



Universitat de Girona

**MÉTODOS DE ESTADÍSTICA ESPACIAL PARA
EVALUAR LA INFLUENCIA DE FACTORES
MEDIOAMBIENTALES SOBRE LA INCIDENCIA
Y MORTALIDAD POR CÁNCER**

Aitana LERTXUNDI MANTEROLA

ISBN: 978-84-690-6047-6

Dipòsit legal: GI-404-2007

**Métodos de estadística espacial para evaluar la influencia
de factores medioambientales sobre la incidencia y
mortalidad por cáncer**

Aitana Lertxundi Manterola

Noviembre, 2006

El doctor **Marc Saez Zafra**, Catedràtic d'Universitat del Departament d'Economia de la Universitat de Girona,

AUTORITZA:

La presentació de la Tesis Doctoral amb títol '**Métodos de Estadística Espacial para Evaluar la Influencia de Factores Medioambientales sobre la Incidencia y Mortalidad por Cáncer**', realitzada per la Sra. **Aitana Lertxundi Manterola**, sota la meua immediata direcció i supervisió i que presenta per obtenir el grau de Doctor per la Universitat de Girona.

Girona, 10 de novembre de 2006

Dr. Marc Saez Zafra
Director

Aitana Lertxundi Manterola
Doctorand

1.- Introducción	
1.1.- Mortalidad e incidencia en España, en Cataluña y en la Regió Sanitària Girona	1
1.1.1.- Mortalidad por cáncer en España, en Cataluña y en la Regió Sanitària Girona	1
Mortalidad por cáncer en España	1
Mortalidad por cáncer en Cataluña	3
Mortalidad por cáncer en la Regió Sanitària Girona	7
1.1.2.- Incidencia de cáncer en España, en Cataluña y en la Regió Sanitària Girona	9
Incidencia de cáncer en España	9
Incidencia de cáncer en Cataluña	11
Incidencia de cáncer en la Regió Sanitària Girona	13
1.2.- Cáncer y medio ambiente	15
1.2.1.- Contaminación atmosférica	18
1.2.2.- Hidrocarburos policíclicos aromáticos (PAH)	19
1.2.3.- Metales	21
1.2.4.- Asbestos	23
1.2.5.- Radiación	24
1.2.6.- Pesticidas	29
1.2.7.- Disruptores endocrinos	30
1.3.- Neoplasias de interés ambiental	32
1.3.1.- Neoplasias Linfoides (CIE-10: C81-C85; C90-C91; C96)	33
1.3.2.- Cáncer de tiroides (CIE-10: C73; CIE-9: 193)	37
1.3.3.- Cáncer de vejiga urinaria (CIE-10: C67; CIE-9: 188)	39
1.3.4.- Cáncer de tráquea, bronquios y pulmón (CIE-10: C34; CIE-9: 162)	41
Referencias	43
Hipótesis	61
Objetivos	62
2.- Modelos estadísticos aplicados a datos espaciales	
2.1.- Estadística espacial	63
2.2.- Tipos de datos espaciales	63
2.2.1.- Datos en rejilla ('lattice data')	64
2.2.2.- Datos geoestadísticos ('geostatistical data')	65
2.2.3.- Procesos puntuales ('point processes')	66
2.3.- Epidemiología espacial	66
2.4.- Variables respuesta: Incidencia y mortalidad por neoplasias de interés medioambiental. Mapas de enfermedades	68
2.4.1.- Casos Brutos y Razones Estandarizadas	69
2.4.2.- Suavización de Razones Estandarizadas	75
2.4.2.1.- Suavización mediante Modelos Lineales Generalizados Mixtos (GLMM)	75
2.4.2.2.- Suavización mediante Modelos Lineales Generalizados Mixtos (GLMM) con efectos espaciales	76
2.4.3.- Estructura de dependencia espacial	83
2.4.3.1.- Modelos autorregresivos de Series Temporales (AR)	83
2.4.3.2.- Autoregresión Espacial Simultánea (SAR)	85
2.4.3.3.- Autoregresión Espacial Condicional (CAR)	87
2.4.3.4.- Autoregresión Espacial Condicional L1 (L1-CAR)	89
2.4.3.5.- Matriz de ponderaciones espaciales	89
2.5.- Variables explicativas	90
2.5.1.- Datos en rejilla	90
2.5.2.- Análisis de datos geoestadísticos	91
2.5.2.1.- Análisis de dependencia espacial	91
2.5.2.2.- Modelos autorizados para el semivariograma	98
2.5.2.3.- Estimación de datos geoestadísticos	103
2.5.2.3.1.- Análisis univariante	103
2.5.2.3.2.- Análisis multivariante. Cokriging	111
2.6.- Estudios de asociación geográfica	117
2.6.1.- Modelos Lineales Generalizados Mixtos con variables explicativas con diseño en rejilla	118
2.6.2.- Modelos Lineales Generalizados Mixtos con variables explicativas con diseño geoestadístico	119
Referencias	124

3.- Metodología. Análisis descriptivo	
3.1.- Población	130
3.2.- Sujetos de estudio	130
3.3.- Diseño	137
3.4.- Variables dependientes	137
3.5.- Variables explicativas	138
-Variables socioeconómicas	138
-Variables medioambientales	146
3.6.- Análisis descriptivo de los datos	169
-Incidencia de las neoplasias de interés medioambiental en la Regió Sanitària Girona	169
-Mortalidad de las neoplasias de interés medioambiental en el Área Metropolitana de Barcelona	176
-Mortalidad de las neoplasias de interés medioambiental en el Área Metropolitana de Bilbo	183
Referencias	190
4.- Metodología. Modelización	
4.1.- Suavización de las Razones Estandarizadas	191
4.1.1.- Especificación del modelo	191
4.1.2.- Estimación del modelo	193
4.1.3.- Resultados de la estimación	195
a.-Incidencia de las neoplasias de interés medioambiental	195
b.-Modelos de asociación geográfica	212
4.1.4.- Resultados de la estimación. Área Metropolitana de Barcelona	225
a.-Mortalidad de las neoplasias de interés medioambiental en el Área Metropolitana de Barcelona	225
b.-Modelos de asociación geográfica	240
4.1.5.- Resultados de la estimación. Área Metropolitana de Bilbo	251
a.-Mortalidad de las neoplasias de interés medioambiental en el Área Metropolitana de Bilbo	251
b.-Modelos de asociación geográfica	266
Referencias	272
5.- Conclusiones y discusión	273
Referencias	283
Anexo 1. Modelos lineales generalizados (GLM)	312
Anexo 2. Modelos lineales generalizados mixtos (GLMM)	320
Anexo 3. El algoritmo Gibbs sampling	335
Anexo 4. Variables explicativas con un diseño en rejilla. Índice de privación e incidencia y mortalidad	338
-Construcción de un índice de privación material para los municipios de la Región Sanitaria Girona	339
-Influencia de la privación sobre la distribución geográfica de la incidencia del cáncer en la Región Sanitaria Girona	364
-Area deprivation and mortality in the capital cities of Andalusia and Catalonia (Spain)	366
Anexo 5. Instrucciones y sintaxis en R y en WinBUGS	382
Referencias	406

Hypotheses

1. In order to know the impact of cancer amongst the population we cannot only limit our research to cancer mortality because:
 - 1.1.– The most lethal tumours are most highly represented.
 - 1.2.– The information contained in death certificates is not accurate enough for all neoplasias.
- 2.– Most neoplasias can be attributed to environmental factors.
- 3.– Environment–related neoplasias are lymphoid neoplasias, thyroid cancer, bladder cancer and throat, bronchial and lung cancer.
- 4.– Environmental factors, like atmospheric pollution and socioeconomic level are the main risk factors in environment–related neoplasias.
- 5.– The observations of all the variables considered in this work show spatial dependence.
- 6.– The data in this work are organised in lattice designs (dependent variables, mortality and incidence, socioeconomic explanatory variables) and spatially continuous (environmental explanatory variables).
- 7.– The spatial variation of the occurrence distribution of an illness is not totally accounted for by the spatial distribution of the known explanatory factors or by random variation.
- 8.– When spatial data are available, the extra–variability (unobserved variability) may break down into spatially unstructured variability (heterogeneity) and spatially structured variability (spatial dependence).
- 9.– In some environment–related neoplasias, case distribution (incidents or deaths) is more dispersed than the Poisson distribution.
- 10.– When exposure is individual but is interpolated ecologically so that it coincides with the dependent variable design, there is a different inter–area variability to the measurement error, which must be controlled by the models.

Objectives

- 1.– To study the relationship between incidence and cancer mortality and environmental factors, particularly atmospheric pollution, using socioeconomic factors as control.
- 2.– To use those methods of spatial statistics suitable for each type of design.
- 3.– To distinguish different sources of spatial extra–variability between the models.
- 4.– To control the problem of excess zeros inherent in some of the environment–related neoplasias.
- 5.– To correctly introduce explanatory variables (according to the design observed in them) of the geographical variability of the environment–related neoplasias in order to avoid possible biases.

Abstract

1.-Introduction

Cancer is one of the main health problems in developed countries. The name 'cancer' incorporates over 100 illnesses with different aetiology, prognosis and treatment, but with a common element, which is that its growth is not controlled by human biological mechanisms.

1.1.1.- Cancer mortality in Spain, in Catalonia and in the Girona Health Region

▪ Cancer mortality in Spain

Spain has an intermediate situation in European countries. There are higher mortality rates amongst Spanish men than their counterparts in the rest of the European Union for the localisations: oral cavity, stomach, liver, larynx, lung, bladder and Hodgkin's disease. Most of these are closely related to tobacco consumption. Spain occupies the first place for larynx and bladder cancer. In women, the global mortality rate for cancer in 1998 was the lowest in Europe. Spanish women only have rates above the European average for liver cancer, uterine cancer and Hodgkin's disease.

▪ Cancer mortality in Catalonia

Lung cancer was the first cause of cancer mortality in men in past decades. In women, breast cancer was the main cause of cancer-related death, which increased until 1991-1992. Colorectal cancer was the second most important cause of cancer mortality both in men and women. The recent trend shows its mortality has now stabilised, following the increases observed after 1980.

▪ Cancer mortality in the Girona Health Region

The greatest number of deaths is due to lung (15.6%) and colorectal cancer (7.9%) in men, and breast (7.5%) and also colorectal cancer (9.9%) in women (UERCG, 2005).

1.1.2.-Cancer incidence in Spain, in Catalonia and in the Girona Health Region

Although Mortality Statistics are widely used to find out the impact of cancer among the population, because they are available in most countries, they are not free from limitations. One of these limitations is inherent in the nature of the data, i.e. that the most lethal tumours are most highly represented. Another limitation is the lack of accuracy of the information contained in the death certificates. For this reason, to make up for these limitations and to improve knowledge of the presence of cancer in the population, population-based Cancer Records have been created, allowing us to estimate the incidence of cancer in the reference population and to identify time-space patterns in the aforesaid populations.

▪ Cancer incidence in Spain

Spain occupies an intermediate position with regards to cancer incidence, with respect to other European countries. Spanish men have higher incidence rates than the remaining European Union countries for the localisations: oral cavity, larynx, lung, bladder. In women, however, the same as with mortality, the incidence rate is below the mean for Europe (Lopez-Abente et al, 2005).

▪ Cancer incidence in Catalonia

In the period from 1996 to 1997, the relative frequency of cancer had a highly differentiated profile according to sex. The most frequent tumours were lung cancer (around 17% of the total numbers of cancer), followed by prostate (15%) and colorectal cancer (14%) in men. In women the most frequent tumour was breast (28%) followed at a certain distance by colorectal cancer (15%).

▪ Cancer incidence in the Girona Health Region

Briefly, the Girona Health Region has the highest cancer incidence with respect to other Spanish records and in some neoplasias (for example , breast cancer) lies in first place.

1.2.-Cancer and the environment

As a group of illnesses, cancer also has multiple factors that determine it. These can be both external (environmental factors, occupational factors and viral infections) and internal (hormones, immune aspects and inherited mutations). This set of causal factors may act jointly or sequentially in order to initiate or stimulate the development of the illness.

Most of the neoplasias are attributed to environmental factors (some 80% according to Higginson, 1969). According to the literature, the environmental factors most highly related to cancer are: atmospheric pollution, polycyclic hydrocarbons, polycyclic aromatic hydrocarbons (PAH), metals, asbestos, radiation, pesticides and endocrine disrupters.

1.2.1.-Atmospheric pollution

Atmospheric pollution is a mixture of substances: volatile organic components (VOC), nitrogen oxides (NO_x), tropospheric ozone (O₃), sulphur dioxide (SO₂), carbon monoxide (CO), particles (PM) and other pollutants (Campbell et al, 2005). Other sources of atmospheric pollution are anthropogenic and natural. The former contributes the most.

Atmospheric pollution can influence the development of cancer through a number of possible mechanisms. Particles (PM) and the ozone can harm DNA by forming reactive oxygens and cause lung cancer. Exposure to PM can even have a carcinogenic effect because it carries other pollutants (metals, PAH, etc) (Brown et al, 2000; Bostrom et al, 2002). Lung cancer is the only cancer this pollutant is clearly linked to (Harrison et al, 2004).

1.2.2.–Polycyclic Aromatic Hydrocarbons (PAH).

Polycyclic aromatic hydrocarbons (PAH) make up a wide range of components formed by two or more benzene rings. The PAH are generally derived from an incomplete combustion of organic matter like carbon, oil, petrol, diesel and tobacco, which are freed into the atmosphere (Boffetta et al, 1997; Bostrom et al, 2002).

Exposure in humans is generally occupational, in particular, from inhalation, intake and through contact with the skin. Other non-occupational sources of PAH are meals, water, cigarette smoke, smoke from kitchens and atmospheric pollution (Kyrtopoulos et al, 2001; WHO, 2004).

Cases of lung cancer have been found in workers exposed to high concentrations of PAH. Other occupational studies also suggest an increase in the risk of bladder cancer through exposure to PAH.

1.2.3.–Metals

Metals and their components are freed into the atmosphere from different industrial sources (the most important sources are foundries, fossil combustion and coal-burning industries), polluting the atmosphere (Jarup, 2003; CDC, 2004). Other important sources of pollution by metals are the timber, pesticide and fertiliser industries, incinerating plants, houses with lead piping, etc (Frumkin et al, 2001; Denkhaus et al, 2002; Jarup, 2003; CDC 2004). Natural sources of pollution are also surface waters, soils and ground-waters (Jarup, 2003; Satarug et al, 2003). For most of the population, exposure occurs mainly when inhaling particles in the atmosphere, drinking water or eating contaminated food, when there is skin contact, or when inhaling contaminated smoke (CDC, 2004).

The most important carcinogens among metals and their components are arsenic, beryllium, cadmium, chrome and nickel. Epidemiological evidence on the carcinogenicity of metals comes mainly from occupational studies. Cohort studies of workers exposed to arsenic, beryllium, cadmium, chrome or nickel compounds from inhalation showed an increase in the risk of lung cancer (Hayes, 1997; ATSDR, 1999).

1.2.4.–Asbestos

Asbestos is the generic name given to a group of six fibrous minerals, made up of magnesium, silicon and other elements that are found in the soil and rocks (ACS, 2003).

Occupational studies have highlighted the association between asbestos and the increase in the risk of mesothelioma (pleural, peritoneal) and lung cancer (ATSDR, 1999). The use of all varieties of asbestos in Spain has been banned since 14th June, 2002.

1.2.5.–Radiation

The electromagnetic spectrum spreads from radiations with small frequencies, and large wavelengths (non-ionizing radiation) to high frequencies with small wavelengths (ionizing radiation).

The most important sources of non-ionizing radiation are those generated in electric power plants. Magnetic and electrical fields (called EMF electromagnetic radiations) have been associated with cancer in several epidemiological studies (Miller et al, 1996; Villeneuve et al, 2000; Ahlborn et al, 2001; Poole et al, 2006; Rösli and Künsli, 2006). There is, however, certain difficulty in assessing exposure to non-ionizing radiation, as a consequence of the great variety of exposure sources, variety of different electromagnetic fields, uncertainty regarding the exposure period and the difficulties in measuring past exposure (Ahlborn et al, 2001).

Ionizing radiation, on the other hand, is a well-known cause of cancer and other adverse effects. It is one of the most widely studied in human carcinogenesis and may account for some 3% of all types of cancer. Ionizing radiation is able to eliminate electrons of atoms and change the molecular structure of cells. These changes are those that can induce the development of cancer, especially thyroid cancer.

1.2.6.–Pesticides

Pesticides are substances created to eradicate or control insects, weeds, fungi or any organism defined as a plague. They are commonly used in agriculture, horticulture and to control insects that cause illnesses in public health programmes (Dich et al, 1997).

Humans can be exposed to pesticides through inhalation, intake or absorption. Between 1977 and 1992 the effect of 47 pesticides were studied on rats and mice, 23 of which were found to be carcinogenic (Dich et al, 1997). The most common tumours in the bioassays were tumours in the kidney, gastrointestinal system and thyroid.

1.2.7.–Endocrine disrupters

Endocrine disrupters are a widespread root of synthetic chemical compounds and natural plants. An endocrine disrupter is defined as an exogenous substance or a mixture that alters the function or functions of the endocrine system and, therefore, causes adverse effects on health (European Commission, 1996).

In general, exposure to endocrine disrupters and the risk of cancer has not been studied in depth. The relationship between breast cancer and exposure to DDT and PCB is an exception, but even here more research is needed.

1.3.–Environment–related neoplasias

This work focuses on those neoplasias in which evidence in the literature suggests a possible relationship with environmental factors and those which are of certain interest in public health. In particular, we are interested in those in which the incidence and/or mortality have substantially increased in the last few years or those in which a certain systematic behaviour (i.e. non random) was found in the observed cases with regard to the geographical pattern. Thus the following (from now on environment–related neoplasias) will be considered: lymphoid neoplasias (CIE–10: C81–C85; C90–C91; C96), thyroid cancer (CIE–10: C73; CIE–9: 193), bladder cancer (CIE–10: C67, CIE–9: 188) and throat, bronchial and lung cancer (CIE–10: C–34; CIE–9: 162).

1.3.1.–Lymphoid Neoplasias (CIE–10: C81–C85; C90–C91; C96).

Lymphoid neoplasias cover a wide variety of malign neoplasias that share a common origin in the cells of the lymphoid immune system, i.e. in the tissues of the organism that house the lymphocytes or their precursors (ganglionic lymphoid tissue).

Based on their cellular origin, WHO has classically grouped lymphoid neoplasias into three categories (Harris et al, 1999; Jaffe et al, 2001; Marcos–Gragera, 2006): non–Hodgkin’s lymphomas (CIE–10: C82–C85; C96, CIE–9: 201), Hodgkin’s lymphomas (or disease) (CIE–10: C81, CIE–9: 202) and myelomas (CIE–10: C90, CIE–9: 203.0).

Table 1 summarises all the risk factors related to lymphoid neoplasias found in the literature.

Risk factor	Type of lymphoma	References
Socioeconomic level	Non-Hodgkin lymphoma	Scherr y Muller, 1996; Harras et al 1996; Krol et al, 2003
Genetic	Non-Hodgkin lymphoma	Linnet y Pottern, 1992; Zhu et al 1998; Potzsch, 1999; Goldin et al, 2004
Inmunologic deficiency	Non-Hodgkin lymphoma	Mueller N, 1999; Knowles DM, 1999; Leval y Harris, 2003; Chang et al 2004a; Domenech et al, 2004; Engels et al, 2004a; Kogevinas et al, 2004
Virus	Non-Hodgkin lymphoma Lymphoma Hodgkin	Mueller N, 1991 ; Blattner WA, 1999 ; Mueller N, 1999; Liu et al, 2003; Chang et al, 2004a; Chang et al, 2004b; Engels et al, 2004a; Engels et al, 2004b ; Morgensztern et al, 2004 ; Negri et al, 2004; Ribera y Navarro, 2004; Rodríguez et al, 2004; Duberg et al, 2005 ; Fritschi et al, 2005 ; Svec et al, 2005
Vaccine	Non-Hodgkin lymphoma	Mekori, 1985; Henry-Amar M, 1992; Holly et al, 1999; Kirk et al, 2001
Bacterium	Non-Hodgkin lymphoma	Vineis et al, 1999; Morgner et al, 2000 ; Takagi et al, 2000
Drugs	Non-Hodgkin lymphoma	Kaldor et al, 1987; Blomberg et al; 1993 ; Kato et al, 2003
Ionizing-radiation	Non-Hodgkin lymphoma	Burnett et al, 1999
Solar radiation	Non-Hodgkin lymphoma	Egan et al. 2005; Smedby et al, 2005
Diet	Non-Hodgkin lymphoma	Ward et al, 1996; Tavani et al, 1997; Chiu et al, 1999 ; Law et al, 1999; Zhang et al, 1999; Naresh et al, 2004
Estilo de vida	Non-Hodgkin lymphoma	Linnet y Pottern, 1992; Freedman et al, 1998; Zhu et al, 1998; Vose et al, 2002

Table 1.–Risk factors of different lymphoid neoplasm. (Source: own work)

1.3.2.–Thyroid cancer (CIE–10: C73; CIE–9:193)

Thyroid cancer is the most common type of malign endocrine tumours and, together with non–Hodgkin’s lymphomas and lung lymphomas, has increased considerably in the last few years (Blattner, 1999; Davies and Welch, 2006). In the literature, risk factors associated with cancer are diverse (follicular and papillary, specifically) of which the most important are: environmental (ionizing radiation), dietary, genetic and occupational factors.

1.3.3.–Bladder cancer (CIE–10 C67; CIE–9: 188).

Bladder cancer is a frequent tumour in men, with a higher incidence in countries with a higher level of economic and industrial development (López–Abente, 1996; Serra, 2002).

The risk factors related to bladder cancer can be seen in Table 2.

Risk factor	References
Tobacco	IARC, 2002
Occupational and chemical exposure, overcoat textile industry	IARC, 1987; Romundstad et al, 2000; Urrutia et al, 2002; Czene et al, 2003; Rafnsson y Sulem, 2003; Bates et al, 2004; Mikoczy et al, 2004; Zeka et al, 2004
PAH exposure	Silverman et al, 1996; Fortuny et al, 1999; Mannetje et al, 1999; Brennan et al, 2000; Serra et al, 2000
Coffee	Pujolar et al, 1993
Diet	Riboli y Kaaks, 1997
Water disinfection by-products	Villanueva et al, 2001
Infection by biological agents	El-Bolkainy y Chu, 1981

Table 2. – Risk factors of bladder cancer. (Source: own work)

1.3.4.–Throat, bronchial and lung cancer (CIE–10 C–34; CIE–9:162).

Lung cancer is the tumour with the highest mortality in the western world. The most frequent cause for the development of lung cancer is related, in 90% of subjects, to the direct consumption of tobacco (active smoker) or exposure to tobacco smoke (passive smoker). Other factors related to the development of lung cancer (see Table 3), such as environmental and occupational exposure to toxic compounds.

Risk factor	References
Environmental exposure	Compuestos orgánicos persistentes (PAH, pesticidas, compuestos orgánicos volátiles, etc) y PM2.5 IARC 1987; Bostrom et al, 2002 ; Chellini et al, 2002; Blais et al, 2003; Boffetta y Nyberg, 2003 ; Sorensen et al, 2003 ; Alavanja et al, 2004 ; Amstrong et al, 2004; Forastiere F, 2004; Harrison et al, 2004
Occupational exposure	Metal : cadmio, plutonio, compuesto de poliuretano, amianto IARC , 1987; Sorahan et al, 2000; Blais et al, 2003; Menvielle et al, 2003; Zeka et al, 2003; Mikoczy et al, 2004; Moshammer y Neuberger, 2004
Radiation	Radiación ocupacional Darbi et al, 2005 ; Richardson y Ashmore, 2005 ; Wing y Richardson, 2005

Table 3.– Risk factors of lung cancer, besides tobacco consumption. (Source: own work)

2.– Statistical models applied to spatial data

2.1.– Spatial statistics

What distinguishes spatial statistics from other branches of statistics is the use of geographical locations in the statistical spatial model. In spatial statistics, observations are not considered independent; quite the opposite, it is implicitly supposed that they are correlated with each other, i.e. there is a spatial dependence.

2.2.– Types of spatial data

When we talk of spatial data, it refers to the measurements or observations carried out at specific localisations or in specific areas. Depending on this classification, a specific analysis methodology must be used: lattice data, geostatistical data and point processes.

2.2.1.– Lattice data

Lattice data are observations from a random process, observed on a countable collection of spatial regions, which can be regularly or irregularly distributed, complemented with what is called neighbour structure, i.e. information on neighbouring regions.

2.2.2.– Geostatistical data

Geostatistical data, also called random field data, are measurements taken at fixed points with continuous localisations in space. The aim is to extend the spatial distribution of the values of an attribute over the whole region under study. Therefore, the analysis of geostatistical data may take into account the modelling of the variability pattern, the determination of the factors with which it may be related, and obtaining a good prediction of the variable at points where it has not been sampled (Bailey and Gatrell, 1995; Kaluzni et al, 1998).

2.2.3.– Point processes

Point processes are so called when the localisations (and not the measurements) are the variables of interest. The aim is to know the intensity variation of occurrences over the region under study and to look for models that help to account for or understand the phenomenon.

2.3.– Spatial epidemiology

The term spatial or geographical epidemiology is used to describe studies on the causes and prevention of illnesses using different analysis perspectives in which the localisation of the occurrences is a fundamental component (Thomas, 1990). One of the main aims is to show which part of the spatial variation of an illness's occurrence distribution is not explained by

either the spatial distribution of known explanatory factors or by random variation. In fact, it is often of interest to find clues to some unknown risk factor of the illness (Saez and Saurina, 2006).

Following Elliot and Watenburg (2204), spatial epidemiology in small areas can be approached from three large points of view: disease maps, geographical association studies and agglomerations of cases or '*clustering*'.

2.4.- Variables response: incidence and mortality from environment-related neoplasias. Disease maps

Disease maps provide a quick visual summary of complex geographical information and allow patterns to be identified in the data which in another way may remain unnoticed in tabular presentations (Elliot and Watenburg, 2004). One of the main aims is to show what part of the spatial variation of an illness's occurrence distribution is not accounted for either by spatial distribution of known explanatory factors or by random variation. In fact, it is often of interest to find clues to some unknown risk factor of the illness (Saez and Saurina, 2006).

2.4.1.-Raw Cases and Standard Ratios

When raw cases are represented, it is generally preferred to represent the ratios that allow the effect on the population at risk to be incorporated. The direct use of the raw ratios does not allow the different areas to be compared, because the differences observed between them can be due to factors that have not been borne in mind. Thus, the ratio must be standardised by those known factors that may have influenced them. The most widely used standardisation in epidemiology is 'age', i.e. to consider the different distribution of the population by age groups in the areas where ratios need to be considered.

- Calculation of the standard incidence/mortality ratio (SIR/SMR)

In the case of the incidence or mortality of an illness, the state of a determined area is usually assessed by firstly determining what incidence mortality would be expected in that area and, then, by comparing it with the observed incidence. The observed/expected ratio, the so-called standard incidence or mortality ratio (SIR/SMR), is no more than a risk estimator regarding the area, that is, of the risk of illness in relation to the considered reference group.

Even though the SIR/SMR are widely used to represent incidence and mortality maps, they have some limitations (Clayton and Kaldor, 1987; Lawson et al, 2003). Firstly, they do not bear in mind how population varies between different geographical areas. In this sense, the extreme SIR/SMR on the map and, therefore, dominant in the apparent geographical pattern that can appear there, are the ones estimated less accurately, simply because they have been calculated using very few cases, either in the numerator (observed cases) or in the

denominator (expected cases). In general, moreover, the variation of the observed cases is usually considerably higher than expected, causing what is called extra-variability.

Part of the problems of the SIR/SMR can be solved by estimating them by means of a generalised linear model (GLM).

Specifically, the observed cases are usually supposed to follow a Poisson-type counter process, with a measurement equal to the product of the relative risk for the expected cases. The SIR/SMR are no more than the estimator of relative risk.

The GLM model supposes that there is no excess risk, i.e. that the observed number of cases will coincide with the expected number (Lawson et al, 2003). In general, however, the variation in the observed cases is not the same as that expected in a Poisson distribution, causing a specification error that shows itself in heteroscedastic residues, also called extra-variability.

When spatial data are available, it is important to distinguish two sources of extra-variation. Firstly, the existence of independent and spatially uncorrelated extra-variation, called 'uncorrelated' or 'non-spatial heterogeneity', must be assumed. Secondly, the most important source of extra-variation is what is called 'correlated' or 'spatial heterogeneity' and is a consequence of the correlation of the spatial unit with neighbouring spatial units. What is most usual is that this dependence is largely due to the existence of uncontrolled variables, i.e. those not included in the analysis.

2.4.2.–Smoothing of Standard Ratios

2.4.2.1.– Smoothing using Generalised Linear Mixed Methods (GLMM)

To solve the problems derived from the direct use of SIR/SMR, mixed models have been used. The idea is to introduce in the model one (or more) random effect which includes the unexplained variability. In this case, a generalised linear mixed model (GLMM) must be specified for the relative risk. The random effect includes both spatial and non-spatial heterogeneity (extra-variation).

2.4.2.1.– Smoothing using Generalised Linear Mixed Models (GLMM) with spatial effects

– Besag, York and Mollié model (BYM)

To smooth the ratios, two random effects are usually used (Besag, York and Mollie, 1991). The model is called '*convolution prior*' (Besag, York y Mollie, 1991; Mollie, 1996). One of the random effects includes spatial heterogeneity, and the other, destructured (non-spatial) extra-variability.

For the spatial component, the most usual approach (and the simplest computationally) is the one known as the conditional autoregressive, CAR. As a general idea, a spatial dependence, for which the risk estimations in a specific area depend on the neighbouring areas, is assumed. Thus the weightings (or spatial weightings) are maximum (equal to the unit) if the areas are adjacent and minimum (equal to zero) in another case.

Often, the distribution of cases is usually more dispersed than the Poisson distribution. That is to say, certain areas do not show any case of any environment-related neoplasia, which means this neoplasia is not distributed exactly like a Poisson. In fact, a differentiation could be made between 'structural zeros', which are unavoidable, and 'sample zeros', which occur at random. In environmental epidemiology, 'zeros' are usually from sample zeros, which means there is no area free from risk. Sometimes, it is also possible for the number of 'zeros' observed to be lower than expected, but it is usually a much less frequent situation in practice.

In any case, we must use alternative models to the Besag, York and Mollié model, based, as is known, on the standard Poisson distribution. The two most important approaches are: finite mixture models and relative risk mixed models.

- Finite mixture models. ZIP models and 'Hurdle' models

Finite mixture models suppose that the population is compartmentalised into k homogeneous components. Each component has a different risk level. Moreover, this risk can depend on different covariates. Usually, there are two homogeneous components, one for zeros (non observed cases) and the other for those that are not zero (positive values from now on), thus leading to a mixed Poisson distribution.

When zeros are structural, the mixed distribution has an extreme shape, leading to models that are much more restrictive than general finite mixture models.

ZIP models

In *Zero-inflated Poisson* (ZIP) models (Mullahy, 1986; Heilbron 1989 and 1994; Lambert, 1992) the variable response is modelled as a mixture of the Bernoulli and Poisson distributions and other distributions of alternative discrete random variables (generalised Poisson distribution and negative binomial).

Hurdle models

These models specify two processes that generate different data, one for zeros and the other for positives. In econometric literature these kinds of models are called *Hurdle models*, (Mullahy, 1986) or also two-part models (Heilbron, 1994). The distribution of positive values is usually a truncated shape (in zero) of a standard discrete distribution, as the Poisson (Cameron and Trivedi, 1998) or the negative binomial distribution (Deb and Trivedi, 2002) could be.

- Mixture relative risk model

The models seen so far are based on smoothing techniques. The problem is that all of them usually smooth in excess, in the sense that they do not maintain those discontinuities of the risk surface, like zeros, which should perhaps be borne in mind (Lawson et al, 2003). In this sense, Lawson and Clarke (2002) proposed a mixture model that simultaneously allows smoothing and representation of discontinuities. In particular, the model assumes that the relative risk (logarithmic) can be decomposed into three components: one fixed component that represents the unstructured heterogeneity, which measures the overdispersion in a particular area; and two random components which represent aspects of the spatial correlation, the spatial random effect and a random effect that allows the discontinuities or the 'leaps'.

2.5.- Explanatory variables

2.5.1.- Lattice data

As an example of explanatory variables with a lattice data design, Annex 4 includes three articles of Lertxundi-Manterola et al (2005 and 2006). A privation index is built in all of them, with a different value for each municipality (the first two) and an electoral district (the last one), which is related to incidence and mortality.

2.5.2.- Geostatistical data analysis

The explanatory variables of interest in this work are environmental, above all those related to atmospheric pollution. The data corresponding to these variables have a geostatistical design.

2.5.2.1.- Spatial dependence analysis

The tools for analysing spatial dependence in geostatistical data are the functions of covariance (covariogram) and semivariogram.

2.5.2.3.- Geostatistical data estimation

- Deterministic interpolation methods

The aim of spatial interpolation methods is to estimate geostatistical data at points on the spatial domain that do not coincide with those observed in the same. There are several techniques which differentiate methods and/or complexity in their hypotheses. The one used in this thesis was *Inverse Distance Weighting (IDW)*.

– Stochastic weighting methods. Kriging

Kriging has two main advantages over deterministic estimators: a) the weightings used in the estimation are determined as a function between the structural distance (the semivariogram) of the observed points and the prediction that is desired; and, most importantly, b) the estimation has a quantification of uncertainty associated to it, the typical kriging deviation (or variance) (Matheron and Kleingeld, 1987).

The information on the environmental exposure used in this thesis was obtained from relatively few spatially dispersed stations. For this reason we used interpolation, which allowed us to assign continuous values of the pollutant to the discrete values of the neoplasia. When the number of stations allowed it, the exposure surface of the pollutants was interpolated using a stochastic method, ordinary kriging. There were two reasons for this choice: firstly, the spatial pollutant variable was isotropic and did not show any spatial trend; secondly, the mean value of this spatial variable was unknown.

2.6.– Geographical association studies

2.6.1.– Generalised Linear Mixed Models with explanatory variables with lattice data

The introduction of an explanatory variable with lattice design in a GLMM model was considered a single fixed effect, i.e. there was no variability within each municipality/electoral district.

2.6.2.– Generalised Linear Mixed Models with explanatory variables with geostatistical design

Still modelling stochastically the exposure surface, the mean value of the pollutant in the corresponding area could be introduced directly. The problem is that it supposes that the possible association observed at the area level (ecological level) is the same for all the individuals resident in an area (individual level) (Wakefield and Shaddick, 2006). This supposition could lead to the problem of ecological fallacy (Selvin, 1958) as a ‘pure specification bias’, which appears when a non linear model is added (Wakefield and Shaddick, 2006). For this reason, intra–area variability was introduced to avoid this kind of bias. But furthermore, this work considered the coefficient associated to the pollutant to be a random effect. This is because we supposed that the estimation error was not exclusively a measurement error, but that, although measured ecologically, the exposure always occurred at individual level (Lawson et al, 2003).

The model finally used was based on that of Jackson et al (2006). In fact, there are other alternatives, such as the complete Bayesian approximation, in which the parameters for health and exposure are estimated simultaneously, or the three–stage model proposed by Wakefield and Shaddick (2006). However, given the large amount of data available, the first is computationally unapproachable, and the Wakefield–Shaddick model has not provided satisfactory results in terms of fit.

3.– Methodology. Descriptive analysis

3.1.– Population

We studied the incidence of cancer in the population of the Girona Health Region (cf. Figure 1, p.131) and mortality in the populations residing in the Metropolitan Areas of Barcelona and Bilbao (cf. Figures 2 and 3, p.135–136).

3.2.– Study subjects

Study subjects were residents of the Girona Health Region diagnosed with one of the environment-related neoplasias between 1994 and 2001; residents in the Metropolitan Area of Barcelona that had died between 1994 and 2003; and residents in the Metropolitan Area of Bilbao that had died between 1996 and 2003.

3.3.–Design

Observational ecological study with lattice data. Cancer incidence in the municipalities of the Girona Health Region and cancer mortality in the electoral districts of the Metropolitan Areas.

3.4.– Dependent variables

Dependent variables were the observed cases of environment-related neoplasias: throat, bronchial and lung cancer (CIE-10: C-34; CIE-9: 162), bladder cancer (CIE-10: C67, CIE-9: 188), thyroid cancer (CIE-10: C73; CIE-9: 193), non-Hodgkin's lymphomas (CIE-10: C82–C85; C96, CIE-9: 201), Hodgkin's lymphomas (or disease) (CIE-10: C81, CIE-9: 202) and myelomas (CIE-10: C90, CIE-9: 203.0).

The reference population for the expected cases for the analysis of the incidence was taken from all the units under study, since it was not possible to know the population that was really exposed. For mortality, the whole Spanish population, according to the Population and Housing Census of 2001, was used as the reference population

3.5.– Explanatory variables

– Socioeconomic variables

The socioeconomic variables considered were unemployment (indicator of social integration), manual workers (level of material well being and social prestige) and insufficient education (indicator associated indirectly with future income and with the population's standard of living).

In order to build a synthetic indicator, these indicators were combined using the Principal Components Analysis (PCA) (Anderson, 1984). As a result of PCA we obtained indexes highly correlated to the percentage of manual workers, unemployment and insufficient education in each of the three areas under study. In the case of the Girona Health Region two factors were withheld that accumulated approximately 80% of the total variance in both sexes (some 45% the first and about 35% the second). The first factor was highly correlated to the unemployment and manual workers variable, and the second with insufficient education. In the Metropolitan Areas of Barcelona and Bilbao, a single factor was withheld in each of them, accumulating some 70% of the total variance in both sexes in Barcelona and 65% in Bilbao.

– Environmental variables

Levels atmospheric pollution of nitrogen dioxide (NO₂) and sulphur dioxide (SO₂) were gathered for the periods 1996–1998, 1999–2001 and 2002–2003. The pollutant recorded in the Girona Health Region (1994–2001) was ozone (O₃).

A station was considered valid if it had lost less than 50% of the data. In this way, for the Metropolitan Area of Barcelona, up to 18 atmospheric pollution measuring stations were chosen, 5 for the Girona Health Region and 62 in the Metropolitan Area of Bilbao.

For each pollutant, area and period, firstly, the lost values in each valid station in one specific year were replaced using multiple imputation (Little and Rubin, 1987; Rubin, 1987; 1996). Specifically, using the EM algorithm (Schafer, 1996) the lost data were estimated by maximum likelihood (Novo and Schafer, 2003). Once the lost data were imputed, the pollutant levels were predicted in the corresponding analysis unit. Previously that long-term time behaviour of the series of pollutants that could distort the spatial dependence was extracted. Then, after extracting the tendency of the corresponding pollutant, and to be able to determine the (spatial) distribution of this pollutant, the spatial correlation between the different stations was estimated. In the case of the two Metropolitan Areas (Barcelona and Bilbao), both for estimating spatial dependence and, in particular, for predicting pollution levels for electoral district, the *kriging* stochastic method was used. Since there were only five measuring stations in the Girona Health Region, a deterministic method, the *IDW* method was used. After estimating the spatial correlation (and checking the goodness of fit), the levels for each pollutant were predicted using ordinary kriging at points other than those observed, i.e. in different districts or municipalities of the studied areas. For each of the analysis units (electoral districts and municipalities) a single value was obtained for the mean of the pollutant, as was its typical deviation (for each area and period) (cf. Figures 13 to 21, p.158–166)

4.– Methodology. Modelling.

4.1.– Smoothing of Standard Ratios

All the SIR/SMR were smoothed using BYM and mixture relative risk model. The SIR of throat, bronchial and lung cancer in men (17.14% zeros) and women (67.14% zeros) were also smoothed using ZINB and Hurdle models. Using the results obtained for SIR, the models were compared by means of different criteria of goodness of fit: Akaike Information Criterion (AIC), BIC and DIC:

The parameters estimated were relative risks, a-posteriori probabilities (PRP) and the typical deviations of the random effects. The PRP were estimated in reality as the probability that the relative risk was greater than one, denoting areas of high risk and, by symmetry, the protection areas were also estimated.

4.1.3.– Estimation results

a.–Incidence of environment-related neoplasias in the Girona Health Region

Table 1 (p.195) shows the results for SIR smoothing of throat, bronchial and lung cancer. The best results (in terms of AIC and BIC) were obtained for the relative risk mixed model, except in the case of men, for which the best fit was obtained with the BYM model. The ZINB and Hurdle models did not provide very satisfactory results in all cases, at least with regard to goodness of fit. So, in the case of few zeros (men), the BYM model was superior, while in the case of excess zeros (women) the best model was the mixture relative risk model. For these reasons, in what follows, the remaining SIR and all the SMR were smoothed using the BYM model (the best, in the case of few zeros) and the mixture relative risk model (in the case of excess zeros). In all the cases the model that provided the best fit (even in terms of DIC) was the mixture relative risk model. In fact, there seems to be a relationship between the percentage of zeros and the 'goodness' of the model. Look at Figure 1 (p.198) to see how, up to 40–50% of zeros, both models provide quite a similar fit. Even with a lower percentage of 30% the BYM model is much better. However, for 40–50% the best model seems to be the mixed one.

SIR were smoothed in all cases (Figures 2 to 13, p.199–210). See the histograms in the lower right-hand corner and, to a lesser extent, the more uniform range of colours in the SIR maps to the left (Figures 26 to 31, p.170–175). The geographical structure hinted at in the preceding chapter (Figure 26 to 31, p.170–175) seems to be confirmed in some cases. Observe the municipalities with excess risk, in red, in the a-posteriori probability maps, in particular, throat, bronchial and lung cancer (both sexes); bladder cancer (men); and thyroid cancer (women). Less clear is the geographical structure in the SIR of non-Hodgkin's lymphoma (both sexes) (a little less clear in men), and in myeloma (both sexes). The remaining neoplasias: bladder (women), thyroid (men), Hodgkin's lymphoma (both sexes) show some isolated municipalities with excess risk, without any systematic behaviour being evident in any case.

b- Geographical association models

b.1.- Explanatory variable: socioeconomic variable. Estimation results

The parameters associated with the privation index (brought together in two factors in the case of Girona Health Region) were statistically significant in approximately half the environment-related neoplasias, specifically in throat, bronchial and lung cancer (both sexes); bladder in women; thyroid cancer in women; and non-Hodgkin's lymphoma in men; Hodgkin's lymphoma in both sexes.

The parameters associated with both factors were jointly significant in throat, bronchial and lung cancer in men (RR=1.2337¹ and RR=1.1815, respectively), and in bladder cancer in women (RR=1.1860¹ and RR=1.3061) (cf. Table 3, p.214, and Figure 14, p.213). The first factor was significant in throat, bronchial and lung cancer in women (RR=1.3297¹), thyroid cancer (women) (RR=1.2726¹), non-Hodgkin's lymphoma (RR=1.1393¹) and myeloma (RR=1.2299¹) (both men). The second factor (only) was significant in Hodgkin's lymphoma in both sexes (RR=1.6491, men and RR=1.5671, women). In any case (cf. Table 3, p.214, and Figure 1, p.213) it can be deduced from the results that greater privation implied a greater incidence of certain environment-related neoplasias, in particular, throat, bronchial and lung cancer (both sexes); Hodgkin's lymphoma (both sexes); bladder cancer (women); thyroid cancer (women); non-Hodgkin's lymphoma (men) and myeloma (men).

b.2.- Explanatory variable: atmospheric pollutants. Estimation results

The pollutant parameter only bore some statistical significance in the case of bladder cancer in men (RR=1.2681) (Table 4, p.222, and Figure 21, p.223).

It must be pointed out that variability and geographical distribution of the municipalities with excess risk were relatively unmodified when the privation index and/or ozone were introduced as explanatory variables (Figures 15 to 20, p. 215–220; Figure 22, p. 224).

4.1.4.- Estimation results. Metropolitan Area of Barcelona

a.-Mortality from environment-related neoplasias in the Metropolitan Area of Barcelona.

The smoothed SMR (as well as the PRP and frequency histograms) are shown in Figures 24 to 35, p.227–238. These seem to show a clear geographical behaviour in throat, bronchial and lung cancer (both sexes), and bladder cancer (both sexes). In the first case, there could be excess risk in the district of Ciutat Vella, Barcelona; in the second, significant high risks were found in El Prat and Sant Just–Sant Feliu. Non-Hodgkin's lymphoma shows districts with excess risk but there was no clear geographical pattern. There was no clear geographical pattern for the remaining neoplasias.

¹ The sign of the first factor was changed to maintain interpretability.

b- Geographical association models

The privation index was statistically significant in throat, bronchial and lung cancer (both sexes) (RR=1.1965, men and RR=1.2276, women); non-Hodgkin's Lymphoma (both sexes) (RR=1.1829, men and RR=1.2589, women); and myeloma (both sexes) (RR=1.5627, men and RR=1.4002, women) (Table 6, p.241, and Figure 36, p.242). Observe that, except in myeloma, the effect of the index on the SMR was greater in the case of women, although sex differences were never statistically significant. The effect of NO₂ was statistically significant in throat, bronchial and lung cancer (women) (RR=2.1711); bladder cancer (women) (RR=1.9059); and non-Hodgkin's lymphoma (men) (RR=1.7651) (Table 6, p.241, and Figure 37, p.242). The effect of SO₂ was never statistically significant and also showed quite large trust intervals (in fact, Figure 38, p.242, has been shown in logarithmic scale so that it can be interpreted). Only in throat, bronchial and lung cancer, in women, and non-Hodgkin's lymphoma, in men, were both the index and NO₂ statistically significant.

It can be seen, as before, that even when introduced explanatory variables were statistically significant, there was little reduction in explained extra-variability (Fig.39-45, p.243-249).

4.1.5.- Estimation results. Metropolitan Area of Bilbao

a.-Mortality from environment-related neoplasias in the Metropolitan Areas of Bilbao

The smoothed SMR are shown in Figures 47 to 58 (p.254-265). Results were repeated, in the sense that there was a clear geographical pattern in throat, bronchial and lung cancer (both sexes) (Figures 47 and 48, p.254-255), and bladder cancer (men) (Figure 49, p.256). The pattern of excess risk differentiated by sex should be highlighted in the case of throat, bronchial and lung cancer. Note that, except Bilbao which has an excess risk in both cases, in men, the districts with the greatest risk were concentrated on the left-hand side of the river; in women, on the other hand, they were on the right-hand side.

b- Geographical association models

The coefficient associated with the privation index was statistically significant in the case of throat, bronchial and lung cancer (both sexes) (RR=1.1294, men and RR=1.2810, women), and in bladder cancer (men) (RR=1.1322) (cf. Table 8, p.267, and Figure 59, p.268). In throat, bronchial and lung cancer, the effect of the index was greater in women than in men. The coefficients associated with the pollutants were statistically significant only in throat, bronchial and lung cancer (RR=1.6693, men and RR=1.1.8224, women) (cf. Table 8, p.267, and Figures 60 and 61, p.268). Note that, as happened with the privation index, the effect was greater in women. The effect of SO₂, however, was only statistically significant in men (RR=1.7161). As in the case of the Metropolitan Area of Barcelona, the standard deviations of the coefficient associated SO₂ were very important.

Introducing the explanatory variables again meant there was little reduction in variability (Figures 62 to 64, p.269-271).

5.–Conclusions and discussion

–Both the incidence and mortality of environment–related neoplasias showed two sources of extra–variation: spatial extra–variation, in which neighbouring units tend to show similar incidence/mortality ratios, and non–spatial heterogeneity.

–In general, spatial extra–variability was much higher (practically twice as much) than non–spatial.

As we have said, although the raw SIR/SMR fit for the population at risk to suffer from environment–related neoplasias, they show a large variability that makes them unstable. We commented that the most important source of extra–variability is spatial. In fact, these ratios have been smoothed using the best model in each case. This was the mixture relative risk model, except in the case of throat, bronchial and lung cancer in men (in which the SIR/SMR were smoothed using a BYM model). It must be pointed out that this neoplasia showed relatively few zeros: 14.6% zeros in the case of incidence and 4.5% Bilbao, and 11.93% Barcelona, in the case of mortality.

–To smooth the SIR/SMR corresponding to variables with a percentage of zeros higher than 40–50%, a model that captures this behaviour must be used.

–The best model, in terms of fit, which recognises excess zeros in the variables of interest was the mixture relative risk model.

Zero inflation, i.e. the existence of many more municipalities/electoral districts in which there were no cases of the environment–related neoplasias which would correspond to a counter process, has implied that the variables probability distribution does not correspond exactly to a Poisson. As said, there are several alternatives: finite mixture models and the mixed relative risk model. With regard to the former, the worst fit was provided by the Hurdle model. In fact, it does not seem reasonable that the process generating zeros should be totally independent of the process generating positive cases, as this kind of model supposes (Ridout et al, 1998, Dalrymple et al, 2003). The best fit (whenever an ‘excess’ of zeros is caused) was given by the mixture relative risk model. The results confirm that this type of model did not smooth in excess, in the sense that it also maintained those discontinuities presented in the data (Lawson et al, 2003).

One of the lines of research that we are involved in at the moment is to find out what the threshold that allows us to define a probability distribution with excess zeros is. We dare to suggest a value of around 40–50%.

One aspect that must be borne in mind is the ‘degree of smoothing’ implied in this kind of model (Elliott and Wartenburg, 2004). This degree establishes a compromise between high sensitivity, areas of correctly identified excess risk, and high specificity, areas that do not have correctly identified excess risk (Richardson et al, 2004). A sensitive but not specific result will generate false positives, while a specific but not sensitive one will fail to identify

those areas with high risk. Richardson et al (2004), after an exhaustive simulation study, showed that at least 5 expected cases are needed for the smoothing method used, with a relative risk of 2 to 3, to provide results that are sensitive enough, in that they detect areas with high risk. As can be seen, the risk is extremely high. For this reason, instead of relative risk, Richardson et al (2004) propose using the a-posteriori probability (PRP) that this risk is bigger than the unit. This was why PRP maps as well as smoothed SIR/SMR were introduced in this work. A PRP greater than 80% (or less than 20% of reduced risk) implies that the smoothing method used was specific enough.

-Smoothed SIR/SMR only showed a clear geographical pattern in some of the environment-related neoplasias.

In both sexes, throat, bronchial and lung cancer are what display a fairly obvious geographical behaviour in the three areas under study. Secondly, it is worth mentioning bladder cancer in men, which shows spatially grouped areas of excess risk in the three cases. Bladder cancer in women might show a certain geographical structure only in the Metropolitan Area of Barcelona. In the same way, the incidence of thyroid cancer in women could also show a certain geographical structure.

As is known, disease maps have been used to generate etiological hypotheses. Although it is not our aim in this thesis, we must highlight the geographical behaviour of, in the first place, throat, bronchial and lung cancer, with relatively higher risks in the more economically deprived areas. Thus, in the Metropolitan Area of Barcelona and also in the Metropolitan Area of Bilbao, excess risk in men is concentrated on the left-hand side of the river, while in women it is on the right-hand side. In the second place, with regard to thyroid cancer in women, there is a geographical concentration in the Baix Empordà region.

This last group will be studied exhaustively in the near future.

-Part of the remaining variability in the smoothed SIR/SMR could be explained by introducing explanatory variables, in particular atmospheric pollution and socioeconomic variables.

-As the atmospheric pollutants were observed in a geostatistical design and the environment-related neoplasias were studied in a lattice design, the exposure surface was modelled.

The information on the environmental exposure was obtained for relatively few spatially dispersed stations. For this reason, we used methods of interpolation that allowed us to assign continuous values of the pollutant to discrete values of the neoplasia. In fact, there are two differentiated situations. One, the Health Region of Girona has very few, highly dispersed station in the region; and the other, the case of the Metropolitan Areas, has quite a lot more stations that covered the area studied. In particular, the 62 territory-wide stations must be highlighted in the case of the Metropolitan Area of Bilbao. Only in the first case was a deterministic interpolation method used. The exposure surface of the pollutants in the

Metropolitan Areas was interpolated using a stochastic method, ordinary kriging. There were two reasons for this choice: firstly, the polluting spatial variable was isotropic and did not display any spatial tendency; secondly, the mean value of this spatial variable was unknown.

Stochastic interpolation (also used by most of the authors, cf. Lawson et al, 2003) allows us to estimate uncertainty, i.e. intra-area variability, in addition to the mean.

-The effect of the pollutant in each municipality/electoral district was approached in the model by introducing the average value in each area and the intra-area variability.

-The effect of the pollutant was considered random, in the sense that it could be different in each area.

Various authors, still modelling the exposure surface stochastically, only introduce the mean value of the pollutant in the corresponding area (Le et al, 1996; Gelfand et al, 2001; Tonellato, 2001; Shaddick and Wakefield, 2002). The problem of all these works is that they suppose that the possible association observed at the area level (ecological level) is the same for all the individuals resident in an area (individual level) (Wakefield and Shaddick, 2006). This supposition could lead to the problem of ecological fallacy (Selvin, 1958) as a 'pure specification bias', which appears when a non linear model is added (Wakefield and Shaddick, 2006)

Deterministic interpolation was used in the case of incidence of environment-related neoplasias, i.e. in the Girona Health Region, but this logically does not provide an estimation of uncertainty. However, making the most of the fact that not only data on space but also on time were gathered, a method of resampling was designed to obtain an estimation of uncertainty.

But besides introducing intra-area variability to avoid this kind of bias, the coefficient associated to the pollutant was considered a random effect in this work. This is because we supposed that the estimation error was not exclusively a measurement error, but that, although measured ecologically, the exposure always occurred at individual level (Lawson et al, 2003).

The model finally used was based on that of Jackson et al (2006). In fact, there are other alternatives, such as the complete Bayesian approximation, in which the parameters for health and exposure are estimated simultaneously, or the three-stage model proposed by Wakefield and Shaddick (2006). However, given the large amount of data available, the first is computationally unapproachable, and the Wakefield-Shaddick model has not provided satisfactory results in terms of fit.

- Socioeconomic conditions were another of the variables that reduced the remaining variability in the smoothed SIR/SMR.

-The socioeconomic variables were summarised from a privation index which, using PCA, showed approximately 70% of the variance of these variables.

Once again, the PCA results for the Metropolitan Areas differ from those obtained for the Girona Health Region. While in the Metropolitan Areas a single factor was enough to approach the privation index, two areas were needed in the Girona Region. There may be two reasons for these results. Firstly, the Girona Health Region is logically a much more dispersed and heterogeneous territory demographically and socioeconomically. Metropolitan areas are usually more homogeneous in this respect. We dare to suggest, however, that the most important reason is related to the nature of the variables included in this index. Of the two factors that approached this index in the Girona Health Region, the first, highly correlated to unemployment (negatively) and with the variable manual workers, may not include the whole concept of material privation. We suggest that this factor, and therefore the resulting index in this case, reflected the concept of 'non urban' population more highly. Thus, using data from the Population and Housing Census of 2001, the population occupied in agriculture, stockbreeding and fishing in the countries of Girona was 5.186% , double the percentage corresponding to Vizcaya (2.625%), practically double that of the rest of Catalonia (3.216%) and over ten times greater than the Metropolitan Area of Barcelona (0.451%).

-Explanatory variables observed with a lattice design, such as the privation index, were introduced in the model as fixed effects.

The privation index is not only observed ecologically but is also built likewise, that is to say, averaging the possible individual values in each area.

-Greater privation, using atmospheric pollutants as control, implied a greater risk of suffering / dying from some of the environment-related neoplasias in the three areas and periods studied.

The privation index (or at least one of the two factors in the case of the Girona Health Region) was always associated with a higher risk of suffering / dying from throat, bronchial and lung cancer. Less solid was the relationship between bladder cancer and the privation index, which was found to be significant in both sexes in the Metropolitan Area of Bilbao and only in women in the Girona Health Region. For the remaining neoplasias, the relationship was statistically significant in some of the areas studied. Hence, thyroid cancer (women), and Hodgkin's lymphoma (both sexes) in Girona, and non-Hodgkin's lymphoma and myeloma (both sexes) in the Metropolitan Area of Barcelona were the neoplasias in which a significant association was found.

There are few studies that relate cancer incidence and socioeconomic conditions (cf. a review in Lertxundi and Saez, 2006, and in Ocaña-Riola et al, 2006) and there are even fewer

studies on the relationship between cancer and socioeconomic conditions using environmental variables as control.

Following an exhaustive review in Medline carried out in August 2006, with regard to the incidence and/or mortality from lung cancer (in several of these works more than one cause is given), we will point out chronologically the works of McWhorter et al (1989), in which they find statistically significant relationships with the racial difference between blacks and whites in the United States; Polednak (1992), with the Puerto Rican immigration in the United States; Gorey and Vena (1995), with poverty in the Metropolitan Areas of the north of the State of New York; Burnley (1997), with variables of the socioeconomic status and manual workers in New South Wales; Lamont et al (1997), with an extreme poverty indicator (the homeless) in the city of Glasgow and in the west of Scotland; Pollock and Vickers (1997), with a privation index in the South Thames Region; Kreiger et al (1999), with an indicator of ethnicity in the United States; Dreassi et al (2003, 2005), with an ad-hoc privation index in Tuscany; Ocaña-Riola et al (2004) with an ad-hoc privation index in Andalusia; Menvielle et al (2005), with occupation and education in France; and Ekberg-Aronsson et al (2006), with the socioeconomic class in the city of Malmo, Sweden.

For bladder cancer, we will quote Ocaña-Riola et al (2004) who finds a significant statistical relationship with the ad-hoc privation index in Andalusia; and Underwood et al (2006), with race in the United States.

With regard to myeloma, we must go back to the work of McWhorter et al (1989), in which they find statistically significant relationships with the racial differences between blacks and whites in the United States; Baris et al (2000), with income, education and social class according to occupation in the United States; and Gebregziabher et al (2006), with social class in the county of Los Angeles.

In Hodgkin's lymphoma, Flavell et al (1999) find significant relationships with the privation index of Townsend in the West Midlands; Glasser et al (2001), with social class in the Metropolitan Area of San Francisco; and Clarke et al (2005), with social class in California.

Finally, McNally et al (2003) find a statistically significant relationship between incidence and mortality of lymphomas and Townsend's privation index in the Northeast of England.

-The effect of privation on the incidence and/or mortality from throat, bronchial and lung cancer, using atmospheric pollutants as control, was greater in women than in men.

The only articles in which a different relationship has been found according to sex were those of Polednak (1992) and Menvielle et al (2005), although in these cases the greatest risk was found for men. In these reports there are at least two differences in relation to our work. Firstly, neither of them is controlled by environmental variables, and secondly, common variables for both sexes were used in both cases.

Although we venture to say that not only the effect but the nature of the index could be different according to sex, it would be necessary to do further research for us to determine the different effects found in this thesis.

-High concentrations of atmospheric pollutants increase the risk of suffering from environment-related neoplasias, using socioeconomic conditions as control.

Nitrogen dioxide was the only pollutant associated consistently and statistically significantly with mortality from throat, bronchial and lung cancer in the Metropolitan Areas of Barcelona and Bilbao. In effect, it was once again greater in women than in men. The association of NO₂ and the remaining neoplasias (bladder and non-Hodgkin's lymphoma in the Metropolitan Area of Barcelona) and the remaining pollutants and neoplasias in other areas (SO₂ and lung cancer in the Metropolitan Area of Bilbao; and ozone and bladder in the Girona Health Region) were not very consistent.

In the case of the Metropolitan Area of Bilbao, the possible association between SO₂ and lung cancer seems to be explained by the high concentrations of this pollutant found at the beginning of the period analysed. We cannot account for the relationship between ozone and the incidence of bladder cancer in the Girona Health Region. Although this relationship was marginally significant, it is obvious that additional research is needed to confirm or rule out this effect.

There are even fewer studies between NO₂ (or, possibly, SO₂) and cancer, particularly lung cancer, although in the above mentioned review, Cohen and Pope (1995) provide a review of the theme (now a little old), as do Pope et al (2002) and Krewski et al (2005) in the United States; Nafstad et al (2006) in Norwegian men; Filleul et al (2006) in France; and finally Vineis et al (2006) in the context of the EPIC (European Prospective Investigation on Cancer and Nutrition) project.

We have not followed the suggestions made by Jackson et al (2006) and analysed the possible interactions between socioeconomic conditions and environmental variables. Heinrich et al (2000) in Germany, Pye et al (2001) in Northern Ireland and Morillo-Frosch and Desdale (2006) in the United States are doing research into partially possible synergies and interactions between both variables.

We believe that the interaction between the socioeconomic index and NO₂ pollution, and maybe SO₂, in the mortality and incidence of lung cancer should be analysed in detail.

In fact, it must be note some insights of non-linearity, specifically in lung cancer and, maybe, in the Bilbao Metropolitan Area (note the different behaviour in both river sides depending on the sex).

-The introduction of explanatory variables, including those statistically significant, did little to reduce the variability explained by the models.

AGRADECIMIENTOS

Este proyecto no sería posible sin la ayuda de muchas personas que he ido conociendo a lo largo de su elaboración durante estos últimos cuatro años.

Al primero al que quiero expresar mi más sincera admiración y mi más profundo agradecimiento es al Dr. Marc Saez Zafra. Además de ser un gran director de Tesis, ha sido capaz de transmitirme su pasión por la estadística, la investigación y la docencia. Debo agradecerle el haber confiado plenamente en mí. Quisiera destacar especialmente el trato recibido por su parte y el cariño y comprensión mostrados en todo momento a lo largo de estos años. Realmente creo que durante estos años, además de la Tesis, hemos conseguido crear una buena amistad. Por todo ello, Marc, bihotz-bihotzez eskerrik asko.

Quisiera agradecer también a todas aquellas personas que de alguna u otra forma han puesto su granito de arena contribuyendo a la realización de este trabajo. Especialmente me gustaría mencionar a la Dra. Carme Saurina, Dr. Rafael Marcos, Dr. Santi Esnaola, Dr. Ricardo Ocaña, Dr. Koldo Cambra, Dra. Marta Blangiardo, Dra. Michela Baccini y Prof. Michael J. Campbell por haber aportado sus conocimientos científicos para la mejora de la Tesis. Muchísimas gracias a todos.

Son muchas las instituciones que han hecho posible la realización de esta tesis doctoral. He de mencionar el respaldo de las dos becas de las que he disfrutado: la beca predoctoral concedida por el Departamento de Educación, Universidades e Investigación del Gobierno Vasco en el último año de la tesis, que me ha permitido terminar y a la vez, disfrutar de la tesis, y la Beca Bayer de Investigación en Economía de la Salud concedida el año 2004. Asimismo agradecer también al Fondo de Investigación Sanitaria (proyectos 02/0735 y 04/0399), del Ministerio de Sanidad gracias a la cual he podido realizar esta tesis doctoral.

Por otra parte, agradecer muy sinceramente tanto a la Unitat d'Epidemiologia i Registre del Càncer de Girona, ICO; al Registre de Mortalitat de Catalunya, Servei d'Informació i Estudis, Direcció General de Recursos Sanitaris, Departament de Salut; al Departament de Medi Ambient; a la Agència de Salut Pública de Barcelona; así como a Eustat, por la aportación de los datos, que sin ellos esta tesis no tendría sentido.

Gràcies a tot el grup GRECS, Carme, Germà, Antònia, Àngels, Lluís, Gemma, Laura, Sonia i Aïna. Ha estat tot un privilegi formar part d'aquest grup. Malgrat de tota la feina que portava cadascú mai ha hagut un moment que no hagi vist un somriure a les vostres cares i això s'agraeix per continuar endavant. Moltes gràcies a tots.

A totes aquelles persones que durant la meva estada a Girona m'hagin aportat part de la seva felicitat, per això agrair-los sincerament. Ha estat tot un plaer conèixer-vos, sempre estareu als meus pensaments.

Y por último, agradecer a mi familia, por toda la paciencia que ha tenido conmigo y haber estado siempre cerca de mí. Eskerrik asko denoi.

1.-Introducción

El cáncer es uno de los principales problemas de salud en los países desarrollados. Bajo el nombre de cáncer se pueden diferenciar más de cien enfermedades con diferente etiología, pronóstico y tratamiento, pero con un elemento común que es su crecimiento no controlado por mecanismos biológicos humanos.

El cáncer puede definirse como un conjunto de enfermedades caracterizadas por el crecimiento y diseminación anómala de células como consecuencia de la acumulación en el ADN de múltiples lesiones genéticas, sean éstas por herencia o adquiridas en los genes que toman parte en los procesos de proliferación, diferenciación y muerte celular. Lo que confiere la característica de malignidad a esta proliferación celular es su capacidad para invadir órganos y tejidos y diseminarse a distancia (Borràs et al, 2001).

1.1.-Mortalidad e Incidencia en España, en Cataluña y en la Regió Sanitària Girona

1.1.1.-Mortalidad por cáncer en España, en Cataluña y en la Regió Sanitària Girona

▪ Mortalidad por cáncer en España

En el año 2003, se produjeron en España 94.100 defunciones debido al cáncer, 59.500 en hombres y 34.600 en mujeres. Los tumores más importantes relacionados con la mortalidad, fueron, para los hombres, el cáncer de pulmón, el cáncer colorrectal y el de próstata, superando todos ellos las 5.000 muertes/año, mientras que en mujeres este número de muertes sólo fue alcanzado por los tumores de mama y el cáncer colorrectal (Lopez-Abente et al, 2005).

En 1998 (último año disponible), para el conjunto de tumores en los hombres, España ocupaba una situación intermedia entre los países europeos. Los varones españoles

presentaban tasas de mortalidad más altas que el conjunto de los países de la Unión Europea para las localizaciones: cavidad oral, estómago, hígado, laringe, pulmón, vejiga y enfermedad de Hodgkin; la mayor parte de las cuales están íntimamente relacionadas con el consumo de tabaco. Los cánceres para los cuales España ocupa el primer puesto fueron el cáncer de laringe y el de vejiga (Ver Figura 1).

En las mujeres, la tasa de mortalidad global por cáncer en 1998 fue la más baja de Europa. Las mujeres españolas sólo presentaban tasas por encima de la media europea para el cáncer de hígado, cuerpo de útero y enfermedad de Hodgkin.

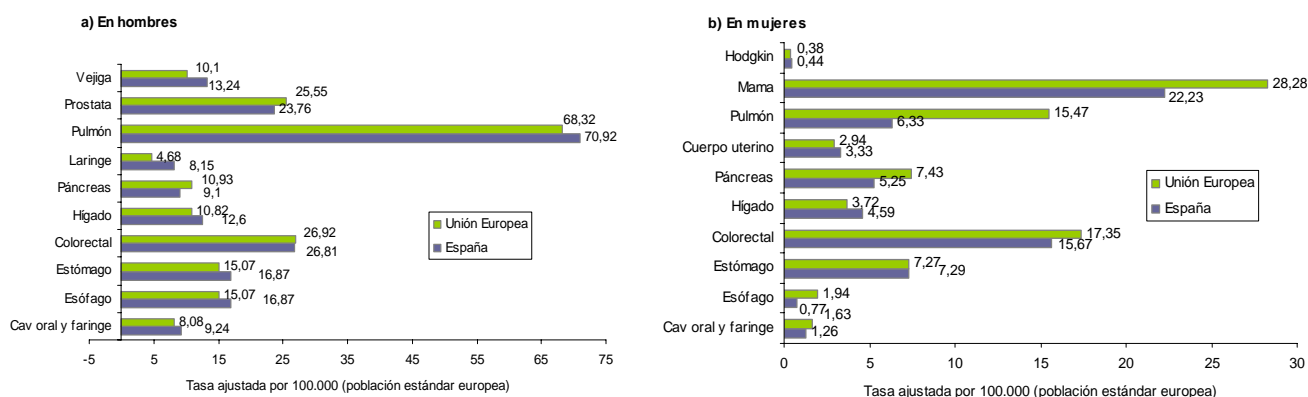


Figura 1.– Mortalidad por cáncer en España y en la Unión Europea en 1998: Tasas ajustadas por 100.000 (Población estándar Europea) (Fuente: Lopez–Abente et al, 2005)

Dentro de España (ver Tabla 1), la mayor mortalidad por cáncer en hombres se observa en Asturias y en el País Vasco, con tasas ajustadas (por edad y sexo) superiores a los 280 casos por 100.000 personas/año, seguidos de Cantabria (Tabla 1a). La mortalidad es menor en Castilla–La Mancha y Melilla (tasas en torno a los 215 casos por 100.000 personas/año). Por localizaciones tumorales, llama la atención las altas tasas de cáncer de cavidad bucal y faringe en Canarias, Asturias, Cantabria, País Vasco y Galicia; de esófago en el País Vasco, Canarias, Galicia y Asturias; de estómago en Castilla y León, de cáncer colorrectal en Asturias, País Vasco y Baleares; de hígado en Madrid, Andalucía y Cataluña; de páncreas en Ceuta y Navarra; de laringe en Asturias; de pulmón en Extremadura, Asturias, Andalucía y Baleares; de próstata en La Rioja; de

vejiga en Murcia y en Andalucía; de tumores del sistema nervioso central en La Rioja y Navarra; y de linfomas no Hodgkin en Canarias.

En mujeres la mayor tasa ajustada de mortalidad global por cáncer corresponde a Canarias (122 casos por 100.000 personas-año), seguido de Baleares y Asturias. La mortalidad es menor en la CCAA de Castilla-La Mancha y en Melilla (ver Tabla 1b).

Por localizaciones, cabe señalar la mayor mortalidad de cáncer de estómago en Castilla-León y en Galicia; de cáncer colorrectal en Baleares; de hígado en Ceuta, Melilla y Andalucía; de páncreas en Navarra y Ceuta; de pulmón en Canarias, País Vasco, Madrid, Asturias y Ceuta; de mama en Baleares y Canarias; de útero en Melilla y Canarias; de ovario en Asturias, La Rioja, Baleares y Aragón; de vejiga en La Rioja; de tumores malignos del sistema nervioso central en Navarra y La Rioja; y de linfomas no Hodgkin en Canarias.

▪ Mortalidad por cáncer en Cataluña

En el período 1996–2000 (últimos años disponibles) se registraron en Cataluña 72.666 muertos, 45.202 hombres y 27.464 mujeres. Es decir, 50 personas murieron diariamente por causa del cáncer, 31 hombres y 19 mujeres. El mayor número de defunciones se debe al cáncer de pulmón (27,8%) y colorrectal (11,7%) en hombres (ver Figura 2), y cáncer de mama (17,9%) y también colorrectal (15,7%) en mujeres (Borràs et al, 2001).



Figura 2.– Mortalidad por cáncer en Cataluña, 1997–2000. Frecuencia relativa (%) de las principales localizaciones tumorales (Fuente: Borràs et al, 2001).

En hombres, la mortalidad global por cáncer (excepto piel no melanoma) fue máxima entre 1991–1992 momento en que se estabiliza y disminuye ligeramente hasta 1998. De la misma manera, la tasa de mortalidad en mujeres ha ido disminuyendo desde 1975 (Fernández et al, 2001) (Ver Figura 3).

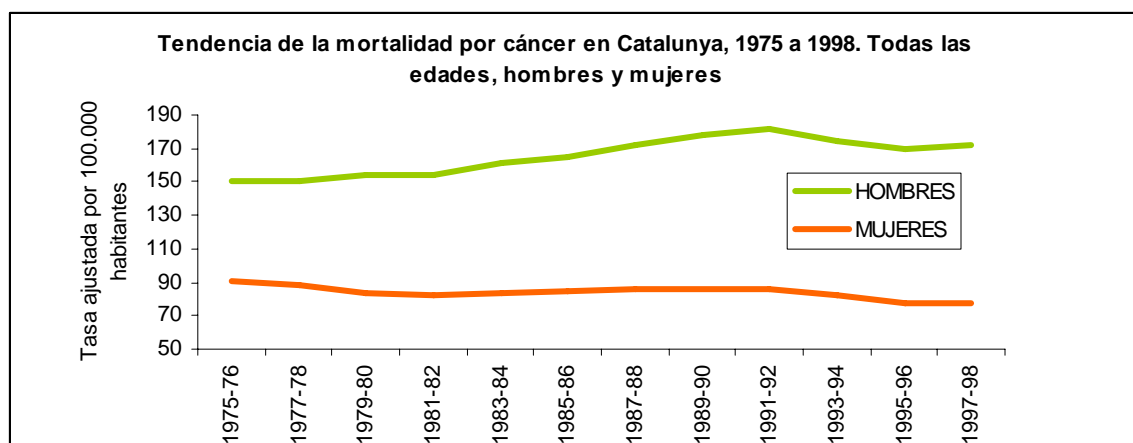


Figura 3.- Mortalidad por cáncer en global (excepto piel no melanoma), período 1975–1998, en Cataluña. Tasa estandarizada por edad (población mundial) (Fuente: Fernández et al, 2001).

El cáncer de pulmón fue la primera causa de mortalidad por cáncer en los hombres en las últimas décadas. El incremento más marcado se produjo entre 1981–1982 y 1989–1990 (4,0% anual), posteriormente se estabilizó y disminuyó hasta 1997–98 (–0,5% anual) (ver Figura 4a y Tabla 2a).

El cáncer colorectal fue la segunda causa de mortalidad por cáncer más importante tanto en hombres como en mujeres. La tendencia reciente indica una estabilización de su mortalidad, después de los aumentos observados desde 1980.

En los hombres, el cáncer de próstata ha pasado a ser la tercera causa de mortalidad por cáncer en 1997–1998, por delante del cáncer de estómago, como un reflejo de la estabilidad de la mortalidad por cáncer de estómago (ver Figura 4a). En mujeres, el cáncer de estómago ha disminuido también durante este período, y se ha situado como la cuarta causa de mortalidad por cáncer junto con los cánceres de útero, que también han disminuido (ver Figura 4b).

La tendencia reciente en la mortalidad por cánceres relacionados con el tabaco (laringe, cavidad oral y faringe, esófago, páncreas, vejiga urinaria, riñón, además del cáncer de pulmón) es favorable, con una reducción anual desde 1989–90 del $-1,1\%$ en hombres, y se ha mantenido estable en las mujeres (Borràs et al, 2001; Fernández et al, 2001).

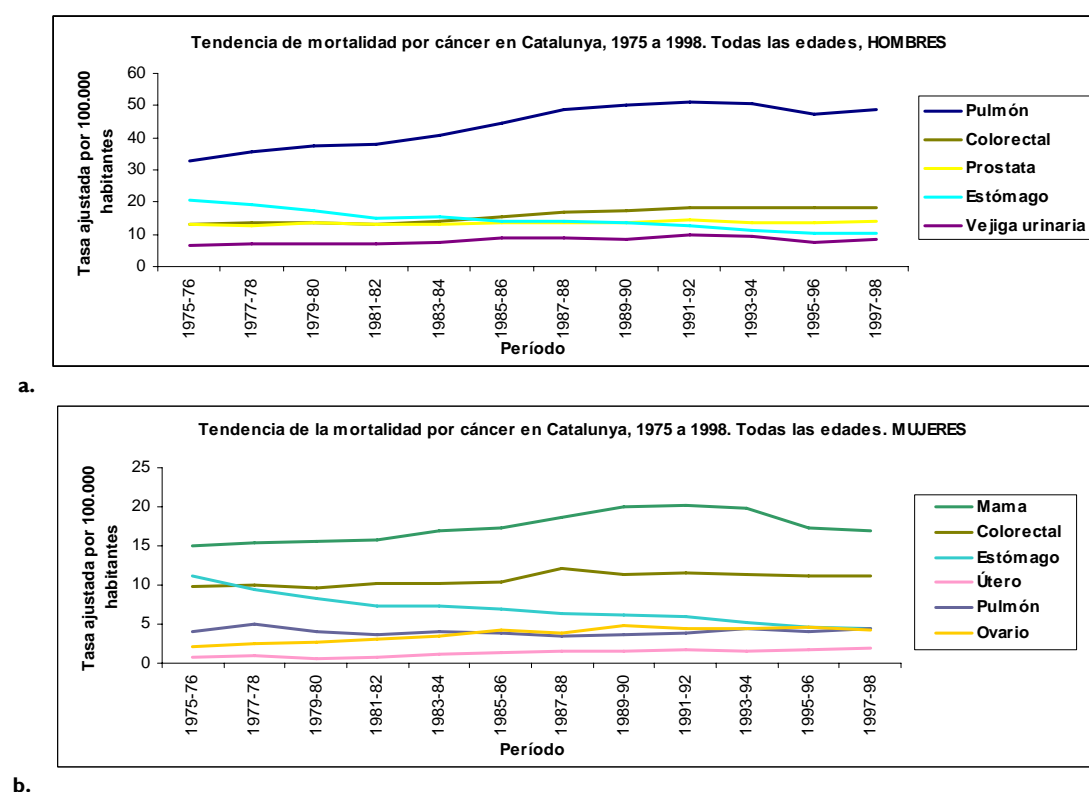


Figura 4.- Mortalidad por cáncer, período 1975–1998, en Cataluña. Diversas localizaciones. Tasa estandarizada por edad en hombres (a) y en mujeres (b) (Población mundial). (Fuente: Borràs et al, 2001; Fernández et al, 2001)

En mujeres (ver Figura 4b y Tabla 2b), el cáncer de pulmón ha pasado a ser la tercera causa de muerte por cáncer, por delante del cáncer de estómago y útero, y después de los cánceres de mama y colorectal. La tendencia más reciente (desde 1989–90) en la mortalidad por cáncer de pulmón en mujeres indica un incremento del $1,9\%$ anual.

El cáncer de mama ha sido la principal causa de muerte por cáncer en las mujeres en las últimas décadas, con un incremento hasta 1991–1992, momento en el que se aprecia una disminución hasta la actualidad.

Mortalidad por cáncer en la Regió Sanitària Girona

Según los datos del Registro de mortalidad de Cataluña durante el año 2001 (últimos años disponibles) murieron 4,996 personas residentes en la Regió Sanitària Girona¹ y 1.408 (877 hombres y 531 mujeres) por cáncer (28,7%). El mayor número de defunciones se debe al cáncer de pulmón (15,6%) y colorrectal (7,9%) en hombres, y cáncer de mama (7,5%) y también colorrectal (5,9%) en mujeres (ver Figura 5) (UERCG, 2005).



Figura 5.– Mortalidad por cáncer en la Regió Sanitària Girona, 2001. Frecuencia relativa (%) de las principales localizaciones tumorales (Fuente: elaboración propia).

La tasa de mortalidad estandarizada por edad (población mundial) de diferentes neoplasias por género se muestra en la Tabla 3.

¹ La Regió Sanitària Girona coincide con la provincia de Girona con la excepción de la comarca de la Cerdanya.

PERÍODO	Cavidad oral y faringe	Esófago	Estómago	Colorectal	Hígado	Vesícula biliar	Páncreas	Laringe	Pulmón	Melanoma	Próstata	Vejiga urinaria	Riñón	Encéfalo	Tiroides	LH	LNH	Mieloma	Leucemias
1975-76	3,6	5,4	20,5	13,3	9,8	1,2	3,9	7,9	32,8	0,4	12,9	6,7	2,0	5,6	0,2	1,5	1,8	1,0	5,4
1977-78	3,4	5,2	19,0	13,6	9,6	0,7	4,0	7,0	35,6	0,5	12,7	7,3	1,8	5,5	0,2	1,1	1,9	0,9	5,4
1979-80	3,2	5,0	17,4	13,5	10,4	1,1	4,3	7,5	37,3	0,5	13,7	7,0	2,0	5,6	0,2	1,0	2,3	1,4	5,4
1981-82	4,4	5,5	14,9	13,2	9,8	1,3	5,1	7,4	38,0	0,5	13,0	6,9	2,2	5,3	0,2	0,7	2,5	1,5	5,5
1983-84	5,8	5,9	15,6	14,2	9,8	1,4	5,2	6,8	41,0	0,7	13,0	7,7	2,3	4,7	0,2	1,2	2,6	1,6	5,9
1985-86	6,4	5,7	13,9	15,4	8,5	1,5	5,4	6,6	44,7	1,1	13,8	8,7	2,2	4,4	0,1	0,9	3,0	1,7	5,6
1987-88	6,4	5,7	14,1	16,7	8,0	1,8	6,2	6,6	48,7	1,0	13,7	9,0	2,9	4,2	0,3	0,8	3,3	1,8	5,7
1989-90	6,5	6,5	13,5	17,4	10,1	1,5	6,0	6,4	50,1	1,1	13,5	8,6	3,2	4,2	0,3	0,8	3,9	2,0	6,2
1991-92	7,2	5,6	12,7	18,3	10,3	1,3	6,2	6,5	51,0	1,3	14,4	9,6	3,2	4,1	0,3	0,7	4,2	2,1	6,0
1993-94	7,3	5,9	11,4	18,5	9,2	1,4	6,2	5,7	50,5	1,1	13,6	9,4	3,0	4,4	0,2	0,5	3,9	1,8	5,6
1995-96	5,9	5,9	10,5	18,1	10,2	1,7	6,4	5,2	47,2	1,3	13,5	7,3	2,8	4,1	0,1	0,7	3,6	2,1	5,1
1997-98	6,9	5,1	10,4	18,5	9,0	1,4	6,5	5,3	48,8	1,1	14,2	8,6	3,1	4,3	0,3	0,9	3,8	1,9	4,9

a.- Hombres

PERÍODO	Cavidad oral y faringe	Esófago	Estómago	Colorectal	Hígado	Vesícula biliar	Páncreas	Laringe	Pulmón	Melanoma	Mama	Útero	Ovario	Vejiga urinaria	Riñón	Encéfalo	Tiroides	LH	LNH	Mieloma	Leucemias
1975-76	0,7	0,7	11,1	9,8	7,3	1,5	2,5	0,3	4,1	0,3	15,1	0,7	2,2	1,3	1,3	3,5	0,3	0,5	1,0	0,6	4,1
1977-78	0,6	0,9	9,3	10,0	6,3	1,9	2,6	0,3	4,9	0,4	15,3	1,0	2,6	1,4	1,1	3,5	0,3	0,7	1,0	0,8	3,8
1979-80	0,6	0,7	8,2	9,5	5,7	1,6	2,5	0,3	4,0	0,4	15,6	0,6	2,7	1,5	0,9	3,5	0,3	0,6	1,4	0,9	3,4
1981-82	0,6	0,7	7,3	10,1	5,0	1,8	2,7	0,1	3,7	0,4	15,7	0,8	3,2	1,4	1,0	3,2	0,4	0,5	1,3	1,3	3,6
1983-84	0,7	0,7	7,4	10,3	4,5	2,2	3,0	0,2	4,1	0,7	17,0	1,2	3,5	1,2	1,3	2,7	0,4	0,6	1,5	1,1	3,4
1985-86	0,9	0,5	6,9	10,4	4,3	2,4	3,2	0,2	3,9	0,6	17,3	1,3	4,2	1,4	0,9	2,7	0,5	0,6	2,0	1,3	3,6
1987-88	0,9	0,6	6,4	12,2	3,6	2,1	3,3	0,1	3,4	0,7	18,7	1,5	3,9	1,4	0,9	2,8	0,4	0,4	2,3	1,3	4,3
1989-90	1,0	0,6	6,2	11,4	4,2	2,0	3,6	0,1	3,7	0,8	20,1	1,5	4,8	1,3	1,0	2,6	0,5	0,5	2,7	1,1	3,7
1991-92	1,0	0,5	6,0	11,6	3,8	2,0	3,2	0,1	3,8	0,7	20,2	1,8	4,4	1,4	1,2	2,6	0,4	0,3	2,3	1,4	3,6
1993-94	0,8	0,5	5,2	11,3	3,2	2,0	3,5	0,1	4,3	0,8	19,7	1,6	4,5	1,4	1,0	2,7	0,4	0,3	2,9	1,4	3,2
1995-96	0,8	0,5	4,7	11,1	3,3	1,7	3,3	0,1	4,0	0,7	17,3	1,7	4,6	1,1	1,3	2,5	0,3	0,4	2,2	1,7	3,1
1997-98	0,8	0,5	4,3	11,1	3,3	1,6	3,3	0,1	4,4	0,8	16,8	1,8	4,1	1,2	1,2	2,6	0,4	0,3	2,6	1,6	3,2

b.- Mujeres

Tabla 2. Tasa de mortalidad estandarizada por edad (población mundial) de los cánceres registrados en Cataluña, en hombres (a) y en mujeres (b), período 1976-1998 (Fernandez E et al, 2001).

Localizaciones tumorales en hombres	Tasa Ajustada		Localizaciones tumorales en mujeres
Pulmón	45,2	18,3	Mama
Colorectal	19,1	12,1	Colorectal
Próstata	12,2	5,5	Ovario
Vejiga urinaria	8,3	4,1	Pulmón
Estómago	7,3	3,8	Leucemias
Hígado	7,0	3,6	Páncreas
Laringe	6,5	2,9	Estómago
Páncreas	6,3	2,9	Sistema nervioso
Esófago	5,7	2,5	Linfoma No Hodgkin
Linfomas No Hodgkin	5,1	2,5	Hígado
Cavidad oral y faringe	5,8	2,1	Útero
Sistema Nervioso	4,7	1,6	Vesícula biliar
Riñón	3,5	1,4	Riñón
Leucemias	2,5	1,3	Vejiga urinaria
Mielomas	1,2	1,2	Melanoma
Vesícula Biliar	0,8	1,1	Cavidad oral y faringe
Melanoma Piel	1,0	0,9	Mieloma
Huesos y articulaciones	0,6	0,7	Esófago
Mesotelioma	0,4	0,4	Tiroides
		0,1	Linfoma Hodgkin

Tabla 3.- Tasa de mortalidad por diferentes localizaciones de cáncer, ajustada por edad, población mundial, en la Regió Sanitària Girona, año 2001 (Fuente: UERCG, 2005)

1.1.2.-Incidencia de cáncer en España, en Cataluña y en la Regió Sanitària Girona

Aunque las Estadísticas de Mortalidad son las más utilizadas para conocer el impacto del cáncer en la población, ya que están disponibles en la mayor parte de los países, no están exentas de limitaciones. Una de estas limitaciones es inherente a la naturaleza de los datos, es decir, que los tumores más letales están más representados. Otra de las limitaciones es la falta de precisión de la información contenida en los certificados de defunción. Por ello, para suplir estas limitaciones y tener un mejor conocimiento de la presentación del cáncer en la población se crearon los Registros de Cáncer de base poblacional que permiten estimar la incidencia de cáncer en la población de referencia, así como identificar patrones temporo-espaciales en dichas poblaciones.

▪ Incidencia de cáncer en España

En el año 2003 existían registros de cáncer poblacionales en Cantabria, País Vasco, Navarra, Zaragoza, Girona, Tarragona, Mallorca, Albacete, Cuenca, Murcia, Granada,

Islas Canarias, Asturias y La Rioja. Los registros más antiguos del país son los de Zaragoza (más antiguo) y Navarra, por lo que tienen series temporales de incidencia de cáncer más amplias que otros registros.

El número de casos/año de cáncer estimado para el periodo 1997–2000 (últimos datos disponibles) fue de 162.000, de los cuales 25.600 casos fueron carcinomas colorrectales. Los siguientes tumores en frecuencia son el cáncer de pulmón con 18.800 casos nuevos y cáncer de mama en mujeres con 15.979 casos nuevos (Lopez–Abente et al, 2005).

España ocupa una posición intermedia en incidencia del cáncer, respecto a otros países europeos. Los hombres españoles presentan tasas de incidencia más altas que el conjunto de los países de la Unión Europea para las localizaciones: cavidad oral, laringe, pulmón, vejiga. En mujeres en cambio, igual que con la mortalidad, la tasa de incidencia se encuentra por debajo de la media Europea (Lopez–Abente et al, 2005).

Por lo que se refiere a los registros españoles (período 1990–1999, último para los que se dispone de datos), Girona y Navarra son las que proporcionan mayores tasas de incidencia de cáncer. Por localizaciones tumorales en hombres, llama la atención las altas tasas de cáncer de cavidad bucal y faringe en Canarias y Granada; estómago en Navarra; colorectal en Girona y Navarra; pulmón en Asturias; próstata y vejiga en Girona (ver Tabla 4a).

En mujeres, Girona, Navarra y Tarragona son los registros que presentaron mayor tasa de incidencia. Por localizaciones tumorales, llaman la atención las neoplasias; estómago en Cuenca y en Navarra; colon en Tarragona y en Girona; mama en Girona, Navarra, Tarragona y en Canarias y tiroides en Navarra (Ver Tabla 4b).

▪ Incidenca de càncer en Catalunya

En Catalunya existen dos registros poblacionales de càncer, el de Tarragona, que cubre las regiones Sanitàrias de Tarragona y de Terres de l'Ebre, y que recoge casos desde 1980 (Borràs et al, 1997a; Borràs et al, 1997b) y el de Girona, que cubre la Regió Sanitària Girona y que recoge casos desde 1994 (Viladiu et al, 2000).

En el período 1996–1997, la frecuencia relativa de los diferentes tipos de càncer presenta un perfil muy diferenciado según el sexo. En las Figuras 6 y 7 se presenta la frecuencia relativa de las principales localizaciones tumorales observada en los registros de Tarragona y Girona correspondientes al período 1994–1997.

17,7	Pulmón			Mama	29,6
15,6	Próstata			Colorectal	15,2
14,6	Colorectal			Útero	6,0
9,9	Vejiga Urinaria			Estómago	4,9
5,9	Estómago			Primario	4,4
5,7	Cavidad oral y faringe			desconocido	4,4
				Ovario	4,1

Figura 6.– Incidenca de càncer en Tarragona, 1994–1997. Frecuencia relativa (%) de las principales localizaciones tumorales (Fuente: Borràs et al, 2001).

16,7	Pulmón			Mama	27,5
13,4	Próstata			Colorectal	14,9
13,0	Colorectal			Útero	5,9
11,7	Vejiga Urinaria			Estómago	4,1
6,6	Cavidad oral y faringe			Ovario	3,8
5,7	Estómago			Primario	3,7
				desconocido	

Figura 7.– Incidenca de càncer en Girona, 1994–1997. Frecuencia relativa (%) de las principales localizaciones tumorales (Fuente: Borràs et al, 2001).

Localizaciones	Albacete, 1993-1997	Cuenca, 1993-1997	Girona, 1994-1997	Granada, 1993-1997	Murcia, 1993-1996	Navarra, 1993-1997	Asturias, 1992-1995	Canarias, 1993-1995	Tarragona, 1993-1997	Zaragoza, 1991-1995
Pulmón	55,4	52,9	73,7	72,6	72,6	71,3	91,6	64,5	67,5	72,0
Próstata	44,7	43,7	62,3	35,9	43,4	64,9	37,1	44,5	45,4	47,1
Vejiga urinaria	36,1	27,9	56,1	46,8	51,4	52,3	42,1	27,8	48,5	40,1
Cavidad oral y faringe	29,4	28,3	23,8	32,3	27,4	26,0	26,2	35,5	25,4	23,5
Estómago	21,8	29,7	21,1	18,9	18,9	31,9	26,8	15,5	21,1	24,8
Colon	21,3	19,1	40,5	23,4	29,5	32,6	30,0	20,7	32,6	25,2
Laringe	15,7	19,5	13,1	18,1	20,9	19,3	21,2	18,7	16,6	24,4
Recto	13,8	12,4	21,9	15,2	21,2	22,6	17,5	16,9	18,2	18,1
Linfomas No Hodgkin	10,7	9,3	17,5	10,1	12,1	12,2	11,9	14,8	12,3	11,2
Hígado	8,4	7,3	14,2	11,4	9,3	12,8	10,0	10,6	9,6	8,9
Riñón	8,1	6,7	10,2	8,2	5,8	10,7	11,4	4,7	8,3	8,3
Páncreas	7,3	6,1	8,6	8,6	8,3	11,9	8,1	9,7	9,3	8,1
Encéfalo	5,7	9,0	8,1	6,8	7,6	10,5	6,4	8,3	9,6	9,5
Esófago	5,6	4,5	8,3	5,5	6,5	9,2	10,2	10,8	7,7	7,1
Mielomas	5,8	4,5	3,8	4,1	4,7	4,5	4,6	3,5	3,2	4,2
Melanoma Piel	4,4	4,4	5,4	5,3	5,5	6,0	3,8	3,2	6,8	3,5
Vesícula	2,8	2,0	3,6	4,0	2,8	4,6	1,7	3,0	3,2	2,9
Linfomas Hodgkin	2,8	2,8	3,6	1,9	2,7	3,3	3,4	2,7	3,5	3,6
Testículo	2,7	1,8	2,3	1,8	2,1	1,7	2,0	1,7	2,1	1,9
Tiroides	1,5	1,7	2,5	1,1	1,7	3,1	2,3	1,7	1,3	1,0
Huesos	1,3	1,3	1,5	1,2	1,4	1,2	1,1	1,4	1,2	1,3
Todos*	344,4	326,8	444,8	378,6	402,0	456,4	425,3	352,3	400,2	383,8

a)

Localizaciones	Albacete, 1993-1997	Cuenca, 1993-1997	Girona, 1994-1997	Granada, 1993-1997	Murcia, 1993-1996	Navarra, 1993-1997	Asturias, 1992-1995	Canarias, 1993-1995	Tarragona, 1993-1997	Zaragoza, 1991-1995
Mama	70,5	63,4	84,7	60,9	69,9	81,7	66,6	80,2	80,4	67,7
Colon	16,3	17,7	26,0	16,0	23,0	19,0	17,9	19,2	26,4	18,3
Ovario	12,9	12,7	10,5	10,6	9,5	11,1	12,5	10,8	10,7	11,9
Cuerpo útero	11,2	14,1	17,1	14,3	16,9	17,2	14,9	15,4	17,2	14,8
Estómago	9,8	13,8	9,3	8,1	9,7	12,7	12,0	7,3	8,2	11,2
Recto	9,4	7,9	9,9	8,9	12,4	10,7	9,2	10,0	9,2	9,1
Cérvix	6,9	4,3	9,3	8,0	9,6	4,8	10,6	10,4	11,4	7,0
Linfoma No Hodgkin	6,9	3,5	7,6	7,9	7,9	9,7	8,6	13,5	10,2	8,2
Encéfalo	6,2	6,4	6,2	5,0	4,7	7,5	5,2	5,5	7,0	6,0
Melanoma	6,2	4,7	8,1	7,7	7,3	7,4	5,5	4,6	6,7	4,1
Vesícula biliar	5,9	6,3	3,8	8,2	4,3	6,0	2,3	5,0	4,4	4,6
Mieloma	5,7	2,6	3,0	2,8	3,3	3,3	2,5	3,4	2,8	2,5
Vejiga urinaria	5,5	4,3	8,6	5,3	5,8	6,0	5,5	4,5	6,8	4,9
Páncreas	5,2	4,1	5,7	4,6	5,2	6,7	4,8	7,2	6,8	5,5
Tiroides	4,8	4,9	5,2	7,3	6,1	9,4	5,4	7,1	5,7	3,8
Pulmón	4,5	3,1	6,6	5,5	5,6	6,3	6,4	8,7	6,7	5,3
Riñón	3,7	3,6	3,6	4,2	2,6	5,4	4,2	2,6	3,4	4,5
Hígado	3,6	5,0	3,7	3,8	3,2	3,9	2,3	5,0	4,0	3,4
Cavidad oral y faringe	3,2	3,7	3,1	3,1	3,9	3,5	3,4	4,4	3,3	2,7
Linfoma Hodgkin	1,0	1,6	1,6	1,6	1,4	2,4	2,4	2,3	2,2	2,3
Huesos	0,7	1,3	0,9	1,1	0,6	1,3	1,1	0,7	0,9	0,8
Esófago	0,6	0,6	1,0	0,9	0,7	0,9	0,7	1,4	0,7	0,7
Laringe	0,5	0,4	0,3	0,2	0,6	0,8	0,4	0,6	0,6	0,5
Todos*	223,8	212,8	262,8	219,4	237,9	263,5	236,4	258,5	262,6	225,4

b)

Tabla 4.- Tasas de incidencia ajustadas por la población Europea por cánceres en los diferentes registros poblacionales (a) en hombres y b) en mujeres) (Fuente: Lopez-Abente et al, 2005).* Todos los tumores malignos excepto cáncer de piel.

Tal y como se puede apreciar en las dos figuras, los tumores más frecuentes fueron el cáncer de pulmón (alrededor del 17% del total de cáncer), seguido del de próstata (15%) y del colorectal (14%) en hombres. Hay que remarcar que el cáncer colorectal ocupaba la segunda posición hasta principios de los noventa (Borràs et al, 1997b) y es a partir del año 1995 cuando es sobrepasado por el de próstata. En las mujeres el tumor más frecuente es el de mama (28%) seguido a una cierta distancia por el cáncer colorectal (15%).

El conjunto de tumores relacionados con el tabaco (cavidad oral y faringe, laringe, páncreas, pulmón, vejiga urinaria y riñón) representaron el 45% del total de tumores en hombres mientras que en las mujeres solamente el 11% como consecuencia de la difusión más reciente del hábito tabáquico en este sexo (Borràs et al, 2000). Por otra parte, los tumores del aparato genital femenino juntamente con el de mama representaron el 42% del cáncer en las mujeres.

En cuanto a las tasas de incidencia de las diferentes neoplasias en los dos registros catalanes, (ver Tabla 4a y 4b) hay que señalar que Girona presenta mayores tasas, tanto en hombres como en mujeres, que Tarragona.

▪ Incidencia de cáncer en Regió Sanitària Girona

En la Regió Sanitària Girona se registraron 3.635 casos (1.520 casos nuevos en mujeres y 2.115 casos nuevos en hombres) de cáncer en el año 2001 (se excluyen los cánceres de piel no melanoma y los carcinomas 'in-situ'). Esto representa una incidencia de 245,6 (dejaría 1 decimal) nuevos casos por 100.000 habitantes/año en hombres y de 157,4 por 100.000 habitantes/año en mujeres (UERCG, 2004). El tumor de mayor incidencia en la Regió es el colorectal, con 425 nuevos casos anuales, seguido del cáncer de próstata con 311, mama con 297 y tráquea, bronquios y pulmón con 288. La Figura 8 representa el crecimiento de casos incidentes de cáncer de mama y próstata (Figura 8a), colorectal (Figura 8b) y pulmón (Figura 8c). Tal y como se

aprecia en las tres figuras, los cuatro cánceres presentan un crecimiento paulatino, excepto el cáncer de colorectal en hombres.

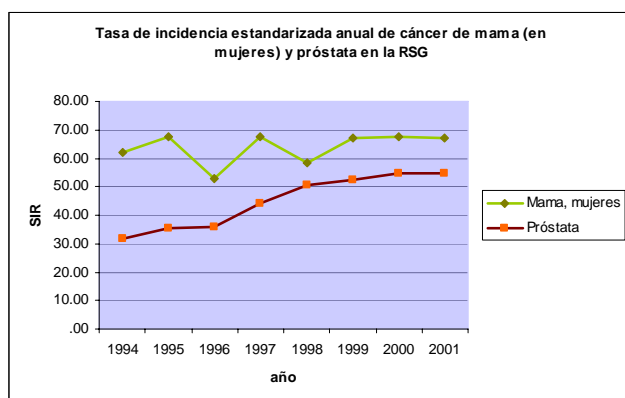


Figura 8a.- Cáncer de mama y próstata

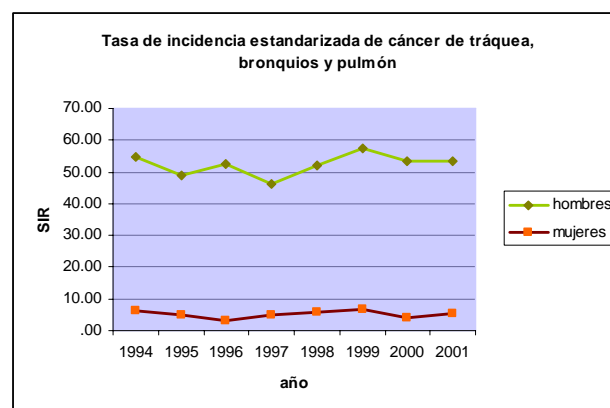


Figura 8b.- Cáncer de tráquea, bronquios y pulmón

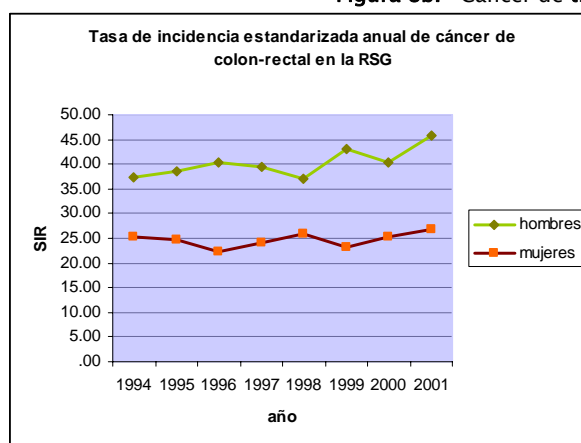


Figura 8c.- Cáncer de colon-rectal

Figura 8.- Tasa de incidencia estandarizada por edad (población Girona) de los cánceres con mayor frecuencia en la Región Sanitaria Girona (período 1994-2001) (Fuente: elaboración propia).

En los datos del año 2001 destaca el aumento de la incidencia del cáncer de próstata que se consolida como primera localización tumoral entre los hombres, por delante del cáncer de pulmón (ver Tabla 5). Cabe destacar que la segunda causa de mortalidad por cáncer entre los hombres es también el cáncer de próstata. Entre las mujeres llama la atención la situación, en cuarto lugar, de los linfomas no Hodgkin, junto a localizaciones como ovario o estómago.

a) Neoplasias en Mujeres	Cruda ¹	Tasa de Estandarización Directa ²	b) Neoplasias en Mujeres	Cruda	Tasa de Estandarización Directa
Todos	432,2	245,6	Todos	286,6	157,4
Próstata	77,9	37,1	Mama	82,3	50,9
Traquea, bronquios, pulmón	74,3	42,7	Colon-rectal	43,4	20,1
Colon-rectal	61,0	32,8	Neoplasias linfoides	20,2	12,0
Vejiga urinaria	37,4	20,0	Linfomas no Hodgkin	10,3	5,9
Neoplasias linfoides	27,3	18,6	Leucemias linfoides	3,5	2,0
Linfomas no Hodgkin	15,1	9,8	Mielomas	4,1	1,9
Leucemias linfoides	4,7	2,7	Linfomas Hodgkin	1,5	1,5
Linfomas Hodgkin	2,6	2,4	Útero	18,7	10,6
Mielomas	4,4	2,4	Estómago	12,2	5,0
Cavidad oral y faringe	20,9	13,7	Ovario	11,6	7,0
Estómago	20,3	10,9	Origen desconocido	11,2	4,1
Origen desconocido	15,6	8,3	Traquea, bronquios, pulmón	7,7	4,1
Hígado	13,2	7,7	Melanoma piel	7,5	4,5
Riñón, otros urinarios	11,5	7,0	Páncreas	7,0	3,0
Sistema nervioso	7,3	5,4	Vejiga urinaria	6,8	2,5
Esófago	8,4	5,2	Tiroides	5,9	4,6
Páncreas	8,2	4,6	Sistema nervioso	5,8	4,4
Melanoma piel	5,1	3,3	Riñón, otros urinarios	5,4	2,7
Vejiga biliar	4,1	2,1	Hígado	4,9	2,0
Conjuntivo y partes blandas	2,6	1,8	Vejiga biliar	4,9	1,8
Testículo	2,3	2,1	Cavidad oral y faringe	3,5	1,7
Tiroides	1,9	1,5	Conjuntivo y partes blandas	2,4	1,7
Kaposi	1,4	1,0	Esófago	1,1	0,5
Huesos y articulaciones	1,3	1,2	Huesos y articulaciones	1,0	0,8
			Kaposi	0,1	0,1

Tabla 5.– Incidencia del cáncer en la Regió Sanitària Girona, tasas ajustadas con la población mundial, período entre 1994–2001 (1,2 ver Capítulo 2) (Fuente: elaboración propia)

En resumen, la Regió Sanitària Girona se situa en las primeras posiciones en relación a la incidencia de cáncer respecto a otros registros españoles y en algunas neoplasias (por ejemplo cáncer de mama) en la primera posición. Cabe señalar que esta posición relativa se efectuó con datos hasta 1999 y que, dado el aumento reciente de la incidencia en Girona, al menos en algunas neoplasias, la situación de la Regió puede estar en posiciones más elevadas.

1.2.–Cáncer y medio ambiente

Como conjunto de enfermedades que es, el cáncer tiene también múltiples factores que lo determinan. Éstos pueden ser tanto externos (factores ambientales, factores ocupacionales e infecciones víricas) como internos (hormonas, aspectos inmunitarios y mutaciones heredadas). Este conjunto de factores causales pueden actuar

conjuntamente o secuencialmente por tal de iniciar o promover el desarrollo de la enfermedad.

Desde el punto de visto cronológico, desde los años setenta, los primeros carcinógenos se encontraron relacionados con la ocupación. De hecho, existe un gran volumen de datos epidemiológicos y experimentales relacionados con la exposición ocupacional (Tardon et al, 1999; Zhang et al, 1999; Acquavella et al, 2004; Alavanja et al, 2004; Chen et al, 2004; Howsam et al, 2004; Kass et al, 2004; Peláez et al, 2004; Bahia et al, 2004; De Roos et al, 2005; Forssén et al, 2005; Kaerlev et al, 2005; Krstev et al, 2005; Reid et al, 2005). Con el fin de sintetizar toda esta información IARC publicó, por primera vez, en el año 1987 una lista de compuestos cancerígenos para el ser humano, agrupados por el tipo de industria y el tipo de compuesto (ver Tabla 6) (IARC, 1987).

Sin embargo, la mayoría de las neoplasias se atribuye a factores de tipo ambiental (un 80% según Higghinson, 1969). De acuerdo con la literatura, los factores ambientales más relacionados con el cáncer son:

- contaminación atmosférica
- hidrocarburos policíclicos aromáticos, PAH
- metales
- asbestos
- radiación
- pesticidas
- disruptores endocriños.

Substancia o compuesto	Ocupación o industria en la cual se encuentra la substancia	IARC (volumen de la monografía, anual)	Tipo de cáncer
Radiación ionizante: Rayos alfa, beta, rayos X y rayos gamma.	Piloto, Asistente de vuelo, Astronauta, Personal Médico o de rayos X, planta industrial, energía nuclear	Vol. 75 (2000a)	Hueso
		Vol. 78 (2001a)	Leucemia, Pulmón Riñón, Tiroides, Tejido de Sarcoma
Radiación no ionizante: Radicación solar (Radiación UV)	Trabajadores expuestos a la radiación solar; agricultores, obreros, entre otros.	Vol. 55 (1992b)	Piel melanoma
Amianto	Mineros, aisladores, astilleros, trabajadores en el metal; industria de cemento del amianto	Suppl. 7 (1987)	Mesotelioma
Erionita (mismo comportamiento que el asbesto)	Tratamiento de escombros, tratamiento de aguas, sistema de control de contaminación atmosférica, agregados de cemento; material de construcción	Suppl.7 (1987)	Mesotelioma
Sílice, cristalina	Industrias del granito y piedras, cerámicas, vidrio e industrias relacionadas, fundiciones e industrias metalúrgicas, abrasivos, construcción	Vol. 68 (1997b)	Pulmón
Polvos de la madera	Leñadores y aserraderos, industria del papel	Vol 62 (1995b)	Cavidad nasal
Metal y componentes del metal	Fusión de metales noferrosos, producción, embalaje	Suppl. 7 (1987)	Piel
Arsénico y componentes del arsénico	Uso de pesticidas que contienen arsénico, mineros expuestos al arsénico	Suppl. 7 (1987)	Pulmón, Riñón
Berilio	Extracción de berilio, industrias aeroespaciales, industrias electrónicas y nucleares, joyeros	Vol. 58 (1993a)	Pulmón
Cadmio y componentes del cadmio	Trabajadores en la fundición del cadmio, trabajadores de producción de batería, producción de pigmentos y tintes	Vol.58 (1993a)	Pulmón
Componentes del cromo	Plantas de producción del cromo, pigmentos y tintes, tratamiento de piel, fotografía, litografía, perfumes sintéticos, pirotecnias, resistencia a la corrosión	Vol. 49 (1990a)	Pulmón
Componentes de níquel, combinaciones de óxido de níquel	Fundiciones y refineries del níquel, soldadura	Vol.49 (1990a)	Pulmón, Cavidad nasal
Benceno	Industria del calzado, industrias químicas, farmacéuticas y caucho, aditivos gasolina	Suppl. 7 (1987)	Leucemia
Alquitrán	Producción que químicos refinados y productos de alquitrán, fundiciones, producción de aluminio	Suppl 7 (1987)	Piel, Pulmón, Vejiga
Aceite mineral	Producción, utilizado como lubricante por los trabajadores del metal, maquinistas, ingenieros, imprenta, utilizado en cosméticos, preparaciones medicinales y farmacéuticas	Suppl. 7 (1987)	Piel, Vejiga, Pulmón, Cavidad nasal
Hollín	Limpiadores de chimeneas, trabajadores de la metalurgia	Vol. 35 (1985)	Piel, Pulmón, Esófago
Monómeros, cloruro de vinilo	Producción de cloruro de polivinilo y polímeros, refrigerante antes del 1974, en los aerosoles	Suppl. 7 (1987)	Riñón
Intermediarios en plásticos e industrias de la goma	Industria del plástico, intercambio de iones y polímeros	Suppl. 7 (1987)	Pulmón
Aminas Aromáticas	Industrias de pigmentos, tintes y colorantes	Suppl. 7 (1987)	Vejiga
Pesticidas	Industria química, agentes esterilizantes, uso de clorofenoles y clorofenoxiles, herbicidas, producción de PCB, producción del papel	Vol. 69 (1997a)	Pulmón, Linfoma no Hodgkin

Tabla 6.- Lista publicada por el IARC de los compuestos cancerígenos para el ser humano (Fuente: IARC, 1987).

1.2.1.-Contaminación atmosférica

La contaminación atmosférica, es decir, la presencia de una o más sustancias en el aire, tiene efectos negativos en humanos, animales, plantas y en la calidad del aire. Existen dos fuentes de contaminación atmosférica; la antropogénica y la natural, de las cuales la antropogénica es la que más contribuye. La contaminación atmosférica es una mezcla de sustancias: compuestos orgánicos volátiles (VOC), óxidos de nitrógeno (NO_x), ozono troposférico (O₃), dióxido de azufre (SO₂), monóxido de carbono (CO), partículas (PM) y otros contaminantes (Campbell et al, 2005). En la Tabla 7 se describe el origen y composición de los contaminantes atmosféricos más nocivos para la salud.

Contaminantes Atmosféricos	Composición	Origen
VOC: Compuestos orgánicos volátiles	Gran variedad de contaminantes diferentes; carbohidratos, componentes orgánicos y disolventes.	Derivan de depósitos de petróleo y gasolina, procesos industriales y combustión de fuel, uso de pinturas y detergentes o actividades de agricultura.
NOX: óxidos de nitrógeno		Emitidos por el tráfico, las instalaciones de combustión, tales como paltes de producción de energía, y las industrias.
SO2: dióxido de azufre		Deriva de la combustión de combustibles fósiles que contienen azufre, tales como petróleo crudo o carbón.
CO: monóxido de carbono		Producido por combustión incompleta de combustibles.
PM:partículas de polvo	Compuestos orgánicos y minerales. Las partículas se clasifican de acuerdo con el tamaño de la partícula (± 2.5 micrómetros).	Fuentes naturales, tales como volcanes, o actividades humanas, tales como procesos de combustión industrial o tráfico.

Tabla 7.- Composición y origen de los diferentes contaminantes atmosféricos (Fuente: elaboración propia).

Existe un número de posibles mecanismos por el cual la contaminación atmosférica puede influenciar en el desarrollo del cáncer. Las partículas (PM) y el ozono pueden dañar el ADN mediante la formación de oxígenos reactivos y provocar cáncer de pulmón. Exposición a PM puede incluso aportar un efecto carcinógeno por su papel de portador de otros contaminantes (metales, PAH, etc.) (Brown et al, 2000; Bostrom et al, 2002).

Antes de 1990 los estudios que se realizaban, con el fin de estudiar la relación entre contaminación atmosférica y cáncer, se basaban en aproximaciones ecológicas (Whitrow et al, 2003). Actualmente, en cambio, se estudia tanto a nivel ecológico como a nivel individual. En todos ellos, el único contaminante atmosférico que se ha encontrado relacionado con el cáncer, ha sido el PM2.5 (Pershagen y Simonato, 1990; Samet y Cohen, 1999). El cáncer de pulmón es el único cáncer con el cual se relaciona claramente con este contaminante (Harrison et al, 2004).

Existen pocos estudios donde se han analizado minuciosamente la relación entre contaminación atmosférica y otro tipo de cáncer que no sea el de pulmón. En general, los datos son inconsistentes para otros cánceres.

No existe ningún estudio que relacione claramente los efectos del SO₂ y el NO₂ en humanos o en animales de laboratorio (ATSDR, 1999).

Ningún estudio del IARC (Internacional Agency for Research on Cancer) ni de la EPA (Environmental Protection Agency) clasifica a los contaminantes atmosféricos como potenciales carcinógenos, existiendo poca información sobre el posible mecanismo carcinogénico.

Por último, el carácter agregado de las medidas de incidencia y de exposición, así como la falta de control de posibles confusores, limitan mucho la inferencia que pueda ser derivada de la relación entre contaminantes atmosféricos y el cáncer.

1.2.2.-Hidrocarburos Policíclicos Aromáticos (PAH)

Los hidrocarburos policíclicos aromáticos (PAH) representan un amplio rango de componentes formados por dos o más anillos de benceno. PAH son generalmente derivados por una combustión incompleta de materiales orgánicos como el carbón,

aceite, gasolina, diesel y el tabaco, que se liberan a la atmósfera (Boffetta et al, 1997; Bostrom et al, 2002).

La exposición en humanos es generalmente ocupacional, en particular, a partir de la inhalación, ingestión, y contacto en la piel. Por lo que se refiere a otras fuentes no ocupacionales de los PAH son las comidas, el agua, el humo del tabaco, el humo de las cocinas y la contaminación atmosférica (Kyrtopoulos et al, 2001; WHO, 2004).

De todas estas fuentes (no ocupacionales), las que se han estudiado con mayor relevancia son la dieta, el humo del tabaco y la contaminación atmosférica.

La dieta es la mayor fuente de PAH. Los PAH se producen al quemar la carne u otras comidas. La deposición atmosférica en la cosecha es otra de las fuentes del PAH encontrado en la comida (Bostrom et al, 2002). Un estudio encontró una mayor concentración de 1-hidroxi-pirano en la orina de los bomberos y en aquellas personas que tienen una dieta basada en la carne cocinada a la brasa (Buckley y Liroy, 1992; Kang et al, 1995).

Ciertos estudios han proporcionado evidencias sobre el aumento de niveles de PAH en niños expuestos involuntariamente al humo en el hogar (Crawford et al, 1998; Tang et al, 1999; Tsai et al, 2003).

Otros han investigado la exposición del PAH en la atmósfera. Se han encontrado indicios de un aumento en el nivel de los biomarcadores del PAH (e.j. aductores del ADN o proteínas relacionados con PAH, o 1-hidroxypireno) en zonas más contaminadas (Motykiewicz et al, 1998; Whyatt et al, 1998) comparado con zonas con bajo nivel de contaminación (Autrup et al, 1999; Scherer et al, 2000; Georgiadis et al, 2001).

Boffetta et al. señalan que ya en 1775 se encontró un aumento en el riesgo de cáncer de escroto en deshollinadores (Boffetta et al, 1997) y en trabajadores del textil (las máquinas de hilar se utilizaban con aceite mineral donde PAH es un importante contaminante) (Tolbert PE, 1997).

Muchos PAH, particularmente aquellos con 4 o 7 anillos de benceno, pueden provocar efectos mutagénicos y carcinógenos en roedores (Georgiadis et al, 2001).

Estudios ocupacionales evidencian el efecto carcinógeno de los PAH en los humanos. Se han encontrado casos de cáncer de pulmón en trabajadores expuestos a altas concentraciones de PAH. Trabajadores en hornos de coque forman el grupo de mayor riesgo de padecer cáncer de pulmón. Otras ocupaciones, como trabajadores en el acero y hierro, en la gasificación del carbón, y trabajadores expuestos a la emisión de diesel, también presentan un riesgo mayor de padecer cáncer de pulmón (Boffetta et al, 1997).

Otros estudios ocupacionales también sugieren un aumento del riesgo de cáncer de vejiga por la exposición a PAH. Las personas que trabajan en industrias del aluminio presentan un riesgo mayor de sufrir cáncer de vejiga. También existen indicios de un mayor riesgo de cáncer de piel en aquellas personas expuestas a los PAH (Campbell et al, 2005).

1.2.3.-Metales

Los metales son elementos que pueden ser descritos por sus propiedades físicas y químicas, como la propiedad de conducción del calor y electricidad, por ser maleables y dúctiles, y formar cationes y óxidos básicos (Atkins et al, 1997). Se pueden encontrar solos o formando compuestos, de los cuales muchos se crean de manera natural en el ambiente. El metal se ha utilizado durante miles de años, destacando su alta producción a partir de la industrialización, a mediados del siglo XIX (Jarup, 2003). Los

metales se consideran como contaminantes por su característica persistente en el ambiente. Éstos están clasificados, por el IARC, como carcinógenos del grupo 1 o grupo 2. El arsénico y el antimonio son dos semi-metales que están dentro de estos grupos.

Los metales y sus compuestos son liberados al ambiente a partir de diferentes fuentes industriales (las fuentes más importantes son las plantas de fundición, combustión de fósiles y plantas de carbón), formando parte de la contaminación atmosférica (Jarup, 2003; CDC, 2004). Otras fuentes importantes de contaminación por metales son las industrias de madera, pesticidas, fertilizantes, plantas de incineración, hogares con tuberías de plomo, etc (Frumkin et al, 2001; Denkhaus et al, 2002; Jarup, 2003; CDC 2004). Las fuentes naturales de contaminación son también las aguas superficiales, los suelos y las aguas freáticas (Jarup, 2003; Satarug et al, 2003). Para la mayoría de la población, la exposición ocurre principalmente al inhalar partículas en el ambiente, al beber agua o comer comida contaminada, al tener contacto dermal, o al inhalar humos contaminados (CDC, 2004).

Los carcinógenos más importantes entre los metales y sus compuestos son el arsénico, berilio, cadmio, cromo y níquel. Evidencia epidemiológica sobre la carcinogenicidad de los metales proviene sobre todo de los estudios ocupacionales. Estudios cohorte de trabajadores expuestos a arsénico, berilio, cadmio, cromo o compuestos de níquel a partir de la inhalación presentaron un aumento del riesgo del cáncer de pulmón (Hayes, 1997; ATSDR, 1999).

El cromo y el níquel, además, fueron asociados con el cáncer nasal. La evidencia de una asociación causal entre cáncer y arsénico es alta (Moore et al, 2002; Browne et al, 2005).

En el caso del cadmio no es tan evidente la relación. El níquel, como sus compuestos, está reconocido como carcinógeno para el ser humano (Barceloux, 1999; Seilkop et al, 2003).

La evidencia de carcinogenicidad de otros metales es limitada. El IARC recientemente reevaluó la carcinogenicidad de los compuestos del plomo, y fueron reclasificados como probables carcinógenos para los humanos (IARC, 2004).

Aunque la evidencia de la carcinogenicidad en humanos es todavía limitada, los cánceres que probablemente estén relacionados con plomo y componentes del plomo son el pulmón, estómago, riñón, y el cerebro (Steenland et al, 2000; IARC, 2004). Un estudio reciente demuestra la carcinogenicidad de los compuestos de cobalto; especialmente cuando se combina con tungsteno, fue reclasificado por el IARC, como probable carcinógeno para el ser humano (para el pulmón) (Lison et al, 2001). El antimonio se ha estudiado menos extensamente, probablemente debido a su uso limitado en la industria (De Boeck, 2003). Con la excepción del arsénico, la relación entre exposición ambiental a metales y el riesgo de cáncer no está bien establecida. Pocos estudios han examinado la exposición ambiental a metales y cáncer en la población general (Cantor, 1997) ya que es difícil cuantificar la dosis expuesta.

1.2.4.-Asbestos

El asbesto es un nombre genérico referido a un grupo de seis minerales fibrosos, hecho de magnesio, silicio y otros elementos que se encuentran en el suelo y rocas (ACS, 2003).

Las fibras tienen propiedades valiosas que incluyen solidez, durabilidad, y resistencia al calor, y como resultado, ha sido utilizado considerablemente por la industria para materiales aislantes y fricción (ACS, 2003). Las fibras de asbestos son derivados de dos amplias categorías mineralógicas. La chrysotile es la forma más utilizada

comercialmente. El asbesto es común en el ambiente: pequeñas concentraciones de sus fibras se propagan en el aire, agua y tierra, a partir de las dos fuentes; natural y antropogénica (ACS, 2003). La mayor ruta de exposición es la inhalación de fibras, el cual forma un polvo que fluye en el aire. El asbesto se puede encontrar también en aguas contaminadas de manera natural o también por fuentes industriales.

A partir de estudios ocupacionales, se ha puesto en evidencia la asociación del asbesto y el aumento del riesgo de mesotelioma (pleural, peritoneo,..) y cáncer de pulmón (ATSDR, 1999). Ha sido confirmado en numerosos estudios experimentales en ratones, ratas y jamsters (EPA, 2005; IARC, 2002). Como resultado, el asbesto está clasificado como carcinógeno humano por varias agencias, incluidas IARC, EPA y el programa de toxicología nacional de estados unidos (IARC, 2002; EPA, 2005). Existe cierta evidencia que el asbesto aumenta el riesgo de otros cánceres, incluyendo aquellos del sistema gastrointestinal y laringe, aunque la evidencia no es tan fuerte como el mesotelioma y el cáncer de pulmón (ATSDR, 1999; Goodman et al, 1999; Browne et al, 2000; Griffiths et al, 2003).

En la actualidad el uso de todas las variedades de amianto está prohibido en España desde el 14 de junio de 2002.

1.2.5.-Radiación

La energía de la radiación, proviene de fuentes naturales como artificiales, en forma de ondas electromagnéticas. El espectro electromagnético se extiende desde radiaciones con frecuencias pequeñas, y grandes longitudes de onda, hasta frecuencias altas con pequeñas longitudes de onda.

- La radiación no ionizante

Los campos electromagnéticos (EMF), infrarrojos, visibles, el radón y la radiación ultravioleta (UVR) poseen el nombre de radiación no ionizante por emitir en frecuencias bajas y grandes longitudes de onda. La radiación ionizante, por el contrario, es de alta frecuencia. Además, en relación a la radiación no ionizante, ha sido más estudiada, teniendo un efecto carcinógeno es más fácil de apreciar y, por tanto, ha sido más controlada.

Las fuentes más importantes de radiación no ionizante son las generadas en las plantas de energía eléctrica. Por otra parte, la exposición de la radiación en hogares es debida a líneas de alta tensión, cableado y aparatos eléctricos. De hecho, cualquier sistema de cargas eléctricas produce un campo y su fuerza disminuye a medida que se aleja de la fuente (radiaciones llamadas ELF, Extra Low Frequency).

Los campos magnéticos y los campos eléctricos (llamadas radiaciones electromagnéticas EMF) se han asociado con el cáncer en diversos estudios epidemiológicos (Miller et al, 1996; Villeneuve et al, 2000; Ahlborn et al, 2001; Poole et al, 2006; Rösli y Künsli, 2006).

Fuentes de campos EMF incluyen algunos equipamientos industriales y médicos, monitores de los ordenadores, televisores, retransmisiones de radio y televisión, y radar. Los móviles son actualmente grandes fuentes de exposición para el público en general; debido a su extenso uso, cualquier pequeño aumento en el riesgo de tumores en áreas expuestas (cerebro y otros tejidos de la cabeza y el cuello) podría tener grandes repercusiones en la salud pública.

En la Tabla 8 se muestran las referencias de la literatura sobre radiación ELF, EMF y cáncer (Ahlborn et al, 2001; IARC, 2002).

Existe sin embargo cierta dificultad para evaluar la exposición a ELF y EMF. En concreto, gran variedad de fuentes de exposición, variedad de diferentes campos electromagnéticos, incertidumbre sobre el período de exposición y las dificultades en medir la exposición pasada (Ahlborn et al, 2001).

Agencia	Conclusiones
<i>World Health Organization (WHO) (Organización mundial de la salud)</i>	
*IARC. Agencia internacional para la investigación del cáncer, 2002	-Leucemia juvenil: evidencia limitada -Cánceres juveniles/adultos: evidencia inadecuada -Campos magnéticos ELF: posiblemente carcinogénico para humanos -Campos eléctricos ELF: carcinogeneidad para humanos -Relación causal entre EMF-cáncer: no establecido; evidencia consistente de leucemia juvenil a altas exposiciones.
*Comisión Internacional para la Protección de la Radiación No-Ionizante, Comité en Epidemiología, 2001	-Campos magnéticos ELF: posiblemente carcinogénico para humanos -Campos eléctricos ELF: carcinogeneidad para humanos -Relación causal entre EMF-cáncer: no establecido; evidencia consistente de leucemia juvenil a altas exposiciones.
<i>Nacional Radiological Protection Board (NRPB)</i>	
*Advisory Group on Non-ionising Radiation (AGNIR), 2001	-Leucemia juvenil a altas exposiciones: cierta evidencia. En adultos: no está establecido.
<i>Nacional Institute of Environmental Health Sciences (NIEHS)</i>	
*EMF-RAPID Program, 1999	Campos electromagnéticos ELF: posible carcinógeno humano. Leucemia juvenil, cierta evidencia. Leucemia linfática crónica: cierta evidencia (exposición ocupacional).

Tabla 8.- Cánceres relacionados con los EMF según WHO, NRPB y NIEHS (Fuente: elaboración propia).

La radiación UVR se define con el rango de longitud de onda, definido entre 100 y 400nm. La radiación UVR se divide en tres bandas; UVA, UVB y UVC. La radiación UVA y UVB tienen efectos nocivos en la salud del ser humano; UVB por absorción directa y UVA a partir de la generación de especies reactivas de oxígeno. Pueden, asimismo, activar oncógenos y enzimas implicados en la carcinogénesis. UVA, UVB y UVC son carcinógenos en animales experimentales. La UVR (radiación ultravioleta) ocasiona más casos de cáncer en los humanos que cualquier otra radiación. La agencia internacional para la investigación del cáncer (IARC, 2002) evaluó la radiación solar como carcinogénica para humanos (Grupo 1) mientras que en el Programa Nacional de Toxicología de Estados Unidos (U.S. NTP) lo consideran como carcinógeno humano conocido. Las radiaciones UVA, UVB, UVC y el uso de lámparas UV y camas UV son probablemente carcinogénicos según el IARC (grupo 2); el U.S.NTP clasifica

similarmente los rayos UVA, UVB y UVC como probables carcinógenos humanos, pero lámparas y camas UV los identifica como carcinógenos conocidos (IARC, 2002; NTP, 2002).

La mayoría de enfermedades relacionadas con UVR son cánceres de células escamosas y cánceres de células basales de la piel. Estos cánceres forman el grupo de cáncer de piel no-melanoma. El riesgo de melanoma, aunque sea poco común en el cáncer de piel, está asociado con un exceso de exposición a UVR.

En cuanto a la exposición ocupacional a UVR existen evidencias entre la relación UVR y cáncer de labio y ojos. Estudios recientes, proporcionan evidencia entre melanoma ocular y UVR solar (Vajdic et al, 2002).

Otra fuente de radiación no ionizante es el radón. El radón es un gas radioactivo producido por las rocas de uranio en la tierra, siendo la principal fuente de exposición de la radiación ionizante para la población en general (Redicha et al, 2006).

Es posible apreciar en toda la literatura la relación entre el cáncer y la radiación ionizante (IARC, 2002), a partir del radón. El radón se inhala y se absorbe a través de los pulmones; los efectos nocivos son origen de sus componentes, los cuales son partículas sólidas y pueden depositarse en los tejidos del pulmón. El radón está compuesto por partículas que emiten rayos alfa (2 protones y 2 neutrones). Aunque éstos no penetran a través del tejido, aportan energía suficiente para causar cambios en la estructura del DNA, mutaciones genéticas, y transformación en las células que alcanzan (IARC, 2002).

El radón está reconocido como carcinógeno humano (Grupo 1) (IARC, 2002). El radón está asociado con el cáncer de pulmón en trabajadores de minas de uranio y otros, debido a productos que transportan radón.

- La radiación ionizante

La radiación ionizante es una causa conocida del cáncer y otros efectos adversos. Es uno de los más estudiados en la carcinogénesis humana y puede contar cerca de 3% de todos los cánceres. La radiación ionizante es capaz de eliminar electrones de los átomos y cambiar la estructura molecular de las células. Estos cambios son los que pueden inducir al desarrollo del cáncer, especialmente el cáncer de tiroides.

Existe alto riesgo de cáncer tiroides en aquellas personas que en la época de infancia estuvieron expuestos a la radiación por radioterapia en las zonas de la cabeza y cuello. La exposición a la radiación durante la infancia puede ser peligrosa debido a que las células son más susceptibles a sufrir un aumento del tamaño.

El peligro de la radiación no fue reconocido hasta 1920, por lo que hasta el período, los radiólogos utilizaron los rayos-X intensivamente.

Entre 1920 y 1960, la radioterapia fue utilizada asiduamente para cierto tipo de tratamiento; como el tinea capitis, linfadenopatía cervical, hipertrofia tonsillar y desórdenes de la piel. Estudios de estas poblaciones, así como a pacientes tratados para un cáncer primario, y supervivientes de la bomba Hiroshima como del accidente nuclear de Chernobil (Jackson et al, 2002; Moysich et al, 2002; Dickman et al, 2003; Kesniiniene et al, 2003; Grossman et al, 2003; Mans et al, 2003; Rahu, 2003; Shakhtarin et al, 2003; Takahashi et al, 2003; Verger et al, 2003; Abdulmughni et al, 2004; Farahti et al, 2004; Mahoney et al, 2004; Jereczek-Fossa et al, 2004; Kahn et al, 2004; Tondel et al, 2004; Hakansson et al, 2005; Zablotska et al, 2005; Davis et al, 2006) , han contribuido mucho para conocer la radiación inducida al cáncer de tiroides. Aunque los estudios realizados utilizaron diferentes metodologías y partieron de diferentes poblaciones, todos han demostrado un riesgo significativo mayor del carcinoma tiroideo en la infancia.

1.2.6.-Pesticidas

Los pesticidas son sustancias creados para erradicar o controlar los insectos, malas hierbas, hongos, o cualquier organismo definido como plaga. Su uso es común en la agricultura, horticultura y en el control de los insectos que causan enfermedades en los programas de salud pública (Dich et al, 1997). El uso de insecticidas químicos, como hidrocarburos clorados que incluyen diclorodifeniltricloroetano (DDT), cloranos, y otros, fue muy difundido en el año 1940 para hacer frente a la malaria y otras enfermedades (Dich et al, 1997). En Norte América el uso del DDT entre los años 1940 y 1960 fue frecuentemente utilizado para el control de los bosques, agricultura, y protección de edificios (Calle et al, 2002).

Al mismo tiempo, herbicidas clorofenoxiles como 2-4-Ácido Diclorofenoxiacético (2-4D) y 2,4,5-ácido triclorofenoxiacético (2,4,5-T) fueron utilizados para controlar las malas hierbas en los campos de cosechas. Insecticidas organofosforados, como el malation, fueron introducidos en el año 1950. En el año 1970, la mayoría de los países desarrollados prohibió el uso de DDT y otros compuestos clorados debido a su potencial biocumulador y toxicidad a largo período (Gaffney et al, 2005).

La exposición humana a pesticidas puede ser a partir de la inhalación, ingestión o absorción. Los residuos de pesticidas se pueden encontrar en el aire, suelo, comida fresca y en las superficies de los hogares. La exposición exterior es principalmente de la agricultura, césped, o sprays utilizados para las plagas. La exposición interior, dentro del hogar, suele ser debida a productos utilizados para las plagas o a la ropa de trabajo utilizado en el campo (Quackenboss et al, 2000; Ward et al, 2006). Se han encontrado compuestos de insecticidas de organofosforados entre el polvo, dentro de las granjas (Simcox et al, 1995). Una de las dificultades al estudiar los pesticidas en la población humana es el cálculo de la exposición.

El Programa de Bioensayos del Instituto de Cáncer Nacional de Estados Unidos, y el Programa de Toxicología Nacional de Estados Unidos, han calculado la carcinogeneidad de varios pesticidas. Entre los años 1977 y 1992 estudiaron el efecto de 47 pesticidas en ratas y ratones y encontraron 23 de ellos carcinógenos (Dich et al, 1997). Los tumores más comunes en estos bioensayos fueron tumores del riñón, del sistema gastrointestinal y tiroides. Muchos, pero no todos, de los pesticidas analizados encontraron efectos mutagénicos en ratones y ratas pero ninguno de ellos en los seres humanos.

1.2.7.-Disruptores endocrinos

Los disruptores endocrinos son una extensa raiz de compuestos químicos sintéticos y plantas naturales. Un disruptor endocrino está definido como una sustancia exógena o una mezcla que altera función(es) del sistema endocrino y por consiguiente, causa efectos adversos en la salud (Comisión Europea, 1996).

Debido a su similitud química a hormonas naturales, los disruptores endocrinos alteran la funcionalidad endocrina de diferentes maneras: Imitando los efectos de las hormonas naturales, bloqueando la vinculación natural de las hormonas y receptores, y evitando la transmisión de los señales entre células, alterando la producción de hormonas naturales, o modificando la producción y la función de los receptores hormonales. Los efectos de los disruptores endocrinos en la salud son dependientes del ritmo de vida, duración y dosis de exposición, características químicas de los compuestos, y numerosas interacciones entre factores endógenos y exógenos (IPCS, 2002). Los efectos adversos en la salud de los disruptores endocrinos en el ambiente son un debate continuo en el mundo científico y político (Stone, 1994; Kavlock et al, 1996; Sharpe, 1999; Daston et al, 2003).

En ciertas partes de Norte América y Europa, ciertas neoplasias de tipo endocrino (mama, útero, próstata y testículos) han aumentado, argumentando la causa al incremento a la exposición a disruptores endocrinos en el ambiente (Davis et al, 1993; Felsot, 1994).

En la literatura, existen más de 30 estudios caso-control desde 1980 que investigan la relación entre DDT y PCB y el riesgo de cáncer de mama (IPCS, 2002; Tsuda et al, 2003; Safe y Zacharewski, 1997). En 1993, Wolf et al (Wolff et al, 1993) encontraron mayores riesgos de cáncer de mama en mujeres con mayor concentración de diclorodifeniletano (DDE) y niveles de PCB en la sangre. Otros estudios de Méjico, Colombia (Olaya-Contreras et al, 1998) y Canadá (Campbell et al, 2005) también confirman la asociación.

En general, la exposición a disruptores endocrinos y el riesgo de cáncer no ha sido estudiada en profundidad. La relación entre cáncer de mama y exposición a DDT y PCB es una excepción, pero incluso aquí, es necesario más investigación.

En resumen se muestra en la Tabla 9 los factores de riesgo ambientales explicados en este apartado.

Evidencia del riesgo de cáncer en el ser humano	Exposición	Comentarios
La evidencia sostiene una asociación de un cáncer, como mínimo	Contaminación atmosférica	Existe cierta evidencia de una relación entre cáncer de pulmón y contaminación atmosférica. (Los datos que existen son insuficientes para valorar la relación con otro tipo de cáncer)
	Arsénico en el agua sanitaria	Altas concentraciones de arsénico aumenta el riesgo de cánceres de piel, pulmón y vejiga, aunque la relación entre dosis y riesgo no está definido.
	Asbestos	La relación con mesotelioma, originalmente establecido a través de estudios ocupacionales, ha sido confirmado en estudios ambientales donde existe un alto riesgo.
	Campos electromagnéticos de baja frecuencia	Existe evidencia de una asociación entre leucemia juvenil y campos de alta intensidad. (La evidencia en los adultos no es clara)
	Radiación solar	Exposición a UVR aumenta al cáncer de célula escamosa, exposición interrumpido provoca carcinoma de célula basal y melanoma.
	Radón	Evidencia clara entre exposición al radón y cáncer de pulmón.
Existe evidencia inconsistente de una asociación	Disruptores endocrinos, ciertos pesticidas	Aunque existe cierta evidencia de una asociación entre exposición ambiental a TDCC y PCB y cáncer de mama, todos los estudios han sido inconsistentes.
Los datos son insuficientes para evaluar la asociación	Metales, PAH	La relación con cáncer para ciertos metales y sus compuestos, y para PAH, ha sido investigado en diferentes estudios ocupacionales. Es difícil cuantificar la dosis de exposición por lo que todos estos estudios son inconsistentes.
	Campos de Radiofrecuencias	A pesar de que ciertos estudios hayan encontrado aumento del riesgo de cáncer de sistema nervioso, leucemia y linfomas en los grupos ocupacionales, no existe evidencia de la ausencia y presencia del riesgo. Los estudios sobre teléfonos móviles todavía son inconsistentes.

Tabla 9.– Resumen de la exposición ambiental y el riesgo de cáncer (Fuente: Campbell et al, 2005; elaboración propia).

1.3.–Neoplasias de interés ambiental

Este trabajo se centra en aquellas neoplasias para las que la literatura sugiere mayor evidencia de una posible relación con factores ambientales y aquellos que tienen cierto interés en la salud pública. Además existen ciertas razones de oportunidad para escoger estas neoplasias. En particular, aquellas en las que la incidencia en la Regió Sanitària Girona ocupa la primera posición, aquellas en las que la incidencia o mortalidad ha aumentado en los últimos años de forma sustancial o aquellas otras en las que se encontró cierta conducta sistemática (es decir no aleatoria) en los casos observados incidentes o muertes por lo que se refiere a su patrón geográfico. Así se considerarán (a partir de ahora neoplasias de interés medioambiental):

- 1.3.1.– Neoplasias Linfoides (CIE-10: C81–C85; C90–C91; C96)
- 1.3.2.– Cáncer de tiroides (CIE-10: C73; CIE-9: 193)
- 1.3.3.– Cáncer de vejiga urinaria (CIE-10: C67, CIE-9: 188)
- 1.3.4.– Cáncer de tráquea, bronquios y pulmón (CIE-10: C-34; CIE-9: 162)

1.3.1.-Neoplasias linfoides (CIE-10: C81-C85; C90-C91; C96).

Las neoplasias linfoides abarcan una amplia variedad de neoplasias malignas que tienen en común su origen en las células del sistema inmune linfoide, es decir, en los tejidos del organismo que albergan los linfocitos o sus precursores (tejido linfoide ganglionar) y que, en condiciones normales, defienden contra agentes u organismos ajenos al cuerpo. Entre ellos reflejan diferencias entre sí, según sea su origen celular en el sistema inmunológico. Esos rasgos de distinción se han hecho más patentes en la actualidad, al conocer que también tienen diferencias biológicas y genéticas.

En base al origen celular, las neoplasias linfoides según la OMS, se han agrupado de manera clásica en tres categorías (Harris et al, 1999; Jaffe et al, 2001; Marcos-Gragera, 2006):

-Linfomas no Hodgkin (CIE-10: C82-C85; C96, CIE-9: 201)

-Linfomas (o enfermedad) de Hodgkin (CIE-10: C81, CIE-9: 202)

-Mielomas (CIE-10: C90, CIE-9: 203.0)

Para subclasificar a los linfomas no Hodgkin (LNH) se hace referencia a linfomas B o linfomas T (linfocitos B o T). Los linfomas Hodgkin (LH) se caracterizan por la presencia en el tejido tumoral de la célula Reed-Sternberg. El mieloma es una proliferación maligna de células plasmáticas, afectando, sobre todo, a individuos de edad avanzada (75 o más años).

Los linfomas hodgkinianos (LH) son la variedad de linfomas más frecuentes en la edad pediátrica. Constituyen el tercer tipo más frecuente de cáncer infantil (edad comprendido entre 0 y 14 años), después de las leucemias agudas y los tumores del sistema nervioso central (SNC).

Las neoplasias linfoides representan aproximadamente el 4% de todos los tumores malignos del ser humano, aunque con una gran variación de unas regiones a otras en el mundo, pero con el dato común de un aumento progresivo, en especial en las zonas más desarrolladas (Bertollini et al, 1992). El LNH presenta mayores tasas en Norte América, Europa (sobre todo, países nórdicos) y en África; y menores tasas en Asia y en Sudamérica (Bertollini et al, 1992). El linfoma de Burkitt (tipo de LNH) es frecuente en África tropical y en Nueva Guinea, y bastante raro en otros sitios, existiendo un consenso sobre el papel etiológico del virus Epstein-Barr y de cierta interacción con la infección por malaria (Bertollini et al, 1992). El LH es más común en Norte América y en el norte de Italia y muy raro en Japón y en China (Bertollini et al, 1992).

La mortalidad en Cataluña por Linfomas en el periodo 1975–2000 es diferente según sea la etiología (ver Tabla 10 y Figura 9, 10 y 11). El LH es el único (ver Figura 9) linfoma que presenta un descenso en dicho período. El Linfoma no Hodgkin y el Mieloma presentan un aumento más o menos constante durante todo el período, siendo más acentuado el aumento en LNH en el último período (Fernandez et al, 2001).

Años	Linfoma de Hodgkin		Linfoma no hodgkiniano		Mielomas	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
1975-76	1,46	0,53	1,76	0,99	1,02	0,55
1977-78	1,07	0,68	1,86	1,02	0,86	0,84
1979-80	1,02	0,64	2,28	1,42	1,43	0,93
1981-82	0,72	0,53	2,5	1,26	1,48	1,26
1983-84	1,2	0,63	2,59	1,5	1,55	1,13
1985-86	0,88	0,57	3,03	2,01	1,68	1,34
1987-88	0,81	0,38	3,29	2,32	1,76	1,34
1989-90	0,81	0,49	3,89	2,66	2,04	1,14
1991-92	0,68	0,33	4,15	2,31	2,11	1,38
1993-94	0,51	0,28	3,9	2,88	1,75	1,37
1995-96	0,69	0,4	3,62	2,19	2,11	1,73
1996-2000	0,92	0,48	5,58	3,85	3,23	2,51

Tabla 10.– Mortalidad por cáncer en Cataluña en hombres y mujeres de neoplasias linfoides. Tasas estandarizadas por edad (población mundial) para todas las edades, 1975–2000 (Fuente: Borràs et al, 2001; Fernández et al, 2001).

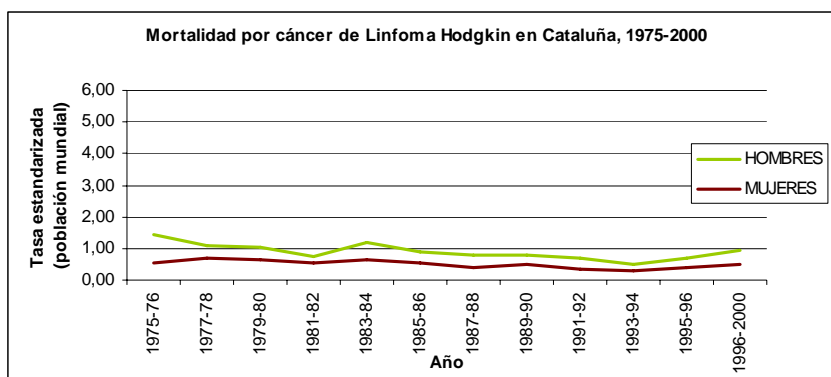


Figura 9.– Mortalidad por cáncer en Cataluña en hombres y mujeres de linfoma Hodgkin. Tasas estandarizadas (población mundial) para todas las edades, 1975–2000 (Fuente: Borràs et al, 2001; Fernández et al, 2001)

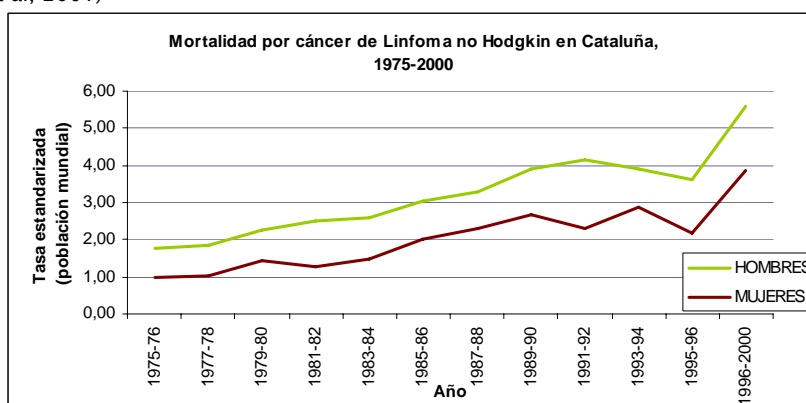


Figura 10.– Mortalidad por cáncer en Cataluña en hombres y mujeres de linfoma no Hodgkin. Tasas estandarizadas (población mundial) para todas las edades, 1975–2000 (Fuente: Borràs et al, 2001; Fernández et al, 2001)

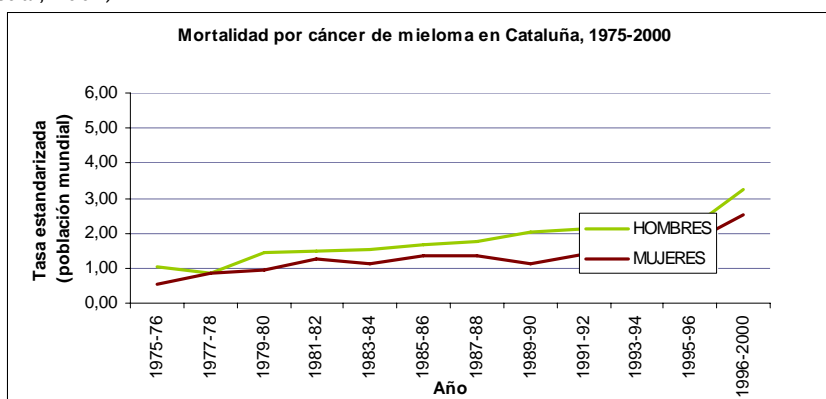


Figura 11.– Mortalidad por cáncer en Cataluña en hombres y mujeres de mieloma. Tasas estandarizadas (población mundial) para todas las edades, 1975–2000 (Fuente: Borràs et al, 2001; Fernández et al, 2001)

La **Tabla 11** hace un resumen de todos los factores de riesgo relacionados con las neoplasias linfoides encontrados en la literatura.

Factores de riesgo	Tipo de neoplasia linfoide	Bibliografía
Nivel socioeconómico	Linfoma No Hodgkin	Scherr y Muller, 1996; Harras et al 1996; Krol et al, 2003
Familiar (genético)	Linfoma no Hodgkin	Linnet y Pottern, 1992; Zhu et al 1998; Potzsch, 1999; Goldin et al, 2004
Déficit inmunológico	Linfoma no Hodgkin	Mueller N, 1999; Knowles DM, 1999; Leval y Harris, 2003; Chang et al 2004a; Domenech et al, 2004; Engels et al, 2004a; Kogevinas et al, 2004
Virus	Linfoma no Hodgkin Enfermedad de Hodgkin	Mueller N, 1991 ; Blattner WA, 1999 ; Mueller N, 1999; Liu et al, 2003; Chang et al, 2004a; Chang et al, 2004b; Engels et al, 2004a; Engels et al, 2004b ; Morgensztern et al, 2004 ; Negri et al, 2004; Ribera y Navarro, 2004; Rodríguez et al, 2004; Duberg et al, 2005 ; Fritschi et al, 2005 ; Svec et al, 2005
Vacuna	Linfoma no Hodgkin	Mekori, 1985; Henry-Amar M, 1992; Holly et al, 1999; Kirk et al, 2001
Bacterias	Linfoma no Hodgkin	Vineis et al, 1999; Morgner et al, 2000 ; Takagi et al, 2000
Medicamentos	Linfoma no Hodgkin	Kaldor et al, 1987; Blomberg et al; 1993 ; Kato et al, 2003
Radiación ionizante	Linfoma no Hodgkin	Burnett et al, 1999
Radiación solar	Linfoma no Hodgkin Enfermedad de Hodgkin	Egan et al, 2005; Smedby et al, 2005
Dieta	Linfoma no Hodgkin	Ward et al, 1996; Tavani et al, 1997; Chiu et al, 1999 ; Law et al, 1999; Zhang et al, 1999; Naresh et al, 2004
Estilo de vida	Linfoma no Hodgkin	Linnet y Pottern, 1992; Freedman et al, 1998; Zhu et al, 1998; Vose et al, 2002
Metales (arsénico)	Mieloma	Boffetta y Nyberg, 2003
Compuestos orgánicos persistentes (PCB, DDT, HCB, Dioxinas...)	Linfoma no Hodgkin Enfermedad de Hodgkin Mieloma	Baris et al, 1998; Linnet y Pottern, 1992; Viel et al, 2000; Floret et al, 2003; Hardell y Eriksson, 2003; De Roos et al, 2005

Tabla 11.– Factores de riesgo asociados a diferentes neoplasias linfoides (Fuente: elaboración propia).

1.3.2.-Cáncer de Tiroides (CIE-10: C73; CIE-9:193)

El cáncer de tiroides es el tipo más común de los tumores endocrinos malignos, y en los últimos años es el cáncer que presenta, junto a linfomas no hodgkinianos y de pulmón, un incremento más acusado (Blattner, 1999; Davies y Welch, 2006). Existen diferentes tipos de histología de los cuales los foliculares y los papilares presentan el 90% de todas las neoplasias tiroideas (DeLellis et al, 2004). En mujeres, es la neoplasia que presenta mayor supervivencia mientras que en hombres es la segunda. En la Figura 12, se puede observar como la tasa de mortalidad en España, entre el período 1978 y 2000, de cáncer de tiroides se ha mantenido más o menos constante y a un nivel bajo.

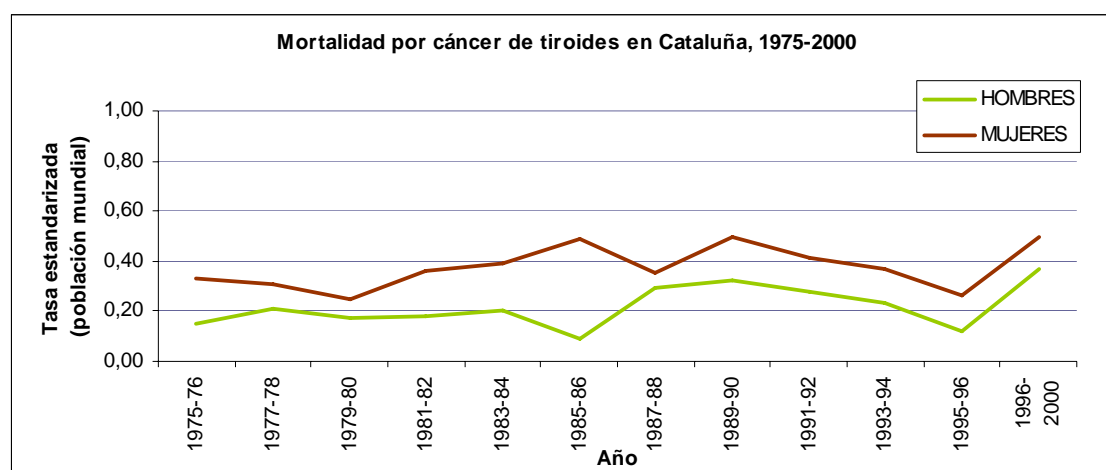


Figura 12.- Mortalidad por cáncer de Tiroides en el período 1975–2000. Tasas estandarizadas (población mundial) para todas las edades, 1975–2000 (Fuente: Borràs et al, 2001; Fernández et al, 2001); EUROCORE (European Cancer Registry-based Study on Survival and Care of Cancer Patient) (Berrino et al., 1995, 1999 y 2003).

El cáncer de tiroides, en España, es 2,6 veces más común en mujeres que en hombres y se suele diagnosticar a una edad comprendida entre 50–65 años (ver Tabla 12) (Lope et al, 2004).

EDAD (años)	HOMBRES		MUJERES	
	Casos	Tasa	Casos	Tasa
0-4	1	0,11	0	0
5-9	1	0,11	0	0
10-14	9	0,75	7	0,65
15-19	3	0,21	18	1,32
20-24	15	0,89	39	2,42
25-29	28	1,67	57	3,52
30-34	38	2,34	81	5,11
35-39	28	1,87	70	4,69
40-44	37	2,79	74	5,55
45-49	50	4,14	78	6,39
50-54	55	4,84	151	12,89
55-59	36	3,77	123	12,24
60-64	50	5,18	138	12,9
65-69	32	3,39	125	11,43
70-74	18	2,36	106	10,97
75 y más	10	1,01	116	6,84

Tabla 12.– Incidencia estimada de cáncer de tiroides en España en 1998 por grupos de edad. Número de casos y tasas específicas por 100.000 personas-año. (Fuente: Lope et al, 2004)

La Tabla 12 representa la tasa de incidencia del cáncer de tiroides, de diferentes poblaciones españolas, ajustada por la población europea. En el caso de la Regió Sanitària Girona, la tasa de incidencia (período 1994–2001), es de 1,5 en los hombres (1,5 casos por 100.000 habitantes/año) y 4,6 en las mujeres (4,6 casos por 100.000 habitantes/año). Navarra (7,9 casos en mujeres por 100.000 habitantes/año y 2.6 casos por 100.000 habitantes/año) es la provincia con mayor tasa de incidencia presentó entre el período 1993 y 1997 (Lope et al, 2004).

	Hombres	Mujeres
Albacete (1993-1997)	1.2	4.2
Asturias (1993-1997)	1.8	4.7
Islas Canarias (1993-1997)	1.5	5.7
Cuenca (1993-1997)	1.4	5.2
Girona (1994-2001)	1.5	4.6
Granada (1993-1997)	0.8	6.2
Mallorca (1993-1997)	1.1	3.8
Murcia (1993-1997)	1.5	5.2
Navarra (1993-1997)	2.6	7.9
Tarragona (1993-1997)	1.1	4.8
Zaragoza (1993-1997)	0.9	3.3

Tabla 13.– Tasas de incidencia de cáncer de tiroides ajustadas con la población europea en los diferentes registros poblacionales españoles (1991–1997) (Fuente: Fernández et al, 2001)

En la literatura, los factores de riesgo asociados al cáncer de tiroides son diversos (foliculares y papilares, en concreto) de los cuales los más importantes son; factores ambientales, factores dietéticos, factores genéticos y factores ocupacionales (ver Tabla 14).

Factores de riesgo	Bibliografía
Radiación Ionizante	Radiación Terapéutica Lundell et al, 1994; Nagataki y Nystrom, 2002; Dickman et al, 2003; Shakhhtarín et al, 2003 Jereczek-Fossa et al, 2004; Schafer et al, 2004.
	Radiación Ambiental Jackson et al, 2002; Moysich et al, 2002; Nagataki et al, 2002; Kesniiniene et al, 2003; Mans et al, 2003; Rahu M, 2003; Tukiendorf et al, 2003; Larijani et al, 2004; Mahoney et al, 2004; Stezhko et al, 2004; Tondel et al, 2004; Hakansson et al, 2005.
Dieta	Alimentos bociógenos Preston-Martin et al, 1993; Markaki et al, 2003; (deficientes de yodo o Lope et al, 2004. exceso de yodo)
Hereditario	Bocio endémico Lope et al, 2004
COP (Compuestos Orgánicos Persistentes)	DDT, PCB, HCB entre Grimalt et al, 1994; Langer P, 1998; Dickman et al, 2003
Actividad volcánica	Lava basáltica Kung et al, 1981; Arnbjörnsson y Olafsson, 1986.
Anticonceptivo oral	Malkesian y Mayberry, 1970; Chan et al, 1975; Weeke y Hansen, 1975
Menarquía tardía	Truong et al, 2005

Tabla 14.- Factores de riesgo asociados al cáncer de tiroides (Fuente: elaboración propia).

1.3.3.-Cáncer de Vejiga Urinaria (CIE-10 C67; CIE-9: 188).

El cáncer de vejiga urinaria es un tumor frecuente en los hombres, con una incidencia superior en países con un mayor nivel de desarrollo económico e industrial (López-Abente, 1996; Serra, 2002). En las últimas décadas, algunas áreas de Europa Occidental han mostrado tasas de incidencia entre los más elevados del mundo (Parkin et al, 1992).

En 1990, España, junto a Dinamarca y Polonia era el país con mayor tasa de cáncer de vejiga urinaria. (Vineis et al, 1999). Actualmente, España también ocupa un lugar

destacado en Europa, respecto a la incidencia (primer país) y mortalidad (segundo país). En España durante 2000 se produjeron 3.198 muertes por cáncer de vejiga en hombres y 720 en mujeres, siendo la tasa de mortalidad ajustada por la población europea de 13,4 en hombres y de 1,9 en mujeres por 100.000 habitantes (Lopez-Abente et al, 2005).

Tal y como se puede apreciar en la Figura 13, el cáncer de vejiga urinaria en Cataluña presenta un crecimiento paulatino en los hombres (excepto a partir de 1995) mientras que en las mujeres más o menos se ha mantenido constante.

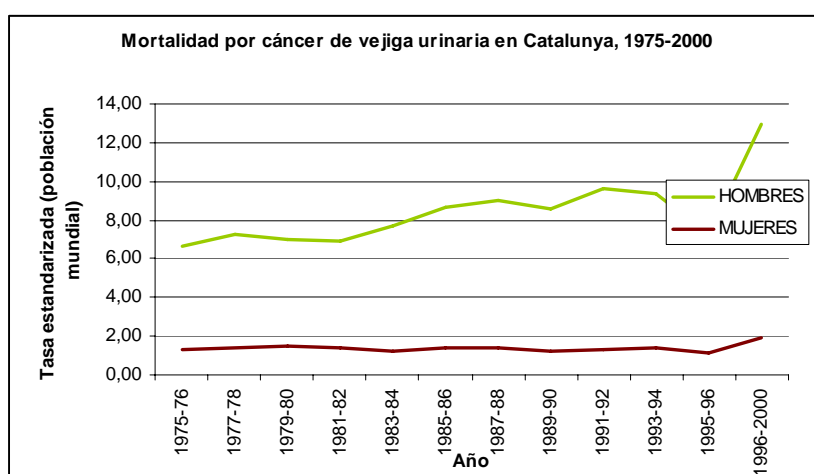


Figura 13.- Mortalidad por càncer de vejiga urinaria, perído 1975–2000, en Catalunya. Tasa estandarizada por edad (Población mundial) (Fuente: Fernandez et al, 2001); ; EUROCARE (European Cancer Registry-based Study on Survival and Care of Cancer Patient) (Berrino et al., 1995, 1999 y 2003).

Al contrario que ocurría con el cáncer de tiroides, en la Regió Sanitària Girona para el periodo 1994–2001, la incidencia de cáncer de vejiga urinaria fue más elevada en hombres (37,4 casos por 100.000 habitantes/año) que en mujeres (6,8 casos por 100.000 habitantes/año).

Los factores de riesgo relacionados con el cáncer de vejiga urinaria se pueden observar en la Tabla 15.

Factores de riesgo	Bibliografía
Tabaco	IARC, 2002
Exposiciones ocupacionales y químicas; sobre todo industria del textil	IARC, 1987; Romundstad et al, 2000; Urrutia et al, 2002; Czene et al, 2003; Rafnsson y Sulem, 2003; Bates et al, 2004; Mikoczy et al, 2004; Zeka et al, 2004
Exposición a aminas aromáticas	Silverman et al, 1996; Fortuny et al, 1999; Manette et al, 1999; Brennan et al, 2000; Serra et al, 2000
Café	Pujolar et al, 1993
Dieta	Riboli y Kaaks, 1997
Cloración de aguas	Villanueva et al, 2001
Infección por agentes biológicos	EI-Bolkainy y Chu, 1981

Tabla 15.- Factores de riesgo asociados al cáncer de vejiga urinaria (Fuente: elaboración propia).

1.3.4.-Cáncer de tráquea, bronquios y pulmón (CIE-10 C-34; CIE-9:162).

El cáncer de pulmón es el tumor más importante en cuanto a mortalidad en el mundo occidental. En España supone unos 18.500 casos nuevos al año y ha sido responsable de 17.308 muertes en el año 2000 (Lopez-Abente et al, 2003). El cáncer (carcinoma) de pulmón abarca una variedad de formaciones malignas (crecimiento de células sin control) que tienen en común su origen en alguno de los tejidos del organismo que componen el tracto respiratorio pulmonar; es decir, el formado por los pulmones y los bronquios (árbol traqueobronquial).

En Cataluña, entre el período comprendido 1975 y 2000 (ver Figura 14), la tasa de incidencia en el caso de los hombre presenta un crecimiento significativo mientras que en el caso de las mujeres un descenso (Fernandez et al, 2001).

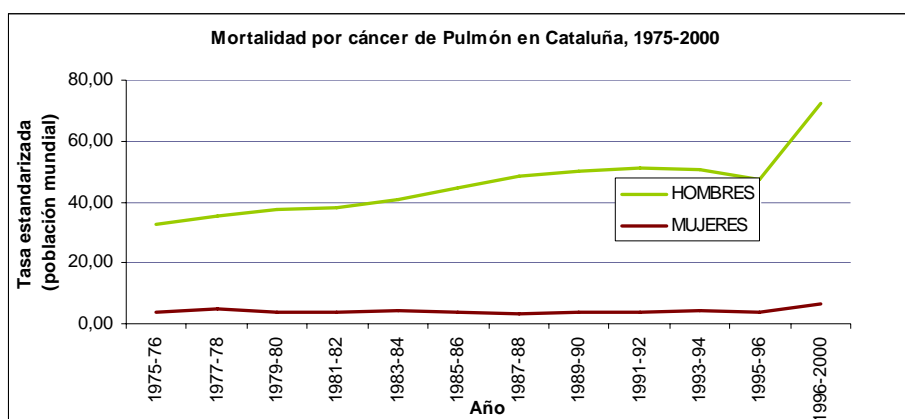


Figura 14.- Mortalidad por cáncer de tráquea, bronquios y pulmón, período 1975-2000, en Cataluña. Tasa estandarizada por edad (Población mundial). (Fuente: Fernandez et al, 2001); EURO CARE (European Cancer Registry-based Study on Survival and Care of Cancer Patient) (Berrino et al., 1995, 1999 y 2003).

Por lo que se refiere a la Regió Sanitària Girona, entre 1994 y 2001, Tabla 5, como era de esperar por su relación al hábito tabáquico (la prevalencia de hábito tabáquico, según ESCA del 2002, en Girona, en hombres fue de 39,1% y en mujeres 22,5%, y en Catalunya, en hombres fue de 37,6% y en mujeres 26,4%), el cáncer de pulmón es mucho más incidente entre los hombres, ocupando el segundo puesto respecto a la incidencia (74,3 casos por 100.000 habitantes/año), que entre las mujeres (7,7 casos por 100.000 habitantes/año). Aunque, cabe decir, que en éstas, ocupa la onceava posición.

La causa más frecuente para el desarrollo del cáncer de pulmón está relacionada en el 90% de los sujetos con el consumo directo de tabaco (fumador activo) o exposición al humo del tabaco (fumador pasivo). El riesgo de desarrollar el tumor se va incrementando a medida que aumenta el número de cigarrillos que se consumen al día, así como el tiempo (años) de exposición a ese hábito.

Existen otros factores relacionados con el desarrollo del cáncer de pulmón (ver Tabla 16), como son la exposición ambiental y ocupacional a compuestos tóxicos.

	Factores de riesgo	Bibliografía
Exposición ambiental	Compuestos orgánicos persistentes (PAH, pesticidas, compuestos orgánicos volátiles, etc) y PM2.5	IARC 1987; Bostrom et al, 2002 ; Chellini et al, 2002; Blais et al, 2003; Boffetta y Nyberg, 2003 ; Sorensen et al, 2003 ; Alavanja et al, 2004 ; Armstrong et al, 2004; Forastiere F, 2004; Harrison et al, 2004
Exposición ocupacional	Metal : cadmio, plutonio, compuesto de poliuretano, amianto	IARC , 1987; Sorahan et al, 2000; Blais et al, 2003; Menvielle et al, 2003; Zeka et al, 2003; Mikoczy et al, 2004; Moshammer y Neuberger, 2004
Radiación	Radiación ocupacional	Darbi et al, 2005 ; Richardson y Ashmore, 2005 ; Wing y Richardson, 2005

Tabla 16.- Factores de riesgo asociados al cáncer de pulmón, a parte del tabaco (Fuente: elaboración propia).

Referencias

Abdulmughni YA, Al-Hureibi MA, Al-Hureibi KA, Ghafoor MA, Al-Wadan Ah, Al-Hureibi YA. Thyroid cancer in Yemen. *Saudi Medicine Journal* 2004; 25: 55–59.

Acquavella JF, Delzell E, Cheng H, Lynch CF, Johnson G. Mortality and cancer incidence among alachlor manufacturing workers 1968–99. *Occupational Environmental Medicine* 2004; 61: 680–685.

Acquavella American Cancer Society (ACS). Asbestos, 2003.
http://www.cancer.org/docroot/PED/content/PED_1_3X_Asbestos.asp?sitearea=WHO [último acceso feb 05]

Agency for Toxic Substances and Disease Registry (ATSDR). Toxfaqs for Sulfur Dioxide. 1999. <http://www.atsdr.cdc.gov/tfacts116.html> [último acceso 23 Marzo 2005]

Ahlbom A, Cardis E, Green A, Linet M, Savitz D, Swerdlow A, et al. Review of the epidemiologic literature on EMF and Health. *Environmental Health Perspectives* 2001;109 Suppl 6:911–33.

Alavanja MCR, Dosemeci M, Samanic C, Lubin J, Lynch C, Knott C, Barker J, Hoppin JA, Sandler DP, Coble J, Thomas K, Blair A. Pesticides and Lung Cancer Risk in the Agricultural Health Study Cohort. *American Journal of Epidemiology* 2004; 160: 876–885.

Armstrong B, Hutchinson E, Unwing J, Fletcher T. Lung Cancer Risk after Exposure to Polycyclic Aromatic Hydrocarbons: A Review and Meta-Analysis. *Environmental Health Perspectives* 2004; 112: 970–975.

Arnbjörnsson E, Arnbjörnsson A, Olafsson A. Thyroid cancer incidence in relation to volcanic activity. *Archives of Environmental Health* 1986, 61: 267–269.

Atkins P, Lo Jones B. Chemistry – Molecules, Matter and Change. 3rd ed. New York: W.H. Freeman; 1997.

Autrup H, Daneshvar B, Dragsted LO, Gamborg M, Hansen M, Loft S, et al. Biomarkers for exposure to ambient air pollution—comparison of carcinogen-DNA adduct levels with other exposure markers and markers for oxidative stress. *Environmental Health Perspectives* 1999;107:233–8.

Bahia SHA, Echenique I, Koifman S. Cancer and wood-related occupational exposure in the Amazon region of Brazil. *Environmental Research* 2005; 99 (1): 132–140.

Barceloux DG. Nickel. *Journal Toxicology Clinical Toxicol* 1999; 37:239–58.

Baris D, Zahm SH, Cantor KP, Blair A. Agricultural use of DDT and risk of non-Hodgkin's lymphoma: pooled analysis of three case-control studies in the United States. *Occupational Environment Medicine* 1998; 55: 522-527.

Bates MN, Rey OA, Biggs ML, Hopenhayn C, Moore LE, Kalman D, Steinmaus C, Smith AH. Case-Control Study of Bladder Cancer and Exposure to Arsenic in Argentina. *American Journal Epidemiology* 2004; 159: 381-389.

Berrino, F., Sant, M. y cols. (Eds.). *Survival of cancer patients in Europe. The EURO CARE study*. IARC Sci Publ 1995; 132: 1-463.

Berrino, F., Capocaccia, R. y cols. (Eds.). *Survival of cancer patients in Europe. The EURO CARE-2 study*. IARC Sci Publ 1999; 151: 1-572.

Berrino, F., Capocaccia, R. y cols. (Eds.). *Survival of cancer patients in Europe: the EURO CARE-3 Study*. *Ann Oncol* 2003; 14(Supl. 5): 1-118.

Bertollini R, Douglas M, Saracci R, Savitz DA. *Exposure of Disease, Environmental Epidemiology* 1992, edit Lewis (WHO).

Blais JM, Froese KL, Kimpe LE, Muir DC, Backus S, Comba M, Schindler DW. Assessment and characterization of polychlorinated biphenyls near a hazardous waste incinerator: analysis of vegetation, snow, and sediments. *Environmental Toxicology Chemistry* 2003; 22: 126-133.

Blattner, WA. Human retroviruses: their role in cancer. *Proceeding of the Association of American Assotiations*, 1999; 111: 563-572.

Blomberg J, Moller T, Olsson H, Anderson H, Jonsson M. Cancer morbidity in blood recipients--results of a cohort study. *European Journal of Cancer* 1993; 15: 2101-2105.

Boffetta P, Jourenkova N, Gustavsson P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes Control* 1997;8:444-72.

Boffetta P, Nyberg F. Contribution of environmental factors to cancer risk. *British Medical Bulletin* 2003; 68: 71-94.

Borràs JM, Galceran J, Bosch X. Cancer Registry of Tarragona, Catalonia, Spain. A: Parkin DM, Whelan SL, Ferlay J, Raymond L, Young L (eds). *Cancer Incidence in five continents*, vol VII. IARC Sci Publ No. 143, Lyon: Internacional Agency for Research on Cancer; 1997a.

Borràs JM, Borràs J, Viladiu P, Bosch FX, eds. *Epidemiologia i prevenció del càncer a Catalunya: 1975-1992*. Barcelona: ICO, Generalitat de Catalunya; 1997b.

Borràs JM, Fernández E, Schiaffino A, Borrell C, La Vecchia C. Pattern of smoking initiation in Catalonia (Spain) from 1948 to 1992. *American Journal Public Health* 2000; 90: 1459–62.

Borràs JM, Borràs J, Bosch XF, Fernández E, Galceran J, Gispert R, González JR, Izquierdo A, Marcos R, Moreno V, Peris M, Puig X, Sánchez V, Viladiu P. *Càncer Catalunya 2001*; Granollers; Servei de Prevenció i Control del Càncer, ICO, 2001.

Bostrom CE, Gerde P, Hanberg A et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environmental Health Perspectives* 2002; 110 Suppl 3: 451–388.

Brennan P, Bogillot O, Cordier S et al. Cigarette smoking and bladder cancer in men: a pooled analysis of 11 European case-control studies. *International Journal of Cancer* 2000; 86: 289–294.

Brown DM, Stone V, Findlay P, MacNee W, Donaldson K. Increased inflammation and intracellular calcium caused by ultrafine carbon black is independent of transition metals or other soluble components. *Occupational Environmental Medicine* 2000; 57: 685–91.

Browne K, Gee JB. Asbestos exposure and laryngeal cancer. *British Occupational Hygiene Society* 2000; 44: 239–50.

Browne ML, Varadarajulu D, Lewis-Michl EL, Fitzgerald EF. Cancer incidence and asbestos in drinking water, Town of Woodstock, New York, 1980–1998. *Environmental Research*, 2005; 98(2): 224–32.

Buckley TJ, Liou PJ. An examination of the time course from human dietary exposure to polycyclic aromatic hydrocarbons to urinary elimination of 1-hydroxypyrene. *British Journal Industrial Medicine* 1992; 49: 113–24.

Burnett C, Robinson C, Walker J. Cancer mortality in health and science technicians. *American Journal Ind. Medicine* 1999; 36: 155–158.

Calle EE, Frumkin H, Henley SJ, Savitz DA, Thun MJ. Organochlorines and breast cancer risk. *CA: Cancer Journal for Clinicians* 2002; 52 (5): 301–9.

Campbell P, Fehring G, Halaph E, Sloan P, Theis B, Kreiger N. Cancer Care Ontario: Insight on Cancer. *Environmental Exposures and Cancer*. Toronto: Canadian Cancer Society (Ontario Division), 2005.

Cantor KP. Drinking water and cancer. *Cancer Causes Control* 1997; 8: 292–308.

Centers for Disease Control and Prevention (CDC). Heavy Metals, 2004.
http://www.envirohealthaction.org/upload_files/Heavy_Metals.pdf [último acceso 08 Febrero 2005].

Chan V, Paraskevaides CA, Hale JF. Assessment of thyroid function during pregnancy. *British Journal of Obstetrics and Gynaecology* 1975; 82: 137–141.

Chang KC, Huang GC, Jones D, Tsao CJ, Lee IYY, Su IJ. Distribution and prognosis of WHO Lymphoma Subtypes in Taiwan Reveals a Low Incidence of Germinal–Center Derived Tumors. *Leukaemia and Lymphoma* 2004a; 45: 1375–1384.

Chang ET, Zheng T, Weir EG, Borowitz M, Mann RB, Spiegelman D, Mueller NE. Childhood Social Environment and Hodgkin's lymphoma: New Findings from a Population–Based Case–Control Study. *Cancer Epidemiology, Biomarkers & Prevention* 2004b; 13(8): 1361–1370.

Chellini E, Cherubini M, Chetoni L, Biggeri A, Vannucchi G. Risk of Respiratory cancer around a Sewage Plant in Prato, Italy. *Environmental Health* 2002; 57:548–553.

Chen A, Zhang J, Zhou L, Gao E, Chen L, Rogan WJ, Wolff MS. DDT serum concentration and menstruation among young Chinese women. Article in press, *Environmental Research* 2004.

Chiu BC, Cerhan JR, Gapstur SM, Sellers TA, Zheng W, Lutz CT, Wallace RB, Potter JD. Alcohol consumption and non–Hodgkin lymphoma in a cohort of older women. *British Journal Cancer* 1999; 80: 1476–1482.

Crawford FG, Mayer J, Santella RM, Cooper TB, Ottman R, Tsai WY, et al. Biomarkers of environmental tobacco smoke in preschool children and their mothers. *Journal National Cancer Institute* 1994;86:1398–402.

Czene K, Tiikkaja S, Hemminki K. Cancer risks in hairdressers: assessment of carcinogenicity of hair dyes and gels. *International Journal on Cancer* 2003; 20:108–112.

Darby S, Hill D, Auvinen A, Barros–Dios JM, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case–control studies. *British Medicine Journal* 2005; 63:1–6.

Daston GP, Cook JC, Kavlock RJ. Uncertainties for endocrine disrupters: our view on progress. *Toxicology Sciences* 2003;74: 245–52.

Davies L, Gilbert H. Increasing Incidence of Thyroid Cancer in the United States, 1973–2002. *Journal American Medical Association* 2006, 295(18):2164:2167.

Davis DL, Bradlow HL, Wolff M, Woodruff T, Hoel DG, Anton–Culver H. Medical hypothesis: xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives* 1993;101:372–7.

Davis S, Day RW, Kopecky KJ, Mahoney MC et al. International Consortium For Research On The Health Effects Of Radiation. Childhood leukaemia in Belarus, Russia, and Ukraine following the Chernobyl power station accident: results from an international collaborative population-based case-control study. *International journal of Epidemiology* 2006; 35:386–396.

De Boeck, Kirsch-Volders M, Lison D. Cobalt and antimony: genotoxicity and carcinogenicity. *Mutation Research* 2003;533:135–52.

DeLellis RA, Lloyd RV, Heitz PU, Eng C. Pathology and Genetics of Tumours of Endocrine Organs. *WHO Classification of Tumours*. Edit. IARC, octubre 2004

Denkhaus E, Salnikow K. Nickel essentiality, toxicity, and carcinogenicity. *Critical Reviews in Oncology/Hematology* 2002;42:35–56.

De Roos AJ, Blair A, Rusiecki JA, Hoppin JA, Svec M, Dosemeci M, Sandler DP, Alavanja MC. Cancer Incidence among Glyphosate-Exposed Pesticide Applicators in the Agricultural Health Study. *Environmental Health Perspectives* 2005; 113: 49–53.

Dich J, Zahm SH, Hanberg A, Adami H-O. Pesticides and cancer. *Cancer Causes Control* 1997; 8:420–43.

Dickman PW, Holm LE, Lundell G, Boice JD Jr, Hall P. Thyroid cancer risk after thyroid examination with 1311: a population-based cohort study in Sweden. *International Journal Cancer* 2003; 106: 580–587.

Domenech E, Kennedy S, Martínez-Maza O, Sanjose S. Occupational Exposure to Immunologically Active Agents and Risk for Lymphoma. *Cancer Epidemiology Biomarkers Prevention* 2004, 13(11): 1814–1818

Duberg AS, Nordström M, Törner A, Reichard O, Strauss R, Janzon R, Bäck E, Ekdahl K. Non-Hodgkin's Lymphoma and Other Nonhepatic Malignancies in Swedish Patients with Hepatitis C Virus Infection. *Hepatology* 2005; 41(3): 652–659.

Egan KM, Sosman JA, Blot WJ. Sunlight and Reduced Risk of Cancer: Is The Real Story Vitamin D. *Journal of the National Cancer Institute* 2005; 97(3): 161–163.

El-Bolkainy MN, Chu EW, eds. *Detection of bladder cancer associated with Schistosomiasis*. The National Cancer Institute. El Cairo, 1981.

Engels EA, Chatieriee N, Cerhan JR, Davis S, Cozen W, Severson RK, Whitby D, Colt JS, Hartge P. Hepatitis C virus infection and Non-hodgkin lymphoma: Results of the NCI-SEER multi-center case-control study. *International Journal of Cancer* 2004a; 111: 76–80.

Engels EA, Chen J, Viscidi RP, Shah KV, Daniel RW, Chaltermjee N, Klebanoff MA. Poliovirus Vaccination during Pregnancy, Maternal Sero-conversion to Simian Virus 40, and Risk of Childhood Cancer. *American Journal Epidemiology* 2004b; 160: 306–316.

Enquesta de Salut de Catalunya (ESCA). Enquesta de salut, 2002. Generalitat de Catalunya. <http://www.gencat.net/salut/depsan/units/sanitat/html/ca/plasalut/index.html>

EPA. Asbestos. <http://www.epa.gov/ttnatw01/hlthef/asbestos.html> [último acceso 08 Febrero 2005].

European Commission. European workshop on the impact of endocrine disruptors on human health and wildlife. Weybridge, UK, Report No. EUR 17549, *Environment and climate research programme*, DG XXI. Brussels, Belgium: European Commission, 1996.

Farahti J, Geling M, Mader U, Morti M, Luster M, Muller JG, Flentje M, Reiners C. Changing trends of incidence and prognosis of thyroid carcinoma in the Franconia, Germany, from 198–1995. *Thyroid* 2004; 14:141–147.

Felsot A. Pesticides, estrogen activity and breast cancer: new cause for regulatory concern? *Agrichem Environ News* 1994;97:8–10.

Fernández E, González JR, Borràs JM, Sánchez V, Moreno V, Peris M. Evolución de la mortalidad por cáncer en Cataluña (1975–1998). *Medicina clínica* 2001;116: 605–609.

Floret N, Mauny F, Challier B, Arveux P, Cahn JY, Viel JF. Dioxin Emissions from a Solid Waste Incinerator and Risk of Non-Hodgkin Lymphoma. *Epidemiology* 2003;14:392–8.

Forastiere F. Fine particles and lung cancer. *Occupational Environmental Medicine* 2004; 61: 797–798.

Forssén UM, Rutqvist LE, Ahlbom A, Feychting M. Occupational Magnetic Fields and Female Breast Cancer: A Case-Control Study using Swedish Population Registers and New Exposure Data. *American Journal of Epidemiology* 2005; 161:250–259.

Fortuny J, Serra C, Bonfill X, Kogevinas M, Sunyer J. Diferencias sociales en el cáncer de vejiga urinaria en Cataluña. *Gaceta Sanitaria* 1999; 13: 208–217.

Freedman DS, Tolbert PE, Coates R, Brann EA., Kjeldsberg CR. Relation of cigarette smoking to non-Hodgkin's lymphoma among middle-aged men. *American Journal Epidemiology* 1998; 148: 833–841.

Fritschi L, Benke G, Hughes AM et al. Occupational Exposure to Pesticides and Risk of Non-Hodgkin's Lymphoma. *American Journal of Epidemiology* 2005, 162: 849–857.

Frumkin H, Thun MJ. Arsenic. *CA Cancer Journal Clinique* 2001;51:254–62.

Gaffney SH, Curriero FC, Strickland PT, Glass GE, Helzlsouer KJ, Breyse PN. Influence of Geographic Location in Modeling Blood Pesticide Levels in a Community Surrounding a U.S. Environmental Protection Agency Superfund Site. *Environmental Health Perspectives* 2005; 113(12):1712–1716.

Georgiadis P, Topinka J, Stoikidou M et al. Biomarkers of genotoxicity of air pollution (the AULIS project): bulky DNA adducts in subjects with moderate to low exposures to airborne polycyclic aromatic hydrocarbons and their relationship to environmental tobacco smoke and other parameters. *Carcinogenesis* 2001;22: 1447–57.

Goldin LR, Pfeiffer RM, Gridley G, Gail MH, Li X, Møller J, Olsen JH, Hemminki K, Linet MS. Familial Aggregation of Hodgkin Lymphoma and Related Tumors. *Cancer* 2004; 100: 1902–1908.

Goodman M, Morgan RW, Ray R, Malloy CD, Zhao K. Cancer in asbestos-exposed occupational cohorts: a meta-analysis. *Cancer Causes Control* 1999;10:453–65.

Griffiths H, Malony NC. Does asbestos cause laryngeal cancer? *Clinical Otolaryngology* 2003;28:177–82.

Grimalt JO, Sunyer J, Moreno V et al. Risk excess of soft-tissue sarcoma and thyroid cancer in a community exposed to airborne organochlorinated compound mixtures with a high hexachlorobenzene content. *Int. Journal on Cancer* 1994;56:200–203.

Grossman CM, Nussbaum RH, Nussbaum FD. Cancers among residents downwind of the Hanford, Washington, plutonium production site. *Arch Environmental Health* 2003; 58: 267–274.

Hakansson N, Stenlund C, Gustavsson P, Johansen C, Floderus B. Arc and resistance welding and tumours of the endocrine glands: a Swedish case-control study with focus on extremely low frequency magnetic fields. *Occupational Environmental Medicine* 2005; 62: 304–308.

Hardell L, Eriksson M. Is the decline of the increasing incidence of non-Hodgkin lymphoma in Sweden and other countries a result of cancer preventive measures?, *Environmental Health Perspectives* 2003; 111: 1704–1706.

Harris A, Edwards BK, Blot WS, Ries LAG. Cancer Rates and Risks. *National Institutes of Health. National Cancer Institute*, Bethesda 1996; 96, 691.

Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting—Airlie House, Virginia, November 1997. *Journal of Clinical Oncology* 1999; 17(12): 3835–3849.

Harrison RM, Smith DJT, Kibble AJ. What is responsible for the carcinogenicity of PM_{2.5}? *Occupational Environmental Medicine* 2004; 61: 799–805.

Hayes RB. The carcinogenicity of metals in humans. *Cancer Causes Control* 1997;8:371–85.

Henry-Amar M. *Second cancer after the treatment for Hodgkin's disease: a report from the International Database on Hodgkin's Disease. Ann.Oncology. 1992; 3: 117–128.*

Higginson, J. Present trends in cancer epidemiology. In: Morgan JF, ed. Proceedings of the Eighth Canadian Cancer Research Conference. New York: *Pergamon*;1969. p. 40–7.

Holly EA, Lele C, Bracci PM, McGrath MS. Case-control study of non-Hodgkin's lymphoma among women and heterosexual men in the San Francisco Bay Area, California. *American Journal Epidemiology* 1999; 150: 375–389.

Howsam M, Grimalt JO, Guinó E, Navarro M, Martí-Ragué J, Peinado MA, Capellá G, Moreno V. Organochlorine Exposure and Colorectal Cancer Risk. *Environmental Health Perspectives* 2004; 112:1460–1466.

IARC. Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1 to 42. IARC *Monographs on the evaluation of carcinogenic risks to Humans* 1987; supply 7.

IARC. IARC Monographs on the evaluation of carcinogenic risks to humans, Vol 83, *Tobacco smoke and involuntary smoking*. Lyon: IARC, 2002.

International Agency for Research on Cancer (IARC). Non-Ionizing Radiation, Part 1: Static and Extremely Low-Frequency (ELF) Electric and Magnetic Fields. *Monography Evaluation Carcinogenesis Risks Human*: Vol 80. Lyon, 2002.

International Agency for Research on Cancer (IARC). IARC monographs on the evaluation of carcinogenic risks to humans. Inorganic and organic lead compounds (Vol. 87, 2004). <http://monographs.iarc.fr/htdocs/announcements/vol87.htm> [último acceso, 8-II-05].

International Programme on Chemical Safety (IPCS). Global assessment of the state-of-the-science of endocrine disruptors. Geneva, Switzerland: International Programme on Chemical Safety, *World Health Organization*, 2002. <http://ehp.niehs.nih.gov/who/> [último acceso 28 de febrero 2005].

Jackson RJ, Delozier DM, Gerasimov G, Borisova O, Garbe PL, Goultchenko L, Shakarishvili G, Hollowell JG, Miller DT. Chernobyl and iodine deficiency in the Russian Federation: an environmental disaster leading to a public health opportunity. *Journal Public Health Policy* 2002; 23: 453–470.

Jaffe ES, Harris NL, Stein H, Vardiman J. *World Health Organisation Classification of tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon: IARC Press, 2001.

Jarup L. Hazards of heavy metal contamination. *Br. Med. Bulletin* 2003;68:167–82.

Jereczek-Fossa BA, Alterio D, Jassem J, Gibelli B, Tradati N, Orecchia R. Radiotherapy-induced thyroid disorders. *Cancer Treatment Reviews*, 2004; 30:369–384.

Kaerlev L, Hansen J, Hansen HL, Nielsen PS. Cancer incidence among Danish seafarers: a population based cohort study. *Occ. Environmental Medicine* 2005; 62: 761–765.

Kahn LH, von Hippel F. Nuclear power plant emergencies and thyroid cancer risk. What New Jersey physicians need to know. *Neurological Journal Medicine* 2004; 101:22–27.

Kaldor JM, Day NE, Band P, Choi NW, Clarke EA, Coleman MP, Hakama M, Koch M, Langmark F, Neal FE. Second malignancies following testicular cancer, ovarian cancer and Hodgkin's disease: an international collaborative study among cancer registries. *International Journal of Cancer* 1987; 39: 571–585.

Kang DH, Rothman N, Poirier MC, Greenberg A, Hsu CH, Schwartz BS, et al. Interindividual differences in the concentration of 1-hydroxypyrene-glucuronide in urine and polycyclic aromatic hydrocarbon-DNA adducts in peripheral white blood cells after charbroiled beef consumption. *Carcinogenesis* 1995;16:1079–85.

Kass DE, Thier AL, Leighton J, Cone JE, Jeffery NL. Developing a Comprehensive Pesticida Health Effects Tracking System for an Urban Setting: New York City's Approach. *Environmental Health Perspectives* 2004; 112: 1419–1423.

Kato I, Koenig KL, Baptiste MS, Lillquist PP, Firzzeria G, Burke JS, Watanabe H, Shore RE. History of antibiotic use and risk of Non-Hodgkin's lymphoma. *International Journal Cancer* 2003; 107: 99–105.

Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, et al. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. *Environmental Health Perspectives* 1996;104 Suppl 4: 715–40.

Kesniiniene A, Cardis E, Tenet V, Ivanov VK, Kurtinaitis J, Malakhova I, Stengrevics A, Tekkel M. Studies of cancer risk among Chernobyl liquidators: materials and methods. *Journal Radiology Protection* 2003; 85: 323–329.

Kirk O, Pedersen C, Cozzi-Lepri A, Antunes F, Miller V, Gattell JM, Kallana C, Lazzarui A, Skinhej P, Barton SE. Non-Hodgkin Lymphoma in HIV-infected patients in the era of highly active antiretroviral therapy. *The American Society of Hematology* 2001; 93: 12.

Knowles DM. Immunodeficiency-associated lymphoproliferative disorders. *Modern Pathology: an official journal of the United States and Canadian Academy of Pathology*, 1999;12: 200–217.

Kogevinas M, Zock JP, Alsvaro T, Garcia-Villanueva M, Domingo-Domenech E, Kennedy S, Martínez-Maza O, Sanjose S. Occupational Exposure to Immunologically Active Agents and Risk for Lymphoma. *Cancer Epidemiology Biomarkers Prevention* 2004, 13(11): 1814–1818.

Krol ADG, Cessie SL, Snijder , Kluin JC, Kluin PHM, Noordijk EM. Non-Hodgkin's Lymphoma in The Netherlands: Results from a Population Based Registry. *Leukemia and Lymphoma*, 2003; 44: 451–458.

Krstev S, Dosemeci M, Lissowska J, Chow W-H, Zatonski W, Ward MH. Occupation and risk of stomach cancer in Poland. *Occ. Environmental Medicine* 2005; 62: 318–324.

Kung T-M, Ng W-L, Gibson JB. Volcanoes and carcinoma of the thyroid: a possible association. *Archives of Environmental Health* 1981; 36: 265–267.

Kyrtopoulos SA, Georgiadis P, Autrup H et al. Biomarkers of genotoxicity of urban air pollution. Overview and descriptive data from a molecular epidemiology study on populations exposed to moderate-to-low levels of polycyclic aromatic hydrocarbons: the AULIS project. *Mutation Research* 2001;496:207–28.

Langer P. Polychlorinated biphenyls and the thyroid gland. *Endocrine regulations* 1998; 32: 193–203.

Larijani B, Shirzad M, Mohagheghi MA, Haghpanah V, Mosavi-Jarrahi AR, Tavangar SM, Vassigh AR, Hossein-Nezhad A, Bandarian F, Baradar-Jalili R. Epidemiologic analysis of the Tehran Cancer Institute Data System Registry (TCIDSR). *Asian Pacific Journal of Cancer Prevention* 2004; 5:36–39.

Law G, Parslow R, McKinney P, Cartwright R. Non-Hodgkin's lymphoma and nitrate in drinking water: a study in Yorkshire, United Kingdom. *Journal Epidemiology Community Health* 1999; 53: 383–384.

Leval L, Harris NL. Variability in immunophenotype in diffuse large B-cell lymphoma and its clinical relevance. *Histopathology*, 2003; 43: 509–528.

Linnet, MS. Pottern, LM. Familial aggregation of hematopoietic malignancies and risk of non-Hodgkin's lymphoma. *Cancer Research.*, 1992; 52:5468s–5473s.

Lison D, De Boeck M, Verougstraete V, Kirsch-Volders M. Update on the genotoxicity and carcinogenicity of cobalt compounds. *Occ. Environmental Medicine* 2001;58: 619–25.

Liu S, Semenciw R, Mao Y. Increasing incidence of non-hodgkin's lymphoma in Canada, 1970–1996: Age–period–cohort analysis. *Hematol oncology* 2003; 21: 57–66.

Lope V, Pollán M, Pérez–Gómez B et al. Incidencia y mortalidad por cáncer de tiroides en España. *Boletín Epidemiológico* 2004; 12: 161–172.

Lopez–Abente G, Pollán M, Escolar A, Errezola M, Abaira V. *Atlas de mortalidad por cáncer y otras causas en España, 1978–1992*. Madrid: fundación Científica de la Asociación Española contra el Cáncer; 1996.

Lopez–Abente G, Pollán M, Aragonés N, Perez–Gomez B, Hernandez–Barrera V, Lope V, Suarez–Rodriguez B, Cárdbaba M, Cerdà T, Salas D. *La situación del cáncer en España*, Madrid; Centro de Publicaciones, Ministerio de Sanidad y Consumo, 2005.

Lundell M, Hakulinen T, Hala LE. Thyroid Cancer after Radiotherapy for Skin Hemangioma in Infancy. *Radiation Research* 1994; 140: 334–339.

Mahoney MC, Lawvere S, Falkner KL, Averkin YI, Ostapenko VA, Michalek AM, Moysich KB, McCarthy PL. Thyroid cancer incidence trends in Belarus: examining the impact of Chernobyl. *International Journal Epidemiology* 2004; 33:1025–1033.

Mannetje A't, Kogevinas M, Chang–Claude J et al. Occupation and bladder cancer in European women. *Cancer Causes Control* 1999; 10:209–217.

Mans DR, Mohamedradja RN, Hoebal AR, Rampadarath R, Joe SS, Wong J, Ramautar P, Mahabier R, Vrede MA. Cancer incidence in Suriname from 1980 through 2000 a descriptive study. *Tumori* 2003; 89:368–376.

Marcos–Gragera R. *Epidemiología de las neoplasias linfoides. Incidencia atribuible al sida*. Tesis Doctoral, Universitat Autònoma de Barcelona, Barcelona 2006.

Markaki I, Linos D, Linos A. The influence of dietary patterns on the development of thyroid cancer. *European Journal Cancer* 2003; 39: 1912–1919.

Mekori YA, Weitzman GL, Galli SJ. Reevaluation of reserpine–induced suppression of contact sensitivity. Evidence that reserpine interferes with T lymphocyte function independently of an effect on mast cells. *Journal of Experimental Medical sciences* 1985; 162: 1935–1953.

Menvielle G, Luce D, Févotte J, Bugel I, alomon C, Goldberg P, Billon–Galland P. Occupational exposures and lung cancer in New Caledonia. *Occupational Environmental Medicine* 2003; 60: 584–589.

Mikoczy Z, Welinder H, Tinnerberg H, Hagmar L. Cancer incidence and mortality of isocyanate exposed workers from the Swedish polyurethane foam industry: updated findings 1959–98. *Occupational Environmental Medicine* 2004; 61: 432–437.

Miller AB, To T, Agnew DA, Wall C, Green LM. Leukemia following occupational exposure to 60-Hz electric and magnetic fields among Ontario electric utility workers. *American Journal Epidemiology* 1996; 144:150–60.

Moore LE, Lu M, Smith A. Childhood Cancer Incidence and Arsenic Exposure in Drinking Water in Nevada. *Environmental Health* 2002; 57: 201–206.

Motykievicz G, Michalska J, Pendzich J, Malusecka E, Strozyk M, Kalinowska E, et al. A molecular epidemiology study in women from Upper Silesia, Poland. *Toxicology Letters* 1998; 96–97:195–202.

Moysich BM, Ravi J M, Michalek AM. Chernobyl-related ionising radiation exposure and cancer risk: an epidemiological review. *Lancet Oncology* 2002; 3: 269–279.

Malkasian GD, Mayberry WE. Serum total and free thyroxine and thyrotropin in normal and pregnant women, neonates, and women receiving progestogens. *American journal of obstetrics and gynecology* 1970; 108: 1234–1238.

Morgensztern D, Rosado MF, Silva O, Santos ES, Abdullah SA, Goodman M, Hamilton-Nelson K, Rosenblatt J, Lossos I. Prevalence of Hepatitis C Infection in Patients with Non-Hodgkin's Lymphoma in South Florida and Review of the Literature. *Leukemia & Lymphoma* 2004; 45(12): 2459–2464.

Morgner A, Bayerdorffer E, Neubauer A, Stolte M. Gastric MALT lymphoma and its relationship to Helicobacter pylori infection: management and pathogenesis of the disease. *Microscopy Research and Technique*, 2000; 48: 349–356.

Moshhammer H, Neuberger. Lung cancer and dust exposure: results of a prospective cohort study following 3260 workers for 50 years. *Occupational Environmental Medicine* 2004; 61: 157–162.

Moysich BM, Ravi J M, Michalek AM. Chernobyl-related ionising radiation exposure and cancer risk: an epidemiological review. *The Lancet Oncology* 2002; 3: 269–279.

Mueller N. The epidemiology of HTLV-I infection, *Cancer Causes Control*, 1991; 2:37–52.

Mueller, N. Overview of the epidemiology of malignancy in immune deficiency, *Journal of Acquired Immune Deficiency Syndromes* 1999; 21:5–10.

Nagataki S, Nystrom E. Epidemiology and primary prevention of thyroid cancer. *Thyroid* 2002; 12: 889–96.

Naresh KN, Agarwal B, Nathwani BN et al. Use of the World Health Organization (WHO) classification of non-Hodgkin's lymphoma in Mumbai, India: a review of 200 consecutive cases by a panel of five expert hematopathologist. *Leukaemia and Lymphoma* 2004; 45: 1569-1577.

National Toxicology Program (NTP). Report on carcinogens, Tenth Edition. U.S. Department of Health and Human Services. Public Health Service, National Toxicology Program, Diciembre 2002.

Negri E, Little A, Boiocchi M, Vecchia C, Franceschi S. B-Cell Non-Hodgkin's Lymphoma and hepatitis C virus infection: a systematic review. *International Journal of Cancer* 2004; 111: 1-8.

Olaya-Contreras P, Rodriguez-Villamil J, Posso-Valencia HJ, Cortez JE. Organochlorine exposure and breast cancer risk in Colombian women. *Cadernos de saude publica* 1998; 14 Suppl 3:125-32.

OMS. Clasificación Internacional de Enfermedades (CIE) 9ª rev. Barcelona: OMS, 1985.

OMS. Clasificación Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE) 10ª revisión. Washington: OPS, OMS, 1995.

Parkin DM, Muir SL, Whelan Y-T, Ferlay J, Powell J. Cancer incidente in five continents. *IARC Scientific Publications*. Lyon; 1992, n 120.

Peláez S, Hierro I, Oña S, Alonso L, Matilla A. Relación entre la exposición a pesticidas y el desarrollo de carcinoma urotelial vesical superficial de bajo grado. *Medicina Clínica* 2004; 123: 571-574.

Pershagen G, Simonato L. Epidemiological evidence on outdoor air pollution and cancer. In: Tomatis L, editor. *Air Pollution and Human Cancer*. Berlin: Springer-Verlag; 1990. pp. 63-94.

Poole C, Greenland S, Luetters C, Kelsey JL, Mezei G. Socieconomic status and childhood leukaemia: a review. *Int. Journal of Epidemiology* 2006; 35: 370-384.

Potzsch C, Schaefer, HE, Lubbert, M. Familial and metachronous malignant lymphoma: absence of constitutional mutations, *Am. Journal Hematology* 1999; 62: 144-149.

Preston-Martin S, Jun F, Duda MJ, et al. A case-control study of thyroid cancer in women under age 55 in Shanghai (People's Republic of China). *Cancer Causes Control* 1993; 4: 431-440.

Pujolar EA, Gonzalez CA, Lopez Abente G, Errezola M, Izarzugaza I, Nebot M, Riboli E. Bladder cancer and coffee consumption in smokers and non-smokers in Spain. *International Journal of Epidemiology* 1993;22:38-44.

Quackenboss JJ, Pellizzari ED, Shubat P et al. Design strategy for assessing multi-pathway exposure for children: the Minnesota Children's Pesticide Exposure Study (MNCPEs). *J of Exposure Analysis and Environmental Epidemiology* 2000;10:145-58.

Rafnsson V, Sulem P. Cancer incidence among marine engineers, a population-based study (Iceland). *Cancer Causes Control* 2003; 14(1): 29-35.

Rahu M. Health effects of the Chernobyl accident: fears, rumours and the truth. *European Journal Cancer* 2003; 39: 295-299.

Reid A, Klerk N, Ambrosini G, Olsen N, Pang SC, Musk AW. The additional risk of malignant mesothelioma in former workers and residents of Wittenoom with benign pleural disease or asbestosis. *Occ Environmental Medicine* 2005, 62: 665-669.

Rericha V, Kulich M, Rericha R, Shore DL, Sandler DP. Incidence of Leukemia, Lymphoma, and Multiple Myeloma in Czech Uranium Miners: A Case-Cohort Study. *Environmental Health Perspectives* 2006, 114(6):818-822.

Ribera JM, Navarro JT. Linfomas en pacientes con infección por el VIH. Las cosas han cambiado para bien. *Enfermedad Infección Microbiológica Clínica* 2004;22(6):313-314.

Riboli E, Kaaks R. The EPIC projects: Rationale and Study design. *International Journal of Epidemiology* 1997; vol26: S8-S14.

Richardson DB, Ashmore JP. Investigating time patterns of variation in radiation cancer associations. *Occupational Environmental Medicine* 2005; 62: 551-558.

Romundstad P, Haldorsen T, Andersen A. Cancer incidence and cause specific mortality among workers in two Norwegian aluminum reduction plants. *American Journal of Industrial Medicine* 2000; 37(2): 175-183.

Röösli M y Künzli N. Comentario: Magnetic field exposure and childhood leukaemia—moving the research agenda forward. *International Journal of Epidemiology* 2006;35: 407-408.

Safe SH, Zacharewski T. Organochlorine exposure and risk for breast cancer. *Progress in clinical and biological research* 1997;396: 133-45.

Salud (CIE) 10^a revisión. Washington: OPS, OMS, 1995.

Samet JM, Cohen AJ. Air pollution and lung cancer. In: Holgate ST, Samet JM, Koren HS, Maynard R, eds. Air pollution and health. London: Academic Press, 1999: 841–64.

Satarug S, Baker JR, Urbenjapol S, Haswell–Elkins M, Reilly PEB, Williams DJ, et al. A global perspective on cadmium pollution and toxicity in non–occupationally exposed population. *Toxicology Letters* 2003; 137: 65–83.

Schafer DW, Lubin JH, Ron E, Stovall M, Carroll RJ. A Reanalysis of Thyroid Neoplasms in the Israeli Tinea Capitis Study Accounting for Dose Uncertainties. *Radiation Research* 2004; 161: 359–368.

Scherer G, Frank S, Riedel K, Meger–Kossien I, Renner T. Biomonitoring of exposure to polycyclic aromatic hydrocarbons of nonoccupationally exposed persons. *Cancer Epidemiology Biomarkers Prevention* 2000;9: 373–80.

Scherr PA, Muller NE. Non–Hodking Lymphomas, 2^a edn. *Cancer epidemiology and prevention*, New York: Oxford University Press, 1996: 920–945.

Seilkop SK, Oller AR. Respiratory cancer risks associated with low–level nickel exposure: an integrated assessment based on animal, epidemiological, and mechanistic data. *Regulation Toxicology Pharmacology* 2003;37:173– 90.

Serra C, Bonfill X, Sunyer J, et al. Working Group on the Study of Bladder Cancer in the County of Vallès Occidental. Bladder cancer in the textile industry. *Scandinavian Journal of Work, Environment and Health* 2000; 26: 474–479.

Serra C. *Ocupació i cancer de bufeta urinària al vallès occidental*. La Salut Laboral Sabadell, Institut Català de la Salut–Ajuntament de Sabadell 2002; 20–26

Shakhtarin VV, Tsyb AF, Stepanenko VF, Orlov MY, Kopecky KJ, Davis S. Iodine deficiency, radiation dose, and the risk of thyroid cancer among children and adolescents in the Bryansk region of Russia following the Chernobyl power station accident. *International Journal Epidemiology* 2003; 32: 584–591.

Sharpe M. Rising hormones. *Journal of Environmental Monitoring: JEM* 1999; 1:5N–7N.

Silverman DT, Morrison AS, Devesa SS. Bladder cancer. In: Schottenfeld D, Fraumeni J (eds). *Cancer epid prev*. New York: Oxford University Press, 1996; 1156–1179.

Simcox NJ, Fenske RA, Wolz SA, Lee IC, Kalman DA. Pesticides in Household Dust and Soil: Exposure Pathways for Children in Agricultural Families. *Environmental Health Perspectives* 1995;103:1126–34.

Smedby KE, Hjalgrim H, Melbye M, Torrany A. Ultraviolet Radiation Exposure and Risk of Malignant Lymphomas. *Journal of National Cancer Institute* 2005; 97: 199–209.

Sorahan T, Esmen NA. Lung cancer mortality in UK nickel-cadmium battery workers, 1947–2000. *Occupational Environmental Medicine* 2004; 61: 108–116.

Sorensen M, Autrup H, Hertel O, Wallin H, Knudsen LE, Loft S. Personal exposure to PM2.5 and biomarkers of DNA damage. *Cancer Epi Biomarkers Prev* 2003;12:191–196.

Steenland K, Boffetta P. Lead and cancer in humans: where are we now? *American Journal of Industrial Medicine* 2000;38:295–9.

Stezhko VA, Buglova EE, Danilova LI et al. A cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: objectives, design and methods. *Radiation Research* 2004; 161:481–492.

Stone, R. Environmental estrogens stir debate (News & Comment). *Science*. 1994;265:308–310.

Study in New Caledonia (South Pacific), a Very High Incidence Area. *American Journal Epidemiology* 2005; 161: 1056–1065.

Svec MA, Ward MH, Dosemeci M, Checkoway H, De Roos AJ. Risk of lymphatic or haematopoietic cancer mortality with occupational exposure to animals or the public. *Occupational Environmental Medicine* 2005; 62: 726–735.

Takagi A, Watanabe S, Igarashi M, Koike J, Hasumi K, Deguchi R, Koga Y, Miwa T. The effect of *Helicobacter pylori* on cell proliferation and apoptosis in gastric epithelial cell lines. *Alimentary Pharmacology & Therapeutics* 2000; 14: 188–192.

Takahashi T, Schoemaker MJ, Trott KR, Simon SL, Fujimori K, Nakashima N, Fukao A, Saito H. The relationship of thyroid cancer with radiation exposure from nuclear weapon testing in the Marshall Islands. *Journal Epidemiology* 2003; 13: 99–107.

Tang D, Warburton D, Tannenbaum SR, Skipper P, Santella RM, Cereijido GS, et al. Molecular and genetic damage from environmental tobacco smoke in young children. *Cancer Epidemiology Biomarkers Prevention* 1999;8:427–31.

Tardon A, Martínez B, Kogevinas M, Cueto A. Cáncer de vejiga en Asturias: incidencia y evolución (1982–1993). *Gaceta Sanitaria* 1999; 13: 346–352.

Tavani A, Pregnolato A, Negri E, Franceschi S, Serraino D, Carbone A, La Vecchia C. Diet and risk of lymphoid neoplasms and soft tissue sarcomas, *Nutritional Cancer* 1997; 27: 256–260.

Tolbert PE. Oils and Cancer. *Cancer Causes Control* 1997;8: 386–405.

Tondel M, Hjalmarsson P, Hardell L, Carlsson G, Axelson O. Increase of regional total cancer incidence in north Sweden due to the Chernobyl accident? *Journal Epidemiology Community Health* 2004; 58: 1011–1016.

Truong T, Orsi L, Dubourdieu D, Rougier Y, Hemon D, Guénel P. Role of Goiter and of Menstrual and Reproductive Factors in Thyroid Cancer : A Population-based Case-Control Study in New Caledonia (South Pacific, a Very High Incidence Area. *American Journal Epidemiology* 2005; 161: 1056–1065.

Tsai HT, Wu MT, Hauser R, Rodrigues E, Ho CK, Liu CL, et al. Exposure to environmental tobacco smoke and urinary 1-hydroxypyrene levels in preschool children. *Kaohsiung Journal Medicine Sciences* 2003;19:97–104.

Tsuda H, Naito A, Kim CK, Fukamachi K, Nomoto H, Moore MA. Carcinogenesis and its modification by environmental endocrine disruptors: in vivo experimental and epidemiological findings. *Japanese Journal of Clinical Oncology* 2003; 33:259–70.

Tukiendorf A, Krasowski G, Rybak Z. Thyroid cancer morbidity in Opole province, Poland, after Czernobyl disaster. *Central European j. public health* 2003; 11: 98–101.

UERC. *Memòria 2003*. Girona: Unitat Epidemiologia Registre Càncer Girona, 2004.

UERC. *Memòria 2004*. Girona: Unitat Epidemiologia Registre Càncer Girona, 2005.

Urrutia G, Serra C, Bonfill X, Bastús R. Incidencia del càncer de vejiga urinaria en un àrea industrializada de España. *Gaceta Sanitària* 2002; 16(4): 291–7.

Vajdic CM, Kricker A, Giblin M, McKenzie J, Aitken J, Giles GG, et al. Sun exposure predicts risk of ocular melanoma in Australia. *International Journal of Cancer* 2002;101:175–82.

Vergier P, Catelinois O, Tirmarche M et al. Thyroid cancers in France and the Chernobyl accident: risk assessment and recommendations for improving epidemiological knowledge. *Health Physic* 2003; 85: 323–329.

Viel JF, Arveux P, Baverel J, Cahn JY. Soft Tissue Sarcoma and Non Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology* 2000;152(1):13–9.

Viladiu P, Izquierdo A, Marcos R, et al. *Epidemiologia del càncer a Girona*: Unitat de Registre del Càncer de Girona, Institut d'Assistència Sanitària, 2000.

Villanueva CM, Kogevinas M, Grimalt JO. Cloració del aua potable en Espanya y càncer de vejiga. *Gaceta Sanitària* 2001; 15: 48–53.

Villeneuve PJ, Agnew DA, Miller AB, Corey PN. Non-Hodgkin's lymphoma among electric utility workers in Ontario: the evaluation of alternate indices of exposure to 60 Hz electric and magnetic fields. *Occupational Environmental Medicine* 2000;57:249-57.

Vineis P, Crosignani P, Sacerdote C et al. Hematopoietic cancer and peptic ulcer: a multicenter case-control study. *Carcinogenesis* 1999; 20:1459-1463.

Vose JM, Chiu BC, Cheson B, Cancey J, Wright J. Update on Epidemiology and Therapeutics for Non-Hodgkin's Lymphoma. *Hematology* 2002; 241-261.

Ward MH, Lubin J, Giglierano J, Colt JS, Wolter C, Begiroglu N, Camann D, Hartge P, Nuckols JR. Proximity to Crops and Residential Exposure to Agricultural Herbicides in Iowa. *Environmental Health Perspectives* 2006, 114(6):893-897.

Weeke J, Hansen AP. Serum TSH and serum T levels during normal menstrual cycles and during cycles on oral contraceptives. *Acta Endocrinológica* 1975; 79:431-438.

Whitrow MJ, Smith BJ et al. Environmental exposure to carcinogens causing lung cancer: Epidemiological evidence from the medical literature. *Resp* 2003;8:513-21.

Whyatt RM, Santella RM, Jedrychowski W, Garte SJ, Bell DA, Ottman R, et al. Relationship between ambient air pollution and DNA damage in Polish mothers and newborns. *Environmental Health Perspectives* 1998;106 Suppl 3:821-6.

Wing S and Richardson DB. Age at exposure to ionising radiation and cancer mortality among Hanford workers: follow up through 1994. *Occ Env Med* 2005; 62: 465-472.

Wolff MS, Toniolo PG, Lee EW, Rivera M, Dubin N. Blood levels of organochlorine residues and risk of breast cancer. *Journal National Cancer Institute* 1993; 85:648-52.

WHO. Chemical hazards in drinking-water. WHO 2004. [cited 16 December 2004] http://who.int/water_sanitation_health/dwg/chemicals/polyaromahydrocarbons/en/ [último acceso Mar 05]

Zablotska LB, Chak A et al. Increased Risk of Squamous Cell Esophageal Cancer after Adjuvant Radiation Therapy for Primary Breast Cancer. *AJE* 2005;161: 330-337.

Zeka A, Eisen EA, Kriebel D, Gore R, Wegman DH. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occupational Environmental Medicine* 2004; 61: 426-431.

Zhang S, Hunter DJ, Rosner BA et al. Dietary fat and protein in relation to risk of non-Hodgkin's lymphoma among women. *J Nat Cancer Inst* 1999;91:1751-8.

Zhu K, Levine RS, Gu Y, Brann, EA, Hall I, Caplan LS, Baum MK. Non-Hodgkin's lymphoma and family history of malignant tumors in a case-control study. *Cancer Causes Control* 1998; 9: 77-82.

Hipótesis

1.- La mortalidad por cáncer presenta algunas limitaciones para conocer el impacto de esta enfermedad en la población:

1.1.- Los tumores más letales están más representados.

1.2.- La información contenida en los certificados de defunción no es lo suficientemente precisa para todas las neoplasias.

2.- La mayoría de las neoplasias se pueden atribuir a factores de tipo medioambiental.

3.- Las neoplasias de interés medioambiental son las neoplasias linfoides, cáncer de tiroides, cáncer de vejiga urinaria y cáncer de tráquea, bronquios y pulmón.

4.- Los factores medioambientales, tales como la contaminación atmosférica y el nivel socioeconómico son los principales factores de riesgo de las neoplasias de interés medioambiental.

5.- Las observaciones de todas las variables consideradas en este trabajo, presentan dependencia espacial.

6.- Los datos en este trabajo se organizan en diseños en rejilla (variables dependientes, mortalidad e incidencia; variables explicativas socioeconómicas) y espacialmente continuos (variables explicativas medioambientales).

7.- La variación espacial de la distribución de la ocurrencia de una enfermedad no está explicada totalmente ni por la distribución espacial de factores explicativos conocidos ni por una variación aleatoria.

8.- Cuando se disponen de datos espaciales, la extravariabilidad (variabilidad no observada) puede descomponerse en variabilidad no estructurada espacialmente (heterogeneidad) y en variabilidad estructurada espacialmente (dependencia espacial).

9.- En algunas neoplasias de interés medioambiental, la distribución de los casos (incidentes o muertes) es más dispersa que la distribución de Poisson.

10.- Cuando la exposición es individual pero se interpola de forma ecológica para que coincida con el diseño de la variable dependiente, existe una variabilidad intra-área distinta del error de medida, que debe ser controlada por los modelos.

Objetivos

- 1.- Estudiar la relación entre la incidencia y mortalidad por cáncer y los factores medioambientales, en particular la contaminación atmosférica, controlando por factores socioeconómicos.
- 2.- Utilizar aquellos métodos de estadística espacial apropiados para cada tipo de diseño.
- 3.- Distinguir en los modelos las diferentes fuentes de extra-variabilidad espacial.
- 4.- Controlar el problema de exceso de ceros inherente a alguna de las neoplasias de interés medioambiental.
- 5.- Introducir adecuadamente, con el fin de evitar posibles sesgos, variables explicativas (según el diseño en que éstas se han observado) de la variabilidad geográfica de las neoplasias de interés medioambiental.

2.- Modelos estadísticos aplicados a datos espaciales

2.1.- Estadística espacial

La estadística espacial tiene como objeto la exploración, la descripción, la visualización y el análisis de los datos considerando sus características de distribución en el espacio. Aquello que distingue la estadística espacial de otras ramas de la estadística es el uso de las localizaciones geográficas en las especificaciones del modelo estadístico.

Como se sabe, las hipótesis subyacentes en el análisis estadístico clásico son que las variables aleatorias siguen una misma distribución de probabilidad y que son independientes entre si. Esta hipótesis permite modelizar la distribución conjunta del modelo como el producto de las distribuciones individuales y permite asumir, como en el análisis de regresión, que los errores sean independientes entre si.

En contraposición con la estadística clásica o convencional, en la estadística espacial las observaciones no se consideran independientes, por el contrario se supone de manera implícita que están correlacionados unas con otras, es decir que existe una dependencia espacial. La característica de dependencia significa una diferencia substancial respecto al análisis estadístico efectuado con datos independientes y va a suponer tanto ventajas (como una predicción más precisa) como también inconvenientes (como una estimación menos precisa) (Saez y Saurina, 2006).

2.2.- Tipos de datos espaciales

Cuando se habla de datos espaciales se refiere a mediciones u observaciones realizadas en localizaciones o en áreas específicas. Las localizaciones pueden referirse a puntos o a áreas. Los datos referidos a un área son los que se utilizan con más frecuencia en epidemiología, tratándose de observaciones agregadas. La localización espacial puede ser regular o irregular, aunque lo habitual en epidemiología espacial es

que sea irregular ya que raramente se trabaja con mediciones obtenidas uniformemente de forma equidistante unas de otras.

Los datos espaciales pueden ser clasificados en tres grandes grupos. En función de esta clasificación, deberá de utilizarse una metodología específica de análisis:

1. Datos en rejilla (*'lattice data'*)
2. Datos espacialmente continuos (*'geostatistical data'*)
3. Procesos puntuales (*'point processes'*)

2.2.1.- Datos en rejilla (*'lattice data'*)

Los datos en rejilla son observaciones procedentes de un proceso aleatorio, observadas sobre una colección contable de regiones espaciales, que pueden estar regular o irregularmente distribuidas, complementados con lo que se denomina estructura de vecindad, es decir información sobre las regiones vecinas.

Matemáticamente una rejilla queda definida como un conjunto de lados y de vértices, es decir, un conjunto de índices de localizaciones con un conjunto asociado de vecinos. Ya que los datos en rejilla están definidos en regiones espaciales las localizaciones concretas especificadas por el vector s suelen referirse al centroide de la región.

Los datos en rejilla son los que habitualmente se utilizan en epidemiología. Así por ejemplo se conoce el número de casos de enfermedad/muerte por comarcas, provincias, municipios o áreas de salud. El objetivo en el análisis de este tipo de datos es el de descubrir y modelizar el patrón espacial existente en los datos y descubrir los determinantes de ese patrón. Las técnicas de representación geográfica de datos en epidemiología suelen ser un paso previo para la visualización de la estructura espacial

y constituyen una herramienta potente para el análisis exploratorio de los datos antes de proceder a su modelización.

Hasta ahora los procedimientos disponibles para el ajuste de datos en rejilla eran muy limitados y hacían prácticamente intratable el problema. Afortunadamente ahora existen soluciones procedentes, en su mayoría, de la confluencia de la estadística Bayesiana y la potencia de computación que posibilitan hacer este tipo de análisis. La difusión de los métodos MCMC (Markov Chain Monte Carlo) y su aplicación mediante el software WinBUGS suponen toda una revolución en este y en muchos otros campos (Gilks et al., 1996; López-Abente e Ibáñez, 2001).

2.2.2.- Datos geoestadísticos (*‘geostatistical data’*)

Los datos espacialmente continuos, denominados también datos de campo aleatorio, son mediciones tomadas en puntos fijos con localizaciones continuas en el espacio. Las mediciones clásicas en salud ambiental, tales como contaminación de suelos y de aire o la radiación natural, suelen pertenecer a esta categoría. La variable medida, sin embargo, puede ser tanto continua como discreta. El objetivo es el de, dados los valores tomados en puntos de muestreo fijos, extender la distribución espacial de los valores de un atributo sobre la región total de estudio. Por lo tanto el análisis de los datos geoestadísticos puede contemplar tanto la modelización del patrón de variabilidad, la determinación de los factores con los que pueda estar relacionado, como la de obtener una buena predicción de la variable en puntos donde no se ha muestreado (Bailey y Gatrell, 1995; Kaluzni et al., 1998). Estos métodos son muy utilizados en estudios del área de geociencias (suelo, clima, hidrología, geología minera, etc) por lo que también suelen ser denominados datos geoestadísticos. En epidemiología pueden encontrarse ejemplos de aplicación de estas técnicas en estudios de incidencia de enfermedades (Oliver et al., 1996) y en aplicaciones en vigilancia epidemiológica (Abellán, 2000).

2.2.3.- Procesos puntuales (*'point processes'*)

Se denominan datos patrones de puntos o procesos puntuales, cuando las localizaciones (y no las mediciones) son las variables de interés. Consisten en un número finito de localizaciones observadas en una región determinada. Un sencillo ejemplo de proceso puntual sería las coordenadas de los domicilios de los enfermos en el caso de una determinada enfermedad. En este tipo de datos se trataría de valorar si existe una tendencia de los eventos a exhibir un patrón sistemático, especialmente alguna forma de regularidad o de agregación. Así, podría ser importante evaluar si los puntos se manifiestan cercanos a un punto del plano donde se sitúa un foco contaminante.

El objetivo de los procesos puntuales es de conocer la variación de la intensidad de los eventos sobre la región de estudio y el de buscar modelos que ayuden a explicar o comprender el fenómeno. Algunos ejemplos de aplicación de estas técnicas pueden verse en Gatrell et al. (1996); en Kelsall y Diggle (1998); y en López-Abente e Ibáñez (2001).

2.3.- Epidemiología espacial

El término epidemiología espacial o geográfica se emplea para describir estudios sobre las causas y la prevención de las enfermedades utilizando diferentes perspectivas de análisis en las que la localización de los eventos es un componente fundamental (Thomas, 1990). Uno de sus principales objetivos es el de mostrar que parte de la variación espacial de la distribución de la ocurrencia de una enfermedad no está explicada ni por la distribución espacial de factores explicativos conocidos ni por una variación aleatoria. De hecho, a menudo, interesa encontrar pistas sobre algún factor de riesgo desconocido de esta enfermedad (Saez y Saurina, 2006).

La epidemiología espacial goza de una larga tradición remontándose a principios del siglo XIX, en el que se realizaron mapas de la ocurrencia de enfermedades en diferentes países con el objetivo de caracterizar la extensión y las posibles causas de brotes de enfermedades infecciosas tales como la fiebre amarilla y el cólera (Walter, 2000).

Con posterioridad, la epidemiología espacial creció en complejidad, sofisticación y utilidad (Elliott y Wartenberg, 2004). La epidemiología espacial extiende la amplia tradición de estudios ecológicos¹ que utilizan las explicaciones de la distribución de enfermedades en diferentes localizaciones geográficas para comprender mejor la etiología de la enfermedad (Doll, 1980; Elliott y Wartenberg, 2004).

Recientes avances en la disponibilidad de datos y en los métodos analíticos para tratarlos proporcionan nuevas oportunidades para extender las investigaciones epidemiológicas tradicionales, a escala nacional o regional, hasta el estudio de las variaciones en las enfermedades a nivel local, o de área pequeña (Walter, 2000). Tales investigaciones deberían contemplar factores de riesgo para la salud con relevancia a nivel local, tales como exposiciones ambientales, la distribución local de condiciones socioeconómicas y los hábitos y estilos de vida (Elliott y Wartenberg, 2004).

Siguiendo a Elliott y Watenberg (2004) la epidemiología espacial en áreas pequeñas puede ser abordada desde tres grandes perspectivas:

- Mapas de enfermedades
- Estudios de asociación geográfica
- Aglomeraciones de casos o '*clustering*'

¹ En epidemiología se distingue entre estudios a nivel individual y a nivel ecológico, según cual sea la unidad de análisis. En los diseños individuales se disponen de observaciones a nivel individual tanto de la variable respuesta (a explicar) como de las posibles variables explicativas. En los diseños ecológicos no se dispone de observaciones individuales, al menos en uno de los dos casos. Por ejemplo, se pueden disponer de datos de ocurrencia de cáncer a nivel individual pero no de posibles factores explicativos de esta ocurrencia.

2.4.- Variables respuesta: Incidencia y mortalidad por neoplasias de interés medioambiental. Mapas de enfermedades

La representación gráfica de la distribución espacial de las enfermedades ha sido y es de un gran interés para mostrar geográficamente las razones de incidencia o de mortalidad con objetivos diversos, abarcando desde los puramente descriptivos (mapas de enfermedades o de mortalidad); la formulación de hipótesis relacionadas con la etiología de la enfermedad que se estudia; y la documentación o establecimiento del marco de estudios de epidemiología ambiental y espacial.

No se trata de un tema nuevo, pues en la literatura pueden encontrarse ejemplos tales como, los mapas sobre la distribución de la fiebre amarilla publicados en 1798; el mapa de John Snow, en 1854 de la distribución geográfica de la mortalidad por cólera en el Soho londinense y la resolución del brote mediante la identificación del foco; y las investigaciones de Palm sobre raquitismo publicadas en 1890 (Walter, 2000).

En los últimos años se ha estado produciendo un gran aumento en este tipo de investigaciones debido al desarrollo de técnicas estadísticas (English, 1992; Cliff y Haggett, 1992; Elliott et al. 1995) y de los programas informáticos de análisis y representación gráfica de los resultados, como los Sistemas Información Geográfica (GIS). Walter y Birmie (1991), por ejemplo, identificaron 49 Atlas regionales, nacionales e internacionales de distintas enfermedades.

Los mapas de enfermedades proporcionan un rápido resumen visual de información geográfica compleja y permiten identificar patrones en los datos que de otro modo podrían pasar inadvertidos en las presentaciones tabulares (Elliott y Wartenberg, 2004). De hecho, los mapas se utilizan con propósitos descriptivos, con el objetivo de generar hipótesis etiológicas; para la vigilancia epidemiológica, a fin de detectar áreas con un aparente mayor riesgo; y como ayuda en la definición de políticas de salud y de

asignación de recursos. Son también útiles para localizar clusters específicos de enfermedades, así como los resultados de otros estudios de focos contaminantes.

En mapas de enfermedades se utilizan fundamentalmente los datos en rejilla. El objetivo en el análisis de este tipo de datos es el de modelizar el patrón espacial en los valores sobre estas áreas y descubrir los determinantes de ese patrón.

2.4.1.-Casos Brutos y Razones Estandarizadas

Se denominan casos brutos a los datos crudos tal y como se observan. Puede considerarse, por ejemplo los casos de incidencia de una de las neoplasias de interés medioambiental, estas observaciones podrían ser representadas directamente sobre la estructura geográfica de una zona, obteniendo un mapa.

Esta sencilla representación no proporciona información de interés. Cualquier interpretación de la estructura de los casos incidentes está limitada por la falta de información sobre la distribución espacial de la población que podría estar 'a riesgo'. Es de esperar que las áreas con más población (a riesgo) tendrán más casos incidentes.

A la representación de los casos brutos, se prefiere en general la representación de razones que permiten incorporar el efecto de la población a riesgo.

La idea más sencilla es la de dividir el número total de casos observados en un área determinada por el total de la población a riesgo en dicha área:

$$p_{ij} = x_{ij} / n_{ij}$$

p_{ij} es la razón 'cruda' el área j y para el grupo de edad i ; x_{ij} indica el número de casos del grupo de edad i en el área j ; y n_{ij} el número de individuos del grupo de edad i en el área j .

Sin embargo el uso directo de las razones crudas no permite la comparación entre distintas áreas, ya que las diferencias observadas entre ellas pueden ser debidas a factores que no hayan sido tenidos en cuenta. Por lo tanto, hay que estandarizar las razones por aquellos factores conocidos que hayan influido sobre ellas. La estandarización más utilizada en epidemiología es la denominada 'por edad', es decir considerar la diferente distribución de la población por grupos de edad en las áreas cuyas razones se quieran comparar.

Existen dos métodos para estandarizar las razones; el método directo e indirecto (véase Saez y Saurina, 2006, para una amplia revisión).

a.- Estandarización de razones por el método directo

La distribución por grupos edad de la población de las áreas geográficas que se quieren comparar se hace corresponder con la de un área 'estándar' o 'de referencia'.

Las razones estandarizadas por el método directo para dos áreas geográficas a y b se calculan como:

$$DSR_a = \frac{\sum p_{ia} n_{is}}{\sum n_{is}} = \sum p_{ia} \left(\frac{n_{is}}{\sum n_{is}} \right) \quad DSR_b = \frac{\sum p_{ib} n_{is}}{\sum n_{is}} = \sum p_{ib} \left(\frac{n_{is}}{\sum n_{is}} \right)$$

siendo x_{ij} el número de casos observados en el grupo de edad i y área j , n_{ij} el número de individuos en el grupo de edad i perteneciente al área j , p_{ij} la razón cruda para el grupo de edad i y el área j , $p_{ij} = x_{ij} / n_{ij}$, y s el subíndice que indica el área de referencia.

Las expresiones obtenidas en cada una de las áreas no son más que una media ponderada de las razones crudas para cada grupo de edad, utilizando las mismas ponderaciones ($\frac{n_{is}}{\sum n_{is}}$) en las dos áreas.

Es evidente que en este tipo de estandarización, la elección del área de referencia es crucial y que los resultados dependen claramente de la elección realizada.

b.- Estandarización de razones por el método indirecto

A veces, la estandarización por el método directo es difícil de calcular. Así, las razones crudas por edad en alguna (o todas) las áreas a comparar (p_{ia}, p_{ib}) o no son conocidas, o son difíciles de obtener. Además, a menudo, las razones crudas están calculadas utilizando un numerador, es decir el número de casos, muy pequeño, lo que las hace extremadamente sensibles.

La estandarización por el método indirecto, en cambio, no requiere del conocimiento de las razones crudas de los grupos de edad en cada área.

Así, definiendo las razones crudas en cada una de las áreas mediante la aplicación de las razones (crudas) específicas por edad del área de referencia:

$$CR_a = \frac{\sum p_{is} n_{ia}}{\sum n_{ia}}$$

siendo s el área de referencia; y a el área específica de estudio.

El factor de estandarización viene expresado por el cociente entre las razones crudas en el área de referencia (s) y el área de estudio (a):

$$FS = \frac{\sum p_{is} n_{is}}{\sum n_{is}} \left(\frac{\sum n_{ia}}{\sum p_{is} n_{ia}} \right)$$

La razón estandarizada por el método indirecto se define \widehat{ISR}_a y es el producto de la razón del área por el factor de estandarización, $CA \times FS$:

$$\widehat{ISR}_a = \left(\frac{\sum p_{ia} n_{ia}}{\sum p_{is} n_{ia}} \right) \left(\frac{\sum p_{is} n_{is}}{\sum n_{is}} \right)$$

siendo $CA = \frac{\sum p_{ia} n_{ia}}{\sum n_{ia}}$

El paréntesis de la derecha no es más que la razón cruda del área de referencia y el de la izquierda se denomina, razón de mortalidad (o de incidencia) estandarizada, RME (o RIE, respectivamente) o SMR (SIR) en sus siglas en inglés.

Algunos autores proponen utilizar únicamente la estandarización directa, (Julious et al., 2001), por cuanto, al basarse en un estándar común, permiten la comparación directa entre diferentes estudios. Sin embargo, Elliott y Wartenberg (2004), entre otros, argumentan que ambos métodos de estandarización proporcionan resultados prácticamente idénticos.

- Cálculo de la Razón de Incidencia/Mortalidad Estandarizada (RIE/RME)

Como se acaba de ver, ni las RME ni las RIE constituyen por si mismas razones estandarizadas por el método indirecto, pero, de hecho, son prácticamente idénticas y, en la práctica, son utilizadas en vez de ellas.

Tal como se ha comentado en el apartado anterior, y centrándose en el caso de la incidencia de una enfermedad, el estado de un área determinada suele evaluarse determinando, en primer lugar, que incidencia se esperarí en esa área y, a

continuación, comparándola con la incidencia observada. El ratio observado/esperado, denominado razón de incidencia estandarizada (RIE), no es más que un estimador del riesgo relativo del área, es decir del riesgo de enfermedad en relación al grupo considerado de referencia.

En concreto, la Razón de Incidencia Estandarizada para el municipio i y para el sexo l , RIE_{il} , se calcula como:

$$RIE_a = \left(\frac{\sum p_{ia} n_{ia}}{\sum p_{is} n_{ia}} \right) \Rightarrow RIE_{il} = \frac{O_{il}}{E_{il}}$$

donde O_{il} son los casos observados de la enfermedad de interés en el municipio i y para el sexo l en un período determinado; y E_{il} son los casos esperados para el mismo sexo, municipio y período.

A su vez, E_{il} se calcula como:

$$E_{il} = \sum_{kl} N_{ikl} p_{kl}$$

donde kl denota un grupo de edad-sexo (por ejemplo: $l =$ hombres, mujeres; $k = \leq 14$; 15-34; 35-44; 45-54; 55-64; 65-74; 75-84; ≥ 85); N_{ikl} es el número de habitantes en el municipio i del sexo l y grupo de edad k ; y p_{kl} es la probabilidad que un individuo de sexo l y grupo de edad k en el *área de referencia* desarrolle la enfermedad de interés. Todo en un período de tiempo determinado.

Las RIE, aun cuando son ampliamente utilizadas en la representación de mapas de incidencia, tienen algunas limitaciones (Clayton y Kaldor, 1987; Lawson et al., 2003). En primer lugar, no tienen en cuenta la variación de la población entre diferentes áreas geográficas. En este sentido, las RIE extremas en el mapa y, por tanto, dominantes en el aparente patrón geográfico que pueda aparecer en el mismo, son las estimadas con

menos precisión, simplemente porque su cálculo se ha basado en muy pocos casos, bien sea en el numerador (incidencia observada) o bien en el denominador (esperados). En general, además, la variación de los casos observados suele ser bastante mayor que la esperada, produciéndose lo que se denomina extra-variabilidad.

– **Estimación de razones estandarizadas (GLM)**

Parte de los problemas de las RIE pueden ser solventados estimándolas mediante un modelo lineal generalizado (GLM) (ver Anexo I).

En concreto, se acostumbra a suponer que los casos observados siguen un proceso contador tipo Poisson:

$$O_i \rightarrow \text{Poisson}(\mu_i E_i)$$

donde μ_i es el *riesgo relativo*, RR_i asociado al municipio i . La RIE_i no es más que su estimador.

En este sentido, un GLM para el riesgo relativo puede ser expresado:

- Función vínculo:

$$\text{Ln}(E(O_i)) = \beta_0 + \text{Ln}(E_i)$$

[1]

- Función varianza:

$$\text{Var}(O_i) = E(O_i)$$

Este modelo supone que no existe exceso de riesgo, es decir que el número observado de casos coincidirá con el número esperado (Lawson et al., 2003). En general, sin embargo, la variación de los casos observados no es la misma que la esperada en una distribución de Poisson, produciéndose un error de especificación que se manifiesta en residuos heteroscedásticos, también denominado extra-variabilidad.

Así, en primer lugar:

- Función varianza:

$$\text{Var}(O_i) = \phi E(O_i) \quad [1b]$$

Cuando se disponen de datos espaciales es importante distinguir dos fuentes de extra-variación. En primer lugar, debe asumirse la existencia de extra-variación independiente e incorrelacionada espacialmente, denominada ‘heterogeneidad incorrelacionada’ o ‘no espacial’. En segundo lugar, la fuente más importante de extra-variación, es la denominada ‘heterogeneidad correlacionada’ o ‘espacial’ y es consecuencia de la correlación de la unidad espacial con unidades espaciales vecinas. Lo más habitual es que esta dependencia sea debida, en gran parte, a la existencia de variables no controladas, es decir no incluidas en el análisis.

2.4.2.–Suavización de Razones Estandarizadas

2.4.2.1.– Suavización mediante Modelos Lineales Generalizados Mixtos (GLMM)

Con el fin de solucionar los problemas derivados de la utilización directa de las RIE se han utilizado modelos mixtos (ver Anexo 2). La idea es introducir en el modelo uno o más efectos aleatorios que recojan la variabilidad no explicada.

En este caso se tiene que especificar un modelo lineal generalizado mixto (GLMM) para el riesgo relativo:

- Función vínculo:

$$\text{Ln}(E(O_i)) = \beta_{0i} + \text{Ln}(E_i)$$

- Función varianza:

$$\text{Var}(O_i) = \phi E(O_i)$$

[2]

- Efecto aleatorio:

$$\beta_{0i} = \beta_0 + v_i \quad v_i \rightarrow iidN(0, \sigma_v^2)$$

El efecto aleatorio recoge tanto la heterogeneidad (extra-variación) espacial como la no espacial. Es importante señalar que, como se ha supuesto, la heterogeneidad suele considerarse (espacialmente) específica. Es decir, se supone que pueden existir agrupaciones de exceso riesgo y que éstas, por tanto, pueden estar localizadas en alguna(s) zona(s) del mapa.

2.4.2.2.- Suavización mediante Modelos Lineales Generalizados Mixtos (GLMM) con efectos espaciales

- Modelo de Besag, York y Mollié (BYM)

Lo más habitual para suavizar las razones, es utilizar dos efectos aleatorios (Besag, York y Mollie, 1991). El modelo se denomina '*convolution prior*' (Besag, York y Mollie, 1991; Mollie, 1996).

Uno de los efectos aleatorios recoge la heterogeneidad espacial (S_i), y el otro la extravariabilidad (no espacial) desestructurada (v_i).

-Función vínculo

$$\ln(E(O_i)) = \beta_{0i} + S_i + \ln(E_i)$$

[3]

-Función varianza

$$\text{Var}(O_i) = \phi E(O_i)$$

-Efecto aleatorio

$$\beta_{0i} = \beta_0 + v_i \quad S_i$$

Respecto a la distribución de los efectos aleatorios, el efecto no espacial se distribuye (idéntica e independientemente) como una normal, $v_i \rightarrow iidN(0, \sigma_v^2)$.

Para la componente espacial, S_i , la aproximación más habitual (y la computacionalmente) es la conocida como modelo condicional autorregresiva, CAR. Como idea general, se asume una dependencia espacial por la que las estimaciones del riesgo en un área determinada dependen de las áreas vecinas. De forma que los pesos (o ponderaciones espaciales) son máximos (igual a la unidad) si las áreas son adyacentes y mínimos (igual a cero) en otro caso.

Siguiendo a Kelsall y Wakefield (2002),

$$S_i | S_{-i} \rightarrow N(\bar{S}_i, \omega_S^2 / m_i)$$

donde $S_{-i} = (S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_n)$; $\bar{S}_i = \sum_{j \in \delta_i} S_j$; δ_i es el conjunto de municipios 'vecinos' (adyacentes) en el municipio i ; y m_i es el número de estos vecinos.

A menudo, la distribución de los casos suele ser más dispersa que la distribución de Poisson. Es decir, ciertas áreas no presenten ningún caso de alguna de las neoplasias de interés medioambiental, por lo que esta neoplasia no se distribuye exactamente como una Poisson.

De hecho, se podrían distinguir entre ‘ceros estructurales’, los cuales son inevitables, y ‘ceros muestrales’, que ocurren por azar. En epidemiología ambiental los ‘ceros’ suelen ser muestrales, por cuanto no existe ninguna zona libre de riesgo.

A veces, también es posible que el número de ‘ceros’ observado sea menor que el esperado, pero suele ser una situación mucho menos frecuente en la práctica.

En todo caso, deben utilizarse modelos alternativos al de Besag, York y Mollié, basada, como se sabe, en la distribución de Poisson estándar. Las dos aproximaciones más importantes se citan a continuación:

– **Modelos de mixturas finitas. Modelos ZIP y modelos ‘hurdle’**

Los modelos de mixturas finitas (*fnite mixture models*) suponen que la población está compartimentada en k componentes homogéneos. Cada componente tiene un nivel de riesgo diferente. Además, este riesgo puede depender de diferentes covariables.

Lo más habitual es que existan dos componentes homogéneos, uno para los ceros (casos no observados) y el otro para aquellos valores que no sean cero (valores positivos a partir de ahora), conduciendo así a una distribución mixta de Poisson. La distribución que más se utiliza como distribución mixta es la binomial negativa (NB) (Greenwood and Yule, 1920; Lawless, 1987).

Sea Y la variable dependiente que toma valores $0,1,2,\dots$, entonces (condicionado a u) $Y \rightarrow Poisson(\mu u)$, la variable aleatoria no observable u , distribuida como gamma con

valor esperado μ y varianza $\alpha = \frac{1}{\theta}$, representa la heterogeneidad no observada. La media y la varianza condicionales se especifican como sigue:

$$E(Y|x) = \mu \quad \text{Var}(Y|x) = \mu + \alpha\mu^{2-k}$$

En la mayoría de los estudios, el parámetro k se supone fijo (Deb y Trivedi, 2002). Existen dos casos particulares. El modelo NB1 se obtiene especificando $k=1$ mientras que el NB2 se obtiene cuando $k=0$.

Feller (1943) y Mullahy (1997) demostraron que la probabilidad de cero en la distribución de Poisson mixta es mayor que la probabilidad calculada a partir de la distribución de Poisson estándar, teniendo en cuenta la misma media. De hecho, otros autores proponen distribuciones diferentes para modelizar variables con excesos de ceros, por cuanto, a diferencia de la binomial negativa, permiten más de una moda, además de la situada en el cero. Hinde y Demétrio (1998) proponen una Neyman Tipo A y una distribución Pólya–Aeppli.

Cuando los ceros sean estructurales (en una proporción w), la distribución mixta presenta una forma extrema, conduciendo a modelos mucho más restrictivos que los modelos generales de mixturas finitas.

Los modelos ZIP

En los modelos *Zero-inflated Poisson* (ZIP) (Mullahy, 1986; Heilbron 1989 y 1994; Lambert, 1992) la variable respuesta se modeliza como una mixtura de las distribuciones de Bernouilli y de Poisson.

Lambert (1992) especifica el modelo ZIP de la siguiente forma:

$$\log(\mu) = X\beta \qquad \log\left(\frac{\omega}{1-\omega}\right) = Z\gamma$$

en la que X y Z son matrices de variables explicativas (que pueden o no coincidir); ω es la proporción de ceros; y β y γ son vectores de parámetros.

Cuando las variables explicativas coinciden, se puede especificar un modelo más parsimonioso suponiendo que los predictores lineales están relacionados de alguna manera. En este sentido, el modelo $ZIP(\tau)$ se podría especificar como:

$$\log(\mu) = X\beta \qquad \log\left(\frac{\omega}{1-\omega}\right) = \tau X\beta$$

en el que τ es un parámetro de escala. Esto implica que $\omega = \frac{1}{1 + \mu^{-\tau}}$.

Se han propuesto modelos *Zero Inflated* para otras distribuciones de variables aleatorias discretas alternativas a la de la Poisson. Por ejemplo Gupta et al (1996) proponen un modelo *Zero Inflated* para la distribución de Poisson generalizada. Greene (1994) especifica un modelo *Zero inflated* para una binomial negativa.

Los modelos Hurdle

Estos modelos especifican dos procesos generadores de datos distintos, uno para los ceros y otro para los positivos. En la literatura econométrica este tipo de modelos se denomina *hurdle models*, (Mullahy, 1986) o también modelos en dos partes (Heilbron, 1994).

Habitualmente la distribución de los valores positivos es la forma truncada (en cero) de una distribución discreta estándar, como podría ser la Poisson (Cameron y Trivedi, 1998) o la distribución binomial negativa (Deb y Trivedi, 2002).

Si se utiliza la distribución de Poisson truncada, el modelo podría ser una simple reparametrización de un modelo ZIP. Sin embargo, en el contexto de los modelos de regresión, los modelos hurdle y ZIP no son tan equivalentes (Ridout et al., 1998). Dalrymple et al. (2003) señalan que los modelos hurdle mantienen separados, estrictamente, los componentes cero y no cero. Este hecho, los diferencia de los modelos ZIP en los que los valores cero podrían ser generados como una distribución de Poisson.

Cuando las variables explicativas son las mismas en los dos procesos, Mullahy (1986) propone el siguiente modelo 'compatible' (porque, según Heilbron, 1994, la probabilidad de tener un valor cero es compatible con la distribución de valores positivos):

$$\log(\mu) = X\beta \qquad \log(-\log(1 - \pi_+)) = X\gamma$$

Siendo $\pi_+ = 1 - \pi_0$ la probabilidad de generar un valor positivo (π_0 es la probabilidad de ocurrencia de un cero).

Este modelo se reduce a uno de Poisson estándar cuando $\beta = \gamma$.

- Modelo mixto de riesgo relativo

Los modelos vistos hasta ahora están basados en técnicas de suavización. El problema es que todos ellos suelen suavizar en exceso, en el sentido que no mantienen aquellas discontinuidades de la superficie de riesgo, como los ceros, que quizás deberían ser tenidas en cuenta (Lawson et al., 2003).

En este sentido, Lawson y Clark (2002) propusieron un modelo mixto que permite, de forma simultánea, la suavización y la representación de discontinuidades. En

particular, el modelo asume que el riesgo relativo (logarítmico) puede ser descompuesto en tres componentes. Un componente fijo que representa la heterogeneidad no estructurada (ν_i), el cual mide la sobredispersión en una área particular y dos componentes aleatorios (S_i y φ_i) que representan aspectos de la correlación espacial, el efecto aleatorio espacial (S_i) y un efecto aleatorio que permite las discontinuidades o los ‘saltos’ (φ_i).

$$\ln(E(O_i)) = \beta_{0i} + p_i S_i + (1 - p_i) \varphi_i + \ln(E_i)$$

El modelo BYM no es más que un caso particular cuando $p_i = 1 \forall i$, mientras que el modelo se convierte en uno de ‘saltos’, cuando $p_i = 0 \forall i$.

Para el efecto aleatorio espacial (S_i), se sigue la especificación CAR del modelo BYM, y para el efecto de ‘salto’ (φ_i), se especifica una distribución que depende de la diferencia (en términos absolutos) entre áreas vecinas (contiguas).

$$\pi(\varphi_1, \varphi_2, \dots, \varphi_m) \propto \frac{1}{\sqrt{\lambda}} \exp\left(-\frac{1}{\lambda} \sum_{i < j} |\varphi_i - \varphi_j|\right)$$

El parámetro λ restringe el modelo y se distribuye como una gamma.

– Estimación de los modelos

Los modelos se suelen estimar siguiendo una aproximación bayesiana.

Se trata de fijar *distribuciones a priori* para los efectos aleatorios, CAR para S_i ; normal (intercambiable) para ν_i ; gamma para φ_i ; así como para el resto de parámetros (tanto de la función vínculo, de la función varianza, como para los parámetros de las distribuciones de probabilidad de los efectos aleatorios).

La estimación de los parámetros desconocidos se realiza a partir de cadenas Markovianas de Monte Carlo, MCMC (Gilks et al., 1996; y, sobre todo, Spiegelhalter et al., 2003) (véase Anexo 3).

2.4.3.-Estructura de dependencia espacial

2.4.3.1.-Modelos Autorregresivos de series temporales (AR)

Antes de abordar el estudio de las estructuras de dependencia espacial se procederá, siguiendo a Saez y Saurina (2006), al análisis de los modelos autorregresivos aplicados a datos temporales. Su estructura y sus propiedades ayudarán al conocimiento del tipo de modelos que serán útiles para modelizar datos correlacionados espacialmente.

Se denomina modelo autoregresivo de orden uno, AR(1), al modelo que expresa la variable dependiente en función de un único retardo temporal de la misma variable:

$$Z_i = \rho Z_{i-1} + \varepsilon_i$$

donde $\varepsilon_i \sim N(0, \sigma^2)$, ε_i independiente de ε_j cuando $i \neq j$, y ε_i independiente de cualquier Z_j donde $j \leq i$.

La sustitución progresiva del retardo en función de retardos anteriores sería de la siguiente manera:

$$\begin{aligned} Z_i &= \rho(\rho Z_{i-2} + \varepsilon_{i-1}) + \varepsilon_i = \rho^2 Z_{i-2} + \rho \varepsilon_{i-1} + \varepsilon_i = \rho^2(\rho Z_{i-3} + \varepsilon_{i-2}) + \rho \varepsilon_{i-1} + \varepsilon_i = \\ &= \rho^3 Z_{i-3} + \rho^2 \varepsilon_{i-2} + \rho \varepsilon_{i-1} + \varepsilon_i = \dots = \rho^k \varepsilon_{i-k} + \rho^{k-1} \varepsilon_{i-k+1} + \rho^{k-2} \varepsilon_{i-k+2} + \dots + \rho^2 \varepsilon_{i-2} + \rho \varepsilon_{i-1} + \varepsilon_i \end{aligned}$$

Es decir se encuentra una forma recursiva simple y es puede expresarlo como un modelo media móvil infinito dependiendo de un único parámetro (MA(∞)).

Si se supone que inicialmente $Z_1 = \varepsilon_1$, el modelo se definiría de la siguiente manera:

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ \vdots \\ Z_n \end{pmatrix} = \rho \begin{pmatrix} 0 \\ Z_1 \\ Z_2 \\ Z_3 \\ \vdots \\ Z_{n-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

o en forma matricial $\mathbf{Z} = \mathbf{B}\mathbf{Z} + \boldsymbol{\varepsilon}$ donde $\mathbf{B} = \rho\mathbf{H}$ y $\mathbf{H} = \begin{pmatrix} \mathbf{0}' & \mathbf{0} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}$.

\mathbf{H} es la matriz que recoge la estructura de vecindad del modelo que en el caso que ocupa, el estudio de una serie temporal, recoge la estructura del retardo que se ha planteado.

Si se quiere considerar un valor μ distinto de cero para la media el modelo se escribirá de manera general como: $\mathbf{Z} - \boldsymbol{\mu} = \mathbf{B}(\mathbf{Z} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}$.

Asimismo los momentos de la distribución conjunta de \mathbf{Z} pueden expresarse como:

$$E(\mathbf{Z}) = \boldsymbol{\mu},$$

$$\text{var}(Z_i) = \frac{\sigma^2}{1 - \rho^2},$$

$$\text{cov}(Z_i, Z_{i-t}) = \frac{\sigma^2}{1 - \rho^2} \rho^t;$$

es decir:

$$\text{var}(\mathbf{Z}) = \boldsymbol{\Sigma} = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{n-3} \\ \rho^3 & \rho^2 & \rho & 1 & \dots & \rho^{n-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 1 \end{pmatrix}$$

Así pues, la matriz de varianzas covarianzas queda absolutamente determinada, a diferencia, como se verá de lo que ocurre en el caso de la dependencia espacial.

Bajo el supuesto de normalidad se puede expresar de manera conjunta al modelo como: $\mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

La expresión del modelo en forma condicional puede expresarse como:

$$f(Z_i | Z_{i-1}, Z_{i-2}, \dots, Z_1) = f(Z_i | Z_{i-1}) = N(\rho Z_{i-1}, \sigma^2).$$

Así pues el modelo se puede especificar de tres maneras equivalentes:

- 1.- Especificación simultánea: $Z_i = \rho Z_{i-1} + \varepsilon_i$.
- 2.- Especificación condicional: $f(Z_i | Z_{i-1}) = N(\rho Z_{i-1}, \sigma^2)$.
- 3.- Especificación conjunta: $\mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Tal y como se puede apreciar, las especificaciones simultánea y condicional permiten reconocer la estructura de vecindad, cuestión clave en el estudio y modelización de datos espaciales, que sería imposible reconocer a través de $\boldsymbol{\Sigma}$. Para trabajar los modelos espaciales deben utilizarse estas especificaciones (simultánea y/o condicional).

2.4.3.2.- Autoregresión Espacial Simultánea (SAR)

Corresponde a la denominada especificación simultánea vista en el apartado anterior.

Muy parecido al modelo AR(1) en el análisis temporal con la particularidad que ahora la dependencia de $Z(\mathbf{s})$ será una función de las observaciones en puntos vecinos en lugar de ser función del pasado. La especificación es:

$$Z_i = \mu_i + \sum_{j=1}^n b_{ij}(Z_j - \mu_j) + \varepsilon_i$$

donde μ_i es la media de Z_i , ε_i son los errores aleatorios no correlacionados y con media cero y b_{ij} es un peso (ponderación) que mide el grado con el que Z_i depende de Z_j . Se asume que si Z_i depende positivamente de Z_j , $b_{ij} > 0$, y si dos ubicaciones no dependen entre sí $b_{ij} = 0$. Por convenio se asigna $b_{ii} = 0$.

La matriz de pesos permitirá recoger la dependencia espacial observada en los datos y su definición dependerá del tipo de análisis que se dé a los datos. En general la matriz de pesos \mathbf{B} no necesita ser simétrica, es decir b_{ij} puede ser distinto que b_{ji} . Aunque la matriz de covarianza debe ser simétrica, pues la correlación espacial entre dos ubicaciones debe ser la misma, podemos interpretar b_{ij} 's como coeficientes de predicción. Es evidente que la predicción sobre j desde un punto i puede ser distinta que la predicción contraria aunque no sea más que por considerar que el número de vecinos en cada sitio es distinto. Aunque se podría pensar en estimar todos estos pesos en general se plantea alguna manera de especificar \mathbf{B} para reducir el número de parámetros desconocidos.

Primero pues, se debe definir una matriz de vecindad \mathbf{H} que debe ser una matriz simétrica, y después se puede ponderar de alguna manera. Por ejemplo

$$w_{ij} = \begin{cases} 1 & \text{si la localización } i \text{ es un vecino de la localización } j \\ 0 & \text{si la localización } j \text{ no es un vecino de la localización } i \\ 0 & i = j \end{cases}$$

y $\mathbf{B} = \rho\mathbf{W}$ donde $\mathbf{W} = \mathbf{H}$ es una matriz con w_{ij} en la i -ésima fila y la j -ésima columna. Esta es la manera más sencilla de especificar la relación.

En notación matricial el modelo SAR puede expresarse como:

$$\begin{aligned} \mathbf{Z} - \boldsymbol{\mu} &= \mathbf{B}(\mathbf{Z} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon} \\ (\mathbf{I} - \mathbf{B})(\mathbf{Z} - \boldsymbol{\mu}) &= \boldsymbol{\varepsilon}. \end{aligned}$$

Con los resultados siguientes:

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu}, \\ \text{var}(\mathbf{Z}) &= (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Lambda} (\mathbf{I} - \mathbf{B}')^{-1}, \\ \text{cov}[\boldsymbol{\varepsilon}, \mathbf{Z}] &= \text{cov}[\boldsymbol{\varepsilon}, (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\varepsilon}] = (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Lambda}, \end{aligned}$$

donde $\boldsymbol{\Lambda} = \text{var}(\boldsymbol{\varepsilon})$ es una matriz diagonal. En este caso $\text{cov}[\boldsymbol{\varepsilon}, \mathbf{Z}]$ no es una matriz diagonal, es decir ε_i no es independiente de la variable Z_j en otra posición. Este hecho marca la principal diferencia respecto el modelo AR(1) de series temporales en la que los errores no están correlacionados con el pasado.

\mathbf{Z} es pues una combinación lineal de los errores ε_i . Si los errores siguen una distribución normal con media cero y matriz de varianzas covarianzas, $\boldsymbol{\Lambda}$, entonces

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Lambda} (\mathbf{I} - \mathbf{B}')^{-1})$$

2.4.3.3.- Autorregresión Espacial Condicional (CAR)

Este modelo especifica la distribución de Z en una posición a partir del conocimiento de la misma en otras posiciones.

Es decir, a partir de $\mathbf{z}_{-i} \{z_j, j \neq i\}$, se quiere conocer la distribución condicional $f(z_i | \mathbf{z}_{-i})$.

Usando la notación similar a los modelos de series temporales se escribirá;

$$E(Z_i | \mathbf{z}_{-i}) = \mu_i + \sum_{j=1}^n c_{ij} (z_j - \mu_j)$$

Si τ_i^2 es la varianza condicional de Z_i conocido \mathbf{z}_{-i} , debe verificarse que $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$, $c_{ii}=0$ y $c_{ij}=0$ para pares de ubicaciones independientes entre si. Si se puede considerar constante la varianza condicional entonces $\tau_i^2 = \tau^2$ y es suficiente que $c_{ij} = c_{ji}$. Si se denomina $\mathbf{C} = [c_{ij}]$ y $\mathbf{M} = \text{diag}(\tau_i^2)$, el modelo CAR puede ser especificado como

$$\mathbf{Z} - \boldsymbol{\mu} = \mathbf{C}(\mathbf{Z} - \boldsymbol{\mu}) + \boldsymbol{\delta}$$

donde $\boldsymbol{\delta} = (\mathbf{I} - \mathbf{C})(\mathbf{Z} - \boldsymbol{\mu})$ es un vector de errores.

De forma equivalente:

$$Z_i = \mu_i + \sum_{j=1}^n c_{ij}(Z_j - \mu_j) + \delta_i$$

Los parámetros del modelo son: $\boldsymbol{\mu}$, \mathbf{C} y \mathbf{M} .

Así $\text{cov}(\boldsymbol{\delta}, \mathbf{Z}) = \text{cov}[(\mathbf{I} - \mathbf{C})(\mathbf{Z} - \boldsymbol{\mu}), \mathbf{Z}] = (\mathbf{I} - \mathbf{C})\text{var}(\mathbf{Z})$. Si la distribución condicional $f(z_i | \mathbf{z}_{-i})$ es gaussiana, se puede escribir

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M})$$

Bajo normalidad se verifica que $\text{cov}(\boldsymbol{\delta}, \mathbf{Z}) = (\mathbf{I} - \mathbf{C})\text{var}(\mathbf{Z}) = \mathbf{M}$, que además es matriz diagonal lo que conlleva que el error δ_i es independiente de las respuestas Z_j en otras posiciones que lo hace parecido a los modelos de series temporales pero distinto a SAR.

En resumen, considerando la distribución condicional $f(z_i | \mathbf{z}_{-i}) = N(0, \sigma_\delta^2)$, y tomando $\mathbf{C} = \rho\mathbf{H}$ para el modelo CAR, la matriz de varianzas-covarianzas para las distribuciones conjuntas de los modelos será:

$$\text{SAR: } \boldsymbol{\Sigma}_S = \text{var}(\mathbf{Z}) = \sigma_\epsilon^2 (\mathbf{I} - \rho\mathbf{H})^{-1} (\mathbf{I} - \rho\mathbf{H})^{-1}$$

$$\text{CAR: } \boldsymbol{\Sigma}_C = \text{var}(\mathbf{Z}) = \sigma_\delta^2 (\mathbf{I} - \rho\mathbf{H})^{-1}$$

2.4.3.4.– Autorregresión Espacial Condicional L1 (L1-CAR)

Algunos autores señalan que la utilización del CAR estándar, al promediar áreas vecinas, puede suavizar en exceso (Richardson et al, 2004). Proponen, de forma alternativa, utilizar una distribución doble exponencial (de hecho, ya fue propuesta por Besag et al. (1991)). Es equivalente a realizar un suavizado local basado en la mediana (norma L1), en lugar de un suavizado basado en la media como en el CAR estándar.

2.4.3.5.– Matriz de ponderaciones espaciales

En todos estos modelos, la elección de la matriz de pesos (c_{ij} para CAR y b_{ij} para SAR) y, por lo tanto la especificación de las ponderaciones requiere un conocimiento a priori del rango (hasta donde están correlacionados en el espacio las áreas) y de la intensidad (en que magnitud están relacionados los áreas vecinas) de la matriz espacial de covarianzas (Griffith, 1995).

Se suelen utilizar, como especificaciones más habituales: la estandarización por filas, la dimensión del vector de vecinos comunes y distintas funciones de la distancia.

Siguiendo a Cressie (1993), y considerando rejillas irregulares, la especificación de la ponderación la realiza en función de la distancia.

Tomando como referencia el modelo CAR se tiene que:

$$c_{ij} = \begin{cases} \rho(d_{ij}^{-k} / C(k)) \left(\frac{n_j}{n_i} \right)^{1/2} & j \in N_i \\ 0 & \text{otro caso} \end{cases}$$

donde ρ es una constante a estimar; d_{ij} es la distancia entre los vecinos i y j ; n_i es el número de casos existentes en el área i ; $C(k) = \max\{d_{ij}^{-k} : j \in N_i; i = 1, \dots, n\}$ (un factor

de escala); y N_i es el conjunto de vecinos para el área i . Cressie (1993) probó diferentes valores para k y encontró que los mejores resultados se obtenían cuando $k=1$. Así, la especificación se simplifica:

$$g_{ij} = \begin{cases} \left(\min\{d_{ij} : i = 1, \dots, n\} / d_{ij} \right) \left(\frac{n_j}{n_i} \right)^{1/2} & j \in N_i \\ 0 & \text{otro caso} \end{cases}$$

2.5.- Variables explicativas

2.5.1.- Datos en rejilla

Como ejemplo de variables explicativas con un diseño de datos en rejilla se incluyen en el Anexo 4 tres artículos de Lertxundi–Manterola et al (2005 y 2006), “Construcción de un índice de privación material para los municipios de la Regió Sanitària Girona” (*Estudios de Economía Aplicada*, 2005); “Influencia de la privación sobre la distribución geográfica de la incidencia de cáncer en la Regió Sanitària Girona” (*Economía y Salud*, 2006); y “Área deprivation a mortality in capital cities of Andalusia and Catalonia (Spain)” (Enviado a *Journal of Epidemiology and Community Health*).

En todos ellos se construye un índice de privación, con un valor distinto para cada municipio (los dos primeros) y sección censal (el último), el cual se relaciona con la incidencia y la mortalidad. El problema, como se verá más adelante en el apartado 2.6, se reduce al de una regresión ecológica (múltiple).

2.5.2.- Análisis de datos geoestadísticos

Las variables explicativas de interés en este trabajo son medioambientales, sobre todo relacionadas con la contaminación atmosférica. Los datos correspondientes a estas variables presentan un diseño geoestadístico.

2.5.2.1.- Análisis de la dependencia espacial

Los instrumentos para analizar la dependencia espacial en datos geoestadísticos son las funciones de covarianza (covariograma) y el semivariograma.

Sea $Z(s)$, es un vector aleatorio en la localización $s \in D$. El momento de primer orden de $Z(s)$ es la esperanza matemática definida como:

$$m(s) = E[Z(s)]$$

Los momentos de segundo orden son:

- La varianza, $\sigma^2(x) = Var[Z(s)] = E\left[\{Z(s) - m(s)\}^2\right]$

- La covarianza o autocovarianza, $C(s_i, s_j) = E\left[\{Z(s_i) - m(s_i)\}\{Z(s_j) - m(s_j)\}\right]$

- El semivariograma, $\gamma(s_i, s_j) = \frac{1}{2} E\left[\{Z(s_i) - Z(s_j)\}^2\right]$

En analogía a las series temporales, una variable aleatoria (espacial) es estrictamente estacionaria si su función de distribución es invariante a cualquier traslación respecto a un vector h (es decir, en el espacio). En la práctica, sin embargo, no se exige que las variables aleatorias espaciales sean estrictamente estacionarias sino que basta que lo sean débilmente. En el caso de la estadística espacial se dice que la variable aleatoria debe cumplir la hipótesis intrínseca.

Las variables aleatorias intrínsecas son aquellas que cumplen las siguientes condiciones:

$$\begin{aligned} E[Z(s+h) - Z(s)] &= 0 \\ \text{Var}[Z(s+h) - Z(s)] &= 2\gamma(h) \end{aligned}$$

Suponiendo la estacionariedad de la variable aleatoria se puede definir entonces:

- El semivariograma, $\gamma(h) = \frac{1}{2} \text{Var}[Z(s) - Z(s+h)] = \frac{1}{2} E[\{Z(s) - Z(s+h)\}^2]$ (Isaaks et al. 1989)

Existe una relación directa entre semivariograma y la función de covarianza.

$$\gamma(h) = C(0) - C(h)$$

Por tanto resulta suficiente usar una de las dos funciones para caracterizar la dependencia espacial, aunque habitualmente se utiliza el variograma.

El estimador más común del semivariograma viene dado por

$$\tilde{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i+h) - Z(s_i)]^2$$

donde $N(h)$ es el número de pares $Z(s_i)$ y $Z(s_i+h)$ separados a una distancia $h = |h|$.

Se trata de un estimador no paramétrico y óptimo, cuando se dispone de una malla regular de muestreo que sea representativa y la distribución sea normal.

En la práctica este estimador produce variogramas erráticos, debido a desviaciones de las condiciones ideales vistas arriba. Además, debido a que $\tilde{\gamma}(h)$ es esencialmente una media muestral, el estimador tiene todas las desventajas comúnmente asociadas a la misma, en particular la no robustez (Cressie y Hawkins, 1980).

Una alternativa robusta fue propuesta por Cressie y Hawkins (1980):

$$\gamma^*(h) = \frac{1}{2} (0.457 + 0.494/N(h))^{-1} \left[\frac{1}{N(h)} \sum_{i=1}^{N(h)} (|Z(s_i + h) - Z(s_i)|)^{1/2} \right]^4$$

Aunque el estimador de Cressie y Hawkins se considera óptimo en condiciones de normalidad, suele infravalorar los valores atípicos.

Como alternativa se han propuesto otros estimadores alternativos, denominados estimadores ponderados, los cuales se basan bien en la función de distribución acumulada o bien en la de densidad. La experiencia práctica con estos estimadores, sin embargo, es muy reducida

▪ **Formas generales del semivariograma muestral.**

En sentido amplio se considera que hay dos tipos principales de semivariogramas. En el primer tipo, la semivarianza aumenta conforme lo hace el valor absoluto del intervalo $|h|$ hasta alcanzar un valor máximo a partir del cuál se mantiene relativamente constante y oscila alrededor del mismo. Estos semivariogramas son conocidos como de tipo transitivo.

El valor del intervalo a partir del cual el semivariograma no se incrementa es conocido como alcance o rango y marca el límite de la dependencia espacial. La varianza máxima es conocida como 'sill' o meseta del semivariograma y teóricamente debe

coincidir con la varianza a priori σ^2 de la muestra de la variable aleatoria $Z(s)$ (ver Figura 1).

Cuando la variable presenta este tipo de semivariograma cumple no sólo la hipótesis intrínseca (es débilmente estacionaria), sino también es estacionaria de segundo orden. Es decir, su valor esperado es constante;

$$E[Z(s)] = m; \forall s$$

la función de covarianza existe y está definida por,

$$\begin{aligned} C(h) &= \text{Cov}[Z(s+h), Z(s)] \\ &\equiv E[\{Z(s) - m\}\{Z(s+h) - m\}] \end{aligned}$$

y el semivariograma está relacionado con las funciones de covarianza y de autocorrelación como sigue:

$$\begin{aligned} \gamma(h) &= C(0) - C(h) \\ \rho(h) &= C(h) / C(0) = 1 - \gamma(h) / (C(0)) \end{aligned}$$

donde $\sigma^2 = C(0)$, es la varianza de la muestra, también conocida como varianza a priori de la variable aleatoria $Z(s)$.

El segundo tipo de variograma aparenta un incremento sin límites, por lo que se denominan no acotados (ver Figura 1).

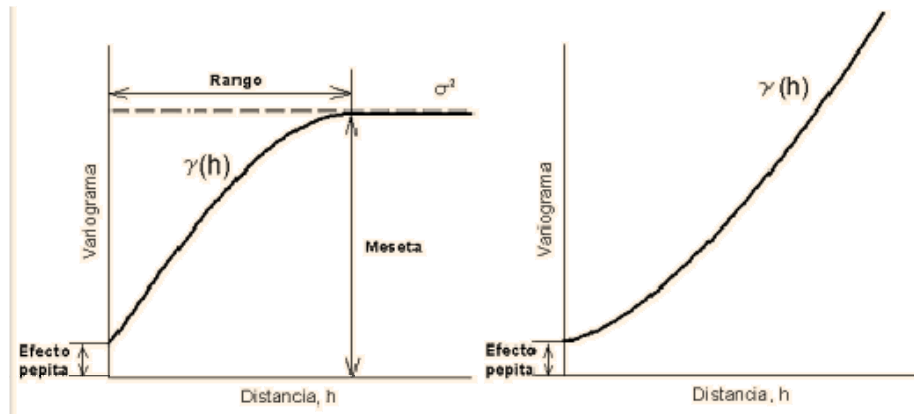


Figura 1.– Los dos tipos variogramas; con meseta, transitivo (izquierda) y sin meseta, no acotado (derecha).

Fuente: Construcción propia.

Por definición debe cumplirse $\gamma(0) = 0$, pero en la práctica el semivariograma muestral ($\gamma^*(h)$) no necesariamente se anula cuando h tiende a cero.

Esto es conocido como efecto '*nugget*' o pepita, y el valor del semivariograma en cero $\gamma(0)$ es conocido como la varianza '*nugget*' o microvarianza (ver Figura 1). En principio esto puede ocurrir solamente si existen discontinuidades en la variable aleatoria. En la práctica su existencia se debe a variación espacial que no puede explicar el variograma, a fenómenos de escala y a errores de medida.

▪ Condiciones que deben cumplir los modelos del semivariograma

En resumen, en la selección de una función adecuada para ser ajustada a un semivariograma muestral, se debe tener en cuenta en la mayoría de los casos hasta tres elementos: la ordenada en el origen, una sección monótonamente creciente y una meseta. Sin embargo, no servirá cualquier función debido a la siguiente razón.

Supóngase que $Z(s)$ es una variable aleatoria estacionaria de segundo orden de la cual se obtiene la variable regionalizada $\{Z(s_i), i = 1, \dots, n\}$, es decir distribuida espacialmente, con semivariograma y función de covarianzas, $\gamma(h)$ y $C(h)$ respectivamente.

Sea Y una combinación lineal de $Z(s_i)$ tal que

$$Y = \sum_{i=1}^n \lambda_i Z(s_i)$$

donde $\lambda_i, i=1, \dots, n$, son ponderaciones arbitrarias.

La variable Y también es una variable regionalizada con varianza:

$$\text{Var}[Y] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i, s_j)$$

que como se sabe la cual no puede ser negativa en ningún caso. Así la función de covarianza en la parte derecha de la expresión anterior debe asegurar que esta condición se cumpla.

La matriz de covarianzas

$$\begin{bmatrix} C(s_1, s_1) & C(s_1, s_2) & \dots & C(s_1, s_n) \\ C(s_2, s_1) & C(s_2, s_2) & \dots & C(s_2, s_n) \\ \dots & \dots & \dots & \dots \\ C(s_n, s_1) & C(s_n, s_2) & \dots & C(s_n, s_n) \end{bmatrix}$$

debe ser definida positiva, es decir, el determinante y todos sus menores principales son positivos o cero. Aquí las covarianzas están definidas como $C(s_i, s_j) = \text{Cov}[Z(s_i), Z(s_j)]$ $i, j = 1, \dots, n$

La función de covarianzas $C(h)$ si existe debe ser semidefinida positiva por lo que serán permisibles funciones que cumplan este criterio (Diaz, 2002).

Las variables que no tengan covarianzas a priori finitas no tendrán definida la función de covarianzas. En el caso de que cumplan la hipótesis intrínseca, entonces estará definido el semivariograma y se cumplirá que:

$$\text{Var}[Y] = C(0) \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(s_i, s_j)$$

Si las ponderaciones λ_i suman 0, es decir $\sum_{i=1}^n \lambda_i = 0$ entonces:

$$\text{Var}[Y] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(s_i, s_j)$$

Esto implica que $-\gamma(h)$ debe ser condicionalmente semidefinida positiva, con la condición $\sum_{i=1}^n \lambda_i = 0$. Esto es equivalente a decir que $\gamma(h)$ debe ser condicionalmente semidefinida negativa.

Como consecuencia, se puede demostrar que el semivariograma debe tener un ritmo de crecimiento inferior a h^2 , es decir se debe cumplir que:

$$\lim_{h \rightarrow \infty} \frac{\gamma(h)}{h^2} = 0$$

Cuando no se satisface esta condición puede ser un indicador de que la función aleatoria no sea estacionaria.

Por otro lado de la definición de semivariograma se infiere que éste debe anularse en el origen, es decir $\gamma(0) = 0$.

Las funciones que cumplen las condiciones anteriores son conocidas como modelos autorizados del semivariograma.

2.5.2.2.- Modelos autorizados para el semivariograma

La variabilidad espacial puede ser diferente dependiendo de la dirección.

a.- Modelos isotrópicos

Los modelos isotrópicos son aquellos en los que la dependencia espacial no varía dependiendo de la dirección espacial que se tome.

Como se dijo, según la forma del semivariograma estos modelos se pueden dividir en transitivos, o acotados, y no acotados.

▪ Modelos transitivos o acotados

Los modelos de variogramas isotrópicos acotados más utilizados son; el modelo lineal con meseta, modelo circular, modelo exponencial y el modelo Gaussiano (Isaaks y Srivastava, 1989).

- Modelo lineal con meseta

La función de autocorrelación de este modelo es la siguiente:

$$\rho(h) = 1 - h/a$$

donde h representa la distancia y a aquella distancia en la que empieza la meseta.

El semivariograma se expresa como (véase Figura 2):

$$\gamma(h) \begin{cases} S(h/a) & \text{para } 0 \leq h \leq a \\ S & \text{para } h > a \end{cases}$$

El modelo lineal solamente es apropiado para el caso unidimensional.

- Modelo esférico

La función de autocorrelación de este modelo es:

$$\rho(h) = 1 - \frac{3h}{2a} + \frac{1}{2} \left(\frac{h}{a} \right)^3$$

y el semivariograma

$$\gamma(h) = \begin{cases} \frac{S}{2} \left\{ 3 \left(\frac{h}{a} \right) - \left(\frac{h}{a} \right)^3 \right\} & \text{para } 0 \leq h \leq a \\ S & \text{para } h > a \end{cases}$$

$$\text{gradiente} = 3S / (2a)$$

Debe señalarse que el modelo esférico es apropiado para el caso de tres dimensiones aunque se puede aplicar para casos de una y dos.

- Modelo exponencial

En este caso el semivariograma es (véase Figura 2):

$$\gamma(h) = S \left(1 - e^{-\frac{h}{r}} \right) \text{ para } h \geq 0$$

$$\text{gradiente} = S / r$$

Se considera como rango efectivo $a=3r$.

Debe señalarse que los procesos autorregresivos de primer orden (y de Markov) conducen a modelos exponenciales.

- Modelo gaussiano

Viene dado por la expresión:

$$\gamma(h) = S \left(1 - e^{-\left(\frac{h}{r}\right)^2} \right) \text{ para } h \geq 0$$

donde r es un parámetro no lineal que determina la escala espacial de la variación, como en el caso exponencial (ver Figura 2). El rango efectivo se considera $a = \sqrt{3r}$, que corresponde al valor 0.95S del variograma.

La hipótesis subyacente a este modelo es que los fenómenos que representa han de ser infinitamente continuos y diferenciables (sumamente continuos). En muchas ocasiones esta hipótesis es difícil de cumplir.

- Modelo de Gneiting

El modelo de Gneiting (1999) es muy parecido al gaussiano, gráficamente prácticamente no distinguible, pero que no tiene ninguna de las desventajas del modelo gaussiano. En concreto, el modelo gaussiano es muy sensible a la existencia de discontinuidades en los datos (Ribeiro y Diggle, 2001)

El modelo es el siguiente:

$C(h) = (1 + 8sh + 25(sh)^2 + 32(sh)^3)(1 - sh)^8 1_{[0,1]}(sh)$ siendo $C(h)$ la covarianza y

$$s = \frac{10\sqrt{2}}{47}$$

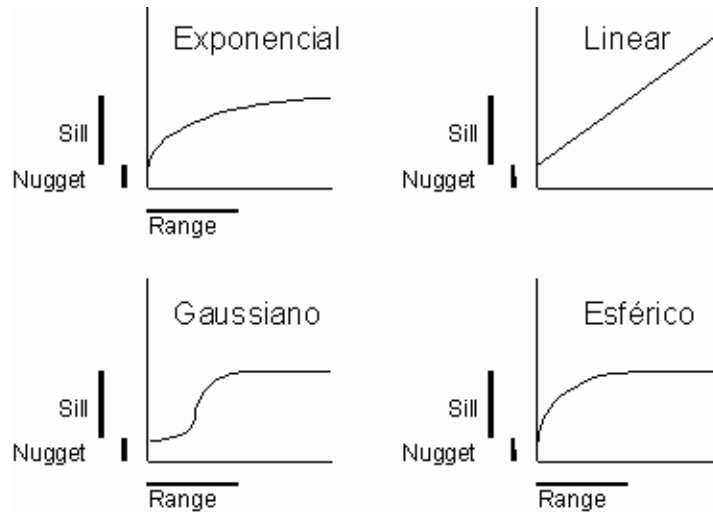


Figura 2. Diferentes modelos isotrópicos acotados (Fuente: Diaz, 2002)

▪ **Modelos no acotados**

- Modelo de Potencia

Existen casos en que la varianza aparenta incrementarse indefinidamente.

De hecho, un movimiento Browniano en una dimensión:

$$Z(s) - Z(s+h) = \varepsilon$$

no es más que una variable aleatoria gaussiana espacialmente independiente.

Su semivariograma es:

$$\gamma(h) = s|h^\theta|; \text{ para } 0 < \theta < 2$$

en la que s se denomina factor de escala.

Este modelo representa fenómenos no estacionarios en los que el comportamiento en el origen depende del valor de θ .

- Modelo de efecto pepita puro

Este modelo representa a un fenómeno completamente aleatorio, en el cual no existe correlación espacial. No importa lo cercano que se encuentren los valores de las variables, siempre estarán incorrelacionados.

$$\gamma(h) = \begin{cases} 0 & \text{si } |h| = 0 \\ s & \text{si } |h| > 0 \end{cases}$$

- Modelo logarítmico (Modelo de Wigsian)

Se define como:

$$\gamma(h) = k \log(h)$$

Puede ser de utilizado cuando el semivariograma experimental se comporta linealmente, utilizando una escala logarítmica para las distancias.

b.- Modelos anisotrópicos

Los modelos hasta ahora descritos son isotrópicos. En la práctica existen numerosas situaciones en que la variación espacial es anisotrópica, por lo que existe un semivariograma diferente para cada una de las direcciones.

En la práctica se estudian cuatro direcciones, estimando los semivariogramas y determinando los rangos para los mismos, y luego se construye el gráfico direccional de los rangos para decidir si hay anisotropía o no.

A veces la anisotropía se puede representar mediante una transformación lineal simple de las coordenadas (en este caso se denomina geométrica o afín).

La transformación se puede expresar como:

$$\Omega(\theta) = \left(A^2 \cos^2(\theta - \phi) + B^2 \sin^2(\theta - \phi) \right)^{1/2}$$

sobre el argumento h de la función semivarianzas en el caso de los modelos transitivos o sobre el gradiente en los modelos no acotados.

En presencia de anisotropía geométrica, el gráfico direccional de los rangos forma una elipse, en la cual el eje menor B es el rango de la dirección de más rápida variación y A, el eje mayor, está en la dirección en que la variabilidad es más lenta. La relación $\lambda = A/B$ es una medida de anisotropía.

2.5.2.3.- Estimación de datos geoestadísticos

2.5.2.3.1.- Análisis univariante

- Métodos de interpolación determinista

El objetivo de los métodos de interpolación espacial es el de estimar datos geoestadísticos en puntos en el dominio espacial que no coinciden con aquellos observados en el mismo. Existen diversas técnicas que difieren en sus hipótesis métodos y/o complejidad.

Análisis tendencial de la superficie (Trend surface analysis)

Es un método sencillo consistente en la aplicación de un análisis de regresión múltiple a fin de obtener una superficie de predicción. Habitualmente la superficie se aproxima por un polinomio, aunque existen gran variedad de funciones alternativas.

Ponderación inversa de la distancia (Inverse Distance Weighted, IDW)

Es el método más utilizado y consiste en una combinación lineal de aquellos puntos observados en el dominio espacial. Como especificidad, el método IDW utiliza como ponderaciones el inverso de la distancia, normalizado de tal forma que la suma de todas las ponderaciones sea igual a la unidad.

$$w(d) = 1/d^p$$

donde w son las ponderaciones de la distancia y el exponente p (en general de 0 a 2, aunque más común 2) se especifica a priori y, al penalizar la distancia debe ser escogido en función de la suavización que se pretenda conseguir.

Splining

Se trata de una técnica determinista para representar curvas de dos dimensiones en superficies de tres dimensiones. Consiste en encontrar un conjunto de funciones que interpolen los puntos observados, minimizando un criterio de suavización (Creutin and Obled, 1982).

En algunas condiciones de las funciones utilizadas y del criterio de suavización finalmente empleado se puede mostrar su equivalencia con el kriging (Mestre, 2000).

- Métodos de interpolación estocástica. Kriging

En los años 50 del siglo XX, el ingeniero sudafricano D. G. Krige (nacido en 1919) diseñó un método de interpolación para determinar la presencia de vetas de oro, a partir de muestras de terreno. La idea básica era que esas predicciones no eran más que medias ponderadas de las cantidades observadas de oro en las muestras. Las ponderaciones de la media dependían de la distancia entre el punto que se quería

predecir y aquellos que se habían observado. Se escogían minimizando la varianza de predicción, es decir que las ponderaciones proporcionaban un estimador ELIO (Estimador, Lineal, Insesgado y Óptimo, *BLUE* en inglés). Por este motivo, el método kriging también se denomina interpolación óptima (Van Beers, 2005).

Así pues, el kriging tiene dos ventajas principales con respecto a los estimadores deterministas: a) las ponderaciones utilizadas en la estimación son determinadas como una función entre la distancia estructural (el semivariograma) de los puntos observados y del que va ser predicho; y, más importante, b) la estimación lleva asociada una cuantificación de incertidumbre, la desviación típica (o la varianza) del kriging (Matheron y Kleingeld, 1987).

Es importante señalar que, como cualquier otro estimador lineal, la realización (muestral) sobre la que se aplique el kriging debe provenir de una variable aleatoria estacionaria.

Sea $Z(s)$ una variable aleatoria, definida en un soporte puntual y estacionario de segundo orden, con:

–un valor esperado,

$$E[Z(s)] = m; \quad \forall s$$

m es una constante generalmente desconocida,

–una función de covarianza centrada

$$C(h) = E[Z(s+h)Z(s)] - m^2$$

– un variograma

$$\text{Var}[Z(s+h) - Z(s)] = 2\gamma(h)$$

Al menos uno de estos dos momentos de segundo orden se supone conocido. En caso de existir el semivariograma únicamente, la variable aleatoria $Z(s)$ debe ser intrínseca.

Denomínense $\{Z(s_i), i=1, \dots, n\}$ los valores observados de la variable aleatoria. Con frecuencia estos valores están definidos en soportes puntuales o casi puntuales, en otros casos son los valores medios $Z_{V_i}(s_i)$ definidos en los soportes V_i centrados en los puntos s_i , donde los n soportes pueden ser todos diferentes.

La estimación del valor medio $Z_{V_i}(s_i)$ en el dominio de V_i se define como:

$$Z_{V_i}(s_i) = \frac{1}{V_i} \int_{V_i} Z(s) ds$$

Es preciso destacar que bajo la hipótesis de estacionariedad el valor esperado es $E[Z(s_i)] = m, \quad \forall i$.

El estimador lineal Z_k^* considerado es una combinación lineal de n valores de datos tal que:

$$Z_k^* = Z^*(s_k) = \sum_{i=1}^n \lambda_i Z(s_i),$$

Las n ponderaciones λ_i se calculan de tal forma que el estimador es insesgado y la varianza de la estimación es mínima, es decir, el estimador Z_k^* es óptimo (ELIO).

– Kriging Simple: Kriging lineal con valores esperados conocidos

El caso más simple se denomina kriging simple y la hipótesis básica es la estacionariedad junto con el hecho de que se asume que la media de la función aleatoria es conocida. Es decir, $E(Z(\mu)) = m$ y m es conocida.

El sistema de ecuaciones del que se deriva el estimador es el siguiente:

$$\begin{cases} \sum_{j=1}^n \lambda_j \sigma_{ij} = \sigma_{i0}, & i = 1, \dots, n \\ \lambda_0 = m(s_0) - \sum_{i=1}^n \lambda_i m(s_i) \end{cases}$$

Siendo el estimador:

$$Z_0^* = \lambda_0 + \sum_{i=1}^n \lambda_i Z(s_i)$$

y su varianza:

$$\sigma_{K_x}^2 = \sigma'_{00} - \sum_{i=1}^n \lambda_i \sigma'_{i0}$$

donde

$$\begin{aligned} m(s_i) &= E[Z(s_i)] \text{ es el valor esperado en el punto } s_i \\ \sigma'_{ij} &= \sigma_{ij} - m(s_i)m(s_j) \text{ – es la covarianza centrada} \end{aligned}$$

– Kriging Ordinario: Kriging lineal con valor esperado estacionario pero desconocido

Generalmente, el valor de la media m es desconocido y por lo tanto el kriging simple no se puede utilizar.

El kriging ordinario establece una condición adicional al sistema de ecuaciones del kriging simple con el fin de controlar la media desconocida.

En este caso el sistema de ecuaciones es:

$$\begin{cases} \sum_{j=1}^n \lambda_j \sigma_{ij} - \sum_{l=1}^L \mu_l \phi_l(s_i) = \sigma_{0i}, & i = 1, \dots, n \\ \sum_{i=1}^n \lambda_i \phi_l(s_i) = \phi_l(s_0), & l = 1, \dots, L \end{cases}$$

El estimador es:

$$Z_0^* = \sum_{i=1}^n \lambda_i Z(s_i)$$

Varianza de la estimación:

$$\sigma_{k_v}^2 = \sigma_{00} - \sum_{i=1}^n \lambda_i \sigma_{i0} + \sum_{l=1}^L \mu_l \phi_l(s_0)$$

-Kriging por Bloques

En el kriging por bloques en lugar de estimar el valor en un punto s_k se considera una región V_k de área A_k con centro en el punto x_k . El estimador tiene la siguiente forma:

$$Z_{V_k}^* = \sum_{i=1}^n \lambda_i Z(s_i)$$

La varianza de la estimación se expresa como

$$\sigma_{V_k}^2 = \sum_{i=1}^n \lambda_i \gamma_{V_k} + \mu - \gamma_{V_k, V_k}$$

donde

$$\gamma_{V_k, V_k} = \frac{1}{A_k^2} \int_{V_k} \int_{V_k} \gamma(s_i - s_j) ds_i ds_j$$

Obviamente la estimación por bloques conduce a estimadores más suavizados que la estimación puntual.

-Kriging en presencia de no estacionariedad

En caso de no estacionariedad se pueden producir dos situaciones: no linealidad y linealidad. En el primer caso, para calcular el estimador $E[Z_k / Z_1, \dots, Z_n]$ sería necesario conocer la distribución conjunta con $n+1$ dimensiones. En la práctica resulta inviable. En el segundo caso, linealidad o variable aleatoria Gaussiana, el mejor predictor es el estimador lineal. Por tanto, debe transformarse adecuadamente la variable si esta no es Gaussiana. La predicción, sin embargo, debe obtenerse en la escala original, por lo que al invertir las transformaciones se deben hacer las correcciones oportunas para el sesgo y la varianza.

A partir de ahora, se supondrá variable aleatorias Gaussianas o transformadas.

Suponiendo media no estacionaria, $E[Z(s)] = m(s)$, donde $m(s)$ es una función espacial, conocida como tendencia o deriva (*drift*), el Kriging más utilizado es el Kriging Universal (Matheron 1969), en el que se supone que $m(x)$ puede ser especificado como un polinomio de orden finito k .

-Kriging universal

En este método deben ser conocidos a priori, el orden k del polinomio que mejor describe o explica la tendencia y el semivariograma, γ , de la variable aleatoria $Z(s)$ (una vez extraída la tendencia).

El problema es que ni k , el orden del polinomio, ni γ , el semivariograma, son conocidos a priori y suele estimarse a partir de los residuos, es decir $R(s) = Z(s) - m(s)$. Lo que se hace es trabajar con unos residuos estacionarios o al menos intrínsecos.

Según Matheron (1969) esto hace que los estimadores sean sesgados y que, al menos que el semivariograma de los residuos sea conocido, dígase a partir de su estimación en una dirección donde la tendencia no sea perceptible, el kriging universal tenga serias dificultades operacionales. Según Matheron (1969), el semivariograma de los residuos estimados $\gamma_R^*(h)$ difiere del semivariograma de los verdaderos residuos $\gamma_R(h)$, con un sesgo función del estimador de $m(s)$. Así en la estimación del semivariograma es donde está la clave del problema. No hay solución conocida. Sólo se pueden hacer suposiciones y tratar de verificarlas mediante el método de prueba y error, por lo que entonces el Kriging Universal se transforma en un procedimiento heurístico que, además puede consumir una gran cantidad de tiempo y no ofrece ninguna garantía de que converja.

-Kriging Residual

El kriging residual fue propuesto por Gambolati y Volpi (1979). Supone conocido el orden de la tendencia $m(s)$, la cual se estima por mínimos cuadrados ordinarios $\hat{m}(s)$. Tras calcular los residuos, $R(s)$ se les aplica kriging ordinario.

El problema es que el kriging residual es inconsistente puesto que el método de mínimos cuadrados ordinarios supone que los residuos son espacialmente independientes.

- Otros métodos de Kriging

Existen muchos otros métodos:

- Según la forma del estimador: lineales (**simple, ordinario, universal y residual.**, en negrita porque han sido explicados previamente), no lineales (disyuntivo, indicador y probabilístico)

- Según el soporte de la medición de los datos: puntual y **en bloques**

-Según se asuma una distribución de probabilidad: paramétrico (multigaussiano, disyuntivo y lognormal) o no paramétrico (**simple, ordinario, universal, residual, indicador y probabilística**)

2.5.2.3.2.- Análisis multivariante. Cokriging

La estimación conjunta de variables aleatorias espaciales, más conocida como cokriging (Kriging conjunto) es el análogo natural del kriging de una variable aleatoria. Mientras que el kriging utiliza la correlación espacial (univariante) para determinar las ponderaciones del estimador lineal, el cokriging utiliza además la correlación entre variables aleatorias espaciales (Cressie N, 1993).

La correlación entre variables espaciales se mide a través del semivariograma cruzado.

-Semivariograma cruzado

Si Z_i y Z_j son variables aleatorias estacionarias o intrínsecas, el semivariograma cruzado se define como:

$$\gamma_{ij}(h) = \frac{1}{2} E\{[Z_i(s+h) - Z_i(s)][Z_j(s+h) - Z_j(s)]\}$$

Como propiedades el semivariograma cruzado tiene las siguientes:

1) $\gamma_{ij}(0) = 0$

2) $\gamma_{ij}(-h) = \gamma_{ij}(h)$

3) $\gamma_{ij}(h) = \gamma_{ji}(h)$

4) Relación con la función de covarianza cruzada

Definiendo la función de covarianza cruzada como:

$$C_{ij}(h) = E\{[Z_i(s+h) - m_i][Z_j(s) - m_j]\}$$

Entonces:

$$\gamma_{ij}(h) = C_{ij}(0) - \frac{1}{2}(C_{ij}(h) + C_{ji}(h))$$

Nótese que la función de covarianza no es necesariamente simétrica.

$$C_{ij}(h) \neq C_{ji}(h)$$

Sin embargo, una práctica común es asumir que la función de covarianza es simétrica. Esto simplifica enormemente los cálculos asociados a la estimación de la función de covarianza conjunta. En ese caso,

$$\gamma_{ij}(h) = C_{ij}(0) - C_{ij}(h)$$

5) Relación de dependencia

Es importante tener presente que entre el semivariograma cruzado y los semivariogramas de cada una de las variables, existe una relación de dependencia. Por ejemplo, se puede demostrar que:

$$|\gamma_{ij}(h)|^2 \leq \gamma_i(h)\gamma_j(h) \quad \text{Desigualdad de Hölder}$$

En particular, el producto de cada uno de las mesetas de los semivariogramas individuales es mayor que el cuadrado de la meseta del semivariograma cruzado.

$$(\sigma_{ij}^2)^2 \leq \sigma_i^2 \sigma_j^2$$

En consecuencia, el modelo para semivariograma cruzado no puede ser escogido independientemente de cada uno de los modelos para los semivariogramas individuales.

6) El modelo de coregionalización lineal

Como se vio más arriba, la varianza de las combinaciones lineales de la variable de interés era positiva cuando se utilizaban los modelos autorizados para el semivariograma. Al incluir más variables, es necesario asegurar que la varianza de combinaciones lineales de estas sea positiva.

Para lograr esto se utiliza el modelo lineal de coregionalización, que establece que los semivariogramas individuales y el cruzado son combinaciones lineales de modelos de semivariogramas. En el caso de dos variables se tiene que:

$$\gamma_i(h) = \mu_1\gamma_1(h) + \mu_2\gamma_2(h) + \dots + \mu_m\gamma_m(h)$$

$$\gamma_j(h) = v_1\gamma_1(h) + v_2\gamma_2(h) + \dots + v_m\gamma_m(h)$$

$$\gamma_{ij}(h) = w_1\gamma_1(h) + w_2\gamma_2(h) + \dots + w_m\gamma_m(h)$$

El uso del modelo de coregionalización lineal implica:

- 1) Toda estructura presente en el semivariograma cruzado debe estar presente en los semivariogramas individuales. El recíproco no es cierto.
- 2) Los semivariogramas individuales tendrán todos los mismos rangos y la forma del semivariograma será la misma. Sólo se diferenciarán en los valores de la meseta.
- 3) Los semivariogramas individuales tendrán todas las mismas direcciones de anisotropía.

El cokriging se suele estimar considerando $Z(u)$ como una combinación lineal de las observaciones disponibles de Z más combinaciones lineales de las observaciones de aquellas variables relacionadas.

Por ejemplo: Z : variable principal S : variable secundaria

$$Z_{cok}^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u)Z(u_{\alpha}) + \sum_{j=1}^K \sum_{\alpha_j=1}^N \beta_{\alpha_j}(u)S_j(s_{\alpha_j})$$

Combinación lineal de la variable principal

Combinaciones lineales de las variables secundarias

- Cokriging simple

El caso más sencillo se denomina cokriging simple y la hipótesis básica es la estacionariedad de todas las variables además de que se asume que las medias de todas las variables son conocidas.

$$E(Z(u)) = m \quad \text{y} \quad m \text{ es conocida}$$

$$E(S_j(u)) = m_j \quad \text{y} \quad m_j \text{ es conocida } \forall j$$

A continuación se obtienen las ecuaciones de cokriging simple en el caso en que se considera solo una variable secundaria. En este caso el estimador propuesto es

$$Z_{cok}^*(u) = m + \sum_{\alpha=1}^N \lambda_{\alpha}(u)(Z(u_{\alpha}) - m) + \sum_{\alpha_1=1}^{N_1} \beta_{\alpha_1}(u)(S_1(s_{\alpha_1}) - m_1)$$

Al igual que en el kriging, las condiciones de optimalidad son:

1) Estimador insesgado $E(Z_{cok}^*(u)) = E(Z(u))$

2) $\text{var}[Z(u) - Z_{cok}^*(u)]$ mínima

La primera condición se obtiene automáticamente:

$$\begin{aligned} E(Z(u_\alpha)) - m &= 0 \\ E(S_1(s_{\alpha_1})) - m_1 &= 0 \end{aligned}$$

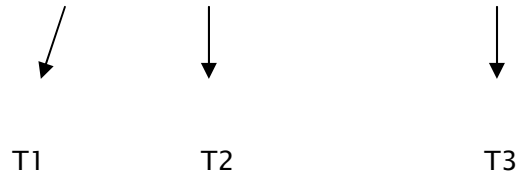
Con lo cual,

$$E(Z_{cok}^*(u)) = m = E(Z(u))$$

La condición de varianza mínima se obtiene derivando respecto a los parámetros α y β e igualando a cero cada una de las derivadas obtenidas.

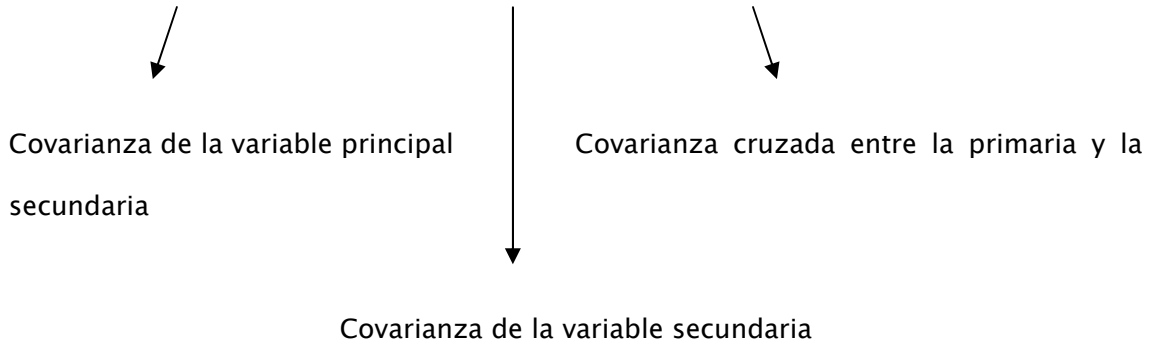
Para calcular explícitamente la expresión de la varianza hay que proceder con cautela debido a que aparecen nuevos términos a tener en cuenta.

$$\text{var}[Z(u) - Z_{cok}^*(u)] = \text{var}[Z(u)] + \text{var}[Z_{cok}^*(u)] - 2\text{cov}(Z(u), Z_{cok}^*(u))$$



$$T1 = \text{var}[Z(u)] = \sigma^2$$

$$\begin{aligned} T2 = & \text{var}[Z_{cok}^*(u)] = \\ & \text{var}\left(\sum \lambda_i (Z(u_i) - m)\right) + \text{var}\left(\sum \beta_j (S(s_j) - m_j)\right) + 2\text{cov}\left(\sum \lambda_i (Z(u_i) - m), \sum \beta_j (S(s_j) - m_j)\right) \\ & = \sum \sum \lambda_i \lambda_j C_Z(u_i + u_j) + \sum \sum \beta_i \beta_j C_S(s_i + s_j) + 2\sum \sum \lambda_i \beta_j C_{ZS}(u_i + s_j) \end{aligned}$$



$$T3 = 2 \text{cov}(Z(u), Z_{cok}^*(u)) = 2 \sum \lambda_i C_Z(u - u_i) + 2 \sum \beta_j C_{ZS}(u - s_j)$$

Una vez calculadas las derivadas respectivas, la expresión detallada del sistema de ecuaciones es:

$$\begin{bmatrix} C_Z & C_{ZS} \\ C_{SZ} & C_S \end{bmatrix} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} = \begin{bmatrix} C_{Zu} \\ C_{Su} \end{bmatrix}$$

- Cokriging ordinario

Al igual que en el caso de kriging ordinario, se asume que las medias de las variables son desconocidas y se imponen condiciones para filtrarlas (Cressie, 1993).

El estimador propuesto es

$$Z_{cok}^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u) Z(u_{\alpha}) + \sum_{j=1}^K \sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j}(u) S_j(s_{\alpha_j})$$

Con lo cual

$$E(Z_{cok}^*(u)) = m \sum \lambda_{\alpha} + \sum_{j=1}^K m_j \sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j}$$

Y se obtienen las condiciones,

$$\sum \lambda_{\alpha} = 1 \qquad \sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j} = 0 \qquad j = 1, 2, \dots, K$$

Con sólo dos variables se requieren cuatro funciones de covarianza. En general, con N variables secundarias se requieren 2^{N+1} funciones de covarianza. Por otra parte, debe existir una correlación lineal entre la variable principal y las variables secundarias. Las

variables secundarias deben poseer un número mucho mayor de observaciones que la variable principal.

Si los datos no son coincidentes es imposible estimar las covarianzas cruzadas. Si son parcialmente coincidentes entonces si que los resultados son satisfactorios.

Cuando las variables están intrínsecamente relacionadas, es decir cuando ocurre que los modelos de semivariograma o covarianza de todas las variables son proporcionales a un mismo modelo de semivariograma o covarianza, entonces el kriging y el cokriging con datos totalmente coincidentes son iguales.

2.6.- Estudios de asociación geográfica

El objetivo en este tipo de estudios es examinar, desde un punto de vista geográfico (y ecológico) la relación entre variables respuesta de salud y exposiciones, tanto a factores medioambientales de tipo ecológico (contaminación atmosférica, del agua y del suelo) como factores de tipo ocupacional; condiciones socioeconómicas y demográficas; o variables relacionados con el estilo de vida (tales como hábito tabáquico y dieta).

Es habitual, sobre todo con datos espaciales, que la sobredispersión (o la extravariabilidad) no desaparezca por completo (de hecho, suele reducirse apenas). Puede ocurrir, por ejemplo, que se trate de un problema de escala. Así, los datos pueden ser 'escasos' (sparseness), es decir no se observe ningún caso en (relativamente) bastantes de las áreas.

Sin embargo, parte de la extra-variabilidad puede ser atribuida a la falta de control de la confusión. De hecho, si se han omitido variables confusoras del riesgo esperado (privación, otras variables socio-económicas), éstas deberían ser incluidas en el análisis.

Los estudios de asociación geográfica pueden utilizar tanto *datos en rejilla* como *procesos puntuales*. En el primer caso, los estudios son considerados como ‘generadores de hipótesis’ (Elliott y Wartenberg, 2004), por cuanto la unidad de observación la constituyen individuos agrupados geográficamente y las asociaciones observadas a este nivel (ecológico) no necesariamente se mantienen a nivel individual².

2.6.1.- Modelos Lineales Generalizados Mixtos con variables explicativas con diseño en rejilla

Podría ocurrir, por ejemplo, que parte de la variabilidad geográfica de la incidencia de las neoplasias de interés medioambiental estuviese explicada por la distribución territorial de las condiciones socioeconómicas (véanse por ejemplo los artículos del Anexo 4). La introducción de una variable explicativa en un modelo GLMM, podría ser con un único efecto fijo, es decir, no existe variabilidad dentro de cada municipio, o con efecto aleatorio, se acepta pequeña variación dentro del municipio.

En primer lugar, con efecto fijo:

-Función vínculo

$$\text{Ln}(E(O_i)) = \beta_{0i} + \beta_1 x_i + \text{Ln}(E_i) \quad [4]$$

-Función varianza

$$\text{Var}(O_i) = \phi E(O_i)$$

-Efecto aleatorio

$$\beta_{0i} = \beta_0 + v_i \quad v_i \rightarrow iidN(0, \sigma_v^2)$$

² Problema conocido como ‘falacia ecológica’ (Piantadosi et al., 1988).

En este modelo, x_i representa una variable (con valores para cada uno de los municipios i), con un riesgo (ecológico) promedio asociado, igual a e^{β_1} .

Introduciendo el coeficiente asociado a la variable explicativa como *efecto aleatorio*.

-Función vínculo

$$\ln(E(O_i)) = \beta_{0i} + \beta_{1i}x_i + \ln(E_i)$$

[5]

-Función varianza

$$\text{Var}(O_i) = \phi E(O_i)$$

-Efecto aleatorio

$$\begin{aligned} \beta_{0i} &= \beta_0 + v_i & v_i &\rightarrow iidN(0, \sigma_v^2) \\ \beta_{1i} &= \beta_1 + \omega_i & \omega_i &\rightarrow iidN(0, \sigma_\omega^2) \end{aligned}$$

En la que el riesgo asociado a las condiciones socioeconómicas varía para cada área i .

2.6.2.- Modelos Lineales Generalizados Mixtos con variables explicativas con diseño geoestadístico

En este trabajo la variable dependiente (incidencia y mortalidad) se observa siempre en un diseño en rejilla. Las variables explicativas medioambientales, sin embargo, se obtienen de un diseño geoestadístico.

La información sobre la exposición medioambiental suele ser obtenida de unas (relativamente) pocas estaciones más o menos dispersas espacialmente. Esta limitación ha conducido recientemente a la modelización de una superficie de exposición (Le et al, 1996; Gelfand et al, 2001; Tonellato, 2001; Shaddick y Wakefield, 2002).

Se necesita, por tanto, algún método de interpolación que permita asignar los valores continuos de la variable medioambiental a los discretos de la dependiente. En este trabajo se ha optado por la interpolación estocástica. Además de por su flexibilidad, es el único que proporciona una estimación de la incertidumbre, que como se verá a continuación, es imprescindible a la hora de especificar los modelos de asociación geográfica, en este caso.

Lawson et al. (2003) proponen utilizar dos tipos de kriging. En primer lugar, el método de kriging por bloques. Éste debe utilizarse cuando se pretende interpolar la variable en la totalidad del área en la que se observa la variable dependiente. Esta estrategia es apropiada cuando se considera que el riesgo, y la variable, varían substancialmente dentro del área. En otro caso, cuando el riesgo y/o la variable tengan una variabilidad más o menos reducida dentro del área, estos autores proponen utilizar el kriging ordinario sobre todo. En este último caso se supone que la variable dependiente se 'sitúa' en el centroide del área y que la variable explicativa se interpola en ese punto.

Una vez obtenida esta superficie de exposición, varios autores han analizado la relación entre ésta y diversos indicadores de salud, como ingresos hospitalarios por causas respiratorias y concentraciones de sulfatos (Le et al, 1996; Zidec et al, 1998); urgencias pediátricas por asma y ozono (Carlin et al, 1999; Zhu et al, 2003).

El problema de todos estos trabajos es que suponen que la posible asociación observada a nivel de área (nivel ecológico) es la misma para todos los individuos residentes en esa área (nivel individual) (Wakefield y Shaddick, 2006). Este supuesto podría conducir al problema de la falacia ecológica (Selvin, 1958) en su versión de 'sesgo puro de especificación' (Wakefield y Shaddick, 2006) que aparece cuando se agrega un modelo no lineal.

Supóngase una región bajo estudio A con k subáreas, A_k , de las que se dispone de información sobre la población, N_k , y sobre algún indicador de salud, Y_k . Supóngase

también que se dispone de información sobre alguna variable de exposición medioambiental, X_n , medida en $m=1, \dots, M$ estaciones situadas en m localizaciones en la región.

$$Y_{ki} | X_{ki}, \beta \longrightarrow \text{Bernouilli}(p(X_{ki}, \beta))$$

Suponiendo que $(p(X, \beta)) = \exp(\beta_0 + \beta_1 X)$.

Así, el parámetro de interés a nivel individual es β_1 . A nivel agregado no se conocen las respuestas individuales Y_{ki} , sino su suma Y_k . Si se denomina X_k al conjunto de exposiciones individuales para los individuos residentes en el área K , entonces;

$$E(Y_k | X_k, \beta) = N_k q_k$$

donde $q_k = \frac{e^{\beta_0}}{N_k} \sum_{i=1}^{N_k} \exp(\beta_1 X_{ki})$ es el riesgo promedio para los individuos residentes en el área K (Wakefield y Shaddick, 2006).

Así, cuando se introduce la exposición 'promedio' en el modelo (ecológico):

$$Y_k | \bar{X}_k, \beta^* \longrightarrow \text{Poisson}(N_k \exp(\beta_0^* + \beta_1^* \bar{X}_k))$$

Se comete un sesgo ecológico ($\beta_1 \neq \beta_1^*$) puesto que la función de riesgo se calcula en la exposición media mientras que lo correcto hubiese sido calcular la media de las funciones de riesgo (Wakefield y Shaddick, 2006). Cabe señalar, no obstante, que si la exposición fuese constante dentro del área no se cometería sesgo ecológico.

Cuando, como en este caso, se disponen de (relativamente) pocas observaciones de la exposición y por tanto debe estimarse la superficie de exposición, se pueden utilizar varias alternativas para evitar el sesgo ecológico.

En primer lugar, utilizando una aproximación bayesiana completa, se podrían estimar simultáneamente los parámetros de salud y de la exposición. El problema es que, puesto que la dimensión del vector estimado de exposición es muy grande esta aproximación es computacionalmente inabordable.

Alternativamente se podría utilizar el modelo en tres etapas propuesto por Wakefield y Shaddick, 2006. El problema es que la implementación de dicha aproximación, tanto en la simulación como en una aplicación práctica de la relación entre mortalidad respiratoria y SO₂ en Londres, no proporciona resultados satisfactorios, según los mismos autores.

En este trabajo se utiliza, alternativamente, la aproximación de Jackson et al. (2006)

-Función vínculo

$$\text{Ln}(E(O_i)) = \beta_{0i} + \beta_{1i}x_i + \frac{\beta_1^2}{2}s_i^2 + \text{Ln}(E_i) \quad [6]$$

-Función varianza

$$\text{Var}(O_i) = \phi E(O_i)$$

-Efecto aleatorio

$$\begin{aligned} \beta_{0i} &= \beta_0 + \nu_i & \nu_i &\rightarrow iidN(0, \sigma_\nu^2) \\ \beta_{1i} &= \beta_1 + \omega_i & \omega_i &\rightarrow iidN(0, \sigma_\omega^2) \end{aligned}$$

Donde x_i es el estimador kriging de la variable observada en el centroide del área, utilizando un diseño geoestadístico y s_i^2 es la variabilidad de esa estimación dentro del área.

Debe señalarse que el sesgo ecológico prácticamente desaparece si el coeficiente individual de interés (β_1) es pequeño y/o la variabilidad de la exposición dentro del área es pequeña (Jackson et al, 2006; Wakefield y Shaddick, 2006). En este sentido, Wakefield y Shaddick (2006) señalan que en el estudio de Zhu et al (2003) el sesgo será reducido por cuanto las áreas bajo estudio son código postales y la principal variación en la exposición es temporal y no tanto espacial.

Obsérvese que β_1 se ha considerado un efecto aleatorio. La razón es que se supone que el error de estimación no es exclusivamente un error de medida sino que, la exposición aunque medida ecológicamente, ocurre siempre a nivel individual (Lawson et al, 2003).

El modelo puede complicarse permitiendo interacción y/o correlación entre las variables explicativas. Asimismo, se puede incorporar tanto variables explicativas observadas en un diseño geoestadístico (medioambientales) como en rejilla (socioeconómicas). Cuando exista correlación entre las variables explicativas un estimador de ésta debe ser introducido como variable explicativa en el modelo [6]. Por ejemplo, se puede estimar una regresión auxiliar tomando como dependiente una de las variables explicativas, y utilizar los valores ajustados en esta regresión como estimadores de la correlación (Jackson et al, 2006). Una alternativa consiste en estratificar los valores de una de las variables explicativas en función de la otra.

Referencias

- Abellán JJ. Evolución espacial de la gripe en la temporada 1998–1999. 2000. <http://dgsp.san.gva.es/SSCC/Epidemiologia/Centinelas/gripe/evo9899.htm>
- Bailey TC, Gatrell AC. *Interactive spatial data analysis*. Essex: Longman Scientific & Technical, 1995.
- Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; 43:1–59.
- Cameron A, Trivedi P. Regression analysis of count data. Cambridge, MA: Cambridge University Press, 1998.
- Carlin BP, Xia H, Devine O, Tolbert P, Mulholland J. Spation–temporal hierarchical models for analyzing Atlanta pediatric asthma ER visit rates. En Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M (eds). *Case Studies in Bayesian Statistics*, volumen IV. Nueva York: Springer, 1999. pp303–320.
- Clayton D, Kaldor J. Empirical Bayes estimates for age–standardized relative risks. *Biometrics* 1987; 43:671–681.
- Cliff A, Haggett P. *Atlas of disease distributions: analytic approaches to epidemiological data*. Oxford: Blackwell, 1992.
- Cressie N. *Statistics for Spatial Data*. Nueva York: John Wiley and Sons, 1993.
- Cressie N y Hawkins J. Robust estimation of the variogram. *Journal Mathematics Geology*, 1980 ; 12(2): 115–126.
- Creutin JD , Obled M. Objective analysis and mapping techniques for rainfall fields: an objective comparison. *Water Resources Research* 1982; 18: 413–431.
- Darrymple ML, Hudson IL, Ford RPK. Finite mixture, zero–inflated Poisson and hurdle models with application to SIDS. *Computational Statistics and Data Analysis* 2003; 41:491–504.
- Deb P, Trivedi PK. The structure of demand for health care: latent class versus two–part models. *Journal of Health Economics* 2002; 21:601–625
- Díaz MA. *Geoestadística aplicada*. México DF: Instituto de Geofísica, UNAM, 2002.
- Doll R. The epidemiology of cancer. *Cancer* 1980; 45:2475–2485.

English D. Geographical epidemiology and ecological studies. En: Elliot P, Cuzick J, English D, Stern R, eds. *Geographical and Environmental Epidemiology*. Oxford University Press, 1992:3–13.

Elliot P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. *Statistical Methods in Medical Research* 1995; 4:137–159.

Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives* 2004; 112(9): 998–1006.

Feller W. On a general class of “contagious” distributions. *Annals of Mathematical Statistics* 1943; 16:319–329.

Gambolatil, G. Volpi G. A conceptual deterministic analysis of the kriging technique in hydrology. *Water Resources. Research*. 1979; 15(3): 625–629.

Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 1996; 21:256–274.

Gelfand AE, Zhu L, Carlin BP. On the change of support problem for spatial–temporal data. *Biostatistics* 2001; 2: 31–45.

Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Nueva York: Chapman & Hall, 1996.

Gneiting T. Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society, Part A*, 1999; 125: 2449–2464.

Greene WH. *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models*. Working Paper EC–94–10, Department of Economics, New York University, 1994.

Greenwood M, Yule GU. An enquiry into the nature of frequency distributions to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society, Series A* 1920; 83:255–279.

Griffith DA. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. En Arlinghaus SL (ed). *Practical Handbook of Spatial Statistics*, pp. 65–72. Boca Raton, Florida: CRC Press, 1995.

Gupta P, Gupta R, Tripathi R. Analysis of zero–adjusted count data. *Computational Statistics and Data Analysis* 1996; 23:207–218.

Heilbron D. *Generalized linear models for altered zero probabilities and overdispersion in count data*. SIMS Technical Report 9, Department of Epidemiology and Biostatistics, University of California, San Francisco, 1989.

Heilbron D. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* 1994; 36:531–547.

Hinde J, Demétrio CGB. Overdispersion: models and estimation. *Computational Statistics and Data Analysis* 1998; 27:151–170.

Isaaks E, Srivastava M. *An introduction to applied geostatistics*. New York, Oxford University Press. 561 p. 1989.

Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. *Statistics in Medicine* 2006; 25(12):2136–2159.

Kaluzny SP, Vega SC, Cardoso TP, Shelby AA. *S+Spatial statistics*. Nueva York: Mathsoft, 1998.

Kelsall JE, Diggle PJ. Spatial variation in risk of disease – a nonparametric binary regression approach. *Journal of the Royal Statistical Association, Series C*, 1998; 47:559–573.

Kelsall J, Wakefield J. Modelling spatial variation in disease risk: a geostatistical approach. *Journal of the American Statistical Association* 2002; 97(459):692–701.

Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. *Statistics in Medicine* 2006; 25(12):2136–2159.

Julious SA, Nicholl J, George S. Why do we continue to use standardized mortality ratios for small area comparisons?. *Journal of Public Health Medicine* 2001; 23:40–46.

Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34:1–14.

Langford IH, Leyland AH, Rasbash J, Goldstein H. Multilevel modelling of the geographical distribution of diseases. *Journal of the Royal Statistical Society, Series C* 1999; 48(2):253–268.

Lawless JF. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics* 1987; 15:209–225.

Lawson AB, Browne WJ, Vidal-Rodeiro C. *Disease Mapping with WinBUGS and MLwiN*. Chichester, West Sussex: John Wiley & Sons, 2003.

Lawson AB, Clark A. Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine* 2002; 21:359–370.

Le ND, Sun W, Sidec JV. Bayesian multivariate spatial interpolation with data missing-by-design. *Journal of the Royal Statistics Society, Series B* 1996; 59:L 501–510.

Lertxundi–Manterola A, Saurina C, Saez M, Ocaña R. Construcción de un índice de privación material para los municipios de la Región Sanitaria Girona. *Estudios de Economía Aplicada* 2005; 23(1):1–25.

Lertxundi–Manterola A, Saez M. Influencia de la privación sobre la distribución de la incidencia de cáncer en la Regió Sanitària Girona. *Economía y Salud*, 2006; 56:10–11.

López–Abente G, Ibáñez C. *Aplicación de técnicas de análisis espacial a la mortalidad por cáncer en Madrid. Documentos Técnicos de Salud Pública* 66, Madrid: Dirección General de Salud Pública, Consejería de Sanidad, Comunidad de Madrid, 2001.

Matheron G. 1969. Le Krigeage Universal. *Cahiers du Centre de Morphologie Mathématique*, No. 1, Fontainebleau, France.

Matheron G, Kleingeld WJ. *The Evolution of Geostatistics*, APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries, Volume 3. Geostatistics. Johannesburg, SAIMM, pp. 9–12.1987.

Mestre A. *Report on rainfall spatialisation*. Madrid: Cost 718 WG1. Report on rainfall spatialisation, 2000.

Mollie A. Bayesian mapping of disease. En Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Nova York: Chapman & Hall, 1996, pp. 359–379.

Mullahy J. Specification and testing of some modified count data models. *Journal of Econometrics* 1986; 33:341–365.

Mullahy J. Heterogeneity, excess zeros, and the structure of count data models. *Journal of Health Econometrics* 1997; 12:337–350.

Ocaña–Riola R, Saurina C, Sanchez–Cantalejo C, Lertxundi A, Fernández–Ajuria A, Saez M, Barceló MA, Daponte A, Ruiz–Ramos M, Martínez JM, Benach J. Area deprivation and mortality in the capital cities of Andalusia and Catalonia (Spain). (Enviado al *Journal of Epidemiology and Community Health*).

Oliver MA, Muir KR, Parkes SE, Webster R. Methods for investigating localized clustering of disease. Geostatistics for determining the risk of rare disease. *IARC Science Publication* 1996; 135:227–247.

Piantadosi S, Byar DP, Green SB. The ecological fallacy. *American Journal of Epidemiology* 1988; 127:893–904.

Ribeiro PJ, Diggle PJ. Geo R: a package for geostatistical analysis *RNEWS*, 2001; 1(2): 15–18.

Ribeiro PJ. The geoR Package. <http://www.est.ufpr.br/geoR> (último acceso 28 de marzo de 2006)

Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives* 2004; 112 (9): 1016–1025.

Ridout M, Demétrio CGB, Hinde J. *Models for count data with many zeros*. International Biometric Conference, Cape Town, December 1998.

Saez M, Saurina C. *Estadística y epidemiología espacial*. Girona: Grup de Recerca en Estadística, Economia Aplicada i Salut, GRECS. Universitat de Girona, 2006.

Selvin HC. Durkheim's 'suicide' and problems of empirical research. *American Journal of Sociology* 1958; 63: 607–619.

Shaddick G, Wakefield JC. Modelling multivariate pollutant data at multiple sites. *Applied Statistics* 2002; 51: 351–372.

Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS User Manual. Version 1.4*. Gener 2003. <http://www.mrc-bsu.cam.ac.uk/bugs>.

Thomas RW. *Spatial epidemiology*. Londres: Pion, 1990.

Tonellato SF. A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Applied Statistics* 2001; 50: 187–200.

Van Beers WCM. Kriging metamodeling in discrete-event simulation: an overview. En Kuhl ME, Steiger NM, Armstrong FB, Joines JA (eds). *Proceedings of the 2005 Winter Simulation Conference*, 2005, pp. 202–208.

Wakefield J, Shaddick G. Health-exposure modelling and the ecological fallacy. *Biostatistics* 2006; 7(3): 438–455.

Walter SD. Disease mapping: a historical perspective. En: Elliott P, Wakefield J, Best N, Briggs DJ (eds). *Spatial epidemiology: methods and applications*. Oxford: Oxford University Press, 223–252, 2000.

Walter SD, Birmie SE. Mapping mortality and morbidity patterns: an international comparison. *International Journal of Epidemiology* 1991; 20:678–689.

Zhu L, Carlin BP, Gelfand AE. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics* 2003; 14: 537–557.

Zidek JV, White R, Lee ND, Sun W, Burnett RT. Imputing unmeasured explanatory variables in environmental epidemiology with application to health impact analysis of air pollution. *Environmental and Ecological Statistics* 1998; 5:99–115.

3.- Metodología. Análisis descriptivo

3.1.- Población

Incidencia de las neoplasias de interés medioambiental

La población estudiada fue la residente en la Regió Sanitària Girona (ver Tabla 1).

Mortalidad por neoplasias de interés medioambiental

En este caso las dos poblaciones estudiadas fueron:

- Población residente en el Área Metropolitana de Barcelona (ver Tabla 2).
- Población residente en el Área Metropolitana de Bilbo (ver Tabla 3).

3.2.- Sujetos de estudio

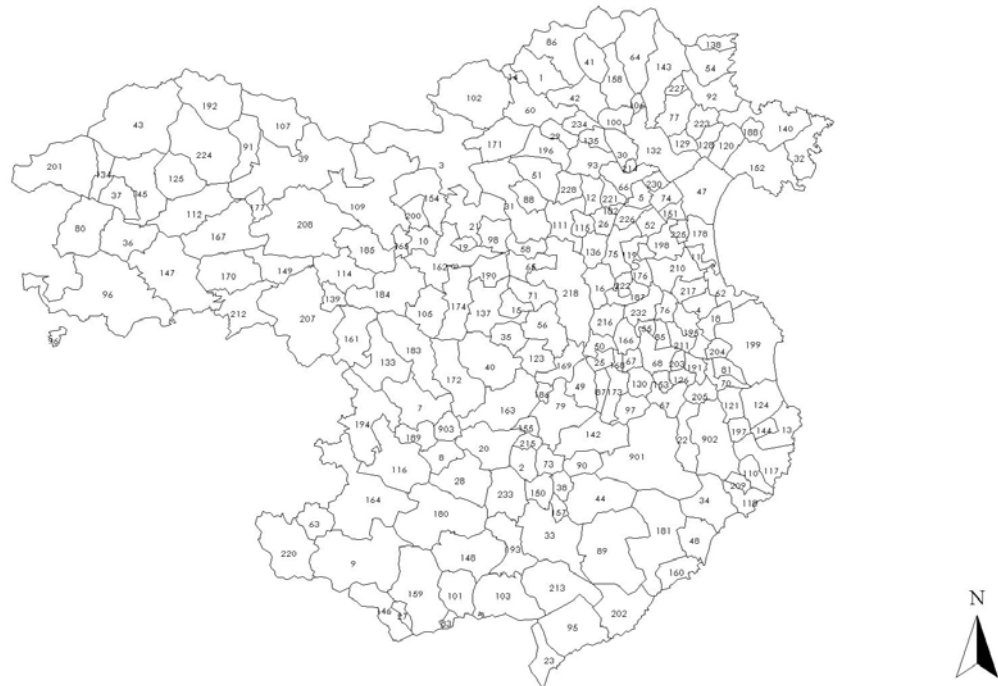
Incidencia de las neoplasias de interés medioambiental

Los sujetos residentes en la Regió Sanitaria Girona que hubieran sido diagnosticados por alguna de las neoplasias de interés medioambiental entre los años 1994 y 2001 (ver Figura 1).

Mortalidad por neoplasias de interés medioambiental

Dos grupos de sujetos:

- Sujetos residentes en Barcelona, L'Hospitalet de Llobregat, Sant Adrià del Besós, Badalona, Cornellà de Llobregat, El Prat de Llobregat, Santa Coloma de Gramenet, Sant Feliu de Llobregat, Montcada i Reixac, Esplugues de Llobregat y Sant Joan Despí (ver Figura 2), que hubieran fallecido entre los años 1994 y 2003.
- Sujetos residentes en Bilbo, Muskiz, Zierbena, Santurtzi, Portugalete, Sestao, Abanto-Zierbena, Ortuella, Trapagaran, Barakaldo, Erandio, Getxo, Leioa, Etxebarri, Sondika y Basauri (ver Figura 3), que hubieran fallecido entre los años 1996 y 2003.



MUNICIPIO	id	MUNICIPIO	id	MUNICIPIO	id	MUNICIPIO	id	MUNICIPIO	id
Agullana	1	Canet d'Adri	36	Gualta	71	Pardines	106	Sant Ferriol	141
Aiguaviva	2	Cantallops	37	Hostalric	72	Parlavà	107	Sant Gregori	142
Aibanyà	3	Capmany	38	Jafre	73	Pau	108	Sant Hilari Sacalm	143
Albons	4	Cassà de la Selva	39	Jonquera (La)	74	Pedret i Marzà	109	Sant Jaume de Llierca	144
Arner	5	Castelló d'Empúries	40	Juià	75	Pera (La)	110	Sant Joan de les Abadesses	145
Anglès	6	Castell-Platja d'Aro	41	Lladó	76	Peralada	111	Sant Joan de Mollet	146
Arbúcies	7	Cellera de Ter (La)	42	Llagostera	77	Planes d'Hostoles (Les)	112	Sant Joan les Fonts	147
Argelaguer	8	Celrà	43	Llambilles	78	Planols	113	Sant Jordi Desvalis	148
Armentera (L)	9	Cervià de Ter	44	Llanars	79	Pont de Molins	114	Sant Julià de Ramis	149
Avinyonet de Puigventós	10	Cistella	45	Llançà	80	Pontós	115	Sant Julià del Llor i Bonmatí	150
Banyoles	11	Colera	46	Llers	81	Porqueres	116	Sant Llorenç de la Muga	151
Begur	12	Colomers	47	Lloret de Mar	82	Port de la Selva (El)	117	Sant Martí de Llímena	152
Bellcaire d'Empordà	13	Cornellà del Terri	48	Llosses (Les)	83	Portbou	118	Sant Martí Vell	153
Besalú	14	Corçà	49	Madremanya	84	Preses (Les)	119	Sant Miquel de Campmajor	154
Bescanó	15	Crespià	50	Maia de Montcal	85	Quart	120	Sant Miquel de Fluvià	155
Beuda	16	Cruïlles, Monells i Sant Sadu	51	Masarac	86	Queralbs	121	Sant Mori	156
Bisbal d'Empordà (La)	17	Darnius	52	Massanes	87	Rabós	122	Sant Pau de Segúries	157
Biure	18	Escala (L)	53	Maçanet de Cabrenys	88	Regencós	123	Sant Pere Pescador	158
Blanes	19	Espinelles	54	Maçanet de la Selva	89	Ribes de Freser	124	Santa Coloma de Farners	159
Boadella d'Empordà	20	Espolla	55	Mieres	90	Riells i Viabrea	125	Santa Cristina d'Aro	160
Bordils	21	Esporonellà	56	Molló	91	Ripoll	126	Santa Logaia d'Iguema	161
Borrassà	22	Far d'Empordà (El)	57	Mollet de Peralada	92	Riudarenes	127	Santa Pau	162
Bàscara	23	Figueras	58	Montagut	93	Riudaura	128	Sarrià de Ter	163
Breda	24	Fiaçà	59	Mont-ras	94	Riudellots de la Selva	129	Saus	164
Brunyola	25	Foixà	60	Navata	95	Riumors	130	Selva de Mar (La)	165
Cabanelles	26	Fontanilles	61	Ogassa	96	Roses	131	Serinyà	166
Cabanes	27	Fontcoberta	62	Ordis	97	Rupià	132	Serra de Daró	167
Cadaqués	28	Forallac	63	Osor	98	Sales de Llierca	133	Setcases	168
Caldes de Malavella	29	Fornells de la Selva	64	Palafrugell	99	Salt	134	Sils	169
Calonge	30	Fortià	65	Palamós	100	Sant Andreu Salou	135	Siurana	170
Camós	31	Garrigoles	66	Palau de Santa Eulàlia	101	Sant Aniol de Finestres	136	Susqueda	171
Campdevànoi	32	Garrigàs	67	Palau-sator	102	Sant Climent Sescebes	137	Tallada d'Empordà (La)	172
Campelles	33	Garriguella	68	Palau-saverdera	103	Sant Feliu de Buixalleu	138	Terrades	173
Camplong	34	Girona	69	Palol de Revardit	104	Sant Feliu de Guixols	139	Torrent	174
Camprodon	35	Gombrèn	70	Pals	105	Sant Feliu de Pallers	140	Torroella de Fluvià	175

Figura 1.-Nombre de los municipios de la Región Sanitaria Girona (Fuente: Construcción propia).

Municipio de residencia	Edad en grandes grupos	TOTAL	Hombres	Mujeres
	TOTAL	14232	6988	7244
17015-Banyoles	Menos de 16	1934	972	962
	16-64	9521	4842	4679
	65 ó más	2777	1174	1603
	TOTAL	8145	4037	4108
17022-Bisbal d'Empordà (La)	Menos de 16	1273	642	631
	16-64	5294	2712	2582
	65 ó más	1578	683	895
	TOTAL	30693	15315	15378
17023-Blanes	Menos de 16	5035	2595	2440
	16-64	21203	10752	10451
	65 ó más	4455	1968	2487
	TOTAL	6650	3330	3320
17034-Calonge	Menos de 16	731	361	370
	16-64	4778	2445	2333
	65 ó más	1141	524	617
	TOTAL	7712	3826	3886
17044-Cassà de la Selva	Menos de 16	1233	626	607
	16-64	4928	2523	2405
	65 ó más	1551	677	874
	TOTAL	6806	3429	3377
17048-Castell-Platja d'Aro	Menos de 16	1030	532	498
	16-64	4902	2487	2415
	65 ó más	874	410	464
	TOTAL	5896	3014	2882
17047-Castelló d'Empúries	Menos de 16	1072	538	534
	16-64	3917	2022	1895
	65 ó más	907	454	453
	TOTAL	5823	2918	2905
17062-Escala (L.)	Menos de 16	834	418	416
	16-64	3957	2016	1941
	65 ó más	1032	484	548
	TOTAL	33064	15966	17098
17066-Figueres	Menos de 16	5566	2831	2735
	16-64	21676	10746	10930
	65 ó más	5822	2389	3433
	TOTAL	74879	35871	39008
17079-Girona	Menos de 16	12421	6316	6105
	16-64	50890	24919	25971
	65 ó más	11568	4636	6932
	TOTAL	5186	2630	2556
17089-Llagostera	Menos de 16	420	206	214
	16-64	3660	1923	1737
	65 ó más	1106	501	605
	TOTAL	20239	10133	10106
17095-Lloret de Mar	Menos de 16	3513	1821	1692
	16-64	14365	7288	7077
	65 ó más	2361	1024	1337
	TOTAL	28060	13636	14424
17114-Olot	Menos de 16	4037	2063	1974
	16-64	17616	8942	8674
	65 ó más	6407	2631	3776
	TOTAL	18322	9197	9125
17117-Palafrugell	Menos de 16	2885	1511	1374
	16-64	12252	6318	5934
	65 ó más	3185	1368	1817
	TOTAL	14842	7328	7514
17118-Palamós	Menos de 16	2340	1233	1107
	16-64	9949	5041	4908
	65 ó más	2553	1054	1499
	TOTAL	10597	5154	5443
17147-Ripoll	Menos de 16	1335	686	649
	16-64	6856	3454	3402
	65 ó más	2406	1014	1392
	TOTAL	12726	6364	6362
17152-Roses	Menos de 16	2360	1225	1135
	16-64	8657	4391	4266
	65 ó más	1709	748	961
	TOTAL	21238	10616	10622
17155-Salt	Menos de 16	3424	1765	1659
	16-64	14564	7492	7072
	65 ó más	3250	1359	1891
	TOTAL	17994	8792	9202
17160-Sant Feliu de Guixols	Menos de 16	2937	1452	1485
	16-64	11848	5969	5879
	65 ó más	3209	1371	1838
	TOTAL	5036	2580	2456
17164-Sant Hilari Sacalm	Menos de 16	753	396	357
	16-64	3355	1785	1570
	65 ó más	928	399	529
	TOTAL	9169	4639	4530
17180-Santa Coloma de Farners	Menos de 16	1490	759	731
	16-64	6137	3189	2948
	65 ó más	1542	691	851
	TOTAL	8244	4127	4117
17199-Torroella de Montgrí	Menos de 16	1380	718	662
	16-64	5456	2789	2667
	65 ó más	1408	620	788
	TOTAL	464452	229256	235196
TOTAL	Menos de 16	86826	37233	36205
	16-64	377341	156589	151941
	65 ó más	101137	35434	47050

Tabla 1.- Número de habitantes (hombres y mujeres) en los municipios con más de 5.000 habitantes de la Regió Sanitària Girona (Fuente: Censo de Población y Viviendas, 2001).

Municipio de residencia	Número de secciones censales	Edad en grandes grupos	TOTAL	Hombres	Mujeres
08015-Badalona	157	TOTAL	205836	101760	104076
		Menos de 16	30708	15855	14853
		16-64	145222	73590	71632
		65 ó más	29906	12315	17591
08019-Barcelona	1491	TOTAL	1503884	705533	798351
		Menos de 16	185301	94916	90385
		16-64	992724	484150	508574
		65 ó más	325859	126467	199392
08073-Cornellà de Llobregat	70	TOTAL	79979	39356	40623
		Menos de 16	10950	5688	5262
		16-64	56118	28228	27890
		65 ó más	12911	5440	7471
08077-Esplugues de Llobregat	29	TOTAL	45127	22150	22977
		Menos de 16	6062	3125	2937
		16-64	32610	16253	16357
		65 ó más	6455	2772	3683
08101-Hospitalet de Llobregat (L)	226	TOTAL	239019	116805	122214
		Menos de 16	29380	15084	14296
		16-64	167142	83777	83365
		65 ó más	42497	17944	24553
08125-Montcada i Reixac	15	TOTAL	28295	13932	14363
		Menos de 16	4596	2324	2272
		16-64	19511	9817	9694
		65 ó más	4188	1791	2397
08169-Prat de Llobregat (E)	36	TOTAL	61818	30924	30894
		Menos de 16	9475	4895	4580
		16-64	44402	22636	21766
		65 ó más	7941	3393	4548
08194-Sant Adrià de Besòs	23	TOTAL	31939	15748	16191
		Menos de 16	5231	2697	2534
		16-64	21802	10991	10811
		65 ó más	4906	2060	2846
08211-Sant Feliu de Llobregat	29	TOTAL	40042	19706	20336
		Menos de 16	5934	3010	2924
		16-64	29232	14688	14544
		65 ó más	4876	2008	2868
08217-Sant Joan Despí	20	TOTAL	28772	14186	14586
		Menos de 16	4833	2447	2386
		16-64	20653	10334	10319
		65 ó más	3286	1405	1881
08245-Santa Coloma de Gramenet	99	TOTAL	112992	56185	56807
		Menos de 16	14906	7659	7247
		16-64	81342	41335	40007
		65 ó más	16744	7191	9553
TOTAL	2195	TOTAL	2377703	1136285	1241418
		Menos de 16	307376	157700	149676
		16-64	1610758	795799	814959
		65 ó más	459569	182786	276783

Tabla 2.- Número de secciones censales y de habitantes (hombres y mujeres) en el Área Metropolitana de Barcelona (Fuente: Censo de Población y Viviendas, 2001).

Municipio de residencia	Número de secciones censales	Edad en grandes grupos	TOTAL	Hombres	Mujeres
48002-Abanto Zierbena	7	TOTAL	9036	4464	4572
		Menos de 16	1250	620	630
		16-64	6280	3225	3055
		65 ó más	1506	619	887
48013-Barakaldo	84	TOTAL	94478	46050	48428
		Menos de 16	10012	5083	4929
		16-64	65008	32723	32285
		65 ó más	19458	8244	11214
48015-Basauri	40	TOTAL	45085	22260	22825
		Menos de 16	4920	2550	2370
		16-64	32110	16179	15931
		65 ó más	8055	3531	4524
48020-Bilbao	288	TOTAL	349972	165805	184167
		Menos de 16	41330	21334	19996
		16-64	235229	115192	120037
		65 ó más	73413	29279	44134
48902-Erandio	20	TOTAL	22422	11092	11330
		Menos de 16	2619	1364	1255
		16-64	16358	8303	8055
		65 ó mas	3445	1425	2020
48029-Etxebarri	5	TOTAL	7043	3572	3471
		Menos de 16	934	486	448
		16-64	5315	2738	2577
		65 o mas	794	348	446
48044-Getxo	64	TOTAL	82285	39136	43149
		Menos de 16	11729	5999	5730
		16-64	58301	28279	30022
		65 ó más	12255	4858	7397
48054-Leioa	20	TOTAL	28381	13929	14452
		Menos de 16	4189	2126	2063
		16-64	21069	10512	10557
		65 o mas	3123	1291	1832
48071-Muskiz	5	TOTAL	6558	3232	3326
		Menos de 16	865	426	439
		16-64	4503	2291	2212
		65 o mas	1190	515	675
48083-Ortuella	6	TOTAL	8684	4224	4460
		Menos de 16	1109	562	547
		16-64	6073	3043	3030
		65 o mas	1502	619	883
48078-Portugalete	40	TOTAL	51066	24880	26186
		Menos de 16	5703	3002	2701
		16-64	35983	17972	18011
		65 o mas	9380	3906	5474
48082-Santurtzi	38	TOTAL	47173	23032	24141
		Menos de 16	5609	2923	2686
		16-64	33174	16589	16585
		65 o mas	8390	3520	4870
48084-Sestao	26	TOTAL	31773	15500	16273
		Menos de 16	3702	1905	1797
		16-64	21874	11079	10795
		65 o mas	6197	2516	3681
48904-Sondika	3	TOTAL	3978	1921	2057
		Menos de 16	503	252	251
		16-64	2613	1314	1299
		65 o mas	862	355	507
48080-Trapagaran	11	TOTAL	12621	6216	6405
		Menos de 16	1475	788	687
		16-64	9017	4548	4469
		65 ó más	2129	880	1249
48913-Zierbena	1	TOTAL	1215	659	556
		Menos de 16	131	80	51
		16-64	853	475	378
		65 ó más	231	104	127
TOTAL	658	TOTAL	801770	385972	415798
		Menos de 16	96080	49500	46580
		16-64	554060	2744462	279598
		65 ó más	151930	62010	89920

Tabla 3.- Número de secciones censales y de habitantes (hombres y mujeres) en el Área Metropolitana de Bilbo (Fuente: Censo de Población y Viviendas, 2001).



Figura 2.- Mapa del Área Metropolitana de Barcelona (Fuente: Construcción propia)



Figura 3.- Mapa del Área Metropolitana de Bilbo (Fuente: Construcción propia)

3.3.- Diseño

Estudio observacional ecológico con datos en rejilla.

Incidencia del cáncer

Las unidades de análisis fueron los municipios de la Región Sanitaria Girona.

Mortalidad por cáncer

Las unidades de análisis fueron las secciones censales de las Áreas Metropolitanas.

3.4.- Variables dependientes

Casos observados de las neoplasias de interés medioambiental:

- Neoplasias Linfoides (CIE-10: C81-C85; C90-C91)
 - Linfomas no Hodgkin (CIE-10: C82-C85; C96, CIE-9: 201)
 - Linfomas (o enfermedad) de Hodgkin (CIE-10: C81, CIE-9: 202)
 - Mielomas (CIE-10: C90, CIE-9: 203.0)

- Tiroides (CIE-10: C73; CIE-9: 193)
- Cáncer de vejiga urinaria (CIE-10: C67, CIE-9: 188)
- Cáncer de tráquea, bronquios y pulmón (CIE-10: C-34; CIE-9: 162)

Fuente: Registro de Cáncer de Girona; Registre de Mortalitat de Catalunya, Servei d'Informació i Estudis, Direcció General de Recursos Sanitaris, Departament de Salut; Construcción propia.

Población de referencia para los casos esperados

Para calcular la probabilidad que un individuo desarrolle la enfermedad de interés, se necesita definir la *población de referencia*.

En el caso del análisis de la incidencia se tomó como población de referencia la incidencia que correspondía al conjunto de unidades bajo estudio, ya que no fue posible conocer la población realmente expuesta. Como denominador se tomó la población del Censo de Población y Viviendas, 2001 (INE, 2004).

En el caso del análisis de la mortalidad se tomó como población de referencia la Española, según el Censo de Población y Viviendas del 2001.

Fuente: Censo de Población y Viviendas, 2001 (INE, 2004); Agència de Salut Pública de Barcelona; Construcción propia.

3.5.- Variables explicativas

- Variables socioeconómicas

Fuente: Censo de Población y Viviendas, 2001 (INE, 2004); Agència de Salut Pública de Barcelona; Construcción propia.

Desempleo: porcentaje de personas de 16 o más años sin empleo, desempleados y desempleados que buscan trabajo por primera vez, respecto del total de la población económicamente activa (hombres y mujeres).

El empleo es la fuente principal de ingresos económicos para la mayoría de personas. Además facilita la integración social de manera que su ausencia puede ser un buen indicador de privación. El porcentaje de desempleados de un área suele ser la variable más utilizada, siendo un reflejo del grado de malestar social de la población (Haynes et al, 1996; Ocaña-Riola et al, 2006).

Trabajadores manuales: porcentaje de personas de 16 o más años, con las siguientes ocupaciones:

- Trabajadores de los servicios de restauración, personales, protección y vendedores de los comercios.
- Trabajadores cualificados en la agricultura y en la pesca
- Artesanos y trabajadores cualificados de las industrias manufactureras, la construcción y la minería, excepto los operadores de instalaciones y maquinaria.
- Operadores de instalaciones y maquinaria y montadores.
- Trabajadores no cualificados.

respecto a la población activa de 16 o más años (hombres y mujeres).

La ocupación es un indicador del grado de bienestar material y prestigio social. Suele estar asociada con el acceso a diferentes recursos sociales, incluidos los servicios sanitarios. En este sentido, el porcentaje de trabajadores manuales es una variable utilizada con frecuencia como indicador indirecto de privación material (Townsend, 1990; Carstairs y Morris, 1991; Ocaña-Riola et al, 2006).

Instrucción insuficiente: porcentaje de personas de 16 años o más respecto a la población de 16 o más años que:

- No saben leer o escribir.
- Saben leer o escribir pero fueron menos de 5 años a la escuela.
- Fueron a la escuela 5 o más años pero sin completar EGB, ESO o Bachillerato Elemental.

La educación está relacionada con el acceso a la información. Es un indicador asociado de forma indirecta con futuros ingresos y con el nivel de vida de la población. El porcentaje de analfabetos, o de instrucción insuficiente, suele ser la variable más

utilizada en este contexto, siendo un buen indicador de privación material extrema (Domínguez–Berjón et al, 2001; Ocaña–Riola et al, 2006)

Estos indicadores se combinaron mediante el Análisis de Componentes Principales (ACP) a fin de construir un indicador sintético (Anderson, 1984). Con el propósito de estabilizar la varianza y aproximarse a una distribución normal, las variables fueron estandarizadas previamente. Las condiciones de aplicación del ACP se comprobaron mediante los contrastes de Barlett y de Kaiser–Meyer–Ollkin (KMO).

En las Tablas 4 a 6 se muestran los descriptivos de las variables socioeconómicas. De las tres zonas estudiadas la Regió Sanitària Girona fue la que presentó una mayor tasa de desempleo y una mayor proporción de trabajadores manuales. El Área Metropolitana de Bilbo fue la que presentó un menos porcentaje de instrucción insuficiente, siendo el Área Metropolitana de Barcelona la que registró un mayor porcentaje de este indicador. La tasa de desempleo fue mayor en mujeres que en hombres, siendo la diferencia particularmente en el Área Metropolitana de Bilbo. El porcentaje de trabajadores manuales, y en menor medida el de instrucción insuficiente, fue mayor en hombres que en mujeres, con la excepción, nuevamente, del Área Metropolitana de Bilbo, en el que el porcentaje de instrucción insuficiente fue mayor en las mujeres. Es de destacar el porcentaje de trabajadores manuales en ambos sexos, mucho mayor en la Regió Sanitària Girona que en el resto de zonas. Estas cifras podrían mostrar un predominio de la población no urbana en dicha región.

Regió Sanitària Girona

	Media	Mediana	Desv. típ.	Mínimo	Máximo	Percentiles	
						25	75
Desempleo, hombres	,0637	,0564	,03640	,00	,29	,0398	,0810
Desempleo, mujeres	,0842	,0753	,04211	,00	,29	,0564	,1052
Trabajadores manuales, hombres	,6222	,6416	,11260	,22	,85	,5603	,6982
Trabajadores manuales, mujeres	,5515	,5638	,10026	,21	,77	,5027	,6171
Instrucción insuficiente, hombres	,1149	,1029	,07166	,00	,51	,0659	,1457
Instrucción insuficiente, mujeres	,0714	,0640	,04048	,00	,27	,0435	,0908

Tabla 4.- Características sociodemográficas de los municipios de la Regió Sanitària Girona (Fuente: Censo de Población y Viviendas, 2001 (INE, 2004); Agència de Salut Pública de Barcelona; Construcción propia).

Área Metropolitana de Barcelona

	Media	Mediana	Desv. típ.	Mínimo	Máximo	Percentiles	
						25	75
Desempleo, hombres	,1000	,0939	,03847	,01	,37	,0751	,1171
Desempleo, mujeres	,1154	,1096	,03671	,03	,36	,0914	,1334
Trabajadores manuales, hombres	,4742	,4835	,17556	,05	,81	,3378	,6295
Trabajadores manuales, mujeres	,4290	,4279	,15326	,09	,77	,3050	,5657
Instrucción insuficiente, hombres	,1236	,1081	,07749	,01	,49	,0627	,1708
Instrucción insuficiente, mujeres	,0930	,0858	,04811	,01	,30	,0564	,1247

Tabla 5.- Características sociodemográficas de las secciones censales de las ciudades del Área Metropolitana de Barcelona (Fuente: Censo de Población y Viviendas, 2001 (INE, 2004); Agència de Salut Pública de Barcelona; Construcción propia).

Área Metropolitana de Bilbo

	Mediana	Desv. típ.	Mínimo	Máximo	Percentiles	
					25	75
Desempleo, hombres	,10975	,042296	,027	,365	,08700	,13663
Desempleo, mujeres	,19935	,056679	,045	,489	,16398	,23853
Instrucción insuficiente, hombres	,0242	,02221	,00	,14	,0129	,0381
Instrucción insuficiente, mujeres	,0438	,03074	,00	,18	,0266	,0633
Trabajadores manuales, ambos sexos	,5727	,18538	,14	,91	,3818	,6776

Tabla 6.- Características sociodemográficas de las secciones censales de las ciudades del Área Metropolitana de Bibo (Fuente: Censo de Población y Viviendas, 2001 (INE, 2004); Agència de Salut Pública de Barcelona; Construcción propia).

Como resultado de ACP se obtuvieron, en cada una de las tres zonas de estudio, índices altamente correlacionados con el porcentaje de trabajadores manuales, el desempleo y la instrucción insuficiente.

En el caso de Regió Sanitària Girona se retuvieron dos factores que acumularon cerca de un 80% de de la varianza total en ambos sexos (un 45% el primero y un 35% el segundo). El primer factor estuvo muy correlacionado con el desempleo y con la variable de trabajadores manuales y el segundo, con la instrucción insuficiente. En las Áreas Metropolitanas de Barcelona y Bilbo se retuvo un único factor en cada una de ellas, acumulando cerca de un 70% de la varianza total en ambos sexos en Barcelona y 65% en Bilbo.

La significación del contraste de esfericidad de Barlett ($p < 0,0001$ en todos los casos) y el elevado valor KMO (mínimo 0,4113, en las mujeres de la Regió Sanitària Girona, máximo 0,68 en mujeres en el Área Metropolitana de Bilbo) mostraron la adecuación del análisis realizado.

En el caso de Regió Sanitària Girona, las zonas más deprimidas, en ambos sexos, se correspondían con la mayor parte de la comarca del Ripollés, Selva interior y Alt Empordà interior. Contrariamente, las zonas menos deprimidas se correspondían con las comarcas costeras y algunos municipios del interior (Camprodon y Queralbs entre otros) (ver Figura 4).

En el caso del Área Metropolitana de Barcelona, se observa como las zonas más deprimidas coincidían con los municipios de El Prat, Cornellà, Hospitalet, Badalona, Sant Adrià y Santa Coloma; y con barrios de Barcelona como Nou Barris, La Mina, el Barri Vell y Raval (ver Figura 5).

La distribución geográfica del índice de privación estuvo más claramente en el Área Metropolitana de Bilbo (ver Figura 6). Las zonas más deprimidas se sitúan en el sur (distritos sur de Bilbo y Barakaldo) y en el oeste (Trapagarán, Ortuella, Abanto y Zierbena y Muskiz). Por el contrario, las zonas menos deprimidas se corresponden con las secciones del centro de Bilbo y el norte del área (Getxo y Leioa). Es de destacar, en el caso de los hombres que entre estas últimas zonas también se sitúa Sondika.

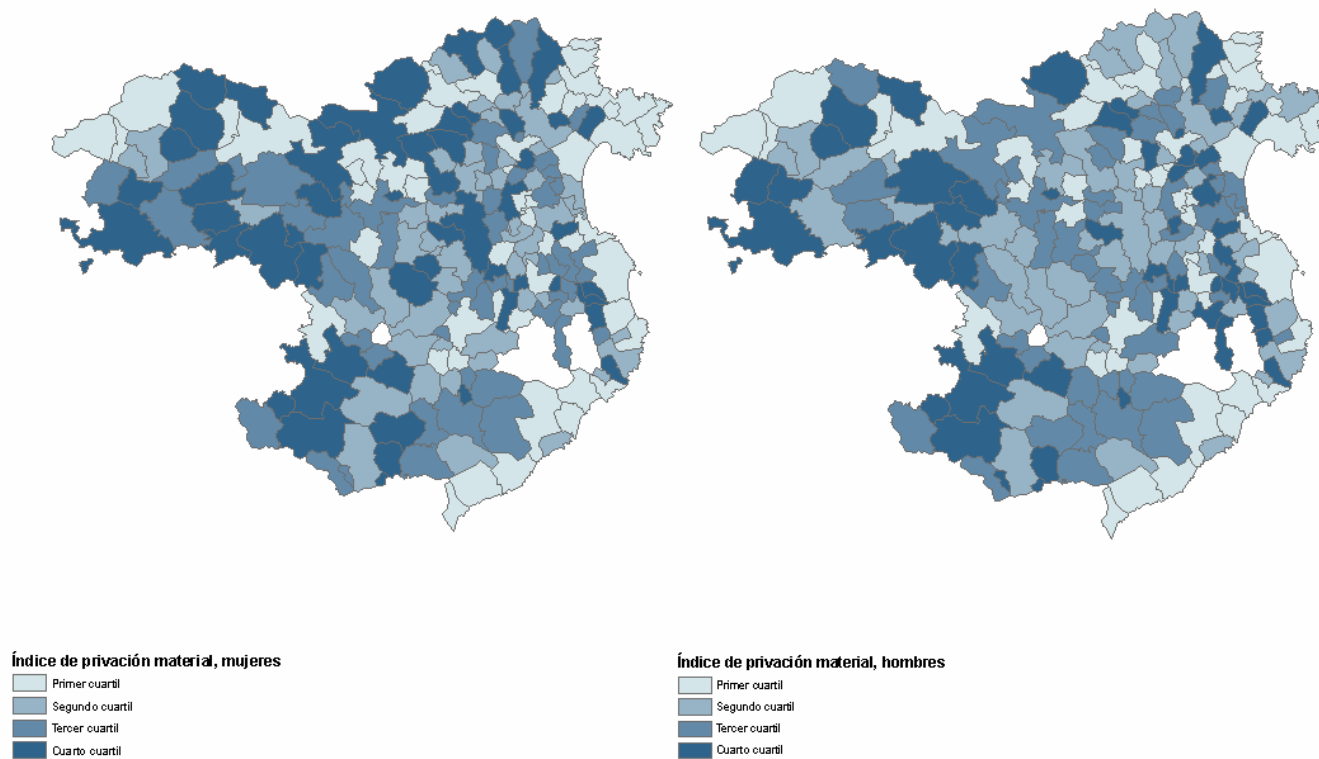


Figura 4.- Primer factor del Índice de privación material, en mujeres y hombres, en la Regió Sanitària Girona (Fuente: Construcción propia).

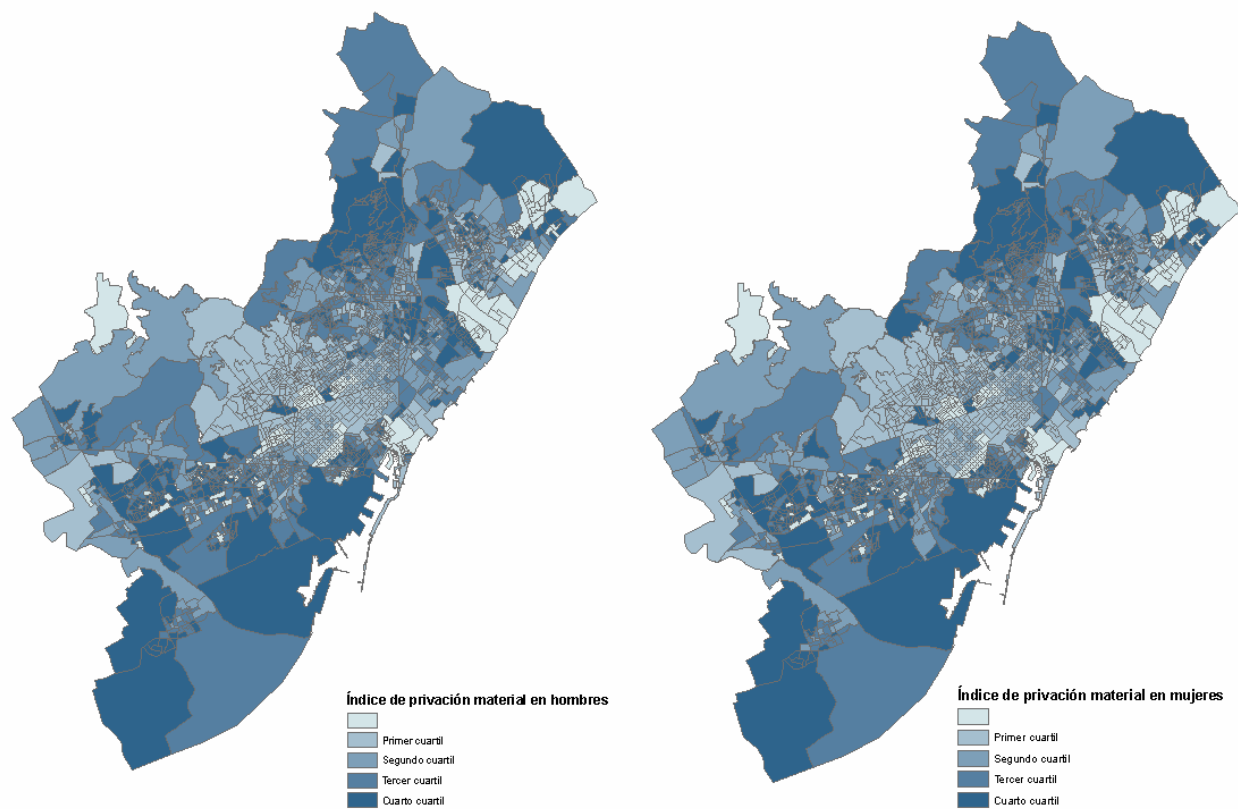


Figura 5.- Índice de privación material, en mujeres y hombres, en el Área Metropolitana de Barcelona (Fuente: Construcción propia).

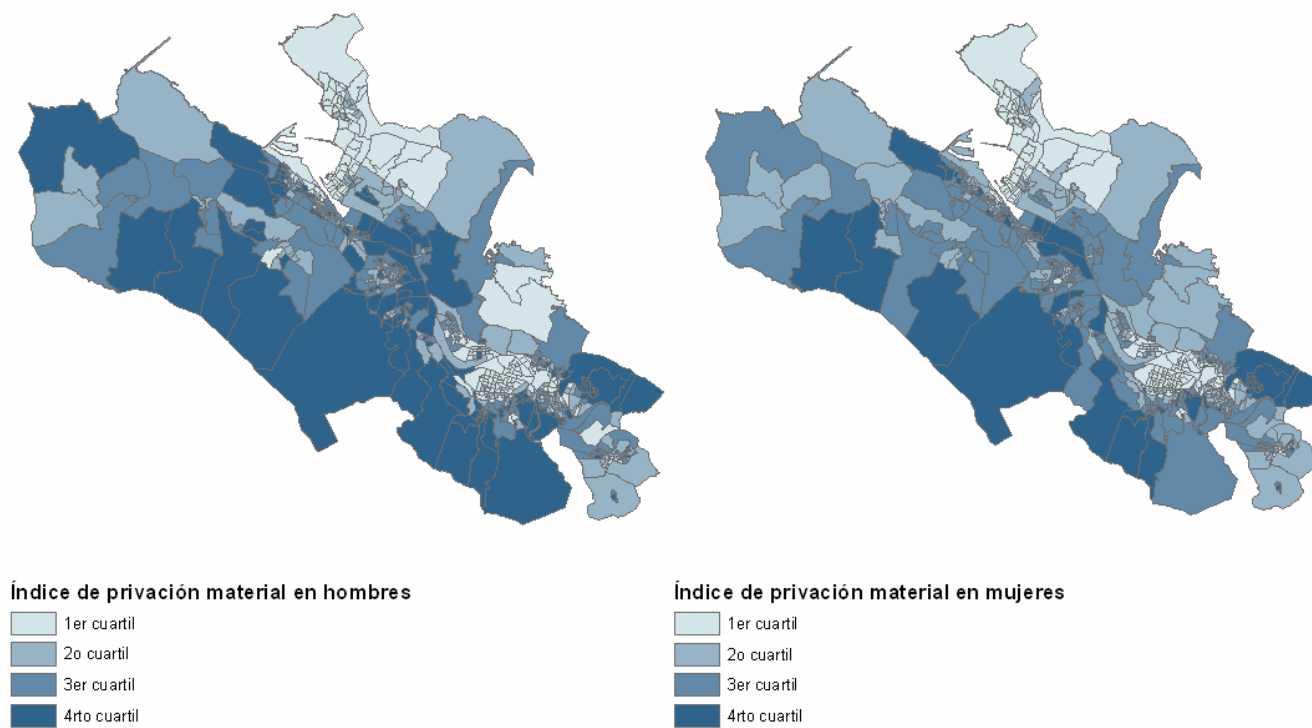


Figura 6.- Índice de privación material, en mujeres y hombres, en el Área Metropolitana de Bilbo (Fuente: Construcción propia).

- Variables medioambientales.

Fuente: Los datos de contaminantes del Área Metropolitana de Barcelona y de la Regió Sanitària Girona se obtuvieron del Àrea d'Immisions, Departament de Medi Ambient i Habitatge, Generalitat de Catalunya. Los datos de contaminantes de Bilbo se obtuvieron del Àrea de Recursos Ambientales de Aire y Ruido del Departamento de Medio Ambiente y Ordenación del Territorio, Gobierno Vasco.

Se recogieron los niveles de contaminación atmosférica de dióxido de nitrógeno (NO₂) y dióxido de azufre (SO₂), para los períodos 1996–1998; 1999–2001 y 2002–2003². El contaminante registrado en la Regió Sanitària Girona (período 1994–2001) fue el ozono (O₃).

Se consideró una estación válida si tenía menos de un 50% de datos perdidos. De esta manera, para el Área Metropolitana de Barcelona se escogieron hasta 18 estaciones medidoras de contaminación atmosférica (7 en Barcelona, 2 en Sant Feliu de Llobregat, 2 en Montcada i Reixac y 1 en cada una de las ciudades restantes) (ver Figura 7), para la Regió Sanitària Girona se escogieron 5 (1 en Agullana, 1 en Pardines, 1 en Begur, 1 en Santa Pau, 1 en Sarrià de Ter y 1 en Cassà de la Selva) y 62 en el Área Metropolitana de Bilbo (21 en Bilbo, 5 en Getxo, 4 en Barakaldo, 4 en Sestao, 4 en Erandio, 2 en Muskiz, 2 en Basauri, 2 en Trapagarán, 2 en Leioa y 1 en cada uno del resto de los municipios).

² NO₂ y SO₂ son los únicos contaminantes para los que existen estaciones con datos válidos en las dos áreas metropolitanas, aunque existen otros contaminantes en los que sólo se disponían de estaciones con datos válidos en una de las áreas (CO, O₃ en Barcelona y partículas en Bilbo)

La situación geográfica de las estaciones en el caso del Área Metropolitana de Barcelona (ver Figura 7), resulta poco óptima para captar la inmisión. Un poco menos de la mitad de las estaciones se sitúan justo al lado de las vías del tren y lejos de las fuentes de inmisión de la ciudad. Además, la estación situada en Sant Adrià del Besòs captura las emisiones de la térmica ya que se localiza a poco más de 100 metros. Sólo tres estaciones son las que verdaderamente capturan la inmisión ya que se sitúan en zonas donde es apreciable la intensidad del tráfico.

En el caso de las estaciones del Área Metropolitana de Bilbo (ver Figura 8), se han representado en el mapa solamente aquellas expuestas en la página web del Gobierno Vasco, no se ha podido fotografiar el resto de las estaciones debido a la difícil accesibilidad. La mayor parte de las estaciones que figuran en el mapa, están ubicadas en parques urbanos o zonas donde la dispersión del aire es alta. La única estación que se sitúa cerca de un foco contaminante, a unos 200 metros, es la de Muskiz, tal y como se puede apreciar en el mapa. En este caso los datos de ciertos contaminantes podrían dar un sesgo de la inmisión.

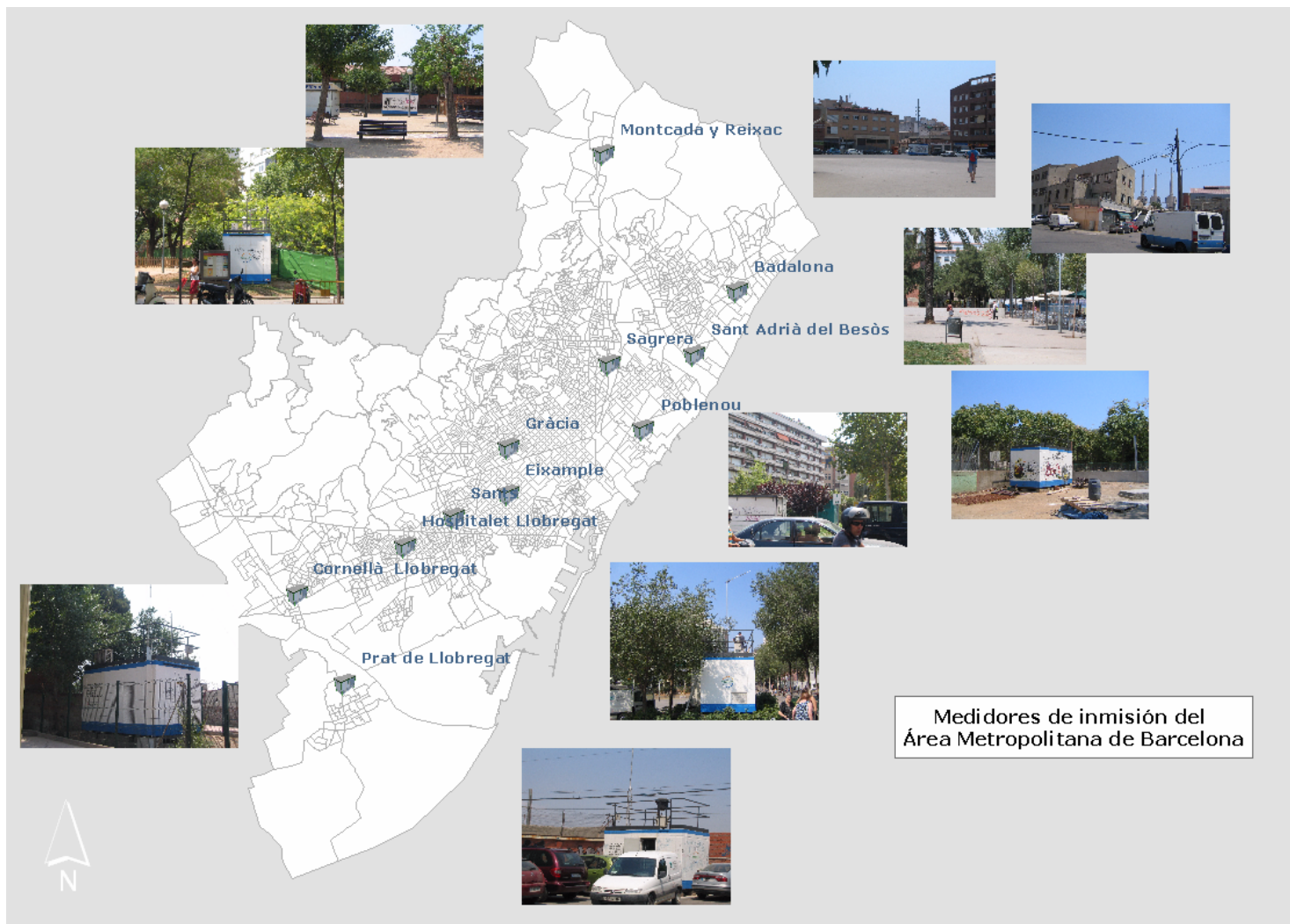


Figura 7.- Medidores de inmisión del Área Metropolitana de Barcelona. Julio 2006. Fuente: construcción propia.

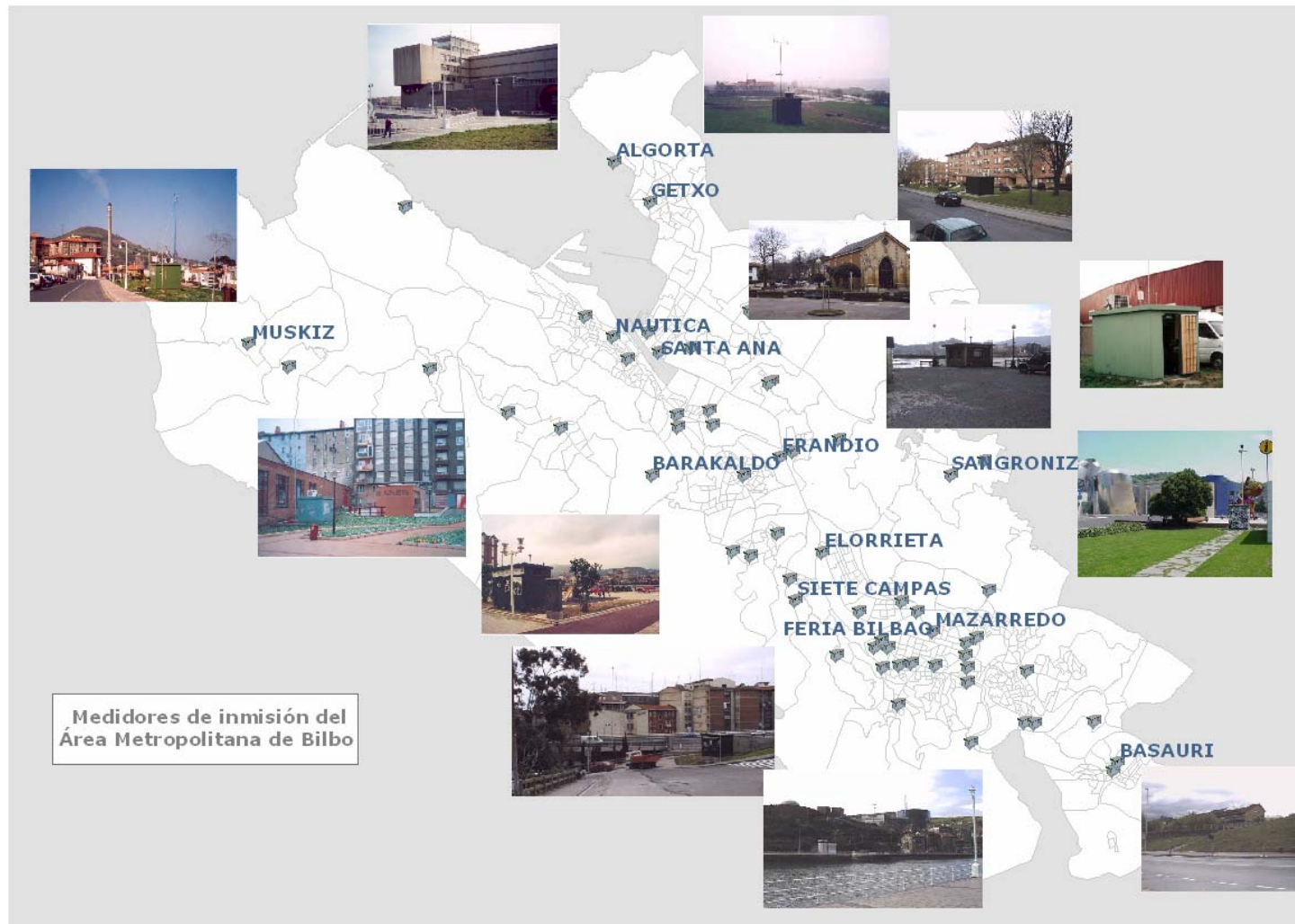


Figura 8.- Medidores de inmisión del Área Metropolitana de Bilbo. Agosto, 2006. Fuente: construcción propia (fotos importadas de la página web: http://www1.euskadi.net/vima_ai_vigilancia).

En la Tabla 7 se muestran estadísticos descriptivos de los contaminantes en las zonas estudiadas.

Área Metropolitana de Barcelona, NO₂

Primer período

no2						Percentiles		
N								
Válidos	Media	Mediana	Desv. tip.	Mínimo	Máximo	25	50	75
1068	45,2844	44,8715	15,03817	10,00	103,50	34,9583	44,8715	54,8953

Segundo período

no2						Percentiles		
N								
Válidos	Media	Mediana	Desv. tip.	Mínimo	Máximo	25	50	75
1095	48,1202	48,1260	14,88319	10,50	115,90	38,1550	48,1260	57,2200

Tercer período

no2						Percentiles		
N								
Válidos	Media	Mediana	Desv. tip.	Mínimo	Máximo	25	50	75
365	46,5451	46,0057	12,86734	18,73	82,81	36,9811	46,0057	56,0486

Área Metropolitana de Bilbo, NO₂

Primer período

no2						Percentiles		
N								
Válidos	Media	Mediana	Desv. tip.	Mínimo	Máximo	25	50	75
1088	46,3419	45,7619	11,71116	19,79	93,53	37,6705	45,7619	53,7375

Segundo período

no2						Percentiles		
N								
Válidos	Media	Mediana	Desv. tip.	Mínimo	Máximo	25	50	75
1096	41,9736	41,6842	12,65193	8,50	91,95	32,7262	41,6842	50,3602

Tercer período

no2						Percentiles		
N								
Válidos	Media	Mediana	Desv. tip.	Mínimo	Máximo	25	50	75
365	36,9235	36,0000	12,23138	11,15	77,75	28,2895	36,0000	45,3559

Área Metropolitana de Barcelona, SO₂

Primer período

so2						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
1026	15,6048	12,1683	10,89594	2,00	79,37	11,0000	12,1683	18,1108

Segundo período

so2						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
1094	9,7183	7,7675	6,77125	1,00	38,62	4,6143	7,7675	13,2903

Tercer período

so2						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
730	3,7407	3,1850	2,17972	1,00	14,64	2,1403	3,1850	4,8296

Área Metropolitana de Bilbo, SO₂

Primer período

so2						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
1079	16,0811	14,6667	6,20156	6,27	47,54	11,7500	14,6667	18,8947

Segundo período

so2						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
1071	15,5952	14,5882	5,27729	6,50	40,41	11,8000	14,5882	18,5000

Tercer período

so2						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
720	11,3578	10,4444	4,32891	4,61	31,50	8,2222	10,4444	13,2864

Regió Sanitària Girona, O₃

Período: 1994–2001

o3						Percentiles		
N								
Válidos	Media	Mediana	Desv. típ.	Mínimo	Máximo	25	50	75
2922	65,5330	64,5395	16,59615	7,50	120,63	52,8590	64,5395	76,7655

Tabla 7.– Descriptivo de los contaminantes (Fuente: Construcción propia).

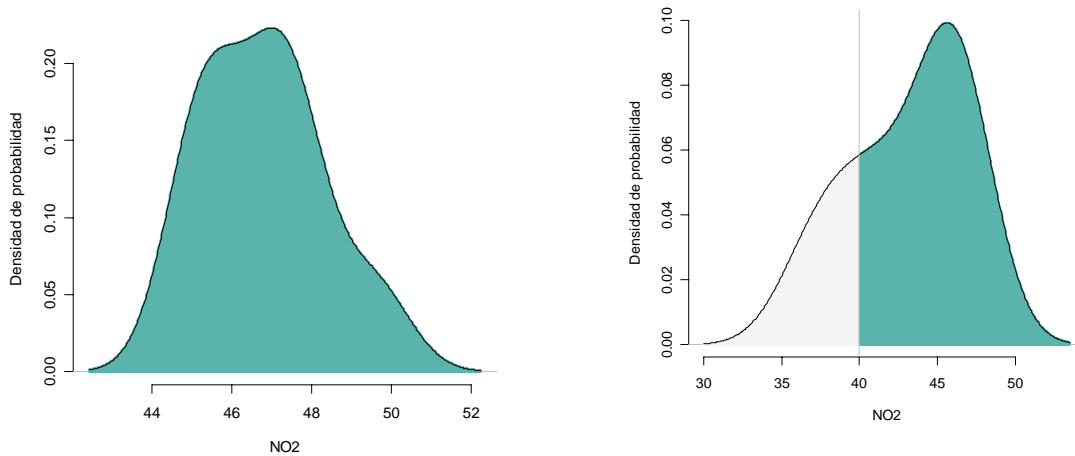


Figura 9.– Función de densidad de las concentraciones de NO₂ en las Áreas Metropolitanas de Barcelona (izquierda) y Bilbo (derecha) (Fuente: construcción propia).

En la Figura 9 se muestran las funciones de densidad de las concentraciones de NO₂ en las dos áreas metropolitanas. El color verde representa la proporción de días que el nivel de NO₂ superó el límite legal fijado en 40µg/m³ (promedio anual), según el Real Decreto 1073/2002. En el área metropolitana de Bilbo, el 70% de los días en el período 1996–2003 superaron el límite legal. En el caso del área metropolitana de Barcelona, se superaron todos los días. De hecho, según el Decreto 226/2006 de la Generalitat de Catalunya, dado que los niveles de calidad del aire por dióxido de nitrógeno son superiores a los admisibles, con un riesgo de aceptación a la salud humana y al medio ambiente, se declararon zonas de protección especial para el contaminante dióxido de nitrógeno y para las partículas todos los municipios que integran el área metropolitana en este trabajo.

Por lo que se refiere a los niveles de SO₂, según se observa en la Figura 10, el máximo se sitúa a bastante distancia del límite legal en las dos áreas (125µg/m³ promedio diario). Debe señalarse que el comportamiento del SO₂ fue bastante errático, sobre todo en el Área Metropolitana de Barcelona. En este sentido esto es en gran parte resultado de una reducción continuada de los niveles de dióxido de azufre en ambas áreas (Figura 11). Parte debido a la restricción legal de la inmisión y principalmente a la pérdida de las zonas industriales en las dos zonas.

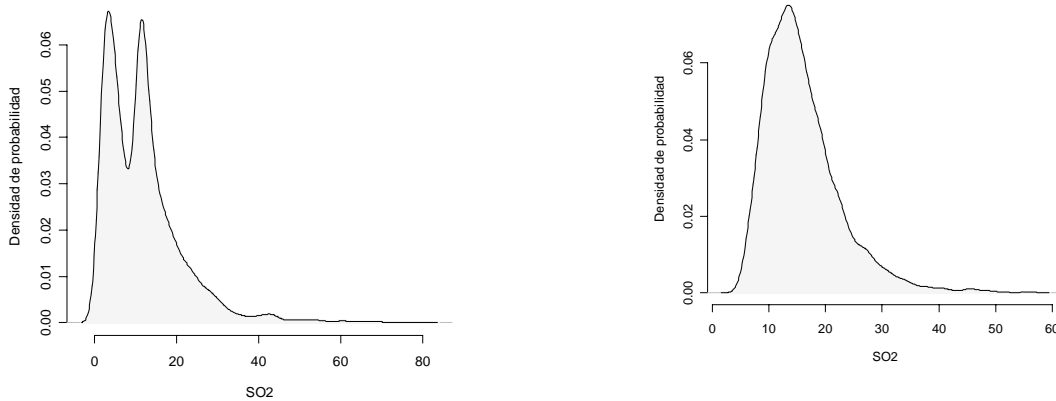


Figura 10.– Función de densidad de las concentraciones de SO₂ en las Áreas Metropolitanas de Barcelona (izquierda) y Bilbo (derecha) (Fuente: construcción propia).

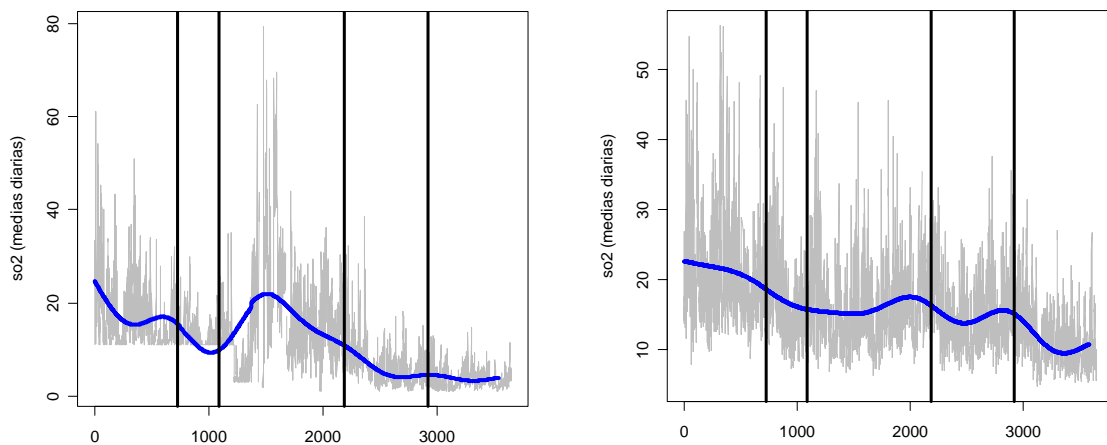


Figura 11.– Serie temporal de las concentraciones de SO₂ en las Áreas Metropolitanas de Barcelona (izquierda) y Bilbo (derecha) en los tres períodos (Fuente: construcción propia).

Por último, en relación al comportamiento de ozono en la Región Sanitaria Girona (aunque sólo se dispone de 5 estaciones de contaminante) en la Figura 12 se muestra que pocos días superaron el límite legal de 120µg/m³ (máximo de las medidas 8-horarias, Real Decreto 1796/2003).

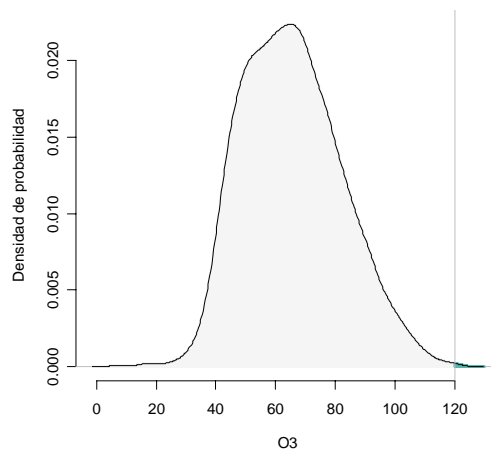


Figura 12.– Función de densidad de las concentraciones de O₃ en la Región Sanitaria Girona (Fuente: construcción propia).

El análisis que se describe de aquí en adelante se realizó para cada contaminante, área y período. Los valores perdidos en cada estación válida en un año determinado fueron substituidos utilizando imputación múltiple (Little y Rubin, 1987; Rubin, 1987; 1996). En concreto, mediante el algoritmo EM (Schafer, 1996)³, los valores perdidos fueron estimados por máxima verosimilitud (Novo y Schafer, 2003).

La idea fue la de utilizar las correlaciones de otras variables con la variable cuyos valores se deseaban imputar. Para cada una de las tres áreas se consideró la serie completa de contaminante y su relación con las series completas del mismo contaminante en otras estaciones del área. El procedimiento se repitió para todos y para cada uno de los años.

Se trata de un método óptimo, siempre y cuando se cumplan dos hipótesis (Little y Rubin, 1987). Se requiere, en primer lugar, que los datos se distribuyan según una normal multivariante. Hipótesis que se cumple en el caso de los contaminantes. En segundo lugar, los valores perdidos deben ser MAR (*missing at random*). El criterio de validez de una estación, en este caso, implica también esta propiedad.

³ Las instrucciones y la sintaxis en R y en WinBUGS se encuentran en el Anexo 5.

Una vez imputados, los datos perdidos, los niveles del contaminante se predijeron en la unidad de análisis correspondiente (secciones censales en el caso de las Áreas Metropolitanas de Barcelona y Bilbo y municipios en el caso de la Regió Sanitària Girona).

Antes de realizar las predicciones se extrajo aquella conducta temporal a largo plazo de las series de los contaminantes que pudiese distorsionar la dependencia espacial. Por este motivo, y utilizando como variable dependiente la serie del contaminante en cada una de las estaciones (para cada área y período), se estimó un modelo GLM de respuesta normal (véase Anexo 1), introduciendo como variables explicativas tantas variables ficticias como años contuviese el período menos uno³.

A continuación, tras extraer la tendencia del contaminante correspondiente (recuérdese para cada área y período), y a fin de poder determinar la distribución (espacial) de este contaminante, se estimó la correlación espacial entre las distintas estaciones. En el caso de las dos áreas metropolitanas (Barcelona y Bilbo), tanto para la estimación de la dependencia espacial, como para, sobre todo, la predicción de los niveles de contaminación por secciones censales, se utilizó el método estocástico *kriging*. Para la Regió Sanitària Girona, puesto que solo se contaba con cinco estaciones medidoras, se utilizó un método determinista, el método *IDW*.

Como se dijo en el Capítulo 2, los métodos estocásticos (*kriging*) suponen que las distintas mediciones de contaminante, es decir las *realizaciones muestrales*, corresponden en realidad a la distribución de una única variable aleatoria distribuida espacialmente, es decir a un *proceso estocástico espacial*. La variable aleatoria distribuida espacialmente debe ser estacionaria (además de normal e isotrópica). Es decir no debe existir ninguna tendencia espacial. Una tendencia espacial implicaría que en realidad no se tiene una única variable aleatoria (o un único proceso estocástico

³ Las instrucciones y la sintaxis en R y en WinBUGS se encuentran en el Anexo 5.

espacial). Si existiese alguna tendencia (espacial), podría ocurrir que determinadas zonas del área correspondiente se comportasen de una forma, y otras, de otra distinta. Si no se tuviese en cuenta la existencia de una tendencia espacial, evidentemente se estimaría de forma sesgada la distribución espacial de la variable. Afortunadamente, en ninguna de las áreas estudiadas se observó una tendencia espacial⁴.

Una vez comprobada la estacionariedad (así como la normalidad y la isotropía⁵), se procedió a la estimación de la dependencia espacial. En vez de utilizar una estrategia gráfica basada en los variogramas, se optó por una estrategia más o menos automática³. La idea fue probar, para cada uno de los días de la serie temporal de cada contaminante (de nuevo, para cada área y período) diferentes estructuras de dependencia espacial: exponencial, esférica, gaussiana y gneiting. Para cada una de las estructuras y día, se obtuvo una medida de bondad del ajuste (raíz del error cuadrático medio, RECM). Promediando los RECMs por año se obtuvo la estructura de dependencia que mejor ajustaba. Nótese que en teoría se permitió que la estructura de dependencia espacial pudiera ser diferente para cada uno de los años. En la práctica, sin embargo, en todos los casos el modelo de variograma con menor RECM fue el modelo gneiting (aunque los parámetros del modelo fueron diferentes).

Tras estimar la correlación espacial (y comprobar la bondad del ajuste) se predijeron (por área y período) los niveles para cada contaminante en otros puntos diferentes de los observados, es decir, en las diferentes secciones o municipios de las zonas estudiadas.

Para ello se utilizó el kriging ordinario. El motivo más importante fue que, puesto que los datos no presentaban ninguna tendencia espacial, no fue necesario recurrir al *kriging universal* (la alternativa más utilizada con datos isotrópicos). El *kriging*

³ Las instrucciones y la sintaxis en R y en WinBUGS se encuentran en el Anexo 5.

⁴ Para comprobarlo se utilizaron métodos gráficos, diagramas de dispersión del contaminante, contra la longitud y latitud, y métodos estadísticos, una regresión no paramétrica con variable dependiente el contaminante y las variables explicativas suavizaciones de la longitud y latitud.

⁵ La normalidad y la isotropía se comprobaron mediante métodos gráficos.

universal, además, presenta varias desventajas. En primer lugar, se asume que la forma funcional de la media no estacionaria es conocida. Como se dijo en el Capítulo 2 es difícil conocer, ni tan siquiera de forma aproximada, la forma funcional de la media. Además según Matheron (1969), proporciona estimaciones sesgadas, puesto que tanto el orden del polinomio por el que se aproxima la forma funcional como, sobre todo el semivariograma, se estiman a partir de unos residuos no estacionarios.

Para la estimación se utilizó un algoritmo Bayesiano (Ribeiro y Diggle, 2001).

En el caso de la Regió Sanitària Girona se utilizó el método determinista conocido como *Inverse Distance Weighted (IDW)*³. Tal y como se dijo en el Capítulo 2 consiste en una combinación lineal de aquellos puntos observados en el dominio espacial, utilizando como ponderaciones el inverso de la distancia, normalizado de tal forma que la suma de todas las ponderaciones fuese igual a la unidad. Como particularidad en este caso, en primer lugar, se extrajo previamente la tendencia temporal como en el kriging, por área y período. Además, a fin de estimar la incertidumbre (necesaria para los modelos de asociación geográfica), se utilizó una técnica de remuestreo. En concreto, se repitió el *IDW* para cada todos y cada uno de los años de los que se disponía.

Los resultados de la predicción se muestran en las Figuras 13 (predicción de ozono para la RSG), 14, 15 (predicción de NO₂ y SO₂ para todo el período del Área Metropolitana de Barcelona), 16, 17, 18 (predicción de NO₂ para cada uno de los períodos en el Área Metropolitana de Bilbo) 19, 20 y 21 (predicción de SO₂ para cada uno de los períodos en el Área Metropolitana de Bilbo).

³ Las instrucciones y la sintaxis en R y en WinBUGS se encuentran en el Anexo 5.

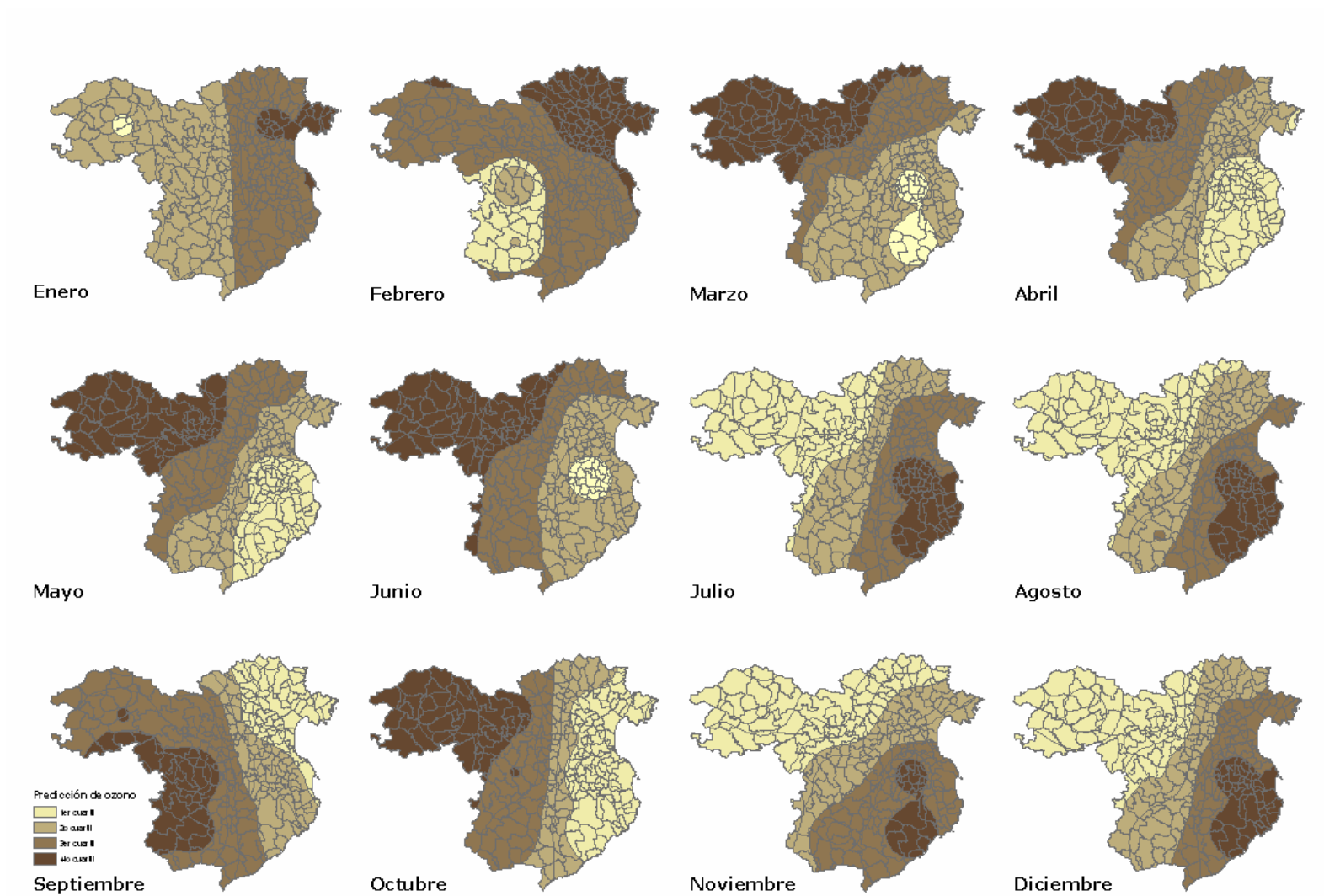


Figura 13.- Niveles de ozono en la Región Sanitaria Girona. Resultados de la predicción (*IDM*). Fuente: construcción propia.

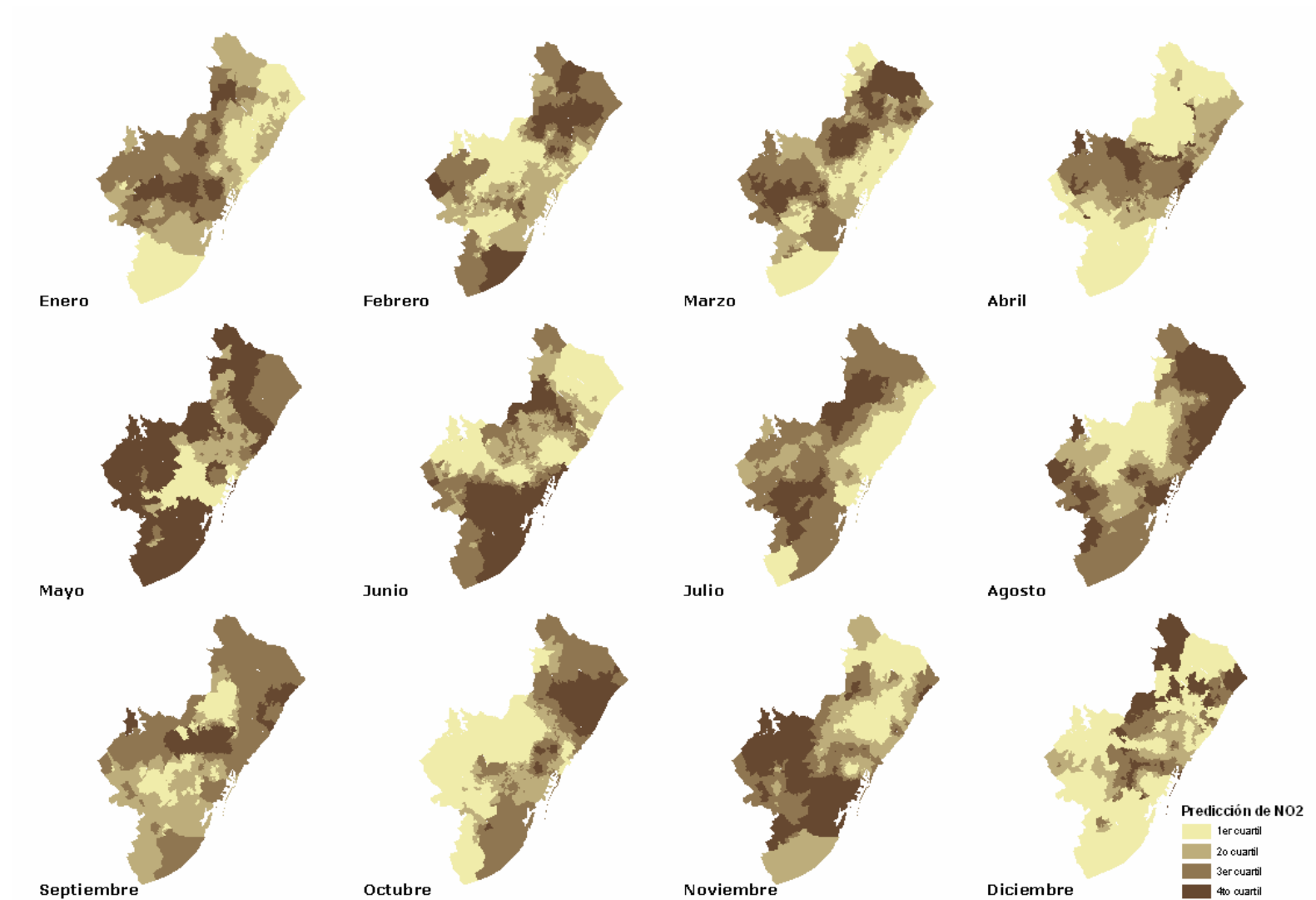


Figura 14.- Niveles de NO2 en el Área Metropolitana de Barcelona. Fuente: construcción propia.

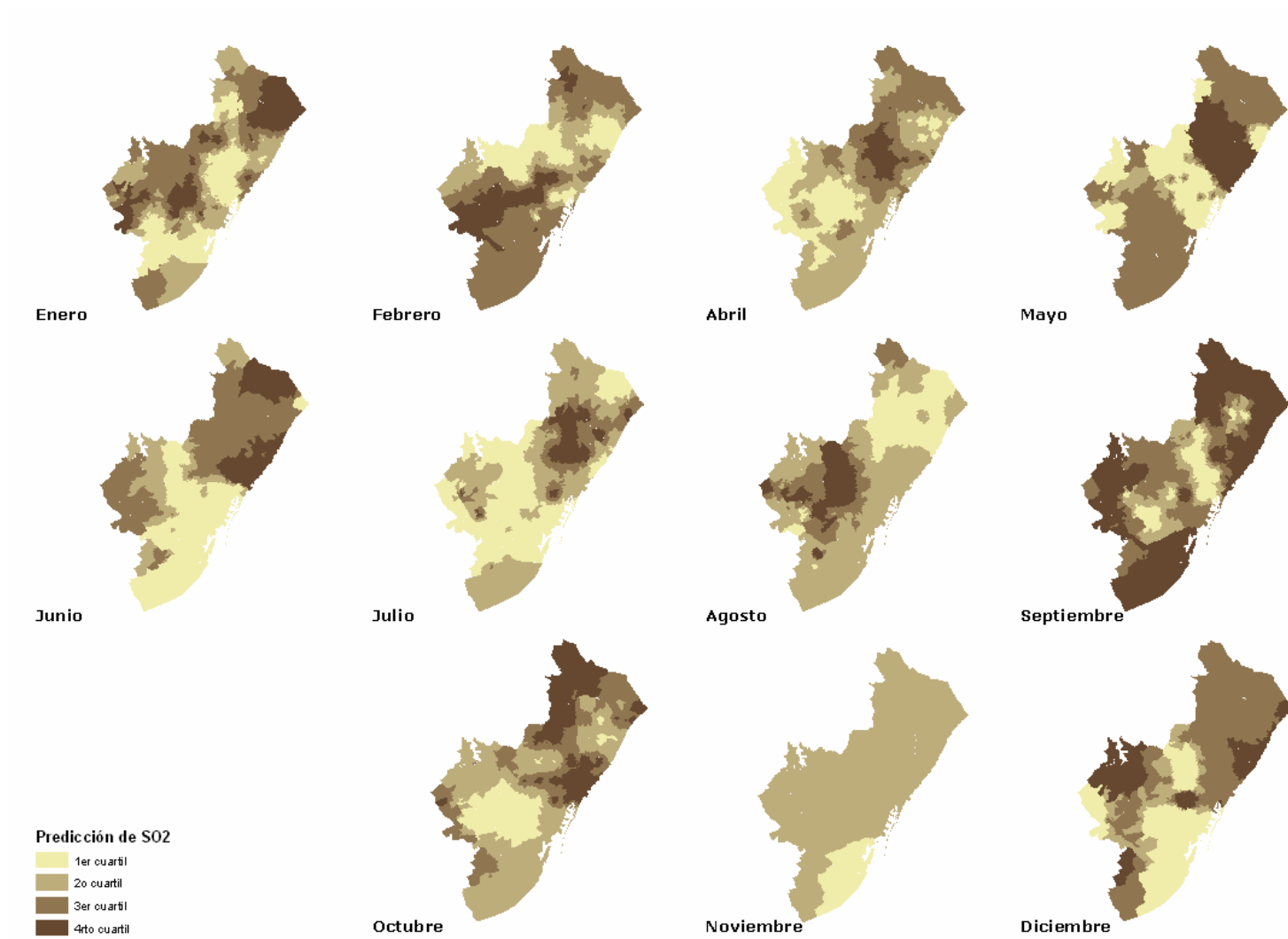


Figura 15.- Niveles de SO2 en el Área Metropolitana de Barcelona. Fuente: construcción propia. La predicción del mes de marzo no se pudo realizar por problemas de variabilidad en los datos.

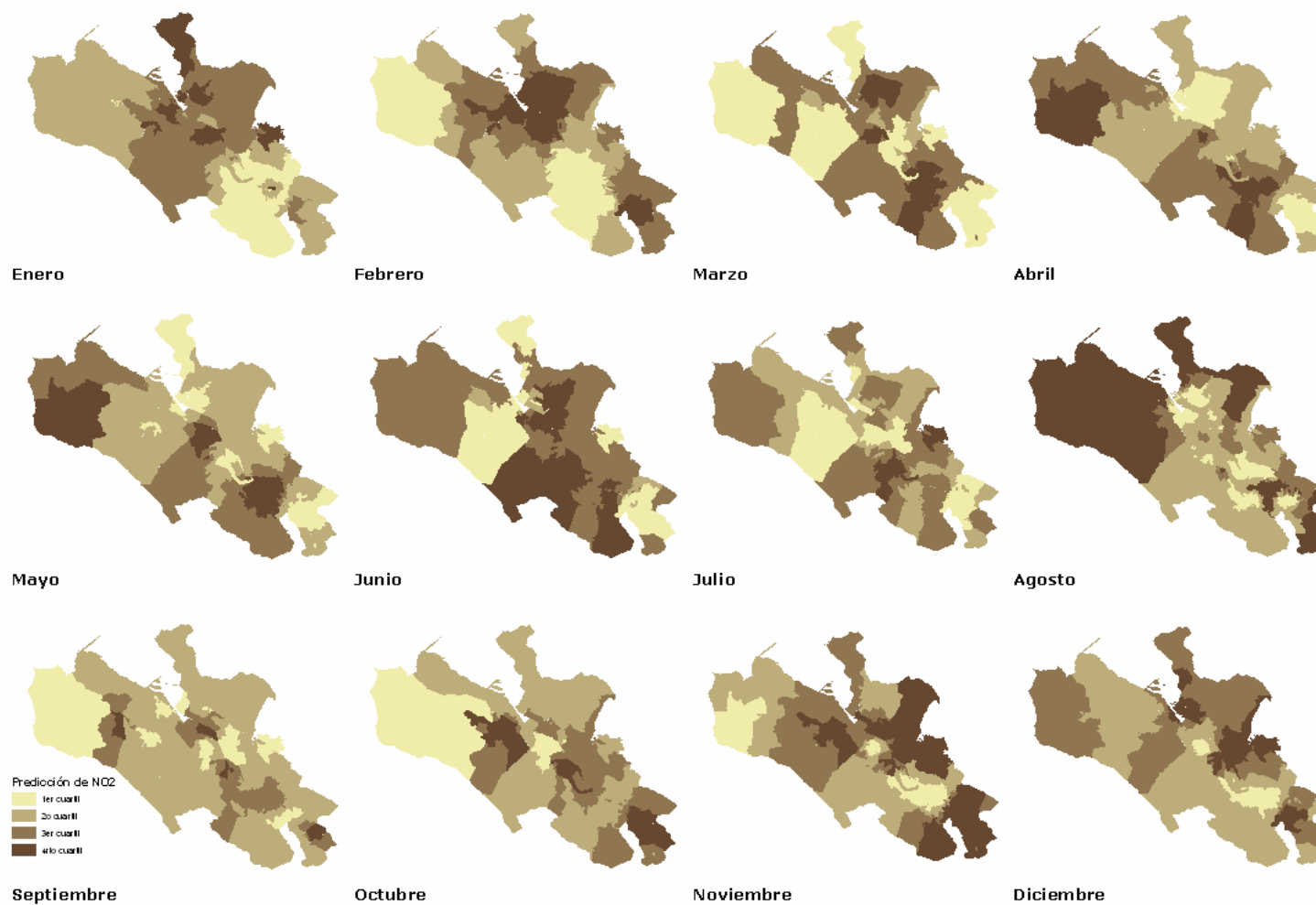


Figura 16.- Niveles de NO2 en el Área Metropolitana de Bilbo, período: 1996-1998. Fuente: construcción propia.

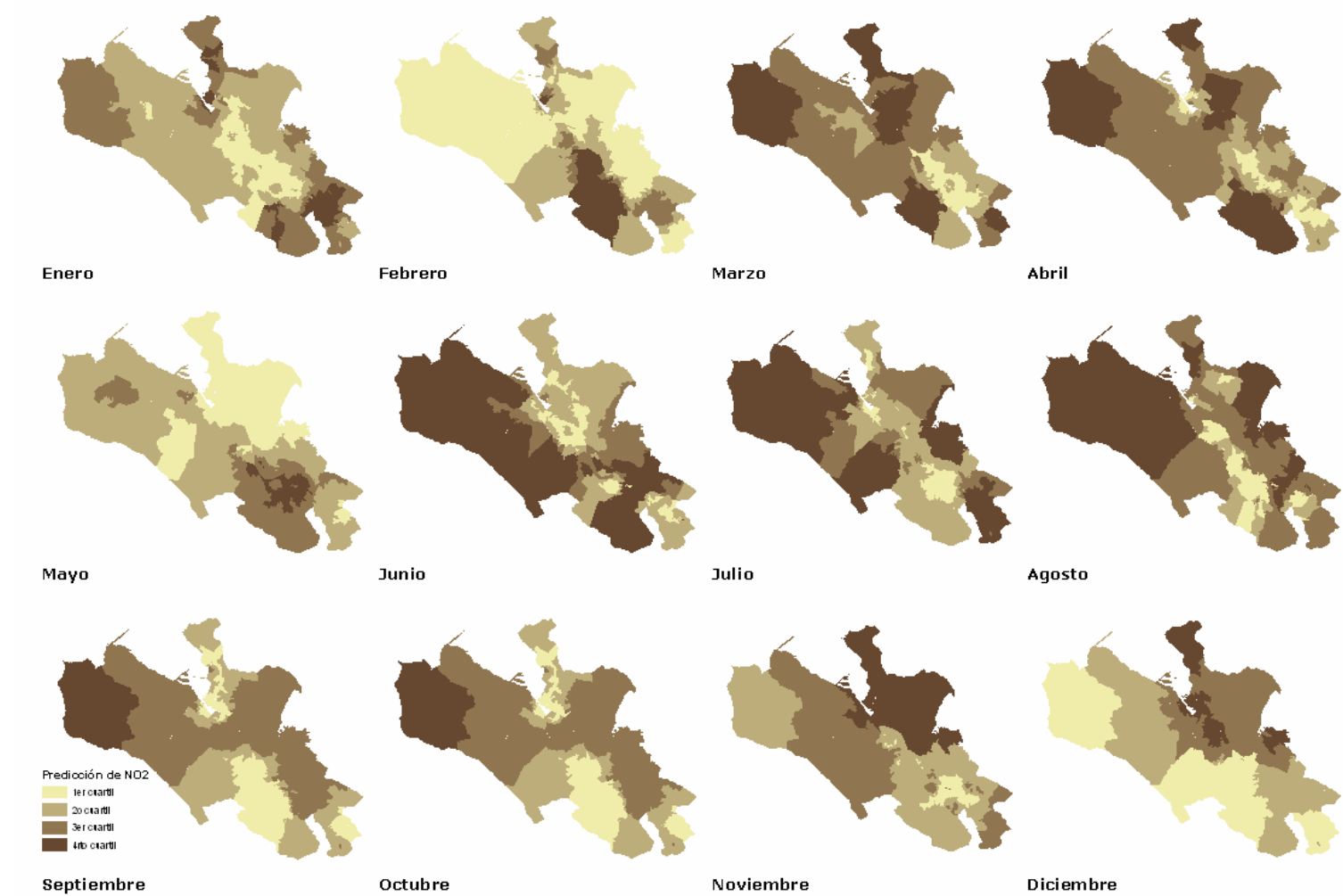


Figura 17.- Niveles de NO2 en el Área Metropolitana de Bilbo, período: 1999-2001. Fuente: construcción propia.



Figura 18.- Niveles de NO2 en el Área Metropolitana de Bilbo, período: 2002-2003. Fuente: construcción propia.



Figura 19– Niveles de SO2 en el Área Metropolitana de Bilbo, período: 1996–1998. Fuente: construcción propia.

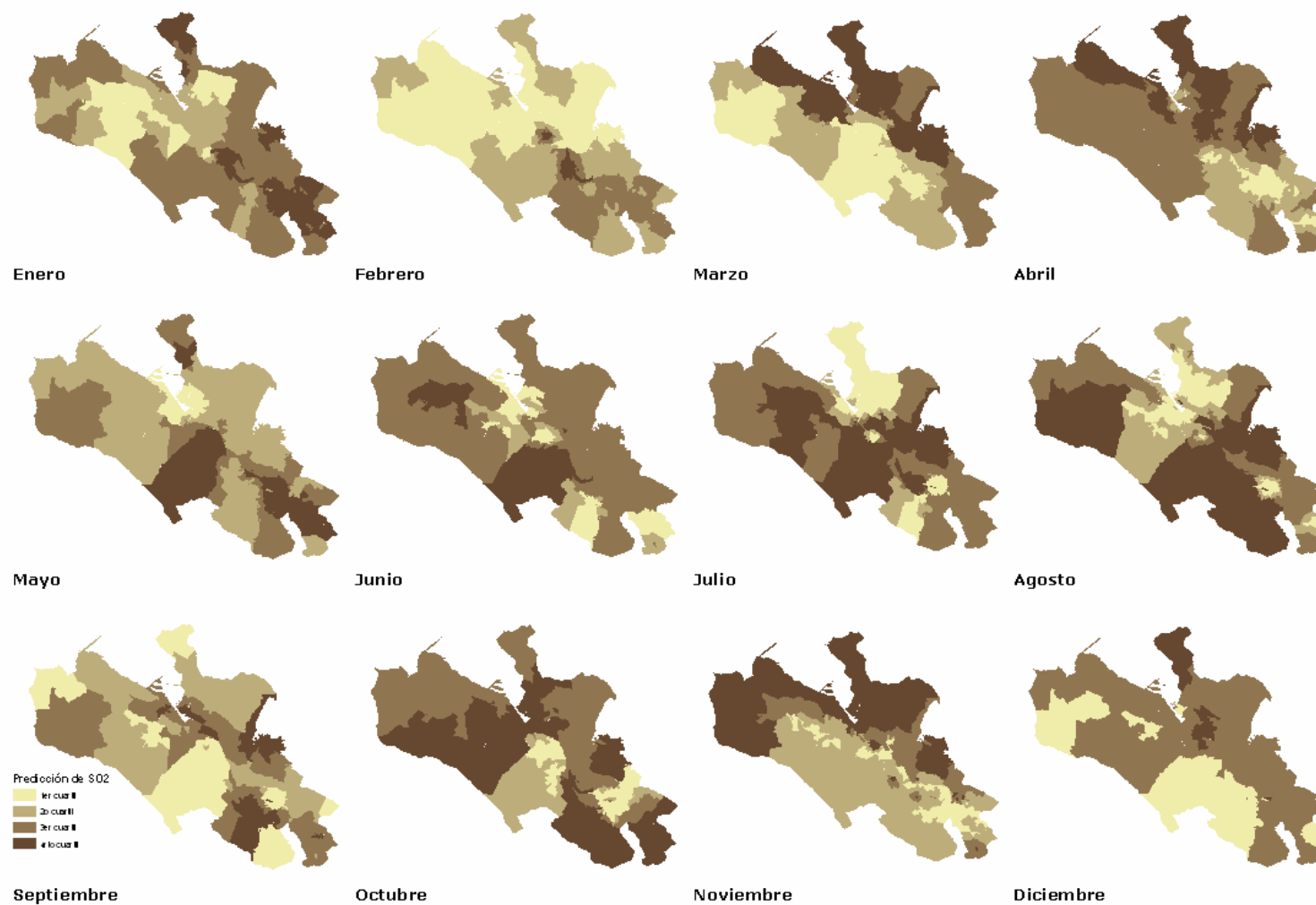


Figura 20.- Niveles de SO2 en el Área Metropolitana de Bilbo, período: 1999–2001. Fuente: construcción propia.

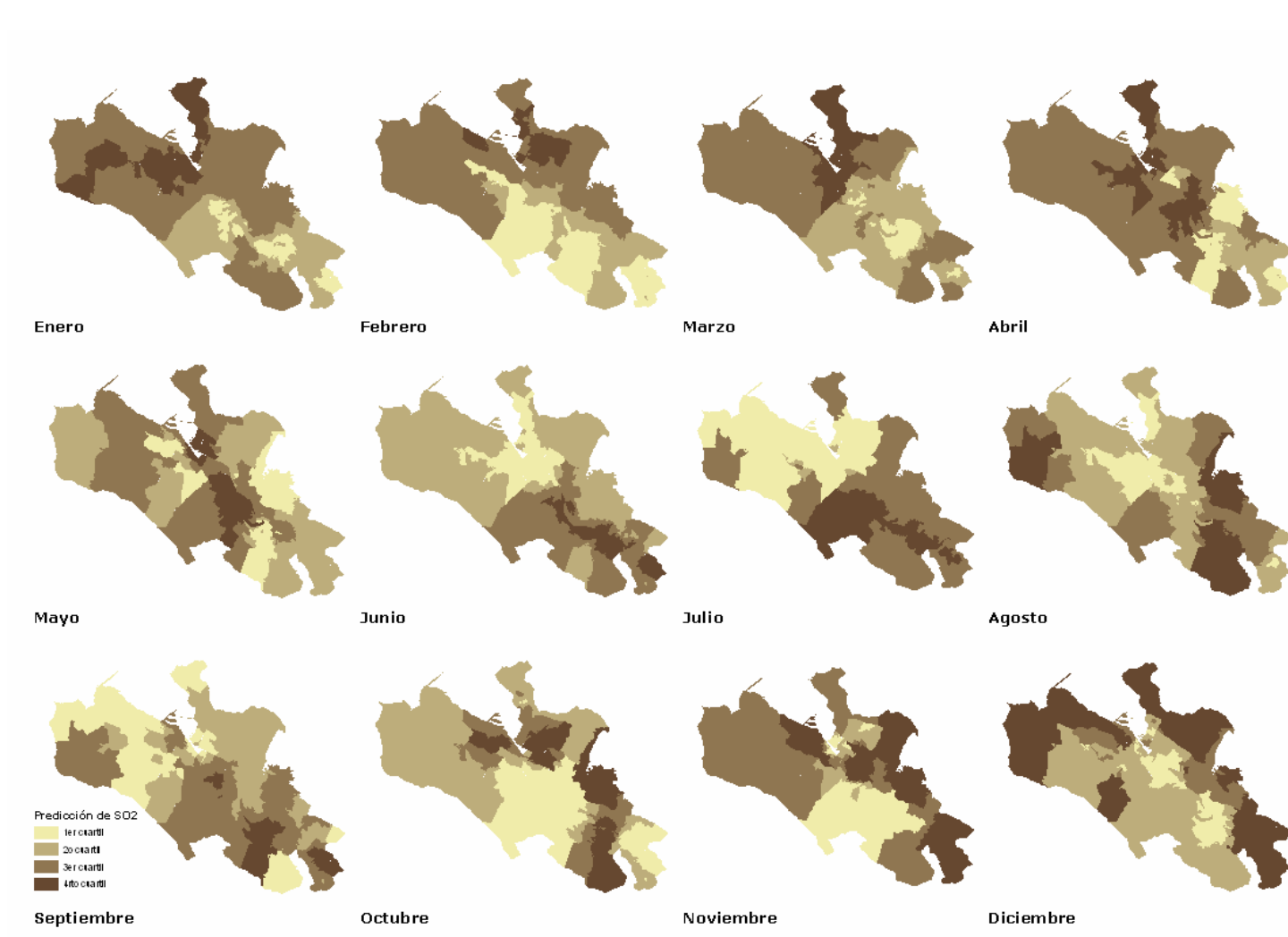


Figura 21.- Niveles de SO2 en el Área Metropolitana de Bilbo, período: 2002–2003. Fuente: construcción propia.

Para cada una de las unidades de análisis (secciones censales y municipios) se obtuvo un único valor para la media del contaminante, así como su desviación típica (para cada área y período).

En las Figuras 22 a 25 se muestra la variabilidad del contaminante dentro del área y período correspondiente.

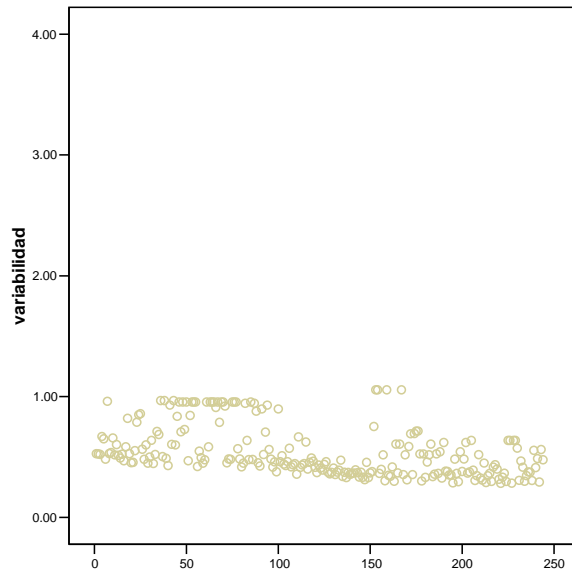
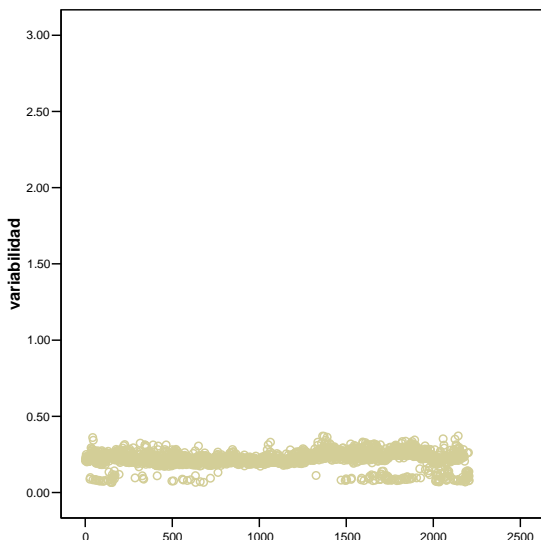
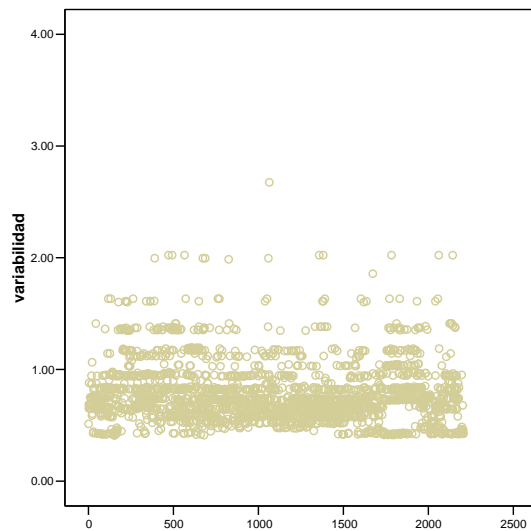


Figura 22.–Variabilidad de Ozono en los municipios en los de la Regió Sanitària Girona, período 1994–2001. Fuente: construcción propia.



a)



b)

Figura 23.– Variabilidad de NO₂ (a) y SO₂ (b) en las secciones censales del Área Metropolitana de Barcelona, período 1996–2003. Fuente: construcción propia.

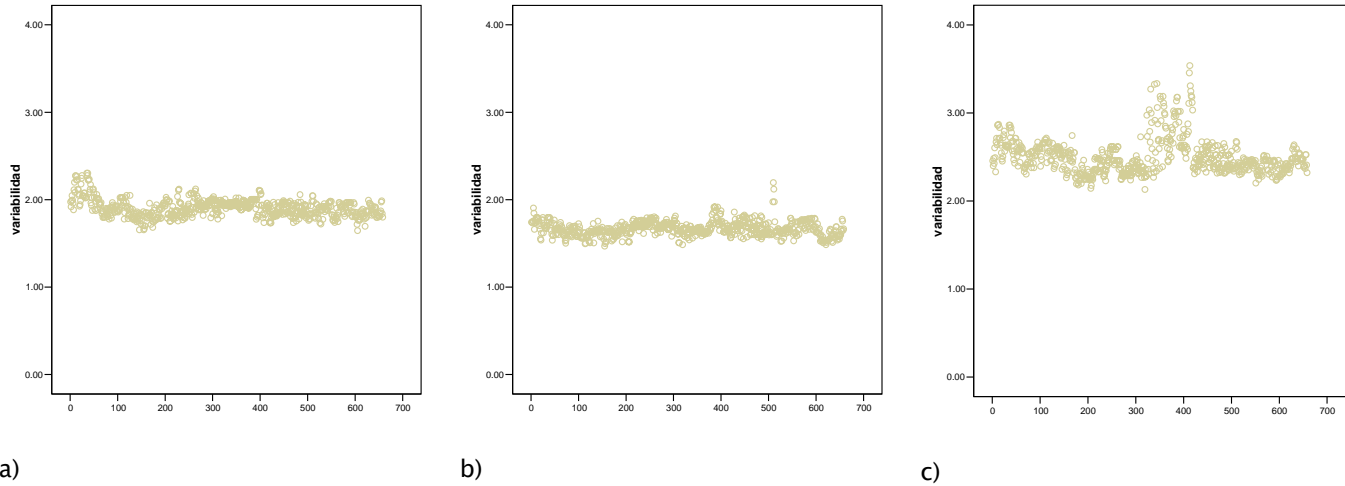


Figura 24.– Variabilidad de NO₂ en las secciones censales de Bilbo, períodos: 1996–1998 (a), 1999–2001 (b) y 2002–2003 (c). Fuente: construcción propia).

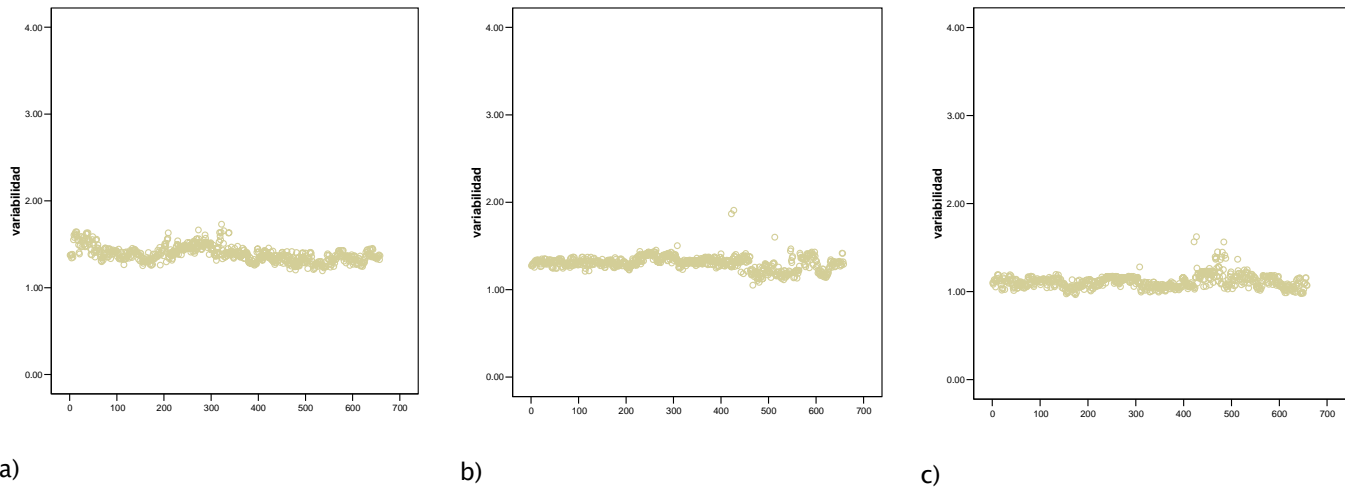


Figura 25.– Variabilidad de SO₂ en las secciones censales de Bilbo, períodos: 1996–1998 (a), 1999–2001 (b) y 2002–2003 (c). Fuente: construcción propia).

La variabilidad del contaminante dentro de las Áreas refleja, en gran parte, el comportamiento del mismo mostrado en las Figuras 9 a 11. El escaso número de estaciones medidoras de la contaminación atmosférica por ozono en la Regió Sanitària Girona implicó una importante dispersión. Los niveles de SO₂ dentro de las Áreas, dentro de las secciones, en el Área Metropolitana de Barcelona concuerdan con el comportamiento errático de este contaminante en los últimos años. En lo que se refiere al Área Metropolitana de Bilbo, la variabilidad dentro de las secciones en el caso del NO₂ es bastante mayor que la de SO₂, destacando el último período.

3.6.- Análisis descriptivo de los datos

Incidencia de las neoplasias de interés medioambiental en la Regió Sanitària Girona

	Pulmón		Vejiga urinaria		Tiroides		Linfoma No Hodgkin		Linfoma Hodgkin		Mieloma	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Observados	1998	208	1003	186	52	160	401	281	70	41	118	109
Media	9,51	0,99	4,78	0,89	0,25	0,76	1,91	1,34	0,33	0,2	0,56	0,52
Desviación típica	23,86	2,76	11,69	2,44	0,93	2,29	5,27	3,99	1,09	0,77	1,79	1,54
Porcentaje de ceros	17,14	67,14	30,48	71,43	87,14	70,00	51,43	61,43	83,33	90,00	75,71	77,62
Primer cuartil	1	0	0	0	0	0	0	0	0	0	0	0
Mediana	2	0	1	0	0	0	0	0	0	0	0	0
Tercer cuartil	7	1	4	1	0	1	1,25	1	0	0	0	0
Esperados	1951	205	980	183	51	157	396	275	69	40	115	108
Media	9,29	0,98	4,67	0,87	0,24	0,75	1,89	1,31	0,33	0,19	0,55	0,52
Desviación típica	21,74	2,49	10,92	2,22	0,6	1,99	4,53	3,37	0,85	0,53	1,28	1,31
Porcentaje de ceros (%)	14,76	79,52	39,05	82,38	94,29	83,33	67,14	74,29	92,86	95,24	88,57	90,00
Primer cuartil	1,25	0,12	0,64	0,11	0,03	0,07	0,24	0,15	0,04	0,02	0,07	0,06
Mediana	2,77	0,26	1,46	0,25	0,07	0,17	0,54	0,34	0,08	0,04	0,17	0,14
Tercer cuartil	7,74	0,78	3,99	0,71	0,2	0,59	1,56	1,04	0,26	0,15	0,46	0,41

N= 210 municipios

Tabla 8.- Incidencia observada y esperada de las neoplasias de interés medioambiental en la Regió Sanitària Girona (Fuente: Construcción propia).

Como se observa, excepto en el caso de pulmón (hombres) y, en menor medida, cáncer de vejiga urinaria (hombres), los datos presentan el problema conocido como exceso de ceros (con un rango que oscila entre el 51,43% en linfoma no Hodgkin, hombres, a 90% en linfoma Hodgkin, mujeres).

Con alguna excepción (cáncer de tiroides en hombres) las RIE (Figuras 26 a 31) muestran una gran variabilidad y, sin embargo (con la debida cautela) también se aprecia algún indicio de cierto patrón geográfico. Obsérvese, sobre todo el cáncer de tráquea, bronquios y pulmón en hombres (Figura 26). Contrariamente no parece apreciarse ninguna conducta sistemática en las RIE de linfoma no Hodgkin (ambos sexos) y en las RIE de cáncer de tiroides (en hombres).

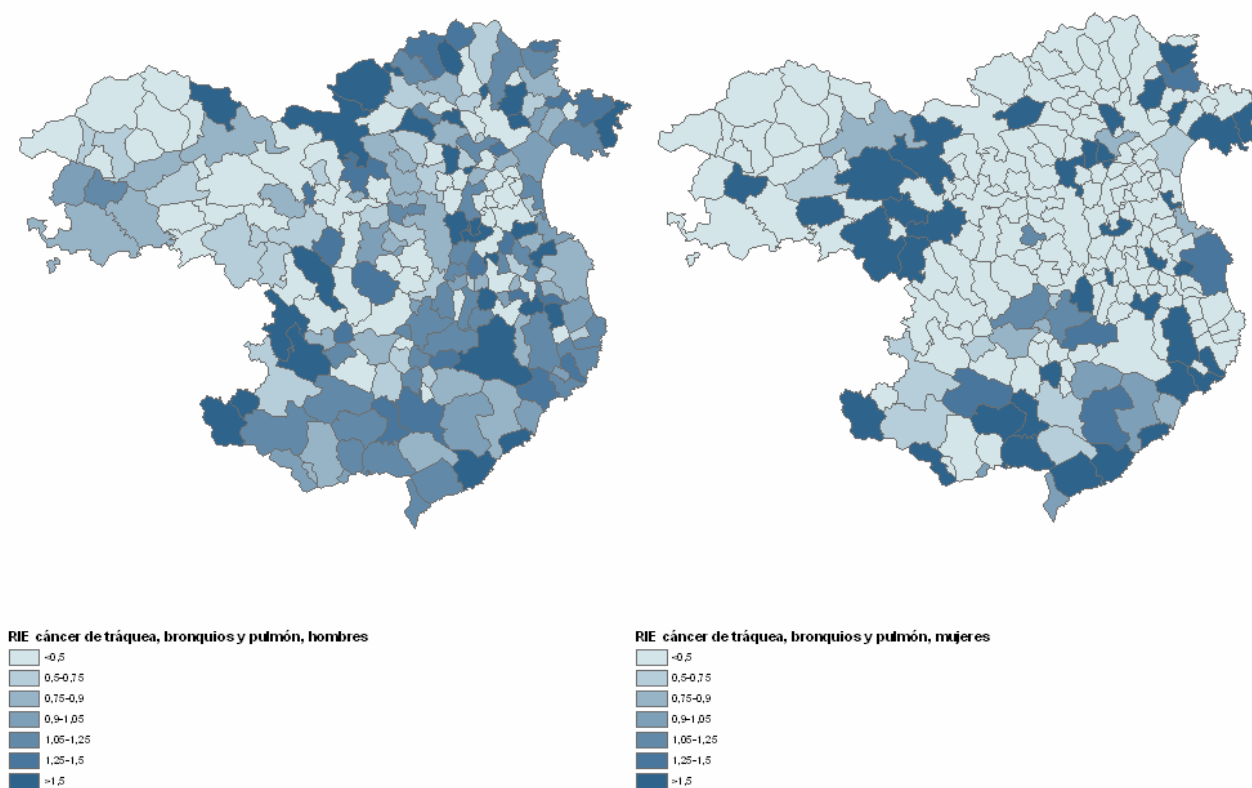


Figura 26.–Razón de Incidencia Estandarizada de cáncer de tráquea, bronquios y pulmón (hombres y mujeres) en la Regió Sanitaria Girona (Fuente: Construcción propia).

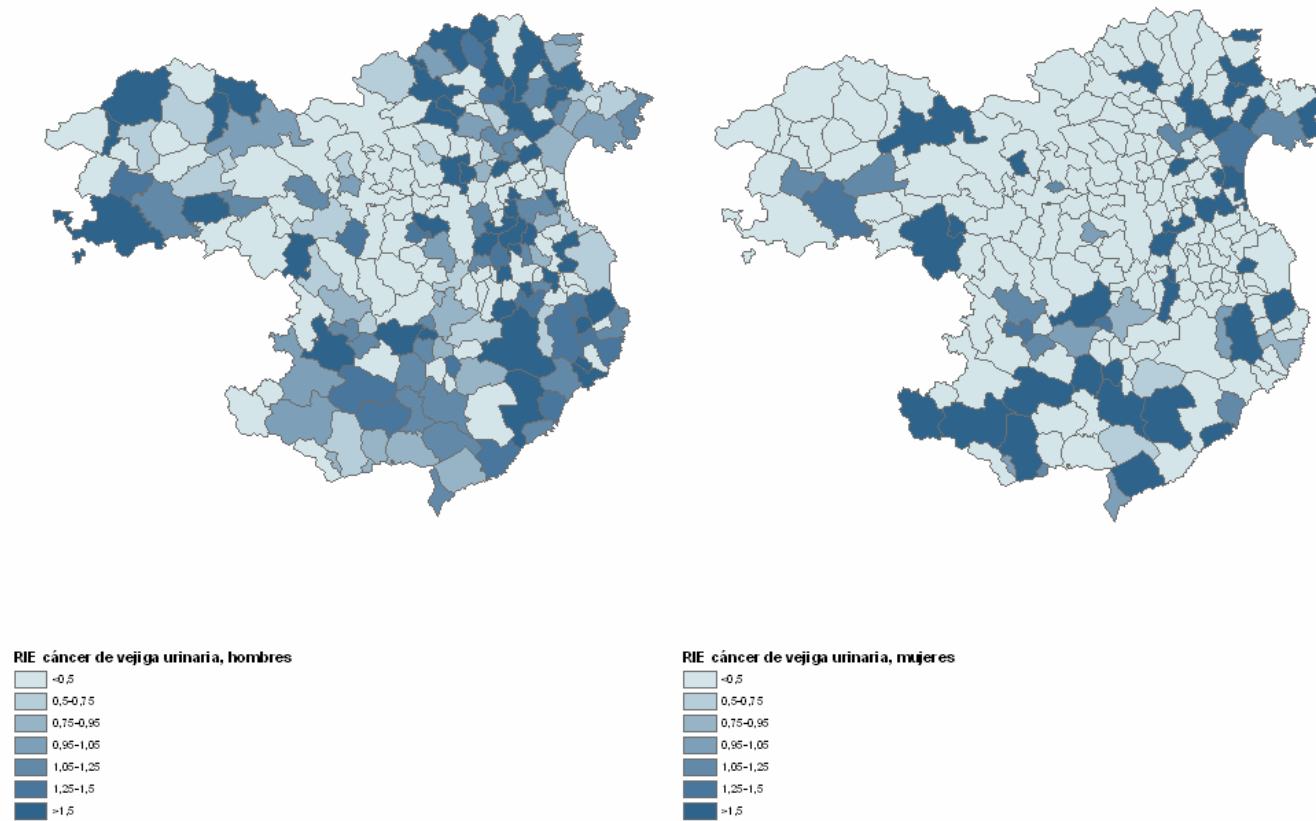


Figura 27.- Razón de Incidencia Estandarizada de cáncer de vejiga urinaria (hombres y mujeres) en la Regió Sanitària Girona (Fuente: Construcción propia).

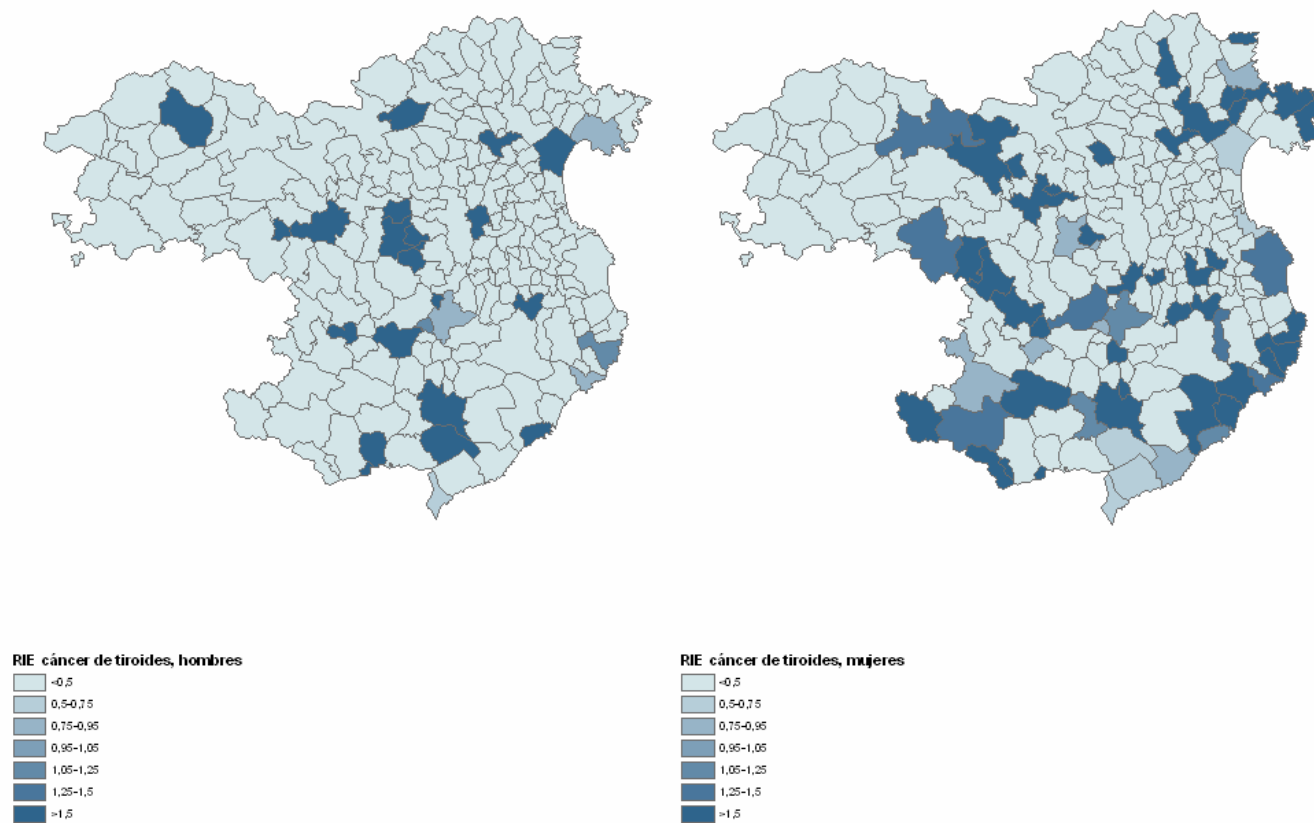


Figura 28.-Razón de Incidencia Estandarizada de cáncer de tiroides (hombres y mujeres) en la Regió Sanitària Girona (Fuente: Construcción propia).

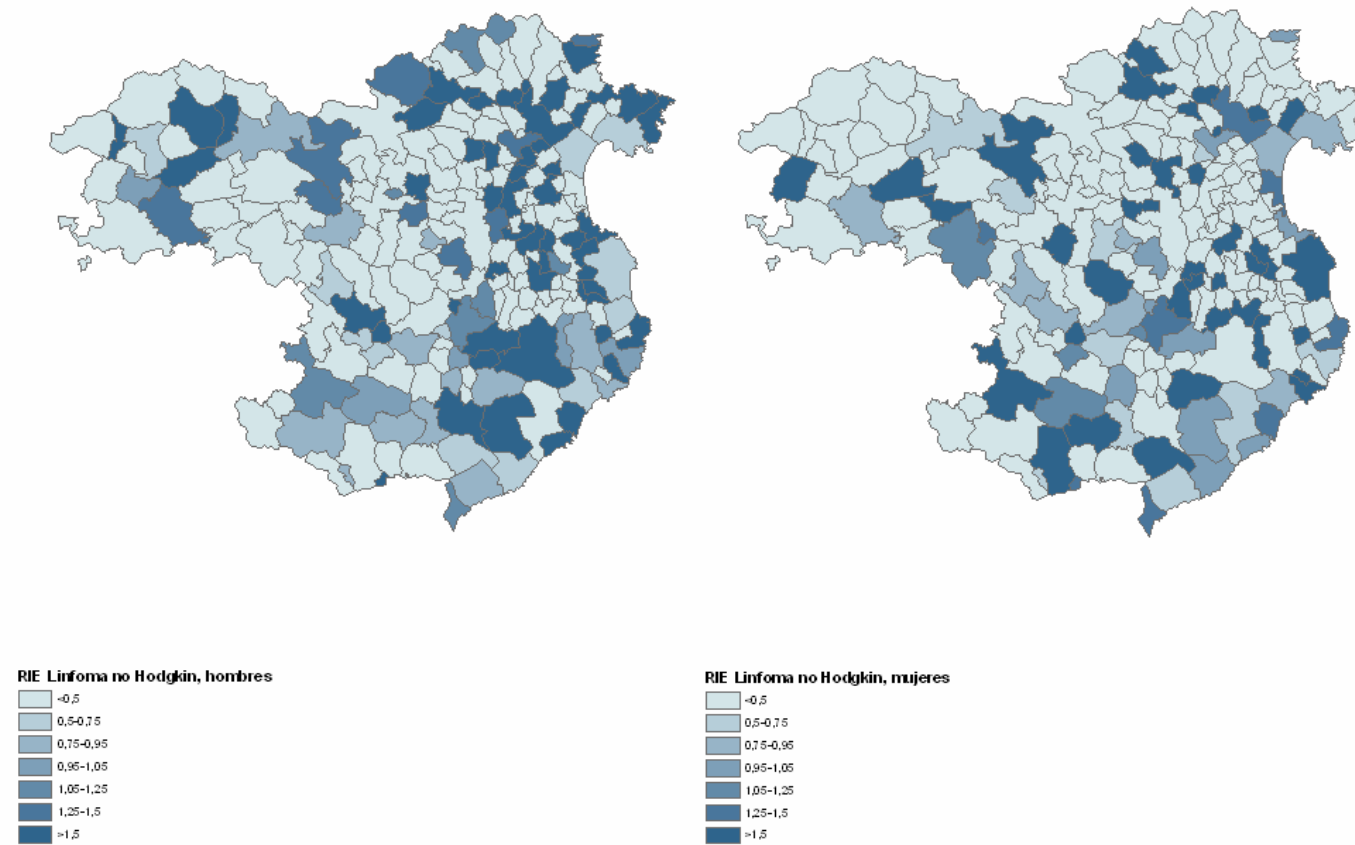


Figura 29.-Razón de Incidencia Estandarizada de Linfoma no Hodgkin (hombres y mujeres) en la Regió Sanitària Girona (Fuente: Construcción propia)

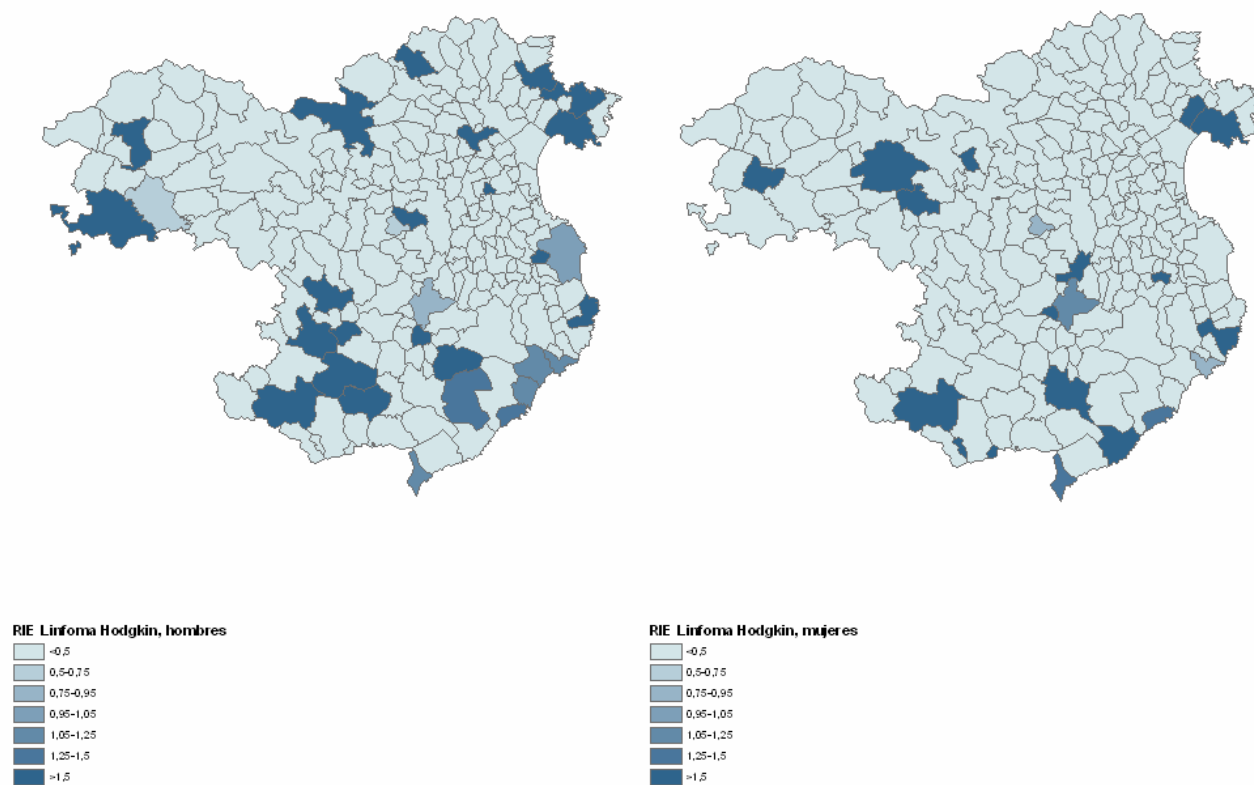


Figura 30.- Razón de Incidencia Estandarizada de Linfoma Hodgkin (hombres y mujeres) en la Regió Sanitària Girona (Fuente: Construcción propia).

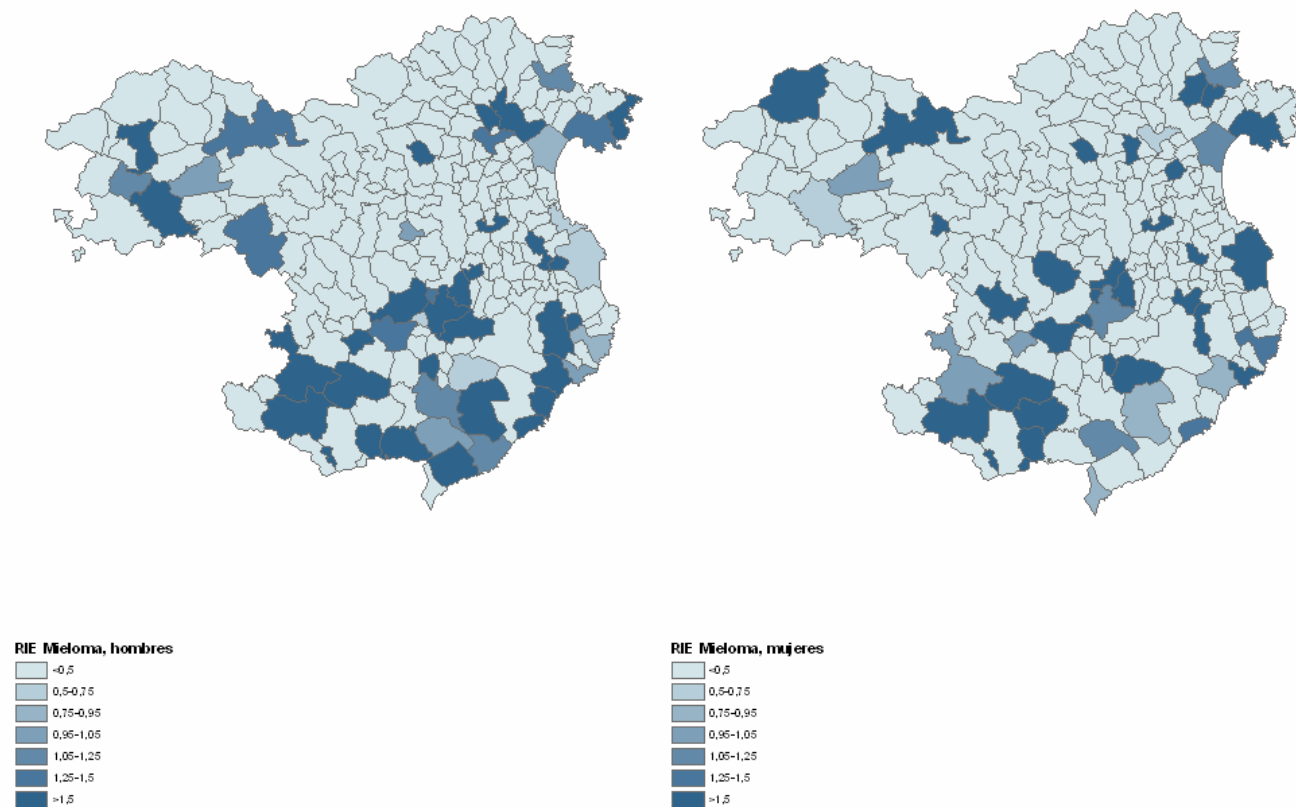


Figura 31.- Razón de Incidencia Estandarizada de mieloma (hombres y mujeres) en la Regió Sanitària Girona (Fuente: Construcción propia).

Mortalidad por las neoplasias de interés medioambiental en el Área Metropolitana de Barcelona

	Pulmón		Vejiga urinaria		Tiroides		Linfoma No Hodgkin		Linfoma Hodgkin		Mieloma	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Observados												
Media	3,33	0,43	0,91	0,22	0,01	0,02	0,16	0,16	0,02	0,02	0,08	0,09
Desv. típ.	5,58	0,87	1,32	0,52	0,12	0,15	0,44	0,45	0,13	0,12	0,29	0,33
Porcentaje de ceros	11,93	66,83	46,06	81,35	98,55	97,82	55,67	86,16	86,16	98,41	98,41	92,83
Primer cuartil	1	0	0	0	0	0	0	0	0	0	0	0
Mediana	2	0	1	0	0	0	0	0	0	0	0	0
Tercer cuartil	4	1	1	0	0	0	0	0	0	0	0	0
Esperados												
Media	4,65	0,60	0,96	0,23	0,02	0,04	0,24	0,22	0,02	0,02	0,14	0,15
Desviación típica	1,70	0,21	0,36	0,10	0,02	0,03	0,18	0,15	0,02	0,01	0,10	0,10
Primer cuartil	3,43	0,45	0,70	0,16	0,01	0,02	0,11	0,11	0,01	0,01	0,06	0,07
Mediana	4,40	0,57	0,91	0,21	0,02	0,03	0,15	0,16	0,01	0,01	0,09	0,11
Tercer cuartil	5,54	0,72	1,14	0,28	0,03	0,05	0,34	0,30	0,03	0,02	0,19	0,19

Tabla 9.- Mortalidad observada y esperada en las neoplasias de interés medioambiental en el Área Metropolitana de Barcelona (Fuente: Construcción propia).

Los datos de mortalidad, por lo que se refiere al problema de exceso de ceros, son muy similares a los de incidencia con la excepción de aquellas neoplasias que presentan alta supervivencia. Obsérvese en el caso de cáncer de tiroides, en ambos sexos, y linfoma no Hodgkin, en mujeres.

La representación de las RME (crudas) (Figuras 32–37) permite apreciar algunas neoplasias que, aparentemente, no presentan ningún patrón geográfico. En concreto, y por orden, el cáncer de tiroides (ambos sexos) (Figura 34) y el linfoma Hodgkin (ambos sexos) (Figura 36). Con bastante cautela (debido a la gran extravariabilidad no capturada) podría existir cierto patrón geográfico en el cáncer de tráquea, bronquios y pulmón (sobre todo hombres) (Figura 32) y en cáncer de vejiga urinaria (hombres) (Figura 33). El resto de causas presentan algunos indicios, por ejemplo algunas zonas de alta RME se repiten en todos los mapas.

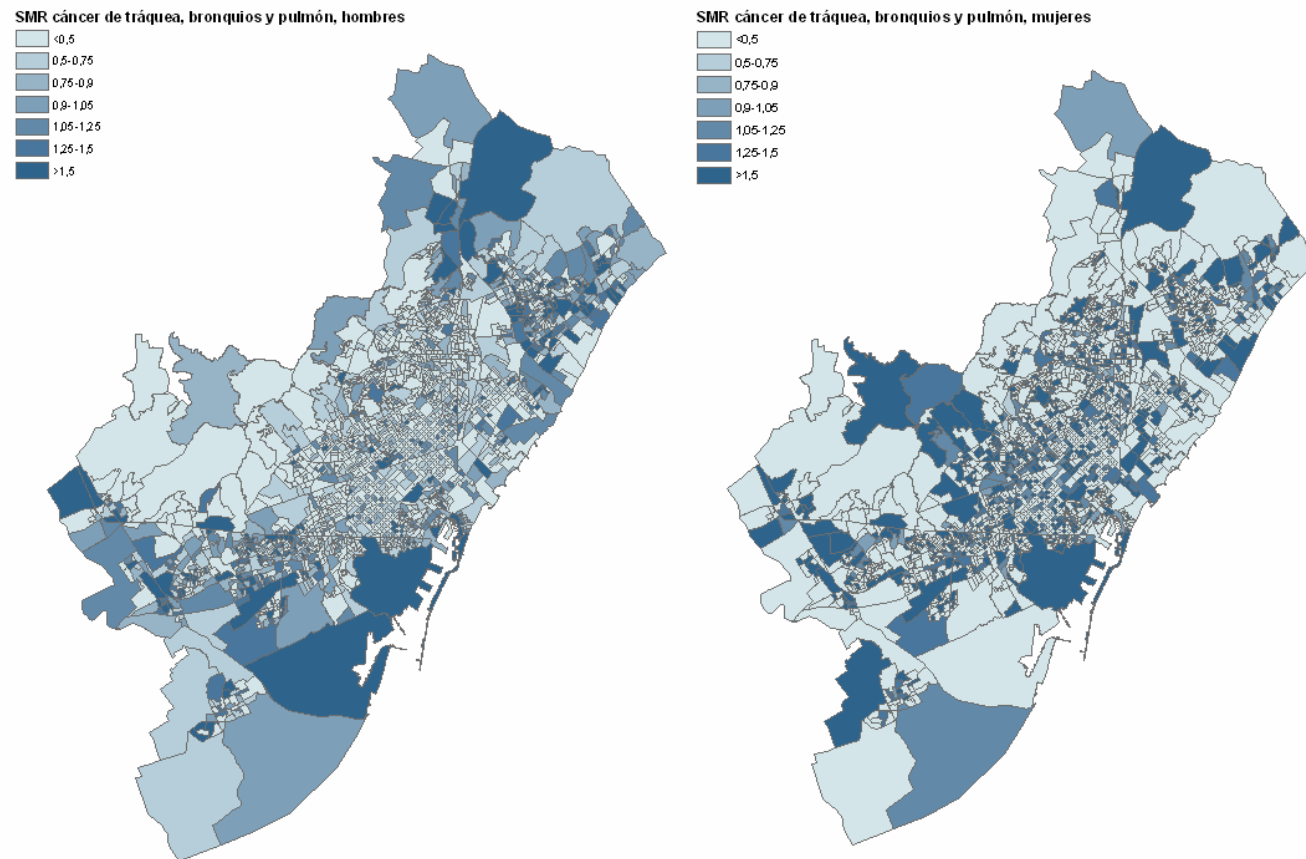


Figura 32.– Razón de mortalidad estandarizada del cáncer de tráquea, bronquios y pulmón, en hombres y en mujeres, en el Área Metropolitana de Barcelona.
Fuente: construcción propia.

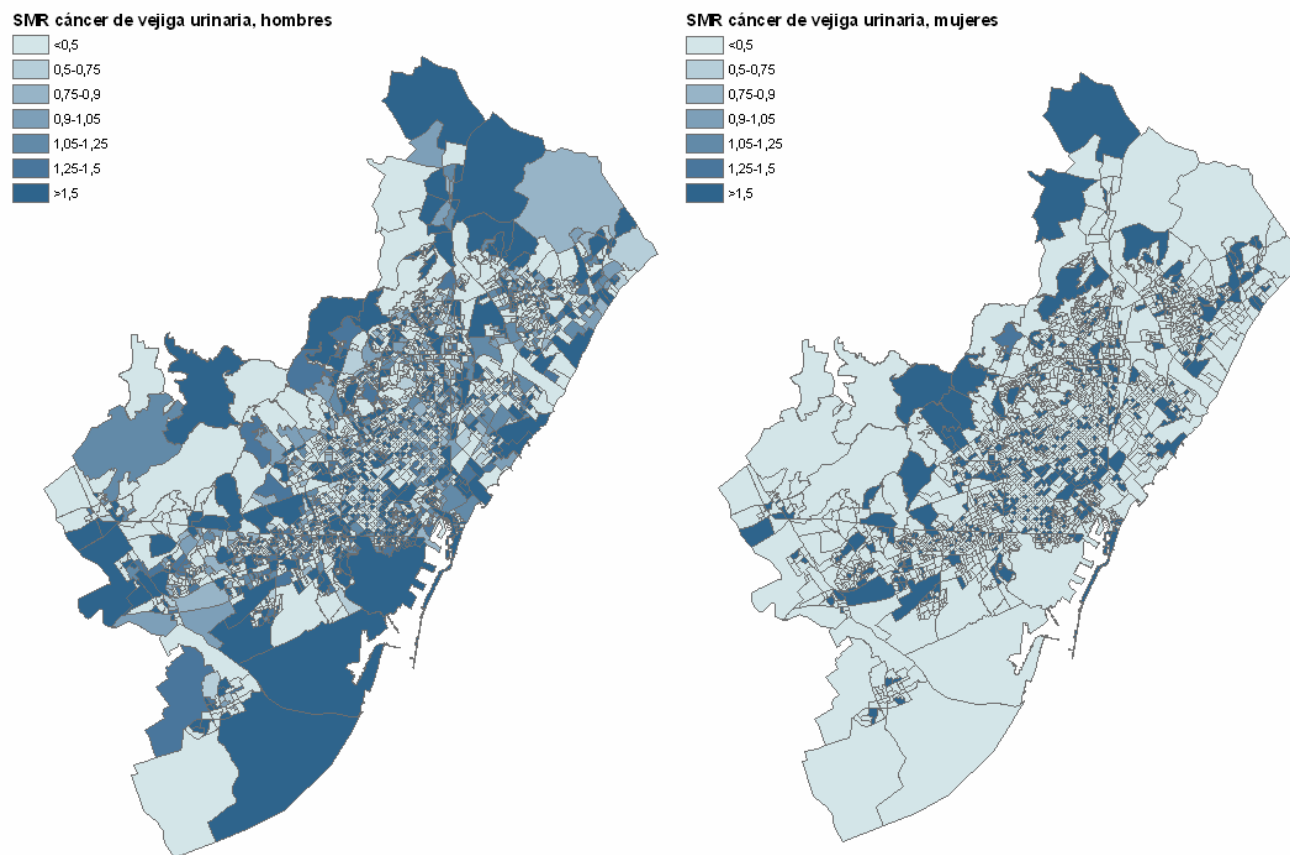


Figura 33- Razón de mortalidad estandarizada del cáncer de vejiga urinaria, en hombres y en mujeres, en el Área Metropolitana de Barcelona. Fuente: construcción propia.

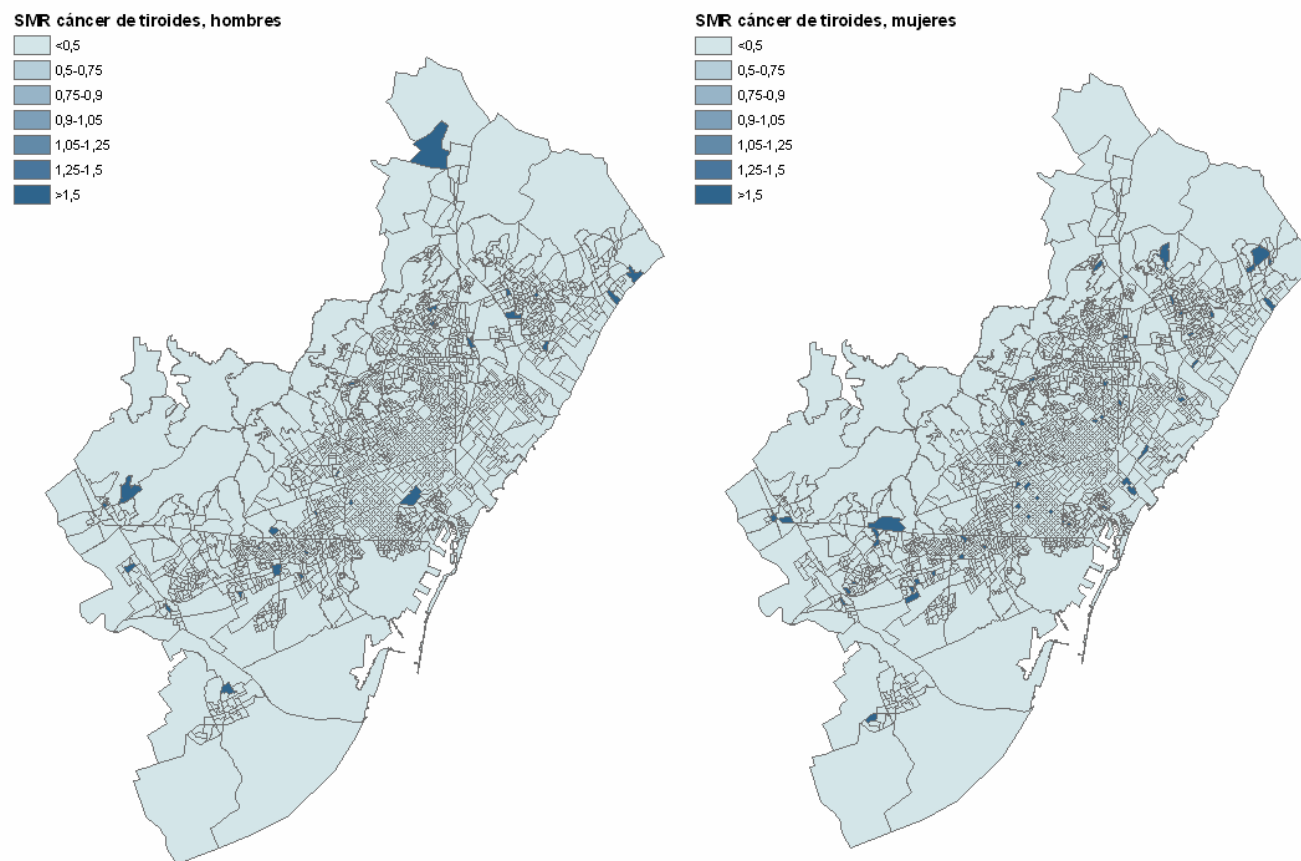


Figura 34.– Razón de mortalidad estandarizada del cáncer de tiroides, en hombres y en mujeres, en el Área Metropolitana de Barcelona. Fuente: construcción propia.

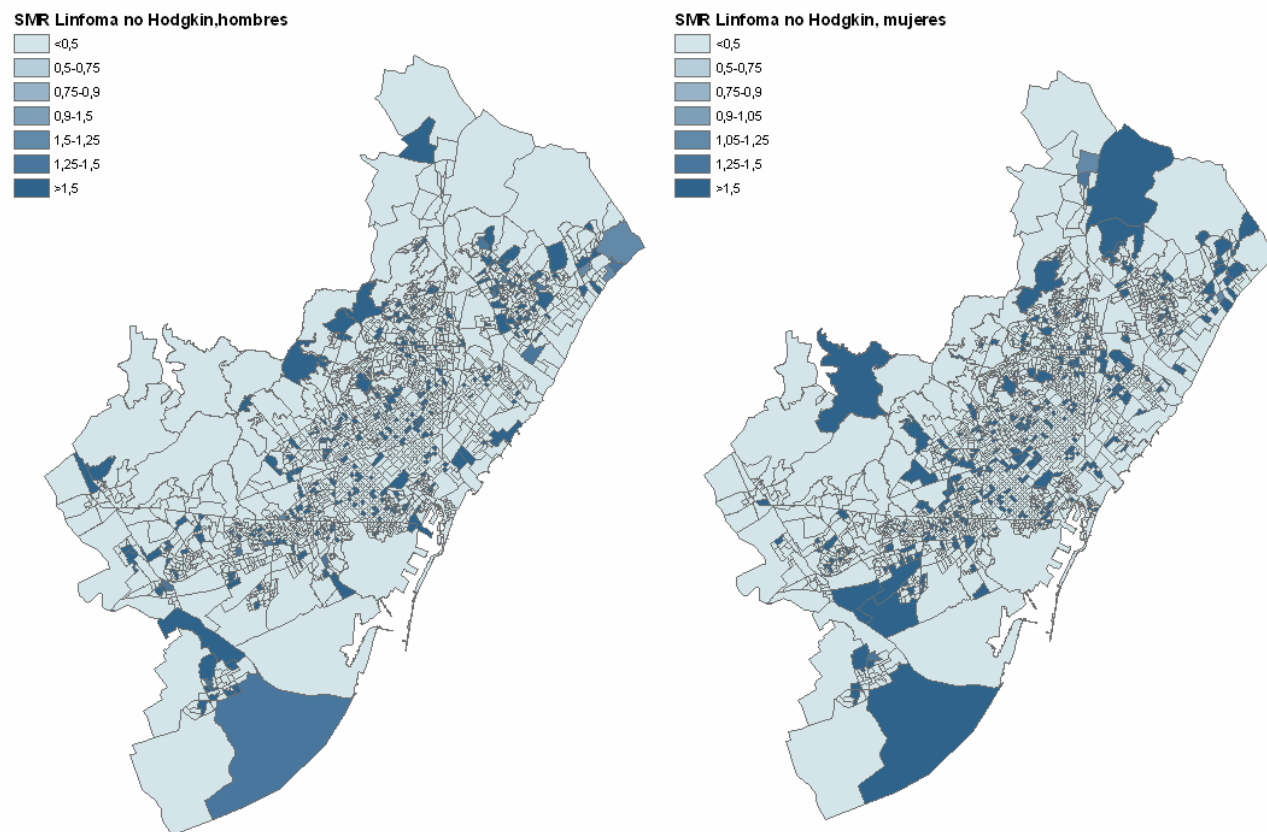


Figura 35.– Razón de mortalidad estandarizada del Linfoma no Hodgkin, en hombres y en mujeres, en el Área Metropolitana de Barcelona. Fuente: construcción propia.

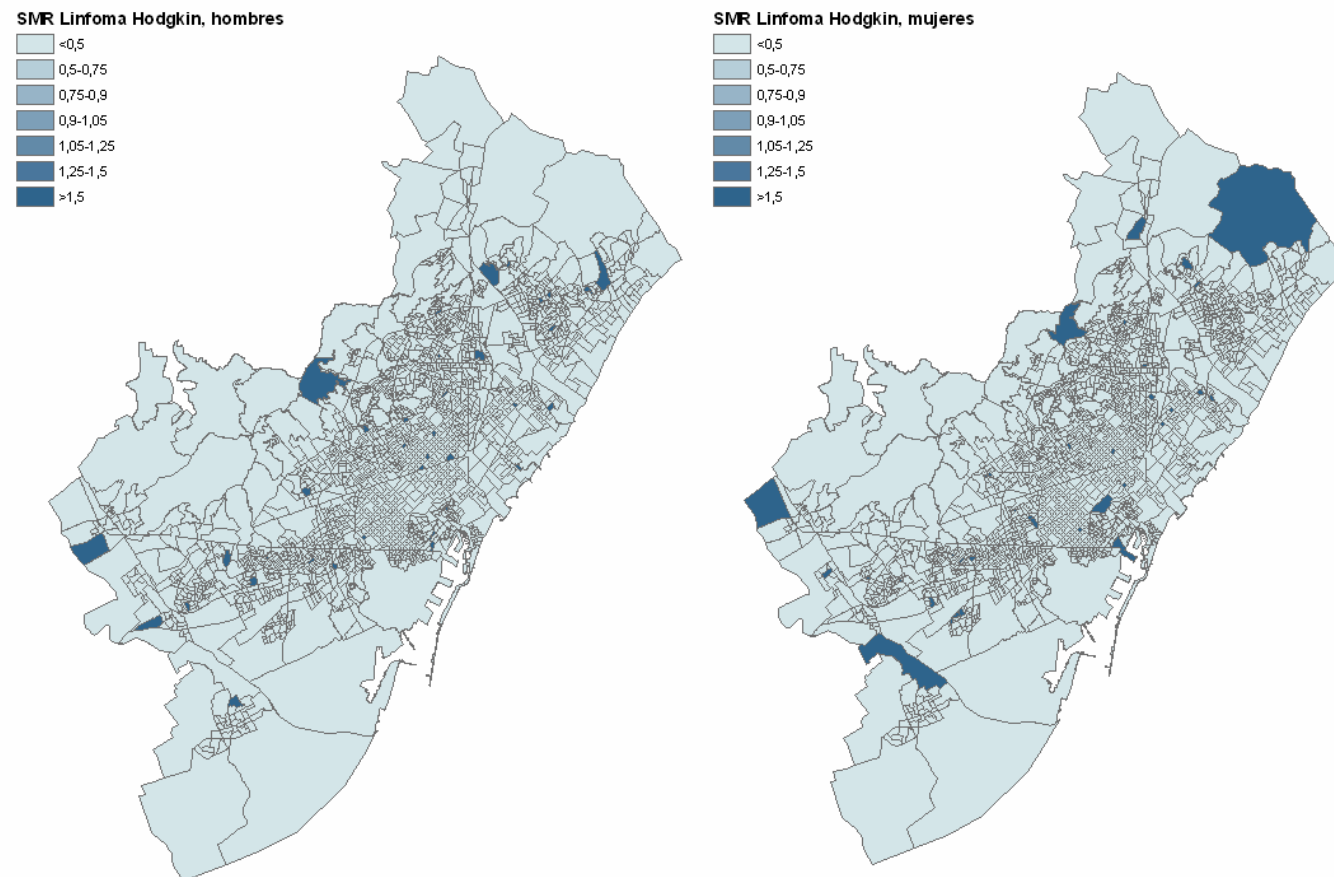


Figura 36.– Razón de mortalidad estandarizada de Linfoma Hodgkin, en hombres y en mujeres, en el Área Metropolitana de Barcelona. Fuente: construcción propia.

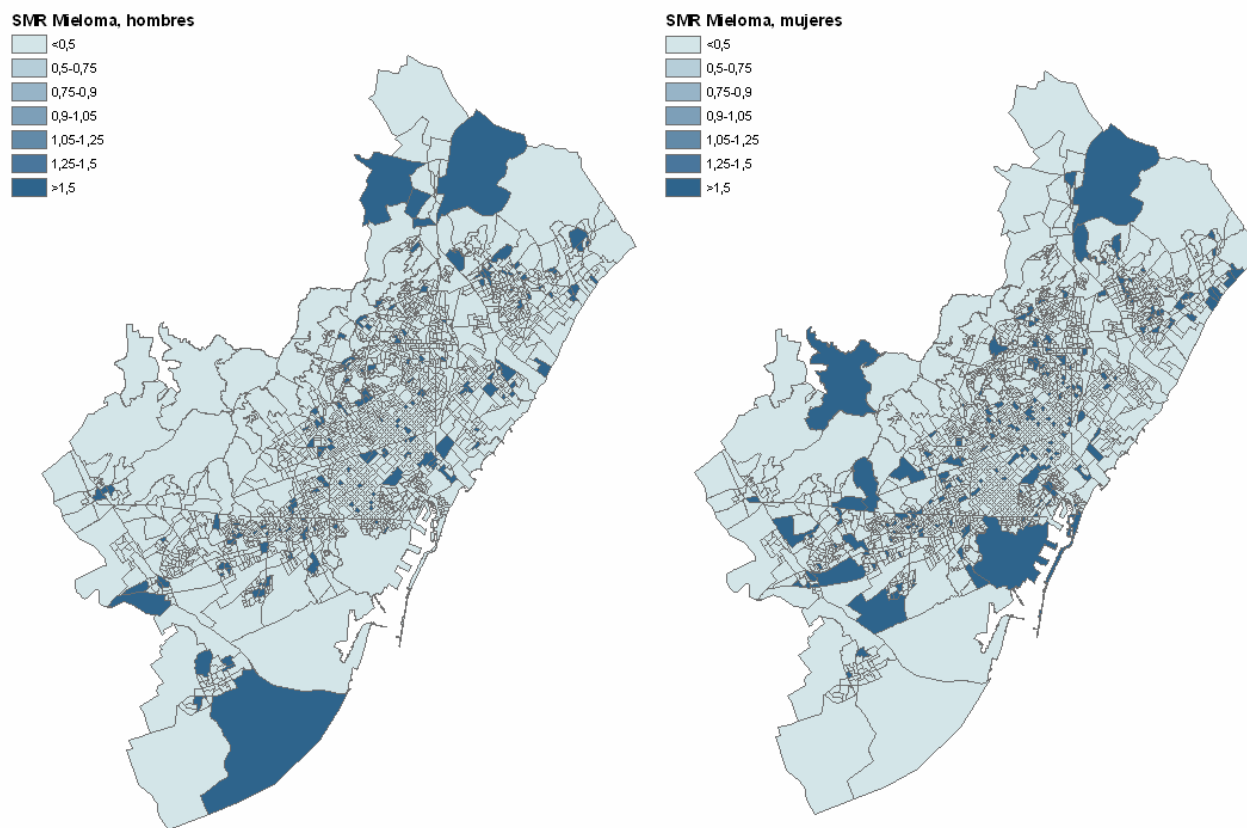


Figura 37.- Razón de mortalidad estandarizada del mieloma, en hombres y en mujeres, en el Área Metropolitana de Barcelona. Fuente: construcción propia.

Mortalidad por las neoplasias de interés medioambiental en el Área Metropolitana de Bilbo

	Pulmón		Vejiga urinaria		Tiroides		Linfoma No Hodgkin		Linfoma Hodgkin		Mieloma	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Observados												
Media	4,28	0,63	0,85	0,16	0,03	0,05	0,04	0,01	0,33	0,28	0,19	0,21
Desv. tip.	2,61	0,84	0,96	0,41	0,18	0,21	0,20	0,12	0,57	0,54	0,44	0,46
Porcentaje de ceros	4,56	54,86	45,14	85,87	97,26	95,29	96,35	98,63	72,04	76,14	83,43	81,91
Primer cuartil	2	0	0	0	0	0	0	0	0	0	0	0
Mediana	4	0	1	0	0	0	0	0	0	0	0	0
Tercer cuartil	6	1	1	0	0	0	0	0	1	0	0	0
Esperados												
Media	4,17	0,53	0,83	0,20	0,04	0,07	0,40	0,38	0,03	0,03	0,23	0,25
Desv. tip.	1,30	0,17	0,29	0,09	0,01	0,02	0,13	0,13	0,01	0,01	0,08	0,09
Primer cuartil	3,24	0,41	0,63	0,13	0,03	0,05	0,31	0,29	0,02	0,02	0,18	0,19
Mediana	3,98	0,51	0,78	0,18	0,04	0,07	0,38	0,37	0,03	0,03	0,22	0,24
Tercer cuartil	4,96	0,62	1,01	0,24	0,05	0,08	0,48	0,46	0,04	0,03	0,28	0,30

Tabla 10.- Mortalidad observada y esperada en las neoplasias de interés medioambiental en el Área Metropolitana de Bilbo (Fuente: Construcción propia).

Al igual que en las otras zonas estudiadas, Bilbo presenta también un alto porcentaje de ceros. El cáncer de tiroides y linfoma no Hodgkin son los que tienen menos casos observados.

En las Figuras 38 al 43 se muestran las RME (crudas) para el Área Metropolitana de Bilbo. Las correspondientes a cáncer de tiroides y linfoma no Hodgkin son los que presentan menor patrón sistemático. En el resto de las neoplasias se puede apreciar (nuevamente con debida cautela) algún indicio de conducta geográfica sistemática. Esta conducta parece presentar diferencias por sexo en los casos de cáncer de tráquea, bronquios y pulmón y cáncer de vejiga urinaria.

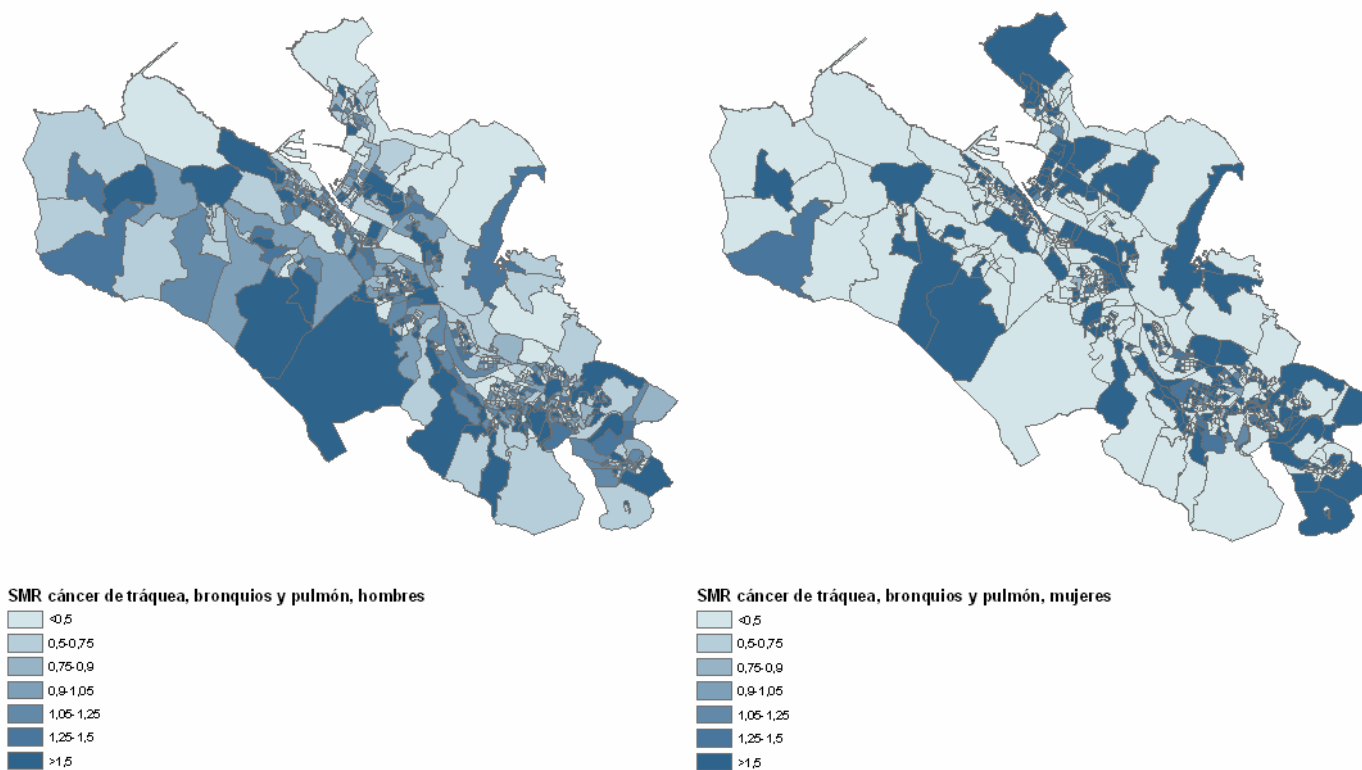


Figura 38.– Razón de mortalidad estandarizada de tráquea, bronquios y pulmón, en hombres y en mujeres, en el Área Metropolitana de Bilbo. Fuente: construcción propia

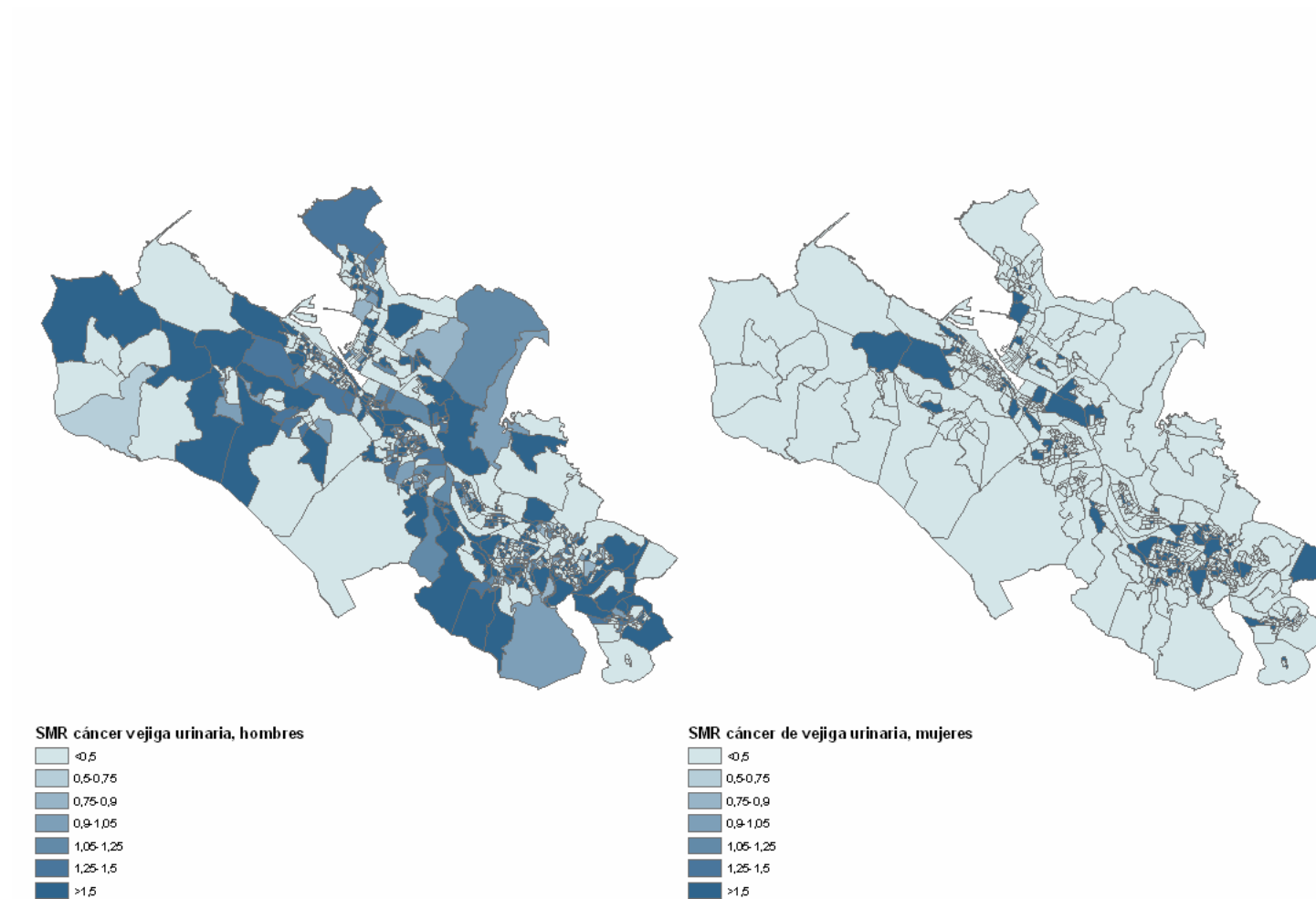


Figura 39.– Razón de mortalidad estandarizada de vejiga urinaria, en hombres y en mujeres, en el Área Metropolitana de Bilbo. Fuente: construcción propia

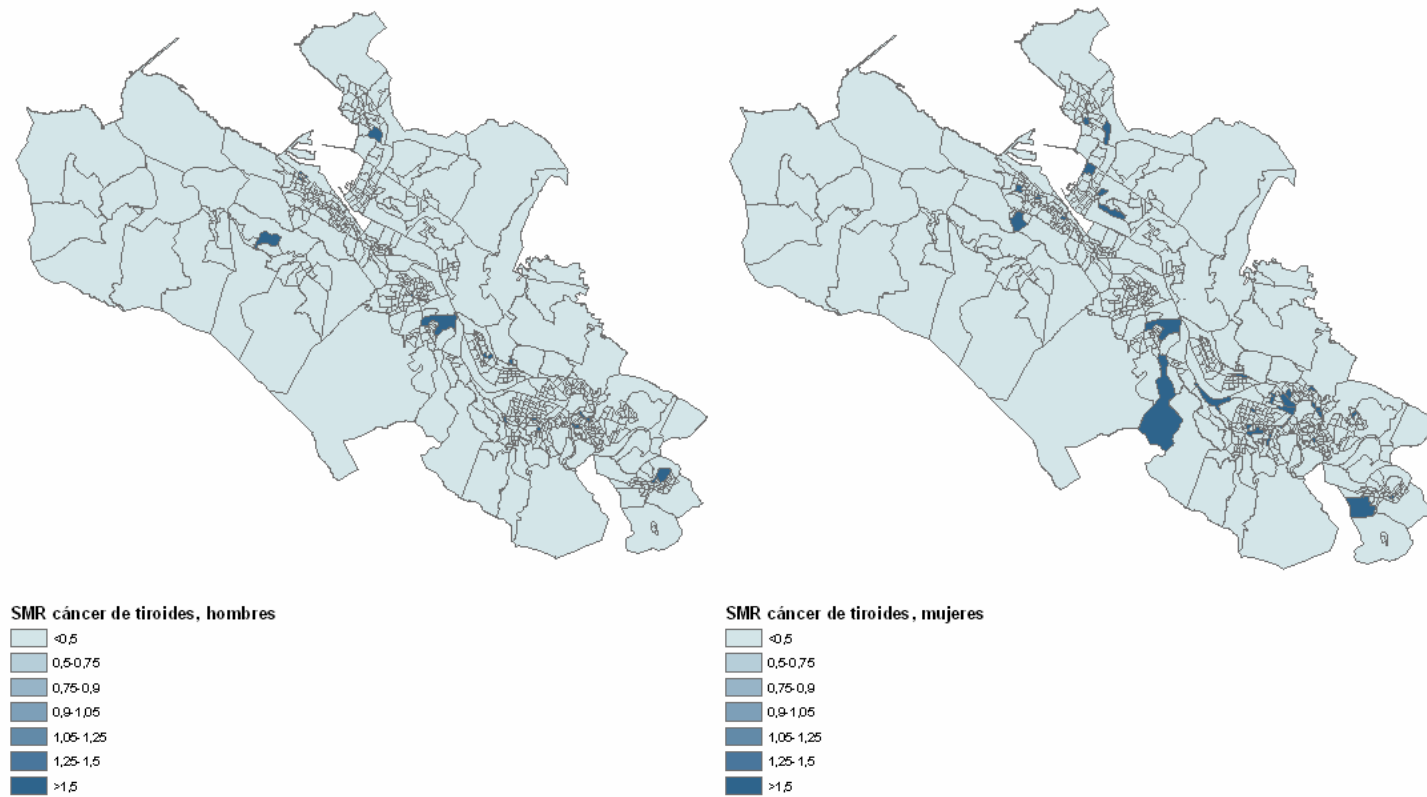


Figura 40.- Razón de mortalidad estandarizada de cáncer de tiroides, en hombres y en mujeres, en el Área Metropolitana de Bilbo. Fuente: construcción propia

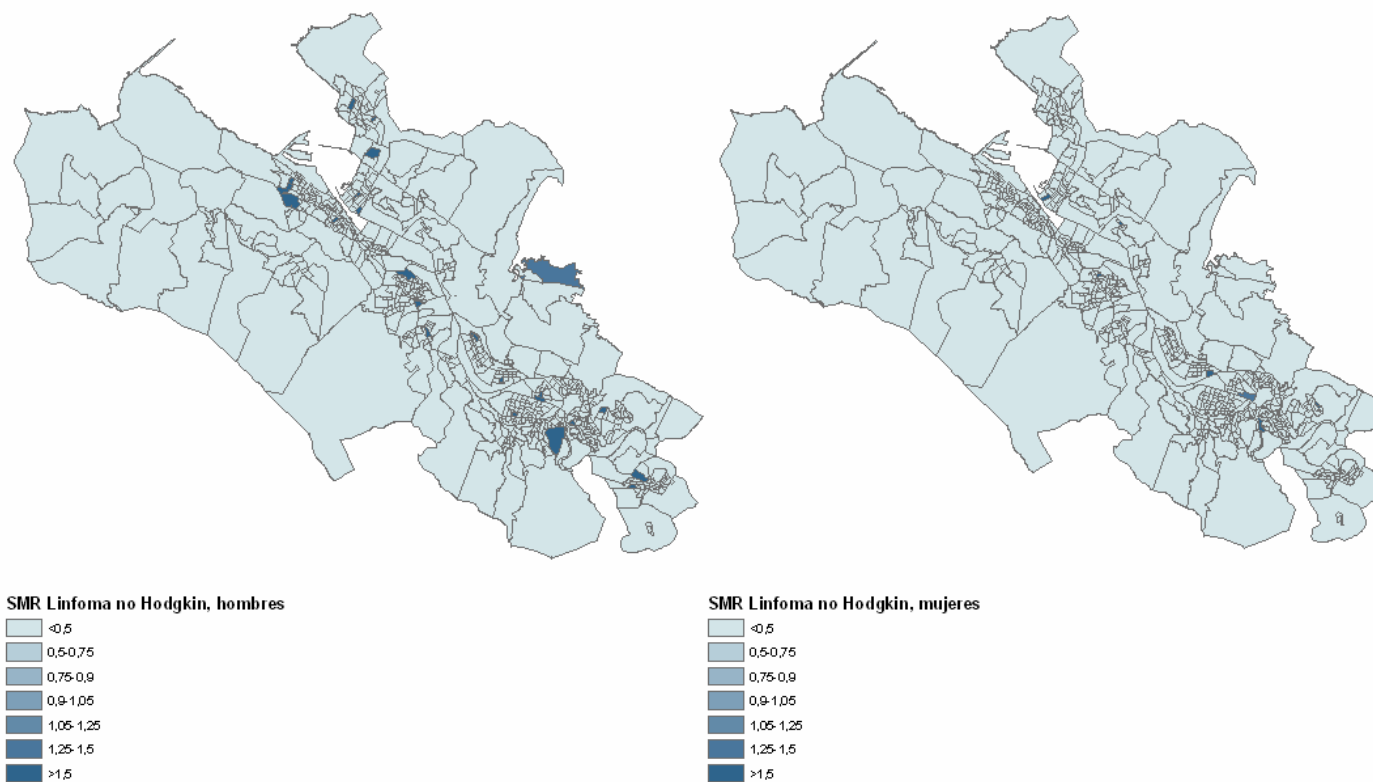


Figura 41.– Razón de mortalidad estandarizada de Linfoma no Hodgkin, en hombres y en mujeres, en el Área Metropolitana de Bilbo. Fuente: construcción propia

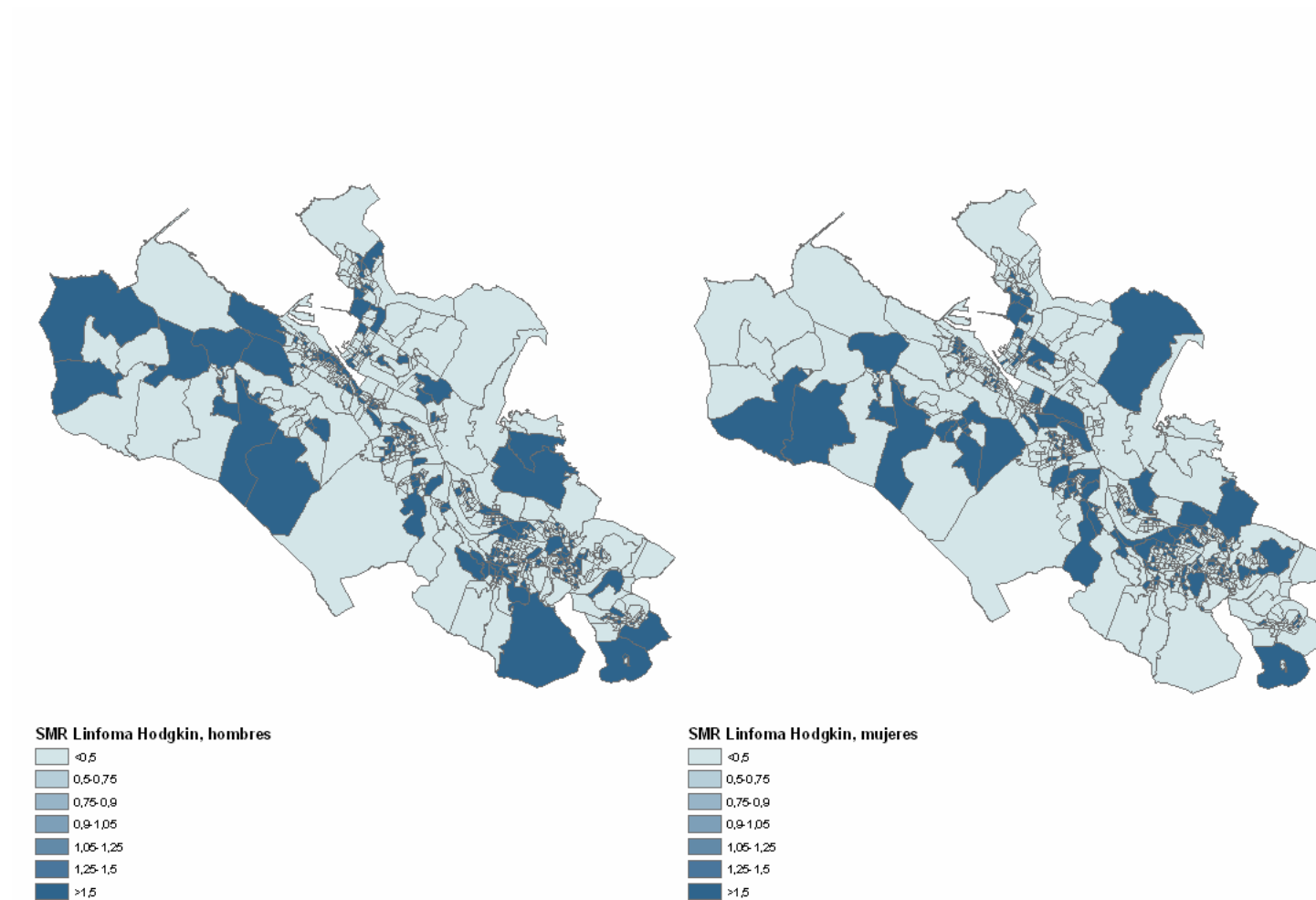


Figura 42.- Razón de mortalidad estandarizada de Linfoma Hodgkin, en hombres y en mujeres, en el Área Metropolitana de Bilbo. Fuente: construcción propia

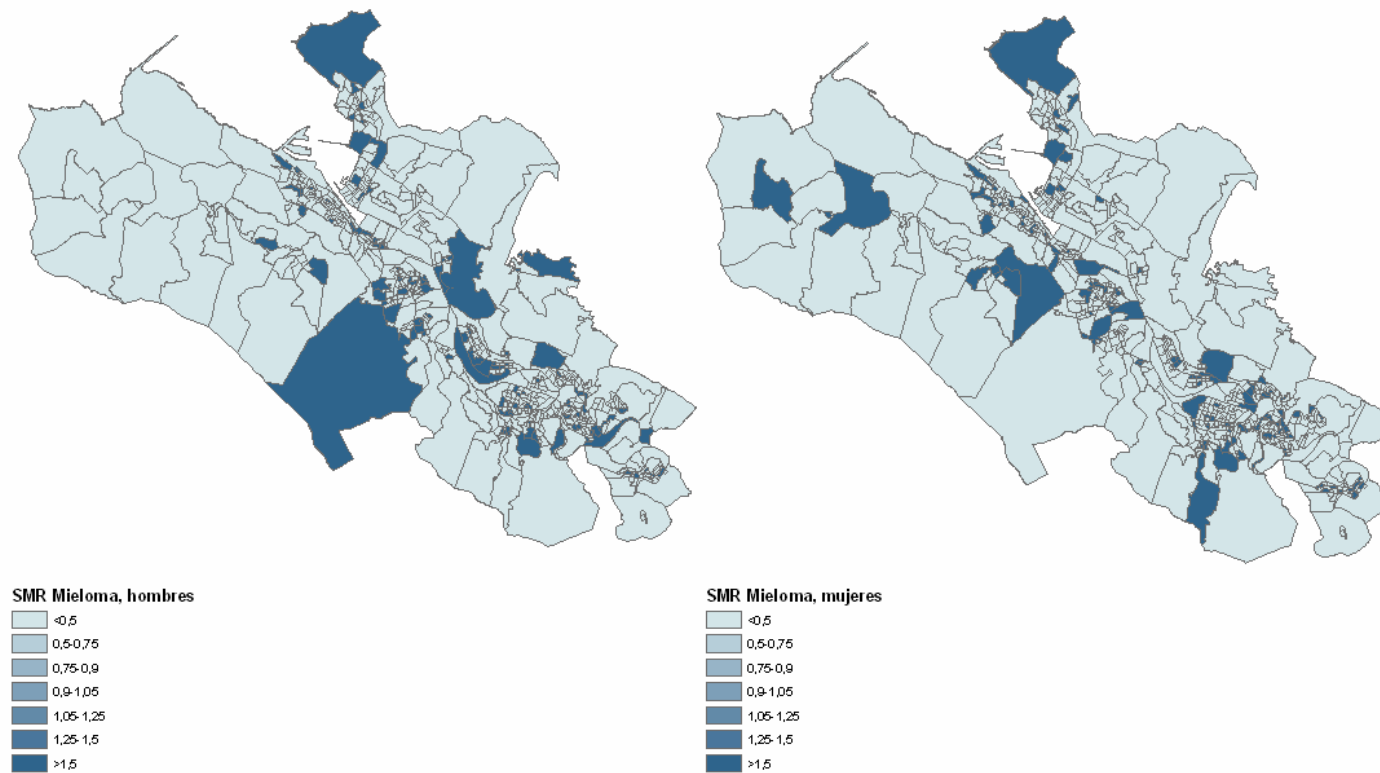


Figura 43.- Razón de mortalidad estandarizada de Mieloma, en hombres y en mujeres, en el Área Metropolitana de Bilbo. Fuente: construcción propia.

Referencias

Anderson, T.W. *An introduction to multivariate statistical analysis*. Nueva York: John Wiley & Sons, 1984.

Carstairs, V, Morris, R. *Deprivation and health in Scotland*. Aberdeen: Aberdeen University Press, 1991.

Dominguez–Berjón, MF, Borrell, C, Benach, J, Pasarín, MI. Medidas de privación material en los estudios de áreas geográficas pequeñas. *Gaceta Sanitaria* 2001; 15(Supl. 4): 23–33.

Haynes, R, Gale, S, Lovett A, & Bentham, G. Unemployment rate as an updatable health needs indicator for small areas. *Journal of Public Health and Medicine* 1996; 18: 27–32.

INE. Censo de población y viviendas, 2001. <http://www.ine.es> Accedido el 6 junio del 2006.

Little RJA, Rubin DB. *Statistical análisis with missing data*. Nueva York: John Wiley & Sons, 1987.

Novo A, Schafer JL. *The norm package*. <http://www.stat.psu.edu/~jls/misoftwa.html#aut>, 2003.

Ocaña–Riola R, Saurina C, Sanchez–Cantalejo C, Lertxundi A, Fernández–Ajuria A, Saez M, Barceló MA, Daponte A, Ruiz–Ramos M, Martínez JM, Benach J. Area deprivation and mortality in the capital cities of Andalusia and Catalonia (Spain). (Enviado al *Journal of Epidemiology and Community Health*).

Ribeiro PJ, Diggle PJ. Geo R: a package for geostatistical analysis *RNEWS*, 2001; 1(2): 15–18.

Rubin DB. Inference and missing data. *Biometrika* 1976; 63: 581–592.

Rubin DB. *Multiple imputation for nonresponse in surveys*. Nueva York: John Wiley & Son, 1987.

Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of American Statistical Association* 1996; 91:473–489.

Schafer JL. *Analysis of Incomplete Multivariate Data*. Nueva Cork: Chapman & Hall, 1996, pp. 359–367.

Townsend, P. Widening inequalities of health in Britain: A rejoinder to Rudolph Klein. *International Journal of Health Services* 1990; 20: 363–372.

4.- Metodología. Modelización

4.1.- Suavización de las Razones Estandarizadas

4.1.1.- Especificación del modelo

Modelo de Besag, York y Mollié³

-Función vínculo

$$\text{Ln}(E(O_i)) = \beta_{0i} + S_i + \text{Ln}(E_i) \quad [3]$$

-Función varianza

$$\text{Var}(O_i) = \phi E(O_i)$$

-Efecto aleatorio

$$\beta_{0i} = \beta_0 + v_i \quad S_i$$

Modelo de Hurdle³

-Función vínculo

$$\log(\mu) = X\beta$$

$$\log(-\log(1 - \pi_+)) = X\gamma$$

Siendo $\pi_+ = 1 - \pi_0$ la probabilidad de generar un valor positivo (π_0 es la probabilidad de ocurrencia de un cero).

Como se dijo este modelo se reduce a uno de Poisson estándar cuando $\beta = \gamma$.

³ La sintaxis y la especificación de los modelos se incluyen en el Anexo 5

Modelo ZINB³-Función vínculo

$$\log(\mu) = X\beta$$

$$\log\left(\frac{\omega}{1-\omega}\right) = \tau X\beta$$

en el que τ es un parámetro de escala. Esto implica que $\omega = \frac{1}{1 + \mu^{-\tau}}$.

La versión del modelo ZI estimada fue la que supone una distribución de probabilidad binomial negativa para los casos positivos. Esto equivale a introducir un efecto aleatorio no espacial en el modelo.

Modelo mixto de riesgo relativo³-Función vínculo

$$\text{Ln}(E(O_i)) = \beta_{0i} + p_i S_i + (1 - p_i) \varphi_i + \text{Ln}(E_i)$$

El modelo BYM no es más que un caso particular de éste cuando $p_i = 1 \forall i$, mientras que el modelo se convierte en un modelo de 'saltos', cuando $p_i = 0 \forall i$.

Todas las RIE/RME fueron suavizadas utilizando los BYM y el Mixto de riesgo relativo. La RIE de cáncer de tráquea, bronquios y pulmón, en hombres (17,14% de ceros) y mujeres (67,14% de ceros), fueron suavizadas utilizando también los modelos ZINB y Hurdle. Utilizando los resultados obtenidos para esta RIE, los modelos fueron comparados mediante distintos criterios de bondad del ajuste: criterio de información de Akaike (AIC), BIC y DIC.

³ La sintaxis y la especificación de los modelos se incluyen en el Anexo 5

$$AIC = Deviance + 2k$$

$$BIC = Deviance + \ln(n) * k$$

$$DIC = Deviance\ media - Deviance$$

Donde k es el número efectivo de parámetros igual a $Var(D)/2$ (Spiegelhalter et al, 1994, 2003; Gelman, 2004); y *Deviance media* es la deviance media a posteriori (Spiegelhalter et al, 2002).

4.1.2.- Estimación del modelo

Se utilizaron dos cadenas y un número de iteraciones tal que garantizase la convergencia (como mínimo 50.000). Las iteraciones de ‘*calentamiento*’ (*burning*) fueron el 10% de las iteraciones finales.

Como priors se escogieron una distribución CAR para el efecto aleatorio espacial y una distribución no informativa para el término independiente común a todos los municipios/secciones censales (coincide con (logaritmo) la RIE/RME promedio bajo estudio).

Cabe señalar, que para las desviaciones típicas de los efectos aleatorios se escogieron priors “vagamente” o “débilmente” informativas. El problema es que la elección de un prior no informativo y “apropiado” (en el sentido que su integral sea la unidad), como una distribución $\text{gamma}(\varepsilon, \varepsilon)$, con ε pequeño y positivo, hubiese implicado que la inferencia posterior dependiese del parámetro ε . Esta (excesiva) sensibilidad a ε , es particularmente problemática si no se tiene suficiente evidencia de la existencia de heterogeneidad entre unidades (Gelman, 2005). Por este motivo se optó por utilizar un prior que no presentase este problema, una distribución uniforme en un rango finito (0 a 5) de la desviación típica (Parks et al, 2004; Gelman, 2005).

La convergencia se comprobó utilizando el estadístico de Brooks–Gelman–Rubin, BGR (tanto gráfica como numéricamente) (Brooks y Gelman, 1998; Gelman et al, 2005). El

valor del estadístico BGR debe situarse en el rango 0,95–1,05, significando que las cadenas diferirán menos de 5 por ciento en valor absoluto, cuando el modelo converja.

En el modelo Hurdle se modelizó de forma ‘separada’ los ceros y las observaciones positivas (véase Capítulo 2).

Como particularidad del modelo mixto de riesgo relativo se especificó también la distribución de probabilidad para las ρ , una distribución beta.

Los parámetros que se estimaron fueron los riesgos relativos, las probabilidades a posteriori (PRP) y las desviaciones típicas de los efectos aleatorios. Las PRP se estimaron en realidad como la probabilidad que el riesgo relativo fuese mayor que uno, denotando áreas de alto riesgo y, por simetría, se estimaron también las áreas protectoras.

4.1.3.- Resultados de la estimación.

a.- Incidencia de las neoplasias de interés medioambiental

Regió Sanitària Girona

En la Tabla 1 se muestran los resultados relativos a la suavización de las RIE del cáncer de tráquea, bronquios y pulmón.

CIE-9: 162: Tráquea, bronquios y pulmón		DESVIACIONES TÍPICAS				
Hombres						
Efectos aleatorios	BYM	HURDLE	ZINB	MIXTURA	Saltos	
Espacial	0,136664(0,08516)	0,126505(0,085135)	0,157834(0,082763)	0,19243(0,11111)	0,37166(0,17555)	
No Espacial	0,643723(0,043896)	0,680798(0,045964)	0,684921(0,046568)	0,17060(0,07001)		
AIC	587,721	1337,362	622,817	855,095		
BIC	1206,167	3143,116	1295,872	1844,696		
DIC	369,943	476,624	378,006	348,375		
Nº iteraciones	50000	90000	70000	100000		
% ceros	14,76%					
Mujeres						
Espacial	0,158269(0,120440)	0,196775(0,145571)	0,165842(0,137167)	0,282335(0,199375)	0,440073(0,227233)	
No Espacial	0,491085(0,092882)	0,602057(0,104237)	0,580904(0,090655)	0,165844(0,098356)		
AIC	409,711	1776,541	413,282	268,654		
BIC	791,342	4311,296	803,753	415,941		
DIC	233,549	433,047	239,697	203,596		
Nº iteraciones	90000	50000	70000	50000		
% ceros	79,52%					

Tabla 1.- Resultados de la estimación para el cáncer de tráquea, bronquios y pulmón en los diferentes modelos. Fuente: construcción propia.

Los mejores resultados (en términos de AIC y BIC) se obtuvieron para el modelo mixto de riesgo relativo, excepto en el caso de los hombres para el que el mejor ajuste se obtuvo con el modelo Besag, York y Mollié.

Según Spiegelhalter (2006) el DIC no puede ser utilizado en modelos que contengan mixturas. Este es el motivo por el cual el DIC fue menor en el modelo Besag, York y Mollié que en el modelo mixto en el caso de los hombres, mientras que ocurrió lo contrario con el AIC y el BIC.

Los modelos ZINB y Hurdle proporcionaron en todos los casos resultados nada satisfactorios, al menos por lo que se refiere a la bondad del ajuste. Así, en el caso de pocos ceros (hombres), el modelo BYM fue muy superior, mientras que en el caso de

exceso de ceros (mujeres) el mejor modelo fue el mixto de riesgo relativo. Por estos motivos, en lo que sigue, el resto de la RIE, y todas las RME, se suavizaron mediante el modelo BYM (el mejor, en el caso de pocos ceros) y el modelo mixto de riesgo relativo (el mejor, en el caso de exceso de ceros).

CIE-9: 188, Vejiga urinaria		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	MIXTURA	Saltos	
Espacial	0,12703(0,08930)	0,21111(0,15032)	0,28243(0,19672)	
No Espacial	0,43111(0,04837)	0,20288(0,05386)		
AIC	507,258	447,958		
BIC	982,600	815,926		
DIC	312,709	281,963		
% ceros	39,05%			
Mujeres				
Espacial	0,18127(0,13570)	0,31121(0,23299)	0,35823(0,24622)	
No Espacial	0,27847(0,14060)	0,16694(0,11317)		
AIC	379,102	267,708		
BIC	684,748	406,162		
DIC	224,769	206,066		
% ceros	82,38%			
CIE-9: 193, Tiroides		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	MIXTURA	Saltos	
Espacial	0,96549(0,64585)	1,40421(0,82559)	0,89984(0,50115)	
No Espacial	0,79442(0,34104)	0,45343(0,27889)		
AIC	251,955	214,988		
BIC	517,589	414,969		
DIC	128,952	125,905		
% ceros	94,29%			
Mujeres				
Espacial	0,29173(0,19827)	0,58176(0,31628)	0,52197(0,52197)	
No Espacial	0,83395(0,10210)	0,15926(0,11644)		
AIC	382,040	271,945		
BIC	752,777	439,019		
DIC	232,724	199,473		
% ceros	83,33%			
CIE-9: 210, Linfoma no Hodgkin		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	MIXTURA	Saltos	
Espacial	0,14036(0,10959)	0,26396(0,18797)	0,38710(0,21807)	
No Espacial	0,48507(0,08062)	0,19167(0,11785)		
AIC	591,979	522,357		
BIC	1.190,215	976,123		
DIC	309,601	293,324		
% ceros	67,14%			
Mujeres				
Espacial	0,21127(0,14155)	0,35061(0,22138)	0,46190(0,23559)	
No Espacial	0,63915(0,09523)	0,23536(0,12047)		
AIC	438,504	393,779		
BIC	855,020	691,027		
DIC	254,048	248,981		
% ceros	74,29%			

CIE-9: 211: Linfoma Hodgkin		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	MIXTURA	Saltos	
Espacial	0,26839(0,23158)	0,57385(0,44958)	0,59466(0,35056)	
No Espacial	0,82810(0,21438)	0,28186(0,17649)		
AIC	395,805	205,213		
BIC	851,651	335,919		
DIC	170,156	147,772		
% ceros	92,86%			
Mujeres				
Espacial	0,68890(0,46614)	0,94422(0,64122)	0,80456(0,39454)	
No Espacial	0,59758(0,33746)	0,43433(0,26611)		
AIC	250,010	195,594		
BIC	518,369	373,934		
DIC	117,888	112,970		
% ceros	95,24%			
CIE-9: 212: Mieloma		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	MIXTURA	Saltos	
Espacial	0,28203(0,19560)	0,45857(0,29366)	0,61428(0,25958)	
No Espacial	0,40598(0,14234)	0,12918(0,09491)		
AIC	304,182	171,425		
BIC	555,415	248,938		
DIC	184,896	141,002		
% ceros	88,57%			
Mujeres				
Espacial	0,29213(0,19459)	0,56587(0,34635)	0,57789(0,31262)	
No Espacial	0,39497(0,17613)	0,20610(0,14924)		
AIC	435,898	250,342		
BIC	785,693	411,151		
DIC	260,148	177,432		
% ceros	90,00%			

Tabla 2.– Resultados de la estimación para el resto de neoplasias de interés medioambiental. Fuente: construcción propia.

Como se ve en todos los casos el modelo que proporcionó el mejor ajuste (incluso en términos de DIC) fue el modelo mixto de riesgo relativo.

De hecho, parece existir una relación entre el porcentaje de ceros y la ‘bondad’ del modelo. Obsérvese en la Figura 1 como hasta el 40–50% de ceros, ambos modelos proporcionan un ajuste bastante similar. Incluso con un porcentaje menor del 30% el modelo BYM es mucho mejor. Sin embargo, a partir de 40–50% el modelo mejor parece ser el Mixto.

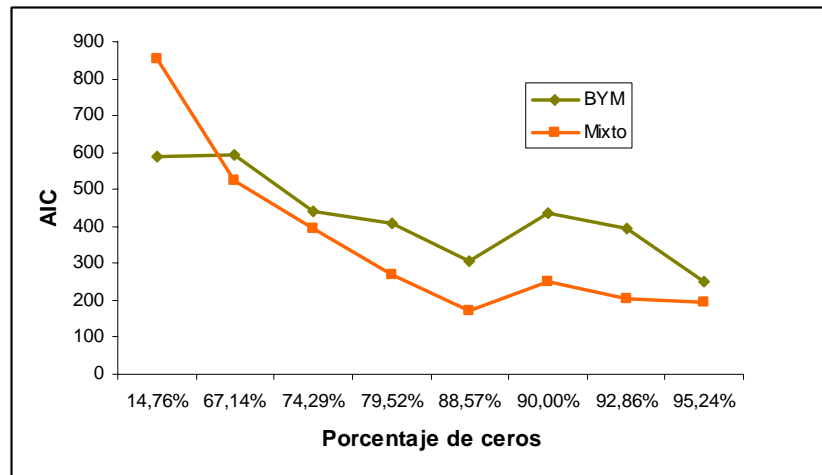


Figura 1.– Relación entre porcentaje de ceros y el estadístico AIC. Fuente: construcción propia.

Observando los mejores modelos (BYM en cáncer de tráquea, bronquios y pulmón, hombres, Tabla 1, y mixto en el resto de los casos, Tablas 1 y 2), aunque las desviaciones típicas de los efectos aleatorios no presentaron un comportamiento sistemático, la desviación de los efectos aleatorios espaciales fue mucho mayor que las otras desviaciones típicas. Nótese que, excepto en dos casos, cáncer de tráquea, bronquios y pulmón, en mujeres, y tiroides, en hombres, las desviaciones típicas del salto, es decir de la discontinuidad, fueron bastante similares a las del otro efecto espacial.

Las Figuras que se presentan a continuación muestran en la parte izquierda las RIE suavizadas (o ajustadas) y la probabilidad a posteriori que éstas sean mayor que la unidad. En rojo se representan las áreas con probabilidad mayor que 0,8, denotando un exceso de riesgo, y en azul menor que 0,2, denotando un área protectora (Richardson et al, 2004). En la parte derecha se muestran, arriba, la función de densidad de la probabilidad a posteriori, con los mismos colores y significado, y abajo el histograma de las RIE crudas y suavizadas (o ajustadas).

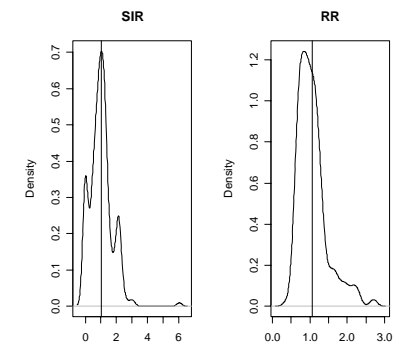
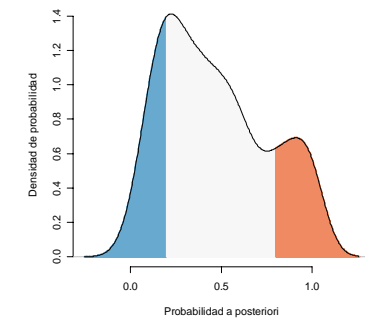
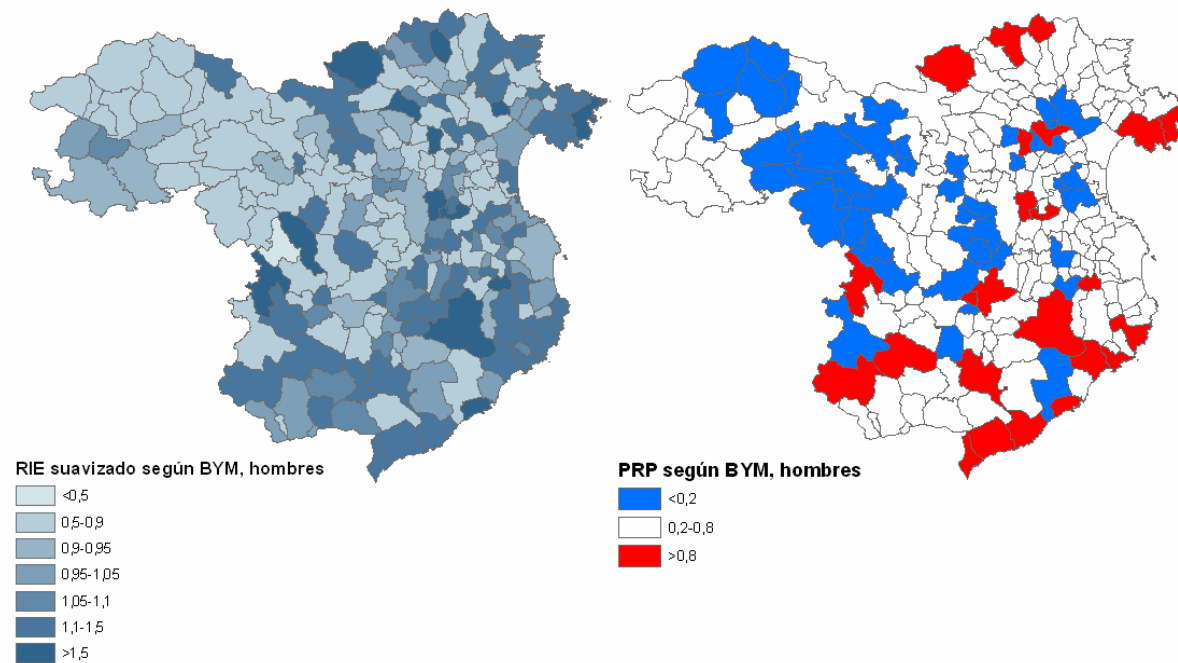


Figura 2.- Cáncer de tráquea, bronquios y pulmón, en hombres. Modelo BYM. Fuente: construcción propia.

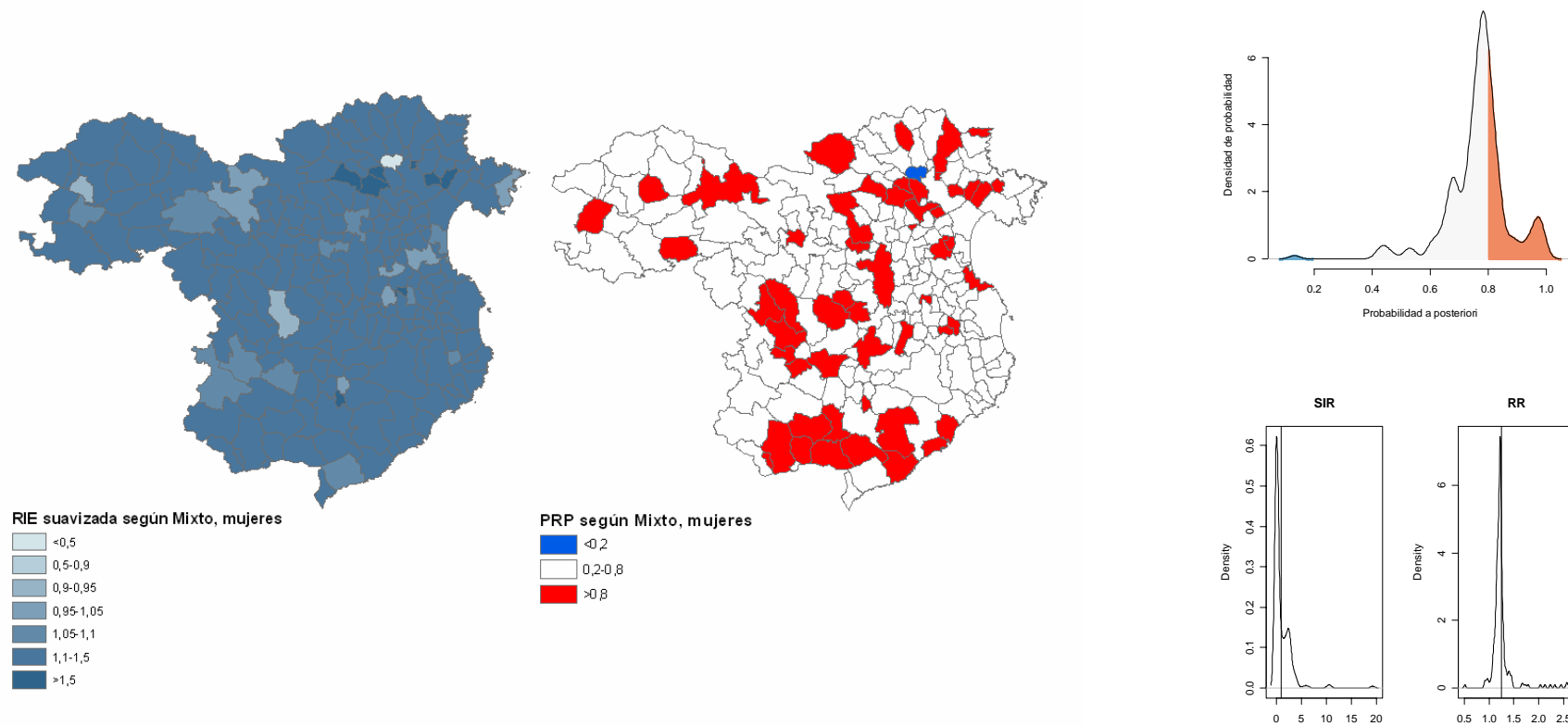


Figura 3.- Cáncer de tráquea, bronquios y pulmón, en mujeres. Modelo Mixto. Fuente: construcción propia.

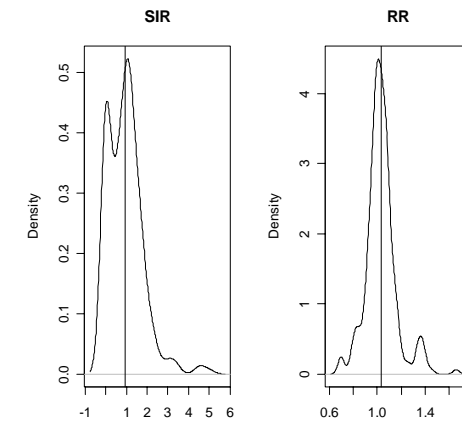
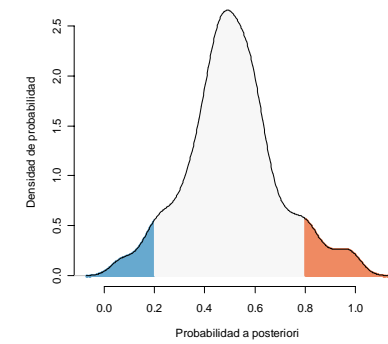
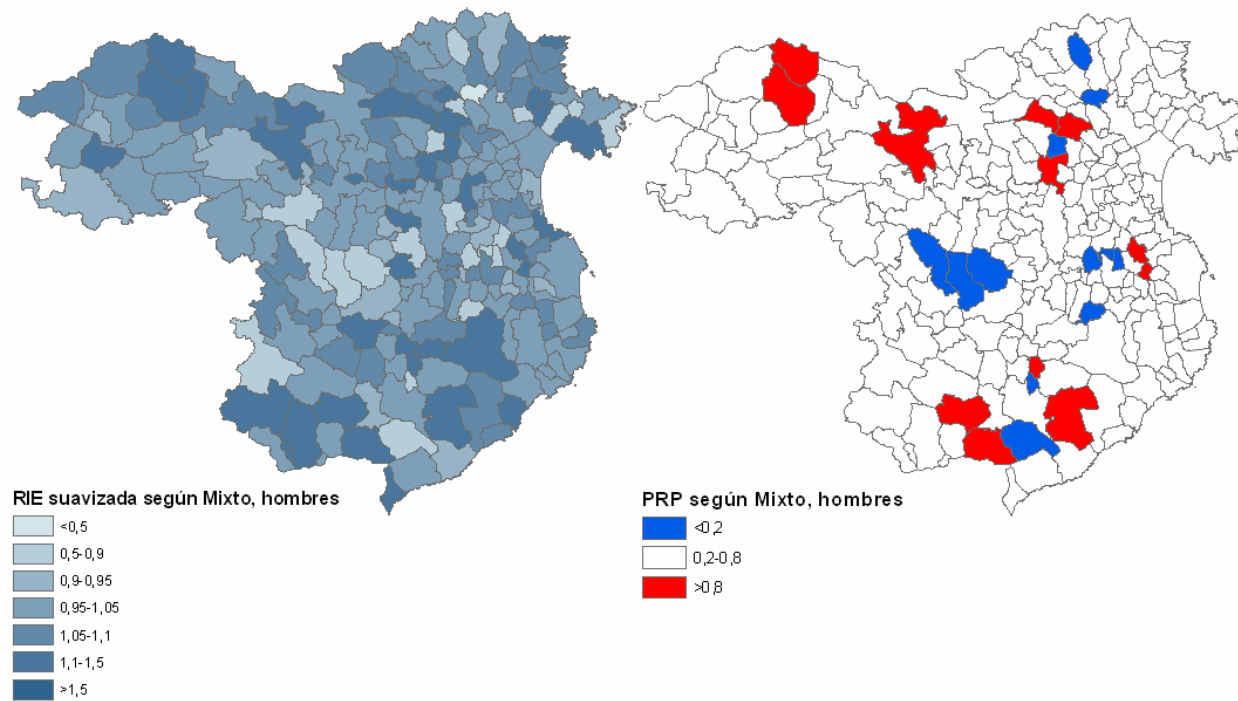


Figura 4.- Cáncer de vejiga urinaria, en hombres. Modelo Mixto. Fuente: construcción propia.

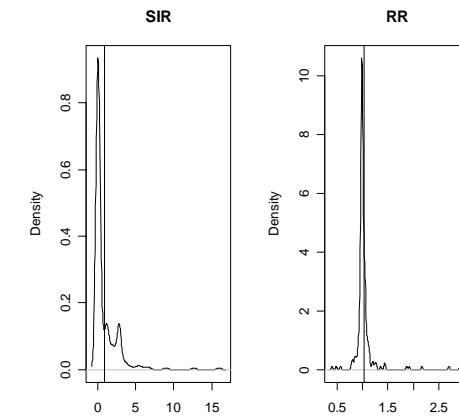
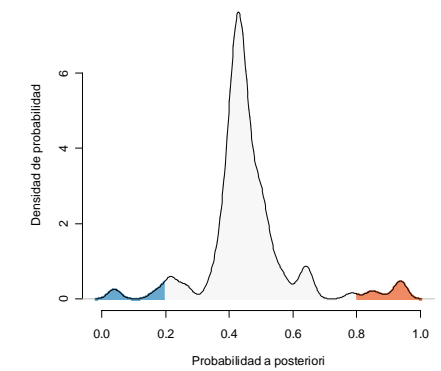
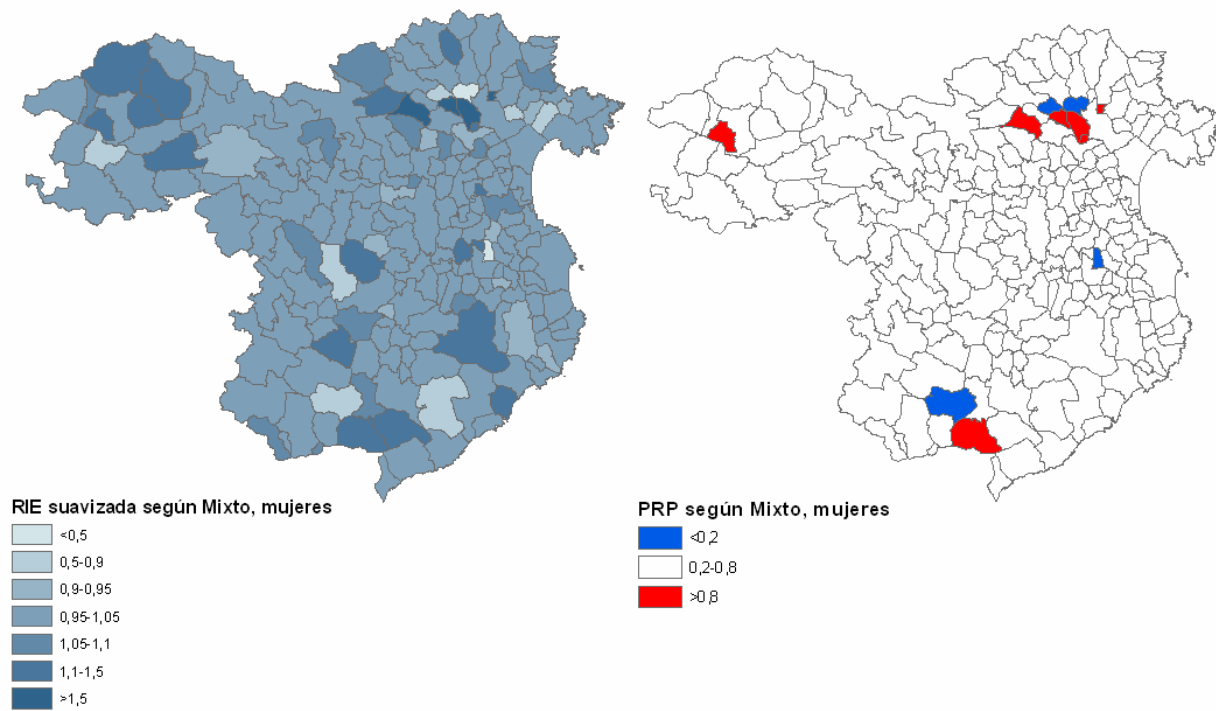


Figura 5.- Cáncer de vejiga urinaria en mujeres. Modelo Mixto. Fuente: construcción propia.

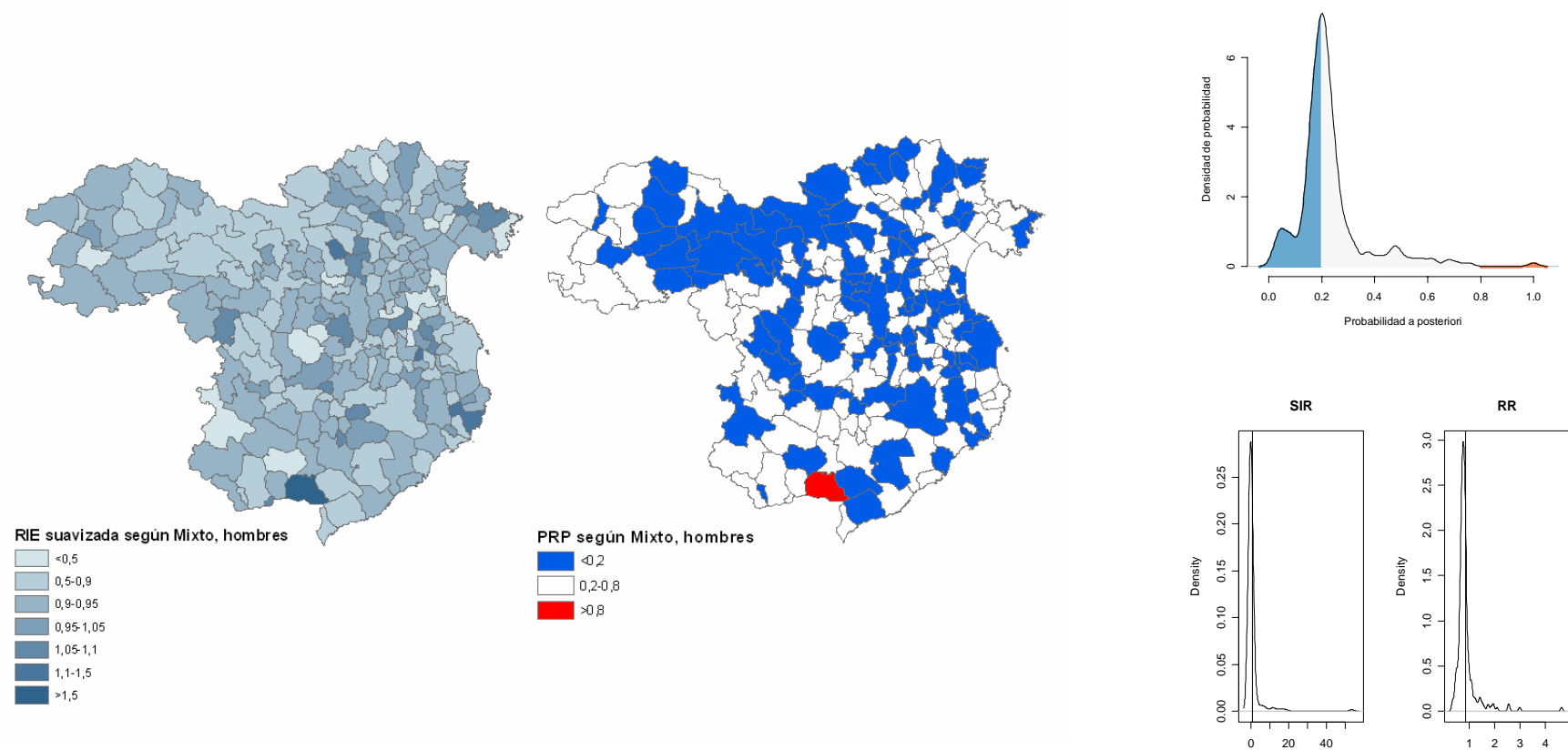


Figura 6.-Cáncer de tiroides, en hombres. Modelo Mixto. Fuente: construcción propia.

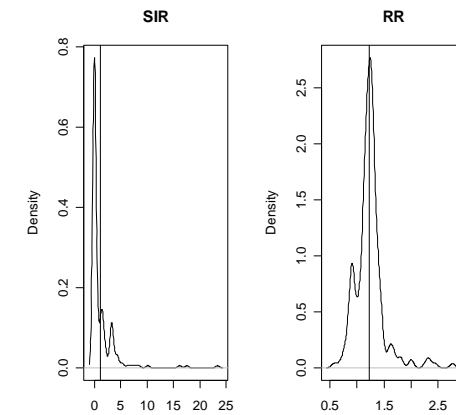
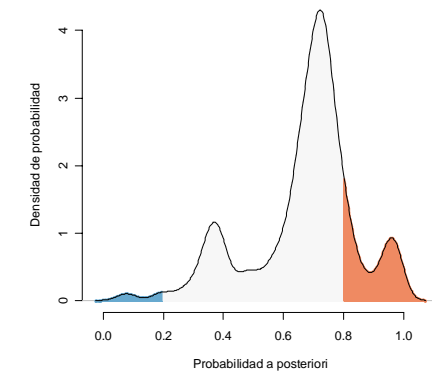
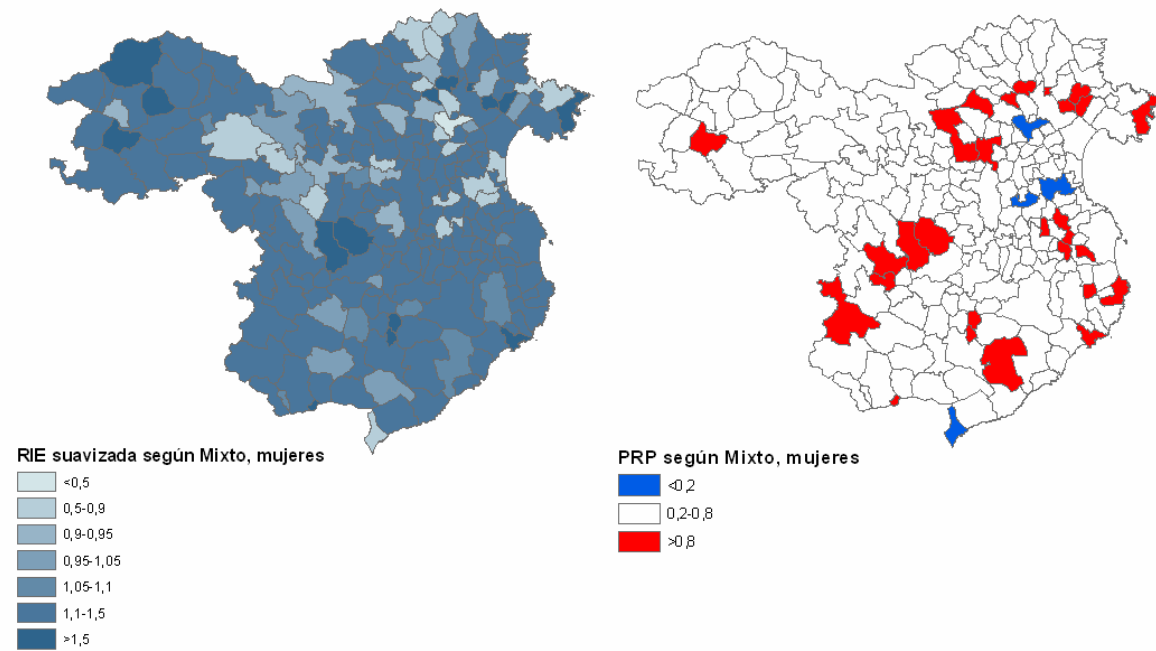


Figura 7.- Cáncer de tiroides, en mujeres. Modelo Mixto. Fuente: construcción propia.

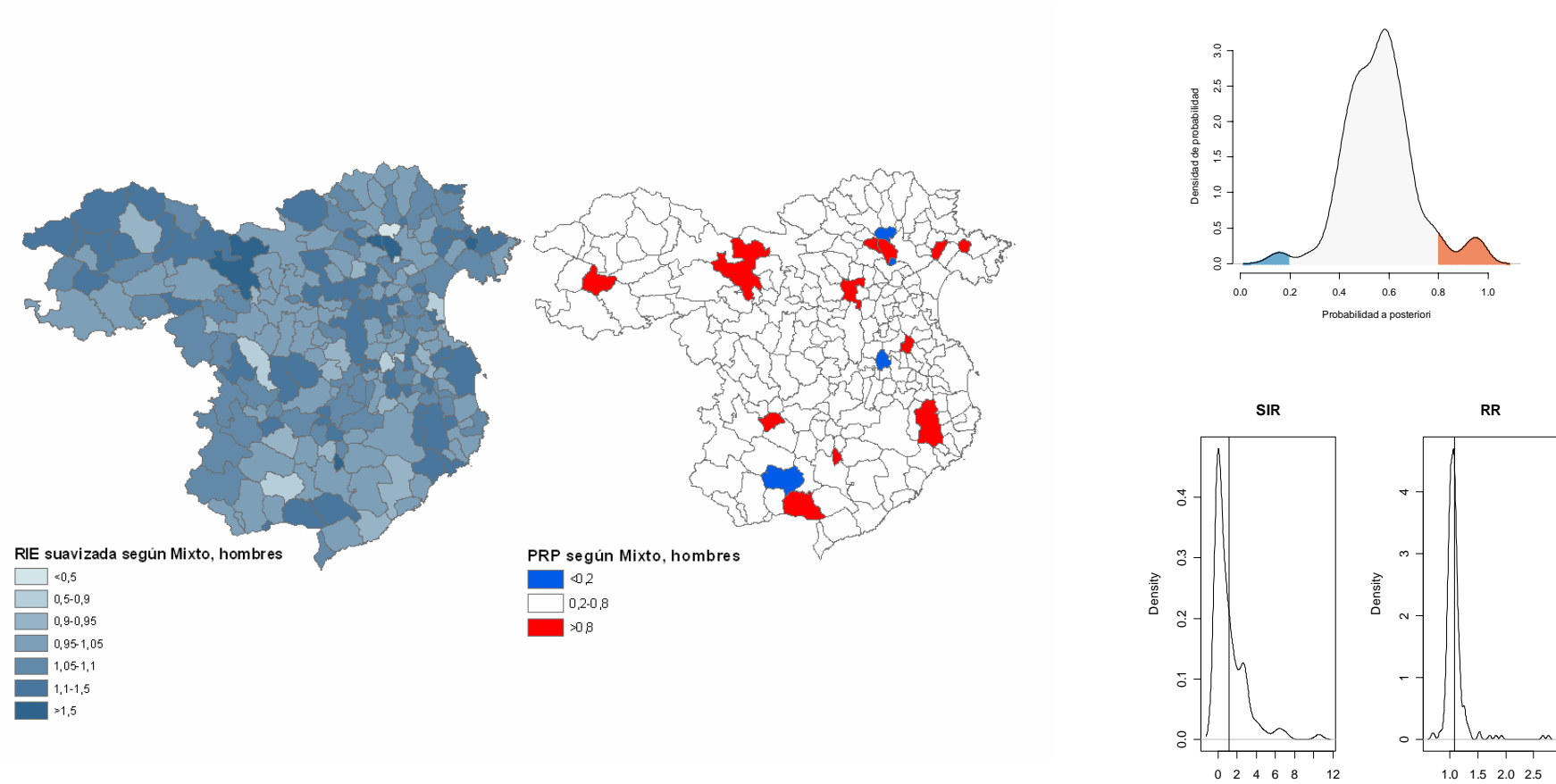


Figura 8.- Linfoma no Hodgkin, en hombres. Modelo Mixto. Fuente: construcción propia.

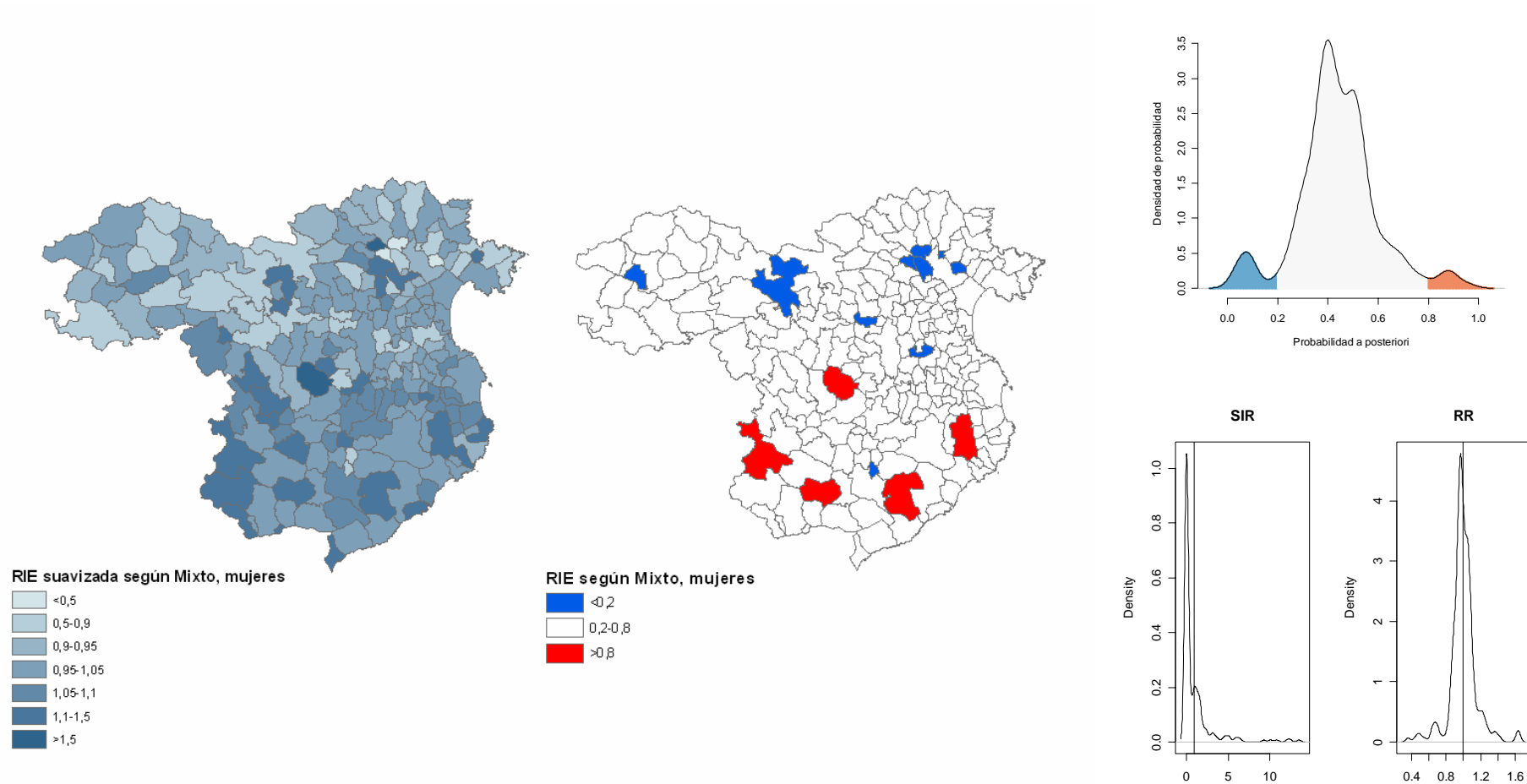


Figura 9.- Linfoma no Hodgkin, en mujeres. Modelo Mixto. Fuente: construcción propia.

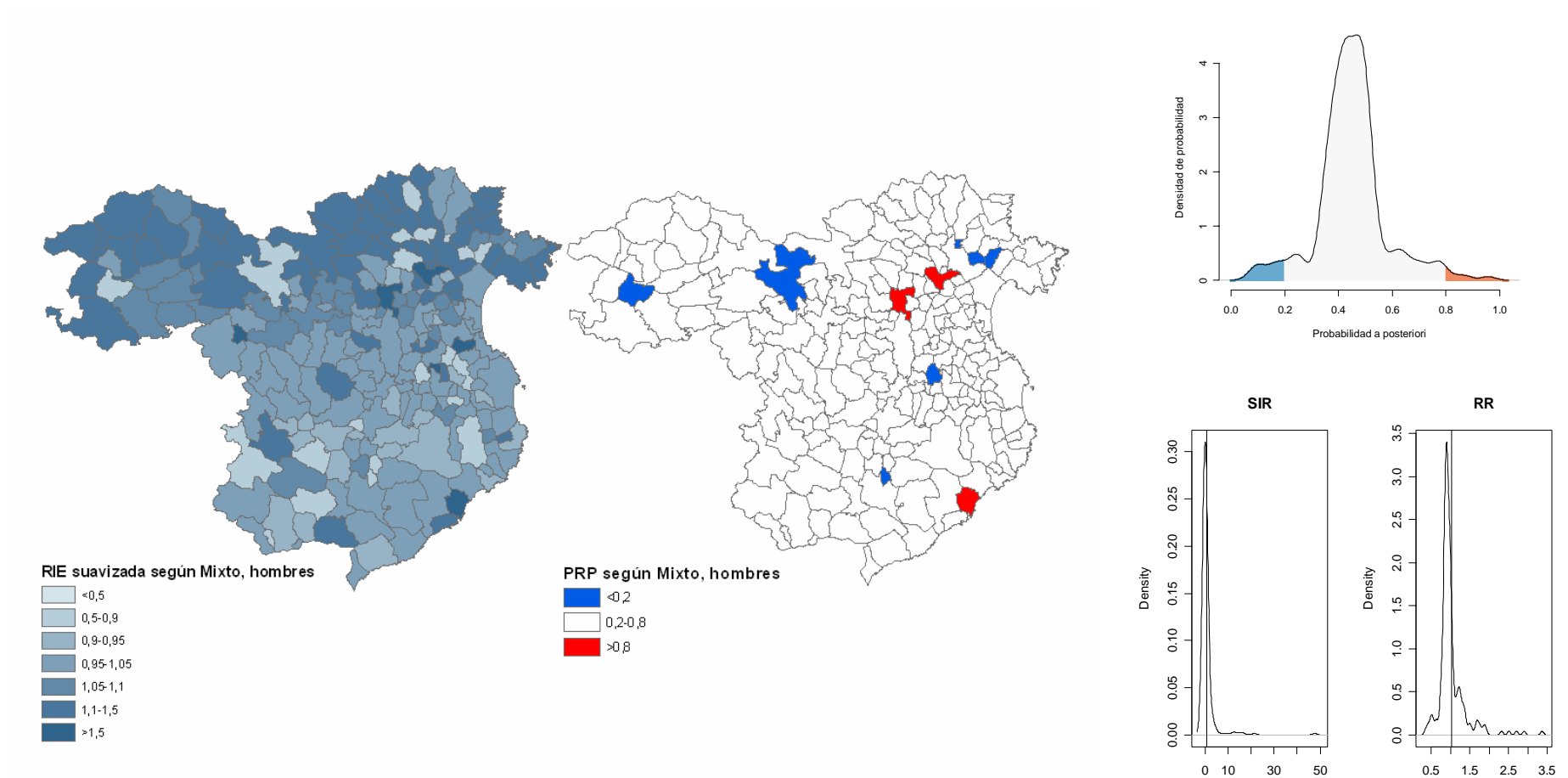


Figura 10.- Linfoma Hodgkin, en hombres. Modelo Mixto. Fuente: construcción propia.

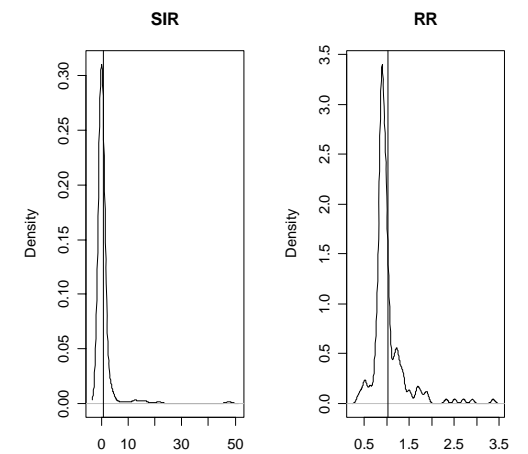
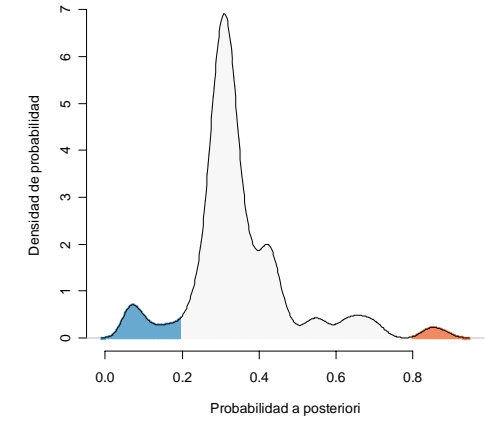
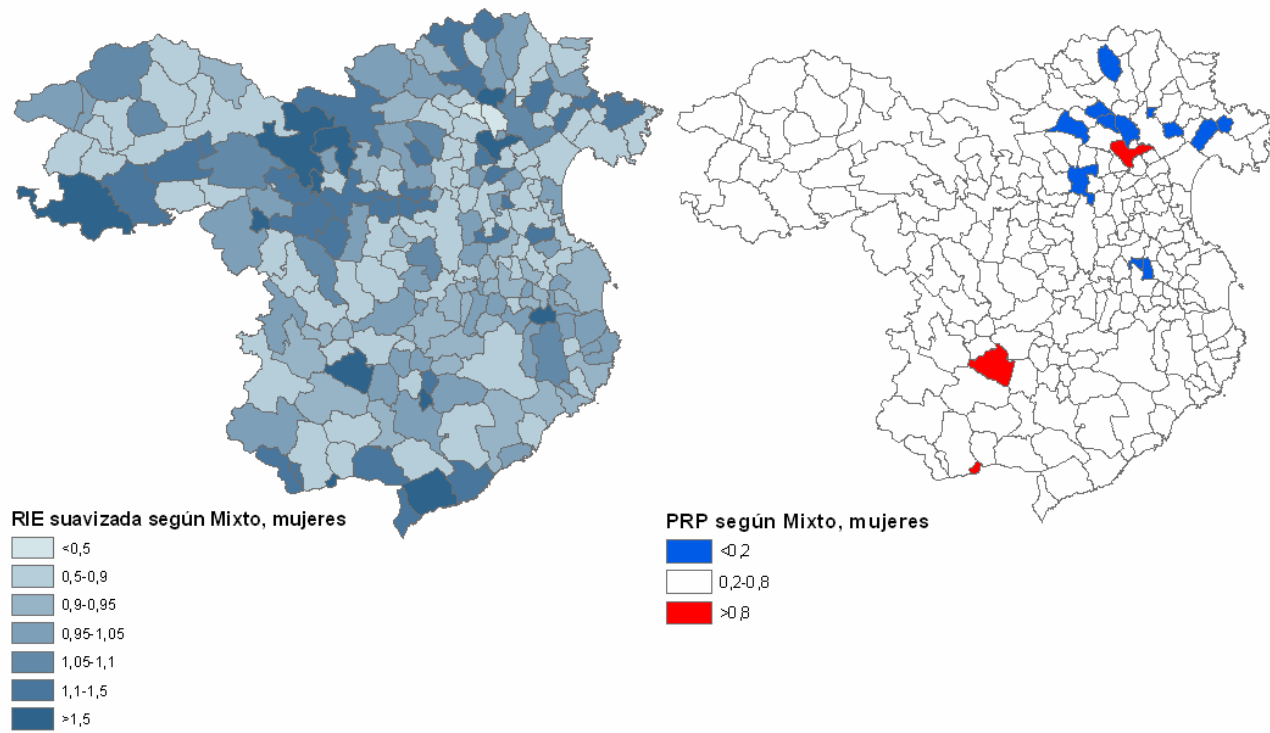


Figura 11.- Linfoma Hodgkin, en mujeres. Modelo Mixto. Fuente: construcción propia.

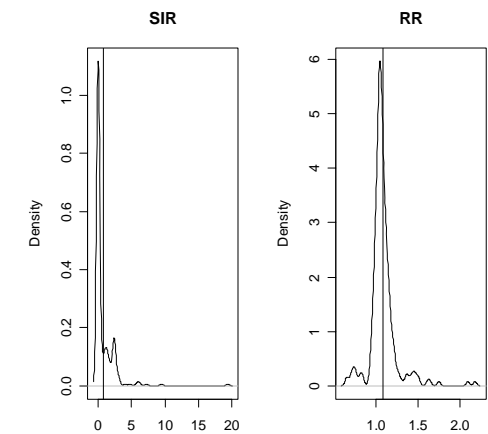
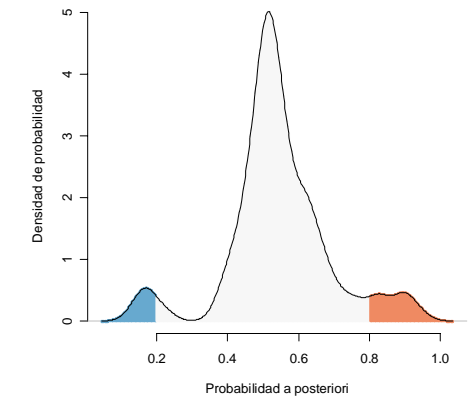
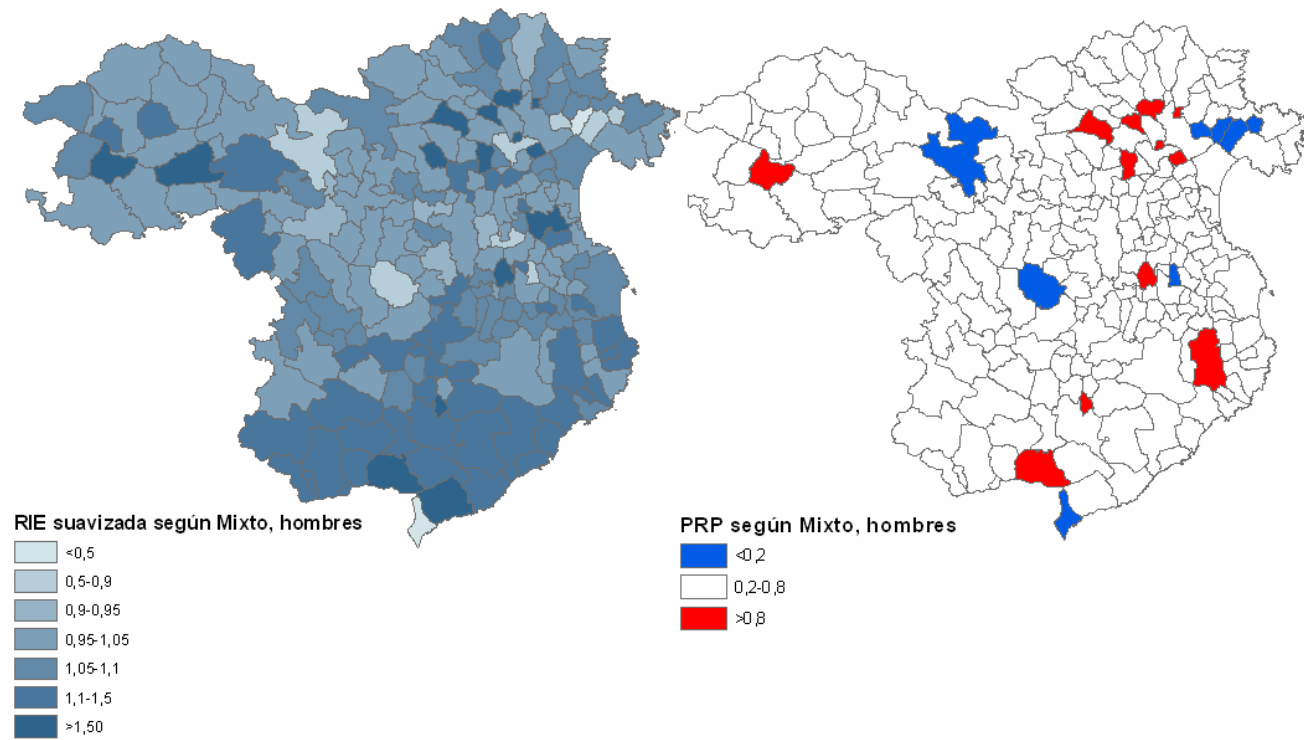


Figura 12.- Mieloma, en hombres. Modelo Mixto. Fuente: construcción propia.

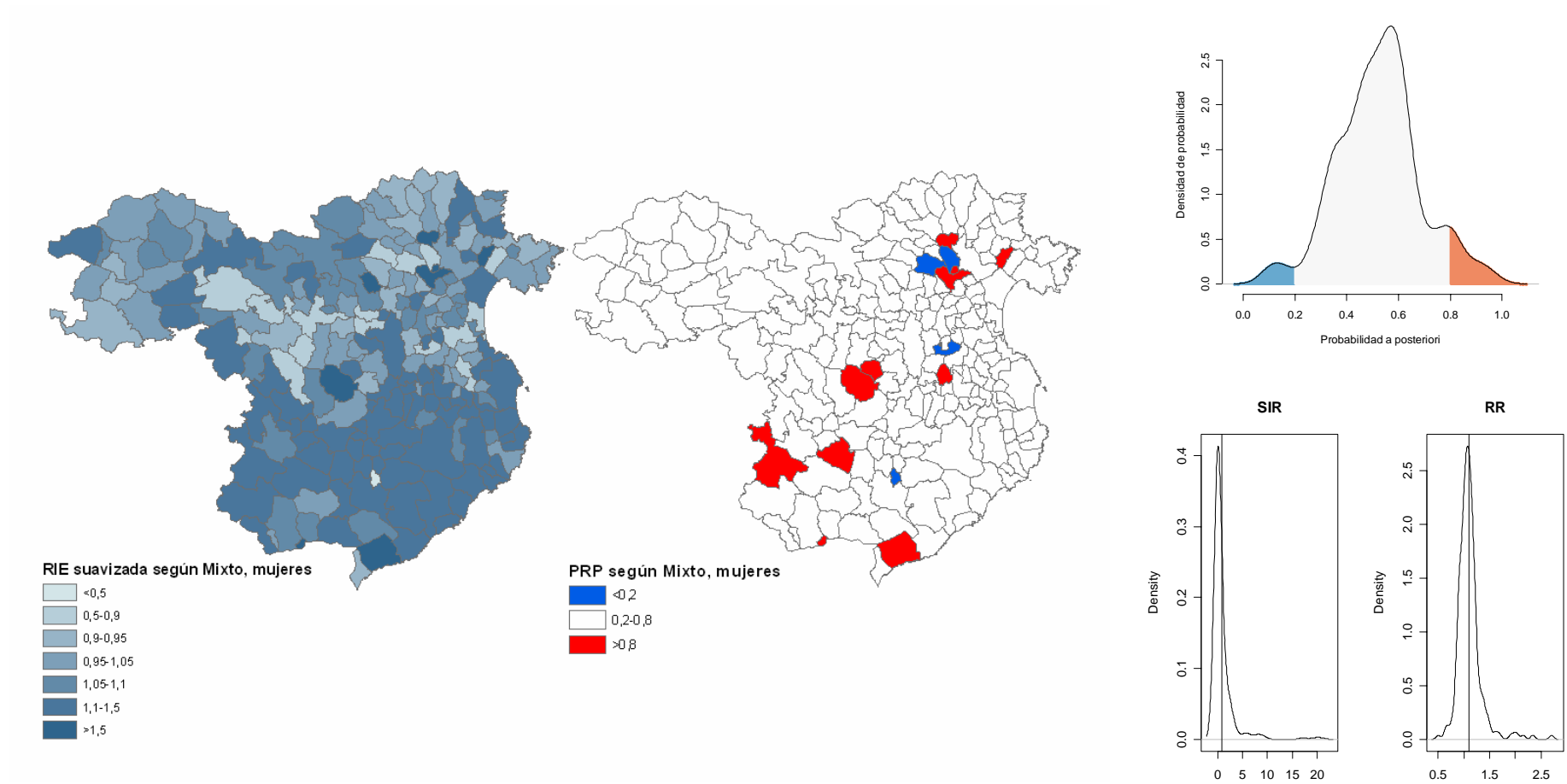


Figura 13.-Mieloma, en mujeres. Modelo Mixto. Fuente: construcción propia.

Como se ve en las Figuras 2 a 13, las RIE se suavizaron bastante en todos los casos. Véanse los histogramas en la esquina inferior derecha y, en menor medida, el rango de colores más uniformes en los mapas de las RIE en la izquierda.

La estructura geográfica intuida en el capítulo anterior (Figuras 26 a 31), parece confirmarse en algún caso. Obsérvense los municipios con exceso de riesgo, en rojo, en los mapas de probabilidad a posteriori. En particular, cáncer de tráquea, bronquios y pulmón, ambos sexos; vejiga urinaria, hombres; y tiroides, mujeres. Menos clara parece la estructura geográfica en las RIE del linfoma no Hodgkin, ambos sexos (un poco más clara en hombres); y en mieloma, ambos sexos.

El resto de neoplasias: vejiga urinaria, mujeres; tiroides, hombres; linfoma Hodgkin, ambos sexos; presentan algunos municipios aislados con exceso de riesgo, sin que en ningún caso se aprecie ninguna conducta sistemática.

b- Modelos de asociación geográfica

b.1.- Variable explicativa: índice socioeconómico. Resultados de la estimación

Regió Sanitaria Girona

Los parámetros asociados al índice de privación (recogidos en dos factores en el caso de la RSG) resultaron estadísticamente significativos en, aproximadamente, la mitad de las neoplasias de interés medioambiental. Concretamente en cáncer de tráquea, bronquios y pulmón, ambos sexos; vejiga urinaria en mujeres; cáncer de tiroides, mujeres; linfoma no Hodgkin, en hombres; linfoma Hodgkin, ambos sexos.

El primer factor, muy correlacionado con el desempleo ($-0,787$ en hombres y $-0,739$ en mujeres) y con trabajadores manuales ($0,828$ en hombres y $0,814$ en mujeres), resumió las variables de ocupación en el municipio. Por este motivo, el signo del parámetro asociado a este factor debe ser interpretado de forma inversa al de privación. El segundo factor resultó muy correlacionado con la variable instrucción insuficiente ($0,949$ en hombres y $0,901$ en mujeres).

Los parámetros asociados a ambos factores fueron conjuntamente significativos en cáncer de tráquea, bronquios y pulmón, en mujeres, y en vejiga urinaria, en hombres (véase Tabla 3 y Figura 14). El primer factor resultó significativo en el cáncer de tráquea, bronquios y pulmón, en mujeres; tiroides, mujeres; y mieloma, mujeres. El segundo factor (únicamente) fue significativo en el linfoma no Hodgkin, ambos sexos.

En todo caso (véase Tabla 3 y Figura 14) de los resultados se desprende que una mayor privación implicó una mayor incidencia de ciertas neoplasias de interés

medioambiental⁶. En particular, cáncer de tráquea, bronquios y pulmón, ambos sexos; vejiga urinaria, mujeres; tiroides, mujeres; linfoma Hodgkin, ambos sexos.

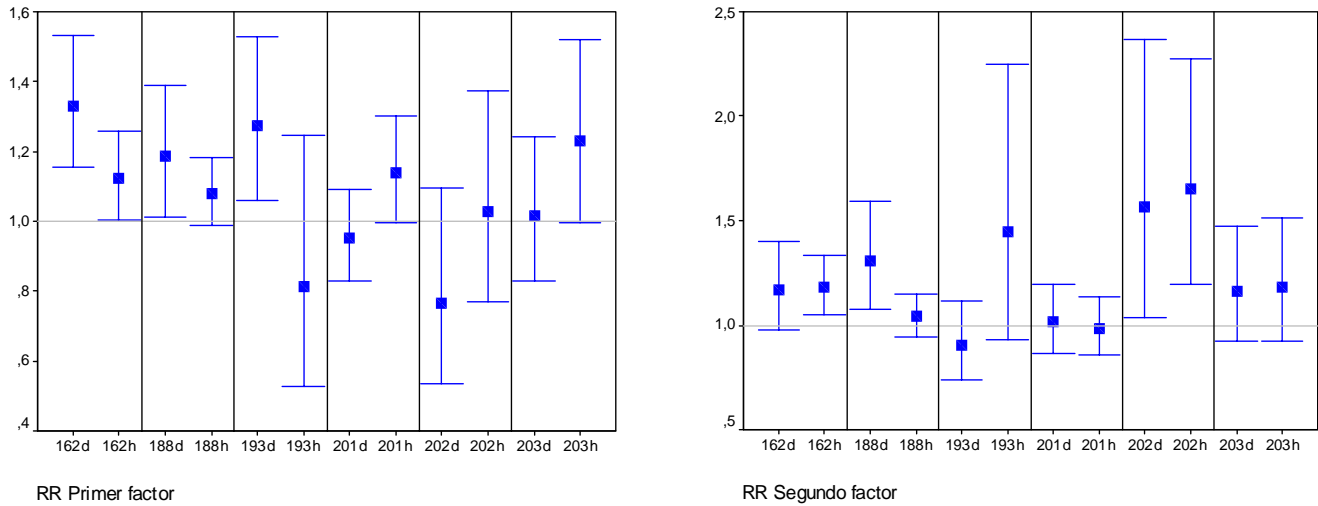


Figura 14.- Riesgo relativo asociado al índice de privación. Fuente: construcción propia.

Sin embargo, aunque en los casos comentados el índice de privación resultó estadísticamente significativo, la variabilidad, tanto espacial como no espacial, apenas vino explicada por este índice. En este sentido, obsérvese que las desviaciones típicas de los efectos aleatorios variaron muy poco cuando se introdujo el índice (Tablas 1 a 3).

⁶ En la Figura 14 se ha invertido el signo del primer factor para facilitar la interpretación.

Cáncer de:	Tráquea, bronquios, pulmón, hombres	Tráquea, bronquios, pulmón, mujeres	Vejiga urinaria, hombres	Vejiga urinaria, mujeres	Tiroides, hombres	Tiroides, mujeres	Linfoma no Hodgkin, hombres	Linfoma no Hodgkin, mujeres	Linfoma Hodgkin, hombres	Linfoma Hodgkin, mujeres	Mieloma, hombres	Mieloma, mujeres
Índice de privación												
beta1	-0,11663(0,05715)	-0,28495(0,07251)	-0,07829(0,04583)	-0,17061(0,08137)	0,20823(0,21921)	-0,24110(0,09935)	-0,13042(0,06864)	0,04923(0,07060)	-0,02768(0,14764)	0,26934(0,18297)	-0,20691(0,10775)	-0,01522(0,10362)
beta2	0,16678(0,06111)	0,15736(0,09162)	0,04073(0,04956)	0,26706(0,10031)	0,37038(0,22504)	-0,09786(0,10600)	-0,01486(0,07150)	0,01619(0,08321)	0,50021(0,16384)	0,44923(0,21004)	0,16531(0,12594)	0,15216(0,11902)
Efectos aleatorios												
Espacial	0,11437(0,08030)	0,39051(0,27131)	0,18249(0,14962)	0,31740(0,25814)	1,42038(0,84009)	0,69636(0,39571)	0,25650(0,19999)	0,37793(0,24745)	0,50671(0,43481)	0,88372(0,64343)	0,55037(0,34888)	0,57627(0,38461)
No Espacial	0,59858(0,04282)	0,15877(0,09884)	0,21172(0,05314)	0,25814(0,20686)	0,47976(0,28881)	0,18563(0,13390)	0,25062(0,11634)	0,22112(0,11893)	0,29077(0,18975)	0,35523(0,27074)	0,13532(0,09478)	0,20112(0,15819)
Salto		0,50811(0,24551)	0,38065(0,20314)	0,45201(0,25499)	0,92829(0,48280)	0,69420(0,34188)	0,36259(0,21555)	0,46172(0,27369)	0,66397(0,37511)	0,58145(0,42546)	0,64898(0,31778)	0,57676(0,32883)
AIC	586,244	501,662	564,813	913,245	1696,395	538,336	604,868	457,983	2891,648	3458,926	336,088	499,143
BIC	1196,747	1010,765	1115,123	2149,867	4441,478	1128,209	1194,667	858,321	7636,345	9252,618	685,474	1081,723
DIC	369,779	238,304	302,876	232,327	158,758	235,328	309,65	261,645	203,569	170,044	158,477	191,259

Tabla 3.– Resultados de la estimación en los modelos de asociación geográfica. Variable explicativa: índice socioeconómico. Fuente: construcción propia.

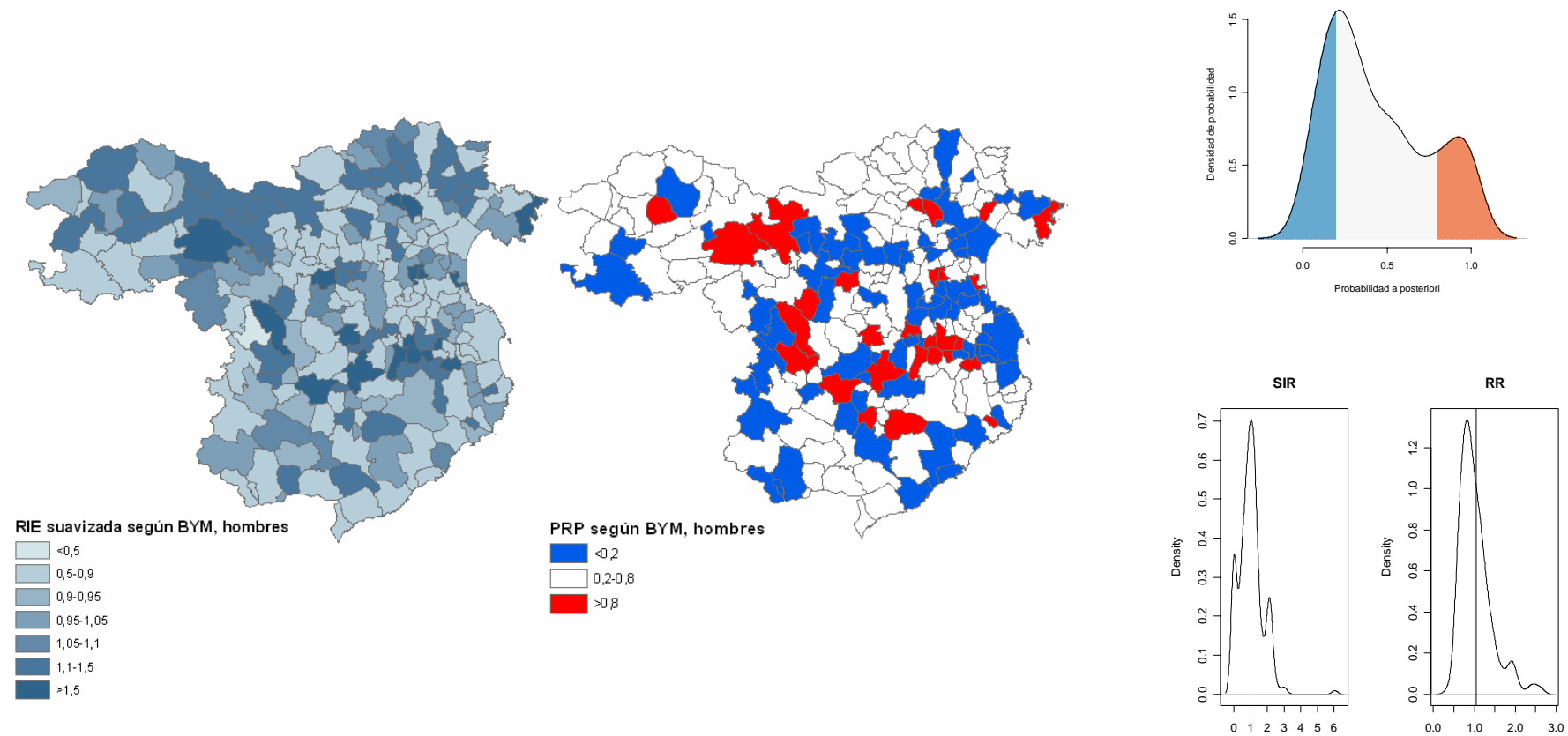


Figura 15.- Cáncer de tráquea, bronquios y pulmón, en hombres, introduciendo el índice socioeconómico (Fuente: construcción propia).

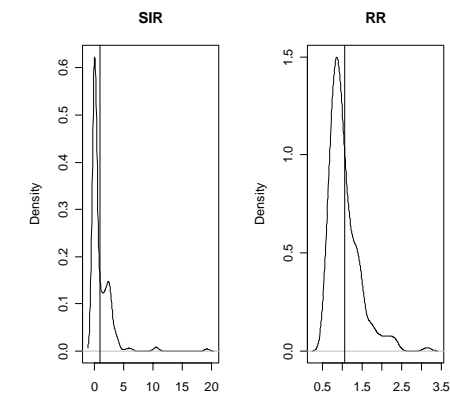
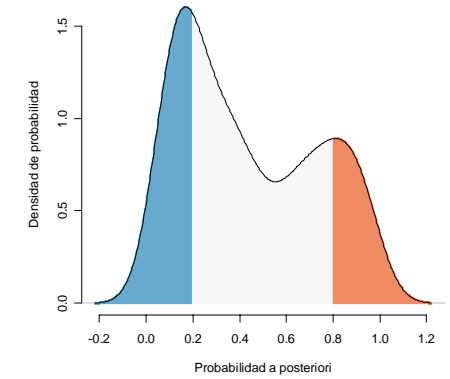
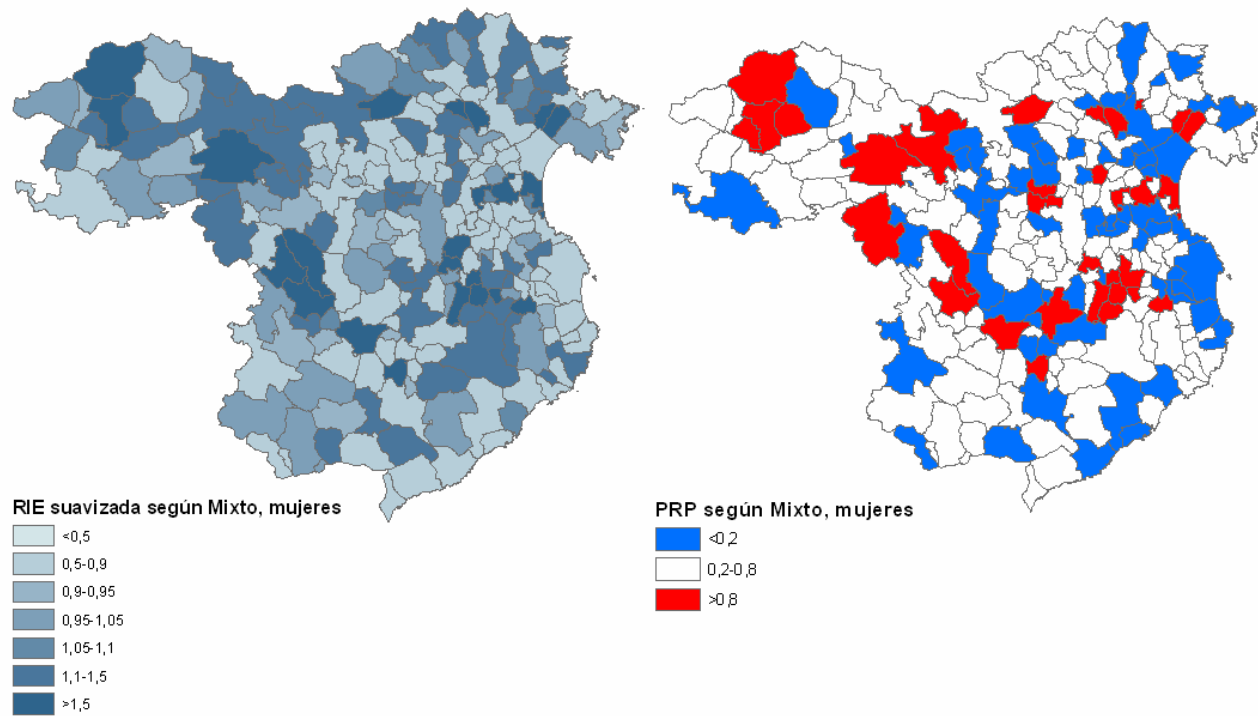


Figura 16.- Cáncer de tráquea, bronquios y pulmón, en mujeres, introduciendo el índice socioeconómico (Fuente: construcción propia).

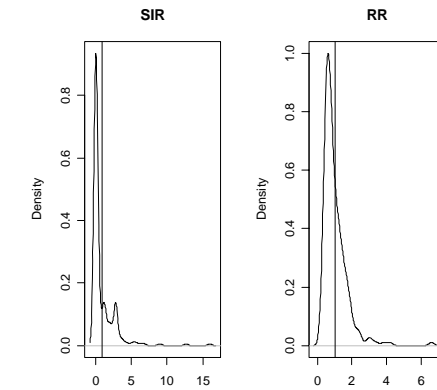
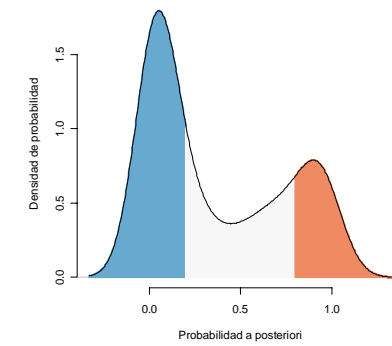
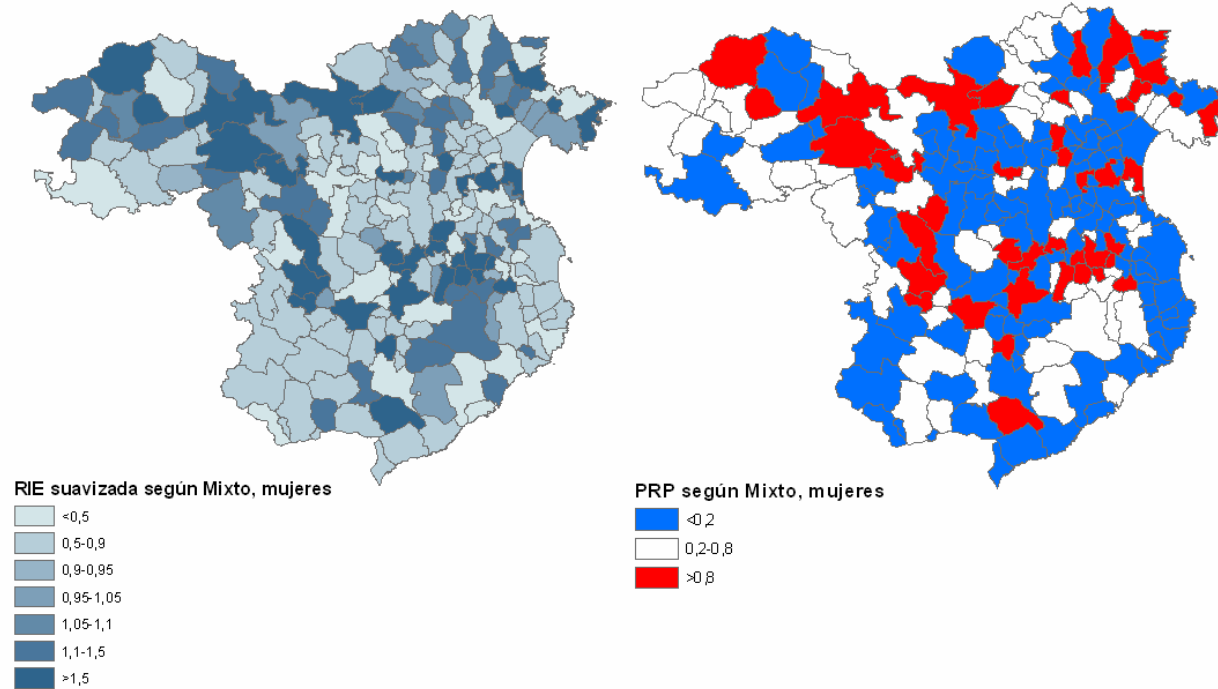


Figura 17.- Cáncer de vejiga urinaria, en mujeres, introduciendo el índice socioeconómico (Fuente: construcción propia).

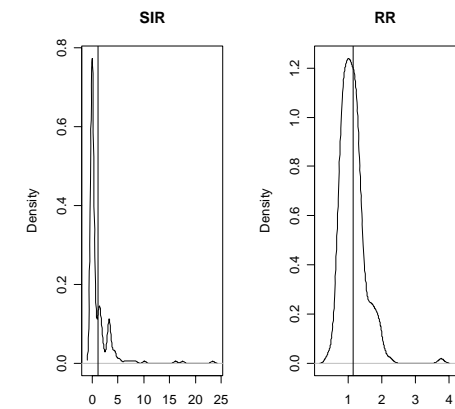
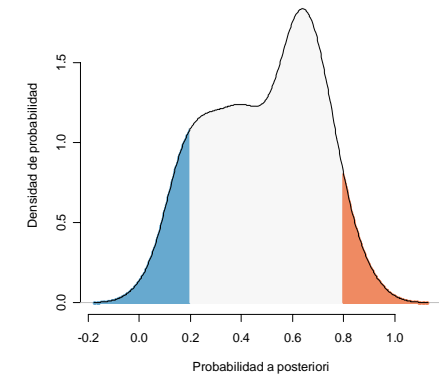
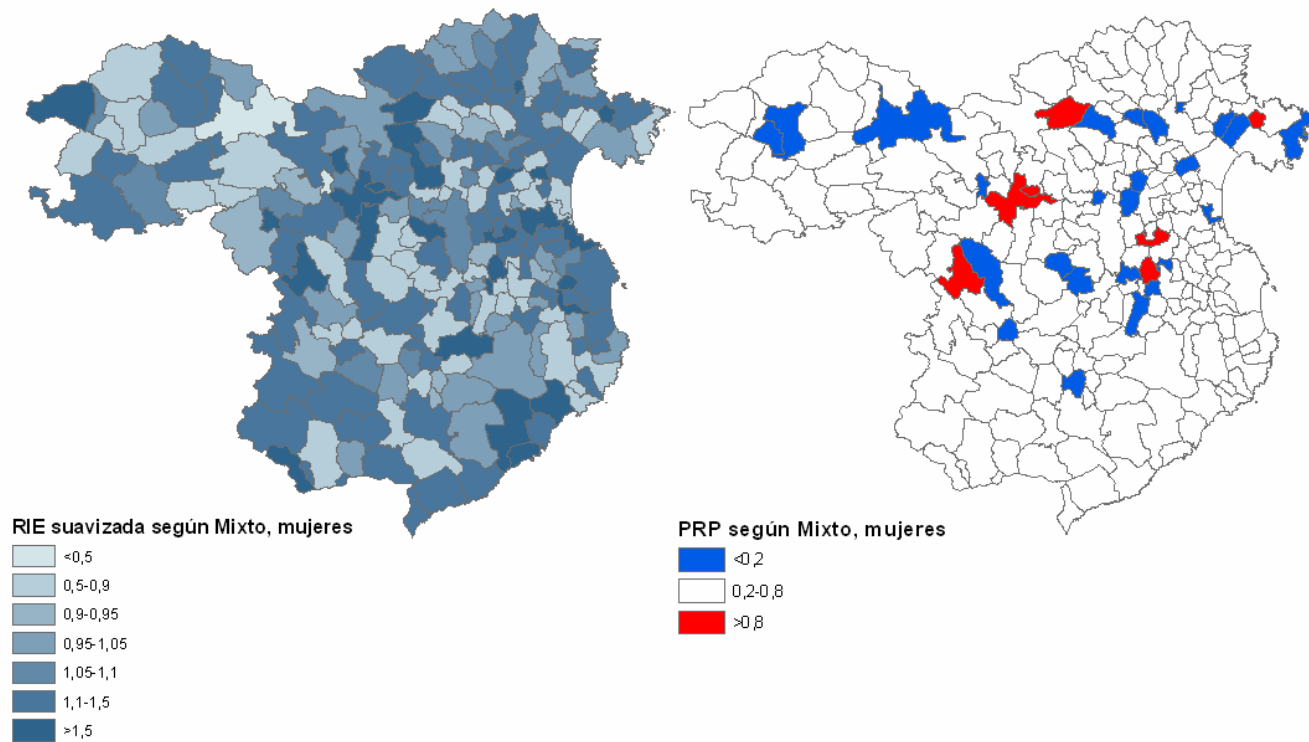


Figura 18.- Cáncer de tiroides, en mujeres, introduciendo el índice socioeconómico (Fuente: construcción propia).

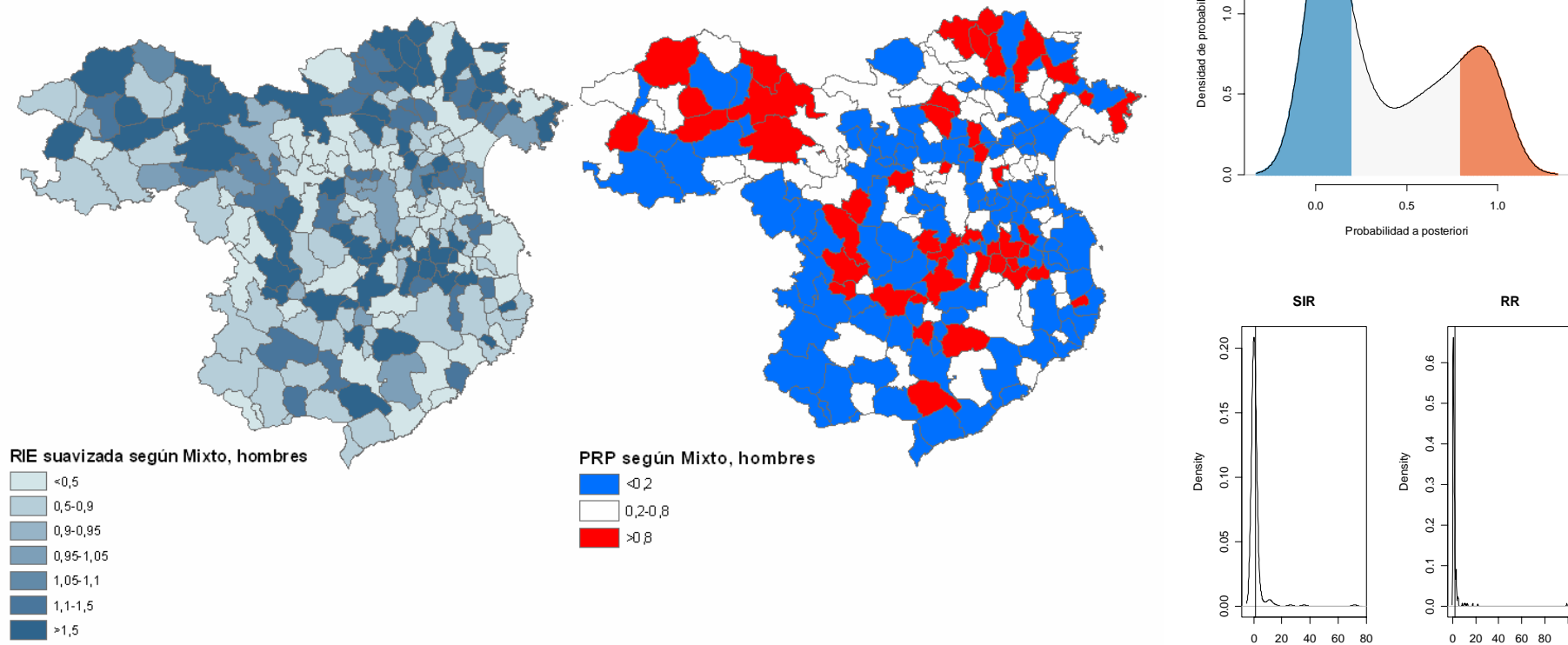


Figura 19.- Cáncer de Linfoma Hodgkin, en hombres, introduciendo el índice socioeconómico (Fuente: construcción propia).

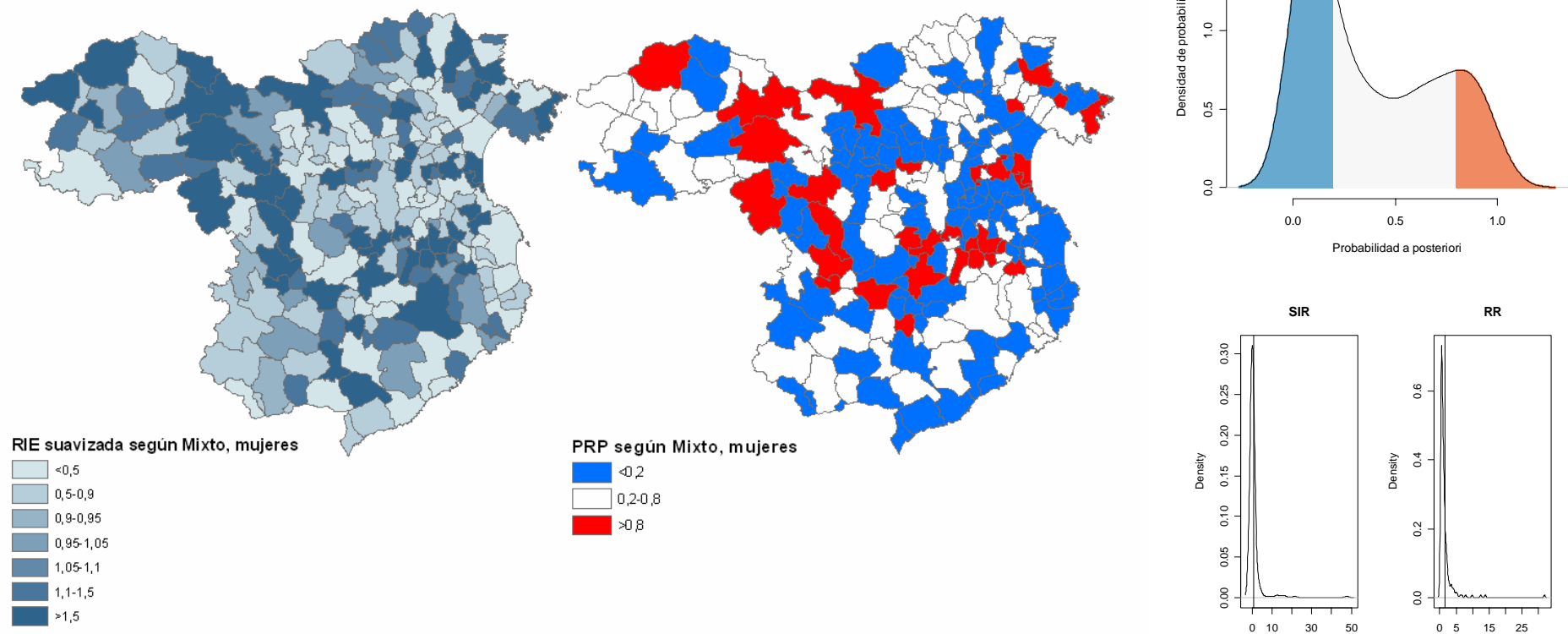


Figura 20.- Cáncer de Linfoma Hodgkin, en mujeres, introduciendo el índice socioeconómico (Fuente: construcción propia).

En las Figuras 15 a 20 se han representado aquellas neoplasias de interés medioambiental asociadas de forma estadísticamente significativa con el índice de privación.

Los patrones vistos en la suavización de las RIE se conservan para aquellas neoplasias con un porcentaje de ceros relativamente pequeño en términos relativos (por ejemplo, cáncer de tráquea, bronquios y pulmón, hombres). El patrón se modifica, llegando incluso a desaparecer el exceso de riesgo en aquellas neoplasias con muchos ceros (por ejemplo, cáncer de tiroides y cáncer de tráquea, bronquios y pulmón, en mujeres).

b.2- Variable explicativa: contaminantes atmosféricos. Resultados de la estimación

Regió Sanitaria Girona

En el caso de la Regió Sanitària Girona, se introdujo el contaminante ozono, O_3 , bien solo, bien junto con el índice de privación en el caso que de éste hubiese sido estadísticamente significativo (cáncer de tráquea, bronquios y pulmón, ambos sexos; vejiga urinaria mujeres; tiroides mujeres; y linfoma Hodgkin ambos sexos).

A fin de minimizar el sesgo ecológico se introdujo la variabilidad del contaminante dentro de cada área y se consideró que el efecto del contaminante pudiese variar en cada una de las áreas (efecto aleatorio). Los resultados de la estimación se muestran en la Tabla 4.

Como se ve (Tabla 4 y Figura 21) el parámetro del contaminante resultó (apenas) estadísticamente significativo únicamente en el caso de cáncer de vejiga urinaria en hombres. Ni la variabilidad, ni la distribución geográfica de los municipios con exceso de riesgo, sin embargo, resultaron apenas modificadas cuando se introdujo el ozono como variable explicativa.

Cáncer de:	Tráquea,bronquios, pulmón, hombres	Tráquea,bronquios, pulmón, mujeres	Vejiga urinaria, hombres	Vejiga urinaria, mujeres	Tiroides, hombres	Tiroides, mujeres	Linfoma no Hodgkin, hombres	Linfoma no Hodgkin, mujeres	Linfoma Hodgkin, hombres	Linfoma Hodgkin, mujeres	Mieloma, hombres	Mieloma,mujeres
contaminante atmosférico, ozono												
beta3	0,59014(0,54644)	0,14782(0,19952)	0,23753(0,10903)	0,23789(0,18649)	0,02359(0,48523)	0,22832(0,39208)	0,37515(0,28411)	-0,15468(0,23467)	0,24979(0,28108)	-0,06516(0,38151)	0,27614(0,29332)	0,09366(0,29644)
Efectos aleatorios												
Espacial	0,16199(0,0889)	0,31806(0,24746)	0,21617(0,16196)	0,33164(0,26132)	1,12082(0,78514)	0,53012(0,33894)	0,24243(0,20295)	0,37110(0,27224)	0,85981(0,63378)	0,82236(0,68383)	0,40608(0,31078)	0,63785(0,47794)
No Espacial	0,13509(0,08516)	0,11165(0,08381)	0,06225(0,04682)	0,12284(0,09339)	0,40451(0,29829)	0,11522(0,09122)	0,08547(0,05967)	0,11999(0,09505)	0,24519(0,19007)	0,31849(0,25918)	0,12592(0,09287)	0,14549(0,12134)
Salto		0,43833(0,25667)	0,26315(0,19369)	0,42565(0,28957)	0,73578(0,46932)	0,49698(0,28553)	0,40354(0,21981)	0,40227(0,27386)	0,63249(0,37741)	0,67717(0,42386)	0,43569(0,29179)	0,12134(0,55099)
AIC	565,092	278,608	401,954	264,261	213,993	263,121	343,679	341,051	212,236	185,8	200,822	238,673
BIC	1137,833	481,770	722,570	460,175	429,345	459,657	580,28	607,612	398,588	362,181	336,711	432,491
DIC	361,233	208,370	289,619	195,413	122,288	194,551	263,774	235,545	140,116	111,435	156,705	162,649

Tabla 4.- Resultados de la estimación en los modelos de asociación geográfica. Variable explicativa: Ozono. Fuente: construcción propia.

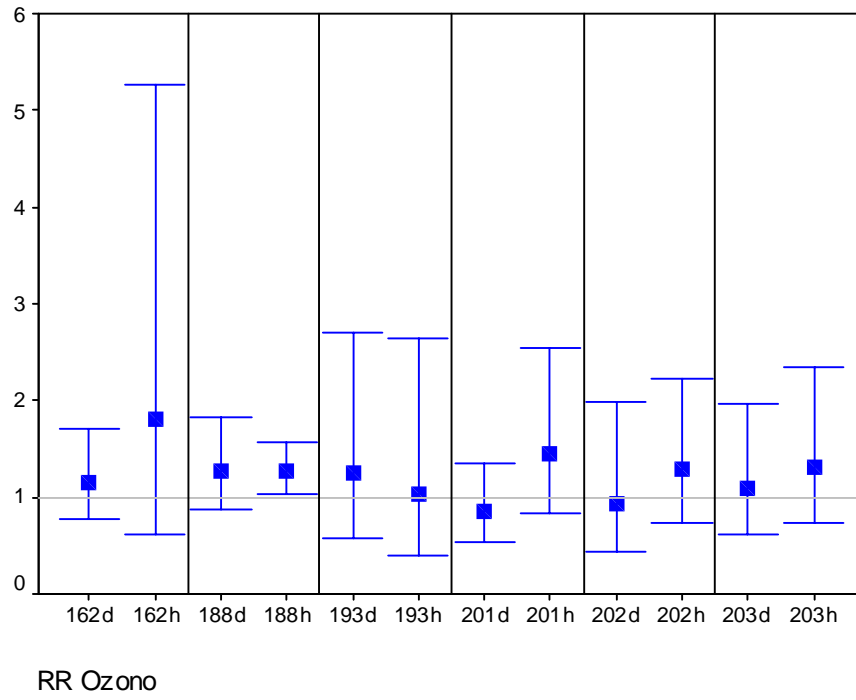


Figura 21.– Riesgo relativo asociado al ozono. Fuente: construcción propia.

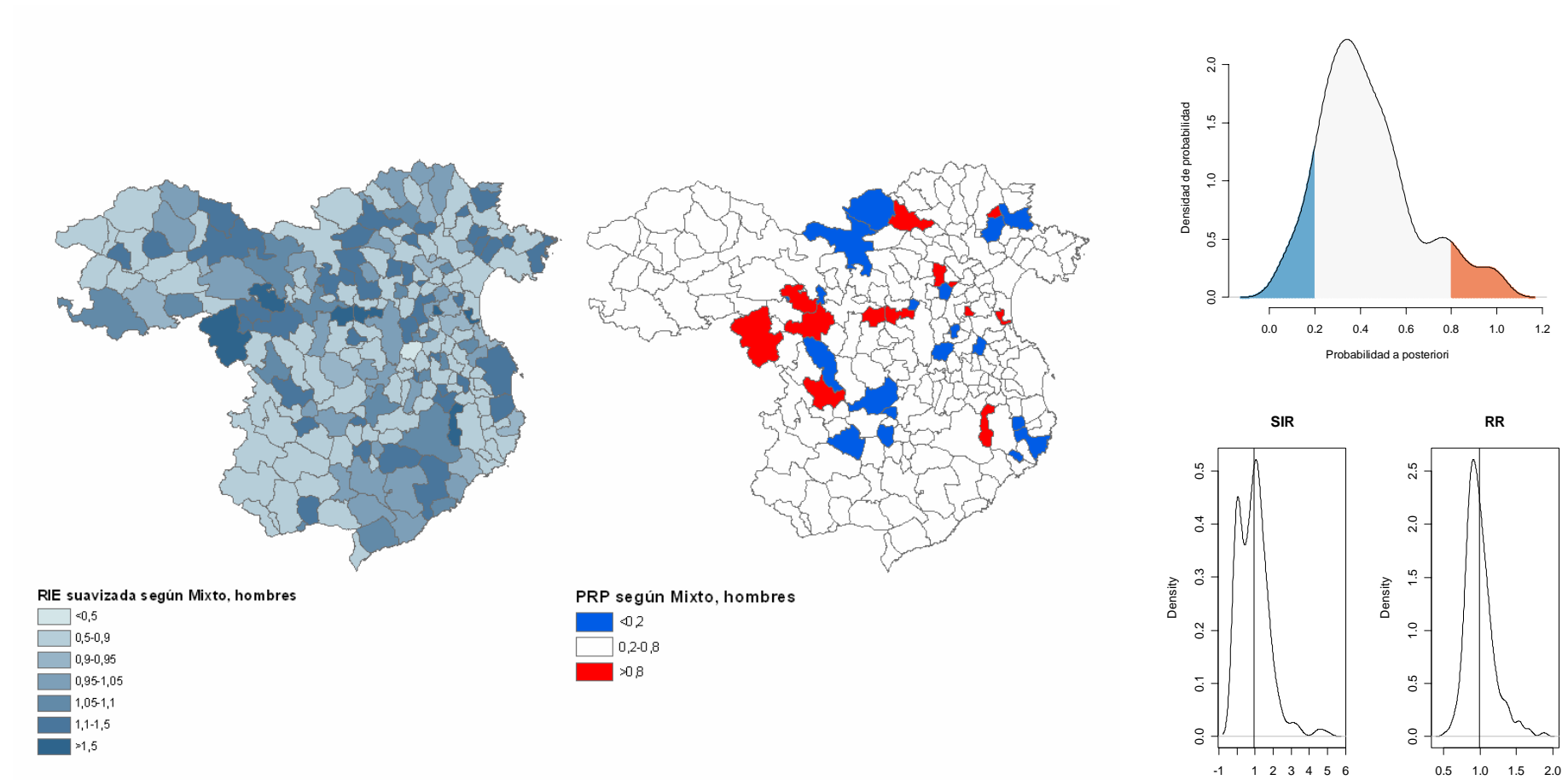


Figura 22.- Cáncer de vejiga urinaria, en hombres, introduciendo el ozono (Fuente: construcción propia).

4.1.4.- Resultados de la estimación. Área Metropolitana de Barcelona

a.-Mortalidad por las neoplasias de interés medioambiental en el Área Metropolitana de Barcelona

Tal y como se observa en la Tabla 5 se repitieron los resultados que se encontraron para la incidencia (Tablas 1 y 2). El mejor ajuste, en términos de AIC y BIC, lo proporciona el modelo mixto de riesgo relativo. Excepto en el caso de tráquea, bronquios y pulmón, en hombres, en el que el mejor ajuste lo proporcionó el modelo BYM. De hecho, la Figura 23 es muy parecida a la Figura 1, en el sentido que a partir del 40-50% de ceros el mejor ajuste lo proporcionó el modelo mixto de riesgo relativo.

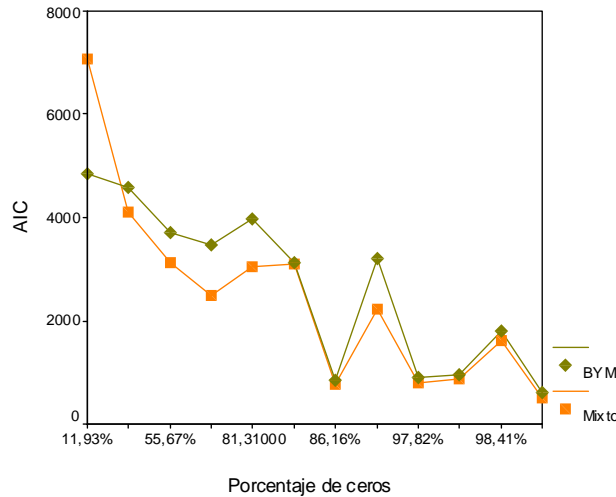


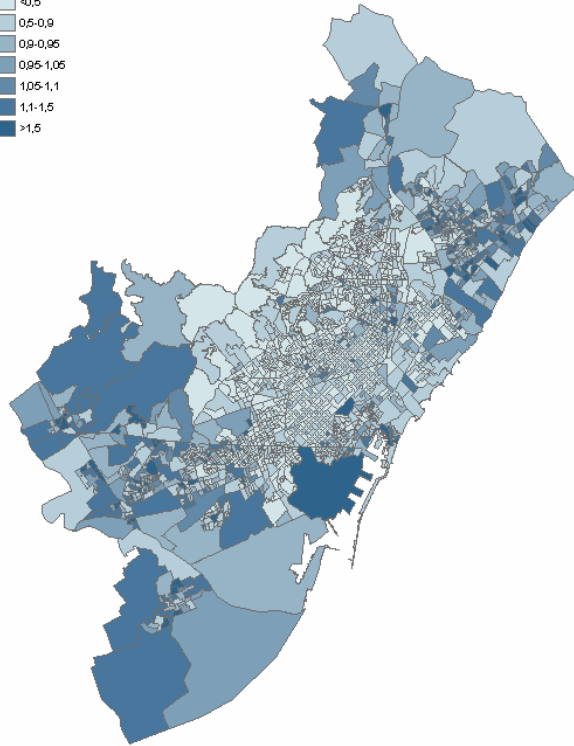
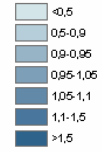
Figura 23.- Relación entre porcentaje de ceros y el estadístico AIC. Fuente: construcción propia.

CIE-9: 162, Tráquea, bronquios y pulmón		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,54219(0,06187)	0,45195(0,06192)	0,30765(0,06192)	
No Espacial	0,53652(0,02135)	0,14219(0,02081)		
AIC	4857,683	7067,61		
BIC	12663,282	19367,06		
DIC	3146,717	2963,26		
% ceros	11,93%			
Mujeres				
Espacial	0,50416(0,10640)	0,45640(0,10438)	0,50630(0,08315)	
No Espacial	0,79142(0,04559)	0,27701(0,04516)		
AIC	3467,693	2496,45		
BIC	8442,985	5532,83		
DIC	2169,785	2164,79		
% ceros	66,83%			
CIE-9: 188, Vejiga urinaria		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,63493(0,09038)	0,73627(0,10029)	0,79219(0,14355)	
No Espacial	0,57494(0,04206)	0,37132(0,06061)		
AIC	4592,317	4118,91		
BIC	11634,137	10348,87		
DIC	2687,908	2486,82		
% ceros	46,06%			
Mujeres				
Espacial	0,45734(0,15837)	0,74977(0,24781)	0,72178(0,28473)	
No Espacial	0,82385(0,08333)	0,41278(0,09478)		
AIC	3962,434	3037,47		
BIC	11240,935	9521,20		
DIC	1706,927	1685,02		
% ceros	81,31%			

CIE-9: 193, Tiroides		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Salto	
Espacial	0,47953(0,34289)	0,57964(0,34489)	0,47163(0,21571)	
No Espacial	0,70892(0,44019)	0,32763(0,11534)		
AIC	596,749	503,88		
BIC	1682,514	1327,72		
DIC	250,475	246,51		
% ceros	98,55%			
Mujeres				
Espacial	0,37319(0,26207)	0,66571(0,22042)	0,61237(0,41769)	
No Espacial	0,59572(0,35163)	0,23481(0,09128)		
AIC	893,598	793,72		
BIC	2461,057	2076,89		
DIC	381,371	377,57		
% ceros	97,82%			
CIE-9: 210, Linfoma no Hodgkin		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Salto	
Espacial	0,49809(0,16016)	0,30109(0,143589)	0,38512(0,16116)	
No Espacial	0,70245(0,11911)	0,18512(0,12658)		
AIC	3713,741	3133,31		
BIC	10806,047	9622,81		
DIC	1416,533	1405,63		
% ceros	55,67%			
Mujeres				
Espacial	0,33521(0,17951)	0,37959(0,14221)	0,42008(0,17134)	
No Espacial	1,06011(0,08057)	0,32008(0,07840)		
AIC	3134,572	3104,28		
BIC	8734,049	8174,60		
DIC	1477,502	1413,82		
% ceros	86,16%			
CIE-9: 211, Linfoma Hodgkin		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Salto	
Espacial	0,79006(0,34311)	0,73516(0,48986)	0,82661(0,34111)	
No Espacial	0,64329(0,52858)	1,02861(0,52459)		
AIC	845,302	770,00		
BIC	2563,562	2468,12		
DIC	285,823	275,67		
% ceros	86,16%			
Mujeres				
Espacial	0,52686(0,34403)	0,69589(0,38087)	0,62432(0,27446)	
No Espacial	1,09360(0,56255)	0,82898(0,60712)		
AIC	958,184	878,95		
BIC	3122,818	2793,71		
DIC	250,087	255,49		
% ceros	98,41%			
CIE-9: 212, Mieloma		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Salto	
Espacial	0,55479(0,20554)	0,36499(0,24457)	0,35454(0,23162)	
No Espacial	0,40703(0,20924)	0,25478(0,13507)		
AIC	1803,768	1604,56		
BIC	4693,196	3951,55		
DIC	848,652	827,71		
% ceros	98,41%			
Mujeres				
Espacial	0,36329(0,21504)	0,44249(0,19221)	0,50151(0,26613)	
No Espacial	0,70574(0,18064)	0,30911(0,16112)		
AIC	3204,624	2215,21		
BIC	9727,321	8688,23		
DIC	1041,853	1031,28		
% ceros	92,83%			

Tabla 5.- Resultados de la estimación para las neoplasias de interés medioambiental. Fuente: construcción propia.

SMR suavizado según BYM, hombres



PRP según BYM, hombres

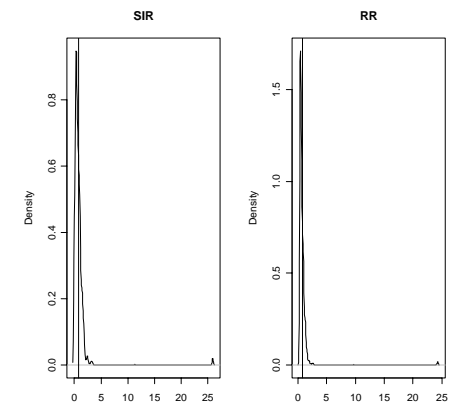
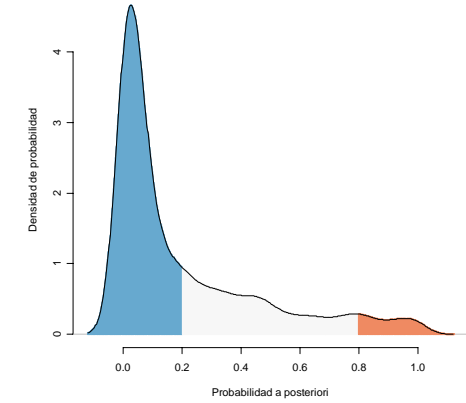
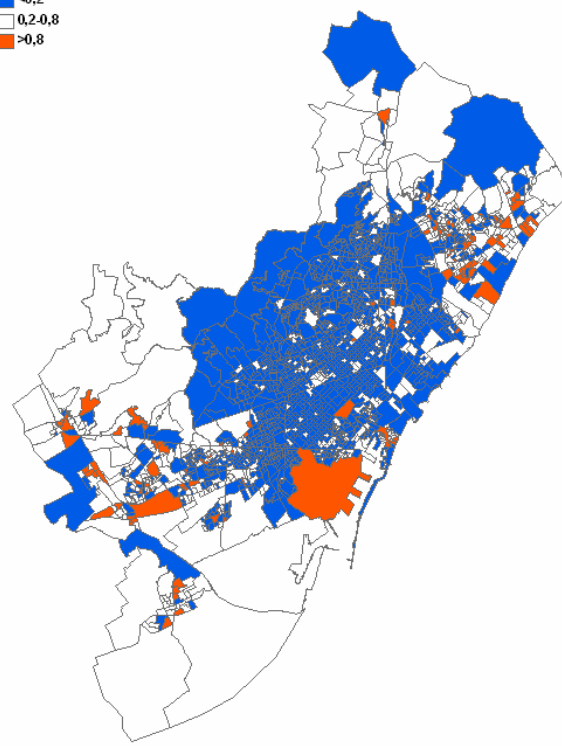
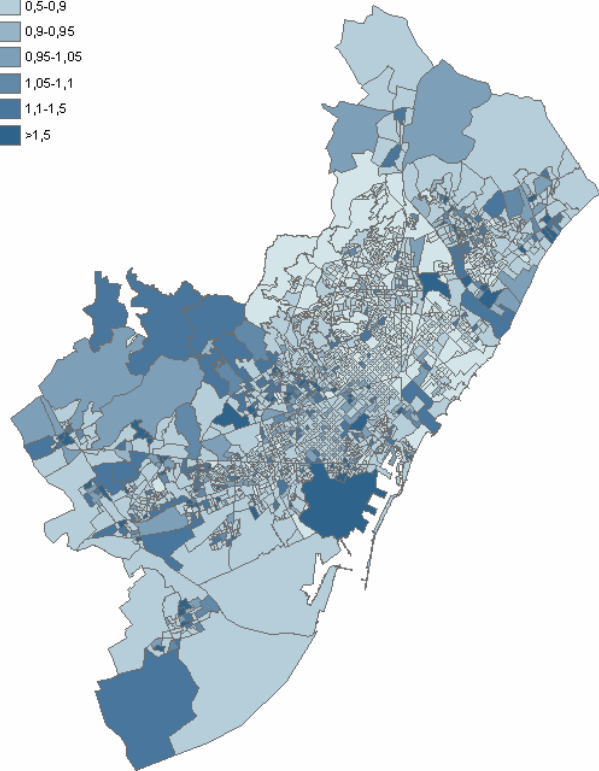
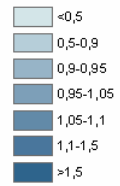


Figura 24.-Cáncer de tráquea, bronquios y pulmón, en hombres. Modelo BYM. Fuente: construcción propia.

SMR suavizado según Mixto, mujeres



PRP según Mixto, mujeres

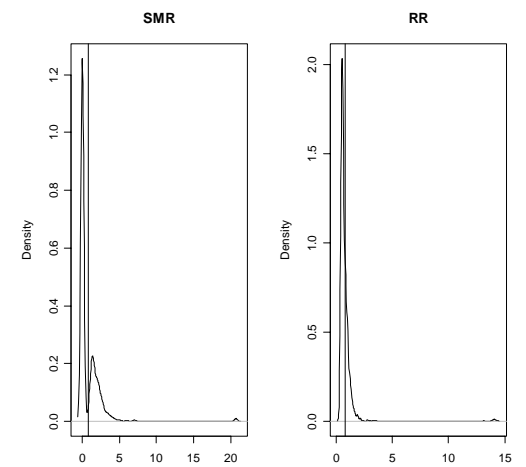
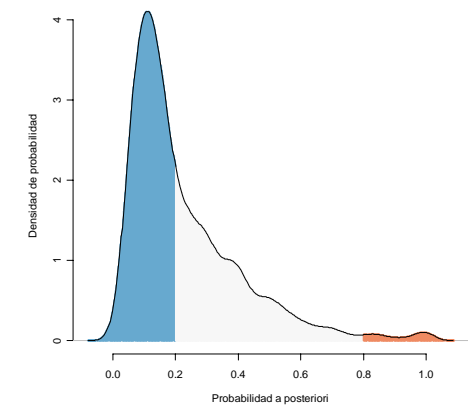
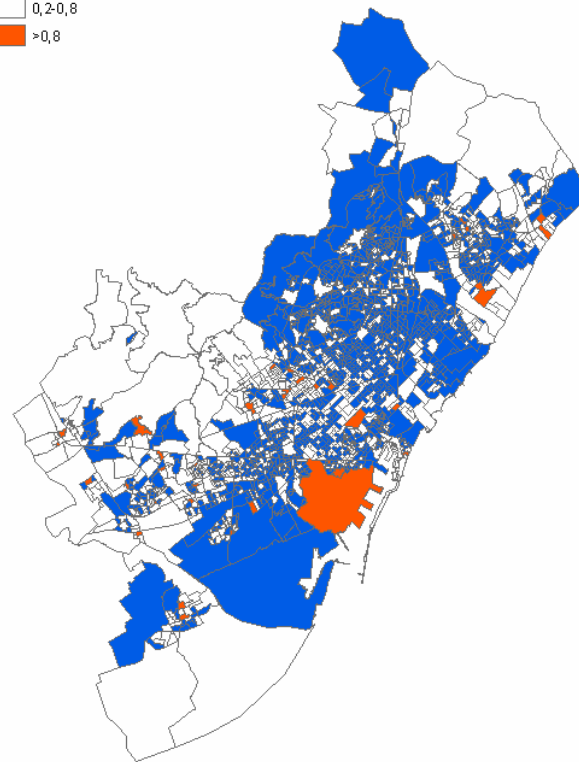
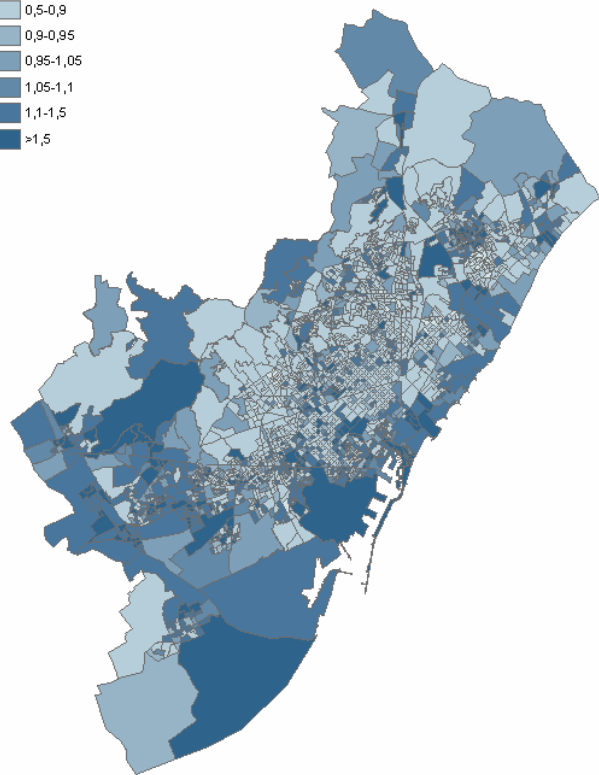
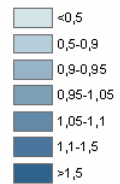


Figura 25.-Cáncer de tráquea, bronquios y pulmón, en mujeres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, hombres



PRP según Mixto, hombres

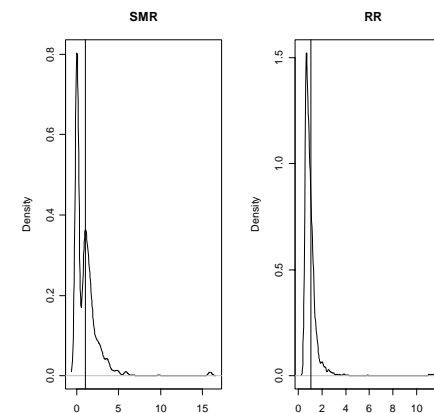
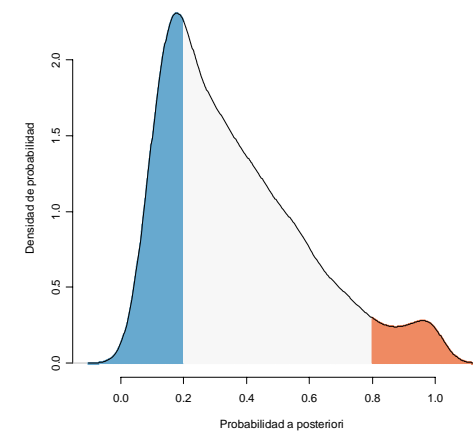
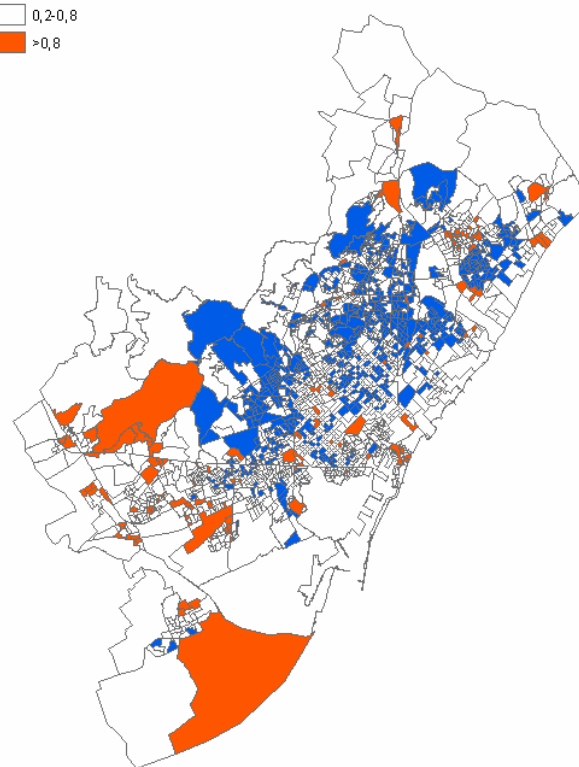
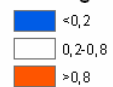
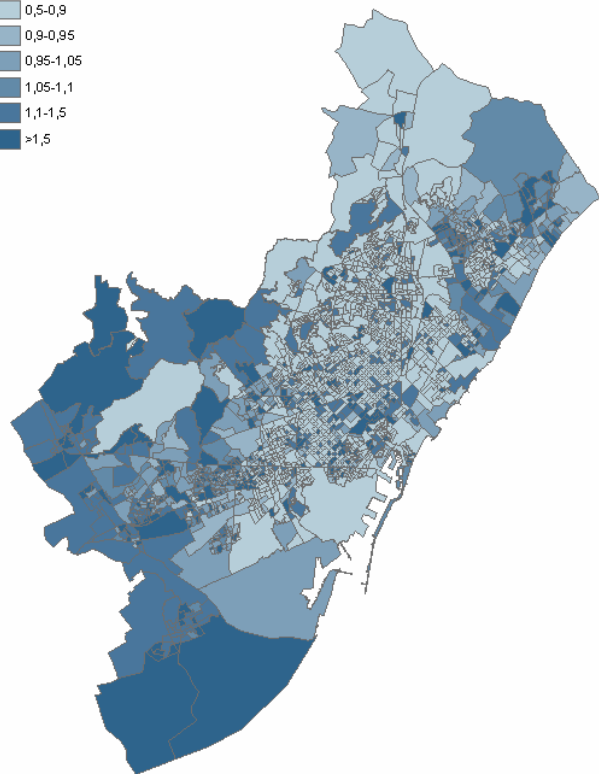
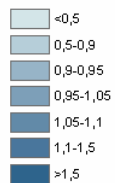


Figura 26.-Cáncer de vejiga urinaria, en hombres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, mujeres



PRP según Mixto, mujeres

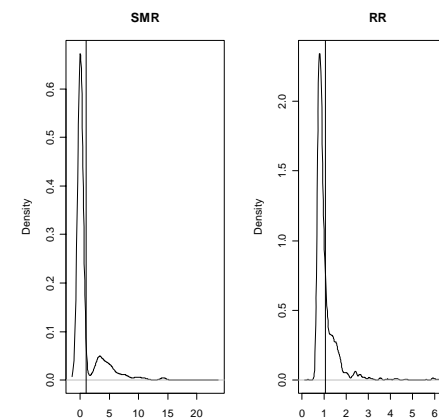
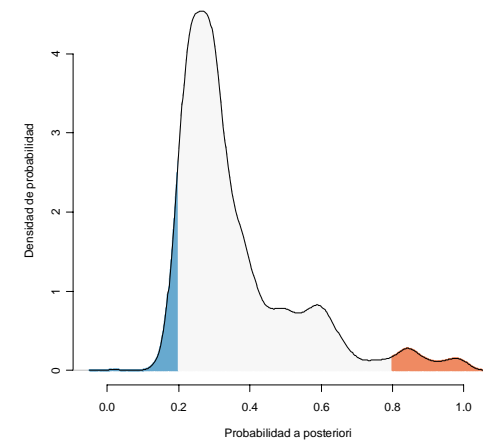
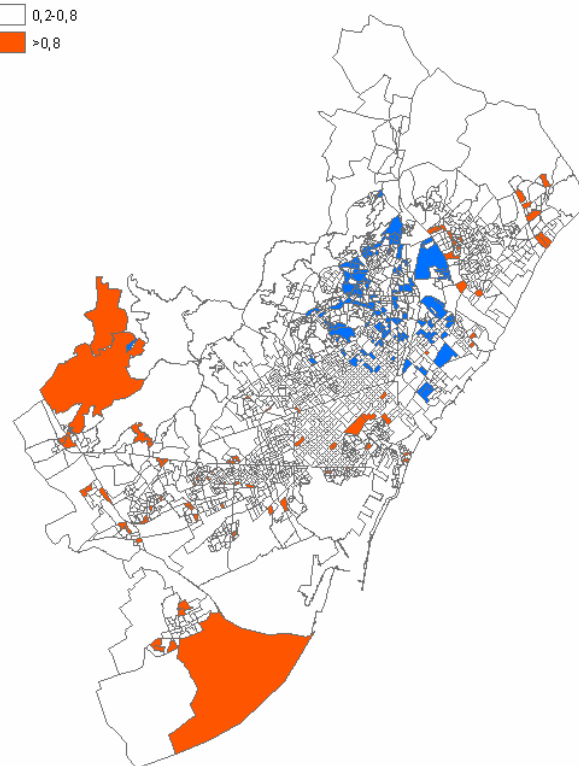
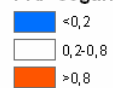
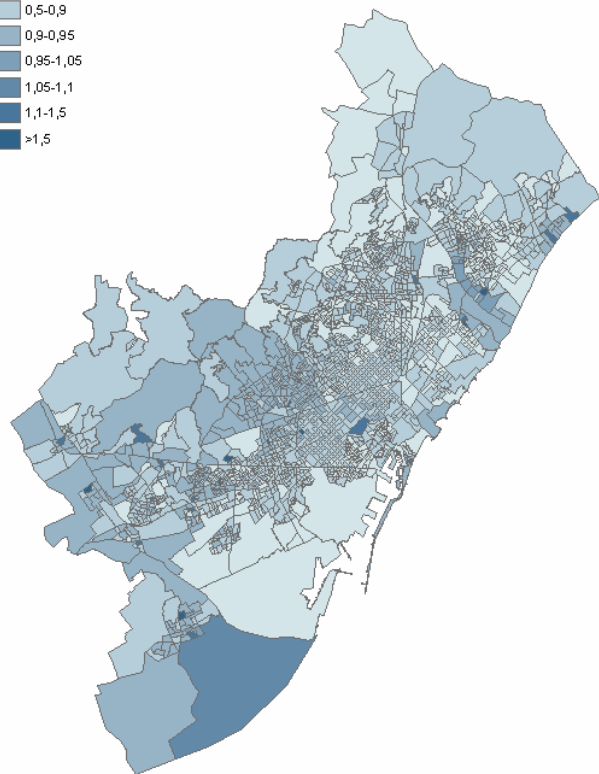
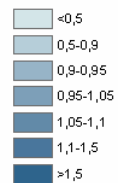


Figura 27.-Cáncer de vejiga urinaria, en mujeres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, hombres



PRP según Mixto, hombres

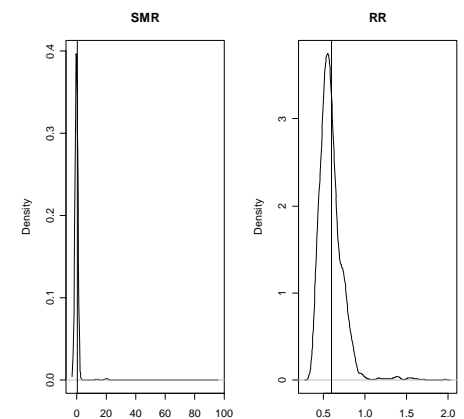
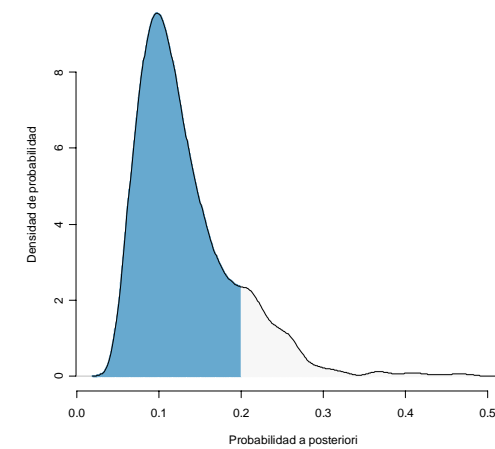
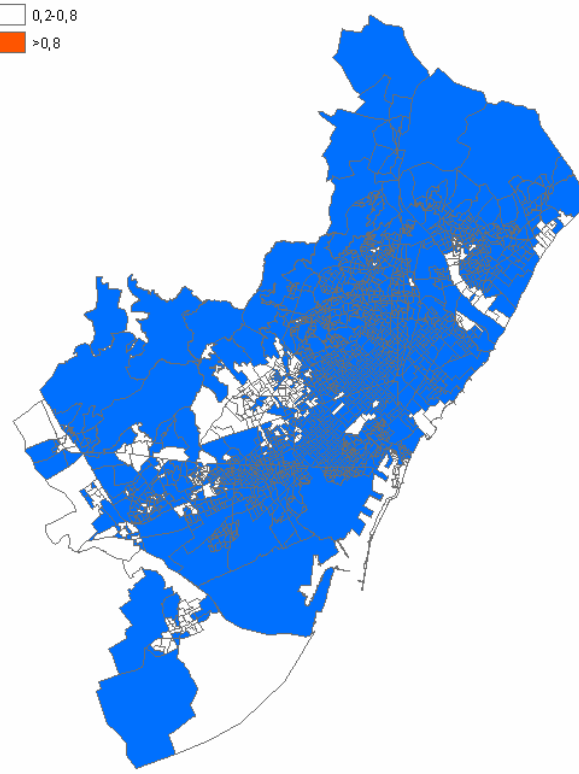
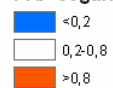
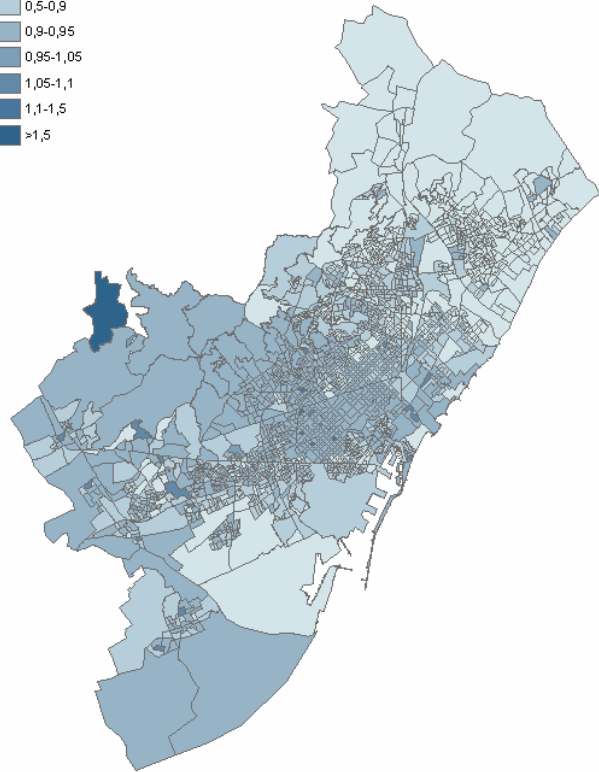
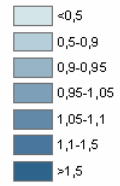


Figura 28.-Cáncer de tiroides, en hombres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, mujeres



PRP según Mixto, mujeres

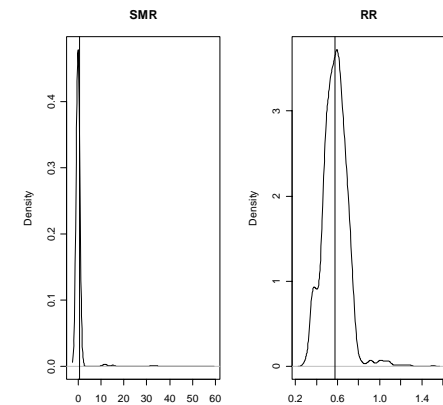
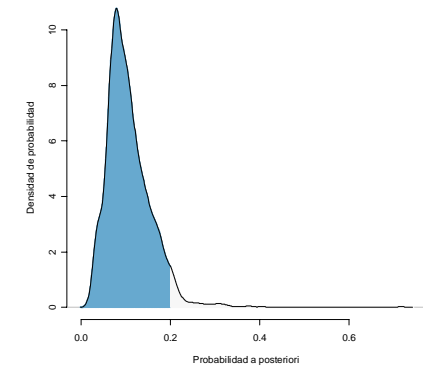
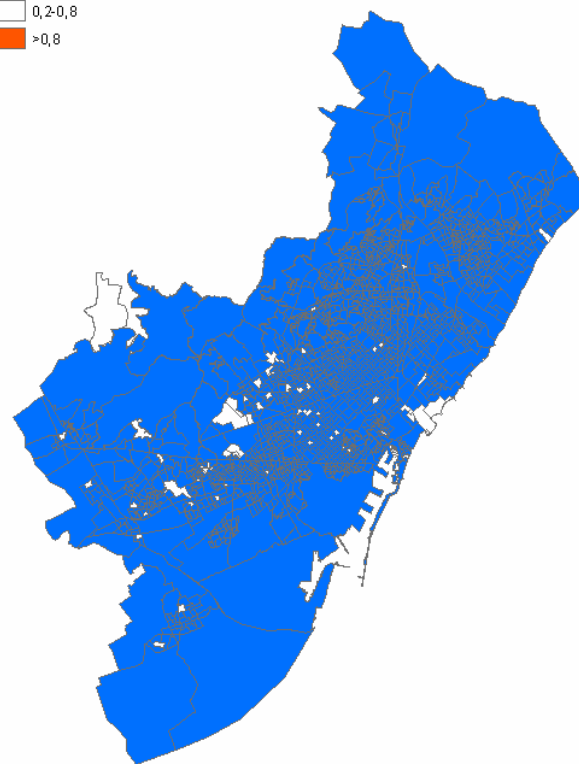
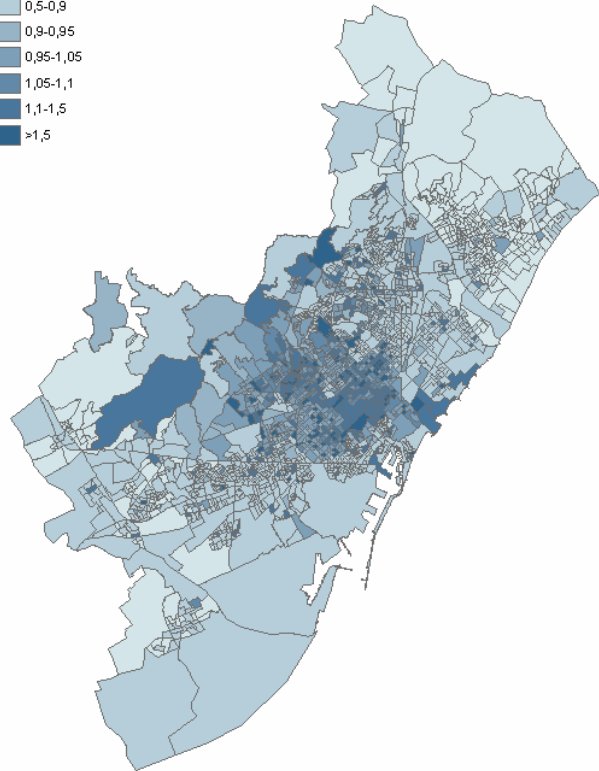
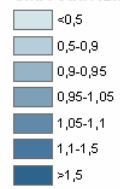


Figura 29.-Cáncer de tiroides, en mujeres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, hombres



PRP según Mixto, hombres

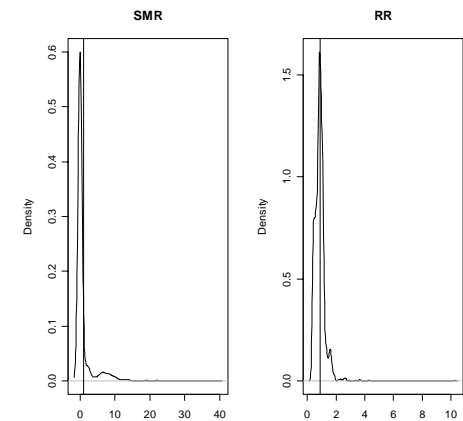
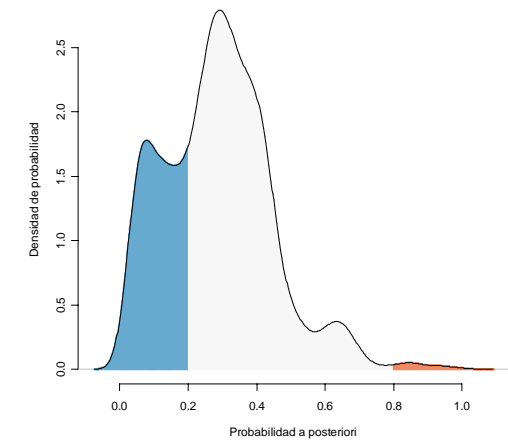
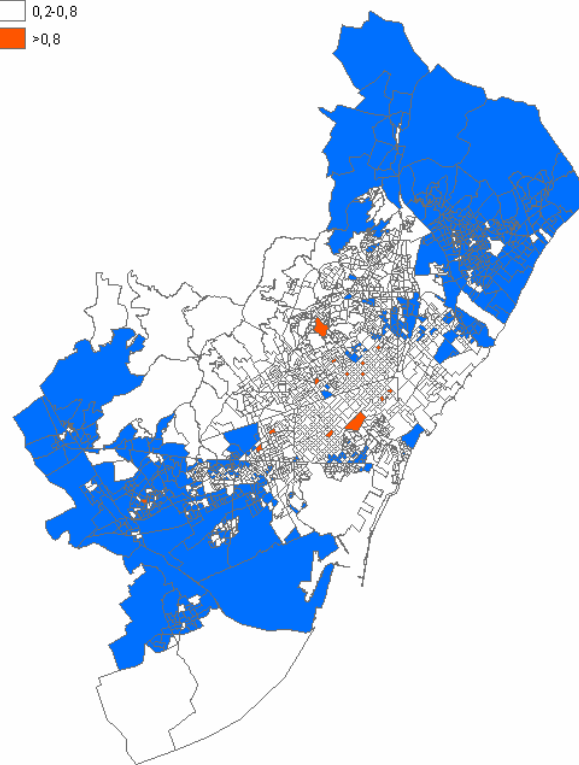
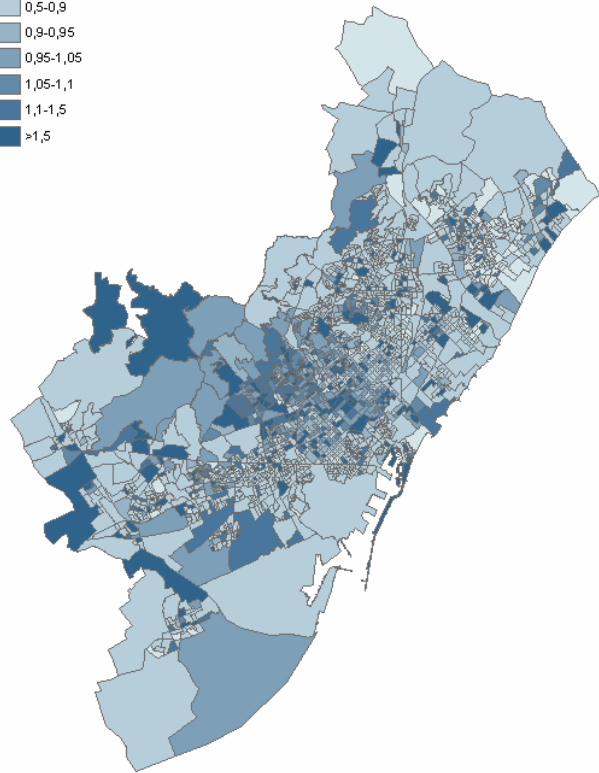
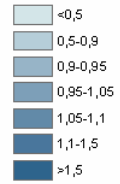


Figura 30.-Linfoma no Hodgkin, en hombres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, mujeres



PRP según Mixto, mujeres

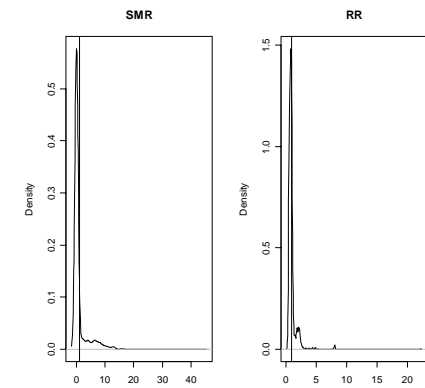
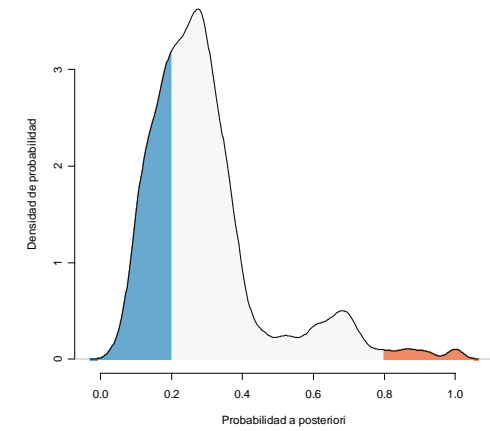
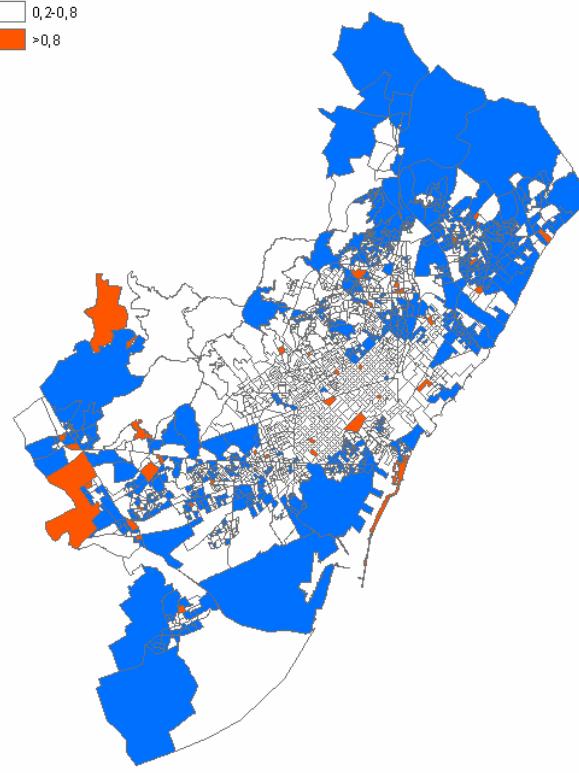
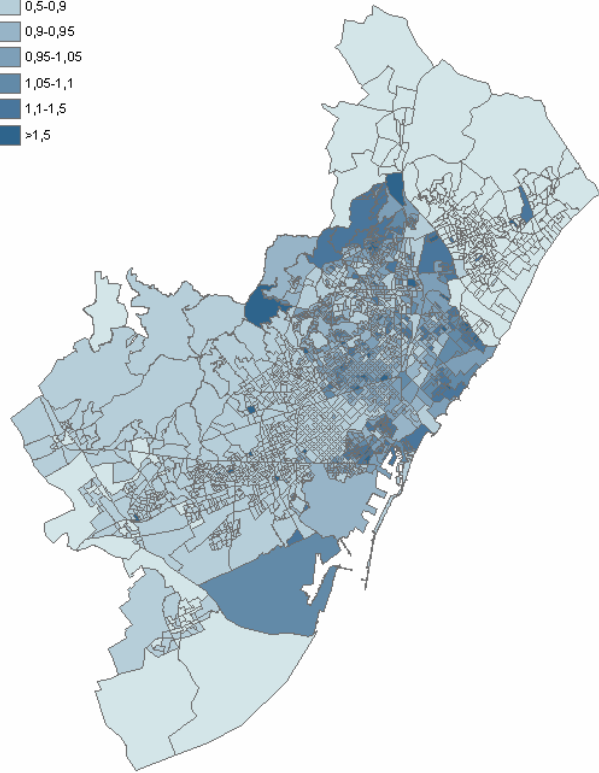
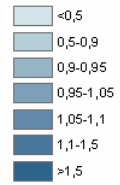


Figura 31.- Linfoma no Hodgkin, en mujeres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, hombres



PRP según Mixto, hombres

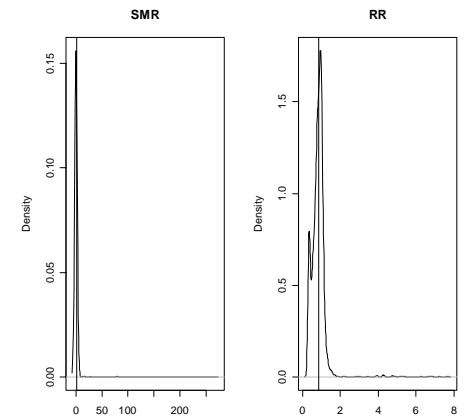
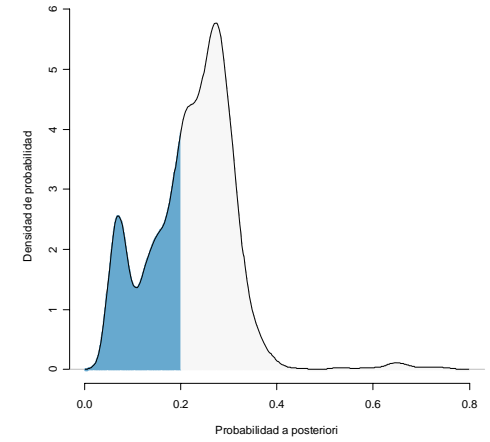
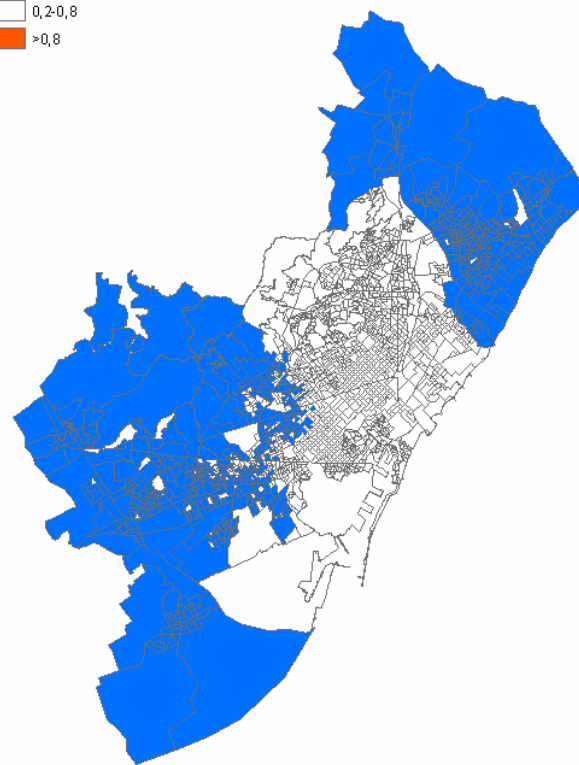
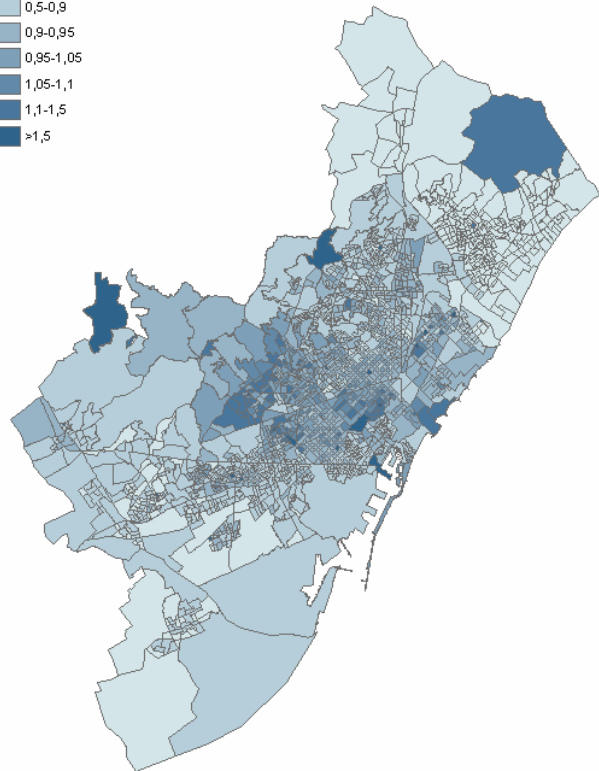
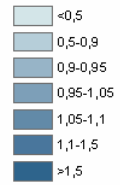


Figura 32.-Linfoma Hodgkin, en hombres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, mujeres



PRP según Mixto, mujeres

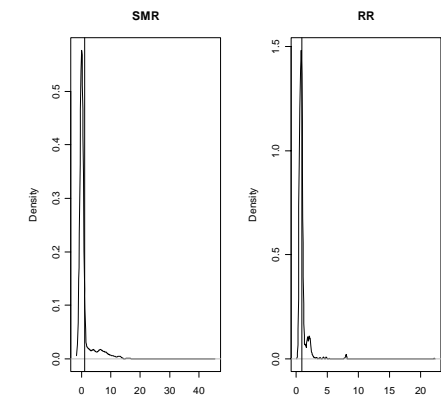
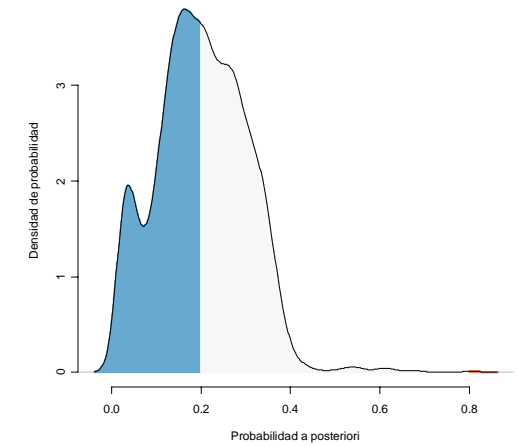
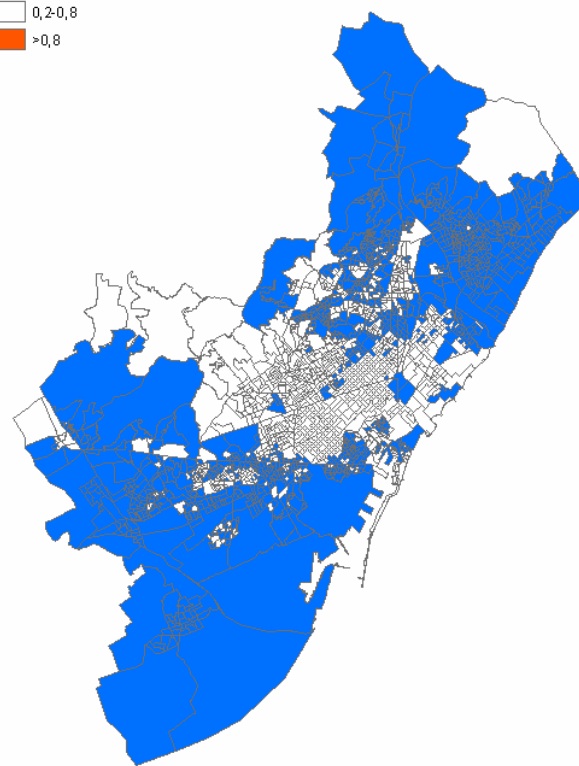
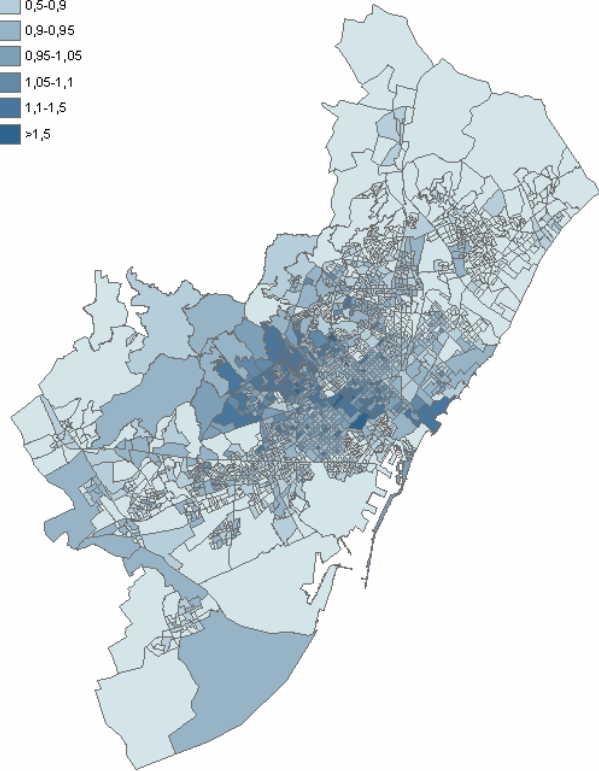
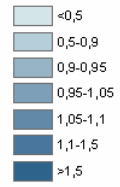


Figura 33.-Linfoma Hodgkin, en mujeres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, hombres



PRP según Mixto, hombres

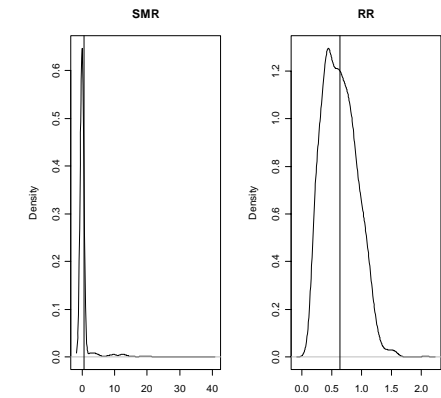
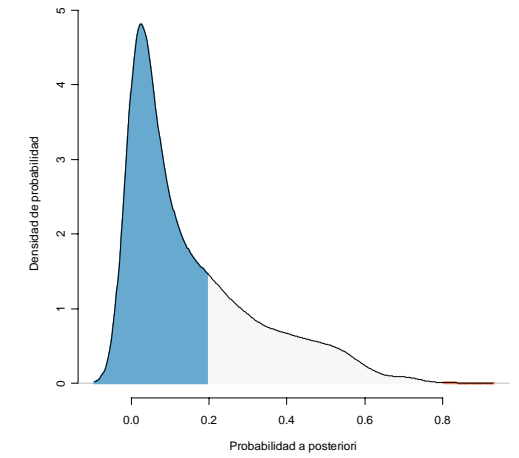
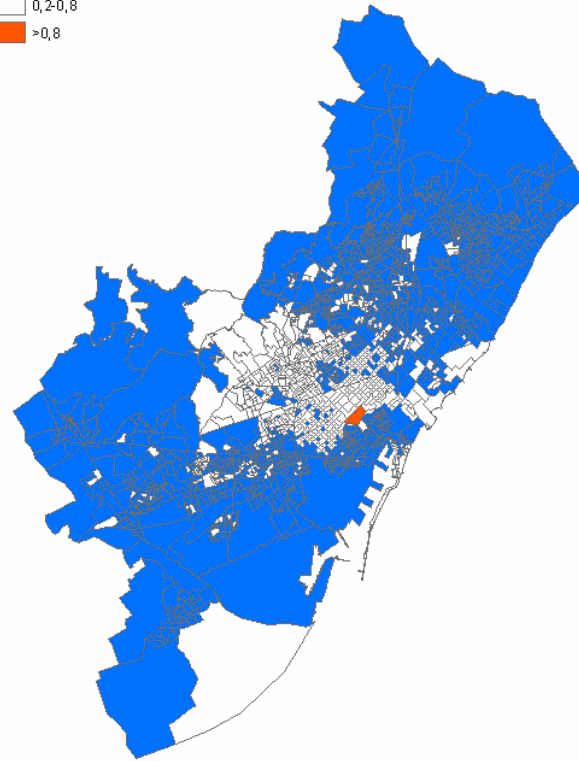
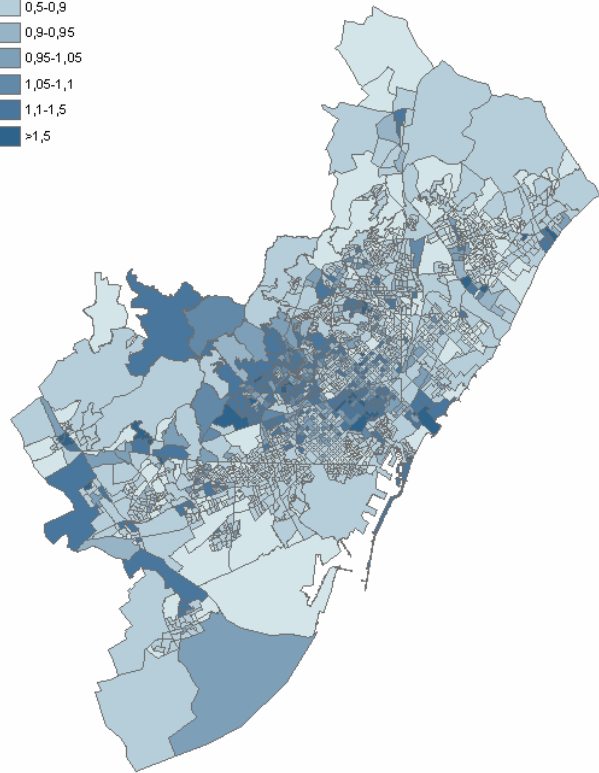
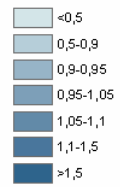


Figura 34.-Mieloma, en hombres. Modelo Mixto. Fuente: construcción propia.

SMR suavizado según Mixto, mujeres



PRP según Mixto, mujeres

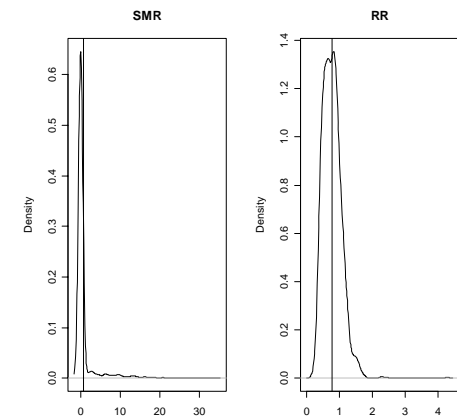
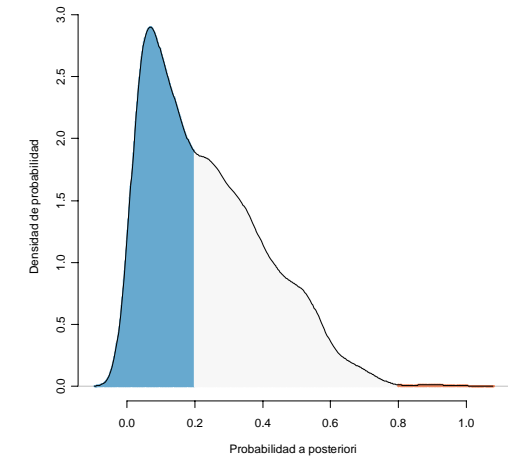
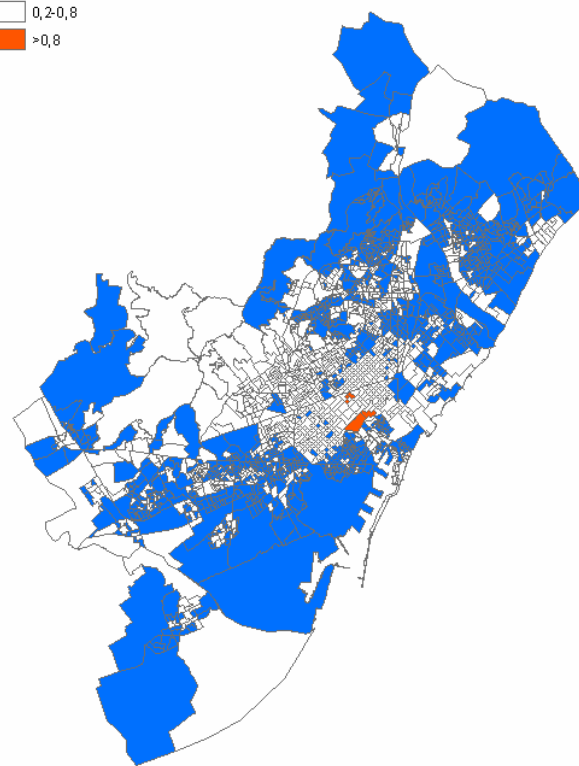


Figura 35.- Mieloma, en mujeres. Modelo Mixto. Fuente: construcción propia.

Las desviaciones típicas del efecto aleatorio espacial fueron mayores que las del efecto no espacial, salvo en el caso de linfoma Hodgkin, hombres, en el que ocurrió lo contrario (véase Tabla 5). La desviación típica del efecto espacial fue prácticamente el doble que las del efecto no espacial, con bastantes excepciones. En concreto, fue sólo algo mayor en las RME del cáncer de tráquea, bronquios y pulmón, hombres; linfoma no Hodgkin, mujeres; linfoma Hodgkin, mujeres; y mieloma, ambos sexos. En todos los casos las desviaciones típicas del salto (la discontinuidad) fueron bastante similares a las del otro efecto espacial.

Las RME suavizadas (así como PRP e histogramas de frecuencias) se muestran en las Figuras 24 al 35. Éstas parecen mostrar una clara conducta geográfica en cáncer de tráquea, bronquios y pulmón, ambos sexos, y en vejiga urinaria, ambos sexos. En el primer caso, podría existir un exceso de riesgo en el barrio de Ciutat Vella, Barcelona; en el segundo, riesgos elevados significativos se encontraron en el Prat y en Sant Just–Sant Feliu. Linfoma no Hodgkin muestra secciones con exceso de riesgo pero no pareció existir ningún patrón geográfico claro. El resto de neoplasias no mostró ningún patrón geográfico.

b- Modelos de asociación geográfica

En este caso se optó por introducir el índice de privación (un único factor) juntamente con el contaminante, bien NO_2 , bien SO_2 , en el mejor modelo en términos de ajuste. Como antes, el índice se introdujo como efecto fijo y el contaminante como efecto aleatorio. Recuérdese, que en el caso del contaminante, a fin de eliminar el sesgo ecológico se introdujo la media y la variabilidad intra-área.

Obsérvese (Tabla 6 y Figura 36) que el índice de privación resultó estadísticamente significativo en los casos de cáncer de tráquea, bronquios y pulmón, ambos sexos; linfoma no Hodgkin, ambos sexos; y mieloma, ambos sexos. Obsérvese que, salvo en mieloma, el efecto del índice sobre las RME fue mayor en el caso de las mujeres, aunque las diferencias por sexo no resultaron nunca estadísticamente significativas.

El efecto del NO_2 resultó estadísticamente significativo en el cáncer de tráquea, bronquios y pulmón, mujeres; vejiga urinaria, mujeres; y linfoma no Hodgkin, hombres (Tabla 6 y Figura 37).

El efecto del SO_2 nunca resultó estadísticamente significativo presentando, además intervalos de confianza bastante considerables (de hecho la Figura 38 se ha representado en escala logarítmica para que se pueda interpretar).

Sólo en el cáncer de tráquea, bronquios y pulmón, en mujeres; y linfoma no Hodgkin, hombres; el índice y el NO_2 fueron ambos estadísticamente significativos.

Cáncer de:	Tráquea, bronquios y pulmón		Vejiga urinaria		Tiroides		Linfoma no Hodgkin		Linfoma Hodgkin		Mieloma	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Índice de privación	0,17943(0,02299)	0,20506(0,04470)	-0,03290(0,03393)	-0,05502(0,05469)	-0,22904(0,20039)	-0,14331(0,15483)	0,16798(0,06176)	0,23022(0,05823)	0,07386(0,20590)	-0,27711(0,23159)	0,44639(0,09615)	0,33665(0,07569)
NO2	0,04295(0,09306)	0,77522(0,11853)	0,07547(0,05754)	0,64498(0,16009)	-0,00069(0,59063)	0,25583(0,59938)	0,56820(0,10925)	0,36333(0,60043)	-0,64492(0,78307)	-0,40837(0,90091)	0,64328(0,38942)	0,20188(0,41979)
SO2	0,16003(0,15621)	1,56644(1,21934)	1,07811(0,94551)	2,05144(1,87590)	5,94106(4,73466)	2,67263(1,92163)	-0,57981(1,17627)	0,90096(0,68574)	-0,54878(1,73057)	-2,36747(2,27075)	1,68377(0,96162)	1,56620(1,12346)
Efectos aleatorios												
Espacial	0,48325(0,05947)	0,44703(0,11104)	0,63085(0,08737)	0,42141(0,15886)	0,38931(0,39473)	0,42565(0,30394)	0,31472(0,13803)	0,11761(0,12617)	0,66255(0,49216)	-2,36747(2,27075)	0,42668(0,17940)	0,29148(0,18458)
No Espacial	0,13417(0,09229)	0,31766(0,14977)	0,15379(0,10056)	0,33015(0,28473)	0,54846(0,35954)	0,35164(0,29941)	0,19757(0,15959)	0,40218(0,24341)	0,87089(0,47637)	0,83500(0,63253)	0,25311(0,15933)	0,29631(0,16578)
Salto		0,41711(0,11146)	0,65421(0,09835)	0,42166(0,24218)	0,37385(0,15223)	0,62112(0,40011)	0,40578(0,11667)	0,15862(0,13178)	0,76655(0,41654)	0,53469(0,23124)	0,53411(0,36665)	0,31977(0,11632)
AIC	4828,485	2494,029	4487,949	3000,396	542,329	912,718	3108,396	3101,079	941,032	1041,229	1592,265	2056,972
BIC	12600,610	5525,122	11248,083	9244,006	1536,915	2611,99	8713,177	8106,727	3038,164	3499,812	3045,955	5386,409
DIC	3125,779	2118,123	2688,564	1602,822	238,945	377,161	1390,627	1409,175	269,141	239,525	826,925	1013,979

Tabla 6.– Resultados de la estimación en los modelos de asociación geográfica. Variables explicativas: Índice socioeconómico, NO₂ y SO₂. Fuente: construcción propia.

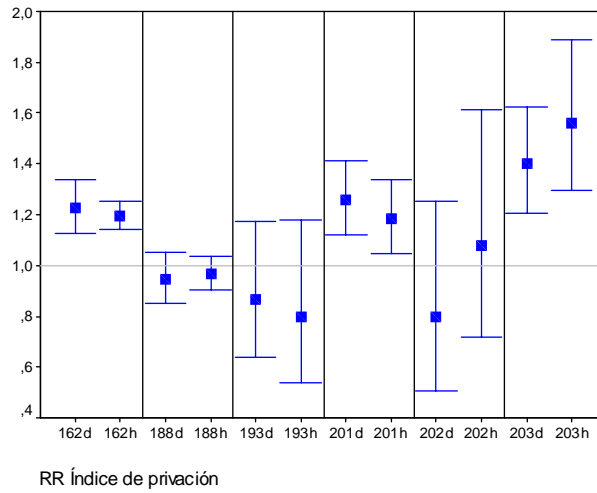


Figura 36.–Riesgo relativo asociado al índice de privación. Fuente: construcción propia.

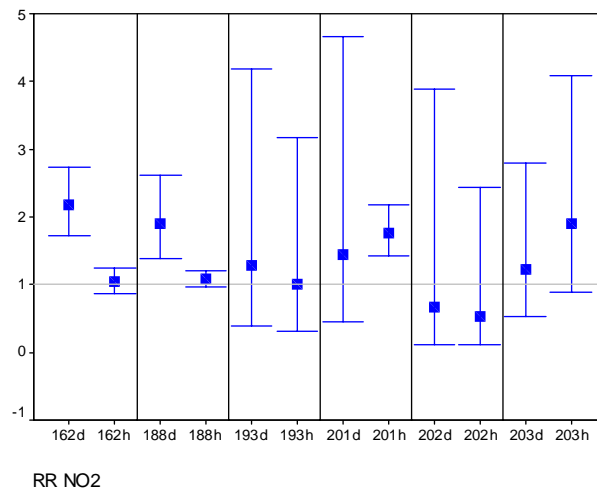
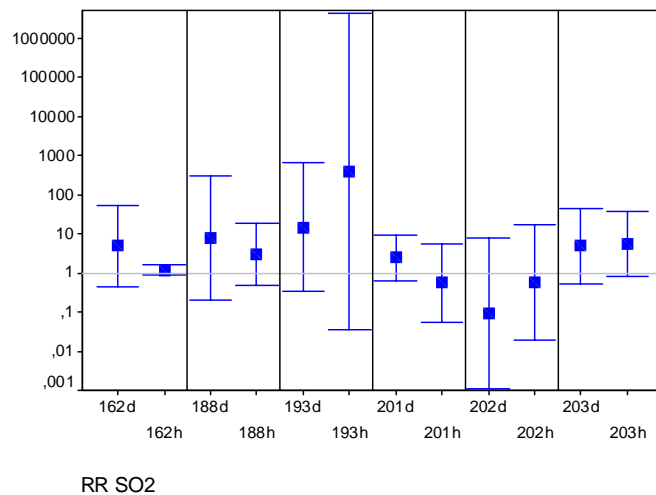


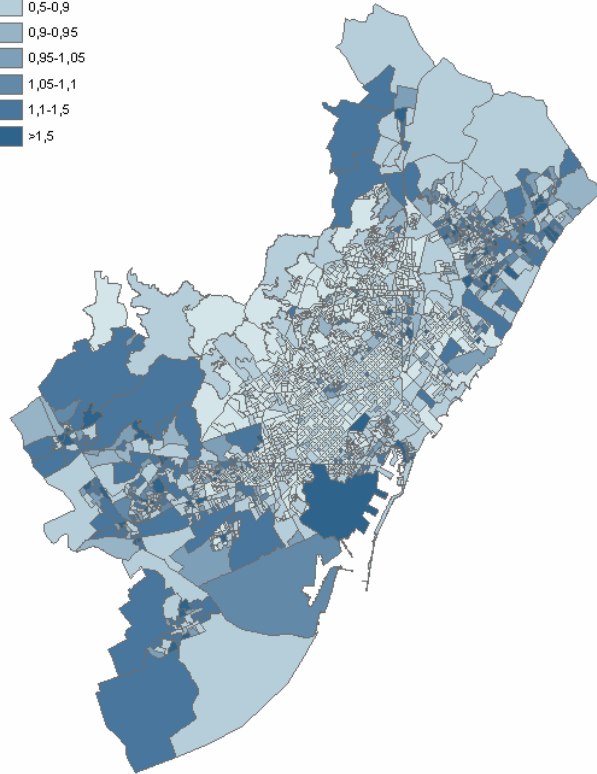
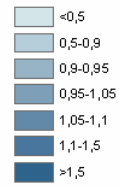
Figura 37.–Riesgo relativo asociado al NO₂. Fuente: construcción propia.



Escala logarítmica

Figura 38.–Riesgo relativo asociado al SO₂. Fuente: construcción propia.

SMR suavizada según BYM, hombres



PRP según BYM, hombres

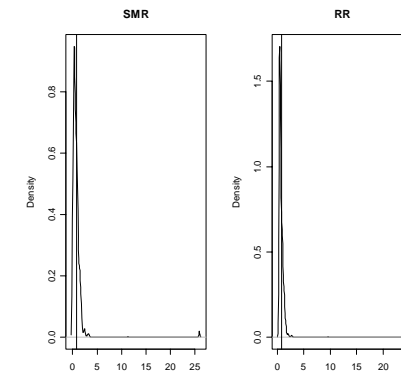
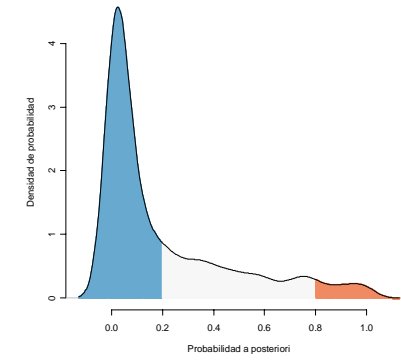
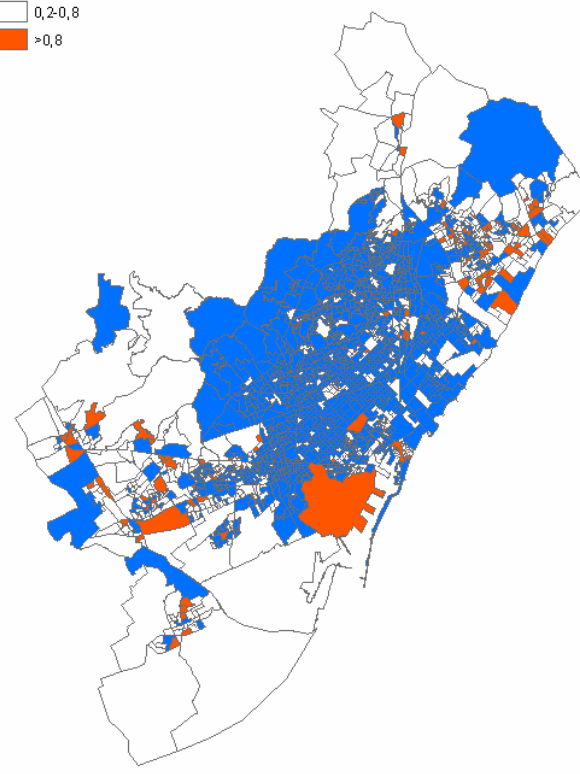
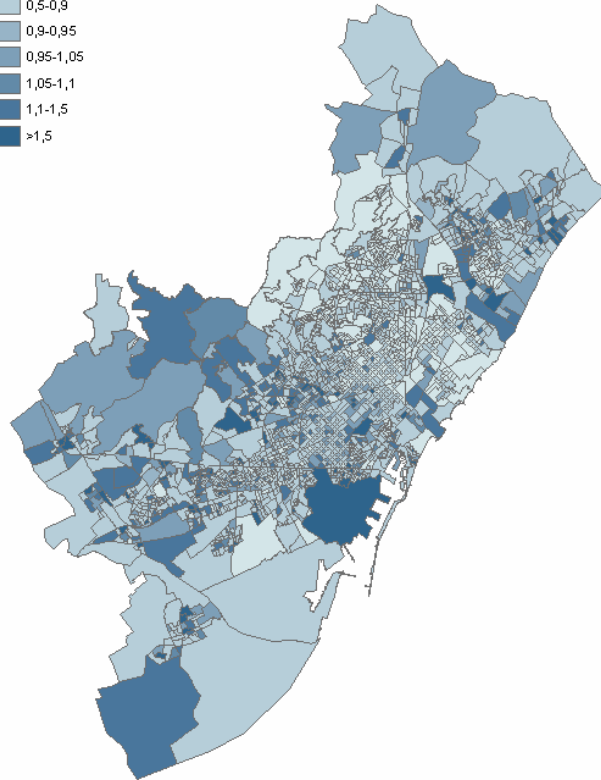
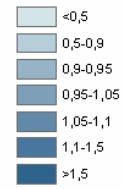


Figura 39.- Cáncer de tráquea, bronquios y pulmón introduciendo variables explicativas, modelo BYM, hombres. Fuente: construcción propia.

SMR suavizada según Mixto, mujeres



PRP según Mixto, mujeres

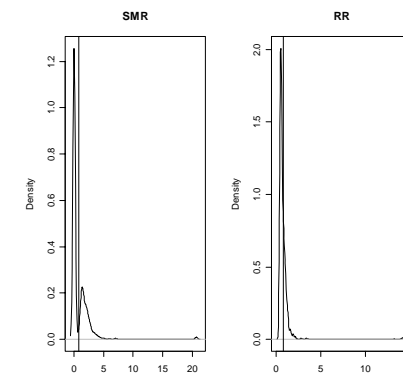
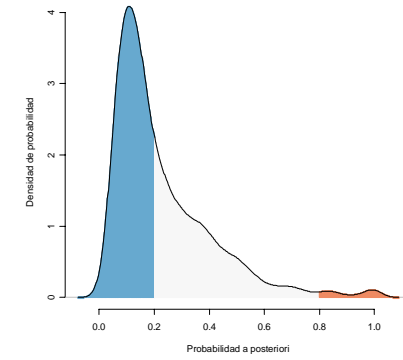
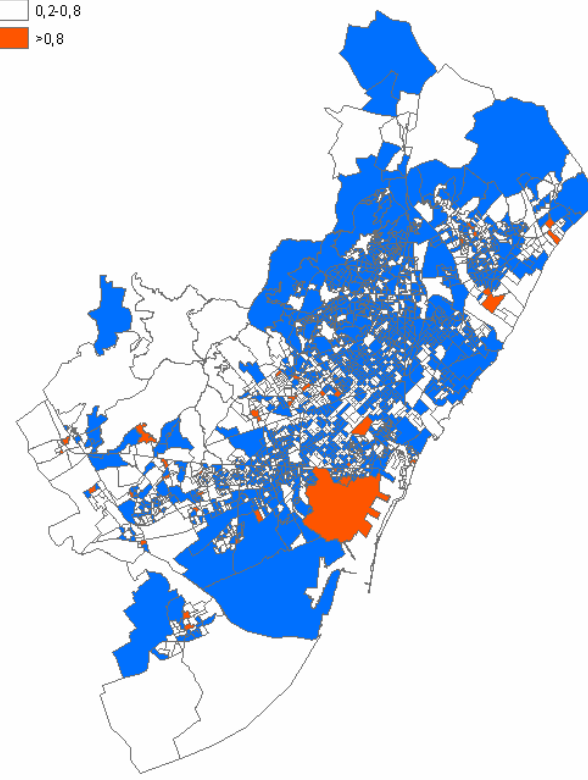
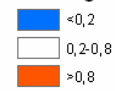
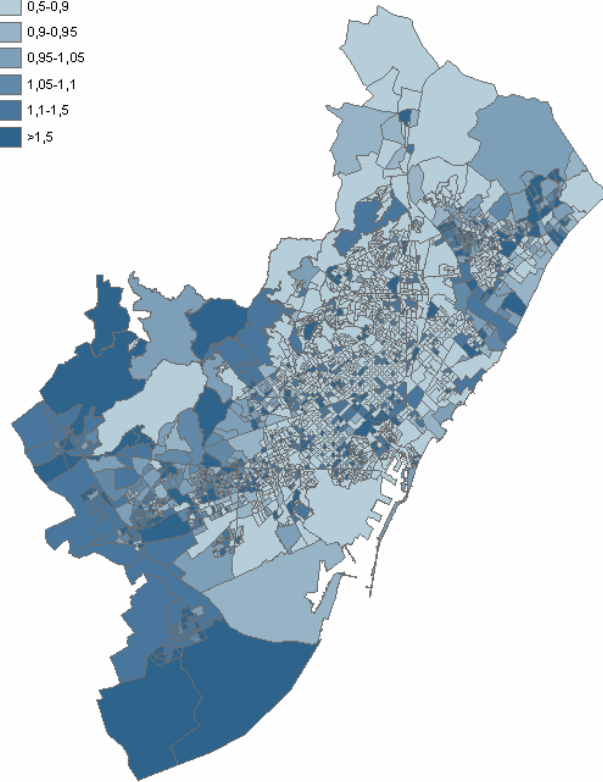
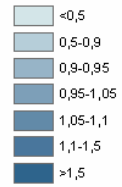


Figura 40.- Cáncer de tráquea, bronquios y pulmón, introduciendo variable explicativa, modelo mixto, mujeres. Fuente: construcción propia.

SMR suavizada según Mixto, mujeres



PRP según Mixto, mujeres

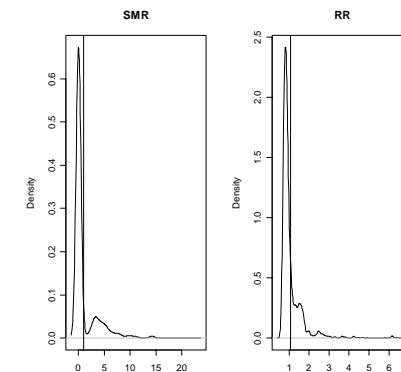
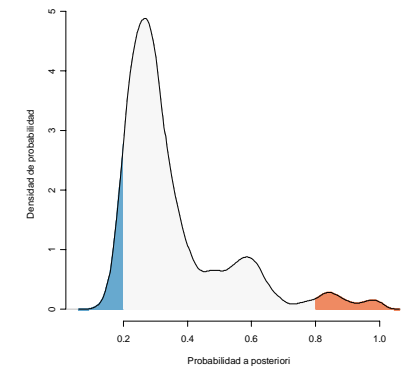
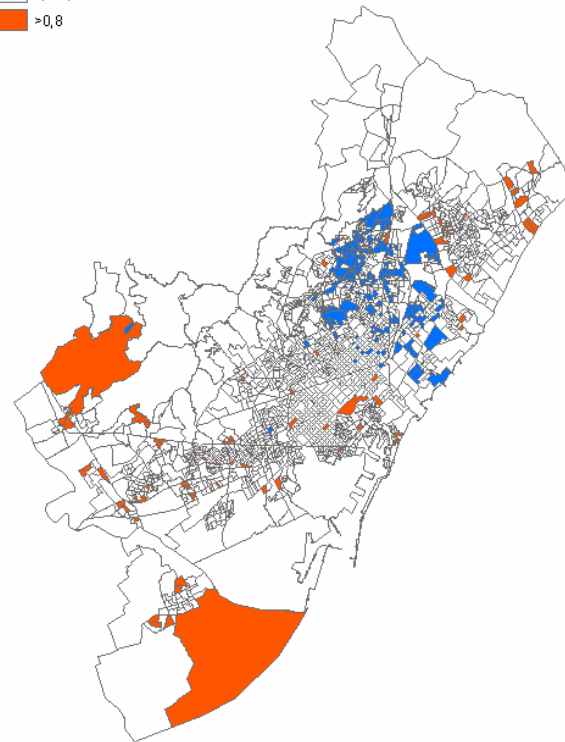
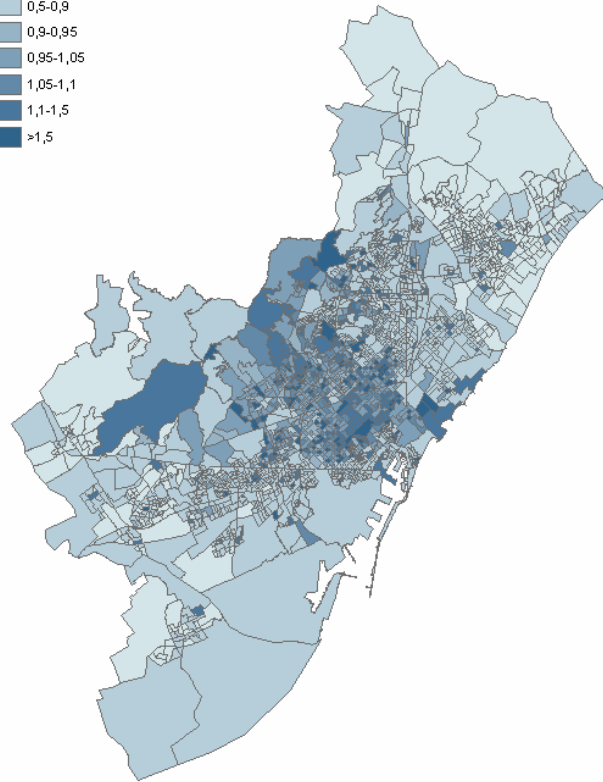
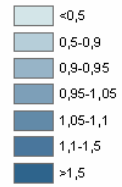


Figura 41.-Cáncer de vejiga urinaria, introduciendo variable explicativa, modelo mixto, mujeres. Fuente: construcción propia.

SMR suavizada según Mixto, hombres



PRP según Mixto, hombres

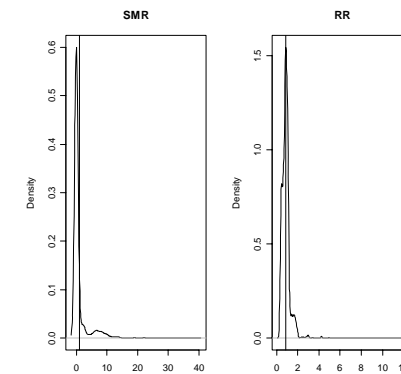
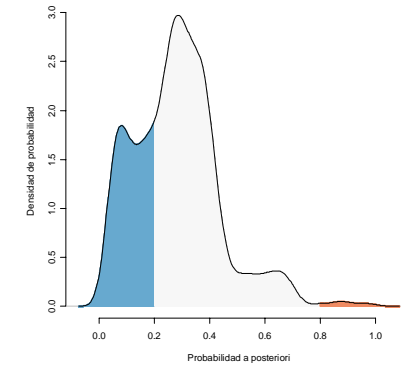
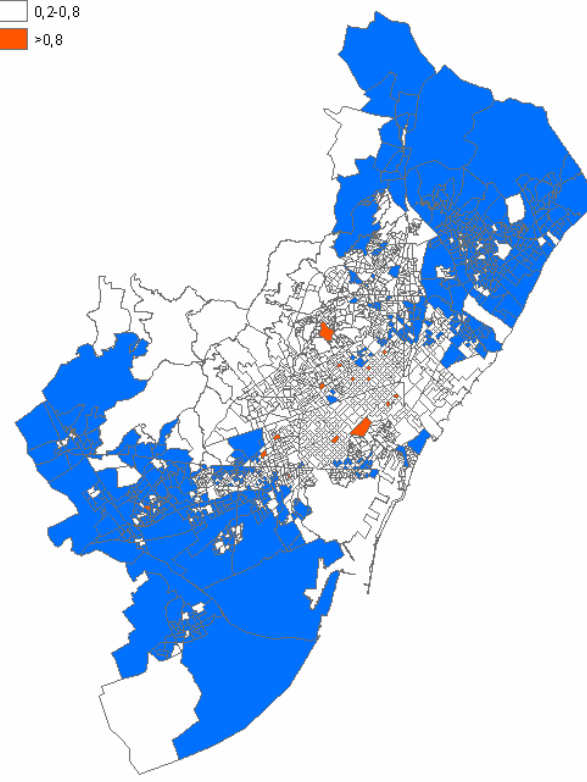
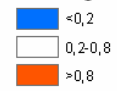
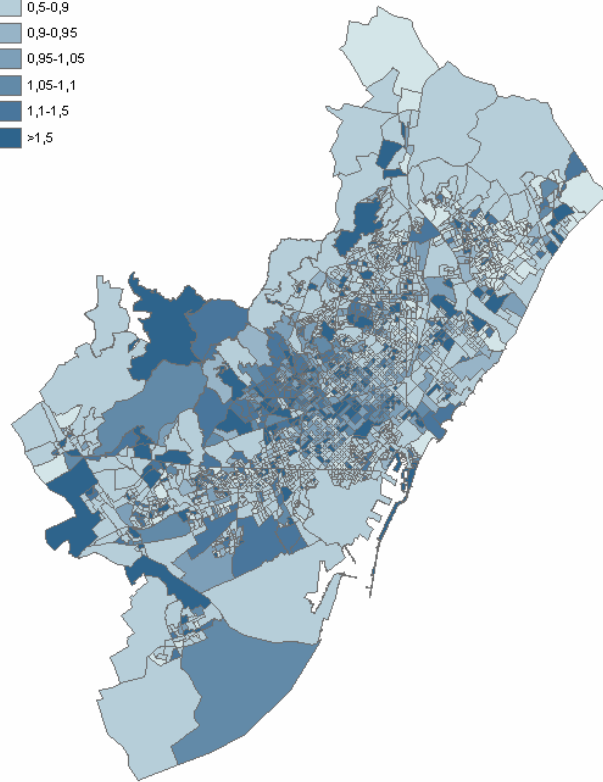
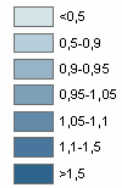


Figura 42.-Linfoma no Hodgkin, introduciendo variable explicativa, modelo mixto, hombres. Fuente: construcción propia.

SMR suavizada según Mixto, mujeres



PRP según Mixto, mujeres

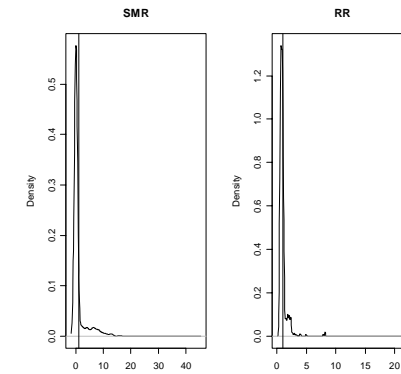
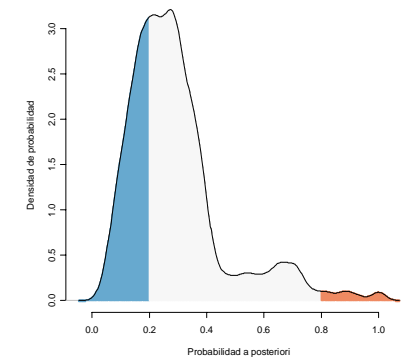
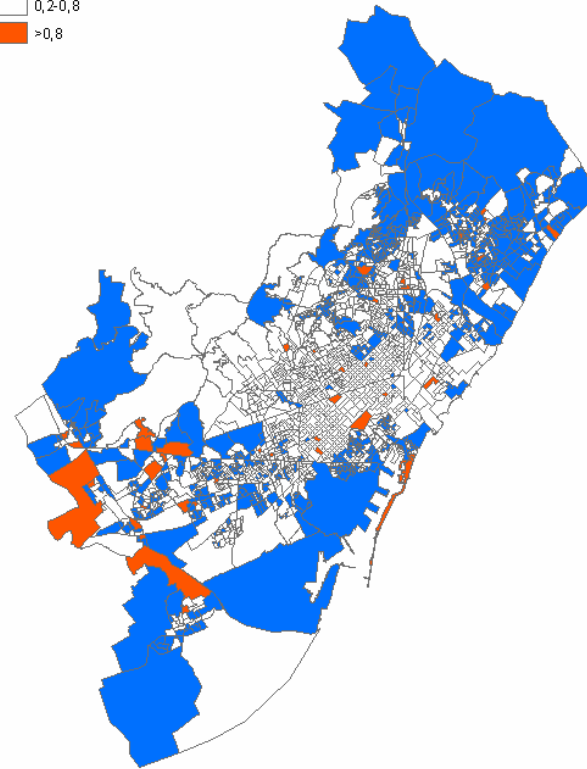
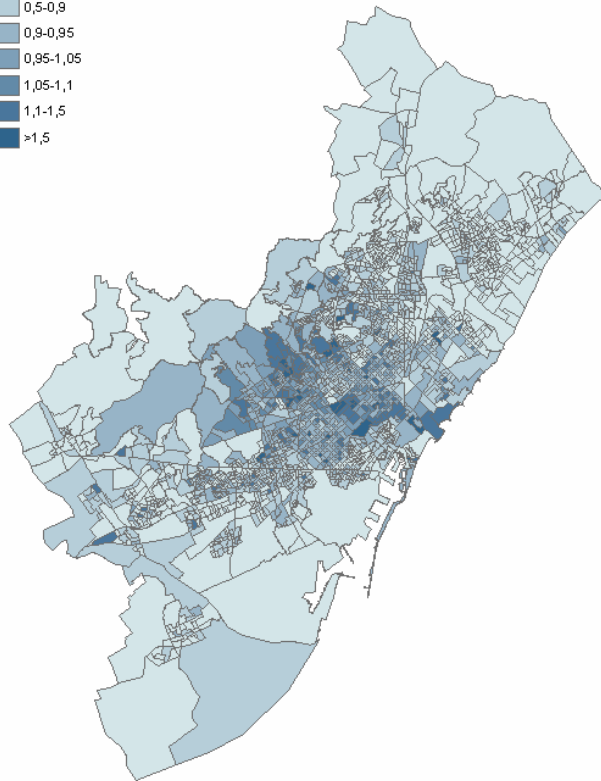
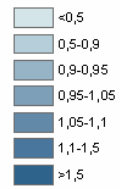


Figura 43.-Linfoma no Hodgkin, introduciendo variable explicativa, modelo mixto, mujeres. Fuente: construcción propia.

SMR suavizada según Mixto, hombres



PRP según Mixto, hombres

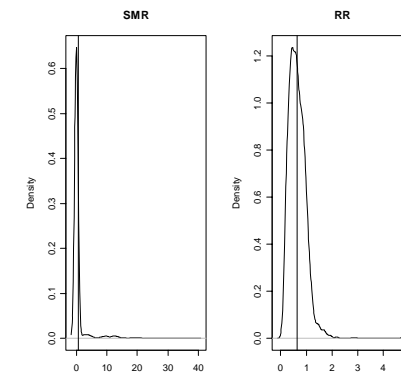
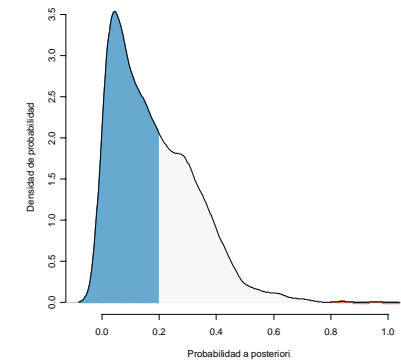
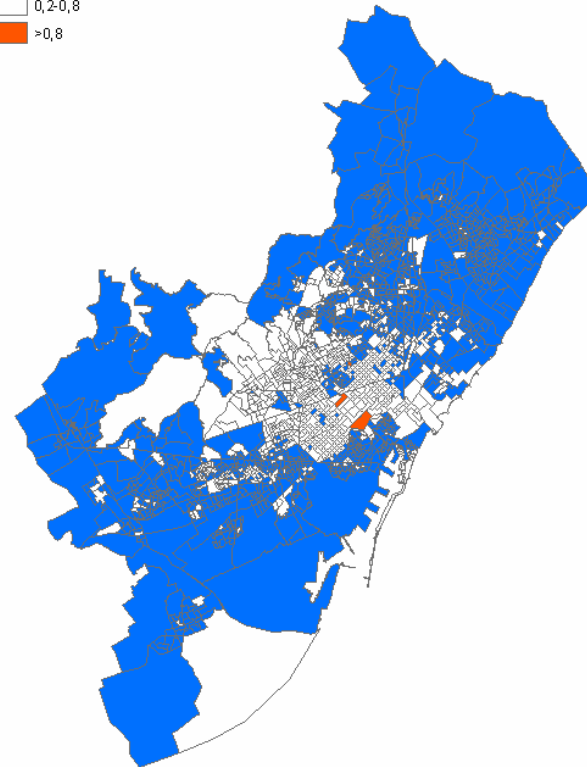
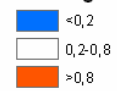
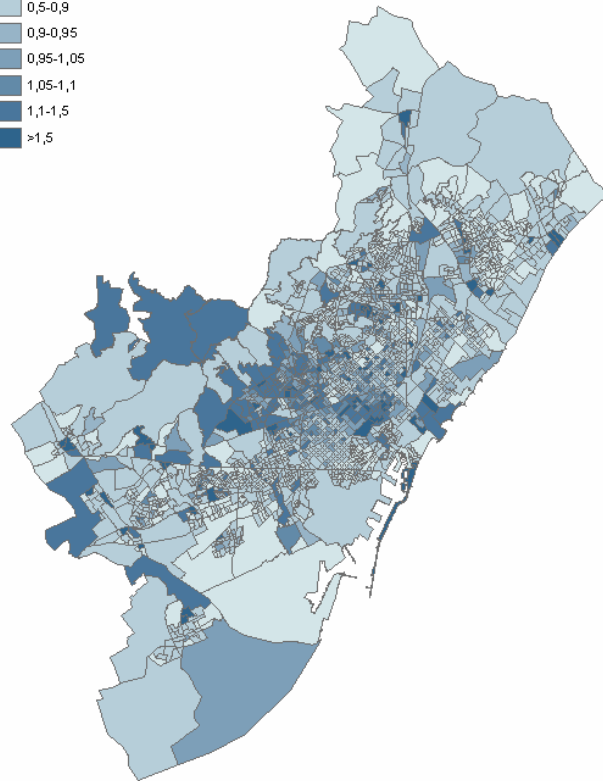
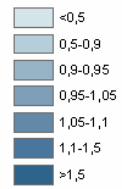


Figura 44.-Mieloma, introduciendo variable explicativa, modelo mixto, hombres. Fuente: construcción propia.

SMR suavizada según Mixto, mujeres



PRP según Mixto, mujeres

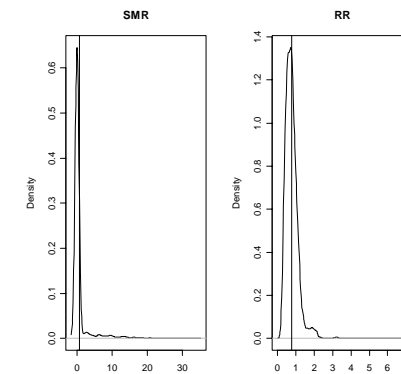
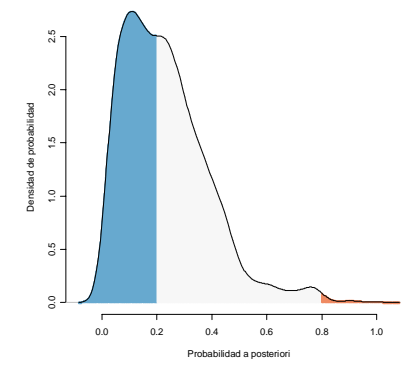
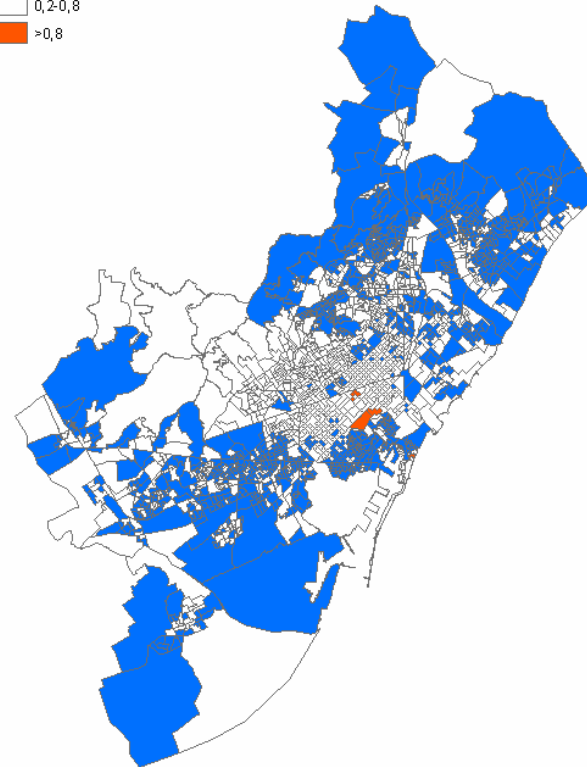


Figura 45.-Mieloma, introduciendo variable explicativa, modelo mixto, mujeres. Fuente: construcción propia.

La introducción de las variables explicativas, cuando resultaron estadísticamente significativas, redujo la desviación típica del efecto aleatorio espacial, aunque no se observó ninguna conducta sistemática (Tabla 6).

De hecho en las Figuras 39 a 45 se observa, como antes, que la introducción de las variables explicativas, incluso cuando éstas resultaron estadísticamente significativas, apenas redujo la extraviabilidad explicada.

4.1.5.- Resultados de la estimación. Área Metropolitana de Bilbo

a.-Mortalidad por las neoplasias de interés medioambiental en el Área Metropolitana de Bilbo

Como se encontró previamente, con la excepción del cáncer tráquea, bronquios y pulmón en hombres, el mejor ajuste lo proporcionó el modelo mixto de riesgo relativo (Tabla 7 y Figura 46). De nuevo nótese el parecido entre la Figura 46 y las Figuras 1 y 23.

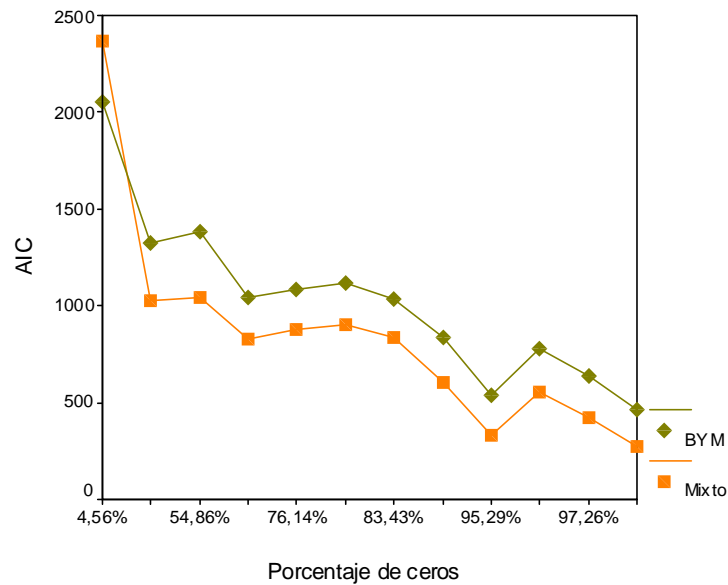


Figura 46.- Relación entre porcentaje de ceros y el estadístico AIC. Fuente: construcción propia.

Las desviaciones típicas del efecto aleatorio espacial fueron mayores que las del efecto no espacial en todos los casos (véase Tabla 7). En general, estas diferencias fueron sustanciales, con la excepción de mieloma, ambos sexos; linfoma Hodgkin, mujeres; y vejiga urinaria, mujeres; en los que la diferencia fue muy reducida. Las desviaciones típicas del salto (discontinuidad) fueron más grandes que las del otro efecto espacial.

CIE-9: 162, Tráquea, bronquios y pulm		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,25198(0,06213)	0,07231(0,05710)	0,14775(0,07112)	
No Espacial	0,17431(0,05630)	0,18540(0,04759)		
AIC	2051,969	2367,73		
BIC	4998,191	5362,68		
DIC	860,391	840,26		
% ceros	4,56%			
Mujeres				
Espacial	0,43748(0,20041)	0,28584(0,17031)	0,38311(0,14866)	
No Espacial	0,15699(0,10783)	0,16338(0,10004)		
AIC	1378,849	1039,88		
BIC	2361,883	1907,99		
DIC	696,487	685,73		
% ceros	54,86%			
CIE-9: 188, Vejiga urinaria		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,16105(0,11112)	0,14417(0,10205)	0,17431(0,07653)	
No Espacial	0,14383(0,10062)	0,128807(0,09690)		
AIC	1323,281	1027,90		
BIC	2016,073	1713,84		
DIC	751,343	745,24		
% ceros	45,14%			
Mujeres				
Espacial	0,29113(0,19838)	0,31635(0,20072)	0,30613(0,13112)	
No Espacial	0,35481(0,21242)	0,30274(0,20226)		
AIC	837,936	602,91		
BIC	1246,795	1125,34		
DIC	390,777	391,88		
% ceros	85,87%			
CIE-9: 193, Tiroides		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,85873(0,74567)	0,99899(0,68073)	0,78915(0,47886)	
No Espacial	0,92225(0,61014)	1,10425(0,59379)		
AIC	635,541	425,79		
BIC	1162,498	1141,35		
DIC	142,037	139,56		
% ceros	97,26%			
Mujeres				
Espacial	0,56108(0,38159)	0,49926(0,32432)	0,29212(0,14162)	
No Espacial	0,47980(0,35518)	0,50391(0,36206)		
AIC	537,673	331,64		
BIC	701,659	678,70		
DIC	196,427	195,45		
% ceros	95,29%			

CIE-9: 210, Linfoma no Hodgkin		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,59919(0,42817)	0,61354(0,42292)	0,73512(0,29861)	
No Espacial	0,88712(0,53772)	0,93990(0,54249)		
AIC	780,146	552,67		
BIC	1557,333	1466,20		
DIC	180,667	179,04		
% ceros	96,35%			
Mujeres				
Espacial	1,19566(0,88827)	1,37895(1,09163)	1,08412(0,83517)	
No Espacial	1,15695(0,82707)	1,03644(0,90468)		
AIC	459,73	273,31		
BIC	701,712	747,64		
DIC	85,179	85,34		
% ceros	98,63%			
CIE-9: 211, Linfoma Hodgkin		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,14580(0,11370)	0,16141(0,11407)	0,24232(0,14123)	
No Espacial	0,23912(0,15095)	0,24269(0,14605)		
AIC	1045,134	823,90		
BIC	1506,305	1436,88		
DIC	572,199	573,55		
% ceros	72,04%			
Mujeres				
Espacial	0,46017(0,27921)	0,47979(0,28017)	0,59128(0,18112)	
No Espacial	0,24126(0,15662)	0,23489(0,14698)		
AIC	1082,663	876,91		
BIC	1777,207	1759,57		
DIC	517,958	516,98		
% ceros	76,14%			
CIE-9: 212, Mieloma		DESVIACIONES TÍPICAS		
Hombres				
Efectos aleatorios	BYM	LAWSON	Saltos	
Espacial	0,31582(0,26633)	0,40278(0,26399)	0,44714(0,23115)	
No Espacial	0,34959(0,22893)	0,38063(0,20129)		
AIC	1035,575	833,15		
BIC	1783,002	1781,42		
DIC	444,209	445,25		
% ceros	83,43%			
Mujeres				
Espacial	0,27503(0,20975)	0,21405(0,19005)	0,19116(0,09577)	
No Espacial	0,33133(0,22317)	0,37814(0,21379)		
AIC	1120,615	904,45		
BIC	2005,472	1948,85		
DIC	469,378	468,74		
% ceros	81,91%			

Tabla 7.- Resultados de la estimación para las neoplasias de interés medioambiental. Fuente: construcción propia.

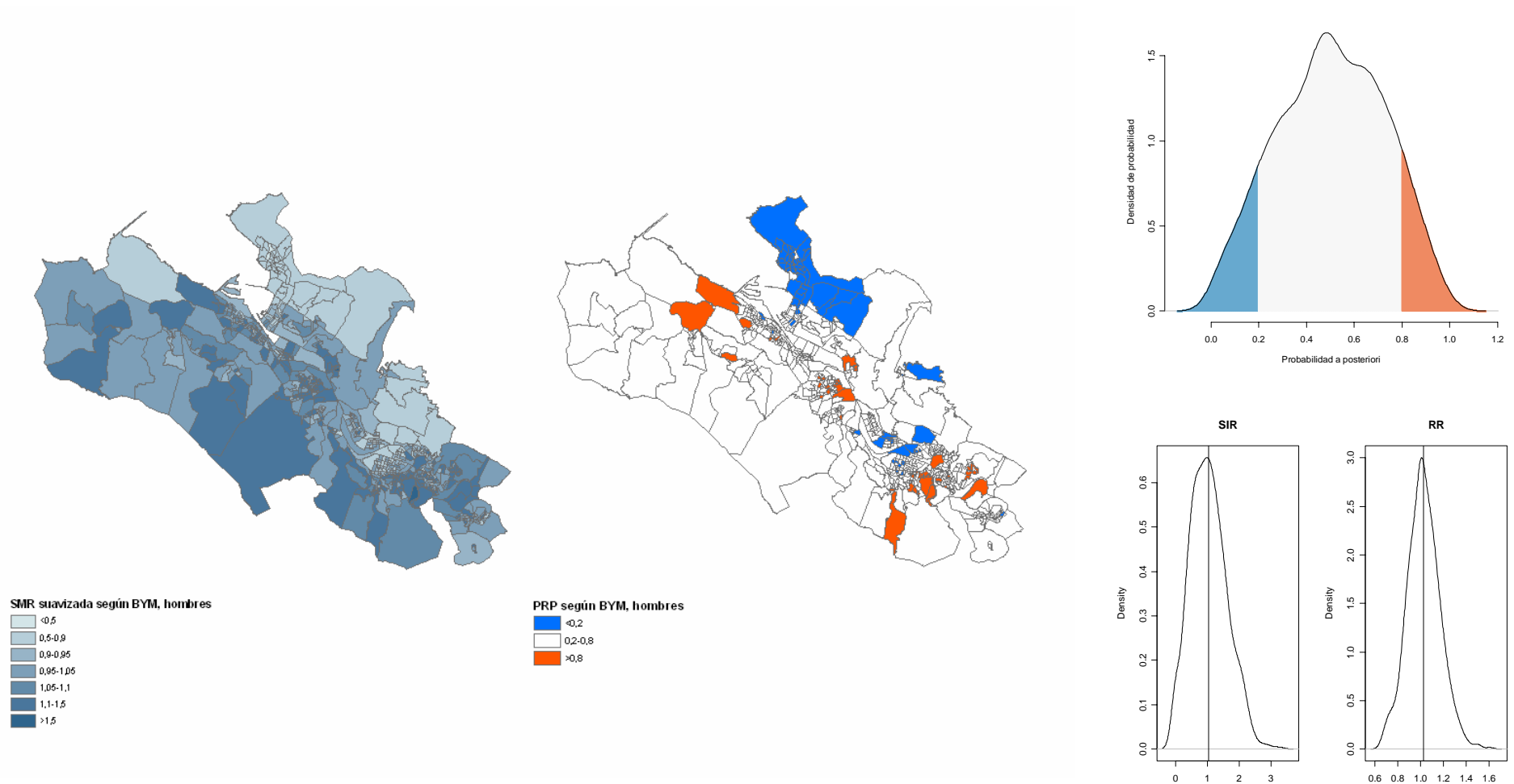


Figura 47.-Cáncer de tráquea, bronquios y pulmón, en hombres, modelo BYM. Fuente: construcción propia.

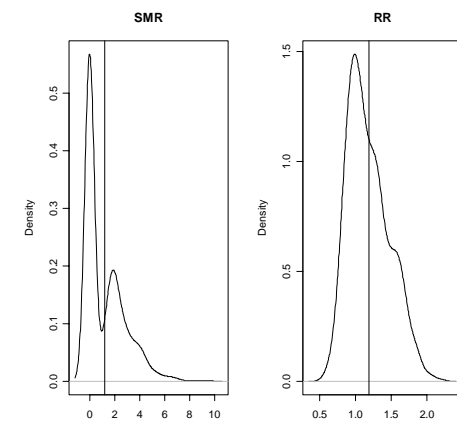
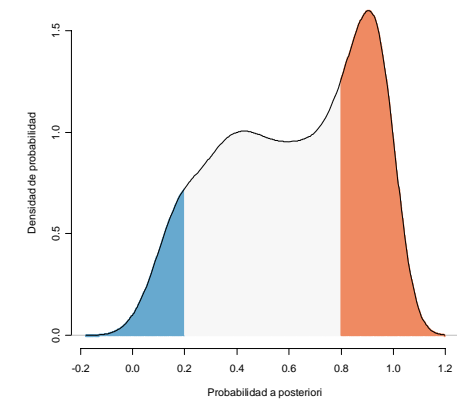
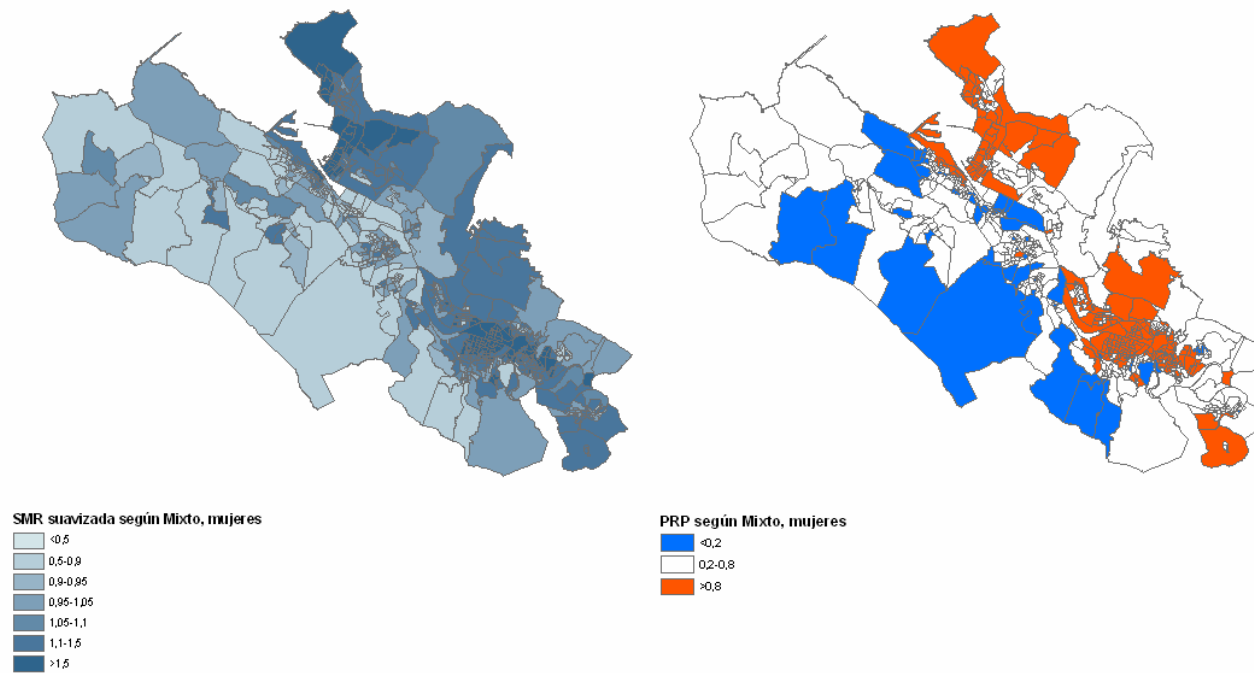


Figura 48.–Cáncer de tráquea, bronquios y pulmón, en mujeres, modelo Mixto. Fuente: construcción propia.

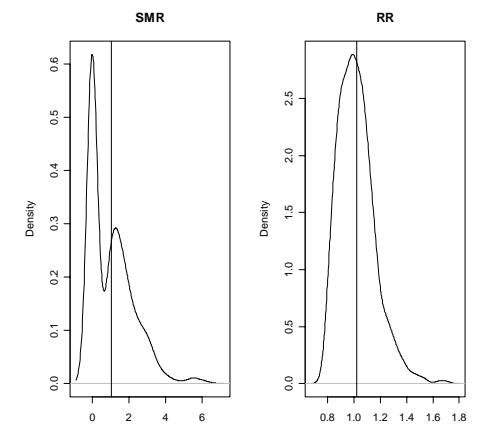
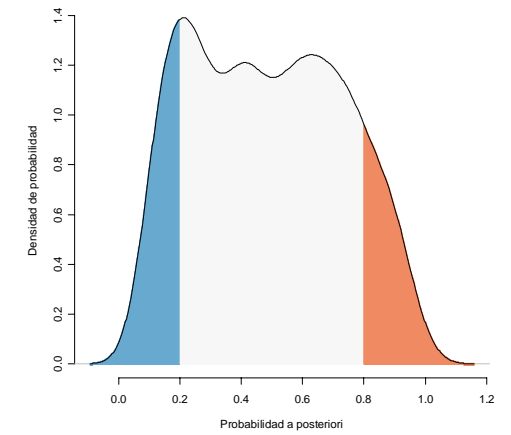
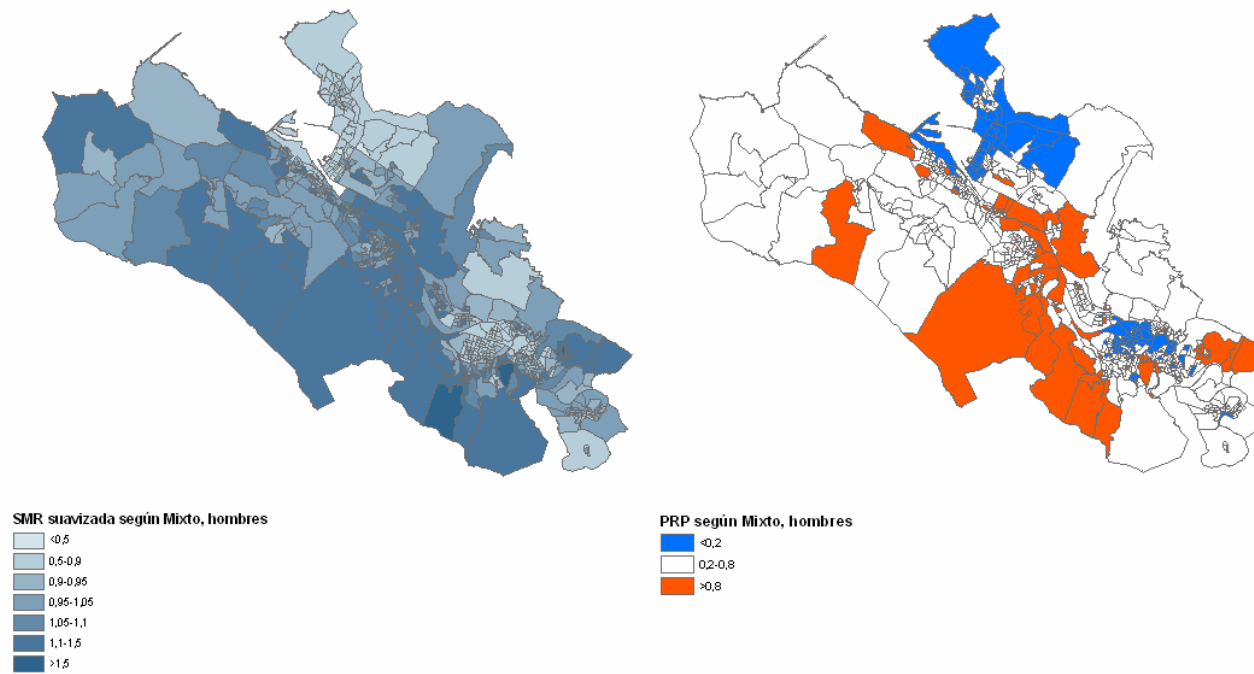


Figura 49.-Cáncer de vejiga urinaria, en hombres, modelo Mixto. Fuente: construcción propia.

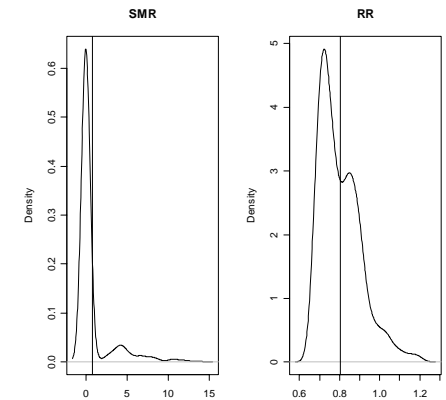
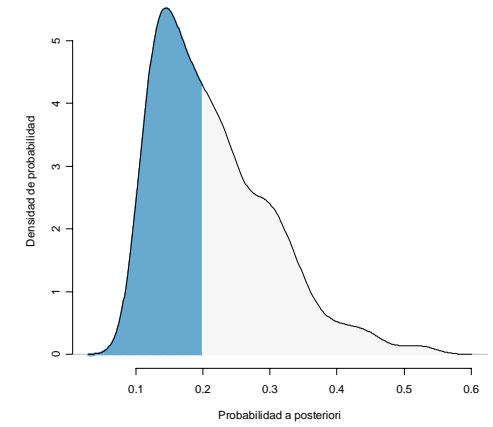
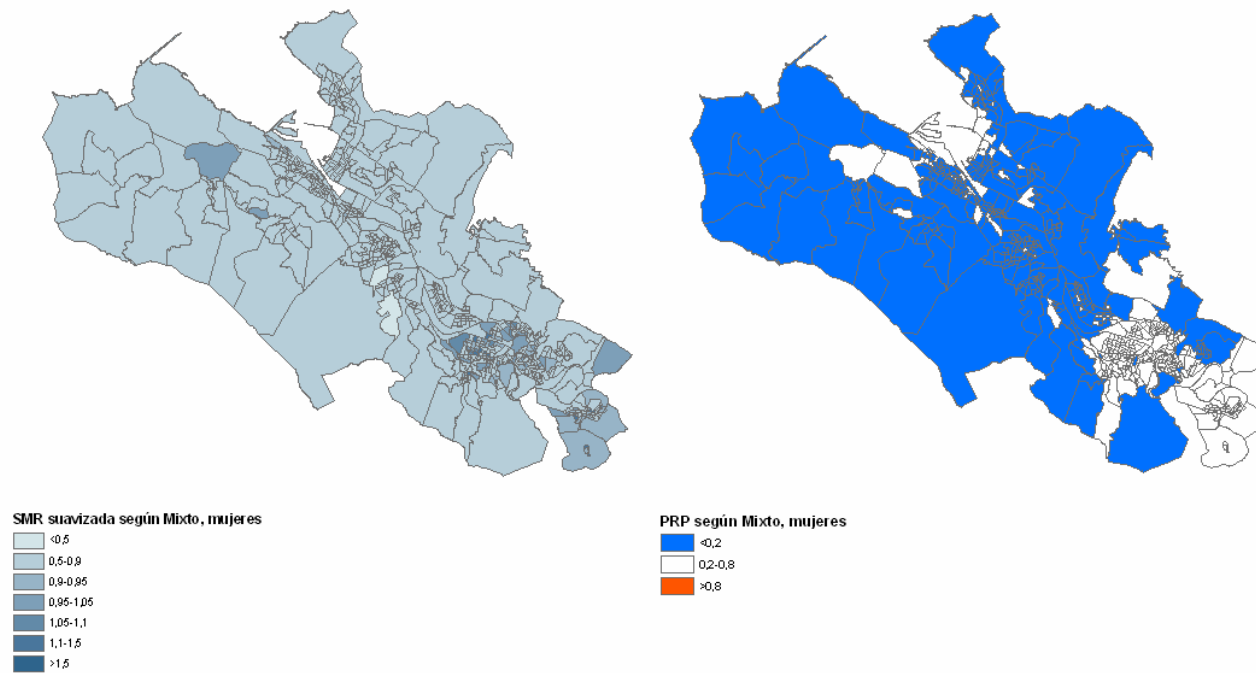


Figura 50.-Cáncer de vejiga urinaria, en mujeres, modelo Mixto. Fuente: construcción propia.

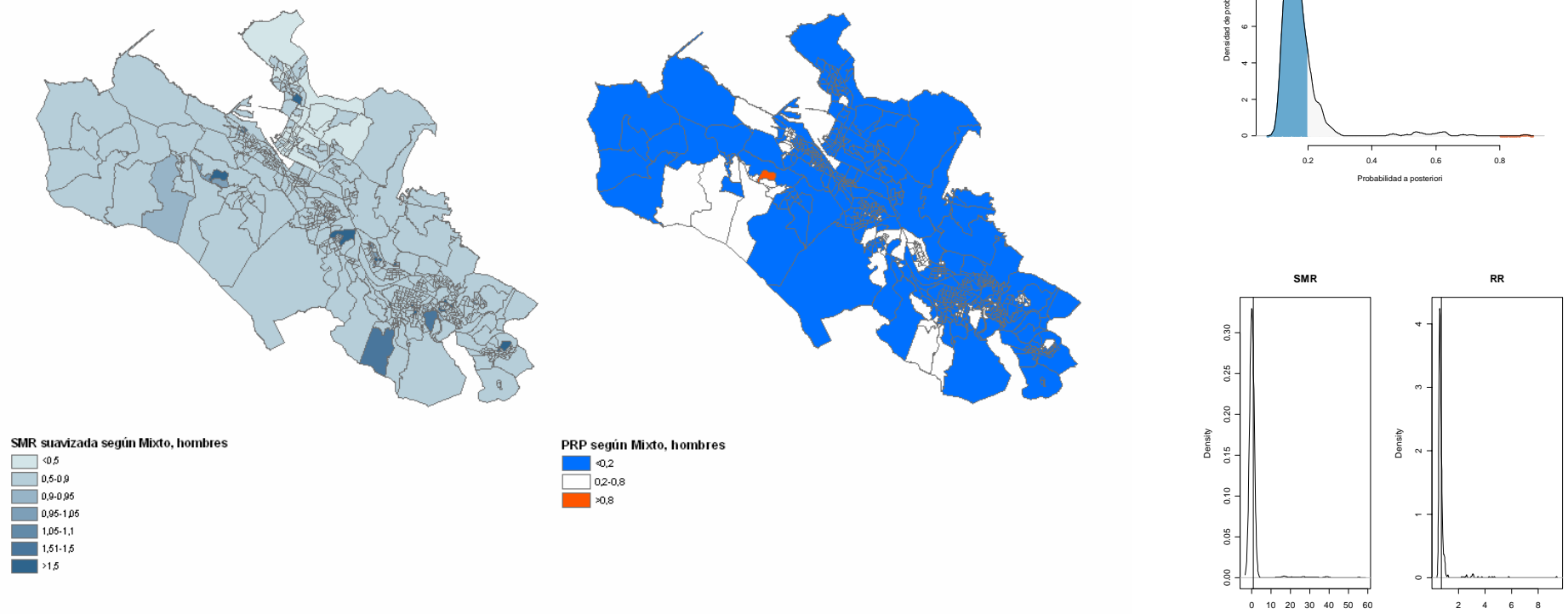


Figura 51.-Cáncer de tiroides, en hombres, modelo Mixto. Fuente: construcción propia.

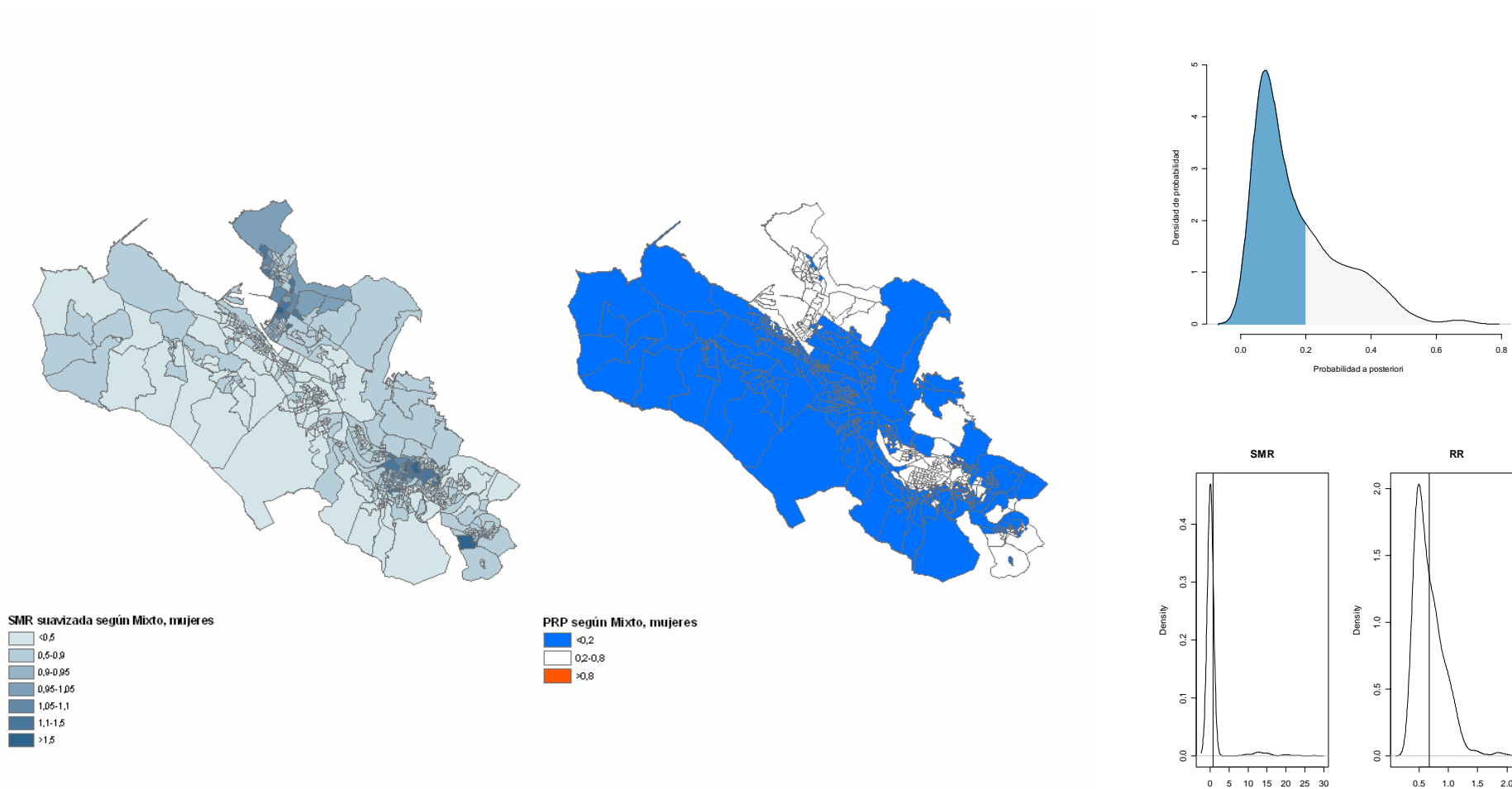


Figura 52.-Cáncer de tiroides, en mujeres, modelo Mixto. Fuente: construcción propia.

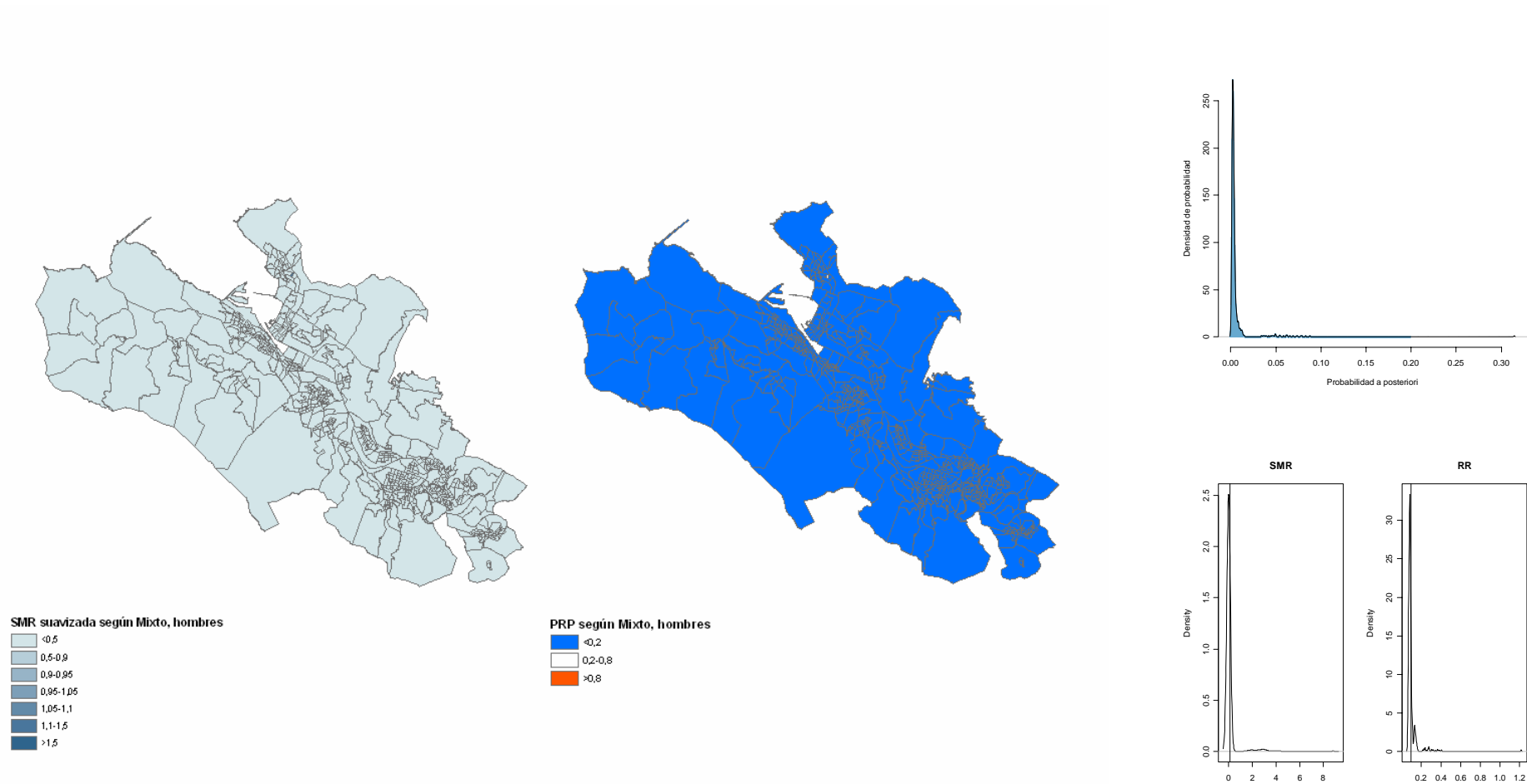


Figura 53.-Linfoma no Hodgkin, en hombres, modelo Mixto. Fuente: construcción propia.

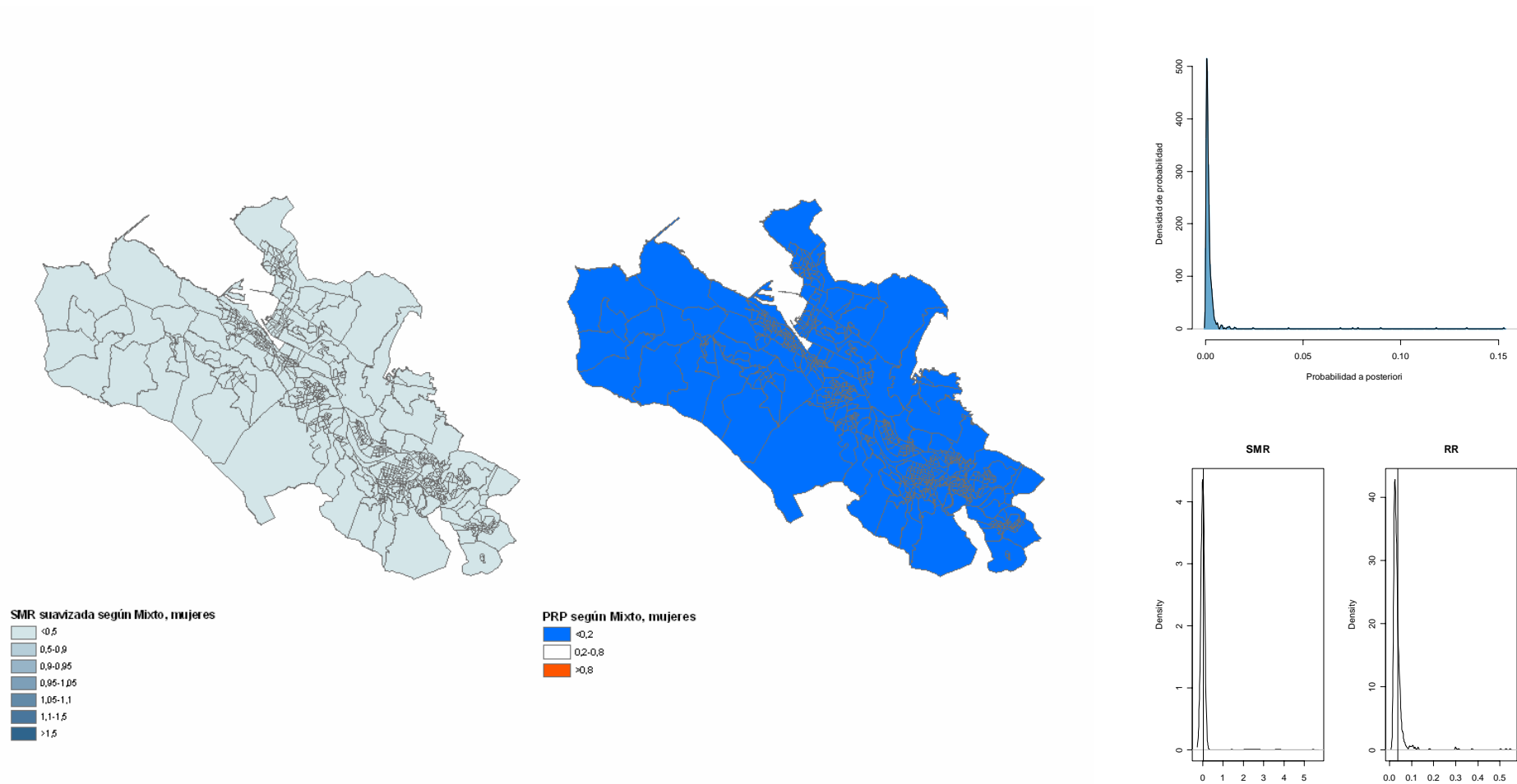


Figura 54.-Linfoma no Hodgkin, en mujeres, modelo Mixto. Fuente: construcción propia.

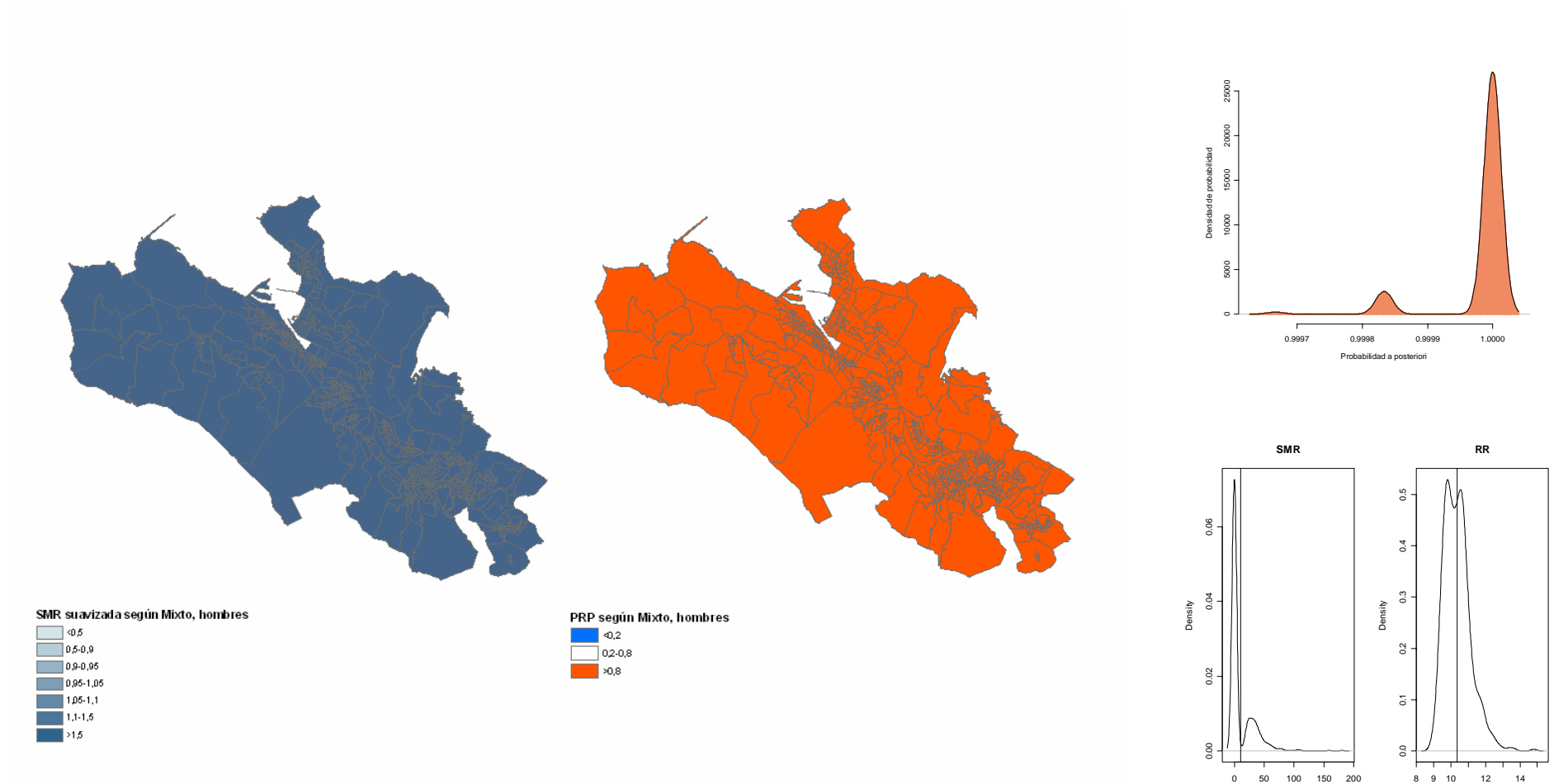


Figura 55.- Linfoma Hodgkin, en hombres, modelo Mixto. Fuente: construcción propia.

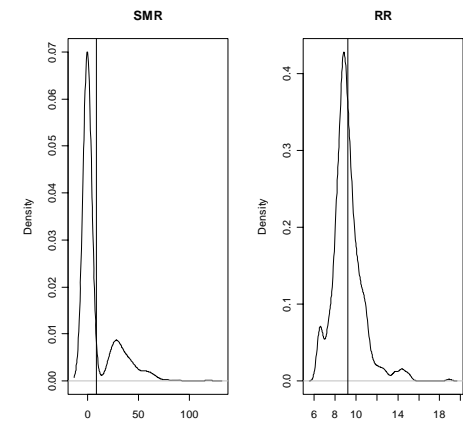
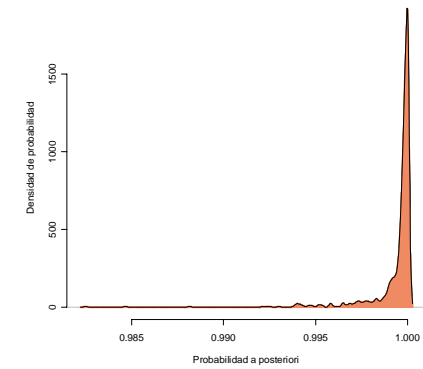
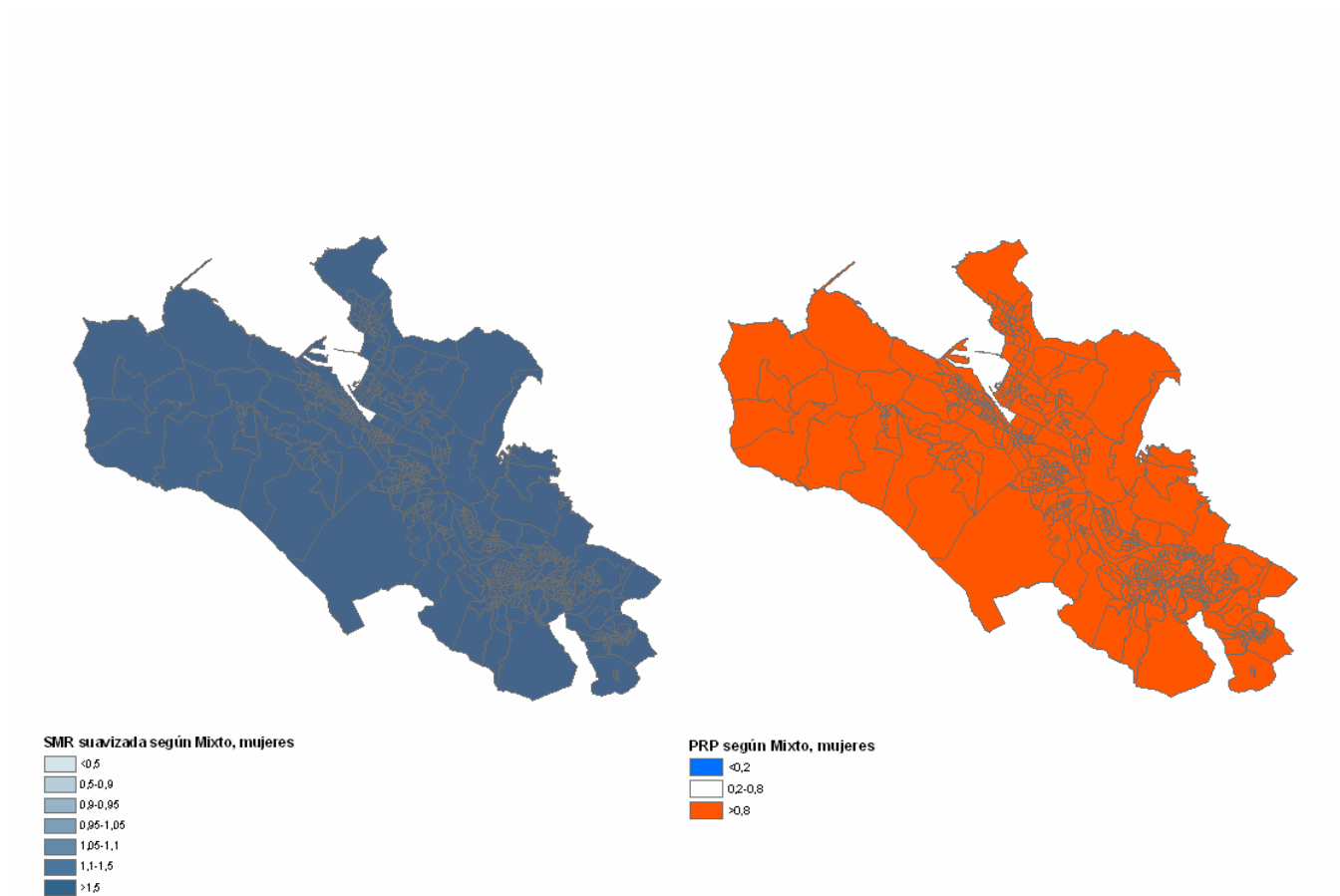


Figura 56.- Linfoma Hodgkin, en mujeres, modelo Mixto. Fuente: construcción propia.

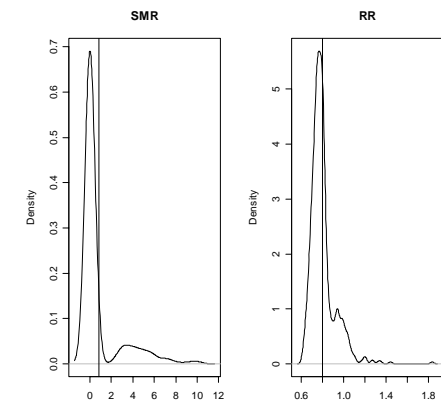
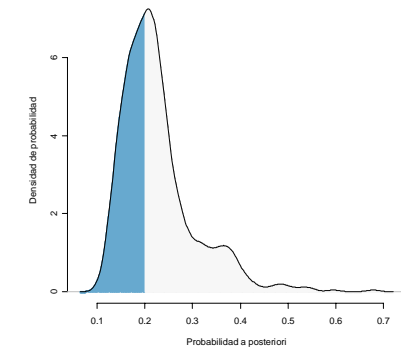
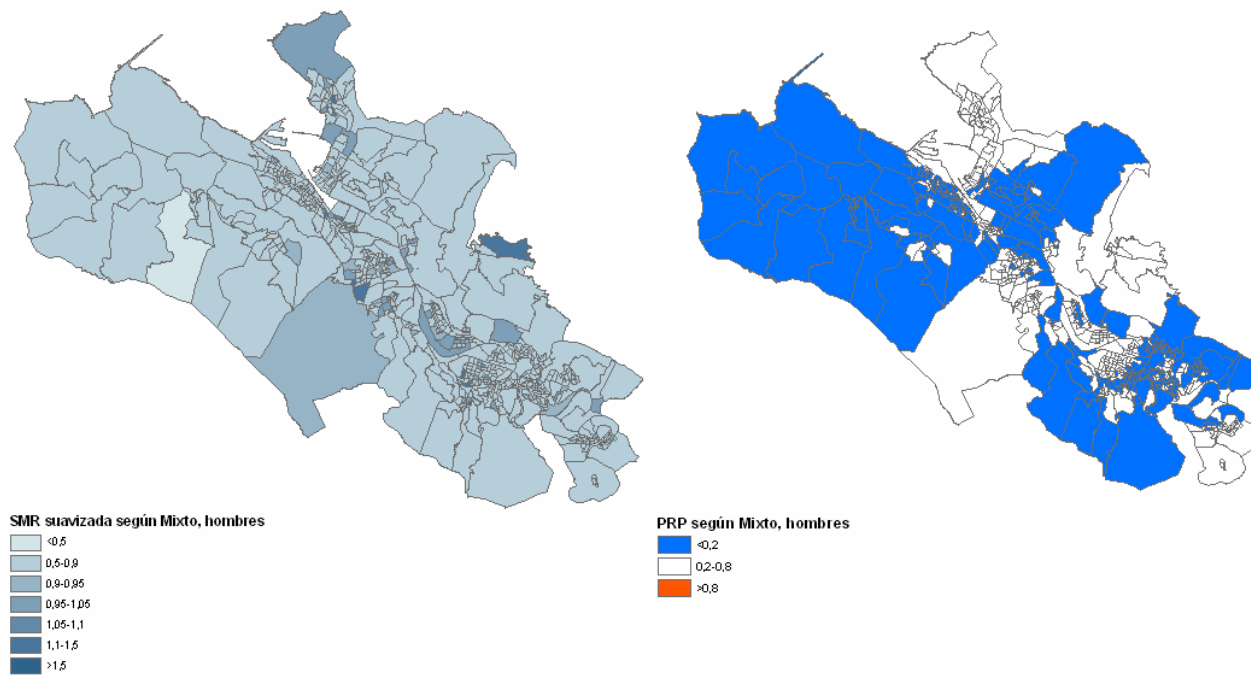


Figura 57.-Mieloma, en hombres, modelo Mixto. Fuente: construcción propia.

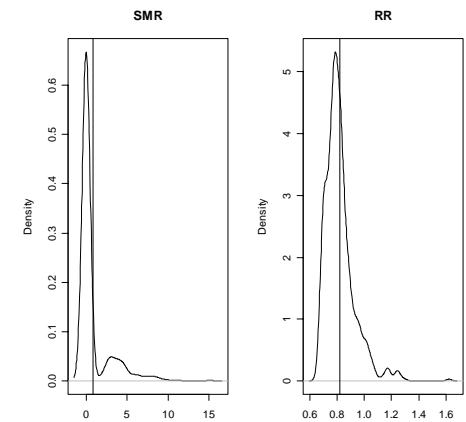
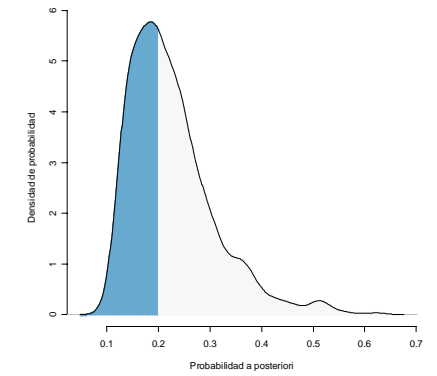
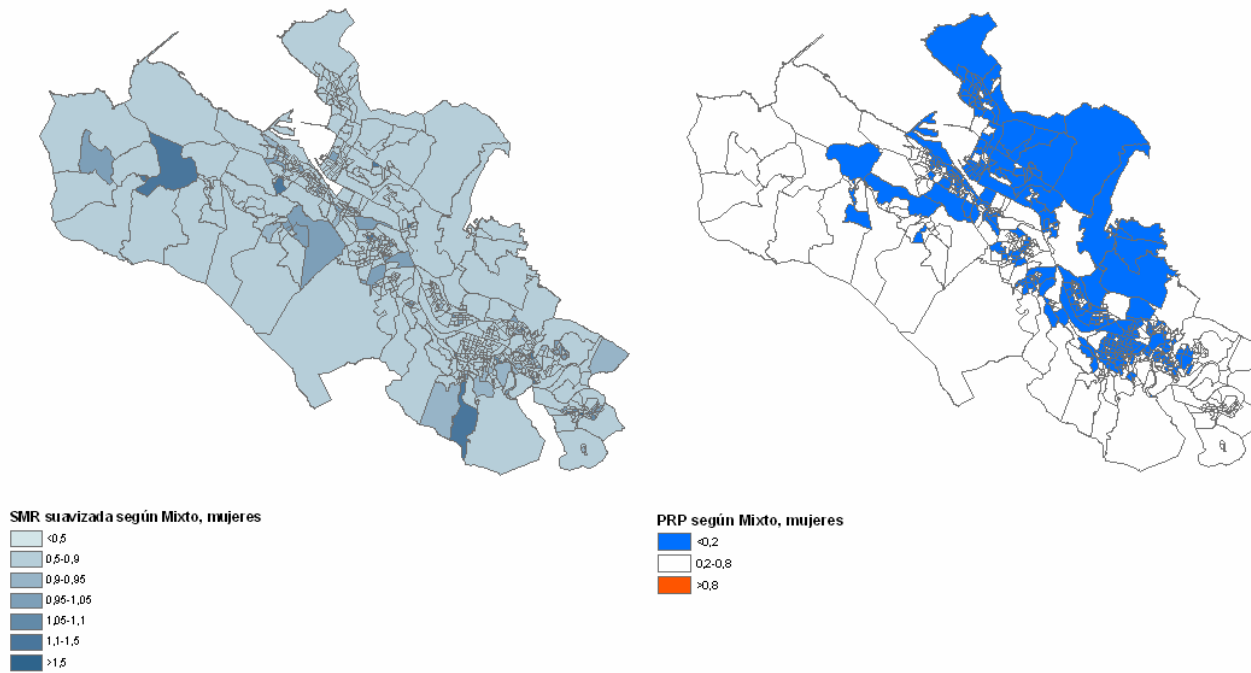


Figura 58.-Mieloma, en mujeres, modelo Mixto. Fuente: construcción propia.

Las RME suavizadas se muestran en las Figuras 47 a 58. Se repitieron los resultados, en el sentido de la existencia de un patrón geográfico claro, en el cáncer de tráquea, bronquios y pulmón, ambos sexos (Figuras 47 y 48); y vejiga urinaria, hombres (Figura 49). Cabe señalar el patrón del exceso de riesgo diferenciado por sexo en el caso de tráquea, bronquios y pulmón. Nótese que, excepto Bilbo con un exceso de riesgo en ambos casos, en hombres las secciones con mayor riesgo se concentraron en la parte izquierda de la ría y en las mujeres, en cambio, en la parte derecha.

b- Modelos de asociación geográfica

Como en el caso del Área Metropolitana de Barcelona, el índice de privación y el contaminante (NO_2 o SO_2) fueron introducidos conjuntamente en el (mejor) modelo, el índice como efecto fijo y el contaminante (media y variabilidad intra-área) como efecto aleatorio.

El coeficiente asociado al índice de privación resultó estadísticamente significativo en el caso de tráquea, bronquios y pulmón, ambos sexos; y en vejiga urinaria, hombres (ver Tabla 8 y Figura 59). El efecto del índice, en el cáncer de tráquea, bronquios y pulmón, resultó mayor en las mujeres que en los hombres.

Los coeficientes asociados a los contaminantes resultaron estadísticamente significativos únicamente en tráquea, bronquios y pulmón (véase Tabla 8 y Figuras 60 y 61). Nótese, como ocurrió con el índice de privación, que el efecto fue mayor sobre las mujeres. El efecto del SO_2 , sin embargo sólo resultó estadísticamente significativo en los hombres. Como en el caso del Área Metropolitana de Barcelona las desviaciones típicas del coeficiente asociado al SO_2 fueron muy importantes.

La introducción de variables explicativas, de nuevo, apenas redujo la variabilidad (Figuras 62 a 64).

Cáncer de:	Tráquea, bronquios y pulmón		Vejiga urinaria		Tiroides		Linfoma no Hodgkin		Linfoma Hodgkin		Mieloma	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Índice de privación	0,12167(0,02427)	0,24768(0,06036)	0,12416(0,05047)	-0,05046(0,10932)	0,13562(0,28477)	-0,35051(0,20978)	-0,18184(0,24447)	0,52660(0,42786)	-0,04100(0,07863)	-0,04530(0,08946)	-0,06215(0,10992)	0,04183(0,09777)
NO₂	0,51842(0,21640)	0,60014(0,21583)	0,07432(0,43385)	0,16501(0,43205)	-0,81881(1,42072)	1,09182(0,95898)	1,30501(1,83437)	-1,40836(1,29953)	0,00426(0,44181)	0,48257(0,41211)	-0,02036(0,55975)	0,52475(0,41186)
SO₂	0,54004(0,20918)	0,31587(0,53973)	0,44252(0,39238)	1,64439(1,25172)	1,42072(4,67769)	1,91886(1,69491)	1,74607(2,08303)	1,74865(1,04709)	-0,51813(0,51342)	0,33569(0,70889)	-0,20371(0,73079)	-0,08994(0,60341)
Efectos aleatorios												
Espacial	0,11025(0,05881)	0,30801(0,13834)	0,15875(0,11023)	0,34926(0,20241)	4,67769(3,96998)	0,47571(0,34640)	0,50472(0,44331)	1,571298(1,14776)	0,16325(0,11989)	0,40754(0,26000)	0,33265(0,26511)	0,24600(0,18334)
No Espacial	0,06112(0,02915)	0,12316(0,07959)	0,09051(0,0749)	0,27624(0,18300)	0,84983(0,73579)	0,44307(0,30558)	0,75419(0,57308)	1,19911(0,98950)	0,17811(0,12546)	0,45173(0,14122)	0,27673(0,18809)	0,26446(0,18183)
Salto		0,37134(0,11142)	0,18122(0,06276)	0,30123(0,11223)	0,99807(0,61461)	0,39155(0,16173)	0,65112(0,18156)	1,34358(0,65456)	0,23453(0,11762)	0,18411(0,12755)	0,37145(0,13231)	0,20217(0,10052)
AIC	1487,567	1014,321	931,641	647,879	409,752	364,753	550,565	272,441	922,106	869,545	816,775	864,533
BIC	3294,233	1839,981	1260,509	1293,836	1109,51	820,009	1488,788	767,275	1828,461	1784,583	1778,403	1885,569
DIC	850,135	694,540	756,825	396,228	133,227	193,942	173,861	76,174	578,626	524,299	446,119	470,502

Tabla 8.– Resultados de la estimación en los modelos de asociación geográfica. Variables explicativas: Índice socioeconómico, NO₂ y SO₂. Fuente: construcción propia.

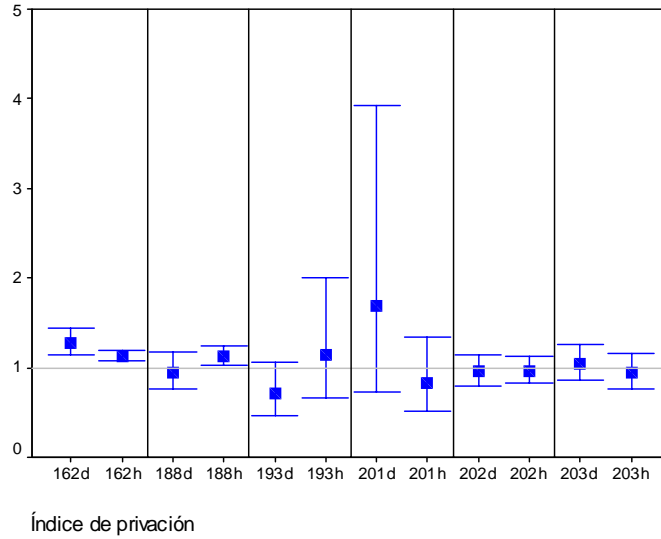


Figura 59.–Riesgo relativo asociado al índice de privación. Fuente: construcción propia.

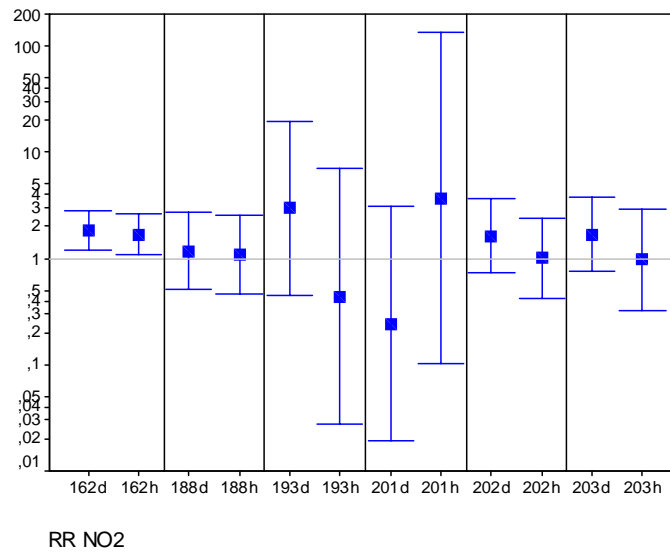


Figura 60.–Riesgo relativo asociado al NO₂. Fuente: construcción propia.

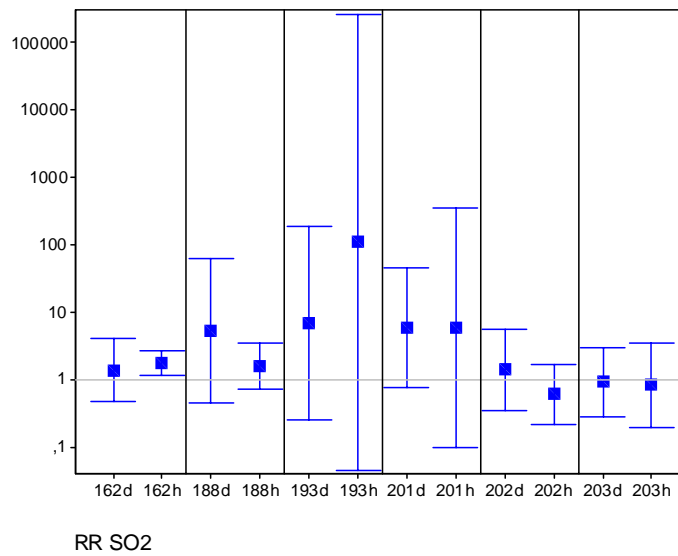


Figura 61.–Riesgo relativo asociado al SO₂. Fuente: construcción propia.

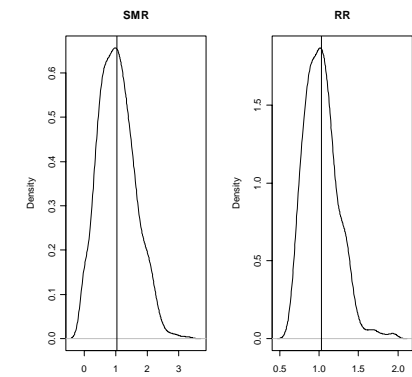
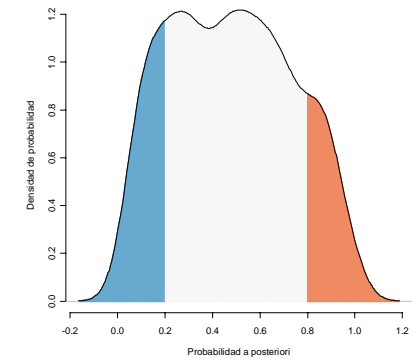
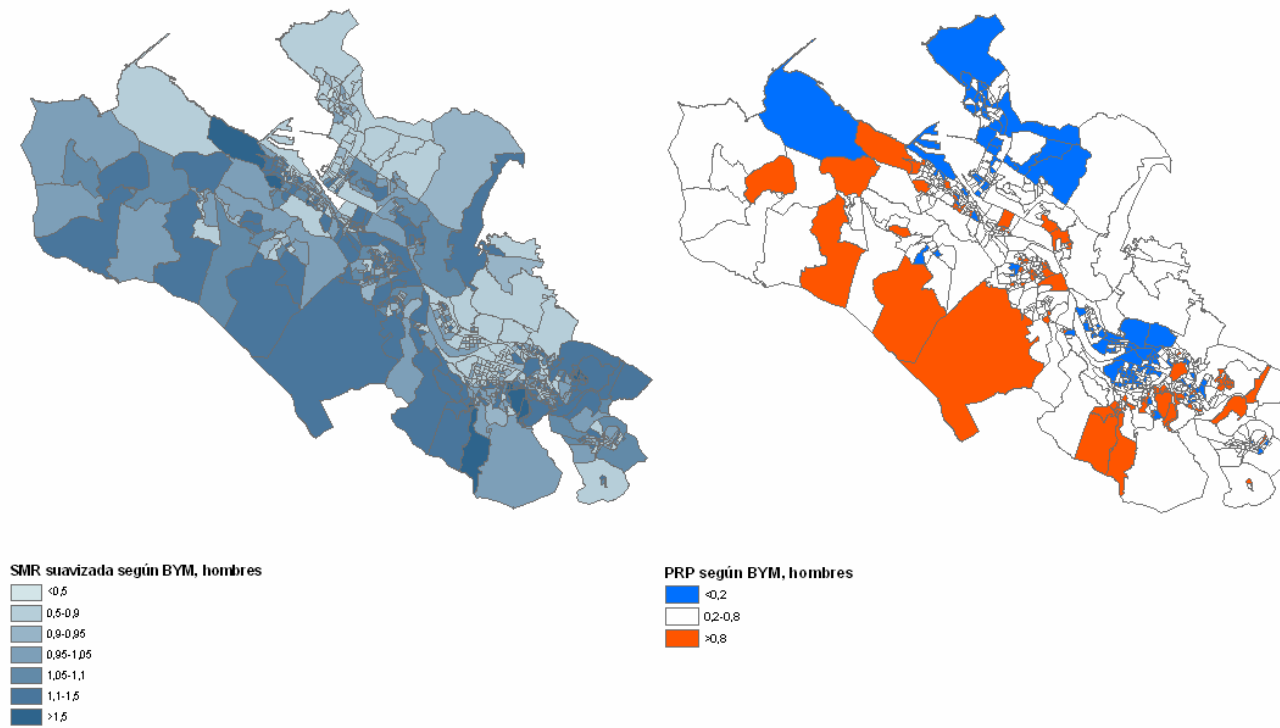


Figura 62.-Cáncer de tráquea, bronquios y pulmón, introduciendo variable explicativa, modelo BYM, hombres. Fuente: construcción propia.

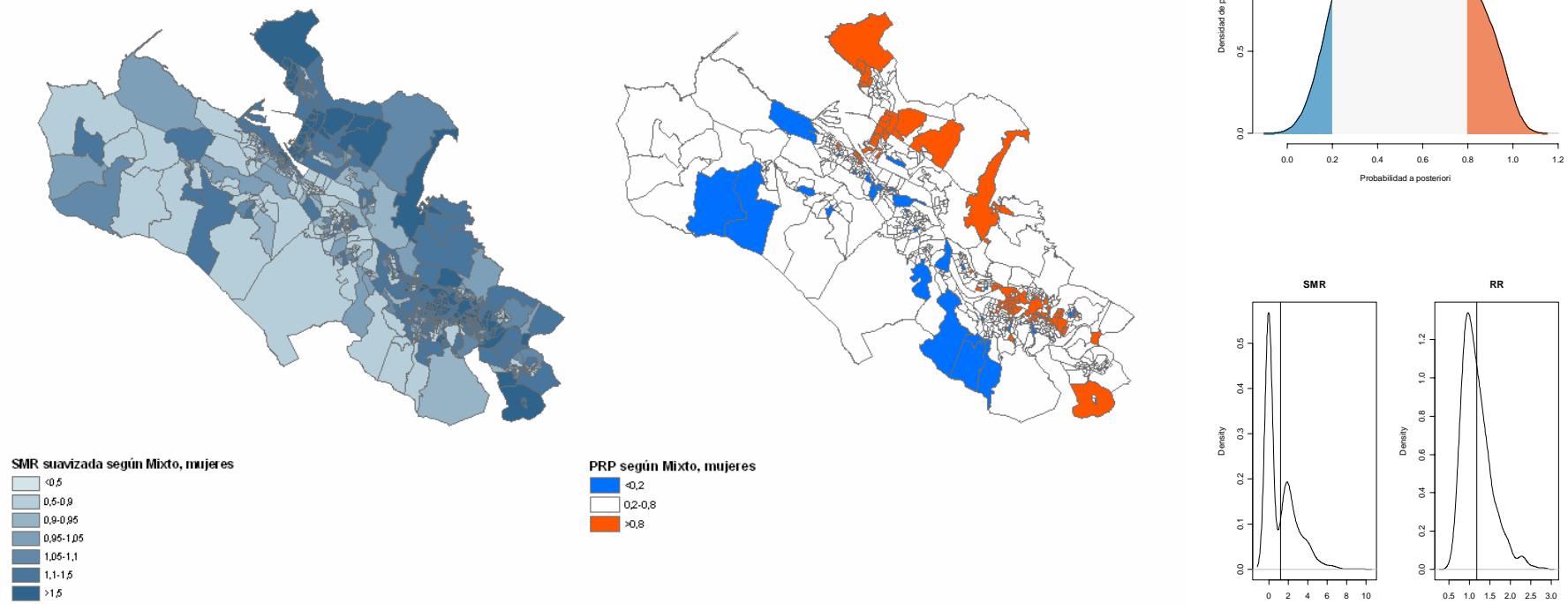


Figura 63.-Cáncer de tráquea, bronquios y pulmón, introduciendo variable explicativa, modelo mixto, mujeres. Fuente: construcción propia.

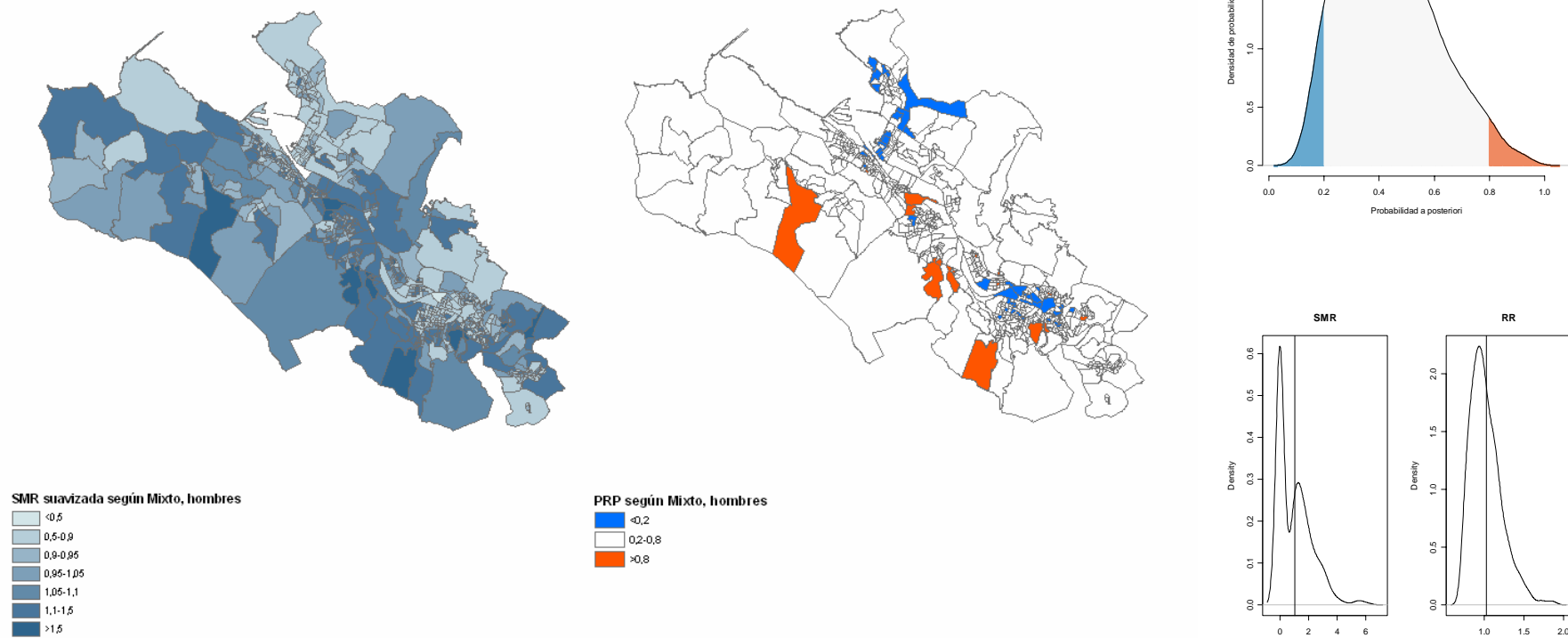


Figura 64.-Cáncer de vejiga urinaria, introduciendo variable explicativa, modelo mixto, hombres. Fuente: construcción propia.

Referencias

Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; 43:1–59.

Brooks Sp, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* 1998; 7: 434–455.

Gelman A. Parameterization and bayesian modeling. *Journal of American Statistical Association* 2004; 99:537–345.

Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2005; 1(2): 1–19.

Gelman A, Carlin JC, Stern H and Rubin DB. *Bayesian Data Analysis*. Chapman and Hall, Nueva York, 2005 (2nd edition).

Mollie A. Bayesian mapping of disease. En Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Nueva York: Chapman & Hall, 1996, pp. 359–379.

Park D, Gelman A, Bajumi J. Bayesian Multilevel Estimation with Poststratification: State–Level Estimates form National Polls. *Political Anlaysis* 2004; 12:375–385.

Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease–mapping studies. *Environmental Health Perspectives* 2004; 112 (9): 1016–1025.

Spiegelhalter D. Two brief topics on modelling with WinBUGS. *IceBUGS: Finland, 11th–12th February* 2006.

Spiegelhalter DJ, Thomas A, Best NG, Gilks WR, Lunn D. *BUGS: Bayesian inference using Gibbs sampling*. Cambridge, UK: MRC Biostatistics Unit, 1994, 2003.

Spiegelhalter DJ, Best NG, Carlin BP, Van del Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002; 64: 583–63.

5.-Conclusiones y discusión

-Tanto la incidencia como la mortalidad de las neoplasias de interés medioambiental presentaron dos fuentes de extravariación. La extravariación espacial, por la que unidades vecinas tienden a presentar razones de incidencia/mortalidad similares, y la heterogeneidad no espacial.

-En general la extravariabilidad espacial ha resultado ser mucho mayor (prácticamente el doble) que la no espacial.

Como hemos venido comentando las RIE/RME crudas, aunque ajustan por la población a riesgo de padecer las neoplasias de interés medioambiental, presentan una importante variabilidad que provoca su inestabilidad. Comentamos que la fuente más importante de extravariabilidad es la espacial. De hecho, estas razones se han suavizado utilizando el mejor modelo en cada caso. Éste resultó ser el modelo mixto de riesgo relativo excepto por lo que se refiere al cáncer de tráquea, bronquios y pulmón en hombres (en el que las RIE/RME se suavizaron utilizando un modelo BYM). Cabe señalar que esta neoplasia presentó relativamente muy pocos ceros. Un 14,6% de ceros en el caso de la incidencia y un 4,56%, Bilbo, y un 11,93%, Barcelona, en el caso de la mortalidad.

-Para suavizar las RIE/RME correspondientes a variables con un porcentaje de ceros superior al 40-50%, debe utilizarse un modelo que capture este comportamiento.

-El mejor modelo en términos de ajuste para recoger el exceso de ceros en las variables de interés ha resultado ser el modelo mixto de riesgo relativo.

El exceso de ceros, es decir, la existencia de muchos más municipios/secciones censales en los que no se produce ningún caso de las neoplasias de interés medioambiental que lo que correspondería a un proceso contador ha implicado que la

distribución de probabilidad de las variables no se corresponda exactamente con una Poisson. Como comentamos, existen varias alternativas, modelos de mixturas finitas y modelo mixto de riesgo relativo. Por lo que se refiere al primer tipo de modelos el peor ajuste lo proporciona el modelo Hurdle. De hecho no parece razonable que el proceso generador de los ceros sea totalmente independiente del proceso generador de los casos positivos tal y como supone este tipo de modelo (Ridout et al, 1998, Dalrymple et al, 2003). El mejor ajuste (siempre y cuando se produce un 'exceso' de ceros) ha resultado ser el modelo mixto de riesgo relativo. Los resultados confirman que este tipo de modelos no suavizó en exceso, en el sentido de mantener también aquellas discontinuidades presentes en los datos (Lawson et al, 2003).

Una de las líneas de investigación que estamos realizando en estos momentos es averiguar cual es el umbral que permite definir una distribución de probabilidad con exceso de ceros. Nos atrevemos a sugerir un valor en torno a 40-50%.

Un aspecto que debe ser muy tenido en cuenta es el 'grado de suavización' que comporta este tipo de modelos (Elliott y Wartenberg, 2004). Este grado determina un compromiso entre una alta sensibilidad, áreas de exceso de riesgo correctamente identificadas, y una alta especificidad, áreas que no tienen exceso de riesgo correctamente identificadas (Richardson et al, 2004). Un resultado sensible pero no específico generará falsos positivos, mientras que uno específico pero no sensible, no identificará aquellas áreas con un riesgo alto. Richardson et al (2004), tras un exhaustivo estudio de simulación, mostraron que se requieren como mínimo 5 casos esperados para que, con un riesgo relativo de 2 a 3, el método de suavización utilizado proporcione resultados suficientemente sensibles, en el sentido de detectar áreas con un riesgo alto. El riesgo, como se ve, es muy elevado. Por ello, Richardson et al (2004) proponen utilizar en vez del riesgo relativo, la probabilidad a posteriori (PRP) que este riesgo sea mayor que la unidad. Ésta fue la razón por la que introdujimos en este trabajo mapas de la PRP además de las RIE/RME suavizadas. Una PRP mayor del

80% (o menor del 20% para el riesgo reducido) implica que el método de suavización utilizado fue lo suficientemente específico.

-Las RIE/RME suavizadas presentan un patrón geográfico claro sólo en algunas de las neoplasias de interés medioambiental.

El cáncer de tráquea, bronquios y pulmón, en ambos sexos, es el que presenta un comportamiento geográfico bastante evidente en las tres zonas de estudio. En segundo lugar, cabe mencionar el cáncer de vejiga urinaria en hombres, que en los tres casos muestra áreas con exceso de riesgo agrupadas espacialmente. El cáncer de vejiga urinaria en mujeres podría mostrar cierta estructura geográfica sólo en el Área Metropolitana de Barcelona. Del mismo modo, la incidencia del cáncer de tiroides en mujeres podría presentar cierta estructura geográfica.

Como es sabido, los mapas de enfermedades han sido utilizados para generar hipótesis etiológicas. Aunque no es nuestro objetivo en esta tesis, debemos señalar la conducta geográfica de, en primer lugar, el cáncer de tráquea, bronquios y pulmón, con riesgos relativos más elevados en zonas más deprimidas económicamente. Así, en el Área Metropolitana de Barcelona podría existir un exceso de riesgo en el barrio de Ciutat Vella, Barcelona; o, también en el Área Metropolitana de Bilbo, el exceso de riesgo en hombres se concentra en la parte izquierda de la ría y en las mujeres, en cambio, en la parte derecha. En segundo lugar, por lo que se refiere al cáncer de tiroides en mujeres, existe una agrupación geográfica en la comarca del Baix Empordà.

Esta última agrupación será objeto de un estudio exhaustivo en un futuro próximo.

-Parte de la variabilidad remanente en las RIE/RME suavizadas pudo ser explicada mediante la introducción de variables explicativas, en particular la contaminación atmosférica y variables socioeconómicas.

-Como los contaminantes atmosféricos fueron observados en un diseño geoestadístico y las neoplasias de interés medioambiental lo fueron en un diseño en rejilla se modelizó la superficie de exposición.

La información sobre la exposición medioambiental se obtuvo de unas, relativamente, pocas estaciones dispersas espacialmente. Por este motivo se utilizaron métodos de interpolación que permitieron asignar los valores continuos del contaminante a los discretos de la neoplasia. De hecho existieron dos situaciones diferenciadas. Una, la Regió Sanitària Girona con muy pocas estaciones y muy dispersas espacialmente; y otra, las Àreas Metropolitanas, con bastante más estaciones que cubrían la mayor zona estudiada. Cabe destacar en especial el Àrea Metropolitana de Bilbo con 62 estaciones que cubren ampliamente el territorio. Sólo en el primer caso se utilizó un método de interpolación determinista. La superficie de exposición de los contaminantes en las Àreas Metropolitanas se interpoló utilizando un método estocástico, kriging ordinario. Dos fueron las razones que condujeron a esta elección: en primer lugar, la variable espacial contaminante era isotrópica y no presentaba ninguna tendencia espacial; en segundo lugar, el valor de la media de esta variable espacial era desconocido.

La interpolación estocástica (utilizada también por la mayoría de los autores, véase Lawson et al, 2003) permite estimar, además de la media, la incertidumbre, es decir, la variabilidad intra-àrea.

-El efecto del contaminante en cada municipio/sección censal se aproximó introduciendo en el modelo el valor promedio en cada àrea y la variabilidad intra-àrea.

-El efecto del contaminante se consideró aleatorio, en el sentido de que podría ser diferente en cada una de las àreas.

Varios autores, aún modelizando de forma estocástica la superficie de exposición, introducen en el modelo únicamente el valor promedio del contaminante en el àrea

correspondiente (Le et al, 1996; Gelfand et al, 2001; Tonellato, 2001; Shaddick y Wakefield, 2002). El problema de todos estos trabajos es que suponen que la posible asociación observada a nivel de área (nivel ecológico) es la misma para todos los individuos residentes en esa área (nivel individual) (Wakefield y Shaddick, 2006). Este supuesto podría conducir al problema de la falacia ecológica (Selvin, 1958) en su versión de 'sesgo puro de especificación' que aparece cuando se agrega un modelo no lineal (Wakefield y Shaddick, 2006).

En el caso de la incidencia de las neoplasias de interés medioambiental, es decir, en la Regió Sanitària Girona, se utilizó una interpolación determinista que, como es lógico, no proporciona una estimación de la incertidumbre. Sin embargo, y aprovechando que los datos no sólo se recogieron espacialmente sino también en el tiempo, se diseñó un método de remuestreo para obtener una estimación de la incertidumbre.

Pero además de introducir la variabilidad intra-área para evitar este tipo de sesgo, en este trabajo el coeficiente asociado al contaminante se ha considerado un efecto aleatorio. La razón es que se ha supuesto que el error de estimación no ha sido exclusivamente un error de medida sino que la exposición aunque medida ecológicamente ocurrió siempre a nivel individual (Lawson et al, 2003).

El modelo finalmente utilizado se basó en el de Jackson et al, (2006). De hecho, existen otras alternativas, como utilizar una aproximación Bayesiana completa, en los que se estimen simultáneamente los parámetros de salud y de la exposición; o el modelo en tres etapas propuesto por Wakefield y Shaddick (2006). Sin embargo, el primero, dada la gran cantidad de datos disponibles, resulta computacionalmente inabordable; mientras que el modelo de Wakefield y Shaddick no ha proporcionado resultados satisfactorios en términos de ajuste.

-Las condiciones socioeconómicas fueron otra de las variables que redujeron la variabilidad remanente en las RIE/RME suavizadas.

-Las variables socioeconómicas fueron resumidas a partir de un índice de privación que, utilizando ACP, recogió cerca del 70% de la varianza de estas variables.

De nuevo los resultados del ACP para las Áreas Metropolitanas difieren del obtenido para la Regió Sanitària Girona. Mientras que en las Áreas Metropolitanas bastó un único factor para aproximar el índice de privación se necesitaron dos en el caso de la Regió. Dos podrían ser las razones de estos resultados. En primer lugar, la Regió Sanitària Girona implica, lógicamente, un territorio mucho más disperso y heterogéneo desde el punto de vista demográfico y socioeconómico. Las áreas metropolitanas suelen ser más homogéneas en este sentido. Nos atrevemos a sugerir que la razón más importante, sin embargo, está relacionada con la naturaleza de las variables incluidas en el índice. De los dos factores que aproximaban este índice en la Regió Sanitària Girona, el primero, muy correlacionado con el desempleo (de forma negativa) y con la variable trabajadores manuales, podría no recoger del todo el concepto de privación material. Sugerimos que este factor, y por tanto el índice resultante en este caso, reflejó más el concepto de población 'no urbana'. Así, utilizando datos del censo de población y viviendas del 2001, la población ocupada en agricultura, ganadería y pesca en las comarcas de Girona fue el 5,186%, el doble que el porcentaje correspondiente a Bizkaia (2,625%), prácticamente el doble al del resto de Catalunya (3,216%) y más de diez veces mayor que el del Área Metropolitana de Barcelona (0,451%).

-Las variables explicativas observadas con un diseño en rejilla, como el índice de privación, se introdujeron en el modelo como efectos fijos.

El índice de privación no sólo se observa de forma ecológica sino que también se construye de la misma forma, es decir, promediando los eventuales valores individuales en cada área.

-Mayor privación, controlando por contaminantes atmosféricos, implicó un mayor riesgo de padecer/morir en algunas de las neoplasias de interés medioambiental en las tres zonas y períodos estudiados.

El índice de privación (o al menos uno de los dos factores en el caso de la Regió Sanitària Girona) estuvo siempre asociado con un mayor riesgo de padecer/morir de cáncer de tráquea, bronquios y pulmón. Menos consistente fue la relación entre el cáncer de vejiga urinaria e índice de privación, encontrándose significativo en ambos sexos en el Área Metropolitana de Bilbo y únicamente en mujeres en la Regió Sanitària Girona. Para el resto de las neoplasias la relación fue estadísticamente significativa en alguna de las zonas estudiadas. Así, cáncer de tiroides, mujeres, y linfoma hodgkin, ambos sexos, en Girona; y linfoma no Hodgkin y mieloma, ambos sexos, en el Área Metropolitana de Barcelona fueron las neoplasias en las que se encontró una asociación significativa.

Son pocos los estudios que relacionan la incidencia por cáncer y las condiciones socioeconómicas (véase un revisión en Lertxundi y Saez, 2006; y en Ocaña-Riola et al, 2006) y son muchos más escasos los que estudian la relación cáncer y condiciones socioeconómicas controlando por variables medioambientales.

Tras una exhaustiva revisión en Medline realizada en agosto del 2006, por lo que se refiere a la incidencia y/o mortalidad por cáncer de pulmón (en varios de estos trabajos se presentan más de una causa), señalaremos por orden cronológico los

trabajos de McWhorter et al (1989), en el que encuentran relaciones estadísticamente significativas con la diferencia racial entre blancos y negros en Estados Unidos; Polednak (1992) con la inmigración portorriqueña en Estados Unidos; Gorey y Vena (1995) con la pobreza en las Áreas Metropolitanas del Norte del Estado de Nueva York; Burnley (1997) con variables de estatus socioeconómico y trabajadores manuales en Nueva Gales del Sur; Lamont et al (1997) con un indicador de la pobreza extrema (individuos sin techo) en la ciudad de Glasgow y en el oeste de Escocia; Pollock y Vickers (1997) con un índice de privación en la Región de Támesis del Sur; Kreiger et al (1999) con un indicador de la etnia en Estados Unidos; Dreassi et al (2003, 2005) con un índice de privación ad-hoc en la Toscana; Ocaña-Riola et al (2004) con un índice de privación ad-hoc en Andalucía; Menvielle et al (2005) con ocupación y educación en Francia; y Ekberg-Aronsson et al (2006) con la clase socioeconómica en la ciudad de Malmo, Suecia.

Para el cáncer de vejiga urinaria citaremos a Ocaña-Riola et al (2004) que encuentra una relación estadísticamente significativa con un índice de privación ad-hoc en Andalucía; y Underwood et al (2006) con la raza en Estados Unidos.

En cuanto al Mieloma cabe remontarse al trabajo de McWhorter et al (1989), en el que encuentran relaciones estadísticamente significativas con las diferencias raciales entre blancos y negros en Estados Unidos; Baris et al (2000) con la renta, educación y clase social según la ocupación en Estados Unidos; y Gebregziabher et al (2006) con clase social en el condado de Los Angeles.

En el linfoma de Hodgkin, Flavell et al (1999) encuentra relaciones significativas con el índice de privación de Townsend en West Midlands; Glasser et al (2001) con la clase social en el Área Metropolitana de San Francisco; y Clarke et al (2005) con clase social en California.

Por último McNally et al (2003) encuentran una relación estadísticamente significativa entre la incidencia y mortalidad de los linfomas y el índice de privación de Townsend en el Noroeste de Inglaterra.

-El efecto de la privación sobre la incidencia y/o mortalidad por cáncer de tráquea, bronquios y pulmón, controlando por contaminantes atmosféricos, fue mayor en las mujeres que en los hombres.

Los únicos artículos en los que se ha encontrado una relación diferente según el sexo han sido el de Polednak (1992) y Menvielle et al (2005), aunque en estos casos el riesgo mayor fue encontrado para los hombres. En estos artículos existen al menos dos diferencias respecto a nuestro trabajo. En primer lugar, ninguno de ellos se controla por variables medioambientales y en segundo lugar, en ambos casos utilizan variables comunes para ambos sexos.

Aunque aventuramos que no sólo el efecto sino la naturaleza del índice podría ser diferente según el género, serían precisas, investigaciones adicionales que permitiesen determinar los diferentes efectos encontrados en esta tesis.

-Altas concentraciones de contaminantes atmosféricos aumentan el riesgo de padecer neoplasias de interés medioambiental, controlando por condiciones socioeconómicas.

El dióxido de nitrógeno fue el único contaminante asociado de forma consistente y estadísticamente significativa con la mortalidad por cáncer de tráquea, bronquios y pulmón en las Áreas Metropolitanas de Barcelona y Bilbo. El efecto, nuevamente, fue mayor en mujeres que en hombres. La asociación de NO₂ y el resto de neoplasias (vejiga urinaria y linfoma no Hodgkin en el Área Metropolitana de Barcelona) o del resto de contaminantes y neoplasias en otras zonas (SO₂ y cáncer de pulmón en el Área Metropolitana de Bilbo; y ozono y vejiga urinaria en la Regió Sanitària Girona) resultaron poco consistentes.

En el caso del Área Metropolitana de Bilbo, la posible asociación entre SO₂ y cáncer de pulmón parece venir explicada por las altas concentraciones de este contaminante al principio del período analizado. No tenemos ninguna explicación a la relación entre el ozono y la incidencia de cáncer de vejiga urinaria en la Regió Sanitària Girona. Aunque esta relación fue marginalmente significativa es evidente que se requieren investigaciones adicionales para confirmar o descartar este efecto.

Son muchos más escasos los estudios entre NO₂ (o, eventualmente, SO₂) y cáncer, de pulmón en particular, aunque en la revisión antes mencionada, encontramos los trabajos de Cohen y Pope (1995), los cuales proporcionan una revisión del tema (ahora un poco antigua); así como, Pope et al (2002) y Krewski et al (2005) en Estados Unidos; Nafstad et al (2006) en hombres noruegos; Filleul et al (2006) en Francia; y finalmente Vineis et al (2006) en el contexto del proyecto EPIC (European Prospective Investigation on Cancer and Nutrition).

No hemos seguido las sugerencias de Jackson et al (2006) de analizar las posibles interacciones entre condiciones socioeconómicas y variables medioambientales. Heinrich et al (2000), en Alemania, Pye et al (2001) en Irlanda del Norte y Morillo-Frosch y Desdale (2006) en Estados Unidos, investigan parcialmente posibles sinergias e interacciones entre ambas variables.

Creemos que la interacción entre el índice socioeconómico y la contaminación por NO₂, y quizás SO₂, en la mortalidad e incidencia de cáncer de pulmón convendría ser analizada en detalle.

-La introducción de las variables explicativas, incluso aquellas estadísticamente significativas, apenas redujo la variabilidad explicada por los modelos.

REFERENCIAS

Abdulmughni YA, Al-Hureibi MA, Al-Hureibi KA, Ghafoor MA, Al-Wadan Ah, Al-Hureibi YA. Thyroid cancer in Yemen. *Saudi Medicine Journal* 2004; 25: 55–59.

Abellán JJ. Evolución espacial de la gripe en la temporada 1998–1999. 2000. <http://dgsp.san.gva.es/SSCC/Epidemiologia/Centinelas/gripe/evo9899.htm>

Acquavella JF, Delzell E, Cheng H, Lynch CF, Johnson G. Mortality and cancer incidence among alachlor manufacturing workers 1968–99. *Occupational Environmental Medicine* 2004; 61: 680–685.

Acquavella American Cancer Society (ACS). Asbestos, 2003. http://www.cancer.org/docroot/PED/content/PED_1_3X_Asbestos.asp?sitearea=WHO [último acceso 08 Febrero 2005].

Agency for Toxic Substances and Disease Registry (ATSDR). Toxfaqs for Sulfur Dioxide. 1999. [último acceso 23 Marzo 2005] <http://www.atsdr.cdc.gov/tfacts116.html>

Ahlbom A, Cardis E, Green A, Linet M, Savitz D, Swerdlow A, et al. Review of the epidemiologic literature on EMF and Health. *Environmental Health Perspectives* 2001;109 Suppl 6:911–33.

Alavanja MCR, Dosemeci M, Samanic C, Lubin J, Lynch C, Knott C, Barker J, Hoppin JA, Sandler DP, Coble J, Thomas K, Blair A. Pesticides and Lung Cancer Risk in the Agricultural Health Study Cohort. *American Journal of Epidemiology* 2004; 160: 876–885.

Anderson, T.W. *An introduction to multivariate statistical analysis*. Nueva York: John Wiley & Sons, 1984.

Armstrong B, Hutchinson E, Unwing J, Fletcher T. Lung Cancer Risk after Exposure to Polycyclic Aromatic Hydrocarbons: A Review and Meta-Analysis. *Environmental Health Perspectives* 2004; 112: 970–975.

Arnbjörnsson E, Arnbjörnsson A, Olafsson A. Thyroid cancer incidence in relation to volcanic activity. *Archives of Environmental Health* 1986, 61: 267–269.

Atkins P, Lo Jones B. Chemistry – Molecules, Matter and Change. 3rd ed. New York: W.H. Freeman; 1997.

- Astrup H, Daneshvar B, Dragsted LO, Gamborg M, Hansen M, Loft S, et al. Biomarkers for exposure to ambient air pollution--comparison of carcinogen-DNA adduct levels with other exposure markers and markers for oxidative stress. *Environmental Health Perspectives* 1999;107:233-8.
- Bahia SHA, Echenique I, Koifman S. Cancer and wood-related occupational exposure in the Amazon region of Brazil. *Environmental Research* 2005; 99 (1): 132-140.
- Bailey TC, Gatrell AC. *Interactive spatial data analysis*. Essex: Longman Scientific & Technical, 1995.
- Barceloux DG, Nickel. *Journal Toxicology Clinical Toxicol* 1999; 37:239-58.
- Baris D, Zahm SH, Cantor KP, Blair A. Agricultural use of DDT and risk of non-Hodgkin's lymphoma: pooled analysis of three case-control studies in the United States. *Occupational Environment Medicine* 1998; 55: 522-527.
- Baris D, Morris L, Silverman D, Hayes R, Hoover R, Swanson GM, Dosemeci M, Schwartz G, Liff JM, Schoenberg JB, Pottern LM, Lubin J, Greenberg RS, Fraumeni JF. Socioeconomic Status and Multiple Myeloma Among US Blacks and Whites. *American Journal of Public Health* 2000; 90(8): 1277-81.
- Bates MN, Rey OA, Biggs ML, Hopenhayn C, Moore LE, Kalman D, Steinmaus C, Smith AH. Case-Control Study of Bladder Cancer and Exposure to Arsenic in Argentina. *American Journal Epidemiology* 2004; 159: 381-389.
- Berrino, F., Sant, M. y cols. (Eds.). *Survival of cancer patients in Europe. The EURO CARE study*. IARC Sci Publ 1995; 132: 1-463.
- Berrino, F., Capocaccia, R. y cols. (Eds.). *Survival of cancer patients in Europe. The EURO CARE-2 study*. IARC Sci Publ 1999; 151: 1-572.
- Berrino, F., Capocaccia, R. y cols. (Eds.). *Survival of cancer patients in Europe: the EURO CARE-3 Study*. Ann Oncol 2003; 14(Supl. 5): 1-118.
- Bertollini R, Douglas M, Saracci R, Savitz DA. *Exposure of Disease, Environmental Epidemiology* 1992, edit Lewis (WHO).
- Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; 43:1-59.
- Blais JM, Froese KL, Kimpe LE, Muir DC, Backus S, Comba M, Schindler DW. Assessment and characterization of polychlorinated biphenyls near a hazardous waste incinerator: analysis of vegetation, snow, and sediments. *Environmental Toxicology Chemistry* 2003; 22: 126-133.

- Blattner, WA. Human retroviruses: their role in cancer. *Proceeding of the Association of American Assotiations*, 1999; 111: 563–572.
- Blomberg J, Moller T, Olsson H, Anderson H, Jonsson M. Cancer morbidity in blood recipients--results of a cohort study. *European Journal of Cancer* 1993; 15: 2101–2105.
- Boffetta P, Jourenkova N, Gustavsson P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes Control* 1997;8:444–72.
- Boffetta P, Nyberg F. Contribution of environmental factors to cancer risk. *British Medical Bulletin* 2003; 68: 71–94.
- Borràs JM, Galceran J, Bosch X. Cancer Registry of Tarragona, Catalonia, Spain. A: Parkin DM, Whelan SL, Ferlay J, Raymond L, Young L (eds). *Cancer Incidence in five continents*, vol VII. IARC Sci Publ No. 143, Lyon: Internacional Agency for Research on Cancer; 1997a.
- Borràs JM, Borràs J, Viladiu P, Bosch FX, eds. *Epidemiologia i prevenció del càncer a Catalunya: 1975–1992*. Barcelona: Institut Català d'Oncologia, Generalitat de Catalunya; 1997b.
- Borràs JM, Fernández E, Schiaffino A, Borrell C, La Vecchia C. Pattern of smoking initiation in Catalonia (Spain) from 1948 to 1992. *American Journal Public Health* 2000; 90: 1459–62.
- Borràs JM, Borràs J, Bosch XF, Fernández E, Galceran J, Gispert R, González JR, Izquierdo A, Marcos R, Moreno V, Peris M, Puig X, Sánchez V, Viladiu P. *Càncer Catalunya 2001*; Granollers; Servei de Prevenció i Control del Càncer, Institut Català d'Oncologia, 2001.
- Bostrom CE, Gerde P, Hanberg A et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environmental Health Perspectives* 2002; 110 Suppl 3: 451–388.
- Brennan P, Bogillot O, Cordier S, Greiser E, Vineis P, González CA, Tzonou A, Chang-Claude J, Bolm-Audorf U, Jöckel KH, Donato F, Porru S, Serra C, Wahrendorf J, Hours M, Mannetje A't, Kogevinas M, Boffetta P. Cigarette smoking and bladder cancer in men: a pooled analysis of 11 European case-control studies. *International Journal of Cancer* 2000; 86: 289–294.
- Brooks Sp, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* 1998; 7: 434–455.

- Brown DM, Stone V, Findlay P, MacNee W, Donaldson K. Increased inflammation and intracellular calcium caused by ultrafine carbon black is independent of transition metals or other soluble components. *Occupational Environmental Medicine* 2000;57:685–91.
- Browne K, Gee JB. Asbestos exposure and laryngeal cancer. *British Occupational Hygiene Society* 2000;44:239–50.
- Browne ML, Varadarajulu D, Lewis–Michl EL, Fitzgerald EF. Cancer incidence and asbestos in drinking water, Town of Woodstock, New York, 1980–1998. *Environmental Research*, 2005; 98(2): 224–32.
- Buckley TJ, Lioy PJ. An examination of the time course from human dietary exposure to polycyclic aromatic hydrocarbons to urinary elimination of 1–hydroxypyrene. *British Journal Industrial Medicine* 1992;49:113–24.
- Burnett C, Robinson C, Walker J. Cancer mortality in health and science technicians. *American Journal Ind. Medicine* 1999; 36: 155–58.
- Burnley IH. Disadvantage and male cancer incidence and mortality in New South Wales 1985–1993. *Social Sciences Medicine* 1997; 45(3): 465–76.
- Calle EE, Frumkin H, Henley SJ, Savitz DA, Thun MJ. Organochlorines and breast cancer risk. *CA: Cancer Journal for Clinicians* 2002;52 (5):301–9.
- Carlin BP, Xia H, Devina O, Tolbert P, Mulholland J. Spation–temporal hierarchical models for analyzing Atlanta pediatric asthma ER visit rates. En Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verdine I, West M (eds). *Case Studies in Bayesian Statistics, volumen IV. Nueva York: Springer*, 1999:303–320.
- Cameron A, Trivedi P. Regression analysis of count data. Cambridge, MA: Cambridge University Press, 1998.
- Campbell P, Fehring G, Halaph E, Sloan P, Theis B, Kreiger N. Cancer Care Ontario: Insight on Cancer. *Environmental Exposures and Cancer*. Toronto: Canadian Cancer Society (Ontario Division), 2005.
- Cantor KP. Drinking water and cancer. *Cancer Causes Control* 1997;8:292–308.
- Carstairs, V, Morris, R. *Deprivation and health in Scotland*. Aberdeen: Aberdeen University Press, 1991.
- Centers for Disease Control and Prevention (CDC). Heavy Metals, 2004. http://www.envirohealthaction.org/upload_files/Heavy_Metals.pdf [último acceso 08 Febrero 2005].

- Chan V, Paraskevaides CA, Hale JF. Assessment of thyroid function during pregnancy. *British Journal of Obstetrics and Gynaecology* 1975; 82: 137–141.
- Chang KC, Huang GC, Jones D, Tsao CJ, Lee IYY, Su IJ. Distribution and prognosis of WHO Lymphoma Subtypes in Taiwan Reveals a Low Incidence of Germinal–Center Derived Tumors. *Leukaemia and Lymphoma* 2004a; 45: 1375–1384.
- Chang ET, Zheng T, Weir EG, Borowitz M, Mann RB, Spiegelman D, Mueller NE. Childhood Social Environment and Hodgkin's lymphoma: New Findings from a Population–Based Case–Control Study. *Cancer Epidemiology, Biomarkers & Prevention* 2004b; 13(8): 1361–1370.
- Chellini E, Cherubini M, Chetoni L, Biggeri A, Vannucchi G. Risk of Respiratory cancer around a Sewage Plant in Prato, Italy. *Environmental Health* 2002; 57:548–553.
- Chen A, Zhang J, Zhou L, Gao E, Chen L, Rogan WJ, Wolff MS. DDT serum concentration and menstruation among young Chinese women. Article in press, *Environmental Research* 2004.
- Chiu BC, Cerhan JR, Gapstur SM, Sellers TA, Zheng W, Lutz CT, Wallace RB, Potter JD. Alcohol consumption and non–Hodgkin lymphoma in a cohort of older women. *British Journal Cancer* 1999; 80: 1476–1482.
- Clarke CA, Glaser SL, Keegan T, Stroup A. Neighborhood Socioeconomic Status and Hodgkin's Lymphoma Incidence in California. *Cancer Epidemiology, Biomarkers & Prevention* 2005; 14(6):1441–47.
- Clayton D, Kaldor J. Empirical Bayes estimates for age–standardized relative risks. *Biometrics* 1987; 43:671–681.
- Cliff A, Haggett P. *Atlas of disease distributions: analytic approaches to epidemiological data*. Oxford: Blackwell, 1992.
- Cohen AJ, Pope CA. Lung Cancer and Air Pollution. *Environmental Health Perspectives* 1995; 219–224.
- Crawford FG, Mayer J, Santella RM, Cooper TB, Ottman R, Tsai WY, et al. Biomarkers of environmental tobacco smoke in preschool children and their mothers. *Journal National Cancer Institute* 1994; 86:1398–402.
- Cressie N. *Statistics for Spatial Data*. Nueva York: John Wiley and Sons, 1993.
- Cressie N y Hawkins J. Robust estimation of the variogram. *Journal Mathematics Geology*, 1980 ; 12(2): 115–126.

- Creutin JD , Obled M. Objective analysis and mapping techniques for rainfall fields: an objective comparison. *Water Resources Research* 1982; 18: 413–431.
- Czene K, Tiikkaja S, Hemminki K. Cancer risks in hairdressers: assessment of carcinogenicity of hair dyes and gels. *International Journal on Cancer* 2003; 20:108–112.
- Darby S, Hill D, Auvinen A, Barros-Dlos JM, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *British Medicine Journal* 2005; 63:1–6.
- Darrymple ML, Hudson IL, Ford RPK. Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics and Data Analysis* 2003; 41:491–504.
- Daston GP, Cook JC, Kavlock RJ. Uncertainties for endocrine disruptors: our view on progress. *Toxicology Sciences* 2003;74: 245–52.
- Davies L, Gilbert H. Increasing Incidence of Thyroid Cancer in the United States, 1973–2002. *Journal American Medical Association* 2006, 295(18):2164:2167.
- Davis DL, Bradlow HL, Wolff M, Woodruff T, Hoel DG, Anton-Culver H. Medical hypothesis: xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives* 1993;101:372–7.
- Davis S, Day RW, Kopecky KJ, Mahoney MC, McCarthy PL, Michalek AM, Moysich KB, Onstad LE, Stepanenko VF, Voillequé PG. International Consortium For Research On The Health Effects Of Radiation. Childhood leukaemia in Belarus, Russia, and Ukraine following the Chernobyl power station accident: results from an international collaborative population-based case-control study. *International journal of Epidemiology* 2006; 35:386–396.
- De Boeck, Kirsch-Volders M, Lison D. Cobalt and antimony: genotoxicity and carcinogenicity. *Mutation Research* 2003;533:135–52.
- Deb P, Trivedi PK. The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics* 2002; 21:601–625
- DeLellis RA, Lloyd RV, Heitz PU, Eng C. Pathology and Genetics of Tumours of Endocrine Organs. *WHO Classification of Tumours*. Edit. IARC, octubre 2004
- Denkhaus E, Salnikow K. Nickel essentiality, toxicity, and carcinogenicity. *Critical Reviews in Oncology/Hematology* 2002;42:35–56.

- De Roos AJ, Blair A, Rusiecki JA, Hoppin JA, Svec M, Dosemeci M, Sandler DP, Alavanja MC. Cancer Incidence among Glyphosate-Exposed Pesticide Applicators in the Agricultural Health Study. *Environmental Health Perspectives* 2005; 113: 49-53.
- Díaz MA. *Geoestadística aplicada*. México DF: Instituto de Geofísica, UNAM, 2002.
- Dich J, Zahm SH, Hanberg A, Adami H-O. Pesticides and cancer. *Cancer Causes Control* 1997; 8:420-43.
- Dickman PW, Holm LE, Lundell G, Boice JD Jr, Hall P. Thyroid cancer risk after thyroid examination with 1311: a population-based cohort study in Sweden. *International Journal Cancer* 2003; 106: 580-587.
- Domenech E, Kennedy S, Martínez-Maza O, Sanjose S. Occupational Exposure to Immunologically Active Agents and Risk for Lymphoma. *Cancer Epidemiology Biomarkers Prevention* 2004, 13(11): 1814-1818
- Doll R. The epidemiology of cancer. *Cancer* 1980; 45:2475-2485.
- Dominguez-Berjón, MF, Borrell, C, Benach, J, Pasarín, MI. Medidas de privación material en los estudios de áreas geográficas pequeñas. *Gaceta Sanitaria* 2001; 15(Supl. 4): 23-33.
- Dreassi E. A space-time analysis of the relationship between material deprivation and mortality for lung cancer. *Environmetrics* 2003; 14; 511-21.
- Dreassi E, Biggeri A, Catelan D. Space-time models with time-dependent covariates for the analysis of the temporal lag between socioeconomic factors and lung cancer mortality. *Statistics in medicine* 2005; 24: 1919-1932.
- Duberg AS, Nordström M, Törner A, Reichard O, Strauss R, Janzon R, Bäck E, Ekdahl K. Non-Hodgkin's Lymphoma and Othr Nonhepatic Malignancies in Swedish Patients with Hepatitis C Virus Infection. *Hepatology* 2005; 41(3): 652-659.
- Egan KM, Sosman JA, Blot WJ. Sunlight and Reduced Risk of Cancer: Is The Real Story Vitamin D. *Journal of the National Cancer Institute* 2005; 97(3): 161-163.
- Ekberg-Aronsson M, Nilsson PM, Nilsson JA, Pehrsson K, Lofdahl CG. Socio-economic status and lung cancer risk including histologic subtyping-a longitudinal study. *Lung Cancer* 2006; 51(1):21-9.
- El-Bolkainy MN, Chu EW, eds. *Detection of bladder cancer associated with Schistosomiasis*. The National Cancer Institute. El Cairo, 1981.

Elliot P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. *Statistical Methods in Medical Research* 1995; 4:137–159.

Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives* 2004; 112(9): 998–1006.

Engels EA, Chatieriee N, Cerhan JR, Davis S, Cozen W, Severson RK, Whitby D, Colt JS, Hartge P. Hepatitis C virus infection and Non-hodgkin lymphoma: Results of the NCI-SEER multi-center case-control study. *International Journal of Cancer* 2004a; 111: 76–80.

Engels EA, Chen J, Viscidi RP, Shah KV, Daniel RW, Chalderjee N, Klebanoff MA. Poliovirus Vaccination during Pregnancy, Maternal Sero-conversion to Simian Virus 40, and Risk of Childhood Cancer. *American Journal Epidemiology* 2004b; 160: 306–316.

English D. Geographical epidemiology and ecological studies. En: Elliot P, Cuzick J, English D, Stern R, eds. *Geographical and Environmental Epidemiology*. Oxford University Press, 1992:3–13.

Enquesta de Salut de Catalunya (ESCA). Enquesta de salut, 2002. Generalitat de Catalunya.

<http://www.gencat.net/salut/depsan/units/sanitat/html/ca/plasalut/index.html>

Environmental Protection Agency (EPA). Asbestos.

<http://www.epa.gov/ttnatw01/hlthef/asbestos.html>

[último acceso 08 Febrero 2005].

European Commission. European workshop on the impact of endocrine disruptors on human health and wildlife. Weybridge, UK, Report No. EUR 17549, *Environment and climate research programme*, DG XXI. Brussels, Belgium: European Commission, 1996.

Farahti J, Geling M, Mader U, Morti M, Luster M, Muller JG, Flentje M, Reiners C. Changing trends of incidence and prognosis of thyroid carcinoma in the Franconia region, Germany, from 198–1995. *Thyroid* 2004; 14:141–147.

Feller W. On a general class of “contagious” distributions. *Annals of Mathematical Statistics* 1943; 16:319–329.

Felsot A. Pesticides, estrogen activity and breast cancer: new cause for regulatory concern? *Agrichem Environ News* 1994;97:8–10.

Fernández E, González JR, Borràs JM, Sánchez V, Moreno V, Peris M. Evolución de la mortalidad por cáncer en Cataluña (1975–1998). *Medicina clínica* 2001; 116: 605–609.

- Filleul L, Rondeau V, Vandentorren S, Moual N, Cantagrel A, Annesi-Maesano I, Charpin D, Declercq C, Neukirch F, Paris C, Vervloet D, Brochard P, Tessier JF, Kauffmann F, Baldi I. Twenty five year mortality and air pollution: results from the French PAARC survey. *Occupational Environmental Medicine* 2005; 62: 453–60.
- Flavell K, Constandinou C, Lowe D, Scott K, Newey C, Evans D, Dutton A, Sommons S, Smith R, Crocker J, Young LS, Murray P. Effect of material deprivation on Epstein-Barr virus infection in Hodgkin's disease in the West Midlands. *British Journal of Cancer* 1999; 80 (3–4): 604–8.
- Floret N, Mauny F, Challier B, Arveux P, Cahn JY, Viel JF. Dioxin Emissions from a Solid Waste Incinerator and Risk of Non-Hodgkin Lymphoma. *Epidemiology* 2003; 14: 392–398.
- Forastiere F. Fine particles and lung cancer. *Occupational Environmental Medicine* 2004; 61: 797–798.
- Forssén UM, Rutqvist LE, Ahlbom A, Feychting M. Occupational Magnetic Fields and Female Breast Cancer: A Case-Control Study using Swedish Population Registers and New Exposure Data. *American Journal of Epidemiology* 2005; 161:250–259.
- Fortuny J, Serra C, Bonfill X, Kogevinas M, Sunyer J. Diferencias sociales en el cáncer de vejiga urinaria en Cataluña. *Gaceta Sanitaria* 1999; 13: 208–217.
- Freedman DS, Tolbert PE, Coates R, Brann EA, Kjeldsberg CR. Relation of cigarette smoking to non-Hodgkin's lymphoma among middle-aged men. *American Journal of Epidemiology* 1998; 148: 833–841.
- Fritschi L, Benke G, Hughes AM, Krickler A, Turner J, Vajdic CM, Grulich A et al. Occupational Exposure to Pesticides and Risk of Non-Hodgkin's Lymphoma. *American Journal of Epidemiology* 2005, 162: 849–857.
- Frumkin H, Thun MJ. Arsenic. *CA Cancer Journal Clinique* 2001;51:254–62.
- Gaffney SH, Curriero FC, Strickland PT, Glass GE, Helzlsouer KJ, Breyse PN. Influence of Geographic Location in Modeling Blood Pesticide Levels in a Community Surrounding a U.S. Environmental Protection Agency Superfund Site. *Environmental Health Perspectives* 2005; 113(12):1712–1716.
- Gambolati, G. Volpi G. A conceptual deterministic analysis of the kriging technique in hydrology. *Water Resources. Research.* 1979; 15(3): 625–629.
- Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 1996; 21:256–274.

- Gebregziabher M, Bernstein L, Wang Y, Cozen W. Risk patterns of multiple myeloma in Los Angeles County, 1972–1999 (United States). *Cancer Causes Control* 2006; 17(7): 931–8.
- Gelfand AE, Zhu L, Carlin BP. On the change of support problem for spatial–temporal data. *Biostatistics* 2001; 2:31–45.
- Gelman A. Parameterization and bayesian modeling. *Journal of American Statistical Association* 2004; 99:537–345.
- Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2005; 1(2): 1–19.
- Gelman A, Carlin JC, Stern H and Rubin DB. *Bayesian Data Analysis*. Chapman and Hall, Nueva York, 2005 (2nd edition).
- Georgiadis P, Topinka J, Stoikidou M, Kaila S, Gioka M, Katsouyanni K, et al. Biomarkers of genotoxicity of air pollution (the AULIS project): bulky DNA adducts in subjects with moderate to low exposures to airborne polycyclic aromatic hydrocarbons and their relationship to environmental tobacco smoke and other parameters. *Carcinogenesis* 2001;22: 1447–57.
- Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Nueva York: Chapman & Hall, 1996.
- Glaser SI, Clarke CA, Nugent RA, Stearns CB, Dorfman RF. Social class and risk of Hodgkin’s disease in young–adult women in 1988–94. *International Journal of Cancer* 2002; 98(1): 110–7.
- Goldin LR, Pfeiffer RM, Gridley G, Gail MH, Li X, Mellemkjaer L, Olsen JH, Hemminki K, Linet MS. Familial Aggregation of Hodgkin Lymphoma and Related Tumors. *Cancer* 2004; 100: 1902–1908.
- Goodman M, Morgan RW, Ray R, Malloy CD, Zhao K. Cancer in asbestos–exposed occupational cohorts: a meta–analysis. *Cancer Causes Control* 1999;10:453–65.
- Gorey KM, Vena JE. The association of near poverty status with cancer incidence among black and white adults. *Journal Community Health* 1995; 20(4): 359–366.
- Gneiting T. Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society, Part A*, 1999; 125: 2449–2464.

Greene WH. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working Paper EC-94-10, Department of Economics, New York University, 1994.

Greenwood M, Yule GU. An enquiry into the nature of frequency distributions to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society, Series A* 1920; 83:255-279.

Griffith DA. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. En Arlinghaus SL (ed). *Practical Handbook of Spatial Statistics*, pp. 65-72. Boca Raton, Florida: CRC Press, 1995.

Griffiths H, Malony NC. Does asbestos cause laryngeal cancer? *Clinical Otolaryngology* 2003;28:177-82.

Grimalt JO, Sunyer J, Moreno V, Amaral OC, Sala M, Rosell A, Anto JM, Albaines. Risk excess of soft-tissue sarcoma and thyroid cancer in a community exposed to airborne organochlorinated compound mixtures with a high hexachlorobenzene content. *International Journal on Cancer* 1994; 56: 200-203.

Grossman CM, Nussbaum RH, Nussbaum FD. Cancers among residents downwind of the Hanford, Washington, plutonium production site. *Arch Environmental Health* 2003; 58: 267-274.

Gupta P, Gupta R, Tripathi R. Analysis of zero-adjusted count data. *Computational Statistics and Data Analysis* 1996; 23:207-218.

Hakansson N, Stenlund C, Gustavsson P, Johansen C, Floderus B. Arc and resistance welding and tumours of the endocrine glands: a Swedish case-control study with focus on extremely low frequency magnetic fields. *Occupational Environmental Medicine* 2005; 62: 304-308.

Hardell L, Eriksson M. Is the decline of the increasing incidence of non-Hodgkin lymphoma in Sweden and other countries a result of cancer preventive measures?, *Environmental Health Perspectives* 2003; 111: 1704-1706.

Harras A, Edwards BK, Blot WS, Ries LAG. Cancer Rates and Risks. *National Institutes of Health. National Cancer Institute*, Bethesda 1996; 96, 691.

Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *Journal of Clinical Oncology* 1999; 17(12): 3835-3849.

Harrison RM, Smith DJT, Kibble AJ. What is responsible for the carcinogenicity of PM_{2.5}? *Occupational Environmental Medicine* 2004; 61: 799-805.

Hayes RB. The carcinogenicity of metals in humans. *Cancer Causes Control* 1997;8:371–85.

Haynes, R, Gale, S, Lovett A, & Bentham, G. Unemployment rate as an updatable health needs indicator for small areas. *Journal of Public Health and Medicine* 1996; 18: 27–32.

Heilbron D. *Generalized linear models for altered zero probabilities and overdispersion in count data*. SIMS Technical Report 9, Department of Epidemiology and Biostatistics, University of California, San Francisco, 1989.

Heilbron D. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* 1994; 36:531–547.

Henry-Amar M. *Second cancer after the treatment for Hodgkin's disease: a report from the International Database on Hodgkin's Disease*. *Ann.Oncology*. 1992; 3: 117–128.

Higginson, J. Present trends in cancer epidemiology. In: Morgan JF, ed. *Proceedings of the Eighth Canadian Cancer Research Conference*. New York: *Pergamon*; 1969. pp. 40–47.

Hinde J, Demétrio CGB. Overdispersion: models and estimation. *Computational Statistics and Data Analysis* 1998; 27:151–170.

Holly EA, Lele C, Bracci PM, McGrath MS. Case-control study of non-Hodgkin's lymphoma among women and heterosexual men in the San Francisco Bay Area, California. *American Journal Epidemiology* 1999; 150: 375–389.

Howsam M, Grimalt JO, Guinó E, Navarro M, Martí-Ragué J, Peinado MA, Capellá G, Moreno V. Organochlorine Exposure and Colorectal Cancer Risk. *Environmental Health Perspectives* 2004; 112:1460–1466.

IARC. Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1 to 42. *IARC Monographs on the evaluation of carcinogenic risks to Humans* 1987; supply 7.

IARC. IARC Monographs on the evaluation of carcinogenic risks to humans, Vol 83, *Tobacco smoke and involuntary smoking*. Lyon: IARC, 2002.

INE. Censo de población y viviendas, 2001. <http://www.ine.es> Accedido el 6 junio del 2006.

International Agency for Research on Cancer (IARC). Non-Ionizing Radiation, Part 1: Static and Extremely Low-Frequency (ELF) Electric and Magnetic Fields. *Monography Evaluation Carcinogenesis Risks Human*: Vol 80. Lyon, 2002.

International Agency for Research on Cancer (IARC). IARC monographs on the evaluation of carcinogenic risks to humans. Inorganic and organic lead compounds (Vol. 87, 10–17 February 2004).

<http://monographs.iarc.fr/htdocs/announcements/vol87.htm>

[último acceso 08 Febrero 2005].

International Programme on Chemical Safety (IPCS). Global assessment of the state-of-the-science of endocrine disruptors. Geneva, Switzerland: International Programme on Chemical Safety, *World Health Organization*, 2002. <http://ehp.niehs.nih.gov/who/>

[último acceso 28 de febrero 2005].

Isaaks E, Srivastava M. *An introduction to applied geostatistics*. New York, Oxford University Press. 561 p. 1989.

Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data. *Statistics in Medicine* 2006; 25(12):2136–2159.

Jackson RJ, Delozier DM, Gerasimov G, Borisova O, Garbe PL, Goultchenko L, Shakarishvili G, Hollowell JG, Miller DT. Chernobyl and iodine deficiency in the Russian Federation: an environmental disaster leading to a public health opportunity. *Journal Public Health Policy* 2002; 23: 453–470.

Jaffe ES, Harris NL, Stein H, Vardiman J. *World Health Organisation Classification of tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon: IARC Press, 2001.

Jarup L. Hazards of heavy metal contamination. *British Medical Bulletin* 2003;68:167–82.

Jereczek-Fossa BA, Alterio D, Jassem J, Gibelli B, Tradati N, Orecchia R. Radiotherapy-induced thyroid disorders. *Cancer Treatment Reviews*, 2004; 30:369–384.

Julious SA, Nicholl J, George S. Why do we continue to use standardized mortality ratios for small area comparisons?. *Journal of Public Health Medicine* 2001; 23:40–46.

Kaerlev L, Hansen J, Hansen HL, Nielsen PS. Cancer incidence among Danish seafarers: a population based cohort study. *Occupational Environmental Medicine* 2005; 62: 761–765.

Kahn LH, von Hippel F. Nuclear power plant emergencies and thyroid cancer risk. What New Jersey physicians need to know. *Neurological Journal Medicine* 2004; 101:22–27.

Kaldor JM, Day NE, Band P, Choi NW, Clarke EA, Coleman MP, Hakama M, Koch M, Langmark F, Neal FE. Second malignancies following testicular cancer, ovarian cancer and Hodgkin's disease: an international collaborative study among cancer registries. *International Journal of Cancer* 1987; 39: 571–585.

Kaluzny SP, Vega SC, Cardoso TP, Shelby AA. *S+Spatial statistics*. Nueva York: Mathsoft, 1998.

Kang DH, Rothman N, Poirier MC, Greenberg A, Hsu CH, Schwartz BS, et al. Interindividual differences in the concentration of 1-hydroxypyrene-glucuronide in urine and polycyclic aromatic hydrocarbon-DNA adducts in peripheral white blood cells after charbroiled beef consumption. *Carcinogenesis* 1995;16:1079–85.

Kass DE, Thier AL, Leighton J, Cone JE, Jeffery NL. Developing a Comprehensive Pesticida Health Effects Tracking System for an Urban Setting: New York City's Approach. *Environmental Health Perspectives* 2004; 112: 1419–1423.

Kato I, Koenig KL, Baptiste MS, Lillquist PP, Firzzeria G, Burke JS, Watanabe H, Shore RE. History of antibiotic use and risk of Non-Hodgkin's lymphoma. *International Journal Cancer* 2003; 107: 99–105.

Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, et al. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. *Environmental Health Perspectives* 1996;104 Suppl 4: 715–40.

Kelsall JE, Diggle PJ. Spatial variation in risk of disease – a nonparametric binary regression approach. *Journal of the Royal Statistical Association, Series C*, 1998; 47:559–573.

Kelsall J, Wakefield J. Modelling spatial variation in disease risk: a geostatistical approach. *Journal of the American Statistical Association* 2002; 97(459):692–701.

Kesniiniene A, Cardis E, Tenet V, Ivanov VK, Kurtinaitis J, Malakhova I, Stengrevics A, Tekkel M. Studies of cancer risk among Chernobyl liquidators: materials and methods. *Journal Radiology Protection* 2003; 85: 323–329.

Kirk O, Pedersen C, Cozzi-Lepri A, Antunes F, Miller V, Gattell JM, Kallana C, Lazzarui A, Skinhej P, Barton SE. Non-Hodgkin Lymphoma in HIV-infected patients in the era of highly active antiretroviral therapy. *The American Society of Hematology* 2001; 93: 12.

Knowles DM. Immunodeficiency-associated lymphoproliferative disorders. *Modern Pathology: an official journal of the United States and Canadian Academy of Pathology*, 1999;12: 200–217.

- Kogevinas M, Zock JP, Alsvaro T, Garcia-Villanueva M, Domingo-Domenech E, Kennedy S, Martínez-Maza O, Sanjose S. Occupational Exposure to Immunologically Active Agents and Risk for Lymphoma. *Cancer Epidemiology Biomarkers Prevention* 2004; 13(11): 1814–1818.
- Kreiger N, Quesenberry Jr, Peng T, Horn-Ross P, Stewart S, Brown S, Swallen K, Guillermo T, Suh D, Alvarez-Martinez L, Ward F. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, black, Hispanic, and white residents of the San Francisco Bay Area, 1988–92 (United States). *Cancer Causes and Control* 1999; 10; 525–37.
- Krewski D, Murnett R, Jerrett M, Pope CA, Rainham D, Calle E, Thurston G, Thun M. Mortality and long-term exposure to ambient air pollution: ongoing analyses based on the American Cancer Society cohort. *Journal of Toxicology Environmental Health* 2005; 68(13–14): 1093–109.
- Krol ADG, Cessie SL, Snijder , Kluin JC, Kluin PHM, Noordijk EM. Non-Hodgkin's Lymphoma in The Netherlands: Results from a Population Based Registry. *Leukemia and Lymphoma*, 2003; 44: 451–458.
- Krstev S, Dosemeci M, Lissowska J, Chow W-H, Zatonski W, Ward MH. Occupation and risk of stomach cancer in Poland. *Occupational Environmental Medicine* 2005; 62: 318–324.
- Kung T-M, Ng W-L, Gibson JB. Volcanoes and carcinoma of the thyroid: a possible association. *Archives of Environmental Health* 1981; 36: 265–267.
- Kyrtopoulos SA, Georgiadis P, Autrup H, Demopoulos NA, Farmer P, Haugen A, Katsouyanni K, Lamert B, Ovrebo S, Sram R, Stephanou G, Topinka J. Biomarkers of genotoxicity of urban air pollution. Overview and descriptive data from a molecular epidemiology study on populations exposed to moderate-to-low levels of polycyclic aromatic hydrocarbons: the AULIS project. *Mutation Research* 2001;496: 207–28.
- Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34:1–14.
- Lamont DW, Toal FM, Crawford M. Socioeconomic deprivation and health in Glasgow and the west of Scotland—a study of cancer incidence among male residents of hostels for the single homeless. *Journal Epidemiology Community Health* 1997; 51(6): 668–71.
- Langer P. Polychlorinated biphenyls and the thyroid gland. *Endocrine regulations* 1998; 32: 193–203.
- Langford IH, Leyland AH, Rasbash J, Goldstein H. Multilevel modelling of the geographical distribution of diseases. *Journal of the Royal Statistical Society, Series C* 1999; 48(2):253–268.

- Larijani B, Shirzad M, Mohagheghi MA, Haghpanah V, Mosavi-Jarrahi AR, Tavangar SM, Vassigh AR, Hossein-Nezhad A, Bandarian F, Baradar-Jalili R. Epidemiologic analysis of the Tehran Cancer Institute Data System Registry (TCIDSR). *Asian Pacific Journal of Cancer Prevention* 2004; 5:36–39.
- Law G, Parslow R, McKinney P, Cartwright R. Non-Hodgkin's lymphoma and nitrate in drinking water: a study in Yorkshire, United Kingdom. *Journal Epidemiology Community Health* 1999; 53: 383–384.
- Lawless JF. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics* 1987; 15:209–225.
- Lawson AB, Browne WJ, Vidal-Rodeiro C. *Disease Mapping with WinBUGS and MLwiN*. Chichester, West Sussex: John Wiley & Sons, 2003.
- Lawson AB, Browne WJ, Vidal-Rodeiro C. *Disease Mapping with WinBUGS and MLwiN*. Chichester, West Sussex: John Wiley & Sons, 2003.
- Lawson AB, Clark A. Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine* 2002; 21:359–370.
- Le ND, Sun W, Sidec JV. Bayesian multivariate spatial interpolation with data missing-by-design. *Journal of The Royal Statistics Society, Series B* 196; 59; L 501–510.
- Lertxundi-Manterola A, Saez M. Influencia de la privación sobre la distribución de la incidencia de cáncer en la Regió Sanitària Girona. *Economía y Salud*, 2006; 56:10–11.
- Lertxundi-Manterola A, Saurina C, Saez M, Ocaña R. Construcción de un índice de privación material para los municipios de la Región Sanitaria Girona. *Estudios de Economía Aplicada* 2005; 23(1):1–25.
- Leval L, Harris NL. Variability in immunophenotype in diffuse large B-cell lymphoma and its clinical relevance. *Histopathology*, 2003; 43: 509–528.
- Linnet, MS. Pottern, LM. Familial aggregation of hematopoietic malignancies and risk of non-Hodgkin's lymphoma. *Cancer Research.*, 1992; 52:5468s–5473s.
- Lison D, De Boeck M, Verougstraete V, Kirsch-Volders M. Update on the genotoxicity and carcinogenicity of cobalt compounds. *Occupational Environmental Medicine* 2001; 58: 619–25.
- Little RJA, Rubin DB. *Statistical análisis with missing data*. Nueva York: John Wiley & Sons, 1987.
- Liu S, Semenciw R, Mao Y. Increasing incidence of non-hodgkin's lymphoma in Canada, 1970–1996: Age-period-cohort analysis. *Hematological oncology* 2003; 21: 57–66.

- Lope V, Pollán M, Pérez-Gómez B, Aragonés N, Suárez B, Carrasco JM. Incidencia y mortalidad por cáncer de tiroides en España. *Boletín Epidemiológico* 2004; 12: 161–172.
- Lopez-Abente G, Pollán M, Escolar A, Errezola M, Abraira V. *Atlas de mortalidad por cáncer y otras causas en España, 1978-1992*. Madrid: fundación Científica de la Asociación Española contra el Cáncer; 1996.
- López-Abente G, Ibáñez C. *Aplicación de técnicas de análisis espacial a la mortalidad por cáncer en Madrid. Documentos Técnicos de Salud Pública* 66, Madrid: Dirección General de Salud Pública, Consejería de Sanidad, Comunidad de Madrid, 2001.
- Lopez-Abente G, Pollán M, Aragonés N, Perez-Gomez B, Hernandez-Barrera V, Lope V, Suarez-Rodriguez B, Cárdbaba M, Cerdà T, Salas D. *La situación del cáncer en España*, Madrid; Centro de Publicaciones, Ministerio de Sanidad y Consumo, 2005.
- Lundell M, Hakulinen T, Hala LE. Thyroid Cancer after Radiotherapy for Skin Hemangioma in Infancy. *Radiation Research* 1994; 140: 334–339.
- McNally RJ, Alston RD, Cairns DP, Eden OB. Geographical and ecological analyses of childhood acute leukaemias and lymphomas in north-west England. *British Journal of Haematology* 2003; 123: 60–5.
- McWhorter WP, Schatzkin AG, Horm JW, Brown CC. Contribution of socioeconomic status to black/white differences in cancer incidence. *Cancer* 1989; 1;63 (5): 982–987.
- Mahoney MC, Lawvere S, Falkner KL, Averkin YI, Ostapenko VA, Michalek AM, Moysich KB, McCarthy PL. Thyroid cancer incidence trends in Belarus: examining the impact of Chernobyl. *International Journal Epidemiology* 2004; 33:1025–1033.
- Mannetje A't, Kogevinas M, Chang-Claude J, Cordier S, González CA, Hours M, Jöckel KH, Bolm-Audorf U, Lynge E, Porru S, Donato F, Ranft U, Serra C, Tzonou A, Vineis P, Wahrendorf J, Boffetta P. Occupation and bladder cancer in European women. *Cancer Causes Control* 1999; 10:209–217.
- Mans DR, Mohamedradja RN, Hoebal AR, Rampadarath R, Joe SS, Wong J, Ramautar P, Mahabier R, Vrede MA. Cancer incidence in Suriname from 1980 through 2000 a descriptive study. *Tumori* 2003; 89:368–376.
- Marcos-Gragera R. *Epidemiología de las neoplasias linfoides. Incidencia atribuible al sida*. Tesis Doctoral, Universitat Autònoma de Barcelona, Barcelona 2006.
- Markaki I, Linos D, Linos A. The influence of dietary patterns on the development of thyroid cancer. *European Journal Cancer* 2003; 39: 1912–1919.

Matheron G. 1969. Le Krigeage Universal. *Cahiers du Centre de Morphologie Mathematique*, No. 1, Fontainebleau, France.

Matheron G, Kleingeld WJ. *The Evolution of Geostatistics*, APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries, Volume 3. Geostatistics. Johannesburg, SAIMM, pp. 9–12. 1987.

Mekori YA, Weitzman GL, Galli SJ. Reevaluation of reserpine-induced suppression of contact sensitivity. Evidence that reserpine interferes with T lymphocyte function independently of an effect on mast cells. *Journal of Experimental Medical sciences* 1985; 162: 1935–1953.

Menvielle G, Luce D, Févotte J, Bugel I, Alomon C, Goldberg P, Billon-Galland P. Occupational exposures and lung cancer in New Caledonia. *Occupational Environmental Medicine* 2003; 60: 584–589.

Menvielle G, Luce D, Geoffroy-Perez B, Chastang JF, Leclerc A. Social inequalities and cancer mortality in France, 1975–1990. *Cancer Causes Control* 2005; 16(5): 501–13.
Mestre A. *Report on rainfall spatialisation*. Madrid: Cost 718 WG1. Report on rainfall spatialisation, 2000.

Mikoczy Z, Welinder H, Tinnerberg H, Hagmar L. Cancer incidence and mortality of isocyanate exposed workers from the Swedish polyurethane foam industry: updated findings 1959–98. *Occupational Environmental Medicine* 2004; 61: 432–437.

Miller AB, To T, Agnew DA, Wall C, Green LM. Leukemia following occupational exposure to 60-Hz electric and magnetic fields among Ontario electric utility workers. *American Journal Epidemiology* 1996; 144: 150–60.

Mollie A. Bayesian mapping of disease. En Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Nova York: Chapman & Hall, 1996, pp. 359–379.

Moore LE, Lu M, Smith A. Childhood Cancer Incidence and Arsenic Exposure in Drinking Water in Nevada. *Environmental Health* 2002; 57: 201–206.

Morello-Forsch R, Jesdale BM. Separate and Unequal: Residential Segregation and Estimated Cancer Risks Associated with Ambient Air Toxics in U.S. Metropolitan Areas. *Environmental Health Perspectives* 2006; 114(3): 386–93.

Motykwicz G, Michalska J, Pendzich J, Malusecka E, Strozyk M, Kalinowska E, et al. A molecular epidemiology study in women from Upper Silesia, Poland. *Toxicology Letters* 1998; 96–97: 195–202.

- Moysich BM, Ravi J M, Michalek AM. Chernobyl-related ionising radiation exposure and cancer risk: an epidemiological review. *Lancet Oncology* 2002; 3: 269–279.
- Malkasian GD, Mayberry WE. Serum total and free thyroxine and thyrotropin in normal and pregnant women, neonates, and women receiving progestogens. *American journal of obstetrics and gynecology* 1970; 108: 1234–1238.
- Morgensztern D, Rosado MF, Silva O, Santos ES, Abdullah SA, Goodman M, Hamilton-Nelson K, Rosenblatt J, Lossos I. Prevalence of Hepatitis C Infection in Patients with Non-Hodgkin's Lymphoma in South Florida and Review of the Literature. *Leukemia & Lymphoma* 2004; 45(12): 2459–2464.
- Morgner A, Bayerdorffer E, Neubauer A, Stolte M. Gastric MALT lymphoma and its relationship to Helicobacter pylori infection: management and pathogenesis of the disease. *Microscopy Research and Technique*, 2000; 48: 349–356.
- Moshhammer H, Neuberger. Lung cancer and dust exposure: results of a prospective cohort study following 3260 workers for 50 years. *Occupational Environmental Medicine* 2004; 61: 157–162.
- Moysich BM, Ravi J M, Michalek AM. Chernobyl-related ionising radiation exposure and cancer risk: an epidemiological review. *The lancet Oncology* 2002; 3: 269–279.
- Mueller, N. The epidemiology of HTLV-I infection, *Cancer Causes Control*, 1991; 2: 37–52.
- Mueller, N. Overview of the epidemiology of malignancy in immune deficiency, *Journal of Acquired Immune Deficiency Syndromes* 1999; 21:5–10.
- Mullahy J. Specification and testing of some modified count data models. *Journal of Econometrics* 1986; 33:341–365.
- Mullahy J. Heterogeneity, excess zeros, and the structure of count data models. *Journal of Health Econometrics* 1997; 12:337–350.
- Nafstad P, Haheim LL, Oftedal B, Gram F, Holme I, Hjermann I, Leren P. Lung cancer and air pollution: a 27 year follow un of 16209 Norwegian men. *Thorax* 2003; 58: 1071–1076.
- Nagataki S, Nystrom E. Epidemiology and primary prevention of thyroid cancer. *Thyroid* 2002; 12: 889–96.
- Naresh KN, Agarwal B, Nathwani BN, Diebold J, McLennan KA, Muller-Hermelink KH, Armitage JU, Weisenburger DD. Use of the World Health Organization (WHO) classification of non-Hodgkin's lymphoma in Mumbai, India: a review of 200 consecutive cases by a panel of five expert hematopathologist. *Leukaemia and Lymphoma* 2004; 45: 1569–1577.

National Toxicology Program (NTP). Report on carcinogens, Tenth Edition. U.S. Department of Health and Human Services. Public Health Service, National Toxicology Program, Diciembre 2002.

Negri E, Little A, Boiocchi M, Vecchia C, Franceschi S. B-Cell Non-Hodgkin's Lymphoma and hepatitis C virus infection: a systematic review. *International Journal of Cancer* 2004; 111: 1-8.

Neinrich J, Mielck A, Schafer I, Mey W. Social inequality and environmentally-related diseases in Germany: review of empirical results. *Soz Praventivmed* 2000; 45(3): 106-18.

Novo A, Schafer JL. *The norm package*.
<http://www.stat.psu.edu/~jls/misoftwa.html#aut>, 2003.

Ocaña-Riola R, Sanchez-Cantalejo C, Rosell J, Sanchez-Cantalejo E, Daponte A. Socio-economic level, farming activities and risk of cancer in small areas of Southern Spain. *European Journal of Epidemiology* 2004; 19(7): 643-50.

Ocaña-Riola R, Saurina C, Sanchez-Cantalejo C, Lertxundi A, Fernández-Ajuria A, Saez M, Barceló MA, Daponte A, Ruiz-Ramos M, Martínez JM, Benach J. Area deprivation and mortality in the capital cities of Andalusia and Catalonia (Spain). (Enviado al *Journal of Epidemiology and Community Health*).

Olaya-Contreras P, Rodríguez-Villamil J, Posso-Valencia HJ, Cortez JE. Organochlorine exposure and breast cancer risk in Colombian women. *Cadernos de saude publica* 1998; 14 Suppl 3:125-32.

Oliver MA, Muir KR, Parkes SE, Webster R. Methods for investigating localized clustering of disease. Geostatistics for determining the risk of rare disease. *IARC Science Publication* 1996; 135:227-247.

OMS. Clasificación Internacional de Enfermedades (CIE) 9ª revisión. Barcelona: OMS, 1985.

OMS. Clasificación Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE) 10ª revisión. Washington: OPS, OMS, 1995.

Park D, Gelman A, Bajumi J. Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. *Political Analysis* 2004; 12:375-385.

Parkin DM, Muir SL, Whelan Y-T, Ferlay J, Powell J. Cancer incidence in five continents. *IARC Scientific Publications*. Lyon; 1992, n 120.

- Peláez S, Hierro I, Oña S, Alonso L, Matilla A. Relación entre la exposición a pesticidas y el desarrollo de carcinoma urotelial vesical superficial de bajo grado. *Medicina Clínica* 2004; 123: 571–574.
- Pershagen G, Simonato L. Epidemiological evidence on outdoor air pollution and cancer. In: Tomatis L, editor. *Air Pollution and Human Cancer*. Berlin: Springer-Verlag; 1990. pp. 63–94.
- Piantadosi S, Byar DP, Green SB. The ecological fallacy. *American Journal of Epidemiology* 1988; 127:893–904.
- Polednak AP. Cancer incidence in the Puerto Rican-born population of Connecticut. *Cancer* 1992; 1; 70(5): 1172–6.
- Pollock AM, Vickers N. Breast, lung and colorectal cancer incidence and survival in South Thames Region, 1987–1992: the effect of social deprivation. *Journal Public Health Medicine* 1997; 19(3): 288–94.
- Poole C, Greenland S, Luetters C, Kelsey JL, Mezei G. Socioeconomic status and childhood leukaemia: a review. *International Journal of Epidemiology* 2006; 35: 370–384.
- Potzsch C, Schaefer, HE, Lubbert, M. Familial and metachronous malignant lymphoma: absence of constitutional mutations, *American Journal Hematology* 1999; 62: 144–149.
- Preston-Martin S, Jun F, Duda MJ, et al. A case-control study of thyroid cancer in women under age 55 in Shanghai (People's Republic of China). *Cancer Causes Control* 1993; 4: 431–440.
- Pujolar EA, Gonzalez CA, Lopez Abente G, Errezola M, Izarzugaza I, Nebot M, Riboli E. Bladder cancer and coffee consumption in smokers and non-smokers in Spain. *International Journal of Epidemiology* 1993;22:38–44.
- Pye S, Stedman J, Adams M, King K. Further Analysis of NO₂ and PM₁₀ Air Pollution and Social Deprivation. Belfast: Department for Environment, Food and Rural Affairs, The National Assembly for Wales and Department of the Environment in Northern Ireland, December 2001 (report).
- Quackenboss JJ, Pellizzari ED, Shubat P, Whitmore RW, Adgate JL, Thomas KW, Freeman NC, Stroebel C, Lioy PJ, Clayton AC, Sexton K. Design strategy for assessing multi-pathway exposure for children: the Minnesota Children's Pesticide Exposure Study (MNCPEs). *Journal of Exposure Analysis and Environmental Epidemiology* 2000;10:145–58.

Rafnsson V, Sulem P. Cancer incidence among marine engineers, a population-based study (Iceland). *Cancer Causes Control* 2003; 14(1): 29–35.

Rahu M. Health effects of the Chernobyl accident: fears, rumours and the truth. *European Journal Cancer* 2003; 39: 295–299.

Reid A, Klerk N, Ambrosini G, Olsen N, Pang SC, Musk AW. The additional risk of malignant mesothelioma in former workers and residents of Wittenoom with benign pleural disease or asbestosis. *Occupational Environmental Medicine* 2005, 62: 665–669.

Reicha V, Kulich M, Reicha R, Shore DL, Sandler DP. Incidence of Leukemia, Lymphoma, and Multiple Myeloma in Czech Uranium Miners: A Case-Cohort Study. *Environmental Health Perspectives* 2006, 114(6):818–822.

Ribeiro PJ, Diggle PJ. Geo R: a package for geostatistical analysis *RNEWS*, 2001; 1(2): 15–18.

Ribeiro PJ. The geoR Package. <http://www.est.ufpr.br/geoR> (último acceso 28 de marzo de 2006)

Ribera JM, Navarro JT. Linfomas en pacientes con infección por el VIH. Las cosas han cambiado para bien. *Enfermedad Infección Microbiológica Clínica* 2004; 22(6): 313–314.

Riboli E, Kaaks R. The EPIC projects: Rationale and Study design. *International Journal of Epidemiology* 1997; vol26: S8–S14.

Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives* 2004; 112 (9): 1016–1025.

Richardson DB, Ashmore JP. Investigating time patterns of variation in radiation cancer associations. *Occupational Environmental Medicine* 2005; 62: 551–558.

Ridout M, Demétrio CGB, Hinde J. *Models for count data with many zeros*. International Biometric Conference, Cape Town, December 1998.

Romundstad P, Haldorsen T, Andersen A. Cancer incidence and cause specific mortality among workers in two Norwegian aluminum reduction plants. *American Journal of Industrial Medicine* 2000; 37(2): 175–183.

Röösli M y Künzli N. Comentario: Magnetic field exposure and childhood leukaemia—moving the research agenda forward. *International Journal of Epidemiology* 2006;35: 407–408.

- Rubin DB. Inference and missing data. *Biometrika* 1976; 63: 581–592.
- Rubin DB. *Multiple imputation for nonresponse in surveys*. Nueva York: John Wiley & Son, 1987.
- Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of American Statistical Association* 1996; 91:473–489.
- Saez M, Saurina C. *Estadística y epidemiología espacial*. Girona: Grup de Recerca en Estadística, Economia Aplicada i Salut, GRECS. Universitat de Girona, 2006.
- Safe SH, Zacharewski T. Organochlorine exposure and risk for breast cancer. *Progress in clinical and biological research* 1997;396: 133–45.
- Salud (CIE) 10^a revisión. Washington: OPS, OMS, 1995.
- Samet JM, Cohen AJ. Air pollution and lung cancer. In: Holgate ST, Samet JM, Koren HS, Maynard R, eds. *Air pollution and health*. London: Academic Press, 1999: 841–64.
- Satarug S, Baker JR, Urbenjapol S, Haswell–Elkins M, Reilly PEB, Williams DJ, et al. A global perspective on cadmium pollution and toxicity in non–occupationally exposed population. *Toxicology Letters* 2003; 137: 65–83.
- Schafer DW, Lubin JH, Ron E, Stovall M, Carroll RJ. A Reanalysis of Thyroid Neoplasms in the Israelí Tinea Capitis Study Accounting for Dose Uncertainties. *Radiation Research* 2004; 161: 359–368.
- Schafer JL. *Analysis of Incomplete Multivariate Data*. Nueva Cork: Chapman & Hall, 1996, pp. 359–367.
- Scherer G, Frank S, Riedel K, Meger–Kossien I, Renner T. Biomonitoring of exposure to polycyclic aromatic hydrocarbons of nonoccupationally exposed persons. *Cancer Epidemiology Biomarkers Prevention* 2000;9: 373–80.
- Scherr PA, Muller NE. Non–Hodking Lymphomas, 2^a edn. *Cancer epidemiology and prevention*, New York: Oxford University Press, 1996: 920–945.
- Seilkop SK, Oller AR. Respiratory cancer risks associated with low–level nickel exposure: an integrated assessment based on animal, epidemiological, and mechanistic data. *Regulation Toxicology Pharmacology* 2003;37:173– 90.
- Selvin HC. Durkheim’s ‘suicide’ and problems of empirical research. *American Journal of Sociology* 1958; 63: 607–619.

Serra C, Bonfill X, Sunyer J, Urrútia G, Turuguet D, Bastús R, Roqué M, Mannetje A't, Kogevinas M. Working Group on the Study of Bladder Cancer in the County of Vallès Occidental. Bladder cancer in the textile industry. *Scandinavian Journal of Work, Environment and Health* 2000; 26: 474–479.

Serra C. *Ocupació i cancer de bufeta urinària al vallès occidental*. La Salut Laboral Sabadell, Institut Català de la Salut-Ajuntament de Sabadell 2002; 20–26

Shaddick G, Wakefield JC. Modelling multivariate pollutant data at multiple sites. *Applied Statistics* 2002; 51: 351–372.

Shakhtarin VV, Tsyb AF, Stepanenko VF, Orlov MY, Kopecky KJ, Davis S. Iodine deficiency, radiation dose, and the risk of thyroid cancer among children and adolescents in the Bryansk region of Russia following the Chernobyl power station accident. *International Journal Epidemiology* 2003; 32: 584–591.

Sharpe M. Rising hormones. *Journal of Environmental Monitoring:JEM* 1999; 1:5N–7N.

Silverman DT, Morrison AS, Devesa SS. Bladder cancer. In: Schottenfeld D, Fraumeni J (eds). *Cancer epidemiology and prevention*. New York: Oxford University Press, 1996; 1156–1179.

Simcox NJ, Fenske RA, Wolz SA, Lee IC, Kalman DA. Pesticides in Household Dust and Soil: Exposure Pathways for Children in Agricultural Families. *Environmental Health Perspectives* 1995;103:1126–34.

Smedby KE, Hjalgrim H, Melbye M, Torrany A. Ultraviolet Radiation Exposure and Risk of Malignant Lymphomas. *Journal of National Cancer Institute* 2005; 97: 199–209.

Sorahan T, Esmen NA. Lung cancer mortality in UK nickel–cadmium battery workers, 1947–2000. *Occupational Environmental Medicine* 2004; 61: 108–116.

Sorensen M, Autrup H, Hertel O, Wallin H, Knudsen LE, Loft S. Personal exposure to PM_{2.5} and biomarkers of DNA damage. *Cancer Epidemiology Biomarkers Prevention* 2003; 12: 191–196.

Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS User Manual. Version 1.4*. Gener 2003. <http://www.mrc-bsu.cam.ac.uk/bugs>.

Spiegelhalter DJ. Two brief topics on modelling with WinBUGS. *IceBUGS: Finland, 11th–12th February* 2006.

Spiegelhalter DJ, Thomas A, Best NG, Gilks WR, Lunn D. *BUGS: Bayesian inference using Gibbs sampling*. Cambridge, UK: MRC Biostatistics Unit, 1994, 2003.

Spiegelhalter DJ, Best NG, Carlin BP, Van del Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002; 64: 583–63.

Steenland K, Boffetta P. Lead and cancer in humans: where are we now? *American Journal of Industrial Medicine* 2000;38:295–9.

Stezhko VA, Buglova EE, Danilova LI, Drozd VM, Krysenko NA, Lesnikova NR, Minenko VF, Ostapenko VA, Petrenko SV, Polyanskaya ON, Rzhetski VA, Tronko MD, Bobylyova OO, Bogdanova TI, Ephstein OV, Kairo IA, Kostin OV, Likhtarev IA, Markov VV, Oliynik VA, Shpak VM, Tereshchenko VP, Zamotayeva GA, Beebe GW, Bouville AC, Brill AB, Burch JD, Fink DJ, Greenebaum E, Howe GR, Luckyanov NK, Masnyk I J, McConnell RJ, Robbins J, Thomas TL, Voilleque PG, Zablotska LB. A cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: objectives, design and methods. *Radiation Research* 2004; 161:481–492.

Stone, R. Environmental estrogens stir debate (News & Comment). *Science*. 1994;265:308–310.

Study in New Caledonia (South Pacific), a Very High Incidence Area. *American Journal Epidemiology* 2005; 161: 1056–1065.

Svec MA, Ward MH, Dosemeci M, Checkoway H, De Roos AJ. Risk of lymphatic or haematopoietic cancer mortality with occupational exposure to animals or the public. *Occupational Environmental Medicine* 2005; 62: 726–735.

Takagi A, Watanabe S, Igarashi M, Koike J, Hasumi K, Deguchi R, Koga Y, Miwa T. The effect of *Helicobacter pylori* on cell proliferation and apoptosis in gastric epithelial cell lines. *Alimentary Pharmacology & Therapeutics* 2000; 14: 188–192.

Takahashi T, Schoemaker MJ, Trott KR, Simon SL, Fujimori K, Nakashima N, Fukao A, Saito H. The relationship of thyroid cancer with radiation exposure from nuclear weapon testing in the Marshall Islands. *Journal Epidemiology* 2003; 13: 99–107.

Tang D, Warburton D, Tannenbaum SR, Skipper P, Santella RM, Cereijido GS, et al. Molecular and genetic damage from environmental tobacco smoke in young children. *Cancer Epidemiology Biomarkers Prevention* 1999;8:427–31.

Tardon A, Martínez B, Kogevinas M, Cueto A. Cáncer de vejiga en Asturias: incidencia y evolución (1982–1993). *Gaceta Sanitaria* 1999; 13: 346–352.

Tavani A, Pignatelli A, Negri E, Franceschi S, Serraino D, Carbone A, La Vecchia C. Diet and risk of lymphoid neoplasms and soft tissue sarcomas, *Nutritional Cancer* 1997; 27: 256–260.

- Thomas RW. *Spatial epidemiology*. Londres: Pion, 1990.
- Tolbert PE. Oils and Cancer. *Cancer Causes Control* 1997;8: 386–405.
- Tondel M, Hjalmarsson P, Hardell L, Carlsson G, Axelson O. Increase of regional total cancer incidence in north Sweden due to the Chernobyl accident? *Journal Epidemiology Community Health* 2004; 58: 1011–1016.
- Tonellato SF. A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Applied Statistics* 2001; 50: 187–200.
- Townsend, P. Widening inequalities of health in Britain: A rejoinder to Rudolph Klein. *International Journal of Health Services* 1990; 20: 363–372.
- Truong T, Orsi L, Dubourdieu D, Rougier Y, Hemon D, Guénel P. Role of Goiter and of Menstrual and Reproductive Factors in Thyroid Cancer : A Population-based Case-Control Study in New Caledonia (South Pacific, a Very High Incidence Area. *American Journal Epidemiology* 2005; 161: 1056–1065.
- Tsai HT, Wu MT, Hauser R, Rodrigues E, Ho CK, Liu CL, et al. Exposure to environmental tobacco smoke and urinary 1-hydroxypyrene levels in preschool children. *Kaohsiung Journal Medicine Sciences* 2003;19:97–104.
- Tsuda H, Naito A, Kim CK, Fukamachi K, Nomoto H, Moore MA. Carcinogenesis and its modification by environmental endocrine disruptors: in vivo experimental and epidemiological findings. *Japanese Journal of Clinical Oncology* 2003; 33:259–70.
- Tukiendorf A, Krasowski G, Rybak Z. Thyroid cancer morbidity in Opole province, Poland, after Czernobyl disaster. *Central European journal of public health* 2003; 11: 98–101.
- UERCG. *Memòria 2003*. Girona: Unitat Epidemiologia Registre Càncer Girona, 2004. Institut Català d'Oncologia.
- UERCG. *Memòria 2004*. Girona: Unitat Epidemiologia Registre Càncer Girona, 2005. Institut Català d'Oncologia.
- Underwood W, Duna RL, Willinams C, Lee CT. Gender and geographic influence on the racial disparity in bladder cancer mortality in the US. *Journal of American College Surgery* 2006; 202(2): 284–90.
- Urrutia G, Serra C, Bonfill X, Bastús R. Incidencia del cáncer de vejiga urinaria en un área industrializada de España. *Gaceta Sanitaria* 2002; 16(4): 291–7.

Vajdic CM, Kricker A, Giblin M, McKenzie J, Aitken J, Giles GG, et al. Sun exposure predicts risk of ocular melanoma in Australia. *International Journal of Cancer* 2002;101:175–82.

Van Beers WCM. Kriging metamodeling in discrete-event simulation: an overview. En Kuhl ME, Steiger NM, Armstrong FB, Joines JA (eds). *Proceedings of the 2005 Winter Simulation Conference*, 2005, pp. 202–208.

Verger P, Catelinois O, Tirmarche M, Cherie-Challine L, Pirard P, Colonna M, Hubert P. Thyroid cancers in France and the Chernobyl accident: risk assessment and recommendations for improving epidemiological knowledge. *Health Physic* 2003; 85: 323–329.

Viel JF, Arveux P, Baverel J, Cahn JY. Soft Tissue Sarcoma and Non Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology* 2000;152(1):13–9.

Viladiu P, Izquierdo A, Marcos R, et al. *Epidemiologia del càncer a Girona*: Unitat de Registre del Càncer de Girona, Institut d'Assistència Sanitària, 2000.

Villanueva CM, Kogevinas M, Grimalt JO. Cloración del agua potable en Espanya y càncer de vejiga. *Gaceta Sanitaria* 2001; 15: 48–53.

Villeneuve PJ, Agnew DA, Miller AB, Corey PN. Non-Hodgkin's lymphoma among electric utility workers in Ontario: the evaluation of alternate indices of exposure to 60 Hz electric and magnetic fields. *Occupational Environmental Medicine* 2000;57:249–57.

Vineis P, Crosignani P, Sacerdote C, Fontana A, Masala G, Miligi L, Nanni O, Ramazzotti V, Rodella S, Stagnaro E, Tumino R, Vignani C, Vindigni C, Costantini AS. Hematopoietic cancer and peptic ulcer: a multicenter case-control study. *Carcinogenesis* 1999; 20:1459–1463.

Vineis P, HOek G, Krzyzanowski M, Vigna-Taglianti F, Veglia F, Airoidi L, Autrup H, Dunning A, Garte S, Hainaut P, Malaveille C, Matullo G, Overvad K, Raaschou-Nielsen O, Clavel-Chapelon F, Linseisen J, Boeing H, Trichopoulou A, Palli D, et al. Air pollution and risk of lung cancer in a prospective study in Europe. *International Journal of Cancer* 2006; 119(1): 169–74.

Vose JM, Chiu BC, Cheson B, Cancey J, Wright J. Update on Epidemiology and Therapeutics for Non-Hodgkin's Lymphoma. *Hematology* 2002; 241–261.

Wakefield J, Shaddick G. Health-exposure modelling and the ecological fallacy. *Biostatistics* 2006; 7(3): 438–455.

- Walter SD. Disease mapping: a historical perspective. En: Elliott P, Wakefield J, Best N, Briggs DJ (eds). *Spatial epidemiology: methods and applications*. Oxford: Oxford University Press, 223–252, 2000.
- Walter SD, Birmie SE. Mapping mortality and morbidity patterns: an international comparison. *International Journal of Epidemiology* 1991; 20:678–689.
- Ward MH, Lubin J, Giglierano J, Colt JS, Wolter C, Begiroglu N, Camann D, Hartge P, Nuckols JR. Proximity to Crops and Residential Exposure to Agricultural Herbicides in Iowa. *Environmental Health Perspectives* 2006, 114(6):893–897.
- Weeke J, Hansen AP. Serum TSH and serum T levels during normal menstrual cycles and during cycles on oral contraceptives. *Acta Endocrinológica* 1975; 79:431–438.
- Whitrow MJ, Smith BJ, Pilotto LS, Pisaniello D, Nitschke M. Environmental exposure to carcinogens causing lung cancer: Epidemiological evidence from the medical literature. *Respirology* 2003; 8: 513–21.
- Whyatt RM, Santella RM, Jedrychowski W, Garte SJ, Bell DA, Ottman R, et al. Relationship between ambient air pollution and DNA damage in Polish mothers and newborns. *Environmental Health Perspectives* 1998;106 Suppl 3:821–6.
- Wing S and Richardson DB. Age at exposure to ionising radiation and cancer mortality among Hanford workers: follow up through 1994. *Occupational Environmental Medicine* 2005; 62: 465–472.
- Wolff MS, Toniolo PG, Lee EW, Rivera M, Dubin N. Blood levels of organochlorine residues and risk of breast cancer. *Journal National Cancer Institute* 1993; 85:648–52.
- World Health Organization. Chemical hazards in drinking-water. WHO 2004. [cited 16 December 2004] [último acceso 23 Marzo 2005]
http://who.int/water_sanitation_health/dwg/chemicals/polyaromahydrocarbons/en/
- Zablotska LB, Chak A, Das A, Neugut AI. Increased Risk of Squamous Cell Esophageal Cancer after Adjuvant Radiation Therapy for Primary Breast Cancer. *American Journal Epidemiology* 2005; 161: 330–337.
- Zeka A, Eisen EA, Kriebel D, Gore R, Wegman DH. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occupational Environmental Medicine* 2004; 61: 426–431.
- Zhang S, Hunter DJ, Rosner BA, Colditz GA, Fuchs CS, Speizer FE, Willett WC Dietary fat and protein in relation to risk of non-Hodgkin's lymphoma among women. *Journal National Cancer Institute* 1999; 91: 1751–1758.

Zhu K, Levine RS, Gu Y, Brann, EA, Hall I, Caplan LS, Baum MK. Non-Hodgkin's lymphoma and family history of malignant tumors in a case-control study. *Cancer Causes Control* 1998; 9: 77-82.

Zhu L, Carlin BP, Gelfand AE. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics* 2003; 14: 537-557.

Zidek JV, White R, Lee ND, Sun W, Burnett RT. Imputing unmeasured explanatory variables in environmental epidemiology with application to health impact analysis of air pollution. *Environmental and Ecological Statistics* 1998; 5:99-115.

Anexo 1. Modelos lineales generalizados (GLM)

Como su nombre indica, los *modelos lineales generalizados*, *GLM* (McCullagh y Nelder, 1989), generalizan el modelo de regresión lineal permitiendo utilizar el mismo tipo de modelización, especificación, estimación y diagnóstico, aquí radica su atractivo, para variables dependientes no normales, como por ejemplo: variables contadoras (distribución de Poisson); variables dicotómicas (distribución binomial), etc.

El objetivo de los GLM es el de describir la dependencia de la respuesta (o variable dependiente), Y , respecto a las variables explicativas

$$E(Y|x) = \mu$$

Los modelos GLM son miembros de la familia exponencial con función de densidad de la forma:

$$f(y) = e^{\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]}$$

en la que θ es un parámetro de localización (no necesariamente la media) y ϕ un parámetro de dispersión (únicamente aparece en distribuciones que se caracterizan por dos parámetros, como por ejemplo la normal).

La forma de las funciones $a()$, $b()$ y $c()$ difiere para cada una de las distribuciones particulares consideradas.

Distribución	$a(\phi)$	$b(\theta)$	$c(y)$
Bernouilli	ϕ	$\ln(1 + e^\theta)$	1
Binomial	ϕ/n	$\ln(1 + e^\theta)$	$\ln[n!/((ny)!(n-ny)!)]$
Poisson	ϕ	e^θ	$-\ln(y!)$
Poisson con offset	ϕ/n	e^θ	$yn \ln(yn) - \ln(yn!)$

El logaritmo de la función de verosimilitud se expresa como:

$$l(\theta; \Phi; y) = \log f_y(y; \theta, \Phi) = [y\theta - b(\theta)]/a(\Phi) + c(y, \Phi)$$

Se verifica que:

$$\frac{\partial l}{\partial \theta} = [y - b'(\theta)]/a(\Phi) \quad \frac{\partial^2 l}{\partial \theta^2} = -b''(\theta)/a(\Phi)$$

Según la teoría de la verosimilitud se verifica:

$$1) E_\theta \left(\frac{\partial l}{\partial \theta} \right) = 0 \quad 2) E_\theta \left(\frac{\partial^2 l}{\partial \theta^2} \right) + \text{var}_\theta \left(\frac{\partial l}{\partial \theta} \right) = 0$$

donde $\text{var}_\theta \left(\frac{\partial l}{\partial \theta} \right) = E_\theta \left(\frac{\partial l}{\partial \theta} \right)^2$

Así pues,

$$1) E_\theta \left(\frac{\partial l}{\partial \theta} \right) = [\mu - b'(\theta)]/a(\Phi) = 0 \Rightarrow E(Y) = \mu = b'(\theta)$$

2)

$$-\frac{b''(\theta)}{a(\Phi)} + E \left[\frac{Y - b'(\theta)}{a(\Phi)} \right]^2 = -\frac{b''(\theta)}{a(\Phi)} + \frac{E[Y - E(Y)]^2}{a^2(\Phi)} = -\frac{b''(\theta)}{a(\Phi)} + \frac{\text{Var}(Y)}{a^2(\Phi)} = 0 \Rightarrow \text{Var}(Y) = b''(\theta)a(\Phi)$$

Por tanto, la media, μ , no es más que la primera derivada de la función $b(\theta)$ ($\mu=b'(\theta)$), mientras que la varianza de Y es proporcional a la segunda derivada de $b(\theta)$ ($\text{var}(Y)=a(\phi)b(\theta)''$); ϕ se denomina parámetro de dispersión.

Los resultados obtenidos para la media y la varianza de las distintas distribuciones adoptan la forma:

Distribución	Media $\mu=b'(\theta)$	Varianza en términos de θ , $a(\phi)b(\theta)''$	Varianza en términos de μ
Bernouilli	$\frac{1}{1+e^{-\theta}}$	$\phi \frac{e^{\theta}}{(1+e^{\theta})^2}$	$\phi\mu(1-\mu)$
Binomial	$\frac{1}{1+e^{-\theta}}$	$\frac{\phi e^{\theta}}{n(1+e^{\theta})^2}$	$\phi\mu(1-\mu)/n$
Poisson	e^{θ}	ϕe^{θ}	$\phi\mu$
Poisson con offset	e^{θ}	$\frac{\phi}{n} e^{\theta}$	$\phi\mu/n$

Así pues, la especificación del modelo GLM puede expresarse:

$$E(Y | x) = \mu$$

$$g(\mu) = \beta_0 + \sum_i \beta_i x_i \quad (i=1, \dots, p)$$

$$\text{Var}(Y | x) = \phi v(\mu)$$

La función $g()$ se denomina 'vínculo' (link), mientras que la función $v()$ se denomina de 'varianza' y ϕ es un parámetro de escala (o de dispersión).

Se entiende mejor la aportación de los modelos GLM cuando se comparan con los modelos de regresión lineal clásicos. En el cuadro resumen siguiente, se observa como la función vínculo mantiene la aditividad¹ de los efectos, aplicando un cambio de escala cuando es necesario.

Modelo lineal clásico	Modelo lineal generalizado
<p>Parte sistemática $E(y_i) = \mu_i = \sum x_{ij} \beta_j \quad i = 1, \dots, n$</p> <p>$\mu = X\beta \quad \mu = n \times 1 \quad X = n \times p \quad \beta = p \times 1$ X.- matriz modelo β.- vector de parámetros</p> <p>Parte aleatoria Se asume independencia y varianza constante de los errores</p> <p>$\epsilon \sim N(0, \sigma^2)$</p> <p>Resumen $Y \sim N(0, \sigma^2)$ $\mu = E(Y) = X\beta$ Y independientes</p>	<p>1.-Parte sistemática $\eta_i = \sum x_{ij} \beta_j \quad i = 1, \dots, n$</p> <p>2.-Parte aleatoria $Y \sim N(0, \sigma^2)$ Y independientes</p> <p>3.-Predictor lineal η $\mu = \eta$</p> <p>Generalización A) La distribución normal supuesta en 2) para Y puede ser cualquier distribución de la familia exponencial B) Si se escribe $\eta_i = g(\mu_i)$ se tiene $g(\cdot) =$ función link, que puede ser cualquier función monótona diferenciable.</p>

Cuando Y es una variable contadora y, por tanto, viene generada por una distribución de Poisson, la función vínculo es el logaritmo, $g(\mu) = \ln(\mu)$; la varianza, $v(\cdot)$, es la función de identidad y el parámetro de dispersión, ϕ , es igual a 1.

$$E(Y | x) = \mu$$

$$\ln(\mu) = \beta_0 + \sum_i \beta_i x_i \quad (i=1, \dots, p)$$

$$\text{Var}(Y | x) = \mu$$

¹ La aditividad es una propiedad de las respuestas esperadas $E[Y|X]$ ($\hat{\mu} = \sum x_j \hat{\beta}_j$).

A menudo, que la media sea igual que la varianza ($\phi=1$) es una hipótesis demasiado restrictiva, por lo que se permite que exista sobre-dispersión ($\phi>1$) o, menos usual, sub-dispersión ($\phi<1$). En este caso, la varianza se especificará

$$\text{Var}(Y | x) = \phi\mu$$

En el caso de una respuesta dicotómica, distribución binomial,

$$E(Y | x) = \mu$$

$$\text{Ln}\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \sum_i \beta_i x_i \quad (i = 1, \dots, p)$$

$$\text{Var}(Y | x) = \phi\mu(1-\mu)$$

Se ha permitido que el modelo sea sub-disperso (lo más habitual en este caso).

Los modelos GLM se pueden estimar por máxima verosimilitud. Utilizando la forma general de la función de densidad para las distribuciones más habituales, correspondientes a una distribución exponencial de un único parámetro (en cuyo caso a es una constante):

$$f(Y; \theta) = e^{\left[\frac{(y\theta - b(\theta))}{a} + c(Y)\right]}$$

El logaritmo de la función de verosimilitud para un conjunto de N observaciones puede ser expresado como (i denota la observación):

$$\log L = \frac{\sum_i (y_i \theta - b(\theta))}{a_i} + K$$

o en notación matricial:

$$\log L = Y' A^{-1} \theta - b(\theta)' A^{-1} b(\theta)^{1/2} + K$$

en la que $A = \text{diag}(a_i)$, $\theta = (\theta_1, \theta_2, \dots, \theta_N)'$; $b(\theta) = (b(\theta_1), b(\theta_2), \dots, b(\theta_N))'$; K una constante; y $\theta = X\beta$.

Dado que la derivada de $\log L$ con respecto a β no es lineal, el logaritmo de la función de verosimilitud no se puede maximizar directamente. Alternativamente se puede utilizar una de las dos aproximaciones descritas a continuación.

1.- Utilizando el *algoritmo de Newton-Raphson*, en la iteración $(s+1)$ el estimador del vector de parámetros β se puede expresar como:

$$\beta_{s+1} = \beta_s + I^{-1}(\beta_s) U(\beta_s)$$

siendo $U(\beta)$ el vector de primeras derivadas o de *scores*, es decir $U(\beta) = \partial \log L / \partial \beta$ y $I(\beta)$ la matriz de información, $I(\beta) = -H(\beta)$; siendo $H(\beta)$ el Hessiano o matriz de segundas derivadas, $H(\beta) = \partial^2 \log L / \partial \beta \partial \beta'$.

En cada iteración, $s+1=1,2,\dots$ tanto el vector de scores como la matriz de información se evalúan en la anterior iteración s . El proceso iterativo termina bien cuando el cambio en el logaritmo de la función de verosimilitud sea lo suficientemente pequeño, bien cuando lo sea el mayor de los cambios relativos en los estimadores de los parámetros. La varianza de los parámetros se puede calcular en la última iteración, $\text{Var}(\beta) = (B'X'V^{-1}XB)^{-1}$, siendo $B = \text{diag}(g'^{-1}(\mu))$ y $V = \text{Var}(Y|x)$.

2.- Alternativamente se puede utilizar el *método de los mínimos cuadrados ponderados iterativos (IRLS)*.

Partiendo de la función vínculo $\eta = g(\mu) = X\beta$

Se define una pseudo-variable, z , como el desarrollo de primer orden de Taylor de $g(Y)$ alrededor de μ tal que:

$$g(Y) \approx g(\mu) + (y - \mu) g'(\mu) = \eta + (y - \mu) d \eta / d\mu$$

$$z_0 = g(\mu_0) + (Y - \mu_0) g'(\mu_0) = g(\mu_0) + (Y - \mu_0) B^{-1} \quad B = \text{diag}(g'^{-1}(\mu))$$

z es la expansión en series de Taylor de primer orden para $g(Y)$ en torno a μ .

Se trata de estimar, β así como $\text{Var}(\beta)$ en el modelo de regresión:

$$z = X\beta + e$$

Recordando que $\eta = g(\mu) = X\beta$, y que η y μ son valores fijos y conocidos, la varianza de z es igual a:

$$V_z = \text{Var} \left[\eta + (y - \mu) \frac{\partial \eta}{\partial \mu} \right] = \left(\frac{\partial \eta}{\partial \mu} \right)^2 \text{Var} y = \text{Var}(X\beta) + B^{-1} \text{Var}(Y - \mu) B^{-1} = 0 + B^{-1} A B B^{-1} = A B^{-1}$$

El algoritmo IRLS puede expresarse:

1. Tomar Y como valores iniciales de z .
2. En la primera iteración $\beta = (X'V_z^{-1}X)^{-1}X'V_z^{-1}z$.
3. El valor de β se utiliza para calcular nuevos valores de z y de V_z .
4. Repetir el proceso de iteración hasta la convergencia.

La varianza asintótica de β se puede obtener en la iteración final como,

$$\text{Var}(\beta) = (X' V_z X)^{-1}$$

Para los links canónicos ($\eta = \theta$), el valor esperado y el actual en la matriz hessiana coinciden y, por lo tanto, el método desarrollado para IRLS (Fisher scoring method) coincide con el algoritmo de Newton Raphson.

Anexo 2. Modelos lineales generalizados mixtos (GLMM)

En el análisis econométrico aplicado están proliferando los *diseños mixtos*, caracterizados por considerar simultáneamente dos o más dimensiones de análisis (Saez, 2001). Comprenden, por tanto, los diseños multinivel, o de niveles múltiples, también denominados jerárquicos y, por otro, los estudios longitudinales o de medidas repetidas. Los estudios multinivel tienen una estructura jerárquica, con agrupaciones de datos en grupos o clústers. En la literatura esos niveles jerárquicos se denominan nivel uno, nivel dos, etc; estadio uno, estadio dos, etc; nivel individual y nivel poblacional; o nivel individual y nivel clúster. La característica peculiar de los estudios longitudinales es la medición repetida a lo largo del tiempo en cada individuo o sujeto del estudio (Saez, 2001).

Empezando, sin pérdida de generalidad, con una respuesta lineal, un modelo jerárquico puede ser especificado como sigue:

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_p x_{p_{ij}} + \varepsilon_{ij}$$

El subíndice $i=1, \dots, m$ denota áreas (l áreas), por ejemplo, y el subíndice $j=1, \dots, n_i$ ($J=\max(n_i)$, número de individuos) los distintos individuos, como ilustración, residentes en cada área (nótese que el número de individuos, n_i , no tiene porque ser el mismo para todas las áreas).

Denótese como y_{ij} las observaciones de la variable dependiente, o respuesta. Se considera que la respuesta de cada área i viene determinada por p variables explicativas observadas en el individuo j , $x_{p_{ij}}$. Algunas de las variables explicativas son fijas ($x_{p_{ij}} = x_{p_i} \forall j$) y otras son variables dentro del área. Los β son p parámetros desconocidos. Finalmente, ε_{ij} es un término de perturbación con media nula.

Se debe suponer:

$$E(\varepsilon_{ij}) = 0$$

$$E(\varepsilon_{ij}^2) = \sigma_\varepsilon^2 \text{ constante}$$

$$E(\varepsilon_{it} \varepsilon_{js}) = 0 \quad \forall t \neq s \text{ y } \forall i \neq j$$

De forma simplificada, considérense cinco áreas $i=1,2,3,4,5$ ($I=5$); tres individuos $j=1,2,3$ ($J=3$); y dos variables explicativas ($p=2$), x_1 (fija) y x_2 (variable dentro del área).

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$$

En notación matricial se escribiría:

a) Para cada área:

$$Y_i = X_i \beta + \varepsilon_i$$

$$\text{donde } Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}, i=1,2,\dots,5; X_i = \begin{bmatrix} 1 & x_{1i} & x_{2i1} \\ 1 & x_{1i} & x_{2i2} \\ 1 & x_{1i} & x_{2i3} \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}; \text{ y } \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix}.$$

La media y la varianza de la variable dependiente pueden formularse:

$$E(y_{ij}) = \mu_i$$

$$\text{Var}(y_{ij}) = \sigma_{\varepsilon i}^2$$

$$E(Y_i) = \mu_i$$

$$\text{Var}(Y_i) = V_i$$

$$\text{donde } \sigma_{\varepsilon i}^2 \text{ es constante para cada área } i; \text{ y (en este caso) } V_i = \begin{bmatrix} \sigma_{\varepsilon i}^2 & 0 & 0 \\ 0 & \sigma_{\varepsilon i}^2 & 0 \\ 0 & 0 & \sigma_{\varepsilon i}^2 \end{bmatrix}.$$

b) Para toda la muestra:

$$Y = X\beta + \varepsilon$$

$$\text{donde } Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix}; \text{ y } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}.$$

Ahora:

$$E(Y) = \mu$$

$$Var(Y) = V$$

$$\text{siendo } V = \begin{bmatrix} V_1 & 0 & 0 & 0 & 0 \\ 0 & V_2 & 0 & 0 & 0 \\ 0 & 0 & V_3 & 0 & 0 \\ 0 & 0 & 0 & V_4 & 0 \\ 0 & 0 & 0 & 0 & V_5 \end{bmatrix}.$$

Es importante señalar que la matriz V debe ser *siempre* diagonal en bloques, es decir que se asume que las áreas (o grupos) son *independientes* entre sí.

Según el objetivo propuesto los estudios longitudinales pueden aproximarse *marginal* o *condicionalmente*. Se utiliza el enfoque marginal cuando se pretenden realizar inferencias “poblacionales”, es decir, si lo que se quiere es explicar la relación entre la variable dependiente y las variables explicativas con independencia de la variabilidad “intraindividual” (o dentro del área).

Por ejemplo:

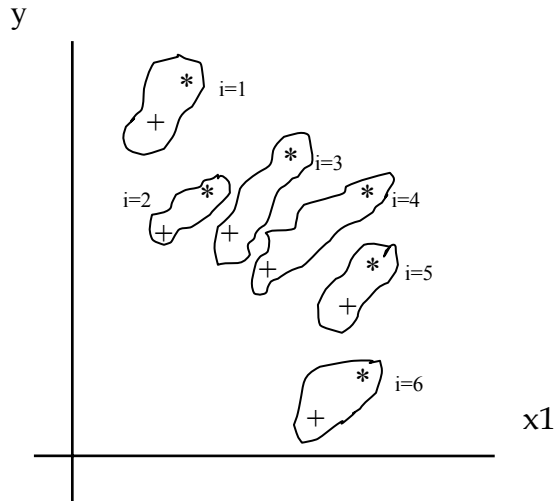
$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$$

La ordenada en el origen (β_0) y los coeficientes asociados a las variables explicativas (β_1 y β_2) son comunes a todas las áreas. No existe heterogeneidad “individual” o, de otra forma, todos los efectos (de las variables explicativas, incluida la ordenada en el origen) son fijos. Se trata de estimar los parámetros correspondientes a la media, β . Muy raramente son de interés los parámetros de la covarianza (la varianza y, en caso de no ser nula, las covarianzas de ε). De hecho éstos suelen ser tratados como un estorbo, controlados, pero no estimados en el enfoque marginal.

El *enfoque condicional* pretende realizar inferencias “individuales”. Para ello, modeliza simultáneamente la media de la variable dependiente (la variabilidad “interindividual”, entre áreas) y la estructura de covarianzas o correlaciones (la variabilidad “intraindividual”, dentro del área). En este enfoque los parámetros que definen la correlación tienen el mismo, y a veces incluso más, interés que los correspondientes a la media.

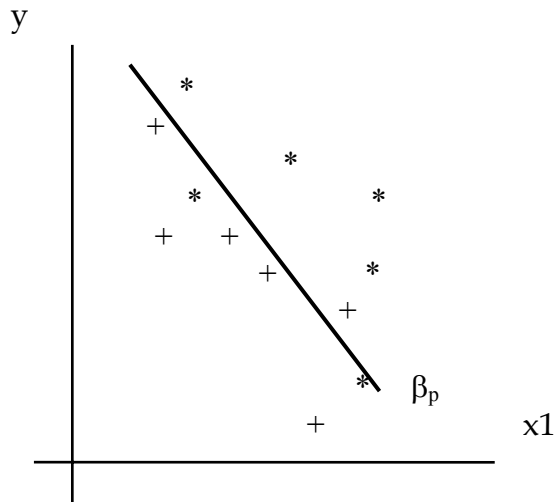
Los *efectos aleatorios* es el enfoque más conocido de entre los condicionales. Supone que la dependencia que pueda existir entre diferentes observaciones de la variable dependiente, es decir entre las respuestas de una misma área, se debe a que los coeficientes de la regresión (de la media) no son los mismos para todas las áreas. De hecho, la interpretación más sencilla, es la que supone que la heterogeneidad individual se debe a factores no observables (o variables omitidas) fijas dentro del área, pero variables entre áreas. Estos factores son los que provocan la dependencia entre las distintas observaciones de la variable dependiente.

A partir del siguiente ejemplo, en el que se han considerado seis áreas y dos individuos (con + se representa el primero y con * el segundo),



Un modelo de efectos fijos sería, por ejemplo, el de una *regresión pooled*

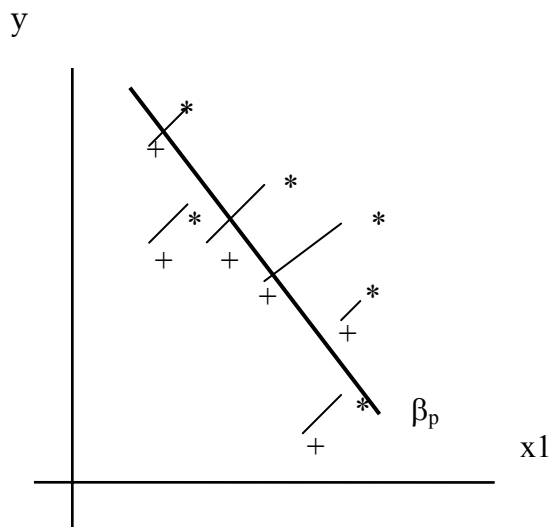
$$y_{ij} = \beta_0 + \beta_p x_{ij} + \varepsilon_{ij}$$



Un modelo de efectos aleatorios podría ser escrito:

$$y_{ij} = \beta_{0i} + \beta_p x_{ij} + \varepsilon_{ij}$$

La ordenada en el origen, β_{0i} , es decir el efecto *basal*, es distinto para cada una de las áreas. El efecto de x_1 , sin embargo, es fijo, aproximadamente el mismo para todas las áreas.



En el siguiente modelo x_1 es el efecto aleatorio.

$$y_{ij} = \beta_0 + \beta_{pi} x_{ij} + \varepsilon_{ij}$$

Este tipo de modelo, en el que el parámetro está asociado a una variable continua se denomina también de *coeficientes aleatorios*. Obviamente el modelo puede combinar efectos y coeficientes aleatorios.

En resumen, los modelos de efectos aleatorios (basales y/o pendientes) intentan modelizar la esperanza condicional de y_{ij} , dados los coeficientes específicos de las áreas, β_i ,

$$y_{ij} = \beta_i x_{ij} + \varepsilon_{ij}$$

El problema es que para cada área suele haber muy pocas observaciones para estimar β_i a partir únicamente de (y_i, x_i) . Por ello se supone que los β_i no son más que realizaciones independientes de alguna distribución de probabilidad con media β :

$$\boxed{} \\ y_{ij} = \beta x_{ij} + (\varepsilon_{ij} + x_{ij} u_i)$$

Ahora, además de:

$$\begin{aligned} E(\varepsilon_{ij}) &= 0 \\ E(\varepsilon_{ij}^2) &= \sigma_\varepsilon^2 \text{ constante} \\ E(\varepsilon_{it} \varepsilon_{js}) &= 0 \quad \forall t \neq s \text{ y } \forall i \neq j \end{aligned}$$

se ha de suponer también:

$$\begin{aligned} E(u_i) &= 0 \\ E(u_{ij}^2) &= \sigma_u^2 \text{ constante} \\ E(\varepsilon_{it} u_j) &= 0 \quad \forall i, t, j \\ E(u_i u_j) &= 0 \quad \forall i \neq j \end{aligned}$$

Las dos últimas hipótesis son particularmente importantes. Los *efectos aleatorios tienen que ser independientes de los fijos* y las *áreas deben ser independientes* entre sí.

Generalizando, y siguiendo con la terminología anterior, un *modelo lineal generalizado mixto*, *GLMM*, puede ser expresado como:

$$E(y_{ij} | x_{ij}, u_{ij}) = \mu_{ij}$$

$$g(\mu_{ij}) = x_{ij}\alpha + z_{ij}'\beta_i$$

$$\text{Var}(y_{ij} | x_{ij}, u_{ij}) = \phi v(\mu_{ij})$$

denotando por α los efectos fijos (x las variables asociadas) y β_i los efectos aleatorios (z las variables asociadas); siendo habitual que los efectos aleatorios se distribuyan normalmente (multivariante) con media nula y matriz de covarianzas G, es decir $\beta_i \sim N(0, G)$.

En términos matriciales:

$$g(\mu) = X\alpha + Z\beta$$

la varianza $V = \text{Var}(Y)$,

$$V = \text{Var}(\mu) + R$$

En particular, cuando la variable dependiente es continua, el modelo mixto puede ser especificado en términos matriciales como:

$$\mu = X\alpha + Z\beta$$

$$\text{Var}(\mu) = \phi$$

La estructura de G, la matriz de covarianzas de los efectos aleatorios puede ser:

Identidad o uniforme

$$\begin{bmatrix} \sigma_c^2 & 0 & 0 \\ 0 & \sigma_c^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{bmatrix}$$

No estructurada o matriz general definida positiva

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$

Diagonal

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

Simetría compuesta

$$\begin{bmatrix} \sigma_c^2 & \theta & \theta \\ \theta & \sigma_c^2 & \theta \\ \theta & \theta & \sigma_c^2 \end{bmatrix}$$

Autorregresiva de orden uno

$$\begin{bmatrix} \sigma_c^2 & \rho & \rho^2 \\ \rho & \sigma_c^2 & \rho \\ \rho^2 & \rho & \sigma_c^2 \end{bmatrix}$$

Toeplitz (AR(p) con p > 1)

$$\begin{bmatrix} \sigma_c^2 & \sigma_1 & \sigma_2 \\ \sigma_1 & \sigma_c^2 & \sigma_1 \\ \sigma_2 & \sigma_1 & \sigma_c^2 \end{bmatrix}$$

Espacial

$$\begin{bmatrix} \sigma_c^2 & \rho^{d_{12}} & \rho^{d_{13}} \\ \rho^{d_{12}} & \sigma_c^2 & \rho^{d_{23}} \\ \rho^{d_{13}} & \rho^{d_{23}} & \sigma_c^2 \end{bmatrix}$$

siendo $d_{ij} = |t_i - t_j|$ y t_i, t_j son los periodos asociados a i y j , respectivamente.

La matriz de covarianzas residual, R, puede adoptar también diversas estructuras: identidad, autoregresiva, exponencial (autoregresiva general en el sentido que los intervalos de tiempo no necesitan ser iguales), espacial exponencial, espacial gaussiana, espacial lineal, de simetría compuesta, Toeplitz.

Media móvil:

$$MA(1) \text{ por ejemplo } \begin{bmatrix} \sigma^2 & \gamma & 0 \\ \gamma & \sigma^2 & \gamma \\ 0 & \gamma & \sigma^2 \end{bmatrix}$$

ARMA:

$$ARMA(1,1) \text{ por ejemplo } \begin{bmatrix} \sigma^2 & \gamma_1 & \gamma_2 \\ \gamma_1 & \sigma^2 & \gamma_1 \\ \gamma_2 & \gamma_1 & \sigma^2 \end{bmatrix}$$

Como era de esperar la varianza de la variable dependiente, la de los efectos aleatorios y la varianza residual están relacionadas (siendo X_i y Z_i no aleatorias y β y ε_i independientes):

$$V_i = \text{Var}(Y_i) = \text{Var}(X_i\alpha) + \text{Var}(Z_i\beta) + \text{Var}(\varepsilon_i) = 0 + Z_i\text{Var}(\beta)Z_i' + \text{Var}(\varepsilon_i) = Z_iGZ_i' + R$$

Por lo que a partir de G y/o R no resulta difícil derivar la estructura de la matriz V_i . Finalmente, tanto la estructuras de G como la de R (y por tanto la de V_i) pueden estar organizada por bloques de áreas y/o de periodos de tiempo.

Por ejemplo, sea un estudio longitudinal en que se siguen dieciocho individuos, nueve casos y sus correspondientes controles. Se ha supuesto una estructura identidad para la matriz de varianzas.

$$V_i = \begin{bmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix} & \begin{pmatrix} \sigma_4^2 & 0 \\ 0 & \sigma_4^2 \end{pmatrix} & \begin{pmatrix} \sigma_7^2 & 0 \\ 0 & \sigma_7^2 \end{pmatrix} \\ \begin{pmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} & \begin{pmatrix} \sigma_5^2 & 0 \\ 0 & \sigma_5^2 \end{pmatrix} & \begin{pmatrix} \sigma_8^2 & 0 \\ 0 & \sigma_8^2 \end{pmatrix} \\ \begin{pmatrix} \sigma_3^2 & 0 \\ 0 & \sigma_3^2 \end{pmatrix} & \begin{pmatrix} \sigma_6^2 & 0 \\ 0 & \sigma_6^2 \end{pmatrix} & \begin{pmatrix} \sigma_9^2 & 0 \\ 0 & \sigma_9^2 \end{pmatrix} \end{bmatrix}$$

Sin embargo, cuando la variable dependiente no es continua (es decir es contadora, dicotómica, etc), μ no es una función lineal de β , por lo que V no puede ser derivada fácilmente a partir de G y R , como en el caso continuo.

En este caso, una aproximación basada en la expansión de series de Taylor de primer orden podría ser expresada (nótese que es tan sólo una aproximación):

$$V \approx BZGZ'B + R$$

B es una matriz diagonal de varianzas determinadas por la función vínculo. Por ejemplo en una Poisson, $B = \text{diag}(\mu_i)$ ($B = \text{diag}(\mu_i/t_i)$ en una Poisson con offset), en una binomial, $B = \text{diag}(\mu_i(1 - \mu_i))$.

La matriz R puede ser la identidad (no existe autocorrelación) en cuyo caso, $R = AB$; o bien podemos permitir la autocorrelación, entonces $R = A^{1/2}B^{1/2}R_sB^{1/2}A^{1/2}$, siendo R_s una de las matrices R definidas más arriba.

Sólo existe una función de verosimilitud completa cuando $R=I$ (es decir no existe autocorrelación), existan o no efectos aleatorios. Se puede escribir:

$$L(\alpha, \beta, \gamma_G; Y) = L(\alpha; Y; \beta) L(\gamma_G; \beta)$$

siendo γ_G los coeficientes en G

Suponiendo una distribución normal para los efectos, $\beta \sim N(0, G)$,

$$L(\alpha, \beta, \gamma_G; Y) \propto L(\alpha, Y | \beta) |G|^{-1/2} e^{-\frac{1}{2} \beta' G^{-1} \beta}$$

Como Y y β son independientes ($R=I$), se puede utilizar la expresión del GLM estándar,

$$L(\alpha; Y | \beta) = e^{Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2} + K} \propto e^{Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2}}$$

en la que $\theta = X\alpha + Z\beta$, $A = \text{diag}(a_i)$ y K es una constante.

La verosimilitud completa puede ser expresada:

$$L(\alpha, \beta, \gamma_G; Y) \propto e^{Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2}} |G|^{-1/2} e^{-\frac{1}{2} \beta' G^{-1} \beta}$$

y el logaritmo de la función de verosimilitud:

$$\text{Ln}L(\alpha, \beta, \gamma_G; Y) \propto Y' A^{-1} \theta - b(\theta)^{1/2} A^{-1} b(\theta)^{1/2} - \frac{1}{2} |G| - \frac{1}{2} \beta' G^{-1} \beta + K$$

En una Poisson, $a_i = 1$ y $b(\theta) = \exp(\theta)$ (en una Poisson con offset, $a_i = 1/t_i$ y $b(\theta) = \exp(\theta)$). En una binomial, $a_i = 1/n_i$ y $b(\theta) = \text{Ln}(1 + \exp(\theta))$.

Cuando R no es la matriz identidad, es decir existe autocorrelación, no existe la verosimilitud completa. Sin embargo se puede definir la cuasi-verosimilitud:

$$QL(\alpha, \beta, \gamma; Y) \propto Ln(f(\alpha, \gamma; Y|\beta)) - \frac{1}{2}|G| - \frac{1}{2}\beta'G^{-1}\beta + K$$

Siendo $\gamma=(\gamma_G, \gamma_R)$ y $f(\alpha, \gamma; Y)$ una función con propiedades similares al logaritmo de la función de verosimilitud, definiéndose tal que,

$$\frac{\partial Ln(f(\alpha, \gamma; Y))}{\partial \alpha} = X'BR^{-1}(Y - \mu)$$

que tiene exactamente la misma forma que la derivada parcial de la verdadera verosimilitud.

La maximización de la cuasi-verosimilitud es extremadamente compleja. Por otra parte, los métodos Bayesianos, a los que se puede recurrir, implican la evaluación total de la cuasi-verosimilitud, por lo que suelen ser muy intensivos computacionalmente.

Alternativamente se puede utilizar la *pseudo-verosimilitud* o *pseudo-verosimilitud penalizada*, PQL (Breslow y Clayton, 1993). El adjetivo pseudo se refiere al hecho que en cada iteración se maximiza la función de verosimilitud de una pseudo-variable, diferente a la correspondiente a los datos originales. Se utiliza la misma pseudo-variable basada en la expansión de series de Taylor de primer orden que definimos más arriba:

$$z = g(\mu) + (y-\mu)B^{-1} = X\alpha + Z\beta + (Y - X\alpha - Z\beta) B^{-1}$$

La varianza de z se define,

$$V_z = \text{Var}(X\alpha + Z\beta) + B^{-1}\text{Var}(Y - \mu) B^{-1} = ZGZ' + B^{-1}R B^{-1}$$

La matriz R puede ser reescrita como $R = A^{1/2}B^{1/2}R_s B^{1/2}A^{1/2}$

$$V_z = ZGZ' + A^{1/2}B^{1/2}R_s B^{1/2}A^{1/2}$$

Esta matriz se parece bastante a la de los datos originales, $V = BZGZ'B + A^{1/2}B^{1/2}R_s B^{1/2}A^{1/2}$

Wolfinger y O'Connell (1993) muestran que $z|\beta$ se distribuyen normalmente,

$$z|\beta \sim N(X\alpha + Z\beta, A^{1/2}B^{1/2}R_s B^{1/2}A^{1/2})$$

por lo que puede ser estimado ponderadamente como un modelo mixto normal con matriz residual R_s y matriz de ponderaciones diagonal, $A^{-1}B$.

Debido a que z y B dependen de los estimadores de α y β , el modelo debe ser estimado iterativamente:

- 0.- Utilizar como valores iniciales de z los datos originales Y (o bien $g(Y)$ evitando posibles valores infinitos). La matriz de ponderaciones inicial puede ser la identidad.
- 1.- Estimar un modelo mixto mediante REML tal y como vimos más arriba.
- 2.- Calcular nuevos valores para z y B utilizando los estimadores de α y β obtenidos en 1. Calcular la matriz de ponderaciones $A^{-1}B$.
- 3.- Volver al paso 1.

El número de iteraciones puede ser elevado ya que dentro de cada etapa se itera a fin de obtener los estimadores REML. Una vez el modelo ha convergido, se calcula la varianza de los estimadores de los efectos fijos:

$$\text{Var}(\hat{\alpha}) = (X'V_z^{-1}X)^{-1}$$

Puesto que V_z es estimada esta varianza puede estar ligeramente sesgada.

Anexo 3. El algoritmo Gibbs sampling

La influencia Bayesiana se basa en la distribución a posteriori:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

Donde θ es un vector de parámetros, habitualmente de gran dimensión.

Mientras que $p(x|\theta)$ (la verosimilitud) y $p(\theta)$ (distribución a priori) son fáciles de obtener, $p(\theta|x)$ suele ser analíticamente inabordable. Alternativamente se intenta obtener bien la distribución muestral a posteriori,

$$p(\theta_i|x) = \int \int \dots \int p(\theta|x) d\theta_{(-i)}$$

Donde $\theta_{(-i)}$ denota el vector θ excluyendo el componente i .

O bien algunas propiedades de $p(\theta_i|x)$ como la media ($\int \theta_i p(\theta_i|x) d\theta_i$) o percentiles ($\int_r^\infty p(\theta_i|x) d\theta_i$).

En cualquier caso resulta vital la integración numérica.

Así pues, se trata de obtener muestra de la distribución conjunta a posteriori $p(\theta|x)$. El muestreo independiente suele ser muy complejo. Sin embargo, el muestreo dependiente a partir de cadenas de Markov con $p(\theta|x)$ como su distribución estacionaria (de equilibrio) es mucho más sencilla.

Una secuencia de variables aleatorias $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots$ forma una cadena de Markov si $\theta^{(i+1)} \longrightarrow p(\theta^{(i+1)}|\theta^{(i)})$; es decir, condicional al valor de $\theta^{(i)}$, $\theta^{(i+1)}$ es independiente $\theta^{(i-1)}, \dots, \theta^{(0)}$.

Existen varios algoritmos para diseñar cadenas de Markov con una distribución estacionaria $p(\theta|x)$.

El algoritmo '*Gibbs sampling*' es uno de los más sencillos de las MCMC. Fue introducido por Geman y Geman (1984) en el contexto de procesamiento de imágenes y, más tarde discutido en el contexto de datos faltantes por Tanner y Wong (1987). El artículo de Gelfand y Smith (1990) aplica el algoritmo a problemas bayesianos.

Sea θ un vector desconocido tal que $\theta = (\theta_1, \theta_2, \dots, \theta_k)$:

1.- Se escogen valores iniciales $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}$.

2.- Se muestrea $\theta_1^{(1)}$ a partir de $p(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

Se muestrea $\theta_2^{(1)}$ a partir de $p(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, x)$

...

Se muestrea $\theta_k^{(1)}$ a partir de $p(\theta_k | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)}, x)$

3.-Se repite la etapa 2 muchas veces (si el número de repeticiones es muy grande se obtendrá una muestra de $p(\theta|x)$).

Más formalmente, sea la distribución condicional conjunta

$$\{\pi(\psi_1 | \psi_2, \dots, \psi_p); \pi(\psi_2 | \psi_1, \psi_3, \dots, \psi_p); \dots, \pi(\psi_p | \psi_1, \dots, \psi_{p-1})\} .$$

Se simula $\{\psi_n\}_{n=1}^N$ de esta distribución conjunta.

Algoritmo (Gibbs Sampling)

$$\psi^{(0)} = (\psi_1^{(0)}, \dots, \psi_p^{(0)})$$

1. Especificar un valor inicial $\psi^{(0)}$:
2. Repetir para $j = 1, 2, \dots, M$.

Generar $\psi_1^{(j+1)}$ de $\pi(\psi_1 | \psi_2^{(j)}, \psi_3^{(j)}, \dots, \psi_p^{(j)})$

Generar $\psi_2^{(j+1)}$ de $\pi(\psi_2 | \psi_1^{(j+1)}, \psi_3^{(j)}, \dots, \psi_p^{(j)})$

⋮

Generar $\psi_p^{(j+1)}$ de $\pi(\psi_p | \psi_1^{(j+1)}, \dots, \psi_{p-1}^{(j+1)})$.

3. Obtener los valores $\{\psi^{(1)}, \psi^{(2)}, \dots, \psi^{(M)}\}$.

Así pues la densidad de transición del estado $\psi_h^{(j)}$ a $\psi_h^{(j+1)}$ es la siguiente

$$\pi(\psi_h | \psi_1^{(j+1)}, \dots, \psi_{h-1}^{(j+1)}, \psi_{h+1}^{(j)}, \dots, \psi_p^{(j)})$$

dado que se ha obtenido el bloque h -ésimo, los $(h-1)$ bloques han sido actualizados. Así, la densidad de transición de la cadena bajo la asunción que π es absolutamente continuo, viene dada por los kernels de transición de cada bloque.

$$K(\psi, \psi') = \prod_{h=1}^p \pi(\psi_h | \psi_1^{(j+1)}, \dots, \psi_{h-1}^{(j+1)}, \psi_{h+1}^{(j)}, \dots, \psi_p^{(j)}) .$$

Anexo 4. Variables explicativas con un diseño en rejilla. Índice de privación e incidencia y mortalidad.

-Construcción de un índice de privación material para los municipios de la Regió Sanitària Girona (*Estudios de Economía Aplicada*, 2005)

-Influencia de la privación sobre la distribución geográfica de la incidencia de cáncer en la Regió Sanitària Girona (*Economía y Salud*, 2006)

-Área deprivation a mortality in capital cities of Andalusia and Catalonia (Spain) (Enviado a *Journal of Epidemiology and Community Health*).

Construcción de un índice de privación material para los municipios de la Región Sanitaria Girona¹

Aitana Lertxundi-Manterola*, Carme Saurina*, Marc Saez**, Ricardo Ocaña-Riola***

* *Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS), Universitat de Girona*

** *Prof. Marc Saez, PhD, CStat. Grup de Recerca en Estadística, Economía Aplicada i Salut (GRECS). Universitat de Girona. Campus de Montilivi, 17071 – Girona.*

*** *Escuela Andaluza de Salud Pública.*

** <http://www.udg.edu/fcee/economia/catala/grecs.htm>; Tfno.: 972-418736; Fax: 972-418032

** E-mail: marc.saez@udg.es

RESUMEN

El objetivo de este trabajo fue el de derivar un índice de privación material que, a la vez que resumiese las condiciones socioeconómicas de los municipios de la Región Sanitaria Girona, permitiese explicar la variabilidad geográfica de la incidencia del cáncer en la Región.

Utilizando como base el índice de privación material de Townsend, y mediante el Análisis de Componentes Principales, construimos otro índice que, a nuestro entender, pareció superior. En primer lugar, nuestro índice resumió mucho mejor las condiciones socioeconómicas de los municipios de la Región Sanitaria Girona. En segundo lugar, tomando como ejemplo el cáncer de estómago en mujeres, el nuestro fue el único índice que fue capaz de capturar (parte de) la variabilidad geográfica de la incidencia en la Región Sanitaria Girona (en el período 1994-2000).

Palabras clave: Privación material, Análisis de Componentes Principales, índice de Townsend, incidencia del cáncer.

Constructing a Privation Index for the Municipalities of the Girona Health Region

ABSTRACT

Our objective was to derive a deprivation index that could summarise the socio-economic conditions of the municipalities of the Girona Health

¹ Agradecimientos

Agradecemos los comentarios de un evaluador anónimo, así como a la Unitat d'Epidemiologia i Registre de Càncer de Girona-ICO, por proporcionarnos los datos. Este trabajo se benefició parcialmente de la Beca Bayer de Investigación en Economía de la Salud, Asociación de Economía de la Salud, 2004, 'Influencia de la privación sobre la distribución geográfica de la incidencia del cáncer en la Región Sanitaria Girona' concedida a Aitana Lertxundi-Manterola y a Marc Saez.

Artículo recibido en agosto de 2004 y aceptado en noviembre de 2004

La referencia electrónica de este artículo en la página www.revista-eea.net, es e-23108

Region and, furthermore, explain the geographical variability of the cancer incidence in the Region.

Using as a basis the Townsend's deprivation index, we built another deprivation index by means of Principal Component Analysis. Firstly, our index provided a better summary of the socioeconomic conditions of the municipalities of the Girona Health Region. Thirdly, using the stomach cancer incidence as an example, our index was the only one that captured (part of) the geographical variability of the incidence in the Girona Health Region (in the period 1994-2000).

Key words: Deprivation, Principal Component Analysis, Townsend's index, cancer incidence.

Clasificación JEL: C10, C43, C82, D30

Aitana Lertxundi-Manterola, Marc Saez Zafra. “Influencia de la privación sobre la distribución geográfica de la incidencia del cáncer en la Región Sanitaria Girona”.

Economía y salud: boletín informativo. Año 19, núm. 56 (mayo 2006): p. 10-11

Title:

Area deprivation and mortality in the capital cities of Andalusia and Catalonia (Spain)

Authors:

Ricardo Ocaña-Riola, Carme Saurina, Alberto Fernández-Ajuria, Aitana Lertxundi, Carmen Sánchez-Cantalejo, Marc Saez, Miguel Ruiz-Ramos, Maria Antònia Barceló, Antonio Daponte, José Miguel Martínez, Joan Carles March, Joan Benach.

Authors' affiliations:

Ricardo Ocaña-Riola, Alberto Fernández-Ajuria, Carmen Sánchez-Cantalejo, Antonio Daponte, Joan Carles March, Escuela Andaluza de Salud Pública, Granada, Spain

Carme Saurina, Aitana Lertxundi, Marc Saez, Maria Antònia Barceló, Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS), Universitat de Girona, Girona, Spain

Miguel Ruiz-Ramos, Consejería de Salud de la Junta de Andalucía, Sevilla, Spain

José Miguel Martínez, Joan Benach, Departament de Ciències Experimentals i la Salut. Unitat de Recerca en Salut Laboral, Universitat Pompeu Fabra, Barcelona, Spain

Correspondence to:

Ricardo Ocaña-Riola
Escuela Andaluza de Salud Pública
Campus Universitario de Cartuja
Cuesta del Observatorio, 4
Apdo. de Correos 2070
18080 Granada (Spain)

Tel.: +34 958 027400

Fax: +34 958 027503

E-mail: ricardo.ocana.easp@juntadeandalucia.es

Keywords:

Spain; Deprivation; Mortality; Small-area analysis; Socioeconomic Factors; Bayesian analysis

Word count:

Abstract: 241

Text: 2971

ABSTRACT

Objective: The purpose of this paper is to study the linkage between material deprivation and mortality from all causes in the capital cities of Andalusia and Catalonia (Spain) in men and women.

Methods: A small-area ecological study was devised using the census section as the unit for analysis. 188 983 deaths occurring in the capital cities of Andalusia and 109 478 deaths recorded in the capital cities of Catalonia were examined. Principal Components Factorial Analysis was used to devise a material deprivation score comprising the percentage of manual labourers, unemployment and illiteracy. A Hierarchical Bayesian Model was used to study the relationship between mortality and area deprivation.

Main results: In most cities, results show an increase in mortality rates among men in the most deprived wards. In Andalusia, the relative risks ranged between 1.24 (Malaga) and 1.40 (Granada), with 95% credibility intervals showing a significant excess risk. In Catalonia, relative risks ranged between 1.08 (Girona) and 1.50 (Tarragona). No evidence was found for an excess of female mortality in most deprived areas in both Autonomous Communities.

Conclusions: The relationship between mortality and deprivation appears to behave similarly in Andalusia and Catalonia. The behaviour of mortality among men and women differs in areas in the same city but with different levels of material deprivation. The continuation of this research must certainly be borne in mind when identifying aetiological factors and when implementing intervention strategies intended to improve the health of the population.

What is already know

- The term material deprivation has been used to refer to socio-economic characteristics of a geographical area which set it apart from others with a higher level of development.
 - Most ecological researches have used the province or municipality as the unit for analysis in order to study the relationship between material deprivation and disease incidence or mortality rates in the areas.
 - There are very few research papers focusing on the relationship between material deprivation and mortality within cities, using the census section as the unit for analysis.
-

What this paper adds

- This is the first research in Spain which study the linkage between material deprivation and mortality from all causes in men and women in all the capital cities of Andalusia and Catalonia.
 - The findings show very similar patterns of association between overall mortality and material deprivation in the cities of two Autonomous Spanish Regions with opposite economic development level.
 - In men, the highest mortality rates arising in the most depressed areas. However, this relationship is weaker for women.
 - The behaviour of mortality among men and women differs in areas in the same city but with different levels of material deprivation.
-

Policy implications

- This kind of study is essential to devise local health policies aimed at promoting the implementation of and improvement to health policies in the urban setting, with particular emphasis on the least advantaged districts.
 - The continuation of this research must certainly be borne in mind when identifying aetiological factors and when implementing intervention strategies intended to improve the health of the population.
-

Anexo 5. Instrucciones y sintaxis en R y en WinBUGS

5.1.- Variables medioambientales

##IMPORTACIÓN DE DATOS

```
setwd("C:/Documents and Settings/ALertxundi/Escritorio/bilbo/kutsaduren datuak/nere
datuak")###con esta instrucción se define el directorio de trabajo
library(foreign)###librería que sirve para importar datos externos
datos<-read.spss("NO2aitana.sav")
attach(datos)###para activar la base de datos llamada 'datos'
```

##IMPUTACIÓN DE DATOS. Se utilizó la imputación múltiple (Little y Rubin, #1987; Rubin, 1987; 1996). En concreto, mediante el algoritmo EM (Schafer, #1996), los valores perdidos fueron estimados por máxima verosimilitud (Novo y #Schafer, 2003). Para cada una de las tres áreas se consideró la serie completa del #contaminante y su relación con las series completas del mismo contaminante en #otras estaciones del área. El procedimiento se repitió para todos y para cada uno #de los años. Se trata de un método óptimo, siempre y cuando se cumpla dos #hipótesis (Little y Rubin, 1987). Se requiere, en primer lugar, que los datos se #distribuyan según una normal multivariante. Hipótesis que se cumple en el caso #de los contaminantes. En segundo lugar, los valores perdidos deben ser MAR #(missing at random). El criterio de validez de una estación, en este caso, implica #también esta propiedad.

```
library(norm)###librería de imputación múltiple para variables gaussianas
rngseed(1234567)###semilla de aleatorización. Necesaria para realizar la #imputación
contodo<-as.matrix(data.frame(datos)[,2:29])###creación de una matriz cuyas #columnas
son las concentraciones del contaminante en cada una de las #estaciones, cada fila corresponde a
los días de la serie temporal
imputacion<- imp.norm(prelim.norm(contodo),em.norm(prelim.norm(contodo)),
contodo)###instrucción para imputación múltiple
```

```
##el siguiente algoritmo sustituye los valores negativos por la mediana
for(i in 1:dim(imputacion)[1]){
  for(j in 1:dim(imputacion)[2]){
    imputacion[i,j]<-ifelse(imputacion[i,j]<0,NA,imputacion[i,j])
    imputacion[i,j]<-
ifelse(is.na(imputacion[i,j]),median(imputacion[i,],na.rm=T),imputacion[i,j])}
imputacion1<-data.frame(año=datos$AÑO,imputacion)
```

#Antes de realizar las predicciones se extrajo aquella conducta temporal a largo #plazo de las series de los contaminantes que pudiese distorsionar la dependencia #espacial. Por este motivo, y utilizando como variable dependiente la serie del #contaminante en cada una de las estaciones (para cada área y período), se estimó #un modelo GLM de respuesta normal, introduciendo como variables explicativas #tantas variables ficticias como años contuviese el período menos uno.

```
attach(imputacion1)## para activar la base de datos llamada 'imputación1'
zbc<-1:nrow(imputacion1)###creación de un indicador llamado zbc
library(splines)###librería que sirve para modelizar de forma paramétrica y no ##paramétrica
```

###Es importante señalar que *na.action=na.exclude* es la opción que se debe utilizar por ##cuanto conserva la dimensión de las matrices.

```
model<-glm(NO27CA~factor(AÑO), na.action=na.exclude)
R1NO27CA<-residuals(model)
```

##La tendencia temporal se extrae mediante un modelo glm en el cual se ##introducen tantas variables ficticias como años tenga la serie menos uno.

##Se trabaja con los residuos. Es decir, una vez extraída la tendencia temporal.

##Este proceso se realiza estación por estación.

```
model<-glm(NO2ABA~factor(AÑO), na.action=na.exclude)
R1NO2ABA<-residuals(model)
model<-glm(NO2ADR~factor(AÑO), na.action=na.exclude)
R1NO2ADR<-residuals(model)
model<-glm(NO2ALG~factor(AÑO), na.action=na.exclude)
R1NO2ALG<-residuals(model)
model<-glm(NO2AST~factor(AÑO), na.action=na.exclude)
R1NO2AST<-residuals(model)
model<-glm(NO2BAS~factor(AÑO), na.action=na.exclude)
R1NO2BAS<-residuals(model)
model<-glm(NO2BCA~factor(AÑO), na.action=na.exclude)
R1NO2BCA<-residuals(model)
model<-glm(NO2DEU~factor(AÑO), na.action=na.exclude)
R1NO2DEU<-residuals(model)
model<-glm(NO2ELO~factor(AÑO), na.action=na.exclude)
R1NO2ELO<-residuals(model)
model<-glm(NO2ERA~factor(AÑO), na.action=na.exclude)
R1NO2ERA<-residuals(model)
model<-glm(NO2GET~factor(AÑO), na.action=na.exclude)
R1NO2GET<-residuals(model)
```

```
model<-glm(NO2HOR~factor(AÑO), na.action=na.exclude)
R1NO2HOR<-residuals(model)
model<-glm(NO2IND~factor(AÑO), na.action=na.exclude)
R1NO2IND<-residuals(model)
model<-glm(NO2ING~factor(AÑO), na.action=na.exclude)
R1NO2ING<-residuals(model)
model<-glm(NO2LEI~factor(AÑO), na.action=na.exclude)
R1NO2LEI<-residuals(model)
model<-glm(NO2MAZ~factor(AÑO), na.action=na.exclude)
R1NO2MAZ<-residuals(model)
model<-glm(NO2MDI~factor(AÑO), na.action=na.exclude)
R1NO2MDI<-residuals(model)
model<-glm(NO2MUS~factor(AÑO), na.action=na.exclude)
R1NO2MUS<-residuals(model)
model<-glm(NO2NAU~factor(AÑO), na.action=na.exclude)
R1NO2NAU<-residuals(model)
model<-glm(NO2PEN~factor(AÑO), na.action=na.exclude)
R1NO2PEN<-residuals(model)
model<-glm(NO2SAN~factor(AÑO), na.action=na.exclude)
R1NO2SAN<-residuals(model)
model<-glm(NO2SGR~factor(AÑO), na.action=na.exclude)
R1NO2SGR<-residuals(model)
model<-glm(NO2SON~factor(AÑO), na.action=na.exclude)
R1NO2SON<-residuals(model)
model<-glm(NO2TRA~factor(AÑO), na.action=na.exclude)
R1NO2TRA<-residuals(model)
model<-glm(NO2TXU~factor(AÑO), na.action=na.exclude)
R1NO2TXU<-residuals(model)
model<-glm(NO2ZIE~factor(AÑO), na.action=na.exclude)
R1NO2ZIE<-residuals(model)
model<-glm(NO2ZOR~factor(AÑO), na.action=na.exclude)
R1NO2ZOR<-residuals(model)
```


##Construcción de la base de datos de los residuos

```

dias<-c(1:365,1:365,1:366,1:365,1:365,1:365,1:366,1:365,1:365)##creación de #la
variable días, para todo el período (en este caso tenemos datos desde 1994 #hasta 2003)
id1<-rep(1:length(dias),27)#id1 será el identificador de cada estación.
R1NO2<-c(R1NO27CA, R1NO2ABA, R1NO2ADR, R1NO2ALG, R1NO2AST, R1NO2BAS,
R1NO2BCA, R1NO2DEU, R1NO2ELO, R1NO2ERA, R1NO2GET, R1NO2HOR, R1NO2IND,
R1NO2ING, R1NO2LEI, R1NO2MAZ, R1NO2MDI, R1NO2MUS, R1NO2NAU, R1NO2PEN,
R1NO2SAN, R1NO2SGR, R1NO2SON, R1NO2TRA, R1NO2TXU, R1NO2ZIE,
R1NO2ZOR)##R1NO2 es el vector de residuos #de todas las estaciones
LON1<-c(LON7CA, LONABA, LONADR, LONALG, LONAST, LONBAS, LONBCA, LONDEU,
LONELO, LONERA, LONGET,LONHOR, LONIND, LONING, LONLEI, LONMAZ, LONMDI,
LONMUS, LONNAU, LONPEN, LONSAN, LONSGR, LONSON, LONTRA, LONTXU, LONZIE,
LONZOR)##LON1 es el vector de las longitudes
LAT1<-c(LAT7CA, LATABA, LATADR, LATALG, LATAST, LATBAS, LATBCA, LATDEU,
LATELO, LATERA, LATGET, LATHOR, LATIND, LATING, LATLEI, LATMAZ, LATMDI,
LATMUS, LATNAU, LATPEN, LATSAN, LATSGR, LATSON, LATTRA, LATTXU, LATZIE,
LATZOR) ##LAT1 es el vector de las latitudes
año<-rep(AÑO,27)#representa el vector de la variable año

```

###creación de una data frame que contenga las variables anteriores, según la ##siguiente estructura

<i>estación</i>	<i>Id1</i>	<i>año</i>	<i>Mes</i>	<i>días</i>	<i>R1no2</i>	<i>Lon1</i>	<i>Lat1</i>
1	1	1994	1	1			
1	2	1994	1	2			
...			
1	3287	2003	12	31			
2	1	1994	1	1			
...			
2	3287	2003	12	31			
...			
27	3287	2003	12	31			

```

ndatos<-data.frame(id=id1,año=año, mes=MES,día=días,
rno2=R1NO2,lon=LON1,lat=LAT1)

```

###definir los meses. Variable que después se utilizará en la predicción

```
enero<-ndatos[ndatos$mes==1,]
febrero<-ndatos[ndatos$mes==2,]
marzo<-ndatos[ndatos$mes==3,]
abril<-ndatos[ndatos$mes==4,]
mayo<-ndatos[ndatos$mes==5,]
junio<-ndatos[ndatos$mes==6,]
julio<-ndatos[ndatos$mes==7,]
agosto<-ndatos[ndatos$mes==8,]
septiembre<-ndatos[ndatos$mes==9,]
octubre<-ndatos[ndatos$mes==10,]
noviembre<-ndatos[ndatos$mes==11,]
diciembre<-ndatos[ndatos$mes==12,]
```

####KRIGING. Predicción de la concentración del contaminante en algunos puntos ###no observados. Método estocástico.

```
library(sp)
library(geoR)###librerías de análisis geoestadístico
```

```
memory.limit(size=4000)###se trata de un procedimiento computacionalmente ###laborioso por lo tanto aumentamos la memoria RAM.
memory.size()
```

###El kriging se realizará por período y mes. Por ejemplo para el primer ##período sería:

```
enero1<-enero[enero$año>1995 & enero$año<1999,]##enero del 96-98
id3<-rep(1:93,27)##se crea un identificador que toma los valores del 1 al 93 ##(todos los días de este período) para cada una de las estaciones
enero1<-data.frame(enero1,id3)
```

##A fin de realizar el kriging es necesario conocer la estructura de dependencia ###espacial que presentan los datos en ese mismo período. Para ello, se utilizará #un procedimiento mixto. La idea es calcular diferentes variogramas. Sin embargo, #no se analizarán gráficamente sino que se escogerá aquel variograma #(dependencia espacial) que presente una raíz del error cuadrático menor, es decir #un mejor ajuste.

##Se probó con 4 estructuras de dependencia espacial; el modelo exponencial, el modelo esférico, el modelo geniting y el modelo gaussiano.

```
matrizenero<-matrix(0,93,3)
for(i in 1:93){
  contGEO<- as.geodata(enero1[enero1$id3==i,], coords.col = 6:7, data.col = 5)
  modelo<-model.control(cov.model="exponential")
  model<-krige.bayes(contGEO, model=modelo)
  matrizenero[i,1]<-mean((model$posterior$phi$phi.marginal[,3]-
  model$posterior$phi$phi.marginal[,2])^2)^.5
  modelo<-model.control(cov.model="spherical")
  model<-krige.bayes(contGEO, model=modelo)
  matrizenero[i,2]<-mean((model$posterior$phi$phi.marginal[,3]-
  model$posterior$phi$phi.marginal[,2])^2)^.5
  modelo<-model.control(cov.model="gneiting")
  model<-krige.bayes(contGEO, model=modelo)
  matrizenero[i,3]<-mean((model$posterior$phi$phi.marginal[,3]-
  model$posterior$phi$phi.marginal[,2])^2)^.5}
  exp<-colMeans(matrizenero)[1]
  esf<-colMeans(matrizenero)[2]
  gne<-colMeans(matrizenero)[3]

vario1<-list("|n","exp",exp, "sphe",esf, "gne",gne, "|n")
vario1
```

###La mejor estructura de dependencia espacial, es decir el mejor variograma, ##será el que presente un menor valor de 'vario1'

###Una vez escogida la estructura de dependencia espacial, se construirá una ##rejilla regular que abarca la zona de estudio a fin de predecir las ##concentraciones del contaminante en esos puntos.

Se crea una rejilla de 15*15 puntos en toda la superficie del mapa ## aitanaGB.shp. El número de puntos a predecir se evaluará en cada caso.

```
library(maptools)
library(tripack)
library(spdep)
library(SparseM) ## Estas librerías permiten importar y trabajar con mapas en
## formato shape
```

**## Se importa el mapa, se representa y se transforma a una matriz de
vecindad espacial**

```
x1<-read.shape("aitanaGB.shp")
plot.Map(x1)
x.2 <- Map2poly(x1)##matriz de vecindad espacial
```

#Creación de la rejilla en base a la matriz de vecindad espacial

```
new.data<-expand.grid(seq(max(attributes(x.2)$maplim$x),
min(attributes(x.2)$maplim$x),l=15),seq(min(attributes(x.2)$maplim$y),max(attributes(
x.2)$maplim$y),l=15))
plot.polylist(x.2,border="grey")
points(new.data)
new.data<-as.matrix(new.data)
```

###conocido el tipo de variograma (el 'gneiting' en todos los casos), se realiza el ##proceso de kriging para la rejilla de 15*15

```
matrizenero1<-matrix(0,225,93)
matrizenerovarianza1<-matrix(0,225,93)
```

##El siguiente algoritmo es el que contiene la instrucción de kriging ordinario

```
for(i in 1:93){
  contGEO<- as.geodata(enero1[enero1$id3==i,], coords.col = 6:7, data.col = 5)
  model<-krige.bayes(contGEO,loc=new.data,
  model=model.control(cov.model="gneiting"))
  matrizenero1[,i]<- model$predictive$mean.simulations
  matrizenerovarianza1[,i]<- model$predictive$variance.simulations}
```

```
Matriz<-matrizenero1
Matriz1<-Matriz[,1:93]
Matrizsd<-matrizenerovarianza1
Matrizsd1<-Matrizsd[,1:93]
```

##La predicción, kriging, se realiza para cada uno de los días. La predicción final ##para cada punto se obtiene mediante una media ponderada de las predicciones ##de cada día.

```
pred1<-matrix(0,225,1)
std1<-matrix(0,225,1)
for(i in 1:225){
  pred1[i,<- weighted.mean(Matriz1[i,],1/(Matrizsd1[i,]),na.rm=T)
  std1[i,<- mean(Matrizsd1[i,]^5,na.rm=T)}
```

##A veces conviene expresar esta predicción de forma que sea interpretable. Se ha ##optado por deshacer la transformación inicial. Es decir, se suman a los residuos ##la tendencia temporal que se ha extraído anteriormente.

```
no2<-data.frame(mes=datos$MES,año=datos$AÑO, no=c(NO27CA, NO2ABA, NO2ADR,
NO2ALG, NO2AST, NO2BAS, NO2BCA, NO2DEU, NO2ELO, NO2ERA, NO2GET,
NO2HOR, NO2IND, NO2ING, NO2LEI, NO2MAZ, NO2MDI, NO2MUS, NO2NAU,
NO2PEN, NO2SAN, NO2SGR, NO2SON, NO2TRA, NO2TXU, NO2ZIE, NO2ZOR))#se
construye una base de datos con los #valores originales del contaminante
```

```
no21<- no2[no2$mes==1,]#se coge solamente enero
no21<-no21[no21$año>1995 & no21$año<1999,]#período 96-98
nuevos.datos1<-data.frame(no2=no21[,3],año=no21[,2])#base de datos #únicamente de
las variables de interés.
```

```
library(nlme)#librería para estimar modelos mixtos
```

#Se calcula la media de la serie temporal del contaminante utilizando el siguiente #modelo mixto. Este modelo permite obtener una media diferente por cada año. Se #ha optado por coger la media del primer año del período.

```
model1<-lme(no2 ~1,random=~1|año,
nuevos.datos1,na.action=na.exclude,weights=varConstPower(form=~fitted(.)))
```

##Esta media se suma a las predicciones de los residuos para obtener la predicción ##final

```
pred.1<- pred1+rep(coef(model1)[[1]][1],225)
pred1<-data.frame(pred.1,std1)
```

```
predicciones1<-data.frame(lon=new.data[,1],lat=new.data[,2],pred1)
```

```
write.table(predicciones1,"prediccionesNO2enero1.txt")
```

5.2.- Suavización y estimación de las RIE/RME

```
setwd("D:/Treballs varies/.../Tesis/Modelos")
library(foreign)
options(contrasts=c("contr.treatment", "contr.poly"))
```

```
library(maptools)
library(tripack)
library(spdep)
library(SparseM)
library(R2WinBUGS)
```

```
datos<-read.spss("mortsAMB.sav")
attach(datos)
```

```
x1 <- read.shape("AMB(206).shp")
```

```
plot.Map(x1,xlab="Lon",ylab="Lat")
```

```
datos$SECCIO<-DISTRITO*10000+SECCION
seccio<-x1$att$DISTRITO*10000+x1$att$SECCION
orden4 <- match(seccio,datos$SECCIO)
```

```
x.2 <- Map2poly(x1)
amb1 <- poly2nb(x.2)
```

```
obs<-OD212[orden4]
esp<-ED212[orden4]
```

MODELO DE BESAG,YORK Y MOLLIE

```
dataAMB <- list(N=length(amb1),O=obs, E=esp, num=sapply(amb1,length),
adj=unlist(amb1),sumNumNeigh=length(unlist(amb1)))
```

```
parametersAMB <- c("RR","PRP","sdS","sdeta","alpha")
```

```
initsAMB <- function(){list(alpha=rnorm(1),
S=rnorm(sapply(amb1,length)),eta=rnorm(sapply(amb1,length)), sdS
=runif(1,0,5), sdeta=runif(1,0,5))}
```

```
memory.limit(4000)
```

```
simu.hierar <- bugs(dataAMB, initsAMB, parametersAMB,
model.file = "ModeloBYM.txt", n.chains = 2, n.iter = 100000 , n.burnin = 10000,
n.thin = 30, debug = TRUE, DIC = FALSE, digits = 5, codaPkg = FALSE,
bugs.directory = "C:/Program Files/WinBUGS14", working.directory = NULL)
```

RESULTADOS

```
print(simu.hierar,digits=5)
```

```
RR<- ifelse(is.na(obs),NA, simu.hierar$mean$RR)
sdRR<- ifelse(is.na(obs),NA, simu.hierar$sd$RR)
PRP<- ifelse(is.na(obs),NA,simu.hierar$mean$PRP)
sdPRP<- ifelse(is.na(obs),NA,simu.hierar$sd$PRP)
resultat<-data.frame(orden4,RR,sdRR,PRP,sdPRP)
write.table(resultat,"GironaRRd212BYM.txt")
```

```
sdS<- simu.hierar$mean$sdS
sd.S<- simu.hierar$sd$sdS
sdeta<- simu.hierar$mean$sdeta
sd.eta<- simu.hierar$sd$sdeta
mu<-t(apply(simu.hierar$sims.list$RR,1,function(x){x*esp}))
D<-apply(mu,1,function(x){-2*sum(obs*log(x)-obs*log(obs+0.000000000001)-
x+obs,na.rm=T)})
Dmedia<-mean(D,na.rm=T)
mumedia<-apply(simu.hierar$sims.list$RR,2,mean)*esp
DenMedia<- -2*sum(obs*log(mumedia)-obs*log(obs+0.000000000001)-
mumedia+obs,na.rm=T)
DIC<-2*Dmedia-DenMedia
```

```

AICs<-Dmedia+2*(var(D,na.rm=T)/2)
BICs<-Dmedia+log(length(obs))*(var(D,na.rm=T)/2)
resultat<-rbind(sdS,sd.S,sdeta,sd.eta,AICs,BICs,DIC)
write.table(resultat,"GironaStatsdBYM212d.txt")

RIE<-obs/esp
SIR<-RIE
par(mfrow=c(1,2))
plot(density(SIR,na.rm=T),xlab="",main="SIR")
abline(v=mean(SIR,na.rm=T))
plot(density(RR,na.rm=T),xlab="",main="RR")
abline(v=mean(RR,na.rm=T))
par(mfrow=c(1,1))

savePlot("DRRPRP212BYM",type = c("wmf"))

library("RColorBrewer")
cortes=c(-1,0.2,0.8,100)
Paleta<-brewer.pal(length(cortes)-1, "RdBu")
mipaleta<-Paleta[length(Paleta):1]
aux<-density(PRP,na.rm=T)
n<-length(aux$x)
qx<-aux$x
qy<-aux$y
colores<-as.numeric(as.factor(cut(aux$x,cortes)))
plot(aux,main="",xlab="Probabilidad a posteriori",ylab="Densidad de probabilidad",
frame.plot=F)
for(i in 1:(length(cortes)-1)){
if(any(colores==i)){segments(qx[colores==i],rep(0,sum(colores==i)),
qx[colores==i],qy[colores==i],lwd=3,col=mipaleta[i])}
lines(aux)

savePlot("D212BYM",type = c("wmf"))

```


###Modelo BYM.txt

```

model
{
  for (i in 1 : N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + S[i] + eta[i]
    RR[i] <- exp(alpha + S[i] + eta[i])
    PRP[i] <- step(RR[i]-1)
    eta[i] ~ dnorm(0, tau.eta)
  }
  # CAR prior distribution for spatial random effects:
  S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }

  # Other priors:
  alpha ~ dflat()
  tau.S <- pow(sdS, -2)
  sdS ~ dunif(0,5)
  tau.eta <- pow(sdeta, -2)
  sdeta ~ dunif(0,5)
}

```

MODELO HURDLE

```
dataGirona <- list(N=length(girona1),O= ifelse(obsN==0,NA,obs), E=esp,  
Y= ifelse(obsN==0,0,1), num=sapply(girona1,length),  
adj=unlist(girona1),sumNumNeigh=length(unlist(girona1)))  
  
parametersGirona <- c("RR","PRP","sdS","sdeta","alpha")  
  
initsGirona <- function(){list(alpha0=rnorm(1), alpha=rnorm(1),  
S=rnorm(sapply(girona1,length)),eta=rnorm(sapply(girona1,length)), sdS=runif(1,0,5),  
sdeta=runif(1,0,5))}  
  
simu.hierar <- bugs(dataGirona, initsGirona, parametersGirona,  
  model.file = "ModeloHurdle.txt",  
  n.chains = 2, n.iter = 50000 , n.burnin = 5000,  
  n.thin = 30, debug = TRUE, DIC = FALSE, digits = 5, codaPkg = FALSE,  
  bugs.directory = "c:/Archivos de programa/WinBUGS14/",  
  working.directory = NULL)
```

###ModeloHurdle.txt

```

model
{
  for (i in 1 : N) {
    Y[i] ~ dbern(p[i])
    logit(p[i]) <- alpha0 + S[i] + eta[i]
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + S[i] + eta[i]
    RR[i] <- exp(alpha0 + S[i] + eta[i]) * (exp(alpha + S[i] +
eta[i])) / (1 + exp(alpha0 + S[i] + eta[i]))
    PRP[i] <- step(RR[i] - 1)
    eta[i] ~ dnorm(0, tau.eta)
  }
  # CAR prior distribution for spatial random effects:
  S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }

  # Other priors:
  alpha ~ dflat()
  alpha0 ~ dflat()
  tau.S <- pow(sdS, -2)
  sdS ~ dunif(0, 5)
  tau.eta <- pow(sdeta, -2)
  sdeta ~ dunif(0, 5)
}

```

MODELO ZINB

```
dataGirona <- list(N=length(girona1),O=obs, E=esp,
num=sapply(girona1,length),
adj=unlist(girona1),sumNumNeigh=length(unlist(girona1)))

parametersGirona <- c("RR","PRP","sdS","sdeta","alpha")

initsGirona <- function(){list(alpha=rnorm(1), p=0.5, index=
ifelse(obsN==0,1,2),S=rnorm(sapply(girona1,length)),eta=rnorm(sapply(girona1,length)
), sdS=runif(1,0,5), sdeta=runif(1,0,5))}

simu.hierar <- bugs(dataGirona, initsGirona, parametersGirona,
  model.file = "ModeloZINB.txt",
  n.chains = 2, n.iter = 50000 , n.burnin = 5000,
  n.thin = 30, debug = TRUE, DIC = FALSE, digits = 5, codaPkg = FALSE,
  bugs.directory = "c:/Archivos de programa/WinBUGS14/",
  working.directory = NULL)
```

###ModeloZINB.txt

```

model {
  for (i in 1 : N) {
    mu[i,1]<-0
    mu[i,2]<-lambda[i]
    mu.i[i]<-mu[i,index[i]]
    index[i] ~ dcat(theta[1:2])
    # mixture
    O[i] ~ dpois(mu.i[i])
    log(lambda[i]) <- log(E[i]) + alpha + S[i] + eta[i]
    RR[i] <- exp(alpha + S[i] + eta[i])
    PRP[i]<-step(RR[i]-1)
    eta[i] ~ dnorm(0, tau.eta)
  }
  p ~ dbeta(1,1)
  theta[1]<-p
  theta[2]<-1-p
  # CAR prior distribution for spatial random effects:
  S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }
  # Other priors:
  alpha ~ dflat()
  tau.S <- pow(sdS,-2)
  sdS ~ dunif(0,5)
  tau.eta <- pow(sdeta,-2)
  sdeta ~ dunif(0,5)
}

```

MODELO MIXTO (Lawson y Clark)

```
dataGirona <- list(N=length(girona1),O=obs,
  E=esp, num=sapply(girona1,length),
  adj=unlist(girona1),sumNumNeigh=length(unlist(girona1)))

parametersGirona <- c("RR","PRP","sdS","sdeta","sdfi","alpha")

initsGirona <- function(){list(alpha=rnorm(1),
  S=rnorm(sapply(girona1,length)),eta=rnorm(sapply(girona1,length)),
  fi=rnorm(sapply(girona1,length)), sdS=runif(1,0,5), sdeta=runif(1,0,5),sdfi=runif(1,0,5) ,
  p=runif(sapply(girona1,length),0.5,1))}

memory.limit(4000)

simu.hierar <- bugs(dataGirona, initsGirona, parametersGirona,
  model.file = "Modelolaw.txt",
  n.chains = 2, n.iter = 70000 , n.burnin = 7000,
  n.thin = 30, debug = TRUE, DIC = FALSE, digits = 5, codaPkg = FALSE,
  bugs.directory = "c:/Archivos de programa/WinBUGS14/",
  working.directory = NULL)
```

###Modelolaw.txt

```

model
{
for (i in 1:N) {
  O[i] ~ dpois(mu[i])
  log(mu[i]) <- log(E[i]) + alpha + eta[i] + p[i]*S[i]+(1-p[i])*fi[i]
  RR[i]<-exp(alpha+eta[i]+p[i]*S[i]+(1-p[i])*fi[i])
  PRP[i]<-step(RR[i]-1)
  eta[i] ~ dnorm(0, tau.eta)
  p[i] ~ dbeta(0.5, 0.5)
}
# CAR prior distribution for spatial random effects:
S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
for(k in 1:sumNumNeigh) {
  weights[k] <- 1 }

# CAR-L 1 prior for the model that models discrete jumps:
fi[1:N] ~ car.l1(adj[], weights[], num[], tau.fi)

# Other priors:
alpha ~ dflat()
tau.S <- pow(sdS,-2)
sdS ~ dunif(0,5)
tau.eta <- pow(sdeta,-2)
sdeta ~ dunif(0,5)
tau.fi <- pow(sdfi,-2)
sdfi ~ dunif(0,5)
}

```

5.3.-Los modelos de asociación geográfica

...

```
obs<-OD212[orden4]
esp<-ED212[orden4]
index1<-INDGID1[orden4]
index2<-INDGID2[orden4]
```

BYM CON VARIABLE EXPLICATIVA EFECTOS FIJOS

```
dataGirona <- list(N=length(girona1),O=obs, E=esp, indice1 =index1,
indice2=index2,num=sapply(girona1,length),
adj=unlist(girona1),sumNumNeigh=length(unlist(girona1)))

parametersGirona <- c("RR","PRP","sdS","sdeta","beta1","beta2")

initsGirona <- function(){list(alpha=rnorm(1), beta1=runif(1,0,1),
beta2=runif(1,0,1),S=rnorm(sapply(girona1,length)),eta=rnorm(sapply(girona1,length)),
sdS =runif(1,0,5), sdeta=runif(1,0,5) )}

simu.hierar <- bugs(dataGirona, initsGirona, parametersGirona,
  model.file = "ModeloBYMvarexpl.txt",
  n.chains = 2, n.iter =90000 , n.burnin = 9000,
  n.thin = 30, debug = FALSE, DIC = FALSE, digits = 5, codaPkg = FALSE,
  bugs.directory = "c:/Archivos de programa/WinBUGS14/",
  working.directory = NULL)
```


###ModeloBYMvarexpl.txt

```

model
{
  for (i in 1 : N) {
    O[i] ~ dpois(mu[i])
    #indice[i] ~ dunif(-10,10)
    log(mu[i]) <- log(E[i]) + alpha + beta1*indice1[i]+beta2*indice2[i]+S[i] + eta[i]
    RR[i] <- exp(alpha + S[i] + eta[i]+beta1*indice1[i]+beta2*indice2[i])
    PRP[i]<-step(RR[i]-1)
    eta[i] ~ dnorm(0, tau.eta)
  }
  # CAR prior distribution for spatial random effects:
  S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1 }

  # Other priors:
  alpha ~ dflat()
  beta1 ~ dnorm(0,0.005)
  beta2 ~ dnorm(0,0.005)
  tau.S <- pow(sdS,-2)
  sdS ~ dunif(0,5)
  tau.eta <- pow(sdeta,-2)
  sdeta ~ dunif(0,5)
}

```

LAWSON CON VARIABLE EXPLICATIVA EFECTOS FIJOS (índice socioeconómico)

```
dataGirona <- list(N=length(girona1),O=obs, E=esp, indice1=index1,
indice2=index2,num=sapply(girona1,length),
adj=unlist(girona1),sumNumNeigh=length(unlist(girona1)))
```

```
parametersGirona <- c("RR","PRP","sdS","sdeta","sdfi","beta1","beta2")
```

```
initsGirona <- function(){list(alpha=rnorm(1), beta1=runif(1,0,1),
beta2=runif(1,0,1),S=rnorm(sapply(girona1,length)),eta=rnorm(sapply(girona1,length)),
fi=rnorm(sapply(girona1,length)), sdS=runif(1,0,5), sdeta=runif(1,0,5),sdfi=runif(1,0,5) ,
p=runif(sapply(girona1,length),0.5,1))}
```

```
simu.hierar <- bugs(dataGirona, initsGirona, parametersGirona,
  model.file = "Modelolawvarexpl.txt",
  n.chains = 2, n.iter = 70000 , n.burnin = 7000,
  n.thin = 30, debug = FALSE, DIC = FALSE, digits = 5, codaPkg = FALSE,
  bugs.directory = "c:/Archivos de programa/WinBUGS14/",
  working.directory = NULL)
```

###Modelolawvarexpl.txt

```

model
{
  for (i in 1:N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + beta1*indice1[i]+beta2*indice2[i]+eta[i] +
    p[i]*S[i]+(1-p[i])*fi[i]
    RR[i]<-exp(alpha+ beta2*indice2[i]+beta2*indice2[i]+eta[i]+p[i]*S[i]+(1-p[i])*fi[i])
    PRP[i]<-step(RR[i]-1)
    eta[i] ~ dnorm(0, tau.eta)
    p[i] ~ dbeta(0.5, 0.5)
  }
  # CAR prior distribution for spatial random effects:
  S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1 }

  # CAR-L 1 prior for the model that models discrete jumps:
  fi[1:N] ~ car.l1(adj[], weights[], num[], tau.fi)

  # Other priors:
  alpha ~ dflat()
  beta1 ~ dnorm(0,0.005)
  beta2 ~ dnorm(0,0.005)
  tau.S <- pow(sdS,-2)
  sdS ~ dunif(0,5)
  tau.eta <- pow(s deta,-2)
  sdeta ~ dunif(0,5)
  tau.fi <- pow(sdfi,-2)
  sdfi ~ dunif(0,5)
}

```

LAWSON CON VARIABLE EXPLICATIVA EFECTOS ALEATORIOS (contaminante atmosférico)

```
dataGirona <- list(N=length(girona1),O=obs, E=esp, indice1=index1,
indice2=index2,cont=contamean, sdcont=contasd,num=sapply(girona1,length),
adj=unlist(girona1),sumNumNeigh=length(unlist(girona1)))
```

```
parametersGirona <- c("RR","PRP","sdS","sdeta","sdfi","beta1","beta2","beta3")
```

```
initsGirona <- function(){list(alpha=rnorm(1), beta1=runif(1,0,1),
beta2=runif(1,0,1),beta3=runif(1,0,1), beta4=runif(1,0,1),
S=rnorm(sapply(girona1,length)),eta=rnorm(sapply(girona1,length)),
beta=rep(0,length(girona1)), fi=rnorm(sapply(girona1,length)), sdS=runif(1,0,5),
sdeta=runif(1,0,5),sdfi=runif(1,0,5), tau.beta=runif(1,0,5),
p=runif(sapply(girona1,length),0.5,1))}
```

```
simu.hierar <- bugs(dataGirona, initsGirona, parametersGirona,
  model.file = "ModelolawvarexpIEA.txt",
  n.chains = 2, n.iter = 70000 , n.burnin = 7000,
  n.thin = 30, debug = FALSE, DIC = FALSE, digits = 5, codaPkg = FALSE,
  bugs.directory = "c:/Archivos de programa/WinBUGS14/",
  working.directory = NULL)
```

###ModelolawwarexpIEA.txt

```

model
{
  for (i in 1:N) {
    O[i] ~ dpois(mu[i])
    indice[i] ~ dunif(-10,10)
    log(mu[i]) <- log(E[i]) + alpha +
beta1*indice1[i]+beta2*indice2[i]+beta[i]*cont[i]+beta3*cont[i]+beta4*sdcont[i]+eta[i]
+ p[i]*S[i]+(1-p[i])*fi[i]
    RR[i]<-exp(alpha+
beta1*indice1[i]+beta2*indice2[i]+beta[i]*cont[i]+beta3*cont[i]+beta4*sdcont[i]+eta[i]+
p[i]*S[i]+(1-p[i])*fi[i])
    PRP[i]<-step(RR[i]-1)
    eta[i] ~ dnorm(0, tau.eta)
    p[i] ~ dbeta(0.5, 0.5)}
  for (i in 1 : N) {
    beta[i]~ dnorm(0, tau.beta)}

# CAR prior distribution for spatial random effects:
S[1:N] ~ car.normal(adj[], weights[], num[], tau.S)
for(k in 1:sumNumNeigh) {
  weights[k] <- - 1 }

# CAR-L 1 prior for the model that models discrete jumps:
fi[1:N] ~ car.l1(adj[], weights[], num[], tau.fi)

# Other priors:
alpha ~ dflat()
beta1 ~ dnorm(0,0.005)
beta2 ~ dnorm(0,0.005)
beta3 ~ dnorm(0,0.005)
beta4 ~ dnorm(0,0.005)
tau.S <- pow(sdS,-2)
sdS ~ dunif(0,5)
tau.eta <- pow(sdeta,-2)
sdeta ~ dunif(0,5)
tau.fi <- pow(sdfi,-2)
sdfi ~ dunif(0,5)
tau.beta ~ dgamma(0.5, 0.0005)
sdbeta <- sqrt(1 / tau.beta)
}

```

Referencias

- Breslow N, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; 88:9–25.
- Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85: 398–409.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12: 609–628.
- McCullagh P, Nelder JA. *Generalized Linear Models*. Second Edition. Londres: Chapman & Hall, 1989.
- Saez M. El problema de las medidas repetidas. Análisis longitudinal en epidemiología. *Gaceta Sanitaria* 2001; 15(4):347–352.
- Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82: 528–549.
- Wolfinger R, O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 1993; 48:223–243.