

Assessment of disease risk in the presence of a putative environmental pollution source: a case study

Un caso di studio sulla valutazione del rischio di malattia in presenza di una presunta sorgente di inquinamento ambientale

Alessio Pollice

Dipartimento di Scienze Statistiche
Università di Bari
apollice@dss.uniba.it

Massimo Bilancia¹

Dipartimento di Scienze Statistiche
Università di Bari
mabil@dss.uniba.it

Riassunto: In questo lavoro viene proposto un modello bayesiano gerarchico per analizzare la distribuzione spaziale di 64 casi di mesotelioma maligno occorsi nel territorio urbano della città di Bari attorno ad una presunta sorgente di rischio .

Keywords: Focused disease clustering, Bayesian hierarchical modelling, Markov Chain Monte Carlo.

1. Introduction

Malignant mesothelioma is a rare and eventually fatal neoplastic disease appearing in all serous membranes (but mostly in the pleura) and associated with the professional and environmental exposure to asbestos (McDonald, 1985). A dataset of 64 residents in the city Bari diagnosed as having malignant mesothelioma between 1980 and 2001 is analyzed.

The spatial distribution of the places of residence mapped in Fig.1 shows some evidence for spatial clustering around an asbestos-cement production plant that ceased a few years ago. Motivated by the (sometimes irrational) media and public interest, an accurate estimation of the disease risk around this putative source of hazardous waste is questioned.

The areal unit chosen for the analysis is the 1991 census tract (see Fig.3); one of the statistical issues coming with this data-set is the extreme prevalence of sampling zeros due to the rareness of the disease (1266 tracts, about 95.2% out of the total). Moreover 99 out of 1430 tracts in the original partitioning of the city of Bari had no resident population in the 1991 census and were excluded from the data base. Even if such areas cannot contribute to the likelihood, it is actually important to provide an estimate of the risk for them, being asbestos an ubiquitous pollutant.

2. Modelling disease risk

Difficulties discussed in the previous section can be dealt with by means of a Bayesian hierarchical mixed Poisson regression model. Let y_i , $i = 1, \dots, 1329$ denote the counts of deaths in the tracts (two tracts, the one where the plant is located and that gathering homeless people, were not considered); as is usual with rare diseases it is assumed that

$$[y_i|\phi_i] \stackrel{\text{ind.}}{\sim}^i \text{Poisson}[\exp(\phi_i) E_i], \quad i = 1, \dots, 1329 \quad (1)$$

¹The authors wish to thank Marina Musti (RENAM-COR Puglia - Dipartimento di Medicina Interna e Medicina Pubblica - Sez. di Medicina del Lavoro, Università di Bari) for kindly supplying the data, and Domenica Cavone (from the same institution) for her helpful assistance.

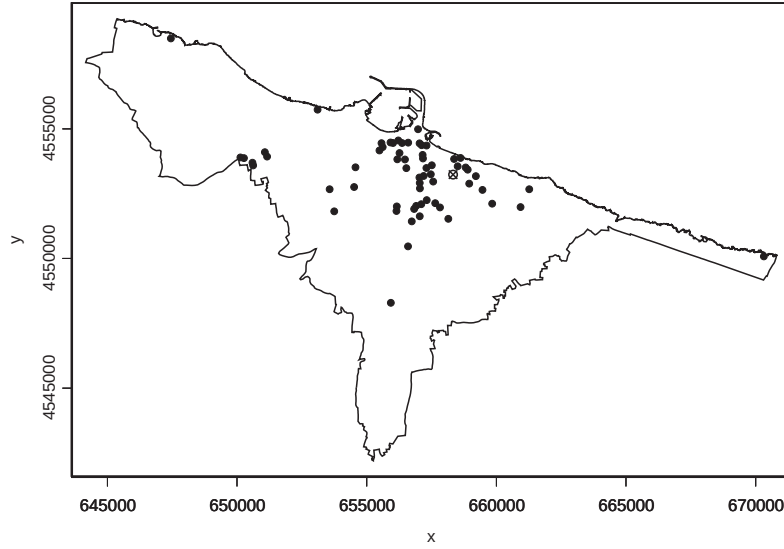


Figure 1: Urban area of Bari: spatial distribution of the places of residence of patients suffering of malignant mesothelioma between 1980 and 2001. The largest filled polygon indicates the location of a plant for the production of asbestos-cement.

Area-specific expected cases E_i s were obtained by internal standardization by means of 1991 census data. At the second level log-relative risks ϕ_i s are modelled as

$$[\phi_i | \alpha, \beta_1, \beta_2, \tau] \stackrel{\text{ind.}}{\sim} N \left[\alpha + \beta_1 d_i + \beta_2 \log(d_i), \tau^{-1} \right], \quad i = 1, \dots, 1329 \quad (2)$$

where d_i is the distance between the centroid of the i th section and that of the section where the asbestos-cement production plant is still located. Lawson (1993) (but see also Morris and Wakefield, 2000) first introduced the expression for the mean in (2) as the linear predictor in a Poisson regression model, in order to conveniently represent the dependence of the disease risk as a function of the distance from the putative source of hazard. In the hierarchical specification we adopted, random effects ϕ_i adjust for influent independent variables not included in the linear predictor and modify the form of the marginal likelihood, accounting for the large number of sampling zeros (under these circumstances the Poisson assumption (1) might be inadequate in a fixed parameters framework).

Independent priors are assumed for dependence parameters α , β_1 and β_2 and the random effects precision τ

$$[\alpha] \sim N(a, b^{-1}), \quad [\beta_1] \sim N(c, d^{-1}), \quad [\beta_2] \sim N(e, f^{-1}) \quad (3)$$

$$[\tau] \sim \text{Gamma}(a_\tau, b_\tau) \quad (4)$$

Within this framework, a natural posterior estimate of the adjusted relative risk in the i th section is given by

$$E \{ \exp [\alpha + \beta_1 d_i + \beta_2 \log(d_i)] | y \}, \quad i = 1, \dots, 1329 \quad (5)$$

This posterior summary can also be used in a predictive way as risk estimates for the tracts excluded from the data base can be easily obtained by calculating (5) with the corresponding distances $d_{i, \text{new}}$, $i = 1, \dots, 99$.

3. Results

Sampling-based estimates of model parameters and area-specific posterior expectations were obtained by Markov Chain Monte Carlo simulations, using Winbugs 1.3 (Spiegelhalter et al., 2000): after a burn-in phase of 2000 sweeps 1000 posterior samples were drawn for each of the 1430 census tracts from 12000 sweeps of an MCMC run, thinning the output to a ratio of 1 out of 10. Hyperparameters in (3) and (4) were set on the ground of non-informativeness to $a = c = e = 0$, $b = d = f = 0.01$ and $a_\tau = b_\tau = 0.01$.

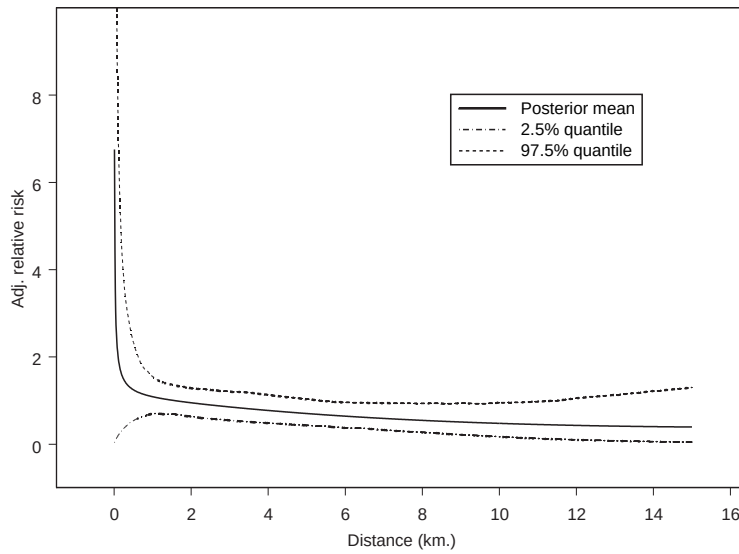


Figure 2: Estimated adjusted risk function based on the proposed model

The adjusted risk function in Fig. 2 was obtained by computing (5) for 1500 artificial distances from 0.01 to 15 km., two consecutive distances being differing of 0.01 km. The steep decrease around the point source suggested by the data is evident: the level of no-increased risk is found at 1.72 km., indicating a local excess of risk over a large area having approximately the form of a 4 km. diameter circle, presumably due to atmospheric diffusion of asbestos fibers.

The same conclusion can be obtained by looking at Fig. 3, drawn accordingly to the deciles of the empirical distribution of posterior relative risk estimates based on (5): the hypothesis of prevalence of the environmental exposure seems to be strengthened by the smooth risk diffusion (it is worth noticing that no spatially structured random effects were used in the proposed model, so that it is unlikely that the smooth appearance of the map in Fig. 3 is an artifact forced by unstructured random effects ϕ_i). Bayesian estimates summarized in Fig. 3 have an important advantage over the visual impact of the map in Fig. 1, in that they provide an accurate spatial quantification of the “increased-risk” area, and this may be helpful to health officers and authorities to decide health precautions in a efficient way.

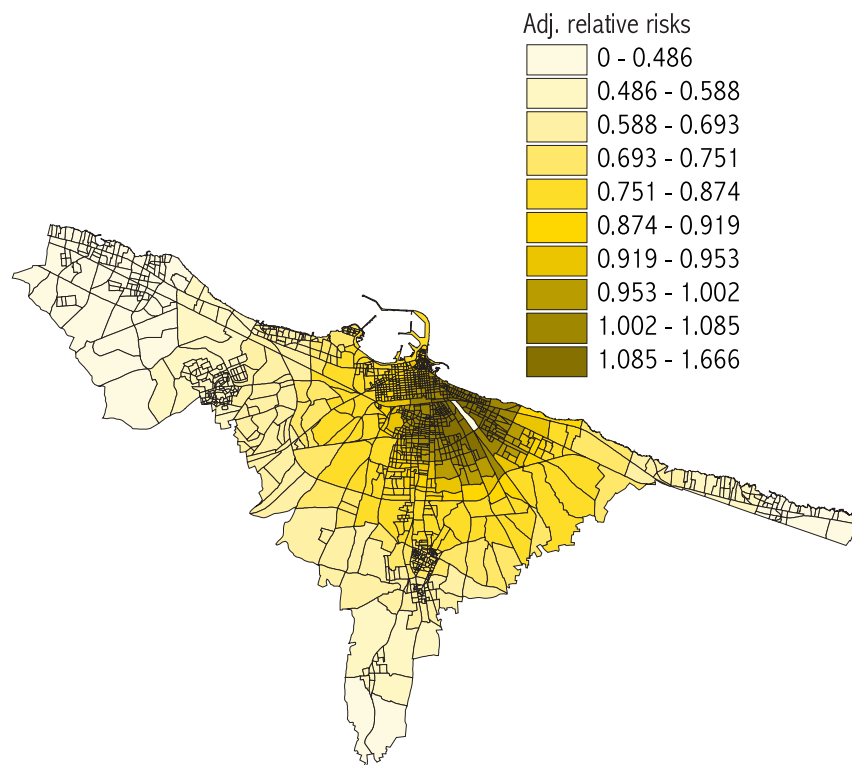


Figure 3: Map based on the deciles of the empirical distribution of sampling-based estimates of (5)

References

- Lawson A.B. (1993) On the analysis of mortality events associated with a prespecified fixed point, *J. Roy. Statist. Soc. A*, 157, 363–77.
- McDonald J.C. (1985) Health implications of enviromental exposure to asbestos, *Envir. Health Persp.*, 62, 319–28.
- Morris S.E. and Wakefield J.C. (2000) Assessment of disease risk in relation to a pre-specified source, in: *Spatial Epidemiology, Methods and Applications*, Elliot P., Wakefield J.C., Best N.G. and Briggs D.J., eds., Oxford University Press, New York, 153–84.
- Spiegelhalter D.J., Thomas A. and Best N.G. (2000) *Winbugs, Version 1.3, User Manual*, Available at: <http://www.mrc-bsu.cam.ac.uk/bugs>.