Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2002/04

A Space-Time analysis of the relationship between Socioeconomic factors and Mortality for Lung Cancer

Emanuela Dreassi

Università degli Studi di Firenze

Applied Statistics

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

A Space-Time analysis of the relationship between Socioeconomic factors and Mortality for Lung Cancer

Emanuela Dreassi

Department of Statistics "G. Parenti", University of Florence, Italy

Summary

The relationship between socioeconomic factors and mortality for lung cancer is investigated. A space-time hierarchical Bayesian model with time dependent covariates, to define the proper lag time on which lung cancer mortality is affected by socioeconomic factors, is adopted. An example on males lung cancer in Tuscany (Italy) during the period 1971-99 is provided. Results confirm the presence of an association between mortality for lung cancer and socioeconomic factors with a temporal lag of ten-fifteen years (latency time).

Key words: Hierarchical Bayesian models; Material deprivation; Space-time models; Time-dependent covariates.

1 Introduction

The relationship between socioeconomic factors and health has been studied in many circumstances, mainly on aggregated data. Some studies suggest that, in terms of control for potential confounding, knowledge of the socioeconomic factors is crucial. In fact, a strong association between these and mortality, on one side, and, between these and exposure to environmental hazards, on the other side, has been repeatedly observed; see for example St Leger (1995).

However it's important to investigate the proper lag time on which socioeconomic factors affect mortality. On mortality for lung cancer socioeconomic factors are assumed to be associated with exposure risk factors (e.g. tobacco smoking habit or occupational) but the process of carcinogenesis suggests that a lag of ten-fifteen years (a latency time) between exposure and mortality is present; for some tips on spatial analysis of cancer mortality and theoretical models that incorporate the process of carcinogenesis, see Glick (1982) and Mayer (1983).

Socioeconomic factors can be summarized by a material deprivation index. Material deprivation indicators usually refer to the prevalence of subject characteristics such as unemployment, low education, living in a small dwelling, overcrowding, not having a car; e.g. see Townsend *et al.* (1988).

The mean goal of this paper is to study the ecological association between lung cancer mortality and material deprivation indicators: we present a space-time hierarchical Bayesian model with time dependent covariates to investigate the contribution of socioeconomic factors to mortality for lung cancer on males in Tuscany in the period 1971-99. The space-time dynamics of the association between mortality and deprivation level has been studied using a different deprivation score which result in meaningful absolute level across the years. We consider socioeconomic factors observed in 1961, 1971, 1981 and 1991 and we look for temporal lag between these and mortality.

2 Data

Lung cancer death certificates are considered for males resident in 287 municipalities of Tuscany Region (Italy) from 1971 to 1999; the population for the 29 years analyzed is a total of 49,684,302 person years and the total number of collected death certificates was 47,343. Data have been made available by the Tuscany Regional Government under the research project *Tuscany Atlas of Mortality 1971-1994* for different calendar period, 1971-74, 1975-79, 1980-84, ..., 1990-94 (see Vigotti *et al.*, 2001) and by the Regional Mortality Register for the period 1995-99. A set of reference rates (Tuscany, 1971-99) have been used to define expected cases for each municipality and period, following direct standardization classifying the population by 18 age classes (0-5, ..., 85 or more). Table 1 describes age and period specific rates ($\times$ 100,000) and the number of cases. For a complete space-time analysis of a subset of this data (males lung cancer in Tuscany from 1971 to 1994), that consider the presence of a heavy cohort effect, see Lagazio *et al.* (2002).

Data on material deprivation have been composed from census data collected from ISTAT on the years 1961, 1971, 1981 and 1991. In particular the used variables (accurately standardized, see Gordon, 1995) are the proportions of population with unfavorable events such: unemployment, low education (less than 6 years of schooling), not having the property of the house, and absence of bathroom in the flat.

3 Space-time models with time dependent covariates

We describe the space-time pattern of mortality risk for lung cancer for the whole region from 1971 to 1999 using a hierarchical Bayesian model with structured random effects on space and time dimension. We adopt the hierarchical Bayesian space-time model defined by Bernardinelli *et al.* (1995) to estimate risk for each period and for each municipality. In this model, the observed cases O_{ij}, in the i-th area ($i = 1, \ldots, 287$) and j-th period ($j =$ 1971-74, 1975-79, 1980-84, 1985-89, 1990-94, 1995-99), are assumed to have a Poisson distribution with mean $E_{ij}\theta_{ij}$, where E_{ij} indicates expected cases under indirect standardization and θ_{ij} the relative risk. A random effects model is assumed for the logarithm of relative risk

$$\log(\theta_{ij}) = u_i + v_i + (z_i + r_i) \quad p_j \tag{1}$$

The term u_i represents an unstructured spatial variability component modelled as Normal (μ_u, λ_u) and v_i a structured spatial variability term modelled, conditional to $v_{\neq i}$ terms, as Normal $(\bar{v_i}, \lambda_v n_i)$, where $\bar{v_i}$ is the mean of the i-th adjacent areas terms and n_i their number (conditional autoregressive model). These terms define the component of the pure spatial model of Besag *et al.* (1991). The term p_j indicates period: equal to zero when $j =$ 1971-74, one and two respectively when $j =$ 1975-79 and $j =$ 1980-84 and so on. Time differences for each i-th area are assumed both spatially structured and unstructured:

$z_i \sim$ Normal (μ_z, λ_z) and $r_i \mid r_{\neq i} \sim$ Normal $(\bar{r_i}, \lambda_r n_i)$, where $\bar{r_i}$ denotes the mean of i-th adjacent areas terms and n_i their number, as before. Non informative Inverse Gamma are assumed for hyperparameters λ and non informative Normal for μ. The posterior distribution for the parameters is approximated by Monte Carlo Markov Chain methods, checks for achieved convergence of the algorithm was performed following Gelman and Rubin (1992); for the software see Spiegelhalter *et al.* (2000).

The effect of socioeconomic factors in different times have been considered by introducing covariates into the hierarchical Bayesian model. A simple descriptive strategy requires to analyze mortality and deprivation for each time span separately. Considering more than one time span needs some caution as there are two space-time misaligned processes.

The model has been specified considering four different temporal lags (0, 5, 10 and 15 years) between census (1961, 1971, 1981 and 1991) and mortality observed in different calendar periods (1971-74, 1980-84, ..., 1995-99). When a temporal lag of zero years has been considered a relation between mortality in calendar period 1971-74 and census 1971, period 1980-84 and census 1981, period 1990-94 and census 1991 have been defined. For a lag of 5 years the model defines a relation between mortality in calendar period 1975-79 and census 1971, period 1985-89 and census 1981, period 1995-99 and census 1991. For a lag of 10 the association has been stated between period 1971-74 and census 1961, period 1980-84 and census 1971 and period 1990-94 and census 1981. For 15 years of lag the association between mortality and material deprivation has been defined between period 1975-79 and census 1961, period 1985-89 and census 1971, period 1995-99 and census 1981. See Table 2 for a schematic view of the association between mortality on each period and material deprivation observed at different censuses defined by the next model (2).

The material deprivation effect has been considered introducing a specific term on model (1)

$$\begin{aligned} \log(\theta_{ij}) &= u_i + v_i + (z_i + r_i) \quad p_j \\ &+\beta \left[(w_0 x_i(j) + w_{10} x_i(j-10))\, I_{0,10}(j) + (w_5 x_i(j-5) + w_{15} x_i(j-15))\, I_{5,15}(j) \right] \end{aligned} \tag{2}$$

where $x_i(t)$ represents the material deprivation index for the i-th area observed at t-th calendar time (censuses 1961, 1971, 1981, 1991) for a lag of $l = 0, 5, 10, 15$ years between mortality and material deprivation. The weights assumed for each temporal lag considered w_0, w_5, w_{10} and w_{15} are assumed uniformly distributed on $[0, 1]$, $I'_{0,10} = (1, 0, 1, 0, 1, 0)$ and $I'_{5,15} = (0, 1, 0, 1, 0, 1)$ represent two indicator vectors.

The coefficient β is modelled as non informative Normal distribution and defines the relation between material deprivation and mortality. An alternative formulation is to consider terms β_j, temporally structured (using a first-order random walk with independent gaussian increments) or β_i, spatially structured; for a review on space varying coefficient models see Assunção (2001).

The model assumes that carcinogenesis process takes a latency time (e.g. a time equal to l), so mortality on time j is associated with a covariate observed in time $t = j - l$. The effect of material deprivation on mortality could be assumed as sum of different time dynamic effects. The model divide the material deprivation effect β on four different time terms. We estimate their weight $(w_0, w_5, w_{10}, w_{15})$ on the coefficient β.

4 Results

Lung cancer mortality in Tuscany (males) exhibits a strong increase in the last thirty years (1971-1999). The distribution of relative risks of mortality for lung cancer is highly spatially structured with north-west areas at higher risk. Figure 1 show the relative risk for lung cancer mortality on six different calendar period obtained from space-time model (1).

Differentials on geographical mortality, after removing space and time effects, could be associated with material deprivation.

Material deprivation scores derived at municipality level for censuses from 1961 to 1991 highlighted areas with earlier industrialization. More recently there is a clustering along the northern boundary and the coastline. Surprisingly, material deprivation index shows a similar strong spatial distribution with higher values in the north-west part of the region and on the coast; see Figure 2.

We first separately analyze the association between mortality and material deprivation for each period and each census using a pure spatial model (as defined by Besag *et al.*, 1991) with a single covariate at a given time period. Let's start with Lung cancer mortality 1971-79. Material deprivation (quintiles) at 1961 census appeared to be associated with lung cancer mortality with a latency of 5-15 years (respectively a RR of 1.48 and 1.38 for the most deprived compared to the less deprived); see figure 3a and 3b. Mortality in the eighties appeared to be strongly associated with deprivation at 1961 (more than 20 years of latency, for a RR of 1.46 and 1.24); see figure 3c and 3d. Again the most evident association in the nineties was between mortality and deprivation level 10-15 years before (RR equal to 1.36 and 1.24), figure 3e and 3f. Material deprivation at 1981 census was associated also with mortality in the eighties, with very short latency time. Since 1981 was a period of economic crisis in Italy, concurrent factors cannot be excluded.

Using the model (2), having achieved convergence of the MCMC algorithm, we retained 5,000 samples and we estimated the posterior distribution. Figure 4 describes the posterior distributions for the β coefficient of the model considered. The mean value for β coefficient is 0.104 (95% confidence interval: 0.072 - 0.140), for a RR of 1.11 (95% confidence interval 1.075 - 1.150). Figure 5 describe the weights for different lags considered. For lung cancer, males, the more appropriate lag resulted to be ten-fifteen years. The mean values of posterior distributions for the weight are: for a lag of zero years 0.1357, for 5 years 0.5804, for ten years 0.8643, for fifteen years 0.4196.

5 Conclusions

It turns out that mortality for lung cancer, males, is associated to material deprivation level of ten-fifteen years before. Nevertheless the study relies on aggregate data and no definite conclusions can be drawn, since ecological biases cannot be excluded. The relationship between an individual's cumulative exposure to a risk agent and the spatial pattern of the associated socioeconomic factors is complex: on the transformation process of an input map of socioeconomic conditions into an output map of cancer mortality, additional distortion must be taken into account.

However, the absence of a clear association between mortality for lung cancer and material deprivation assessed at the same time period (lag zero) suggests that factors related to the Health Systems played a minor role.

The strong evidence of an association with deprivation level at least ten years before mortality being recorded, stems for a correlation between material deprivation and exposure to etiologic factors. Consistently with some recent literature on socioeconomic inequalities in health, our results support the hypothesis that economically disadvantaged population groups are more exposed to risk factors, i.e. tobacco smoke.

The barriers to utilization of health service and to obtain appropriate care for people lying in a poor material condition can probably affect other causes of death than lung cancer (e.g. ischemic hearth diseases), for which effective therapy is limited. For these causes, preference to affluent people is presumed, or, because of an higher instruction, only affluent people can take advantage of these therapy.

References

Assunção, R.M. (2001). Space Varying Coefficient Models for Small Area Data, *Unpublished manuscript*.

Besag, J., York, J., Mollié, A. (1991). Bayesian image restoration, with applications in spatial statistics (with discussion), *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.

Bernardinelli, L., Clayton, D., Pascutto, C. (1995). Bayesian analysis of space-time variation in disease risk, *Statistics in Medicine*, **14**, 2433–2443.

Gelman, A., Rubin, D.R. (1992). Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science*, **7**, 457–511.

Glick, B.J. (1982). The Spatial Organization of Cancer Mortality, *Annals of the Association of American Geographers*, **72**, 4, 471–481.

Gordon, D. (1995). Census based deprivation indeces: their weighting and validation, *Journal of Epidemiology & Community Health*, **49**, suppl. 2, S39–S44.

Lagazio, C., Biggeri, A., Dreassi, E. (2002). Age-Period-Cohort Models on Disease Mapping, *Unpublished manuscript*.

Mayer, J.D. (1983). The role of Spatial Analysis and geographic data in the detection of disease causation, *Social Science & Medicine*, **17**, 1212–1221.

Spiegelhalter, D.J., Thomas, A., Best, N.G., Gilks, W.R. (2000). *WinBugs*, Medical Research Council Biostatistics Unit, Cambridge.

St Leger, S. (1995). Use of deprivation indices in small area studies of environment and health, *Journal of Epidemiology & Community Health*, **49**, 1–88.

Townsend, P., Phillimore, P., Beattie, A. (1988). *Health and deprivation: inequalities and the north*, London: Croom Helm.

Vigotti, M.A., Biggeri, A., Dreassi, E., Protti, M. A., Cislaghi, C. (2001). *Atlas of mortality in Tuscany 1971-94*, Edizioni Plus, Università degli studi di Pisa.

Table 1. Age-period specific mortality rates (× 100,000) and number of deaths. Lung cancer, males, Tuscany (Italy), 1971-1999.

	Period											
Age	1971-74		1975-79		1980-84		1985-89		1990-94		1995-99	
0–4	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)
5–9	0	(0)	0.167	(1)	0.189	(1)	0	(0)	0	(0)	0	(0)
10–14	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)
15–19	0	(0)	0	(0)	0	(0)	0	(0)	0.184	(1)	0	(0)
20–24	0	(0)	0.729	(4)	0.164	(1)	0	(0)	0.156	(1)	0	(0)
25–29	0.846	(4)	1.002	(6)	0.707	(4)	1.287	(8)	0.292	(2)	0	(0)
30–34	2.294	(11)	2.202	(13)	1.182	(7)	2.281	(13)	0.953	(6)	1.139	(8)
35–39	5.690	(27)	7.415	(45)	7.171	(43)	4.345	(26)	5.741	(33)	2.498	(16)
40–44	21.611	(106)	22.764	(135)	19.425	(116)	19.750	(117)	14.356	(86)	8.242	(48)
45–49	43.255	(218)	44.661	(273)	47.661	(278)	46.078	(273)	32.182	(190)	26.567	(161)
50–54	75.662	(342)	91.121	(568)	108.292	(642)	85.201	(489)	76.188	(446)	68.585	(398)
55–59	128.517	(467)	150.718	(780)	187.735	(1112)	180.141	(1031)	143.195	(802)	113.619	(650)
60–64	190.339	(816)	229.643	(1030)	247.445	(1225)	292.444	(1633)	259.192	(1413)	201.356	(1081)
65–69	279.517	(960)	319.244	(1500)	370.082	(1445)	352.777	(1548)	394.758	(2018)	345.721	(1731)
70–74	286.059	(714)	385.879	(1324)	432.871	(1672)	468.204	(1562)	456.730	(1740)	478.094	(2123)
75–79	264.585	(385)	344.589	(743)	443.843	(1092)	495.515	(1449)	521.901	(1378)	518.007	(1537)
80–84	145.496	(128)	270.123	(293)	407.835	(517)	489.365	(756)	487.923	(971)	532.007	(963)
85 +	129.855	(57)	153.766	(92)	257.543	(165)	345.532	(269)	365.508	(385)	445.953	(619)
	61.927	(4235)	78.318	(6807)	96.595	(8320)	107.409	(9174)	111.353	(9472)	109.909	(9335)

Table 2. The different lags on which mortality and material deprivation are associated

	Temporal lag			
Period	lag 0	lag 5	lag 10	lag 15
1971-74	1971	-	1961	-
1975-79	-	1971	-	1961
1980-84	1981	-	1971	-
1985-89	-	1981	-	1971
1990-94	1991	-	1981	-
1995-99	-	1991	-	1981

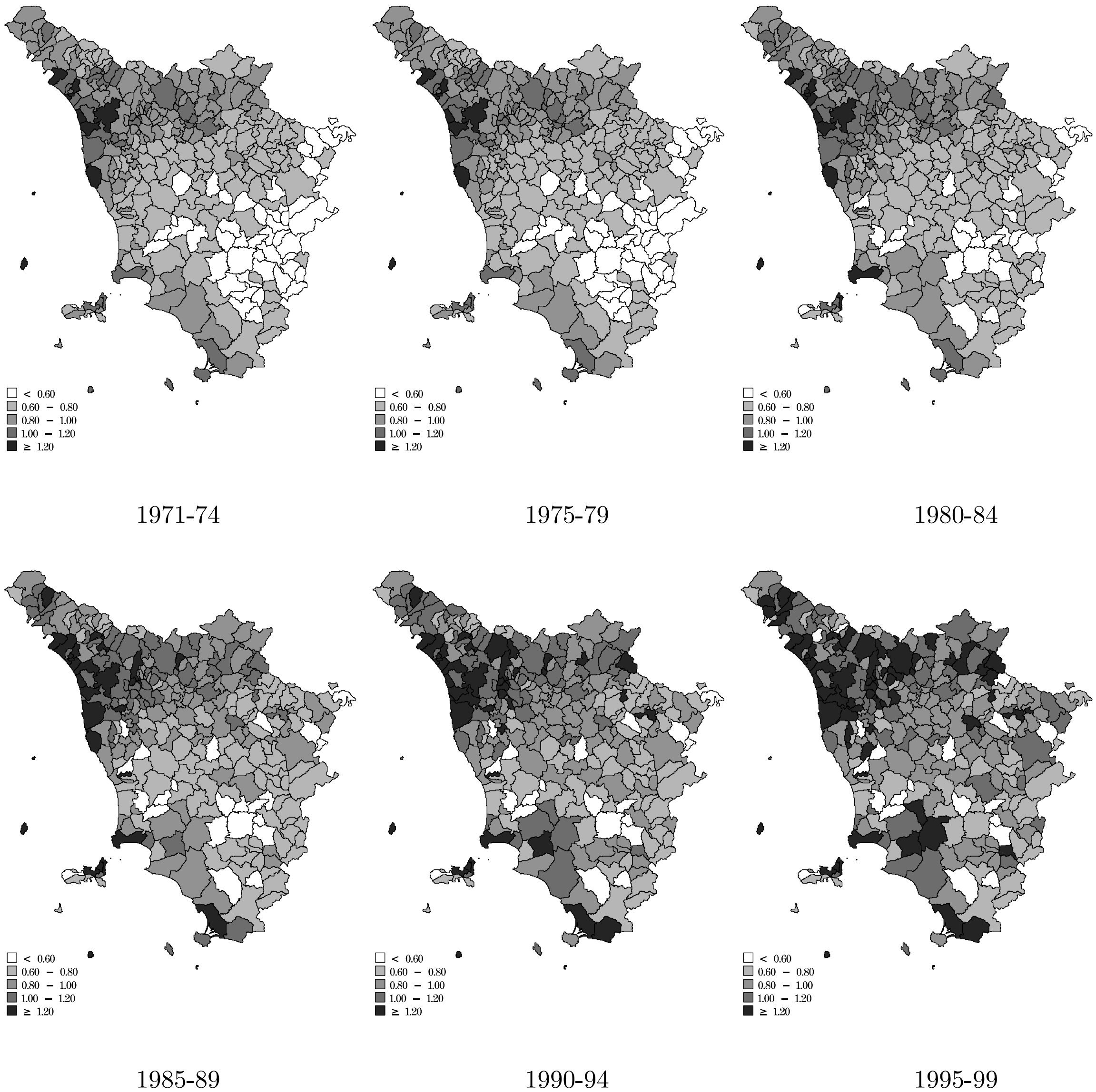


Figure 1: Space-Time distribution of relative risk for lung cancer, males, Tuscany (Italy) from 1971 to 1999

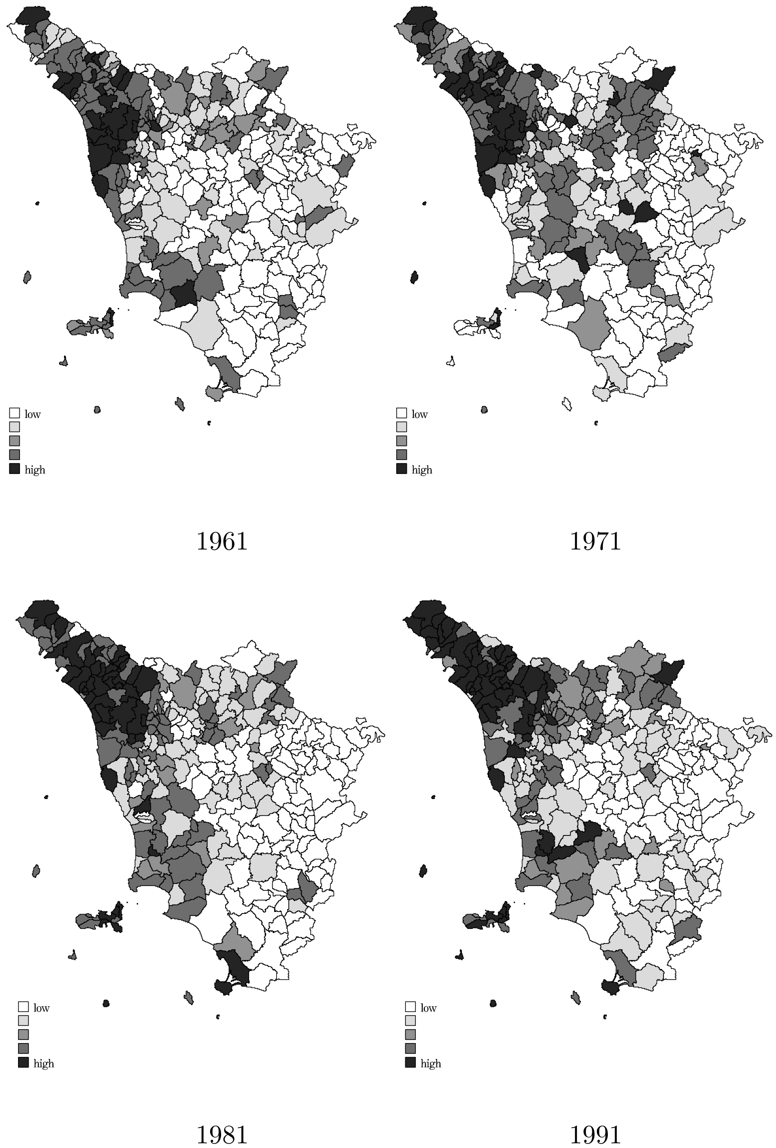
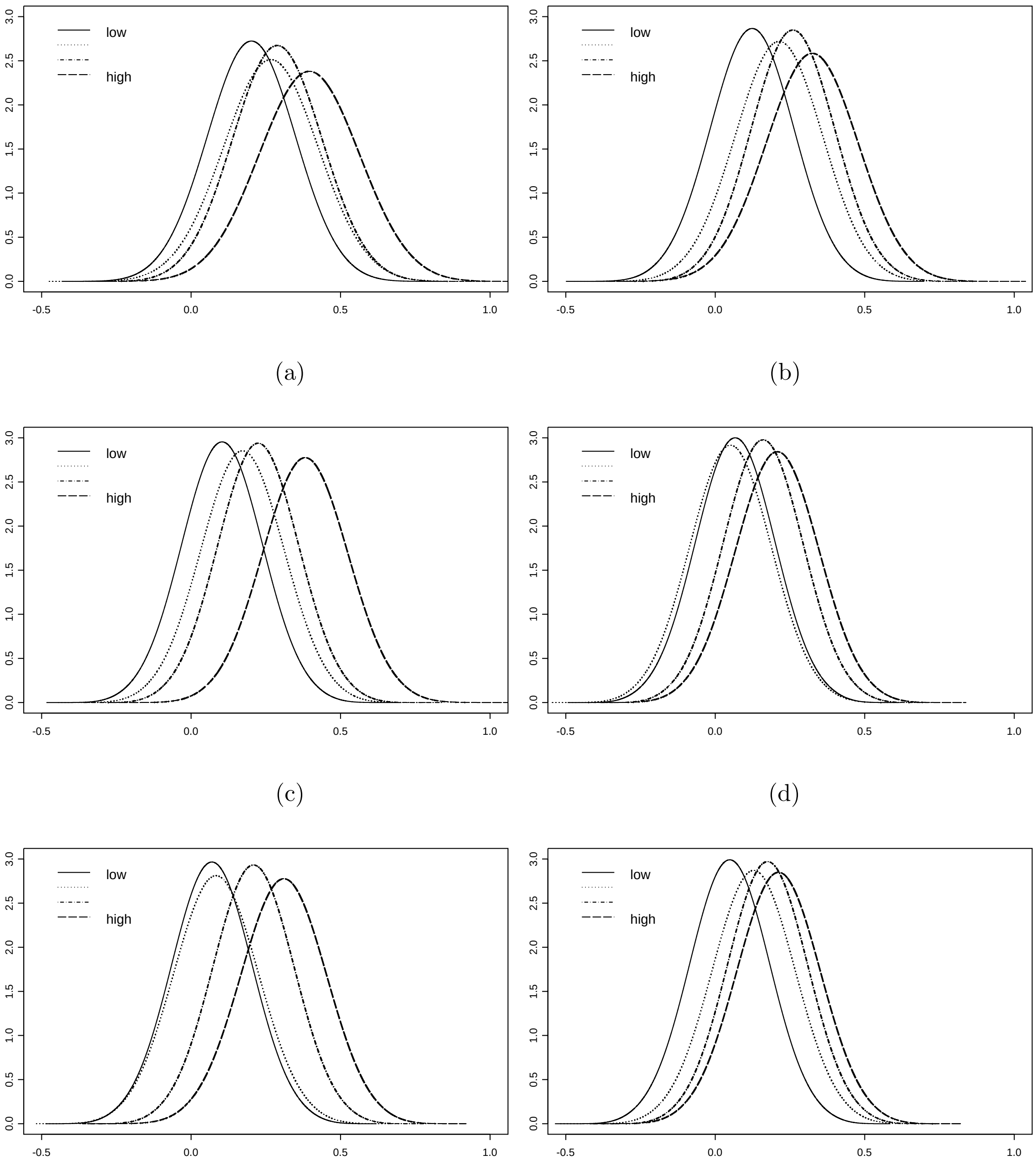


Figure 2: Spatial distribution of material deprivation index in Tuscany (Italy) at the four different censuses



(a) (b)

(c) (d)

(e) (f)

Figure 3: Regression Coefficient for separate pure spatial analyses for material deprivation (quintiles) vs mortality (reference lower material deprivation): (a) period 1971-74 vs 1961's census; (b) period 1975-79 vs 1961's census; (c) period 1980-84 vs 1961's census; (d) period 1985-89 vs 1961's census; (e) period 1990-94 vs 1981's census; (f) period 1995-99 vs 1981's census

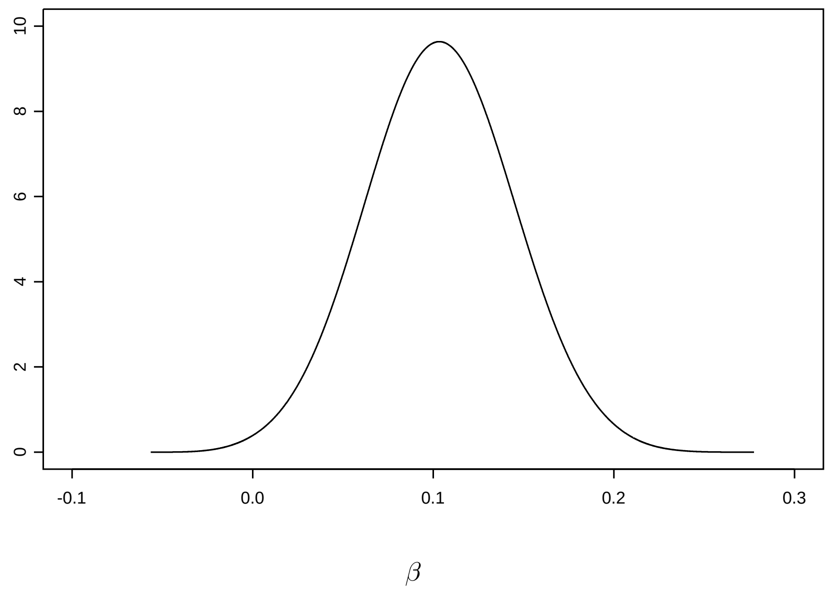


Figure 4: Posterior distribution of the Regression Coefficient β for material deprivation vs mortality for the complete space-time analysis

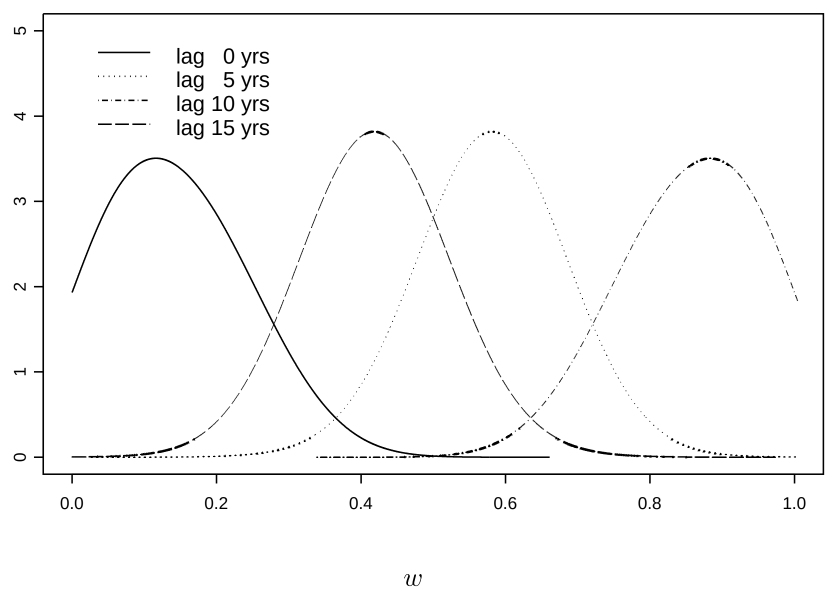


Figure 5: Posterior distribution of weights for different temporal lags between mortality and material deprivation