.993
				1.5	0.903	1.075
				2	0.830	1.170
	DI	Bladder Cancer	1.446	1.1	0.996	1.451
				1.5	0.984	1.470
				2	0.970	1.046
Occupational	Lung Cancer	1.122	1.5	0.996	1.127	
			2	0.992	1.131	
			2.5	0.989	1.135	

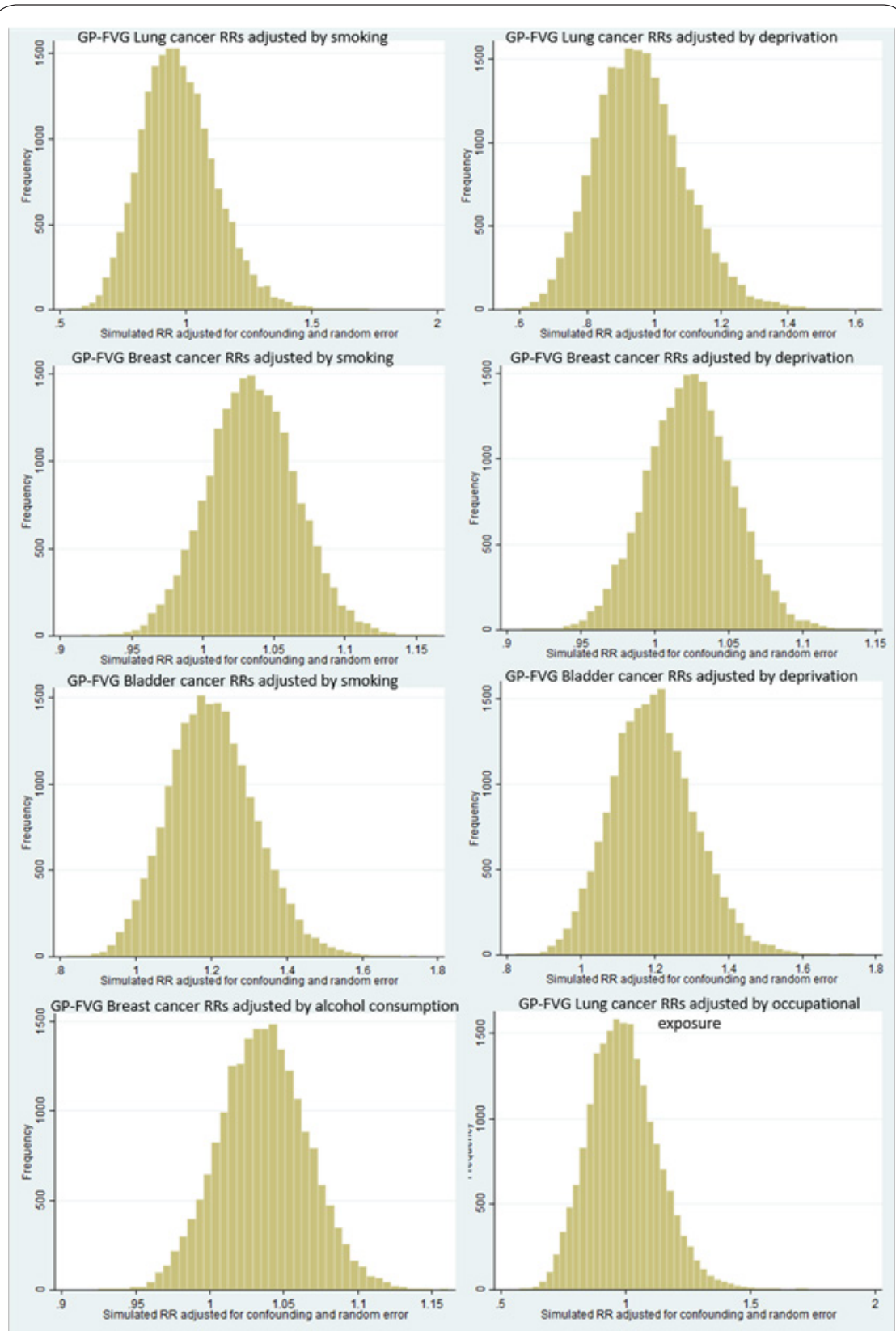


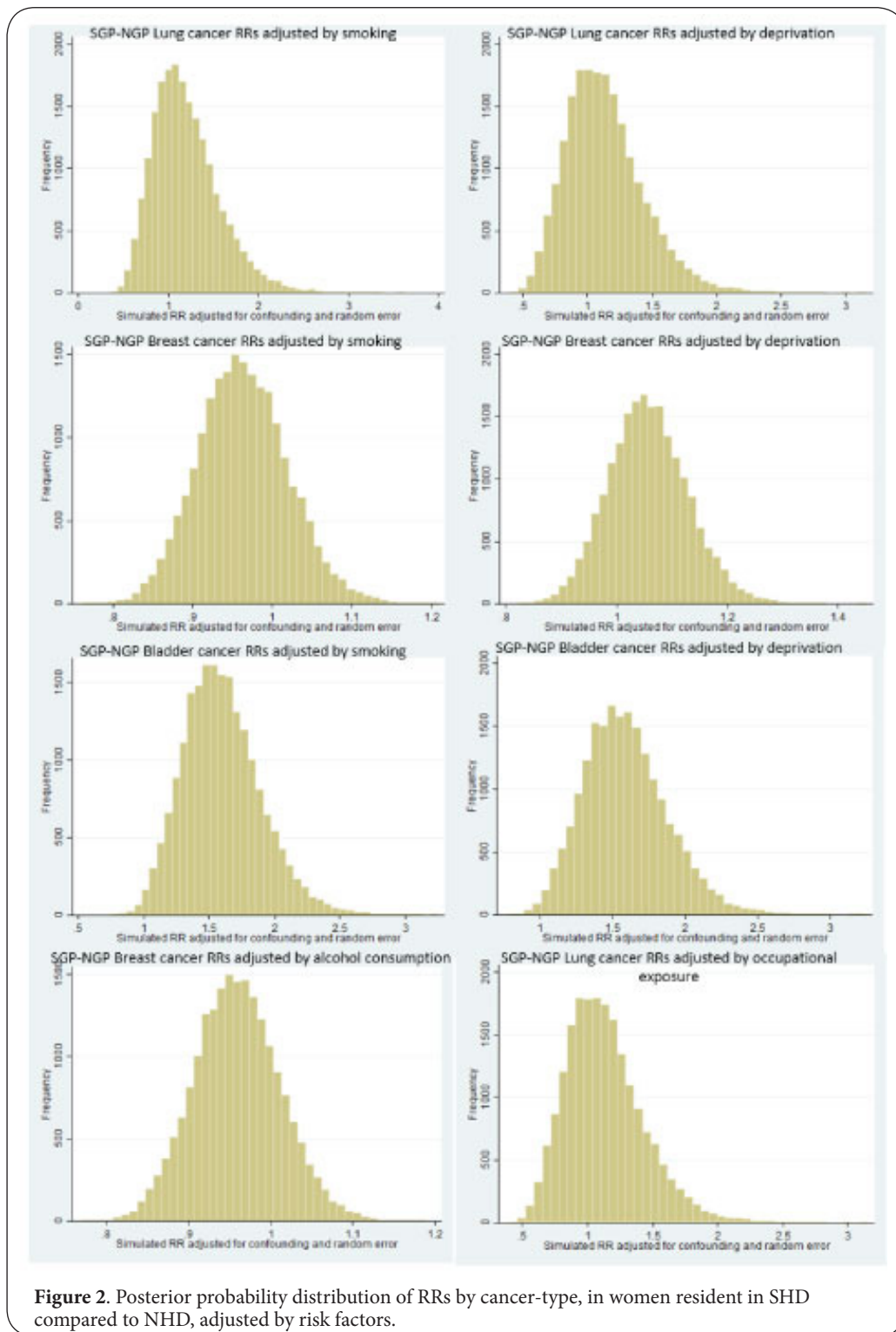
Figure 1. Posterior probability distribution of RRs by cancer-type, in women resident in GP compared to FVG, adjusted by risk factors.

Table 7. Posterior probability distribution of RRs by cancer-type, in women resident in GP compared to FVG, adjusted by risk factors.

Type of adjustment	Lung Cancer (adj. smoking)				Lung Cancer (adj. deprivation)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.73	0.95	1.25	1.73	0.73	0.95	1.25	1.73
Systematic error	0.87	0.96	1.07	1.24	0.95	0.95	0.95	1.00
Systematic and random error	0.71	0.96	1.29	1.81	0.72	0.95	1.25	1.72
Type of adjustment	Breast Cancer (adj. smoking)				Breast Cancer (adj. deprivation)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.98	1.03	1.09	1.12	0.98	1.03	1.09	1.12
Systematic error	1.02	1.03	1.05	1.04	1.02	1.02	1.03	1.02
Systematic and random error	0.97	1.03	1.10	1.13	0.97	1.02	1.08	1.12
Type of adjustment	Bladder Cancer (adj. smoking)				Bladder Cancer (adj. deprivation)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.99	1.20	1.44	1.45	0.99	1.20	1.44	1.45
Systematic error	1.16	1.20	1.23	1.06	1.19	1.19	1.19	1.00
Systematic and random error	0.99	1.20	1.44	1.46	0.99	1.19	1.43	1.45
Type of adjustment	Breast Cancer (adj. alcohol)				Lung Cancer (adj. occupational exposure)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.98	1.03	1.09	1.12	0.73	0.95	1.25	1.73
Systematic error	1.03	1.04	1.04	1.02	0.97	0.98	1.00	1.03
Systematic and random error	0.98	1.04	1.10	1.12	0.75	0.98	1.29	1.73

Table 8. Posterior probability distribution of RRs by cancer-type, in women resident in SGP compared to NGP, adjusted by risk factors.

Type of adjustment	Lung Cancer (adj. smoking)				Lung Cancer (adj. deprivation)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.65	1.09	1.82	2.80	0.65	1.09	1.82	2.80
Systematic error	0.89	1.14	1.53	1.71	1.09	1.10	1.10	1.01
Systematic and random error	0.64	1.15	2.08	3.27	0.65	1.10	1.83	2.80
Type of adjustment	Breast Cancer (adj. smoking)				Breast Cancer (adj. deprivation)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.86	0.95	1.06	1.23	0.86	0.95	1.06	1.23
Systematic error	0.92	0.96	1.01	1.10	0.98	1.05	1.15	1.18
Systematic and random error	0.86	0.96	1.08	1.26	0.92	1.06	1.21	1.31
Type of adjustment	Bladder Cancer (adj. smoking)				Bladder Cancer (adj. deprivation)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	1.09	1.55	2.21	2.03	1.09	1.55	2.21	2.03
Systematic error	1.47	1.57	1.69	1.16	1.56	1.57	1.58	1.01
Systematic and random error	1.09	1.58	2.25	2.06	1.10	1.57	2.23	2.03
Type of adjustment	Breast Cancer (adj. alcohol)				Lung Cancer (adj. occupational exposure)			
	Percentiles		Ratio		Percentiles		Ratio	
	2.5	50	97.5	97.5/2.5	2.5	50	97.5	97.5/2.5
Conventional	0.86	0.95	1.06	1.23	0.65	1.09	1.82	2.80
Systematic error	0.94	0.96	0.98	1.04	1.09	1.09	1.10	1.01
Systematic and random error	0.86	0.96	1.07	1.24	0.65	1.10	1.83	2.80



availability of historical data; b) the level of geographical detail of the information. Especially in the case of cancer, where the induction-latency period can last decades, it is important to have historical information available on the diffusion of behaviors like, for instance, smoking and drinking. Moreover,

the greater the detail (region, province, municipality, census area), the lower the ecological bias related to its use and the closer the measured estimates will be to individual data.

A wider availability of data – for longer periods and with more detailed information – would allow more reliable CRRs

to be obtained. In the light of the systematic and accurate work carried out in particular by cancer registries and of the current availability of frequency indicators for the disease, the evaluation of the weight of confounders is therefore crucial to correctly identify areas with possible risk excesses and to hypothesize the existence of causal relationships with the exposure to pollution sources.

The above-mentioned limits have considerably influenced our estimates since historical data are only available at regional level and, even in the case of recent data, the only study that could be used to calculate smoking and drinking prevalence among women residing in the province of Gorizia, in the Southern and Northern districts and in the rest of FVG was the PASSI survey. Therefore, we assumed that the prevalence differences by area were constant over time. For the deterministic analysis, to simulate different scenarios regarding prevalence variations, the upper and lower limits of the 95% confidence intervals of prevalence were used. The results reported in this paper show the CRR variations at changing of both confounder prevalence and strength of associations between confounder and disease (RR_{CE+}).

The estimate of CRRs made possible to adjust cancer IRRs for different confounders among women residing in the studied areas and showed how these risks changed according to different scenarios based on ecological data.

Although the advantages of the probabilistic sensitivity analysis upon the conventional approach have been already highlighted [2,4-6], it's worth to note that, differently from the deterministic sensitivity analysis, the probabilistic analysis requires the investigator to explicitly define the uncertainty around the bias parameters, through a prior distribution. This prior distribution may arise from available background information as well as from the personal judgment of the researcher. The posterior probability distribution of bias-adjusted estimates reflects this uncertainty about bias.

As another possible limitation, we ran our probabilistic sensitivity analysis only under the hypothesis of uncontrolled confounders assuming no exposure misclassification or selection bias.

Despite all limitations considered, our results seem to suggest that even running several confounding scenarios and probability density functions of the bias parameters, the risk of bladder cancer among women resident in GP and particularly in SGP is higher than in the rest of the Region and this may induce further investigations.

Conclusions

Incidence rates showed an increasing trend over the years for all types of cancers and, in particular, for malignant breast cancer; these variations were seen both in the province of Gorizia and in the rest of the region.

Data on bladder cancer, for the whole period and the 5-year periods, and the respective confidence intervals suggest a higher incidence of this cancer among the women residing in

the province of Gorizia and, in particular, in the SGP. This excess risk arises even after adjustment for different bias parameters.

The use of aggregate data to evaluate the effect of possible bias parameters, through deterministic and probabilistic sensitivity analysis, shows to be promising.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors' contributions	LC	VR	EB	PC	DS	FB
Research concept and design	✓	--	--	--	--	✓
Collection and/or assembly of data	✓	✓	✓	--	--	--
Data analysis and interpretation	✓	✓	✓	--	✓	✓
Writing the article	✓	--	--	--	--	✓
Critical revision of the article	--	--	✓	✓	✓	✓
Final approval of article	✓	✓	✓	✓	✓	✓
Statistical analysis	✓	✓	--	--	--	--

Acknowledgements and funding

The authors are grateful to Lorenzo Richiardi for his valuable suggestions for the design of the study. This research was funded by Friuli Venezia Giulia region (Regional law 27/2012).

Publication history

Editor: Feng Gao, Washington University School of Medicine, USA.
 Received: 18-Oct-2017 Final Revised: 25-Dec-2017
 Accepted: 05-Jan-2018 Published: 27-Jan-2018

References

- Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB and Wynder EL. **Smoking and lung cancer: recent evidence and a discussion of some questions.** *J Natl Cancer Inst.* 1959; **22**:173-203. | [PubMed](#)
- Pandor A, Goodacre S, Harnan S, Holmes M, Pickering A, Fitzgerald P, Rees A and Stevenson M. **Diagnostic management strategies for adults and children with minor head injury: a systematic review and an economic evaluation.** *Health Technol Assess.* 2011; **15**:1-202. | [Article](#) | [PubMed](#)
- Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC and Greenland S. **Good practices for quantitative bias analysis.** *Int J Epidemiol.* 2014; **43**:1969-85. | [Article](#) | [PubMed](#)
- Greenland S. **Multiple-bias modelling for analysis of observational data** Sander. *J. R. Stat. Soc.*, 2005; **168**:267-306.
- Greenland S. **Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment.** *Risk Anal.* 2001; **21**:579-83. | [Article](#) | [PubMed](#)
- McCandless LC and Gustafson P. **A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding.** *Stat Med.* 2017; **36**:2887-2901. | [Article](#) | [PubMed](#)
- Registro Tumori del FVG. 2014. | [Website](#)
- Birri S, Bidoli E and Zucchetto A. **Dati di incidenza , sopravvivenza e prevalenza: aggiornamento al 2007.** 2011.
- Miettinen OS. **Components of the crude risk ratio.** *Am J Epidemiol.* 1972; **96**:168-72. | [PubMed](#)
- Darrow LA and Steenland NK. **Confounding and bias in the attributable fraction.** *Epidemiology.* 2011; **22**:53-8. | [Article](#) | [PubMed](#)
- Miettinen OS. **Stratification by a multivariate confounder score.** *Am J Epidemiol.* 1976; **104**:609-20. | [PubMed](#)
- Miettinen OS and Cook EF. **Confounding: essence and detection.** *Am J Epidemiol.* 1981; **114**:593-603. | [PubMed](#)

13. The PASSI survey. 2015. | [Website](#)
14. National Institute of Statistic. 2015. | [Website](#)
15. Caranci N, Biggeri A, Grisotto L, Pacelli B, Spadea T and Costa G. **Emilia-Romagna, "L'indice di deprivazione italiano a livello di sezione di censimento: definizione, descrizione e associazione con la mortalità.** *Epidemiol. Prev.* 2010; **34**:167-176.
16. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, Scotti L, Jenab M, Turati F, Pasquali E, Pelucchi C, Galeone C, Bellocchio R, Negri E, Corrao G, Boffetta P and La Vecchia C. **Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis.** *Br J Cancer.* 2015; **112**:580-93. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
17. Falk RT, Maas P, Schairer C, Chatterjee N, Mabie JE, Cunningham C, Buys SS, Isaacs C and Ziegler RG. **Alcohol and risk of breast cancer in postmenopausal women: an analysis of etiological heterogeneity by multiple tumor characteristics.** *Am J Epidemiol.* 2014; **180**:705-17. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
18. Nordlund LA, Carstensen JM and Pershagen G. **Cancer incidence in female smokers: a 26-year follow-up.** *Int J Cancer.* 1997; **73**:625-8. | [Article](#) | [PubMed](#)
19. Catsburg C, Miller AB and Rohan TE. **Active cigarette smoking and risk of breast cancer.** *Int J Cancer.* 2015; **136**:2204-9. | [Article](#) | [PubMed](#)
20. Park S, Jee SH, Shin HR, Park EH, Shin A, Jung KW, Hwang SS, Cha ES, Yun YH, Park SK, Boniol M and Boffetta P. **Attributable fraction of tobacco smoking on cancer using population-based nationwide cancer incidence and mortality data in Korea.** *BMC Cancer.* 2014; **14**:406. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
21. Rushton L, Bagga S, Bevan R, Brown TP, Cherrie JW, Holmes P, Fortunato L, Slack R, Van Tongeren M, Young C and Hutchings SJ. **Occupation and cancer in Britain.** *Br J Cancer.* 2010; **102**:1428-37. | [Article](#) | [PubMed Abstract](#) | [PubMed FullText](#)
22. Kuznetsov L, Maier W, Hunger M, Meyer M and Mielck A. **Regional deprivation in Bavaria, Germany: linking a new deprivation score with registry data for lung and colorectal cancer.** *Int J Public Health.* 2012; **57**:827-35. | [Article](#) | [PubMed](#)
23. Mattei F, Guida F, Matrat M, Cenee S, Cyr D, Sanchez M, Radoi L, Menvielle G, Jellouli F, Carton M, Bara S, Marrer E, Luce D and Stucker I. **Exposure to chlorinated solvents and lung cancer: results of the ICARE study.** *Occup Environ Med.* 2014; **71**:681-9. | [Article](#) | [PubMed](#)
24. Barbone F, Filiberti R, Franceschi S, Talamini R, Conti E, Montella M and La Vecchia C. **Socioeconomic status, migration and the risk of breast cancer in Italy.** *Int J Epidemiol.* 1996; **25**:479-87. | [PubMed](#)
25. La Vecchia C, Negri E and Franceschi S. **Education and cancer risk.** *Cancer.* 1992; **70**:2935-41. | [Article](#) | [PubMed](#)
26. Larsen SB, Olsen A, Lynch J, Christensen J, Overvad K, Tjønneland A, Johansen C and Dalton SO. **Socioeconomic position and lifestyle in relation to breast cancer incidence among postmenopausal women: a prospective cohort study, Denmark, 1993-2006.** *Cancer Epidemiol.* 2011; **35**:438-41. | [Article](#) | [PubMed](#)

Citation:

Castriotta L, Rosolen V, Bidoli E, Collarile P, Serraino D and Barbone F. **Cancer occurrence among women resident in the Gorizia province (North-Eastern Italy). May ecological data be used for adjusted epidemiological measures? Deterministic and probabilistic sensitivity analysis.**

J Med Stat Inform. 2018; **6**:1.

<http://dx.doi.org/10.7243/2053-7662-6-1>