Air Pollution and Lung Cancer in Trieste, Italy: Spatial Analysis of Risk as a Function of Distance from Sources

Annabile Biggeri,1 Fabio Barbone,2,3 Corrado Lagazio,1 Massimo Bovenzi,4 and Giorgio Stanta5

1Department of Statistics “G. Parenti,” University of Florence, Florence, Italy; 2Unit of Hygiene and Epidemiology, DPMSC, University of Udine, Udine, Italy; 3Epidemiology Unit, Aviano Cancer Center, Aviano, Italy; 4Institute of Occupational Medicine, University of Trieste, Trieste, Italy; 5Institute of Pathology, Cancer Registry of the Province of Trieste, Trieste, Italy

To investigate the relationship between four sources of environmental pollution (shipyard, iron foundry, incinerator, and city center) and lung cancer risk, we conducted a case–control study of deceased men in Trieste, Italy. We identified 755 cases of lung cancer and 755 controls through the local autopsy registry. Information on smoking habits, occupational history, and place of residence were obtained from the subject’s next of kin. The case–control design was used to properly account for subject-specific confounders, which represent a major problem in geographical analysis. Spatial models were used to evaluate the effect of sources of pollution on lung cancer after adjustment for age, smoking habits, likelihood of exposure to occupational carcinogens, and levels of air particulate. The models are based on distance from the sources and enable estimation of the risk gradient and directional effects separately for each source. The risk of lung cancer was highly related to the city center (p = 0.0243), with an excess relative risk at zero distance of 2.2 and a smooth decrease moving away from the source (-0.015), and related to the incinerator (p = 0.0098), with an excess relative risk of 6.7 in the source and a very steep decrease (-0.176). These results are consistent with findings of previous analyses and provide further evidence that air pollution is a moderate risk factor of lung cancer. Key words: air pollution, epidemiology, geographical analysis, lung cancer. Environ Health Perspect 104:750–754 (1996)

Results of a case–control study on air pollution and lung cancer in Trieste, Italy, were reported by Barbone et al. (1). That study confirmed a moderate elevation in risk of lung cancer in polluted areas and showed a variation by histologic type and category of air pollution. Trieste, which had approximately 250,000 inhabitants in the mid-1980s, is a border city located in the northeast of Italy and is characterized by a major port and a high concentration of industries. Air pollution has been monitored since the early 1970s. Higher total particulate deposition levels (i.e., >0.3 g/m²/day) were documented in the center of the city and in the industrial area in the 1970s. Currently, higher levels of carbon monoxide (monthly average 3.6 mg/m³) and nitrogen oxides (218 μg/m³) are found in the center of the city, and higher levels of ozone (32–39 μg/m³) and sulfur dioxide (50–59 μg/m³) are present near an incinerator and an iron foundry. The presence of suspended asbestos fibers was documented near a shipyard. Here we present analyses of the spatial pattern of risk of lung cancer with regard to four sources, shipyard, iron foundry, incinerator, and the city center, while adjusting for known risk factors.

Geographical investigations are hampered by the difficulties in properly accounting for confounders (2). However, methods based on the case–control design have been proposed in the statistical literature that allow the collection of data at individual level, avoiding the ecologic bias (3). The merit of the analysis presented here is in relaxing the a priori categorization of the subject residence in given areas and in using the distance from a source as a proxy for exposure. Second, the method we used allows for directional effects and estimates the risk gradient in order to properly describe the specific pattern of risk for each source.

Materials and Methods

The Cancer Registry and the Department of Pathology of the Province of Trieste identify 99% of cancer cases and conduct autopsies on approximately 73% of all the deaths of the region. From these institutions, 938 histologically confirmed cases of lung cancer were identified among males resident in the province of Trieste, who died from 1979 to 1981 or from 1985 to 1986. The two enrollment periods were chosen to cover an extended time span at a reasonable cost. The study had been originally designed to investigate environmental and occupational risk factors for lung cancer. This, together with statistical power considerations, was the reason we restricted the study to male cases only. We excluded 182 cases because we failed to trace the next of kin and 1 case because his residence was outside the Province of Trieste.

For each case, one male control resident in the Province of Trieste, who died within the same 6-month period, at the same age (± 2 years), was randomly selected from the same archive at the Department of Pathology. The causes of death of the controls were not chronic lung diseases or cancer of the upper aerodigestive tract, urinary tract, pancreas, liver, or gastrointestinal system. The sampling probabilities for the control series are usually varied according to the proportion of cases by some relevant variable such as age or sex (4,5). The baseline spatial intensity would be therefore distorted, compared to a random sample of death controls. Use of death controls instead of living ones is widely discussed in the epidemiological literature (6). Our choice is justified by minimizing selection biases with special reference to residential history.

The present study was based on 755 case–control pairs, determined by age. Each subject’s next of kin was interviewed within 1–3 years of the subject’s death by means of a structured questionnaire to obtain information on demographic characteristics, smoking habits, occupational history, and last place of residence. Likelihood of exposure to occupational carcinogens was obtained from expert evaluation based on the type of job and also for people working in the iron foundry, shipyard, and incinerator. This summary variable was chosen to increase statistical power, since to include several variables for each job would have led to sparse data and results would have been affected by excess random variation.

Length of residence was not individually assessed; we only assessed if any subject moved from his place of residence in the last 10 years. A detailed description of data collection procedures and exposure coding has been published elsewhere (1).

Geographical. The boundaries of the Province of Trieste were coded using the geographical coordinates (Mercator projection) as provided by the Italian Army Geographical Institute (Florence, Italy; map

Address correspondence to C. Lagazio, Department of Statistics "G. Parenti," University of Florence, Viale Morgagni 59, 50134 Firenze, Italy.

This work was partially supported by the Consiglio Nazionale delle Ricerche (CNR), grant no. 8400604.44 and by the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MURST), 60% grant. We thank the referees for suggestions that led to substantial improvements in the content and organization of the paper.

Received 13 November 1995; accepted 27 March 1996.
produced

For the analysis, we calculated the distance and the angle from each subject location
to each pollution source (north orientation). Maps with point locations were
produced using ARC/INFO 6.1 (7); contour plots of relative risk gradient were con-
structed using Gauss 2.2 (8).

Point-source analysis. The present analysis focuses on the spatial intensity \( \lambda(x) \)
i.e., the frequency of events by unit area at location \( x \). This is the spatial counterpart of
the usual concept of rate, having substituted
unit-time
unittime
average

\[ \lambda(x) = \lambda_p(x)p(x-x_0; \theta) \]

where \( \lambda_p(x) \) indicates the population intensity at the location \( x \) and \( p(x-x_0; \theta) \)
is the risk as a function of the distance \( x-x_0 \) from the location of the source \( x_0 \), model-
ed by the parameters \( \theta \).

The case-control design is used to bypass the task of obtaining valid estimates of the population density at each location
\( x \). The spatial intensity for the control
series (i.e., non-cases) is:

\[ \lambda_{CS}(x) = \lambda_p(x)[1 - p(x-x_0; \theta)] \]

and for the case series:

\[ \lambda_{CS}(x) = \lambda_p(x) \frac{p(x-x_0; \theta)}{1 - p(x-x_0; \theta)} \]

where \( k \) and \( c \) are constants determined by study design (sampling fraction and
and case-control ratio, respectively). The spatial intensity of disease is therefore a function of the
odds of disease (the odds being the probability of being ill over the probability of not being ill). To overcome the difficulty in estimating \( \lambda_{CS}(x) \), Diggle and Rowlingson (3)
proposed conditioning the analysis on the observed case and control locations [further
details are in Lagazio (9)]. We define a logistic regression model in which the odds of
disease is:

\[ \text{odds}[p(x-x_0; \theta)] = w[1 + f(x-x_0; \theta)] \]

assuming an additive scale for the relative risk [where \( w \) is a proportionality factor and \( f() \)
is a function to be defined later]. This is plausible because, with a suitable
choice of \( f() \), the risk is unchanged at infinite
distance from the source. In the case of multiple sources the model becomes:

\[ \text{odds}[p(x-x_0; \theta)] = w[1 + \sum_i f(x-x_{0i}; \theta)] \]

and individual risk factors can be modeled in the following way:

\[ \text{odds}[p(x-x_0; z; \theta, \gamma)] = w \Pi_i \exp[\gamma_i f(x-x_{0i}; \theta)] + \sum_i \gamma_i \]

where \( \gamma_i \) denotes the \( i \)th source and \( \gamma_i \) is the
log odds ratio for the \( i \)th risk factor, \( z_i \). The adjusted excess risk gradient for each source
is modeled as follows:

\[ f(x-x_0; \theta) = \alpha \exp[\beta_0 + \beta_1 \sin(\theta) + \beta_2 \cos(\theta)] \]

where \( \beta_0 \) is the parameter \( \alpha \) models the excess
relative risk at the source location, \( d \) is the (in meters) from the \( i \)th source, and
the parameter \( \beta_1 \) models the exponential
decrease of the excess relative risk for longer distances. To allow for directional
effects, we define the following model for a given source:

\[ f(x-x_0; \theta) = \alpha \exp[\beta_0 + \beta_1 \sin(\theta) + \beta_2 \cos(\theta)] \]

where \( d \) is the distance and \( \theta \) is the angle between the case or control location and the
source location. This is of particular importance when considering a situation like that
in Trieste, where the city is located between the coast (southwest) and hills (northeast).
Although Trieste is famous for a strong northeast to southwest wind (bora) the
moderate winds from the sea toward the hills are more relevant for the spread of air pollution.

The model-based spatial analysis was conducted to allow for the contribution of rele-
vant risk factors. These terms were considered in the multiplicative scale in the model:
age, smoking habit (nonsmoker, 1–19, 20–39, and \( >40 \) cigarettes/day),
and exposure to occupational carcinogens (none, possible, likely). Moreover, we included the
levels of air particulates as defined in a previous
paper (1) (tetrates of distribution, 1972–1977: \( <0.175; 0.175-0.298; \geq0.298 \)
g/m\(^3\)/day). Each subject was assigned the
average value measured by the nearest among
the 28 stations that covered the city.

In the appendix, we report point esti-
mates and likelihood ratio tests for the sig-
nificance of the spatial terms in the model.
The likelihood surface for those parameter
estimates has an odd shape, and therefore
their relative standard errors are poorly esti-
mated. In this situation it is preferable to
rely on likelihood ratios (10). These models
are known as mixed additive-multiplicative
models for excess relative risks and can be fitted using Epicure software (11).

Crude analysis. To describe the observed
pattern of relative risk within the study area,
we estimated the spatial intensity, \( \lambda(x) \)
on-
parametrically, following the suggestions of
Bithell (12) and Lawson and Williams (13).
The spatial intensities for the case and control
series are estimated separately as follows:

\[ \hat{\lambda}(x) = \sum_{i=1}^{n} \hat{h}_i^{-2} \frac{(x-x_i)^2}{h_i^2} \]

where the kernel function, \( G(\cdot) \), has the
Epanechnikov functional form (14). The terms
\( h_i \) are smoothing parameters that allow
for local variation of the degree of smoothing. They are obtained as
\( h_i = \eta h_\text{ave} \)
where \( \eta h_\text{ave} \) is fixed in advance (500 m for
our application), and \( \eta h_\text{ave} \) is a previous estimate obtained using the simple nearest-neighbor
technique (14).

The ratio of the kernel estimates for
cases and non-cases is the odds of being a
case, given the observed sample (this quantity
differs from the odds of being ill
because it also depends on the case-control ratio). To obtain easily interpretable con-
tour plots, we back-transformed it to prob-
bility; i.e.,

\[ \hat{p}(x) = \hat{g}(x)/(1 + \hat{g}(x)) \]

where \( \hat{g}(x) \) represents the odds of being a
case. Because in our study the case-control

Table 1. Relative risks for lung cancer in Trieste: smoking habits, occupational exposures, and levels of air particulates

| Variable          | Cases | Controls | Odds ratio* | 95% CL   |
|-------------------|-------|----------|-------------|---------|
| Smoking           |       |          |             |         |
| (cigarettes/day)  | 0     | 22       | 199         | 1.0     |
| 1–19              | 225   | 272      | 6.7         | 4.2-11  |
| 20–39             | 302   | 198      | 12.8        | 7.9-21  |
| \( >40 \)         | 206   | 86       | 21.3        | 15-36   |
| Occupational exposure to carcinogens |       |          |             |         |
| No                | 255   | 351      | 1.0         |         |
| Possible          | 282   | 279      | 1.4         | 1.1-1.9 |
| Probable          | 218   | 125      | 2.5         | 1.8-3.4 |
| Air particulates  | (g/m\(^3\)/day) |         |             |         |
| \( <0.175 \)     | 188   | 219      | 1.0         |         |
| \( 0.175-0.298 \) | 256   | 274      | 1.1         | 0.8-1.5 |
| \( >0.298 \)     | 311   | 262      | 1.4         | 1.1-1.8 |

*Adjusted for smoking, likelihood of occupational exposure to carcinogens, and air pollution.
Articles

Figure 1. Locations of cases.

Figure 2. Locations of controls.

to the source location. The incinerator was highly significant ($p = 0.0098$), with an excess relative risk of 6.7 and a very rapid decay moving away from the source. No other sources reached statistical significance when city center and incinerator had been included in the model.

Finally, we investigated if there were directional effects with regard to the effect of the incinerator. The appendix shows the results of fitting that model. Although not statistically significant, the point estimates for the directional effects suggested a wind effect from southwest to northeast.

Incidentally, we note the estimates for the levels of particulate: the odds ratios were 1.1 (95% CL, 0.8–1.5) for the second tertile and 1.4 (1.1–1.8) for the highest tertile. When we took into account the distance from the city center and the incinerator, the effect of particulate vanished: second tertile, OR = 1.2 (0.9–1.4); highest tertile, OR = 1.0 (0.7–1.4).

Discussion

The present analysis supports and validates the geographical areas defined in a previous study (1). Indeed, the use of the distance between residential location and sources of pollution as a continuous variable provided a more sensitive approach to spatial modeling of risk than the classification of the residences into four areas on the basis of their proximity to each source. Furthermore, the evidence of higher risk in the neighborhood of the incinerator has been confirmed. The excess relative risk estimated at the city center and at the location of the incinerator appears to be consistent as well as the shallow and steep descent, respectively.

The model adopted is simple, allowing an exponential decrease by distance from the source. Although several alternatives could be specified (15), we chose the model described here because it could be extended to include more than one source. The peculiar spatial location of the four sources complicate the analysis. The sources appear to be highly correlated, and the geography of the city is heavily affected by its proximity to the coast.

For these reasons we adopted a forward strategy to select the best-fitting model. The final model contains terms for spatial effects of the city center and of the incinerators. This could be due to the indistinguishable effects of the shipyard, the city center, and, to a lesser degree, the iron foundry, which lie on the same line along a north–south direction. The incinerator effects retained statistical significance even when adjusting for individual risk factors and spatial effects of the city center.

The previous analysis based on histological subtypes of lung cancer showed higher relative risks for small cell and large cell carcinoma among residents close to the city center, whereas the relative risk for squamous cell carcinoma and adenocarcinoma was elevated among those residents who lived close to the incinerator (1). The presence of a linear trend by level of particulate deposition was significant for small and large cell cancers. In the present study, for all lung cancers there was a significant increase in risk for those resi-
dent in areas in the highest tertile of particulate (>0.296 g/m²/day, OR = 1.4; 95% CI, 1.1–1.8). This effect appeared to be fully explained once distance from city center and incinerator had been included in the model.

Appendix

Excess risk of lung cancer as a function of distance from city center, shipyard, foundry, and incinerator considered separately

Null model: odds[p(x;γ)] = wΠj exp(αjγj)

Includes terms for age, smoking habits, occupational exposure and levels of air particulate.

Model 1: odds[p(x - xj,0;θ,γ)] = wΠj exp(αjγj)[1 + αexβexxj]

αex = risk excess in the source (city center) = 2.209
βex = risk decay moving away from city center = -0.0151
Likelihood ratio statistic model 1 vs. null model = 7.435, df = 2
p = 0.0243

Model 2: odds[p(x - xj,0;θ,γ)] = wΠj exp(αjγj)[1 + αshβshxj]

αsh = risk excess in the source (shipyard) = 2.033
βsh = risk decay moving away from the shipyard = -0.01922
Likelihood ratio statistic model 2 vs. null model = 7.868, df = 2
p = 0.0196

Model 3: odds[p(x - xj,0;θ,γ)] = wΠj exp(αjγj)[1 + αshβshxj]

αsh = risk excess in the source (foundry) = 1.702
βsh = risk decay moving away from the shipyard = -0.01692
Likelihood ratio statistic model 3 vs. null model = 5.273, df = 2
p = 0.0166

Model 4: odds[p(x - xj,0;θ,γ)] = wΠj exp(αjγj)[1 + αshβshxj]

αsh = risk excess in the source (incinerator) = 1.484
βsh = risk decay moving away from the incinerator = -0.01505
Likelihood ratio statistic model 4 vs. null model = 4.736, df = 2
p = 0.0937

Excess risk of lung cancer as a function of distance from city center and from either the shipyard, foundry, or incinerator

Null model: odds[p(x - xj,0;θ,γ)] = wΠj exp(αjγj)[1 + αexβexxj]

Includes terms for age, smoking habits, occupational exposure and levels of air particulate and excess risk as function of distance from the city center.

Model 1: odds[p(x - xj,0;θ,γ)] = wΠj exp(αjγj)[1 + αexβexxj] + αshβshxj

αex = risk excess in the source (city center) = 0.9091
βex = risk decay moving away from city center = -0.01855
αsh = risk excess in the source (shipyard) = 1.242

This study was mainly a geographical investigation with characterization of environmental exposure by adjustment for total particulate deposition and residence location. Although the spatial pattern of the risk was adjusted for relevant confounders, residual confounding due to other unmeasured exposure cannot be excluded. Background radiation should not be a problem in this area because it is known that radiation follows a gradient, with a minimum at the city center and a maximum in the rural area at the boundary of the province. A selection bias due to the chosen frame of cases and controls cannot be excluded in principle; however, it should be noted that the subject list is derived from the Cancer Registry, which guarantees the coverage of the resident population and provides high-quality data, including 73% of all deaths autopsied. It was impossible to obtain a complete residential history for each subject enrolled. Therefore, misclassification bias due to change in residence cannot be excluded (we note that eventually this error would push the risk estimates toward the null value; nondifferential misclassification or selective migration of cases, e.g., of terminally ill people, outside the risk areas). The results shown here are coherent with the hypothesis of an independent effect of residing close to the incinerator and the city center. Further investigations should be undertaken to characterize the types and levels of pollutants from the incinerator and the center of the city.
\[ \beta_{sk} = \text{risk decay moving away from the shipyard} = -0.02208 \]

Likelihood ratio statistic model 1 vs. null model = 1.089, \( df = 2 \)

\[ p = 0.5803 \]

Model 2: \[ \text{odds}[p(x, x; \gamma)] = w \Pi \exp(\gamma_j \gamma_i) [1 + \alpha_\gamma \exp(\beta_{s,k}) + \alpha_\gamma \exp(\beta_{s,k})] \]

\[ \alpha_\gamma = \text{risk excess in the source (city center)} = 1.857 \]
\[ \beta_{s,k} = \text{risk decay moving away from city center} = -0.02439 \]
\[ \alpha_\gamma = \text{risk excess in the source (iron foundry)} = 5.858 \]
\[ \beta_{s,k} = \text{risk decay moving away from the iron foundry} = -0.1615 \]

Likelihood ratio statistic model 2 vs. null model = 4.889, \( df = 2 \)

\[ p = 0.0868 \]

Model 3: \[ \text{odds}[p(x, x; \gamma)] = w \Pi \exp(\gamma_j \gamma_i) [1 + \alpha_\gamma \exp(\beta_{s,k}) + \alpha_\gamma \exp(\beta_{s,k})] \]

\[ \alpha_\gamma = \text{risk excess in the source (city center)} = 1.959 \]
\[ \beta_{s,k} = \text{risk decay moving away from city center} = -0.03523 \]
\[ \alpha_\gamma = \text{risk excess in the source (incinerator)} = 6.740 \]
\[ \beta_{s,k} = \text{risk decay moving away from the incinerator} = -0.1762 \]

Likelihood ratio statistic model 3 vs. null model = 9.241, \( df = 2 \)

\[ p = 0.0098 \]

**References**

1. Barbone F, Bovenzi M, Cavallieri F, Stanta G. Air pollution and lung cancer in Trieste, Italy. Am J Epidemiol 141:1161–1169 (1994).
2. Brenner H, Savitz DA, Jockel KH, Greenland S. Effect of nondifferential exposure misclassification in ecologic studies. Am J Epidemiol 135:85–95 (1992).
3. Diggle PJ, Rowlingson BS. A conditional approach to point process modelling of elevated risk. J R Stat Soc A 157:433–440 (1994).
4. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. J R Stat Soc B 52:73–104 (1990).
5. Biggeri A, Marchi M. Case–control designs for the detection of space clusters of diseases. Environmetrics 6:385–393 (1995).
6. Lasky T, Stolley PD. Selection of cases and controls. Epidemiol Rev 16:6–17 (1994).
7. ESRI. ARC/Info Users manual, rev. 6.1.1. Redlands, WA: Environmental Systems Research Institute, 1993.
8. Aprect Systems. Gauss. System and graphics manual, version 2.2. Kent WA: Aprect Systems, 1991.
9. Lagazio C. Case–control studies for the analysis of association between risk of disease and putative sources of pollution [in Italian] (PhD thesis). Florence, Italy: University of Florence, 1994.
10. Moolgavkar SH, Venzon DJ. General relative risk regression models for epidemiologic studies. Am J Epidemiol 126:949–961 (1987).
11. Preston DL, Lubin JH, Pierce DA, McConney ME. Epicure. Seattle, WA: Hirosoft International, 1993.
12. Bihell JF. Application of density estimation to geographical epidemiology. Stat Med 9:691–701 (1990).
13. Lawson AB, Williams FLR. Application of extraction mapping in environmental epidemiology. Stat Med 12:1249–1258 (1993).
14. Silverman BW. Density estimation. London: Chapman & Hall, 1986.
15. Lawson AB. On the analysis of mortality events associated with a prespecified fixed point. J R Stat Soc A 156:363–377 (1993).