Environmental asbestos pollution and malignant mesothelioma: a spatial case-control study in the city of Bari (Italy). Methodology

Marina Musti[1], Alessio Pollice[2], Domenica Cavone[1], Massimo Bilancia[2]

[1] Department of Internal Medicine and Public Medicine, "Ramazzini" Section of Occupational Medicine,
University of Bari, National Registry of Mesothelioma, Regional Operative Centre (C.O.R.) Apulia, Italy
[2] "Carlo Cecchi" Department of Statistical Sciences, University of Bari, Bari, Italy

Corresponding Author:
Domenica Cavone
Università degli Studi di Bari, Dipartimento di Medicina Interna e Medicina Pubblica -
Sezione Medicina del Lavoro Ramazzini (Registro Nazionale dei casi di Mesotelioma –
Centro Operativo Regione Puglia) - Policlinico Piazza G. Cesare - 70124 Bari, Italy
E-mail: d.cavone@medlav.uniba.it

Summary
Objectives. To present the methodology of a spatial case-control study conducted to estimate the risk of malignant mesothelioma (MM) due to environmental asbestos exposure in people living near an asbestos cement factory in Bari, Italy.
Methods: The complete residential history of cases and controls was analyzed after the metrical coordinates of each address had been obtained and the distance from the putative source of risk computed. Point process-based methods were used to estimate the risk surface. Focused analyses were based on the Cuzick-Edward test and logistic regression considering disease occurrence as a function of raw classes of distance, controlling for sex and age. Explicit testing for the presence of a spatial trend in the risk was also carried out by means of a likelihood test for a model in which the spatial odds depend on distance, versus a null model that did not include the distance effect.
Results: An increase in the odds ratio of MM was found for people living close to the factory (within a radius of up to 500 metres).
Conclusions: The risk decreased with increasing distance from the factory. The test for the presence of a spatial trend was highly significant ($p = 3.2 \times 10^{-5}$).

KEY WORDS: *asbestos cement factory, environmental-neighbourhood exposure, mesothelioma registry; spatial case-control study, methodology, Italy.*

Introduction

In the 1990s, an increased mortality rate due to malignant mesothelioma (MM), related to the massive spread of asbestos use in the 1950s and '60s (1-2), was documented in many European countries, including Italy. The onset of mesothelioma has been related to inhalation of asbestos fibres and there is no evidence of a minimum exposure threshold below which there is no risk (3).

An excess risk has been observed in populations living near asbestos cement factories (AC factories): substantial evidence has been reported in a multicentric study conducted in six areas of Italy, Spain and Switzerland (4).

In Italy, a link has been demonstrated between asbestos production and an extension of the risk of tumours to the non-worker population living near production plants in many geographical areas: Casale Monferrato, Broni and Bari (5-7). The risk excess was found to be inversely correlated with the distance of subjects' homes from the factory.

This study analyzes the spatial distribution of deaths from MM due to environmental (neighbourhood) exposure in the city of Bari, where the source of risk was an AC factory active from 1934 to 1989 and then closed down; the surrounding area has still has not been properly decontaminated and reclaimed.

The aim of this study was to establish whether there is evidence of a greater concentration of MM cases around the AC factory in Bari, and to assess the influence, on mesothelioma onset, of the distance of subjects' homes from the factory.

Materials and methods

Study population

In this study, the residential distribution (distribution of the addresses) of MM patients who lived in Bari at diagnosis (between 1993 and 2003) and died before the start of the study is compared with the residential distribution of controls (a sample of decedents in Bari in the same period).

Cases

The records of all cases of MM registered at the Apulia regional operative centre of the Italian national mesothelioma registry between 1993 and 2003 were analyzed. The registry ensures complete, quality information on exposure thanks to the adoption of operating guidelines agreed at national level. All the cases in the registry had histologically confirmed mesothelioma. Exposure to asbestos (occupational, household, neighbourhood or environmental) was assessed and classified through face-to-face interviews with the subjects affected by mesothelioma (or their relatives); these interviews were conducted using a standardized questionnaire in accordance with ReNaM (National Mesothelioma Registry) guidelines [http://www.ispesl.it/ispesl/sitorenam/lineeguida.htm]. The questionnaire is designed to reconstruct, among other things, demographic characteristics, life habits and patients' lifetime occupational and residential histories. Information on patients' homes includes addresses and descriptions of their various habitations and the relative neighbourhoods.

Controls

The controls were automatically selected by date of death among the deaths from causes other than mesothelioma registered at the births and deaths registry of the city of Bari. Data on: sex, date and place of birth, date, place and cause of death and complete residential history were retrieved for each control from municipality registries. No data on occupational or household asbestos exposure (i.e. relatives' or cohabitants' exposure and domestic exposure) were available.

Matching of cases and controls for date of death, sex and date of birth was considered reliable only for the date of death (Table 1).

For the period 1993-2003, 48 cases whose deaths were attributable to environmental (neighbourhood) exposure (without occupational or household exposure) and 273 non-matched controls were available.

In view of the method of selection of the controls, sex and age were considered as confounding factors and expressed as covariates in the various parametric models used to produce risk estimates.

Residential history and geo-referencing

A complete residential history (addresses throughout the life of the individual, together with the precise periods of time they lived at each address) was available for each individual (cases and controls).

All the addresses were geo-referenced as geographical coordinates (in sexadecimal degrees), precise to the third decimal figure, using the Maporama tool, available on-line at www.maporama.com. The precision achieved by this mapping tool is comparable to that of a GPS system (i.e. about five metres within urban areas). WinDatum conversion software was used to transform the geographical coordinates into metric coordinates (UTM - WGS 84) for graphic representation and calculation of the distances.

In order to attribute each individual with a single res-

Table 1. Distribution of cases and controls and case-to-control ratios by year of death.

Year of death	Cases	Controls	Cases/ Controls
1992-1994	10	45	0.22
1995-1997	9	51	0.18
1998-2000	20	108	0.19
2001-2003	9	69	0.13
Total	48	273	0.18

idential address, two criteria were considered: first, the prevalent address (i.e. the one at which the individual lived for the longest time during her/his life) and second, the address nearest to the source of environmental risk (i.e. closest to the AC factory). Both references are useful because both the duration and the intensity of the exposure influence the probability of contracting MM.

In any case, the residence addresses for the fifteen years immediately prior to the date of diagnosis (cases) or prior to the date of death (controls) were excluded, in view of the minimal latency period of the disease. In this study, wishing to minimize number of false positives, we considered only the results relating to the prevalent addresses.

Statistical methods

The data in a spatial case-control study consist of two series of points corresponding to the spatial locations of cases of a given disease in a given geographical area over a given period of time, and of a sample of controls extracted from the population at risk.

The spatial distribution or pattern of the points is considered the result of a spatial stochastic process (8). The spatial variations in the process generating the data can first be studied by considering its first-order properties: the intensity function $\lambda(x)$ of the process corresponds to the mean number of events for an infinitesimal area dx, varying with the location x. In order to estimate the intensity functions for the two spatial processes corresponding to the cases and the controls, we used a Gaussian kernel density estimator which refers to a common smoothing parameter (the bandwidth) (9). For the critical issue of choosing this smoothing parameter, a standard approach based on the minimization of the mean squared estimation error was used (10). Kernel density estimations can be used to produce smooth maps of the spatial variation in disease risk. While estimation of the intensity of a single point pattern of cases may show areas of low and high risk, its utility is limited because it may, to a great extent, reflect the underlying pattern of the population at risk; the risk surface can thus be assessed by plotting a log of the estimated case-to-control intensity ratio. However, this approach was not suitable for our data given that the control estimated intensity was often zero in underpopulated areas of the city. It was therefore deemed preferable to consider the standardized difference between the two kernel estimates of the intensity functions, defined according to Han et al. (11)

$$\frac{\sqrt{\hat{\lambda}(x)_{cases}}-\sqrt{\hat{\lambda}(x)_{cantrols}}}{s.d.\left(\sqrt{\hat{\lambda}(x)_{cases}}-\sqrt{\hat{\lambda}(x)_{cantrols}}\right)} \qquad [1]$$

The above method makes it possible to produce a map of the relative risk surface and to identify "peaks" and "troughs" in the data. Raw methods for the detection of clustering in relation to a pre-specified source rely on semi-parametric tests which take into account the distance of cases and controls from the source of risk: the Cuzick and Edwards (CZ) test (12), for example, is a very simple procedure based on a count of the number of cases among the K cases and controls nearest to the source. Under the null hyphothesis of no clustering, the test statistic has an exact hypergeometrical distribution: the parameter K is commonly set equal to a fraction of the total sampling dimension (5% or 10%). Other procedures for case-control data do exist (13), although, like the CZ test, they have low statistical power.

Isotropic focused analyses (13) included consideration of four rings of increasing distance (500-1000, 1000-1500, 1500-2000 and >2000 metres) from a centre that coincided with the factory location. The risks (odds ratios) of each band were estimated using a logistic regression model in which the probability of disease occurrence is expressed as a function of the classes of distance (using distances greater than 2000 metres as the reference category), controlling for sex and age. The odds ratios and confidence intervals were obtained by exponential transformation of the corresponding parameter estimates in the logistic regression model.

We also estimated an extension of the logistic regression model, obtained without discretizing the distances from the putative source of risk. We defined the following additive log-linear models for disease odds at location x

$$odds(x;\theta)=\exp\left\{\theta_0+\sum_{j=1}^{p}\theta_j z_j(x)+h(x)\right\} \qquad [2]$$

where $z_j(x)$ represent known risk factors associated with individual spatial locations (in our study, sex and age class) and $h(x)$ is a non-linear function that describes the relation between the spatial odds and the distance from the putative risk source. The distance function adopted in this study had the following expression

$$h(x) = \log\left[1 + \alpha \exp\left(-\|x - x_0\|^2 / \beta\right)\right] \qquad [3]$$

Our spatial exploratory data analysis was based on the standardized difference between the kernel intensity estimations for the cases and controls. A preliminary analysis of the mean squared error of the kernel estimations allowed the optimal bandwidth for the estimations to be determined; to facilitate the comparison the same 360-metre bandwidth was considered for both the cases and the controls. The pattern of the resulting map (Fig. 2) is characterized by the presence of a peak near the factory, in other words, the addresses of the cases show a spatial cluster near the AC factory, a finding not paralleled in the spatial distribution of the controls, which represents overall mortality. However, the estimated risk surface also showed, on close observation, other secondary peaks corresponding to minor disease clusters: it is worth noticing that a cluster of six cases was found located to the east of the plant, close to the "Torre Quetta" urban beach, which, in the 1950s-1970s, was used as an unlicensed waste disposal site. Mortality rates exceeding background mortality levels (light grey areas) can be observed around the city centre, where the population density is highest.

Preliminary focused analysis based on the CZ test did not provide conclusive evidence: p-values for test statistics with K set equal to 5% or 10% of the total sampling dimension were respectively $p_{5\%} = 0.07$ and $p_{10\%} = 0.08$.

To analyze exposure effects as a function of the residential location, we classified subjects as residents in concentric bands increasingly distant from the AC factory. Table 3 shows the risk estimates in the concentric geographical bands, obtained using the log-linear model that controls for the confounding variables (sex and age class). The results revealed a spatial risk trend: risk decreasing with increasing distance from the production plant. A remarkable odds ratio (OR=5.29, 95% CI: 1.18-23.74) was found within the band including distances of up to 500 metres; the other odds ratios were above one, but they failed to reach statistical significance.

Explicit testing for the presence of a spatial trend in the risk of MM was assessed by fitting the model described in expressions [2] and [3], which takes into account the distance from the factory in the calculation of the risk of mesothelioma and adjusts for the effects of the confounding variables (14). The likelihood ratio test, which hypothesizes the absence of an

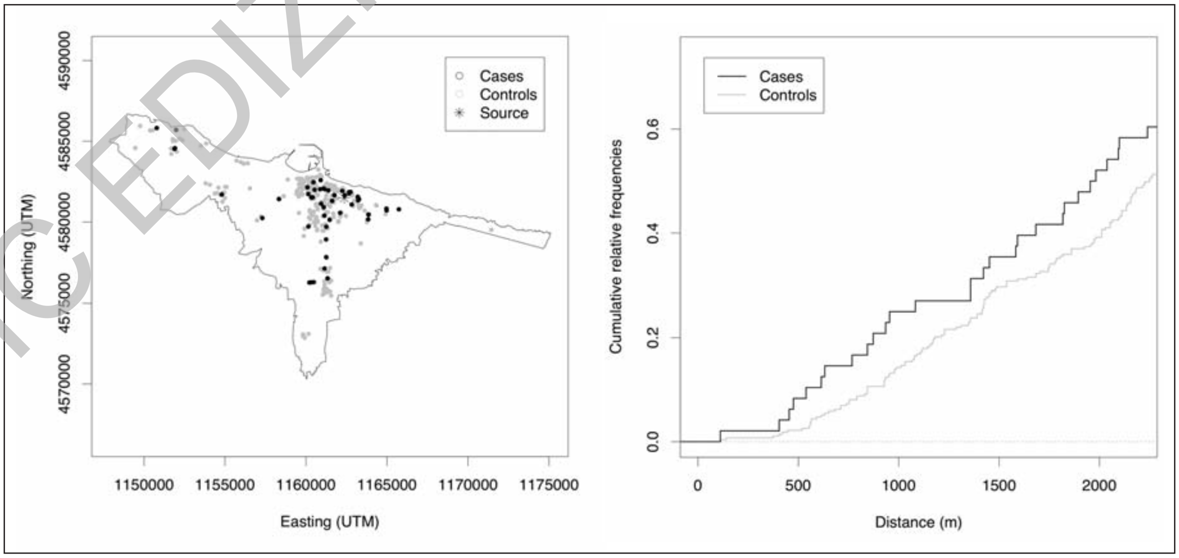

Figure 1a-1b. Map of residence addresses of cases and controls and distribution of cumulative frequencies of the distances from the asbestos cement factory.

Table 2. Distribution of the 48 cases and 273 controls by sex and age at death.

Age	Cases M	Cases F	Total cases	Controls M	Controls F	Total Controls
≤ 49	1 (3.7)	4 (19.0)	5 (10.4)	7 (3.4)	3 (4.5)	10 (3.7)
50-54	1 (3.7)	3 (14.3)	4 (8.3)	4 (1.9)	2 (3.0)	6 (2.2)
55-59	4 (14.8)	1 (4.8)	5 (10.4)	18 (8.7)	2 (3.0)	20 (7.3)
60-64	3 (11.1)	5 (23.8)	8 (16.7)	15 (7.3)	2 (3.0)	17 (6.2)
65-69	2 (7.4)	2 (9.5)	4 (8.3)	20 (9.7)	3 (4.5)	23 (8.4)
70-74	4 (14.8)	1 (4.8)	5 (10.4)	15 (7.3)	6 (8.9)	21 (7.7)
75-79	5 (18.5)	2 (9.5)	7 (14.6)	39 (18.9)	5 (7.4)	44 (16.1)
80-84	4 (14.8)	2 (9.5)	6 (12.5)	37 (18.0)	16 (23.9)	53 (19.4)
85-89	2 (7.4)	0 (0.0)	2 (4.2)	31 (15.1)	12 (17.9)	43 (15.8)
≥ 90	1 (3.7)	1 (4.8)	2 (4.2)	20 (9.7)	16 (23.9)	36 (13.2)
Total	27 (100.0)	21 (100.0)	48 (100.0)	206 (100.0)	67 (100.0)	273 (100.0)

Table 3. Estimated risk by distance from the AC factory, adjusted for sex and age.

Distance (in metres)	Cases		Controls			
	n	%	n	%	OR	95% CI
0-500	4	8.3	6	2.2	5.29	1.18-23.74
500-1000	8	16.7	33	12.1	1.49	0.58-3.82
1000-1500	5	10.4	42	15.4	1.27	0.49-3.71
1500-2000	8	16.7	26	9.5	2.31	0.88-6.06
> 2000[a]	23	47.9	166	60.8	1	

[a] Reference category

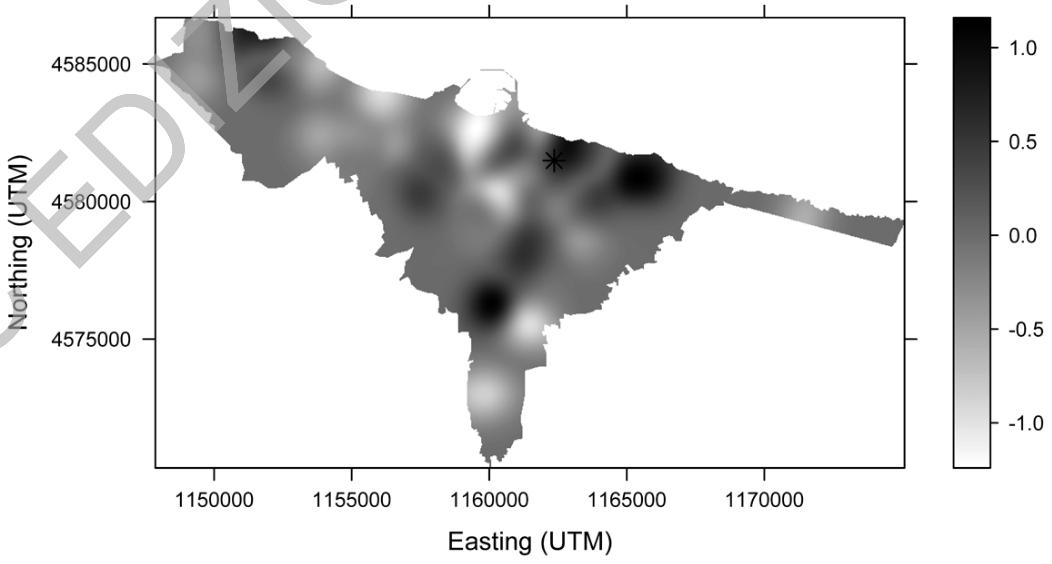

Figure 2. Standardized differences between the kernel estimations of the intensity function for cases and controls.

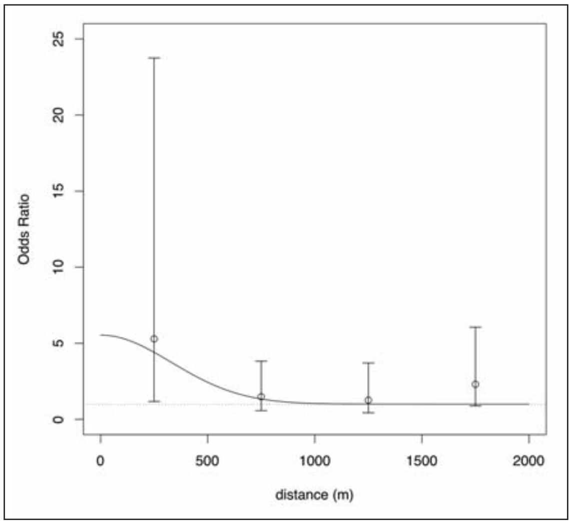

Figure 3. Sex- and age-adjusted mesothelioma risk estimates as a function of the distance from the AC factory, based on Diggle and Rowlingson's model (curve) and odds ratios (circles) with the logistic regression model (vertical lines indicate 95% confidence intervals).

effect of distance from the pollution source under the null hypothesis ($H_0 : h(x) = 0$), showed strong evidence in favour of the model in which the spatial odds depend on the distance ($p = 3.2 \times 10^{-5}$). The graph in figure 3 shows the function describing the behaviour of the risk of mesothelioma, according to the distance from the AC factory, estimated with the above model, together with the odds ratios estimated with the logistic model shown in Table 3.

Discussion

The role of geographical studies in environmental epidemiology and their importance in terms of public health has been firmly established in the literature: general purpose spatial statistical methods are available (16, 17), and these can easily be integrated with in-depth analytical epidemiological studies (18). Geographical information systems (19) as well as new methods for the analysis of disease risk as a function of the distance from a known source of risk (15,20) can help to provide evidence that environmental exposure to industrial sources increases cancer risk among local residents (21).

We used the available information on the residential histories of cases and controls in order to provide a detailed analysis of the spatial variation in MM risk from environmental asbestos exposure; our results were consistent with those of a similar study concerning another Italian area with a large AC factory (22).

Many aspects of our study design are worth discussing. We considered sex and age as confounding factors to be expressed as covariates in the various risk estimation models, in order to get round the limitation of the study due to our inability to match the cases and controls. Another limitation was the fact that the procedure used to select the controls did not enable us to verify lack of occupational exposure to asbestos, even though the controls did not show a tendency to cluster around likely points of asbestos exposure (factories, harbour, etc.). As these limitations can induce an underestimation of mesothelioma risk, our results must be regarded as conservative. Despite the small number of cases in the sample, the availability of lifetime residential history (historical certificates of residence) was a pre-requisite for inclusion, of both cases and controls, in the study (11, 23). Moreover, both cases and controls were included only if they had lived in Bari all their life. We examined various criteria for selecting the addresses to be included in the spatial analysis in order to control for the various types of bias that could be introduced. In order minimize the number of false positives, it was decided to take into consideration each subject's prevalent address (corresponding to the longest period of residence).

Other important issues concern the statistical methods used in the study. The use of a Gaussian kernel to estimate the risk surface was not crucial to our results; it is recognized in the literature on point processes that non-parametric estimates of the intensity function are substantially independent of the choice of kernel function, while the use of other kernel functions yielded results entirely comparable to those presented in this paper (24).

Another choice that may appear arbitrary was the shape of the function [3] used to model the risk on the basis of the distance from the putative source of pollution; however, this shape shows an exponential reduction of the risk with increasing distance, reflecting a flexibility sufficient to fit any real situation. (25).

Moreover, in the model the risk shows the desirable property of tending towards 1 when the distance from the source of pollution tends towards infinity: this baseline value allows immediate identification of any excess or deficiency in the disease odds, within a finite distance from the putative risk source.

Focused analyses were conducted, in accordance with Maule et al. (22), using the distance from the factory as a proxy of the environmental exposure, in order to confirm the hypothesis of an increase of MM incidence due to proximity to the risk source (AC factory). It is worth noticing that the odds ratio for the cases and controls falling within the 0-500 metre radius (OR=5.29, 95% CI= 1.18-23.74) indicated a significant increase in MM risk: this result was strengthened by the highly significant p-value ($p = 3.2 \times 10^{-5}$) obtained for the model [2], wherein the spatial odds depended on the distance from the source, versus the null model which did not include the distance effect. On the whole, the results of our study strengthen the evidence of an increased disease incidence around the AC factory in Bari.

Acknowledgments

We thank Professor Annibale Biggeri for his comments on the manuscript.

Conflict of interest statement: None declared by any author.

References

1. La Vecchia C, Decarli A, Peto J, Levi F, Tomei F, Negri E. An age, period and cohort analysis of pleural cancer mortality in Europe. Eur J Cancer Prev. 2000; 9: 179-184.
2. Marinaccio A, Montanaro F, Mastrantonio M, Uccelli R, Altavista P, Nesti M, Seniori Costantini A, Gorini G. Predictions of mortality from pleural mesothelioma in Italy: a model based on asbestos consumption figures supports results from age-period-cohort models. Int J Cancer 2005; 115: 142-147.
3. Hillerdal G. Mesothelioma cases associated with non-occupational and low dose exposures. Occup Environ Med 1999; 56: 505-513.
4. Magnani C, Agudo A, Gonzalez CA, Andrion A, Calleja A, Chellini E, Dalmasso P, Escolar A, Hernandez S, Ivaldi C, Mirabelli D, Ramirez J, Turuguet D, Usel M, Terracini B. Multicentric study on malignant mesothelioma and non-occupational exposure to asbestos. Br J Cancer 2000; 83: 104-111.
5. Magnani C, Dalmasso P, Biggeri A, Ivaldi C, Mirabelli D, Terracini B. Increased risk of malignant mesothelioma of the pleura after residential or domestic exposure to asbestos: a case-control study in Casale Monferrato, Italy. Environ Health Perspect 2001; 109: 915-919.
6. Amendola P, Belli S, Binazzi A, Cavalleri A, Comba P, Mastrantonio M, Trinca S. Mortality from malignant pleural neoplasms in Broni (Pavia), 1980-1997. Epidemiol Prev 2003; 27: 86-90.
7. Bilancia M, Cavone D, Pollice A, Musti M. Valutazione del rischio di mesotelioma: il caso di una fabbrica per la produzione di cemento-amianto nella città di Bari. Epidemiol Prev 2003; 27: 277-284.
8. Gatrell AC, Bailey TC, Diggle, PJ, Rowlingson BS. Spatial point pattern analysis and its application in geographical epidemiology. Trans Inst Engl Geogr 1996; 21: 256-274.
9. Rogerson PA. Statistical methods for the detection of spatial clustering in case-control data. Stat Med. 2006; 25: 811-823.
10. Berman M, Diggle PJ. Estimated weighted integrals of the second-order intensity of a spatial point process. JRSS B 1989; 51: 81-92.
11. Han D, Rogerson PA, Bonner MR, Nie J, Vena JE, Muti P, Trevisan M, Freudenheim JL. Assessing spatio-temporal variability of risk surfaces using residential history data in a case control study of breast cancer. Int J Health Geogr 2005; 4: 9.
12. Cuzick J, Edwards R. Spatial clustering for inhmogeneous population (with discussion), JRSS B 1990, 52: 73: 104.
13. Morris SE, Wakefield JC. Assessment of disease risk in relation to a pre-specified source. In: Elliot P, Wakefield JC, Best NG, Briggs DJ eds Spatial Epidemiology – Methods and Applications. 2000, Oxford, Oxford University Press: 153-184.
14. Diggle PJ, Rowlingson BS. A conditional approach to point process modelling of elevated risk. JRSS A 1994; 157: 433-440.
15. 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: 750-754.
16. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect 2004; 112: 998-1006.
17. Mather FJ, White LE, Langlois EC, Shorter CF, Swalm CM, Shaffer JG, Hartley WR. Statistical methods for linking health, exposure, and hazards. Environ Health Perspect 2004; 112: 1440-1445.
18. Bianchi F, Biggeri A, Cadum E, Comba P, Forestiere F, Martuzzi M, Terracini B. Environmental epidemiology and polluted areas in Italy. Epidemiol Prev. 2006; 30: 146-152.
19. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. Environ Health Perspect 2004; 112: 1007-1015.
20. Vieira V, Webster T, Aschengrau A, Ozonoff D. A method for spatial analysis of risk in a population-

based case-control study. Int J Hyg Environ Health 2002; 205: 115-120.

21. Comba P, Ascoli V, Belli S, Benedetti M, Gatti L, Ricci P, Tieghi A. Risk of soft tissue sarcomas and residence in the neighbourhood of an incinerator of industrial wastes. Occup Environ Med 2003; 60: 680-683.
22. Maule MM, Magnani C, Dalmasso P, Mirabelli D, Merletti F, Buggeri A. Modeling mesothelioma risk associated with environmental asbestos exposure. Environ Health Perspect 2007,115:1066-1071.
23. Jacquez GM, Meliker JR, Avruskin GA, Goovaerts P, Kaufmann A, Wilson ML, Nriagu J. Case-control geographic clustering for residential histories accounting for risk factors and covariates. Int J Health Geogr 2006;5:32.
24. Diggle P. Statistical Analysis of Spatial Point Patterns, 2nd ed. 2003, London, Hodder Arnold.
25. Dreassi E, Lagazio C, Maule MM, Magnani C, Biggeri A. Sensitivity analysis of the relationship between disease occurrence and distance from a putative source of pollution. Geospat Health. 2008, 2:263-271.