Environment International 37 (2011) 577–585

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Prostate cancer and industrial pollution
Risk around putative focus in a multi-source scenario

Rebeca Ramis [a,b,c,*], Peter Diggle [c], Koldo Cambra [b,d], Gonzalo López-Abente [a,b]

[a] Department of Environmental Epidemiology and Cancer, National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain
[b] Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública — CIBERESP), Madrid, Spain
[c] Division of Medicine, Lancaster University, UK
[d] Public Health Department, Basque Country Government, Spain

ARTICLE INFO

Article history:
Received 31 August 2010
Accepted 3 December 2010
Available online 8 January 2011

Keywords:
Prostate cancer
Environmental exposure
Industrial pollution
Metal industry
Non-linear model

ABSTRACT

Background: Prostate cancer is the second most common type of cancer among men but its aetiology is still largely unknown. Different studies have proposed several risk factors such as ethnic origin, age, genetic factors, hormonal factors, diet and insulin-like growth factor, but the spatial distribution of the disease suggests that other environmental factors are involved. This paper studies the spatial distribution of prostate cancer mortality in an industrialized area using distances from each of a number of industrial facilities as indirect measures of exposure to industrial pollution.

Materials and methods: We studied the Gran Bilbao area (Spain) with a population of 791,519 inhabitants distributed in 657 census tracts. There were 20 industrial facilities within the area, 8 of them in the central axis of the region. We analysed prostate cancer mortality during the period 1996–2003. There were 883 deaths giving a crude rate of 14 per 100,000 inhabitants.

We extended the standard Poisson regression model by the inclusion of a multiplicative non-linear function to model the effect of distance from an industrial facility. The function's shape combined an elevated risk close to the source with a neutral effect at large distance. We also included socio-demographic covariates in the model to control potential confounding.

Results: We aggregated the industrial facilities by sector: metal, mineral, chemical and other activities. Results relating to metal industries showed a significantly elevated risk by a factor of approximately 1.4 in the immediate vicinity, decaying with distance to a value of 1.08 at 12 km. The remaining sectors did not show a statistically significant excess of risk at the source.

Discussion: Notwithstanding the limitations of this kind of study, we found evidence of association between the spatial distribution of prostate cancer mortality aggregated by census tracts and proximity to metal industrial facilities located within the area, after adjusting for socio-demographic characteristics at municipality level.

1. Introduction

Prostate cancer is the second most common type of cancer among men, with higher risk in people aged over 65 years. Its incidence is increasing throughout Europe (Ferlay et al., 2007). However, increases in the late 1980s and 1990s are also associated with the use of the prostate-specific antigen (PSA) test to screen for prostate cancer (Post et al., 1998). The aetiology of the disease is still largely unknown; however, different studies have proposed several risk factors such as ethnic origin, age, genetic factors, hormonal factors, diet and insulin-like growth factor (Grönberg, 2003). Alternatively, occupational studies suggest possible associations with farming, metal working and the rubber industry, and with exposure to hazardous substances such as pesticides, cadmium and polycyclic aromatic hydrocarbons (Parent and Siemiatycki, 2001). The spatial patterns shown in cancer atlases suggest the possibility that other environmental factors may be involved; see López-Abente (2006) (http://www.isciii.es/htdocs/centros/epidemiologia/libros/Atlas_municipal.pdf) and Martinez-Beneito et al. (2005) (http://www.sp.san.gva.es/epidemiologia/atlas).

One such risk factor could be exposure to industrial pollution, but evidence regarding the health risk of living near to polluting factories is limited. Some authors have described associations of lung cancer with metallurgical and other industries (Cambra et al., 2010; Monge-Corella et al., 2008; Parodi et al., 2005; Gottlieb and Carr, 1982). Also, lymphomas and leukaemia are more common near to industrial areas (De Roos et al., 2010; Benedetti et al., 2001; López-Abente et al., 1999; Sharp et al., 1996; Sans et al., 1995). However, other studies have not found any association between particular cancers and proximity to

* Corresponding author. Department of Medicine, Faraday Building, Lancaster University, Lancaster, LA1 4YB, United Kingdom.
E-mail addresses: rramis@isciii.es (R. Ramis), p.diggle@lancaster.ac.uk (P. Diggle), ki.cambra.contin@navarra.es (K. Cambra), glabente@isciii.es (G. López-Abente).

0160-4120/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envint.2010.12.001

industrial facilities and incinerators (Michelozzi et al., 1998; Pekkanen et al., 1995).

Assessment of exposure to environmental agents that are noxious to human health is complicated by the variety of possible exposure measurement strategies, including the use of remote sensors, biomarkers, or estimates of pollutant dispersion using theoretical or statistical models (Nieuwenhuijsen et al., 2006). Studies of elevation in risk close to potentially hazardous sites include Stone (1988), Elliott (1996), Biggeri (1996), Viel (2000), Draper (2005) and Maule (2007). This paper studies the spatial distribution of prostate cancer mortality in an industrialized area using distance, from each of a number of industrial facilities, to provide indirect measures of exposure to industrial pollution.

2. Materials and methods

To assess the possible association between prostate cancer mortality and environmental exposure to industrial pollution we studied Gran Bilbao, the biggest urban area of the Basque Country, sited in the north of Spain. This region is one of the most industrialized parts of Spain. Specifically, there are 20 industrial facilities located within the area that were included in the European Pollutant Emission Register (EPER) (EPER, 2008).

2.1. Data

According to the 2001 official census, the population of Gran Bilbao was 791,519 inhabitants, distributed between 17 municipalities including the city of Bilbao, and 657 census tracts. The total area was 204 km^2, hence the population average density was 3880 inhabitants/km^2.

2.1.1. Cases: mortality data

The mortality data were the number of deaths during the period 1996–2003, in each of the 657 census tracts. In this period there were 883 deaths attributed to prostate cancer with a crude rate per 100,000 inhabitants of 111.55. The populations of the individual census tracts varied between 1000 and 2000 inhabitants.

2.1.2. Expected cases

We calculated the expected number of cases in each census tract using indirect standardization as follows. Firstly, we stratified the population by age and sex, producing 18 strata. Then, in each stratum we obtained the number of person-years by multiplying the population of the reference year 2001 by the number of studied years, namely eight. Finally, we multiplied the overall Spanish mortality rate for prostate cancer in each stratum by the number of person-years and added over strata to give the total number of expected cases in each census tract (Barceló et al., 2008).

2.1.3. Socio-demographic covariates

According to Anand the main risk factor for cancer mortality is lifestyle, including diet, tobacco and obesity (Anand et al., 2008). Another main factor is age. We included age in our models by the use of indirect standardization as described above. Unfortunately there was no available information on lifestyle for our study; however lifestyle is related with socio-demographic characteristics of the population. We used the 1991 Spanish census as our source of this socio-demographic information; although this information was aggregated to the level of municipality rather than census tract we considered it more important to recognise the well-known long latency period of prostate cancer, rather than to exploit the finer spatial resolution offered by more recent census data.

Our municipal-level socioeconomic covariates were *percentage of illiterates*, *percentage of unemployed* and *cohabitants per house* from the 1991 census, and *average income in 1991* provided by the Spanish Credit Bank (BEC, 1993). All of these covariates were standardized at national level. Ideally we would have included prevalence of tobacco smoking as a covariate but, as mentioned above, such information was not available at the required level of spatial aggregation. Therefore, we used the Standardized Mortality Ratio of lung cancer as an approximation to *tobacco smoking prevalence* (López-Abente et al., 2006a,b; Best and Hansell, 2009).

Summarizing, the five socio-demographic covariates included in our analyses were: percentage of illiterates (*Education (−)*), percentage of unemployed (*unemployed*), cohabitants per house (*cph*), income (*income*) and SMR of lung cancer (*tobacco*).

2.1.4. Factories

As our source of information about the industrial facilities we used the European Pollutant Emission Register (EPER) (EPER, 2008). The information available allowed us to identify six types of industrial activity, namely: 1) energy industries; 2) production and processing of metals; 3) mineral industry; 4) chemical industry and chemical installations; 5) waste management; and 6) other activities, including paper and board production, manufacture of fibres or textiles, tanning of hides and skins, slaughterhouses, intensive poultry or pig rearing, installations using organic solvents, and the production of carbon or graphite. Twelve of the 20 factories so identified were located far from all residential areas. We therefore considered only the remaining eight factories that were located along a central axis of the area; three metal factories (smelting), one mineral factory (glass), three chemicals factories (fertilizer, pharmaceutical products, plastic products) and one from the "other activities" group, specifically a food factory (brewery). Our exposure variables were defined as the eight distances between the centroid of the census tract and the location of each factory. Finally, we assumed common parameters associated within each of the four industrial categories so as to provide more precise estimation of model parameters. Fig. 1 shows Gran Bilbao by census tract with the locations of the eight factories represented by circles of different colours according to their industrial categories.

2.2. Model

In epidemiology, the standard method for analysing aggregated data on chronic disease is ecological regression, typically implemented through a Poisson log-linear model. To analyse the effect of the exposure to industrial pollution on the spatial distribution of cancer mortality the log-linear formulation is unrealistic because of the need to combine an elevated risk close to the source with a neutral long-range effect; therefore, we extended the model by the inclusion of a multiplicative non-linear distance effect as proposed by Diggle et al. (1997) for a single source, to give

$$O_i \sim Po(E_i\mu_i)$$
$$\mu_i = \rho \exp[\textstyle\sum_k (\vartheta_k Z_{ik})] \prod_j f\left(d_{ij}\right); \quad f\left(d_{ij}\right) = 1 + \alpha_j \exp\left[-\left(d_{ij}/\beta_j\right)^2\right]. \tag{1}$$

In Eq. (1), ρ is related to the overall risk, θ_k are the parameters of the socio-demographic covariates Z_{ik}; α_j and β_j are the parameters of the distance function, and d_{ij} is the distance between the centroid of the area i and the focus j. The exposure-related parameters α_j and β_j are assumed to take common values within each of the four industrial categories.

The approximate log-likelihood function for this model, up to a constant term, is (Diggle et al., 1997):

$$L(\rho, \vartheta, \alpha, \beta) = -\sum_i \mu_i + \sum_i O_i \log(\mu_i). \tag{2}$$

A detailed description of likelihood-based parameter estimation and model-fitting is given in Appendix A.

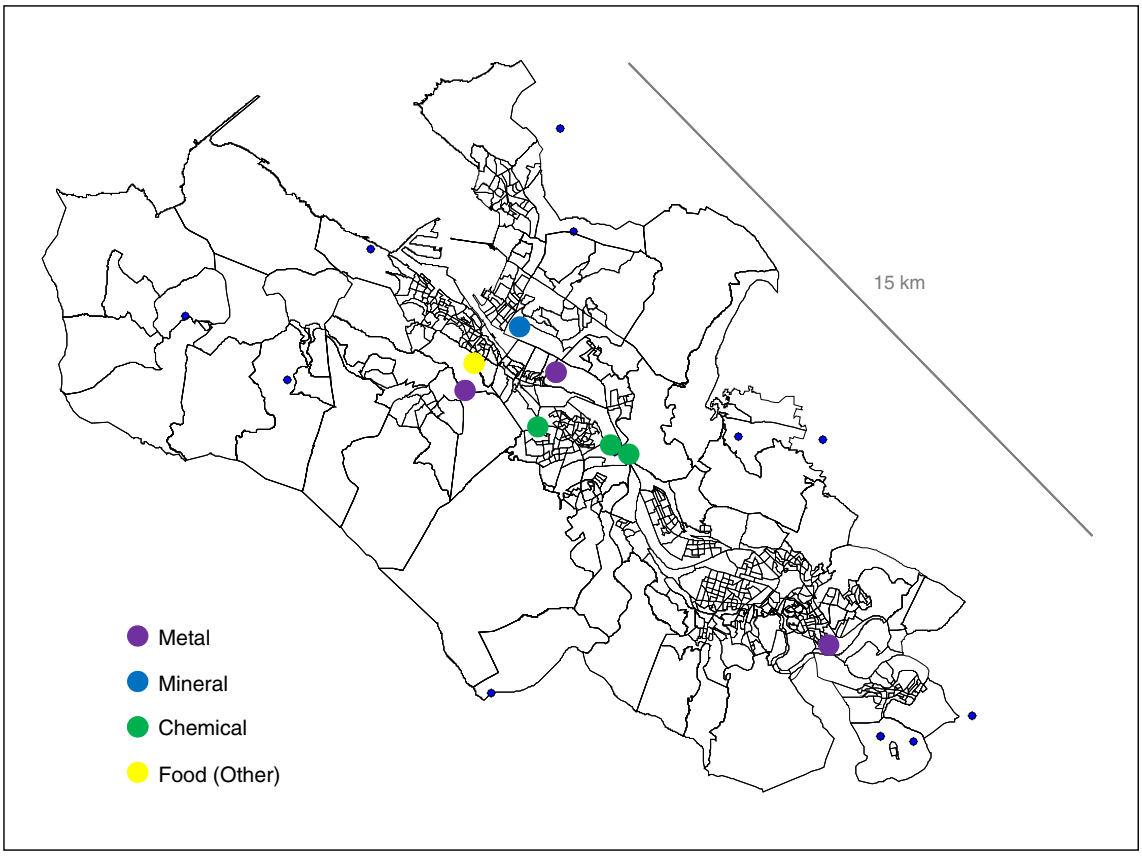

Fig. 1. Gran Bilbao by census tract. Factory locations are shown as circles colour-coded according to industrial categories: purple for metal industry, blue for mineral industry, green for chemical industry and yellow for other activities (food); factories not included in the analysis are shown as small blue dots.

Because of the potential for confounding between socio-demographic factors and proximity to industry, we adopted the conservative strategy of including adjustments for all five socio-demographic variables (*Cov*) before exploring the possible distance-based effects.

3. Results

There were a total of 883 deaths attributed to prostate cancer in Gran Bilbao during the period 1996–2003 (mean 1.34 cases per census tract, standard deviation 1.36). This represent highly significant over-dispersion by comparison with a constant-mean Poisson distribution (chi-squared = 902.8 on 656 degrees of freedom, $p<0.001$).

Table 1 shows the sequence of fitted models with the different distance variables included in each, together with the values of the maximised log-likelihood test and the p-values of the likelihood ratios for each of the models 2 to 14 against Model 1 and Model 3. There was evidence that the inclusion of the distance variables in Model 14 gave a significant improvement over Model 1. The value of the D statistic was 10.42 with a P-value of 0.03. With regard to remaining models, there seemed to be no significant improvement over Model 1 since all P-values are larger than 0.05.

The results for the socio-demographic covariates in our final model showed estimated relative risks bigger than one for income, 1.26 (0.97–1.51); percentage of illiterates (education), 1.19 (0.84–1.69); and RR of lung cancer (tobacco prevalence), 1.08 (1.06–1.10); and estimated risk below one for percentage of

Table 1
Log-likelihood, test statistic, df and P-value for a χ^2_n for the multiple regressions.

Models	Covariates	Log-likelihood	Reference model = Model 1			Reference model = Model 3		
			Test statistic	df	P-value	Test statistic	df	P-value
Model 1	Cov	−537.88						
Model 2	Cov + Metal	−537.81	0.13	1	0.70			
Model 3	Cov + Chemical	−536.43	2.90	1	0.09			
Model 4	Cov + Metal + Mineral	−535.65	4.46	2	0.10			
Model 5	Cov + Metal + Chemical	−535.39	4.97	2	0.08	2.08	1	0.15
Model 6	Cov + Metal + Others	−537.61	0.54	2	0.76			
Model 7	Cov + Chemical + Mineral	−536.61	3.15	2	0.21	−0.36	1	1.00
Model 8	Cov + Chemical + Others	536.04	3.68	2	0.16	0.78	1	0.38
Model 9	Cov + Mineral + Others	−537.86	0.04	2	0.98			
Model 10	Cov + Metal + Mineral + Others	−537.61	0.40	3	0.94			
Model 11	Cov + Metal + Chemical + Others	−535.37	5.03	3	0.17	2.08	2	0.35
Model 12	Cov + Metal + Mineral + Chemical	−534.89	5.98	3	0.11	3.08	2	0.21
Model 13	Cov + Mineral + Chemical + Others	−535.84	4.06	3	0.25	1.18	2	0.55
Model 14	Cov + Metal + Mineral + Chemical + Others	−532.67	10.42	4	0.03	7.52	3	0.05

Cov = Socio-demographic covariates: percentage of illiterates (*Education (−)*), percentage of unemployed (*unemployed*), cohabitants per house (*cph*), income (*income*) and SMR of lung cancer (*tobacco*).

Table 2
Risk distance functions for the industrial sector.

Industrial area	Distance function
Metal	$1+0.4*\exp(-(\text{dist}/10)^2)$
Mineral	$1+0.42*\exp(-(\text{dist}/1810)^2)$
Chemical	$1-0.18*\exp(-(\text{dist}/4)^2)$
Other (food)	$1+13.2*\exp(-(\text{dist}/0.2)^2)$

unemployed, 0.67(0.51–0.95); and cohabitants per house, 0.48(0.32–0.73). The estimated relative risks associated with distance from each factory type are presented in Table 2 and, in a more easily interpretable graphical form, in Fig. 2.

Fig. 2 shows the estimated risk functions and 95% Monte Carlo confidence limits (as described in Appendix A) for the effects of distance from each type of factory. Fig. 2a, relating to metal industries, shows a significantly elevated risk by a factor of approximately 1.4 in the immediate vicinity, decaying with distance to a value of 1.08 at 12 km, whilst the confidence interval includes the null value, 1, at distances above 10 km. In contrast, the risk function relating to distance from a chemical factory (Fig. 2b) has a positive slope with a risk of 0.83 at distance 0 increasing until 1 at 6 km, but with a confidence interval that includes the value 1 for distances greater that 3 km. Fig. 2c shows the estimated risk function relating to distance from a mineral factory to be approximately constant within the plotted range. Finally, the estimated risk function associated with the food factory (Fig. 2d) falls from 2.7 to 1 within 0.5 km, but the confidence interval includes the value 1 at all distances.

4. Discussion

4.1. Prostate cancer

Our analysis suggests an association with distance to some types of factories, i.e. metal and chemical facilities. The estimated function associated with proximity to metal industries decayed with distance from a starting value of 1.4, reaching a value not significantly different from 1.0 at a distance of 10 km. For chemical factories the estimated function associated with proximity increased with distance from a starting value of 0.82 and reached a value not significantly different from 1.0 at a distance of 3 km. The estimated risk functions for the remaining industrial categories, mineral and other (food), did not show association between distance and prostate cancer risk. The estimated risk function for mineral industry was flat, whereas the estimated function for the food factory showed a steep slope but was nevertheless not significantly different from constant. These results suggest that these two industrial categories are not associated with prostate cancer risk. However, unlike the full model (Model 14), the model without those two covariates (Model 5) was not significantly better that the baseline model (Model 1). This could reflect confounding due to unmeasured variables such as diet, genetic or hormonal factors (Rothman, 2002).

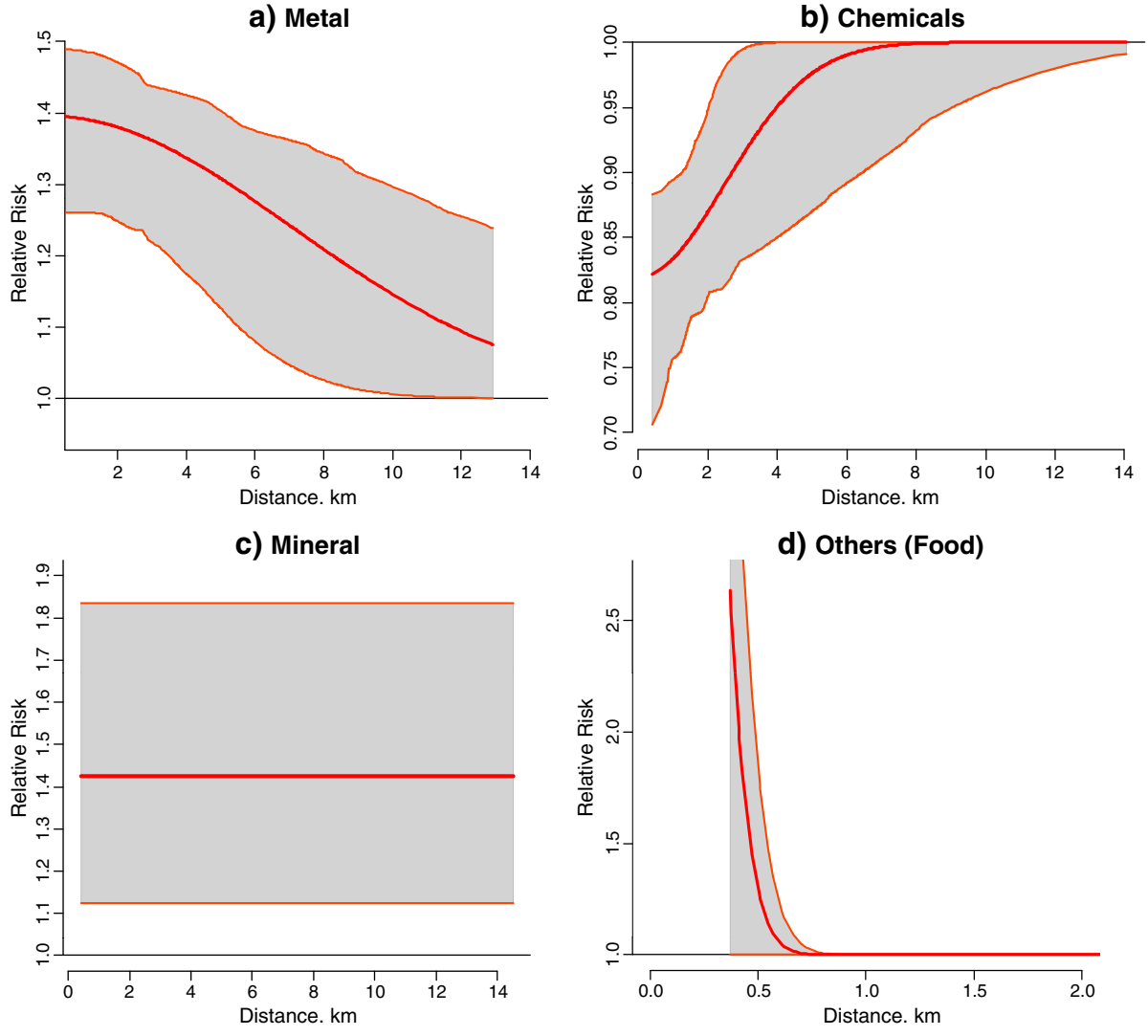

Fig. 2. Risk functions and 95% Monte Carlo confident limits for the distance to: a) metal factories, b) mineral, c) chemical and d) other (food).

The shape of the estimated risk function for chemical industry could be considered counter intuitive. To explain it we consider the geographical position of the chemical factories. Two of these were located close together in the same industrial estate, whereas the third was in the same area. During the nineties and early 00s the zone changed substantially, some factories disappeared and new residential areas were constructed originating the migration of new population. Thus, the resulting risk function is prone to socio-demographic bias and not reliable to explain the effect of exposure pollution from these chemical industries over prostate cancer mortality. Its estimated shape is probably related to the socio-demographic characteristics of the incoming population, particularly to age which is a well-known risk factor (Grönberg, 2003). Although we used indirect standardization to control the age effects this technique does not always eliminate them completely (Miettinen, 1972).

In relation to metal industry an examination of the information contained in the EPER for 2001 showed that the three factories reported emissions above the thresholds that determined their inclusion in the registry for the following compounds: lead, cadmium, zinc, chrome, copper, chlorine and PM10 particulate matter. Taken individually, these industries also reported emissions of CO, nitrogen dioxide, sulphur dioxide, arsenic, mercury, nickel, benzene and polycyclic aromatic hydrocarbons (PAH). IARC classifies as carcinogen 1 group: cadmium, chrome, nickel, arsenic and benzene; and as carcinogen 2A group: lead and copper (IARC, 1993, 1987); among them, cadmium and arsenic are specific carcinogens for prostate cancer. Furthermore, a recent Italian case–control study identifies cadmium as a possible risk factor for prostate cancer (Vinceti et al., 2007). In addition, several occupational studies have reported significant associations between metal industries and prostate cancer. A recent study of aluminium smelter workers showed an excess of mortality risk in production workers but not in office workers (Sim et al., 2009). Another study linked prostate cancer with exposure to metalworking fluids (Agalliu et al., 2005). Finally, an early review of prostate cancer in metal workers and repairmen showed a slightly elevated risk of prostate cancer for these occupations (Gulden, 1997). None of the other facilities reported emissions on heavy metals.

We did not attempt a detailed interpretation of the risk estimates associated with the socio-demographic covariates because, as noted earlier, the spatial aggregation level for these covariates was the municipality rather than the census tract. For the same reason, different aggregation levels, we did not exclude *Income* and *Education* for the model despite their not being statistically significant. Instead, we retained these variables in our baseline model as the best available information associated with lifestyle to control for the possible confounding of distance-based effects with potential lifestyle risk-factors for prostate cancer mortality.

4.2. Methodological issues

This is the first study to analyse the relation between the spatial distribution of prostate cancer mortality in Spain and exposure to industrial pollution, using a distance-based empirical model as a surrogate for exposure at small-area (census tract) level. Thus, it is important to discuss some methodological issues.

According to the literature, cancer's main risk factor is lifestyle. The major contributions to overall cancer risk are: diet 30–35%; tobacco, 25–30%; and obesity, 10–20%. The contribution of infections is estimated to be 15–20%. Finally, genetic predisposition contributes 5–10% (Anand et al., 2008). However, the interaction between these factors may be very important in the development of the disease. In the present study, only ecological data about the socio-demographic status of the population and estimated exposure to industrial pollution has been included, hence potentially important individual-level variation is not being taken into account. Furthermore, only mortality data are included and our analysis necessarily excludes prevalent cases. This lack of information about non-lethal cancer cases may bias the analysis; on the other hand, in Spain tumours with lower survival rates are well represented using death certificates according to Pérez-Gómez and Aragonés, specifically for prostate cancer death certificates have high confirmation and detection rates (Pérez-Gómez et al., 2006). The overall five-year survival rate for prostate cancer in Spain is 77% and we believe there are at most small differences in survival rates or quality of care between regions due to in universal health system established in Spain since 1986.

Another limitation is our use of the centroid of each census tract as the reference point to calculate distances from industrial locations. This will bias the results if there is substantial variation in risk within census tracts (Diggle and Elliott, 1995); note, however, that the average area of a census tract is approximately 0.36 km^2, whereas our estimated exposure effects extend over several kilometres, hence this source of bias is likely to be small. Furthermore, the use of small areas as units in an ecological study reduces the risk of ecological bias.

As in all studies of this kind, we have assumed that place of current residence determines the estimated exposure, hence no allowance is made either for long-term movements between different addresses or short-term movements between home and work. In particular, we have acknowledged above that major migration over the relevant time period may have introduced substantial bias into the estimated risk function for proximity to chemical industry. Also, we have not been able to include other sources of environmental pollution such as traffic or indoor pollution. Exposure to such pollution can contribute to the development of cancers (Belpomme et al., 2007). For example, substances such as polycyclic aromatic hydrocarbons produced by combustion of organic fuels are considered as mutagens (IARC, 1989), whilst indoor pollutants such as benzene or other volatile organic compounds are rated as carcinogens (IARC, 1995b). We have also not been able to include either information on occupational exposures even though we have mentioned several associations between occupational exposures and prostate cancer.

Finally, we note that a new European Pollutant Release and Transfer Register (E-PRTR) replaced the EPER in 2009, allowing enhancement of the validity of a study of this type, with the possibility of evaluating the effects of specific pollutants.

4.3. Model

The model was initially developed by Diggle and Rowlingson (1994) to fit individual-level data with reference to a single point source, and later adapted to be used with aggregated data, such us census tract data by Diggle et al. (1997). That paper also gave the theoretical formulation for multiple sources, but did not include an empirical example. Other authors have formulated different models to address the multi-source problem. Congdon (2003) proposed a model for individual outcomes which uses a similar distance function, but suggested an additive rather than a multiplicative form for the combined effects of multiple sources. Lawson (2001) and Dunn et al. (2001) extended purely distance-based models to include directional effects. In our application, the data were too sparse to justify the additional number of parameters that would be required to fit directional effects; a directional model would require at least two additional parameters for each source, one to identify the principal direction and a second to capture the degree of concentration about the principal direction. A more basic concern, not unique to our study, is the use of distance from the source as a surrogate for exposure. Physically based plume models attempt to address this but typically rely on detailed modelling assumptions that cannot be validated empirically. In our opinion, a better solution in principle is to seek to identify and adjust for spatially varying explanatory variables that can account for spatial variation in exposure that is not adequately described by a relatively simple function of distance; in practice, this

strategy is limited by the number of candidate explanatory variables available. For data that are more abundant and/or more finely resolved than we have in our application, other possibilities are to model the variation in risk around a point source non-parametrically (Rodrigues et al., 2010), or to use geographical information systems to derive spatially continuous explanatory variables. An example of the latter is Fanshawe et al. (2008), who used a map of the local density of domestic chimneys as a surrogate for exposure to atmospheric pollution derived from coal-burning on open fires.

5. Conclusions

Notwithstanding the limitations acknowledged in the discussion, we have found evidence of association between the spatial distribution of prostate cancer mortality aggregated by census tracts and possible exposure to pollutant substances from the metal industrial facilities located within the study area, as expressed by the variation in risk with distance from particular types of industrial location at census tract level after adjusting for socio-demographic characteristics at municipality level.

Acknowledgments

We thank the Basque Government for providing census tract mortality and geographical data needed to carry out this study.

This study was funded by Grant No. FIS: PI040041 from the Carlos III Institute of Health.

Appendix A

Parameter estimation

Parameters were estimated by direct maximisation of the likelihood function using the numerical optimization function "optim", included in the R software (www.r-project.org).

Standard error calculations

A common way to approximate the standard errors of maximum likelihood parameters estimates is through the inverse of the Hessian matrix. Most optimization algorithms, including "optim", provide a numerical approximation to the Hessian matrix, but we found that for point source models of the kind used in this paper, whilst numerically accurate values were returned for the maximum likelihood parameter estimates the associated standard errors derived by inverting the estimated Hessian could be unreliable. We therefore obtained standard errors by combining the R function for direct maximisation of the likelihood with replicated Monte Carlo simulations of the fitted model. Table A1 presents results of an experiment in which we simulated a dataset of observed cases from a model with four socio-demographic covariates and two pollutant sources. The left-most column gives the true values of the parameters. The next columns give different standard errors; the first of these (*Hessian*) uses approximate Hessian matrix, the second (*Monte Carlo*) uses the Monte Carlo method as described above, whilst the third uses Monte Carlos method but with true, rather than estimated, parameter values. The Monte Carlo estimates are unbiased by design, and the 95% confidence limits also shown in the table give an indication of the size of the Monte Carlo error. The results therefore suggested that the Hessian-based standard errors could be unreliable, as in some cases the claimed standard errors lie outside the 95% confidence limits from the simulations. These results were in qualitative agreement with those from earlier studies (Diggle and Rowlingson, 1994; Diggle et al., 1997).

Table A1
Given values, Hessian standard errors, Monte Carlo standard errors using estimated parameter values and standard errors using true parameters.

	Values	Hessian standard errors	Monte Carlo standard errors using estimated parameter values	Standard errors using true parameters
ρ	0.019	0.245	0.216 (0.196–0.239)	0.160 (0.146–0.177)
θ_1	0.111	0.263	0.254 (0.231–0.282)	0.172 (0.156–0.190)
θ_2	0.099	0.125	0.129 (0.117–0.143)	0.100 (0.091–0.110)
θ_3	−0.020	0.101	0.089 (0.081–0.099)	0.066 (0.060–0.073)
θ_4	−0.093	0.100	0.098 (0.089–0.108)	0.074 (0.067–0.082)
α_1	0.100	0.192	0.270 (0.246–0.299)	0.238 (0.216–0.263)
α_2	0.100	0.180	0.208 (0.189–0.230)	0.179 (0.162–0.198)
β_1	0.200	0.123	0.174 (0.159–0.193)	0.232 (0.211–0.257)
β_2	0.400	1.149	0.395 (0.360–0.438)	0.340 (0.308–0.375)

Hypotheses testing. Approximate null distribution of likelihood ratio statistic D

To test the hypotheses about the parameters we used the usual generalized likelihood ratio test statistic;

$$D = 2\left\{L\left(\hat{\phi}\right) - L_0\left(\hat{\phi}_0\right)\right\}$$

where L_0 and L denote the maximised log-likelihoods under the null and alternative hypotheses, respectively.

Previous studies have pointed out that the usual asymptotic properties of the likelihood ratio test may not hold for the models considered in this paper because the distance function is such that when $\alpha = 0$, β is indeterminate (Diggle and Rowlingson, 1994; Diggle et al. 1997). To clarify this point in the context of our application, we ran a simulation experiment generating data from the following models:

1. Null model: $\mu_i = \rho$
2. Distance model: $\mu_i = \rho \prod_j f(d_{ij})$.

For the distance model we contemplated two scenarios, one with three point sources and one with four. For each scenario we ran 100 simulations and calculated the corresponding likelihood ratio statistic D. We considered as candidate reference distributions for the likelihood ratio test χ^2_n and χ^2_{2n}, where n is the number of point sources in the distance model. For the first scenario, Fig. A1 shows three graphs: (a) a density graph of the empirical distribution of D and the two reference distributions; (b) a Q–Q plot of sample D-values against χ^2_3; (c) a Q–Q plot of sample D-values against χ^2_6 (c). In Fig. A1(a), the empirical density of the D-statistic almost overlaps the density of χ^2_3; whereas, the shape of the χ^2_6 density is clearly different from the empirical density. In the QQ-plot for χ^2_3 almost all points are close to the main diagonal, whereas in the second QQ-plot they lie clearly below the main diagonal. Finally, results of formal Kolmogorov–Smirnov goodness-of-fit tests were consistent with the graphical evidence, giving p-values $p = 0.477$ and $p < 0.001$ against the reference distributions χ^2_n and χ^2_{2n}, respectively.

Fig. A2 shows the corresponding graphs for the scenario with four point sources. As in the previous case, the results of the tests supported the use of the χ^2_n distribution, where now $n = 4$, as the reference distribution for the likelihood ratio test statistic. Kolmogorov–Smirnov goodness-of-fit tests gave p-values $p = 0.871$ and $p < 0.001$ against the reference distributions χ^2_n and χ^2_{2n}, respectively.

The results of the simulation experiment suggest that the standard χ^2_{2n} approximation to the null sampling distribution of the generalized likelihood ratio statistic D does not hold, whereas the χ^2_n distribution gives a good approximation. In previous papers (Diggle and Rowlingson, 1994; Diggle et al., 1997) χ^2_{2n} was used as

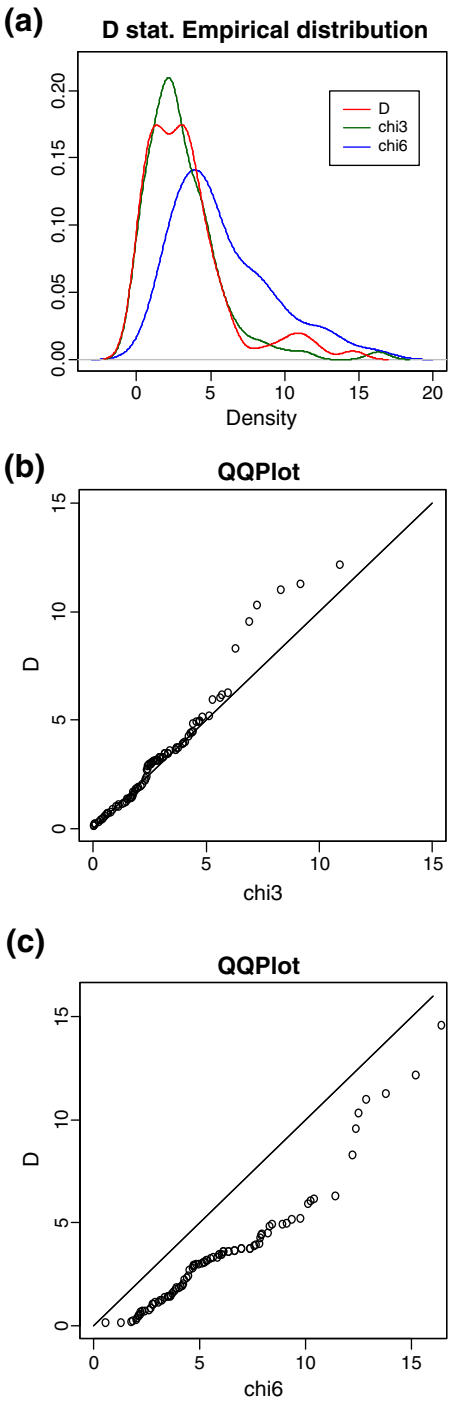

Fig. A1. Comparison between the empirical distribution and the reference distributions in a scenario with three point sources (a). Density graph of the empirical distribution and the two reference distributions); (b) Q–Q plot of sample D-values against χ^2_3; (c) Q–Q plot of sample D-values against χ^2_6.

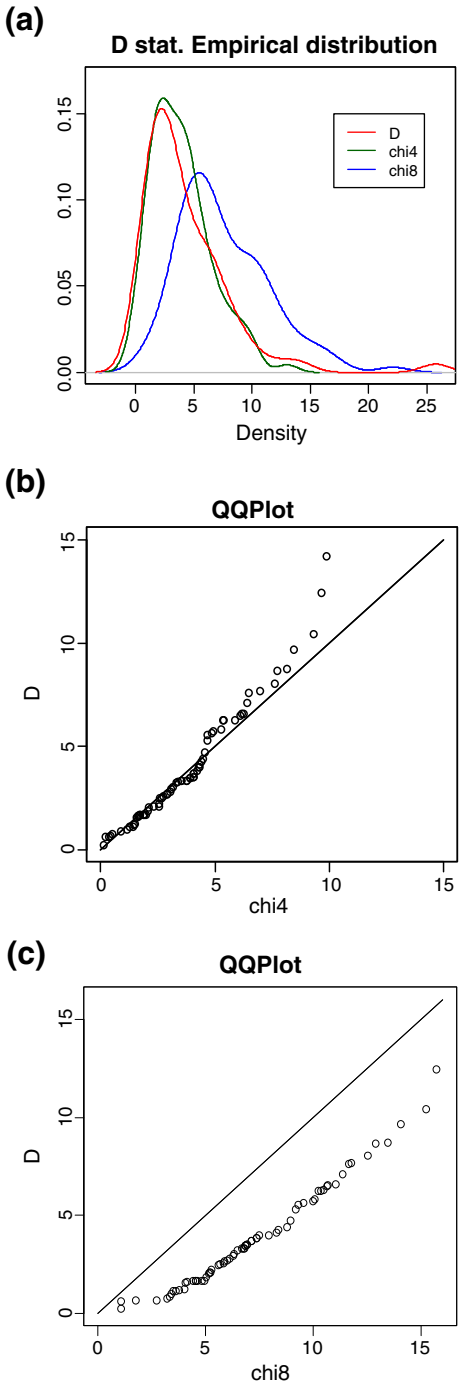

Fig. A2. Comparison between the empirical distribution and the reference distributions in a scenario with four point sources (a) density graph of the empirical distribution and the two reference distributions); (b) Q–Q plot of sample D-values against χ^2_4; (c) Q–Q plot of sample D-values against χ^2_8.

the distribution of the generalized likelihood ratio statistic, This non-standard behaviour of the likelihood ratio test can be attributed to the indeterminacy of β under the null hypothesis $\alpha = 0$, hence the test can be considered to have one effective degree of freedom per source, rather than two. Some theoretical results concerning the failure of the standard asymptotic properties of the generalized likelihood ratio test in non-regular problems are given in Cheng and Traylor (1994).

Residual analysis

Fig. A3 plots deviance residuals for our preferred model against the distance to the nearest factory in each of the four industrial categories. All four cases show an apparently random scatter, consistent with a well-fitting model.

References

Agalliu I, Eisen E, Kriebel D, Quinn M, Wegman D. A biological approach to characterizing exposure to metalworking fluids and risk of prostate cancer (United States). Cancer Causes Control 2005;16:323–31.

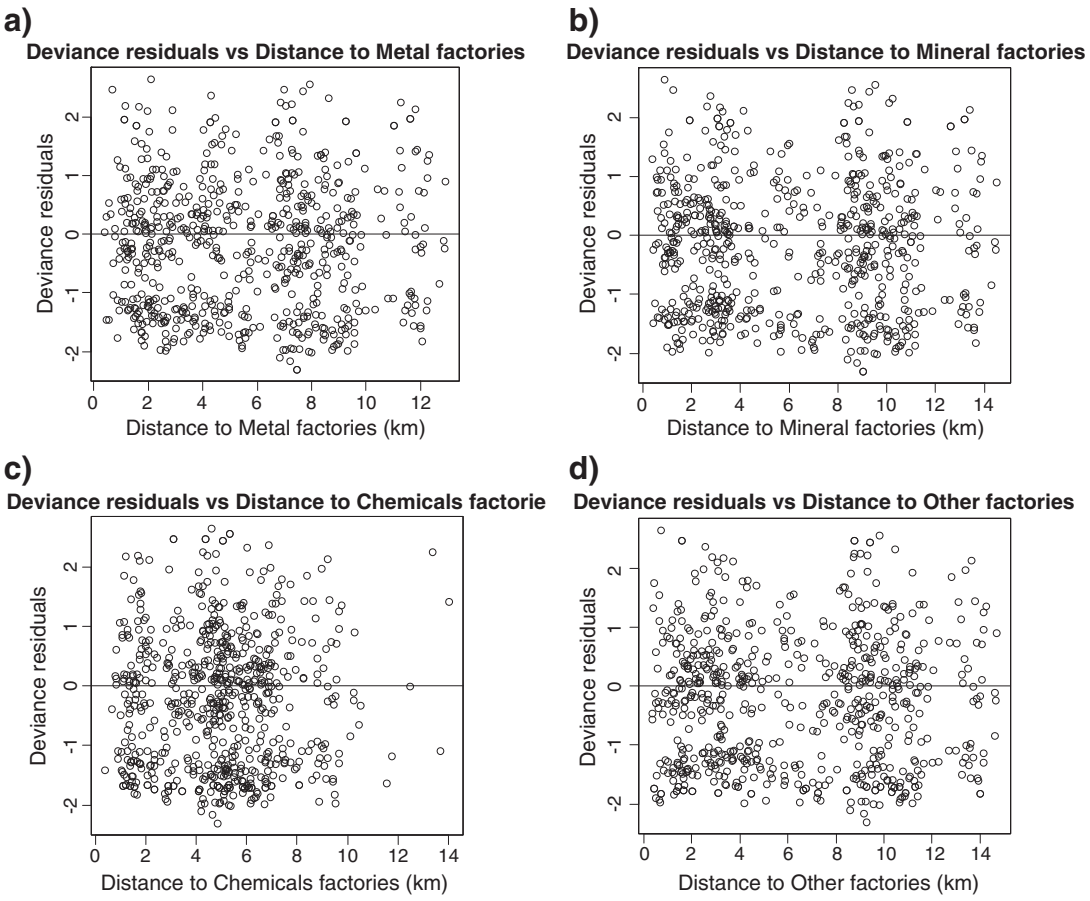

Fig. A3. Deviance residuals vs distance to the nearest factory in each industrial category: (a) metals; (b) minerals; (c) chemicals; and (d) food.

Anand P, Kunnumakara A, Sundaram C, Harikumar K, Tharakan S, Lai O, et al. Cancer is a preventable disease that requires major lifestyle changes. Pharm Res 2008;25(9): 2097–116.

Banco Español de Credito. 1993. Banco Español de Credito (BEC): Anuario del Mercado Español. 1993. Madrid.

Barceló MA, Saez M, Cano-Serral G, Martínez-Beneito MA, Martínez JM, Borrell C, et al. Methods to smooth mortality indicators: application to analysis of inequalities in mortality in Spanish cities [the MEDEA Project]. Gac Sanit 2008;22(6):596–608 Spanish.

Belpomme D, Irigarayb P, Hardellc L, Clappd R, Montagniere L, Epsteinf S, et al. The multitude and diversity of environmental carcinogens. Environ Res 2007;105:414–29.

Benedetti M, Iavarone I, Comba P, Lavarone I. Cancer risk associated with residential proximity to industrial sites: a review. Arch Environ Health 2001;56:342–9.

Best N, Hansell A. Geographic variations in risk. Adjusting for unmeasured confounders through join modelling of multiple diseases. Epidemiology 2009;20:400–10.

Biggeri A, Barbone F, Lagazio C, Bovenzi M, Stanta G. Air pollution and lung cancer in Trieste, Italy: spatial analysis of risk as a function of distance from sources. Environ Health Perspect 1996;104:750–4.

Cambra K, Martínez-Rueda T, Alonso-Fustel E, Cirarda F, Ibañez B, Esnaola S, et al. Mortality in small geographical areas and proximity to air polluting industries in the Basque Country (Spain). J Occup Environ Med 2010. doi:10.1136/oem.2009.048215.

Cheng R, Traylor L. Non-regular maximum likelihood problems. J R Statist Soc B 1994;57:3-44.

Congdon P. Applied Bayesian modelling. West Sussex: Wiley; 2003.

De Roos AJ, Davis S, Colt JS, Blair A, Airola M, Severson RK, et al. Residential proximity to industrial facilities and risk of non-Hodgkin lymphoma. Environ Res 2010;110(1):70–8.

Diggle PJ, Elliott P. Disease risk near point sources: statistical issues for analyses using individual or spatially aggregated data. J Epidemiol Community Health 1995;49:20–7.

Diggle P, Rowlingson B. A conditional approach to point process modelling of elevated risk. J R Statist Soc A 1994;157(3):433–40.

Diggle P, Elliott P, Morris S, Shaddick G. Regression modelling of disease risk in relation to point sources. J R Statist Soc A 1997;160:491–505.

Draper G, Vincent T, Kroll ME, Swanson J. Childhood cancer in relation to distance from high voltage power lines in England and Wales: a case–control study. BMJ 2005;330:1290–4.

Dunn CE, Skingham S, Rowlingson B, Bhopal RS, Cockings S, Foy CJ, et al. Analysing public health data: a comparison of three methodological approaches. Health Place 2001;7:1-12.

Elliott P, Shaddick G, Kleinschmidt I, Jolley D, Walls P, Beresford J, et al. Cancer incidence near municipal solid waste incinerators in Great Britain. Br J Cancer 1996;73:702–10.

European Pollutant Emission Register (EPER). [http://eper.ec.europa.eu/eper/]. 2008.

Fanshawe TR, Diggle PJ, Rushton S, Sanderson R, Lurz PWW, Glinianaia SV, et al. Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. Environmetrics 2008;19:549–56.

Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol 2007;18(3):581–92.

Gottlieb MS, Carr JK. Case–control cancer mortality study and chlorination of drinking water in Louisiana. Environ Health Perspect 1982;46:169–77.

Grönberg H. Prostate cancer epidemiology. Lancet 2003;361:859–64.

Gulden J. Metal workers and repairmen at risk for prostate cancer: a review. Prostate 1997;30:107–16.

IARC (International Agency for Research on Cancer). Overall evaluations of carcinogenicity: an updating of IARC monographs volumes 1 to 42. Monographs on the Evaluation of Carcinogenic Risks to Humans Supplement 7. Lyon: IarcPress; 1987.

IARC (International Agency for Research on Cancer). IARC monographs on the evaluation of carcinogenic risk to humans. Occupational Exposures in Petroleum Refining; Crude Oil and Major Petroleum Fuels, Vol. 45. Lyon: IarcPress; 1989.

IARC (International Agency for Research on Cancer). Beryllium, cadmium, mercury, and exposures in the glass manufacturing industry. IARC monographs on the evaluation of carcinogenic risks to humans. Vol. 58. Lyon: IarcPress; 1993.

IARC (International Agency for Research on Cancer). IARC monographs on the evaluation of carcinogenic risk to humans. Wood dust and formaldehyde, vol. 62. Lyon: IarcPress; 1995.

Lawson A. Statistical methods in spatial epidemiology. West Sussex: Wiley; 2001.

López-Abente G, Aragonés N, Pollán M, Ruiz M, Gandarillas A. Leukemia, lymphomas, and myeloma mortality in the vicinity of nuclear power plants and nuclear fuel facilities in Spain. Cancer Epidemiol Biomark Prev 1999;8(10):925–34.

López-Abente G, Ramis R, Pollán M, Aragonés N, Pérez-Gómez B, Gómez-Barroso D, et al. Atlas municipal de mortalidad por cáncer en España, 1989–1998 (Atlas of municipal cancer mortality in Spain 1989–1998). Madrid: Instituto de Salud Carlos III; 2006a. http://www.isciii.es/htdocs/centros/epidemiologia/libros/Atlas_municipal.pdf.

López-Abente G, Aragonés N, Ramis R, Hernández-Barrera V, Pérez-Gómez B, Escolar-Pujolar A, et al. Possible role of mining and industry in the municipal distribution of bladder cancer mortality in Spain. BMC Public Health 2006b;27:6-17.

Martinez-Beneito M, Lopez-Quilez A, Amador A, Melchor I, Botella-Rocamora P, Abellan J, et al. Atlas de mortalidad de la Comunidad Valenciana 1991–2000. Valencia: Applied Spatial Statistics for Public Health Data; 2005. http://www.sp.san.gva.es/epidemiologia/atlas.

Maule M, Magnani C, Dalmasso P, Mirabelli D, Merletti F, Biggeri A. Modeling mesothelioma risk associated with environmental asbestos exposure. Environ Health Perspect 2007;115(7):1066–71.
Michelozzi P, Fusco D, Forastiere F, Ancona C, Dell'Orco V, Perucci CA. Small area study of mortality among people living near multiple sources of air pollution. Occup Environ Med 1998;55:611–5.
Miettinen OS. Standardization of risk ratios. Am J Epidemiol 1972;96:383–8.
Monge-Corella S, García-Pérez J, Aragonés N, Pollán M, Pérez-Gómez B, López-Abente G. Lung cancer mortality in towns near paper, pulp and board industries in Spain: a point source pollution study. BMC Public Health 2008;8 8: 288.
Nieuwenhuijsen M, Paustenbach D, Duarte-Davidson R. New developments in exposure assessment: the impact on the practice of health risk assessment and epidemiological studies. Environ Int 2006;32:996-1009.
Parent M, Siemiatycki J. Occupation and prostate cancer. Epidemiol Rev 2001;23:1.
Parodi S, Stagnaro E, Casella C, Puppo A, Daminelli E, Fontana V, et al. Lung cancer in an urban area in Northern Italy near a coke oven plant. Lung Cancer 2005;47: 155–64.
Pekkanen J, Pukkala E, Vahteristo M, Vartiainen T. Cancer incidence around an oil refinery as an example of a small area study based on map coordinates. Environ Res 1995;71:128–34.
Pérez-Gómez, B., Aragonés, N., Pollán, M., Suárez, B., Lope, V., Llácer, A., López-Abente, G., 2006. Accuracy of cancer death certificates in Spain: a summary of available information. Gac Sanit 2006 Dec; 20 Suppl 3, 42–51.
Post PN, Kil PJM, Crommelin MA, Schapers RFM, Coebergh JWW. Trends in incidence and mortality rates for prostate cancer before and after prostate-specific antigen introduction. A registry-based study in Southeastern Netherlands, 1971–1995. Eur J Cancer 1998;34(5):705–9.
Rodrigues A, Diggle P, Assuncao R. Semi-parametric approach to point source modelling in epidemiology and criminology. Appl Stat 2010;59:533–42.
Rothman K. Epidemiology. An introduction. New York: Oxford University Press; 2002.
Sans S, Elliott P, Kleinschmidt I, Shaddick G, Pattenden S, Walls P, et al. Cancer incidence and mortality near the Baglan Bay petrochemical works, South Wales. Occup Environ Med 1995;52:217–24.
Sharp, L., Black, R.J., Harkness, E.F., McKinney, P.A., 1996. Incidence of childhood leukaemia and non-Hodgkin's.
Sim MR, Del Monaco A, Hoving JL, Macfarlane E, McKenzie D, Benke G, et al. Mortality and cancer incidence in workers in two Australian prebake aluminium smelters. Occup Environ Med 2009;6:464–70.
Stone R. Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. Statistic Med 1998;7:649–60.
Viel JF, Arveux P, Baverel J, Cahn JY. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. Am J Epidemiol 2000;152(1):13–9.
Vinceti M, Venturelli M, Sighinolfi C, Trerotoli P, Bonvicini F, Ferrari A, et al. Case–control study of toenail cadmium and prostate cancer risk in Italy. Sci Total Environ 2007;373:77–81.