Spatial regression analysis between air pollution and childhood leukaemia in Portugal

M. Martinho · M. C. Freitas

Abstract This study aims to investigate whether known carcinogenic chemical elements in atmospheric deposition might be associated with child mortality due to leukemia in the Portuguese population. A Bayesian hierarchical model was used to explore the association between lichen bio-monitoring measurements of four elements—As, Hg, Ni, Pb—and childhood leukemia death counts taken at small administrative units. This geographical epidemiological study found a non-significant positive association between the risk of childhood leukemia and levels of arsenic, mercury and lead, and a non-significant negative association between the disease and the level of nickel. Lead seems to show a weaker association with childhood leukemia than arsenic and mercury.

Keywords Air pollution · Pollutants · Children · Leukaemia · Spatial regression · Epidemiology

Introduction

Air pollution is a complex mixture of many elements, of which several are known or suspected carcinogens. Arsenic and nickel have been classified as a group 1 human carcinogen by the International Agency for Research and Cancer [1, 2]. Mercury in the human body reacts with –SH groups and inhibits enzymes; fatal brain damage may occur [3]. Lead compounds—used till the 1990s as additives to gasoline in most countries—can evaporate to the air, be absorbed by the human body, cross the blood-brain barrier and lead to death [3]. Indirectly both elements may induce cancer due to changes in the cells, the bones, the nervous system, the lymph nodes, as well as in the functioning of the organs [4].

Previous studies on the association between air pollution and childhood leukemia focused mainly on air pollution from traffic [5, 6], but did not consider the effects of individual elements. In recent decades, a number of techniques have been developed to monitor concentrations of elements in the air. Biomonitoring is one of such techniques. It uses bio-organisms (biomonitors), such as mosses and lichens, to obtain quantitative information on certain characteristics of the biosphere [7, 8]. Lichens can be effectively used as biomonitors because they depend largely on atmospheric depositions for their nutrient supply, thus showing elemental compositions which reflect the gaseous, dissolved and/or particulate elements in the atmosphere [9].

This study investigates the relationship between the geographical variation of death counts of childhood leukemia (ICD9 204-208), taken in small administrative areas, and spatial variation of biomonitoring measurements of arsenic, mercury, nickel and lead. First, established geo-statistical methodology is used to estimate air pollution levels of these four elements for all administrative units on the basis of the collected data on biomonitoring measurements. Second, a spatial regression model is applied to investigate the relation between these estimated measurements and the childhood leukemia death counts, taking into account the spatial relationships among the small administrative areas and the rarity of the disease. All models are developed within a Bayesian framework and implemented with the Markov chain Monte Carlo method.

M. Martinho
United Nations Statistics Division, Two UN Plaza, New York, NY 10017, USA

M. C. Freitas (✉)
Reactor-ITN, Technological and Nuclear Institute, E.N. 10, 2686-953 Sacavém, Portugal
e-mail: cfreitas@itn.pt

In this spatial study, it is not ascertained whether individuals who have been exposed to air pollution are the same individuals who developed the disease. Therefore, statements about causal associations are inappropriate. However, this type of study is relatively inexpensive because it exploits already available health and environmental data and is thus valuable to generate hypotheses that can be pursued in future individual studies.

Experimental

Databases of chemical elements and mortality by leukemia

The data on chemical elements concentrations were obtained in a biomonitoring lichens survey performed in 1993 in continental Portugal. A few previous publications have described in detail the sampling and analytical procedures [10–14]. In summary, the epiphytic lichen *Parmelia sulcata* Taylor was collected in 260 sampling sites from olive trees, at 1.5–2 m away from the soil, and the samples were analyzed by k_0-Instrumental Neutron Activation Analysis (k_0-INAA) and Proton-Induced X-Ray Emission (PIXE). The location of the sampling sites was selected by a 10×10 km grid in the regions closer to the Atlantic coast while a 50×50 km grid was adopted in the interior of the country. The variability found in elemental concentrations at a site was on average 20% [15]. Figure 1a shows the sampling sites, marked as full circle points according to their geographical coordinates, latitude and longitude, over the map of continental Portugal containing the boundaries of the 275 counties.

For the purpose of the present work, the original 1993 lichen survey database [13], revised later for data quality checking [16], containing concentrations for 38 chemical elements is discussed for the well-known carcinogenic chemical elements: lead (Pb) and nickel (Ni), determined by PIXE and arsenic (As) and mercury (Hg), determined by k_0-INAA. The concentrations of these elements were previously reported [10, 11, 13, 14] and the number of results, minimum, maximum, and mean values (mg kg^{-1}), respectively, are as follows: As: 252, 0.4, 31, 2; Hg: 247, 0.09, 1.8, 0.3; Ni: 227, 0.8, 33, 4; Pb: 228, 2, 141, 19.

Recorded leukemia (ICD9 204-208) deaths were provided by the General Board of Health in Portugal (DGS) for the period 1991–1998. Over the eight year period, there were a total of 239 deaths from leukemia in children 0–14 years old in continental Portugal. For each death, age at death, sex, and code of administrative unit of residence (county) were recorded.

Statistical analyses

Since lichen biomonitoring measurements were not available for every county, Bayesian Gaussian kriging models [17] were applied to produce estimates of As, Hg, Ni and Pb levels for the 275 county centroids. These models assume that the data follow a multivariate Gaussian distribution with the elements Σ_{ij} of the correlation matrix expressed as a parametric function of the distance d_{ij} between pairs of points, $\Sigma_{ij} = \exp[-(\phi\ d_{ij})^{\kappa}]$, where the parameters $\phi > 0$, κ in (0, 2), are estimated by the model fitting. Large values of κ lead to greater smoothing. The parameter ϕ controls the rate of decline of the correlation with distance: large values of ϕ indicate rapid decay; small values imply a slow decay. As the distributions of the measurements of the four elements were skewed, the model was applied to the logarithm of the measurements.

To study the influence of the four elements—As, Hg, Ni, Pb—on childhood leukemia deaths, a Bayesian hierarchical model was used. The number of deaths O_i in county i is assumed to be Poisson distributed: $O_i \sim \text{Poisson}(E_i\ \theta_i)$, where the expected number of cases E_i for county i account for different population sizes and age and sex distributions among counties [18]—this standardization is necessary because childhood leukemia rates vary with age and sex. The expected number of cases for each county was calculated on the basis of population sizes from the 1991 census and age- and sex-specific disease rates for the 3–5 year age groups 0–4, 5–9, and 10–14 of the population under study.

The Poisson distribution is appropriate here because the observed number of deaths caused by childhood leukemia is small. The logarithm of θ_i is modeled as a linear function of the logarithm of the estimated exposures to As, Hg, Ni and Pb in area i plus area-specific parameters V_i and U_i:

$$\log(\theta_i) = \alpha + \beta_1\log(As_i) + \beta_2\log(Hg_i) + \beta_3\log(Ni_i) + \beta_4\log(Pb_i) + V_i + U_i. \tag{1}$$

The components U_i and V_i account for the effects of unmeasured or unknown risk factors not included in the model: U_i captures the influence of spatially correlated effects (e.g., other environmental pollutants or socioeconomic conditions) and V_i the effects which are independent across areas (e.g., data inaccuracies). U_i and V_i are assumed independent, V_i is assumed to be a realization from a Gaussian white noise with unknown variance and U_i is assumed to be distributed according to an intrinsic conditional autoregression accounting for dependence between spatially adjacent regions [19]. These random components were used to avoid incorrect inferences about the significance of the effects of the covariates [20].

All models were fitted to the data using the Markov chain Monte Carlo method implemented in WinBUGS 1.4 [21].

Fig. 1 **a** Map of sampling sites over the counties in continental Portugal, in 1993. **b** Map of childhood leukemia standard mortality ratios (SMR) at 275 counties in continental Portugal, in 1991–1998
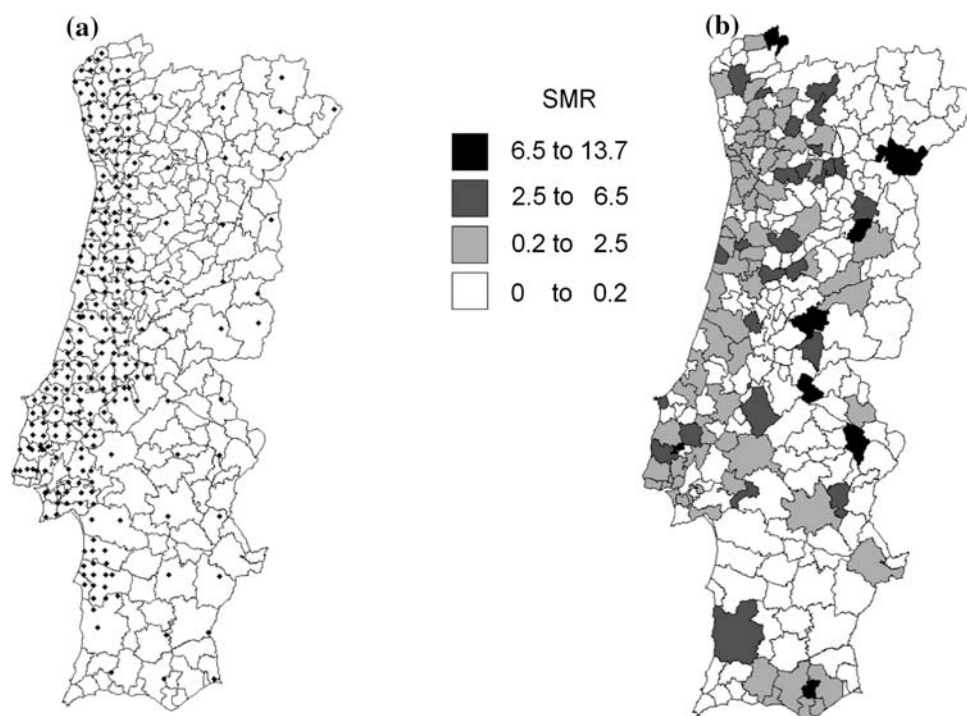
Results and discussion

Figure 1b shows the standardized mortality ratios [18] (SMRi = Oi/Ei) for the 275 counties. Figure 2 shows the posterior means for As, Hg, Ni and Pb values, estimated for the centroids in each county. Table 1 presents the posterior means and the 95% credible intervals (CIs) for the parameters ϕ and κ obtained with the Bayesian Gaussian kriging models. The model fitted to the mercury measurements shows the most rapid decay but also less smoothing, while the model fitted to the nickel concentrations has the slowest decay.

Table 1 also displays the posterior means and 95% CIs for the regression coefficients, as estimated by the Bayesian hierarchical model. The posterior means indicate positive associations with childhood leukaemia deaths for all elements except Ni, but the 95% CIs all include zero and, in this sense, the association is not significant. The lower limits of the 95% CIs for As and Hg are close to zero, thus suggesting a stronger association of the disease with these two elements than with lead.

One of the more sensitive aspects of this study is the measurement of air pollution levels and the estimation of the exposure levels of the population in each county. First, biomonitors are often not available in areas with excessive air pollution, such as cities with high population, which may lead to an underestimation of the pollution levels in the highly urban counties, like Braga, Coimbra, Lisbon and Porto. Second, as measurements were not taken in all counties, estimates were used for

the exposure in counties. Third, the air pollution values were considered as constant within each county, which may not reflect the true variability of the exposure of individuals.

In order to assess the impact of the lack of data in the main urban centers, measurements from alternative techniques for measuring air pollution were analyzed in the four cities listed above and in their vicinity ($PM_{2.5}$ data for atmospheric particles with aerodynamic diameter below 2.5 μm) [22–24]. For the vicinity of the cities, the $PM_{2.5}$ points included were those whose locations match the location of the lichen measurement points closest to the cities. The $PM_{2.5}$ concentrations in the neighbouring points of the cities are 1.2–1.6 times smaller than in urban centres, thus suggesting that data from these points are not enough to estimate the higher pollution levels in the cities. However, it must also be noticed that the counties encompass areas larger than the cities and thus the underestimation may be less marked for the whole county.

Since data were not available for all counties and the locations of the measurements may not necessarily be representative of the total exposure in the respective county, kriging models were used to estimate the county exposure values. Other authors have opted for aggregating the units in order to obtain at least one lichen biomonitoring measurement in each unit [16], but the analysis of larger areas increases the possibilities of bias on the regression coefficients due to non-negligible within area variability of the exposures [25]. Inverse distance weighting methods have also been applied for interpolation of

Fig. 2 Maps of estimated exposure to arsenic, mercury, nickel and lead at 275 counties in continental Portugal, in 1993
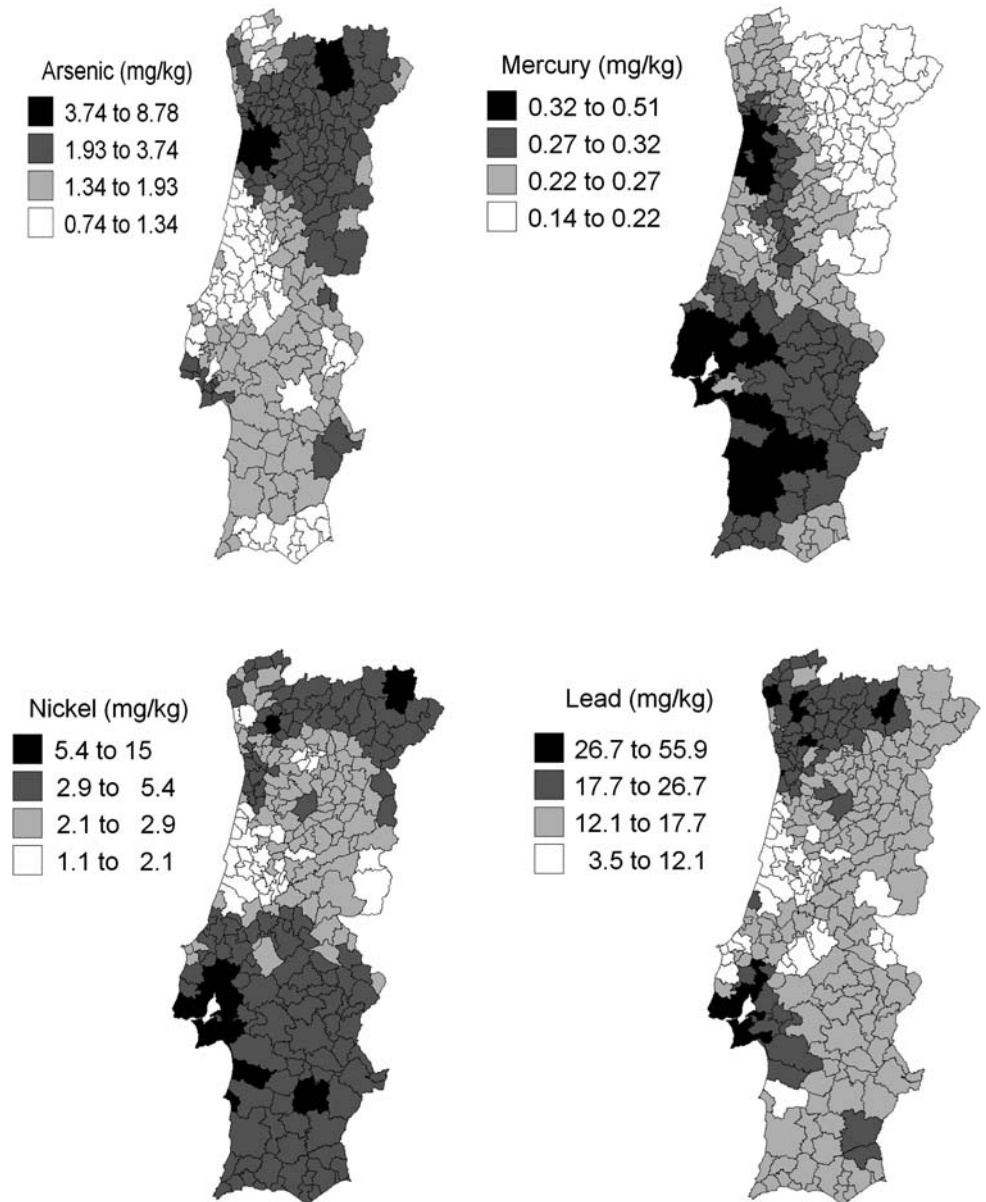

Table 1 Posterior means with 95% credible intervals (CI) in brackets for the parameters ϕ and κ and for the regression coefficients (β)

	Posterior mean and 95% CI		
	Bayesian Gaussian kriging models		Bayesian hierarchical model
	ϕ	κ	Regression coefficient (β)
Arsenic	4.33 (1.01; 9.34)	0.52 (0.12; 0.97)	0.25 (−0.10; 0.60)
Mercury	8.47 (0.35; 36.6)	0.30 (0.15; 0.41)	0.64 (−0.06; 1.36)
Nickel	2.98 (0.50; 7.20)	0.47 (0.31; 0.65)	−0.28 (−0.81; 0.28)
Lead	6.07 (0.76; 14.9)	0.50 (0.28; 0.77)	0.19 (−0.43; 0.76)

lichen biomonitoring measurements, but shown to be less compatible with these data than kriging models [26].

Since this study uses areal data, the exposure of interest is not available for each individual but rather for counties. If the exposures for all individuals within a county are not all the same, there is loss of information and inference on the effects of the exposures may be biased [27]. With the aim of assessing the within county variability of the exposures, the average and the standard deviation of the biomonitoring measurements of As, Hg, Ni and Pb were

calculated for the four counties with more than five air pollution measurements. The ratios of the standard deviation to the average vary between 0.2 and 0.9. Some of this variability is intrinsic to the measurements—the variability found in elemental concentrations at a site was on average 20%—but there seems to be additional variability across the counties. These counties are though the largest counties and less variability is expected in smaller counties.

Despite the risk of ecological bias, areal studies may be preferable than individual studies when these concern the effect of air pollution, as the exposure is difficult to measure individually and may be more reliable as a measurement over a small area.

Conclusions

The analysis reported here suggests a non-significant positive association between the risk of childhood leukemia and levels of arsenic, mercury and lead in the air; and a non-significant negative association between the disease and the level of nickel. Lead seems to show a weaker association with childhood leukemia than arsenic and mercury. The opinions herein are the authors', and not necessarily those of the United Nations.

References

1. IARC Working Group. IARC Monographs on the Evaluation of Carcinogenic Risk to Humans, vol. 23, Suppl. 7 (1987)
2. IARC Working Group. IARC Monographs on the Evaluation of Carcinogenic Risk to Humans, vol. 49 (1990)
3. Baird, C.: Toxic Organic Chemicals. In: W.H. Freeman and Company (ed.) Environmental Chemistry. W.H. Freeman and Company, NY, USA (1995)
4. Bowen, H.J.M.: Environmental Chemistry of the Elements. Academic Press, London (1979)
5. Best, N., Cockings, S., Bennett, J., Wakefield, J., Elliot, P.: Ecological regression analysis of environmental benzene exposure and childhood leukemia: sensitivity to data inaccuracies, geographical scale and ecological bias. J. R. Stat. Soc. A **164**, 155–174 (2001)
6. Raaschou-Nielsen, O., Hertel, O., Thomsen, B.L., Olsen, J.H.: Air pollution from traffic at the residence of children with cancer. Am. J. Epidemiol. **153**, 433–443 (2001)
7. Wolterbeek, H.T., Garty, J., Reis, M.A., Freitas, M.C.: Biomonitors in use: lichens and metal air pollution. In: Markert, B.A., Breure, A.M., Zechmeister, A.G. (eds.) Bioindicators and Biomonitors, pp. 377–419. Elsevier Science B.V., Amsterdam, The Netherlands (2002)
8. Wolterbeek, H.T.: Biomonitoring of trace element air pollution: principles, possibilities and perspectives. Environmental pollution. Environ. Poll. **120**, 11–21 (2002)
9. Bargagli, R.: Lichens as biomonitors of airborne trace elements. In: Trace Elements in Terrestrial Plants: An Ecophysiological Approach to Biomonitoring and Biorecovery, pp. 179–206. Springer, Berlin, Germany, (1998)
10. Reis, M.A., Alves, L.C., Wolterbeek, H.T., Verburg, T., Freitas, M.C., Gouveia, A.: Main atmospheric heavy metal sources in Portugal by biomonitor analysis. Nucl. Instr. Meth. Phys. Res. B **109/110**, 493–497 (1996)
11. Freitas, M.C., Reis, M.A., Alves, L.C., Wolterbeek, H.T., Verburg, T., Gouveia, A.: Biomonitoring of trace-element air pollution in Portugal: qualitative survey. J. Radioanal. Nucl. Chem. **217**, 21–30 (1997)
12. Freitas, M.C., Reis, M.A., Alves, L.C., Wolterbeek, H.T.: Distribution in Portugal of some pollutants in the lichen *Parmelia sulcata*. Environ. Poll. **106**, 229–235 (1999)
13. Reis, M.A.: Biomonitoring and assessment of atmospheric trace elements in Portugal. Methods, response modeling and nuclear analytical techniques. Ph.D. Thesis, Technical University of Delft, The Netherlands (2001)
14. Freitas, M.C., Reis, M.A., Alves, L.C., Wolterbeek, H.T.: Nuclear analytical techniques in atmospheric trace element studies in Portugal. In: Markert, B.A., Friese, K. (eds.) Trace Elements—Their Distribution, Effects in the Environment, pp. 187–213. Elsevier, Amsterdam, The Netherlands (2000)
15. Freitas, M.C., Nobre, A.S.: Bioaccumulation of heavy metals using *Parmelia sulcata* and *Parmelia caperata* for air pollution studies. J. Radioanal. Nucl. Chem. **217**, 17–20 (1997)
16. Sarmento, S., Wolterbeek, H.T., Verburg, T.G., Freitas, M.C.: Correlating element atmospheric deposition and cancer mortality in Portugal: data handling and preliminary results. Environ. Poll. **151**, 341–351 (2008)
17. Diggle, P.J., Moyeed, R.A.: Model-based geostatistics. Appl. Statist. **47**, 299–350 (1998)
18. Estève, J., Benhamou, E., Raymond, L.: Statistical Methods in Cancer Research, vol. IV—Descriptive Epidemiology, International Agency for Research on Cancer (1994)
19. Besag, J., York, J., Mollié, A.: Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Statist. Math. **43**, 1–59 (1991)
20. Cliff, A.D., Ord, J.K.: Spatial Processes: Models and Applications. Pion, London (1981)
21. Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D.: GeoBUGS User Manual, Version 1.2 (2004). http://www.mrc-bsu.cam.ac.uk/bugs/
22. Freitas, M.C., Farinha, M.M., Ventura, M.G., Almeida, S.M., Reis, M.A., Pacheco, A.M.G.: Gravimetric and chemical features of airborne PM_{10} and $PM_{2.5}$ in mainland Portugal. Environ. Monit. Assess. **109**, 81–95 (2005)
23. Pio, C.A., Ramos, M.R., Duarte, A.C.: Atmospheric aerosol and soiling of external surfaces in an urban environment. Atmos. Environ. **32**, 1979–1989 (1998)
24. Portuguese Environment Agency, Portuguese Ministry of Environment (2006). http://www.apambiente.pt
25. Greenland, S., Morgenstern, H.: Ecological bias, confounding, and effect modification. Int. J. Epidemiol. **18**, 269–274 (1989)
26. Wolterbeek, H.T., Verburg, T.G.: Judging survey quality in biomonitoring. In: Wiersma, G.B. (ed.) Environment Monitoring, pp. 583–604. CRC Press, Boca Raton, FL (2004)
27. Robinson, W.: Ecological correlations and the behavior of individuals. Am. Sociol. Rev. **15**, 351–357 (1950)