Detection of spatial aggregation of cases of cancer from data on patients and health centres contained in the Minimum Basic Data Set

Pablo Fernández-Navarro,[1,2] Jose-Miguel Sanz-Anquela,[3,4] Angel Sánchez Pinilla,[5] Rosario Arenas Mayorga,[5] Carmen Salido-Campos,[6] Gonzalo López-Abente[1,2]

[1]*Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid;* [2]*Consortium for Biomedical Research in Epidemiology and Public Health (CIBER en Epidemiología y Salud Pública - CIBERESP);* [3]*Cancer Registry and Pathology Department, Hospital Universitario Principe de Asturias, Alcalá de Henares;* [4]*Department of Medicine and Medical Specialties, University of Alcalá;* [5]*Institute of Statistics, Council of Economy and Technological Innovation, Community of Madrid;* [6]*Admission and Clinical Documentation Department, Hospital Universitario Principe de Asturias, Alcalá de Henares, Spain*

Abstract

The feasibility of the Minimum Basic Data Set (MBDS) as a tool in cancer research was explored monitoring its incidence through the detection of spatial clusters. Case-control studies based on MBDS and marked point process were carried out with the focus on the residence of patients from the Prince of Asturias University Hospital in Alcalá de Henares (Madrid, Spain). Patients older than 39 years with diagnoses of stomach, colorectal, lung, breast, prostate, bladder and kidney cancer, melanoma and haematological tumours were selected. Geocoding of the residence address of the cases was done by locating them in the continuous population roll provided by the Madrid Statistical Institute and extracting the coordinates. The geocoded control group was a random sample of 10 controls per case matched by frequency of age and sex. To assess case clusters, differences in Ripley K functions between cases and controls were calculated. The spatial location of clusters was explored by investigating *spatial intensity* and its ratio between cases and controls. Results suggest the existence of an aggregation of cancers with a common risk factor such as tobacco smoking (lung, bladder and kidney cancers). These clusters were located in an urban area with high socioeconomic deprivation. The feasibility of designing and carrying out case-control studies from the MBDS is shown and we conclude that MBDS can be a useful epidemiological tool for cancer surveillance and identification of risk factors through case-control spatial point process studies.

Introduction

The study of the spatial distribution of disease incidence and mortality is a basic approach to find possible causes. Traditionally, aggregation (clusters) of health events has been recorded for certain large administrative geographical areas, such as the provinces and municipalities in Spain. These methods for geocoding health-related events have been used in most countries. The result is that the specific factors that cause aggregations of cases of disease cannot be investigated unless its size and geographical boundaries match with the spatial units coded. For years, this has been a recognised limitation (Gatrell and Loytonen, 1998). Attempts to overcome them have always clashed with the lack of an alternative method of spatial encoding of the health events. In recent years, the introduction of Google Maps, ggmap, ggplots, R packages, *etc*. and the development of spatial epidemiology (Elliott *et al.*, 2000) radically changed this scenario by allowing connection of geocoding health events to the address point level. With that, the limited and inefficient old coding approach, which assigned indi-

vidual records to areas or predefined regions, is no longer sound.

To know the real cancer incidence in a given population requires having access to a high-quality population-based cancer registry (Forman *et al*., 2013) and this implies that they must have a physical infrastructure and trained personnel. Usually the inclusion of each case requires considerable effort, which can delay the end of a registration period with several years, hampering the epidemiological surveillance of certain diseases. In this connection, the epidemiology services of the Autonomous Communities in Spain are making great efforts in integrating their information systems (cancer incidence registries, mortality registries, health surveys, clinical information) to produce an instrument for assessing morbidity and mortality for any type of pathology and surveillance of the non-communicable diseases (Mayoral Cortes *et al.*, 2016). Examples of these systems in Spain are the Minimum Basic Data Set (MBDS), the hospital cancer registries and the healthcare information systems.

The MBDS records information on the patients and health centres based on the coding of all diagnoses of hospitalised patients discharged each month and sent to the department of health in each local county. The MBDS is the only authority that has both state coverage, mandatory completion and linkage of administrative data with diagnoses regarding all patient information with exception of mortality. Many of the information systems described above, including the MBDS, have the information necessary to carry out geocoding, which allows performing spatial epidemiological investigations such as the detection of outbreaks of cases (Abellán *et al*., 2002; López-Abente and Ibáñez, 2002). This information has been used for reporting epidemiological surveillance with regional character or the study of certain diseases (Gil Prieto *et al.*, 2009; García-García *et al.*, 2010). However, its use in epidemiological surveillance of cancer is rare (Gil *et al*., 2007), which is because there is little information on the capacity or coverage of these systems with respect to case detection and because of the long induction periods of many cancers. In the case of MBDS, it has been used in numerous studies of epidemiological surveillance of various infectious diseases in Spain, but in cancer it has only been used to study the extent of admissions for cervix cancer during the period 1999-2002 (Márquez Cid *et al*., 2006).

The development of information systems in public health services invites the use of this infrastructure for health-related epidemiological studies. In this context, the aim of this paper is to explore the feasibility of MBDS as a tool for epidemiological research on cancer, specifically in monitoring its occurrence through the detection of spatial clusters of cases.

Materials and Methods

Case control studies of cancer patients were designed with special reference to the geographical coordinates of the patients' addresses of residence. These designs are included in the scope of spatial data analysis known as point processes (Diggle, 2003).

Study area

The location of the area of study, the town of Alcalá in Madrid, Spain, can be seen in Figure S1 (Supplementary file). This area is located in the centre of the country near the capital (Madrid).

Data

The MBDS hospital discharge data, obtained directly from the Prince of Asturias University Hospital (HUPA) for the time period between January 2012 and June 2014, was used. This is the only public hospital in the study area, and admissions collected in the MBDS from it are mostly from the region of the study. The HUPA has a hospital cancer register (HCR), which is not the case in all hospitals in Spain. The HCR started its activity in 2008 based on pathology reports only, but in mid-2011 the MBDS was added as source data. The data sources that currently feed the HUPA's HCR are MBDS, pathology and haematology reports, the proceedings/minutes of the hospital committee of tumours and chemotherapy lists. This circumstance allowed us to assess the MBDS as a source of alternative cancer records data using data from a HCR as the reference standard in another study already published in the same region of the study area (Fernández-Navarro *et al.*, 2016).

The MBDS records contain the birth, admission and discharge dates of every patient, as well as the patient's personal data and disease diagnose(s) code according to the International Classification of Diseases (ICD-9-CM), which can be up to 13 different diagnoses. In the case of cancer, diagnoses can correspond to incident tumours as well as to a history of cancer (prevalent and cured). All recorded cancer diagnoses from hospital admissions in the MBDS during the study period were collected. In order to carry out this selection, the 13 different diagnoses included in the MBDS for the whole period in each patient were assessed. Once a diagnosis of cancer (described below) appeared for a patient, he or she was selected as a case. In that way, the same patient can be accounted for different cancer cases groups. For the analysis, we have selected the following tumour sites: stomach cancer, colorectal cancer, lung cancer, breast cancer in women, prostate, bladder and kidney cancers, melanoma and haematological tumours (non-Hodgkin's lymphomas, myeloma and leukaemias).

Cases

Patients older than 39 years and registered in the MBDS during the study period by any of the diagnoses of cancer described above were selected.

Geocoding and selection of control groups

The Statistics Institute of the Community of Madrid in Spain geocode automatically and routinely use the home address of all those enrolled in the municipal register of the region to build a geocoded municipal register. The tools developed are applicable to records containing postal addresses. The geocoding of the cases in our study was carried out by locating them in the annual municipal rolls of all patients registered in the MBDS extracting the coordinates of their address of residence. For this, a file with identifiers but without diagnoses was referred to the institute. All patients non-domiciled in the municipality were excluded.

In this type of study, the control group provides information on the spatial heterogeneity of the population. To select the control group, the continuous register of inhabitants geocoded provided by the Statistical Institute of the Community of Madrid was used. A random sample of 10 controls per case matched by frequency of age and sex of each tumour location was obtained. The recorded variables were: coordinates (x and y) in the ED50/UTM zone 30N projection (European Petroleum Survey Group spatial reference 23030) (http://spatialreference.org/ref/epsg/ed50-utm-zone-30n/),

sex and year of birth. With respect to the coordinates of residence, the last digit of coordinates (x, y) was assigned randomly in order to preserve confidentiality. This change does not alter the results of the spatial analysis since the coordinates are expressed in meters.

Population coverage

The MBDS population coverage has been evaluated in order to have an approximate idea of cases not included in the hospital base of the study (Table 1). For this, the age-specific incidence rates of the population-based Cancer Registries of Cuenca and Tarragona was used to calculate the cases expected in Alcalá if it had an incidence similar to one of these provinces. These rates were obtained from the European population-based cancer registries database, (EUREG) (http://eco.iarc.fr/eureg/). These registries were selected because they are the provinces most similar to the municipality of Alcalá among all those existing in Spain.

Average socioeconomic status

Low socioeconomic levels have been associated with the development of different chronic diseases like cancer. In order to assess if it is a possible cause of the spatial clustering of this type of disease, the average socioeconomic status (ASE) by census track, included in the 2001 Spanish census, were represented on a map. The ASE is a combination of occupation variables, such as activity and professional status of the household, where a high value represents a high socioeconomic level.

Details at: http://www.ine.es/censo_accesible/es/glosario.html.

Statistical analysis

To assess the possible aggregations of cases, we studied the spatial distribution of cases by tumour location through Ripley K function and its difference between cases and controls (Diggle and Chetwynd, 1991) that is the standard procedure for cluster detection using point processes. The assessment of the statistical significance of spatial aggregation was investigated by random labelling and Monte Carlo simulation methods. For spatial location of possible clusters, the distribution of risk in the study area was evaluated through the *spatial intensity* of the process, *i.e.* estimating the frequency of cases at each specific location. The parameters of the scanning window were defined empirically. The ratio of spatial intensity of cases against controls is interpretable as a relative risk (RR) and it allows estimating the surface of risk by detecting the *peaks* exceeding the margin of statistical significance. These surfaces of tolerance, that show areas with excess risk, were obtained by random labelling and Monte Carlo simulation methods (Kelsall and Diggle, 1995). To facilitate this analysis, a simple polygon that contain cases and controls was defined. The estimated ratios of smoothed intensities for each tumour site, using the default kernel functions included in the software used, were represented in maps (Kelsall and Diggle, 1995). The results are shown in figures, each of which includes the k-difference (k cases-k controls) and two lines with the 95% confidence interval (CI) for each type of cancer. When the K-difference (y-axis) crosses the confidence bands is an indication of clustering at the distance in abscissa axis (Distance in meters). Moreover, surfaces of tolerance and Kernel ratio of the intensity (Relative Risks, RR) of cases and controls will be shown in maps to locate the possible spatial clusters for each type of cancer.

Cartography and software

For the mapping representation the municipal and provincial census, sectioning of the National Statistics Institute of Spain (INE) (http://www.ine.es/en/welcome.shtml) from official cartography, was used. All analyses and cartographic representations were made using software R and *maptools* and *splancs* libraries (R Development Core Team, 2005).

Results

The HUPA registered 283.796 patients between 1997 and 2011, 75.3% of whom (213.526) were found in the geocoded municipal register. Out of those 98.9% (211.165) had assigned coordinates, while 1.1% were without geocodes. However, 70.270 patients were not located in the municipal rolls (24.7%) and may to a large part correspond to non-residents in the municipality.

Table 1 shows the estimation of the population coverage for all analyzed tumour locations. In general, there was a good coverage

Table 1. The Minimum Basic Data Set (MBDS) population registry coverage (2008-2014) in the study region by reference to the incidence rates of the population cancer registries of Cuenca (2004) and Tarragona (2005).

	Actual no. of cases MBDS	Expected no. of cases Cuenca[a]	Expected no. of cases Tarragona[a]	MBDS % coverage Cuenca[a]	MBDS % coverage Tarragona[a]
Lung	340	518	558	65,64	60,93
Bladder	474	288	411	164,58	115,33
Colorectal	512	520	797	98,46	64,24
Stomach	148	272	168	54,41	88,10
Melanoma	52	55	120	94,55	43,33
Breast (women)	442	544	706	81,25	62,61
Prostate	486	626	586	77,64	82,94
Kidney	143	109	183	131,19	78,14
Haematological	258	358	341	72,07	75,66
Leukaemia	77	139	122	55,40	63,11
NHL[b]	134	111	170	120,72	78,82
Myeloma	47	108	49	43,52	95,92
Total	2,855	3,290	3,870	86,78	73,77

[a]To calculate the cases expected in the study region, the age-specific incidence rates of the population-based cancer registries of Cuenca and Tarragona were used; [b]NHL: Non-Hodgkin's lymphoma.

over what was expected using the incidence rates of the population cancer registries, but there were tumour locations that are over-represented, such as bladder cancer. The total coverage was about 74-87% with data reference from the register of Tarragona and Cuenca, respectively. Table 2 shows the number of cases recorded in the MBDS and the number of cases and controls included in the analysis. The study included 2,683 cases and 27,825 controls where the most frequent tumour was found to be colorectal cancer followed by breast cancer (in women) and prostate cancer.

Clusters of cases

Figure 1 shows the distribution of cases of lung cancer and its controls over the mapping of the sectioned census of the study region by way of example. The grey points correspond to the control group representing the spatial heterogeneity of the population. In this case, it seems that there is a high density of cases of lung
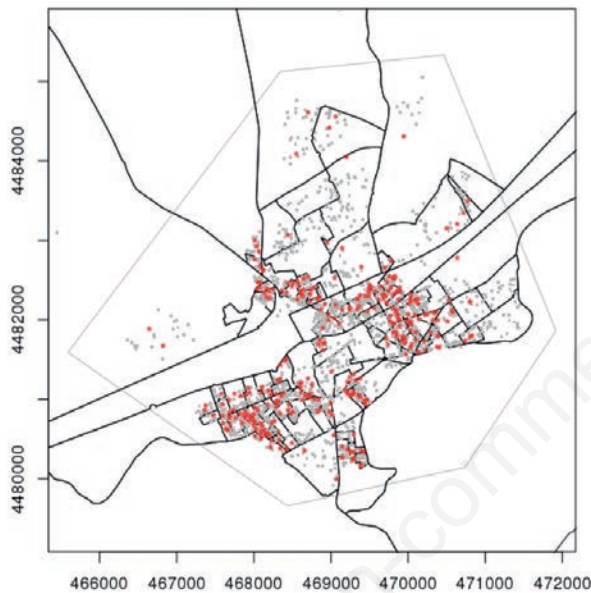
cancer in the Centre and in the South of the study region.

Figures 2-4 show the result of the detection of spatial clustering of cases relative to controls for the different tumour sites stud-
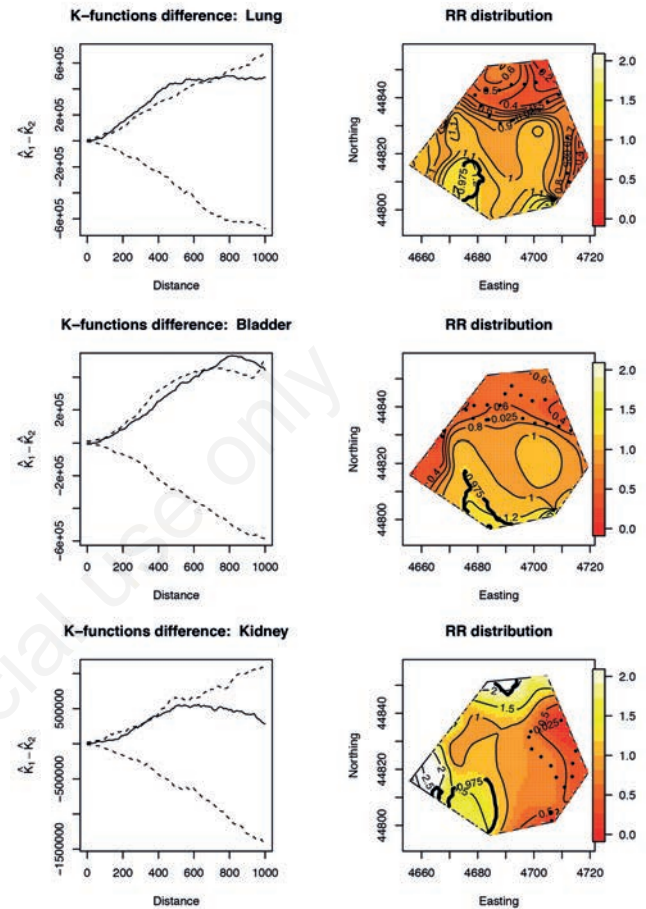

Figure 2. Lung, bladder and kidney cancer: estimation of the difference of K-functions.
On the left side: estimation of the difference of K-functions (k_1-k_2), *i.e.* controls subtracted from cases; On the right side: map surfaces of tolerance and intensity, *i.e.* kernel ratios of relative risk (RR); bold, solid lines indicate possible case cluster areas; dotted lines indicate the least RR; distance in the difference of K-functions plots in meters and the RR distribution maps axes in hectometers.

Figure 1. Distribution of lung cancer cases (red asterisks) and controls (gray dots) in the town of Alcala de Henares. Black lines correspond to the spatial boundaries of census sections and the gray line corresponds to the spatial polygon defined, which includes the subjects selected for the analysis.

Table 2. Number of cases registered in the Minimum Basic Data Set (MBDS) and the number of cases and controls included in the analysis.

	Cases MBDS[a]	Cases 40+ years	Cases included[b]	Controls included[b]
Lung	340	338	3,380	327	3,376
Bladder	474	471	4,710	455	4,707
Colorectal	512	499	4,990	472	4,990
Stomach	148	146	1,460	140	1,459
Melanoma	52	45	450	44	449
Breast (women)	442	422	4,220	408	4,218
Prostate	486	482	4,820	466	4,818
Kidney	143	139	1,390	139	1,389
Haematological	270	267	2,670	232	2,419
Total	2,867	2,809	28,090	2,683	27,825

[a]All the cases in the MBDS; [b]Number of cases and controls included in the analysis respectively.

ied. Specifically, Figure 2 shows the results for cancers of the lung, bladder and kidney, where the main risk factor is smoking. These three tumours show signs of spatial aggregation of cases on the difference of the K functions. Comparison of the spatial density of cases and controls point to an area in the Southwest of the municipality with a higher incidence than determined by random and marked with a thick continuous stroke in Figure 2. In contrast, there is no such kind of relevant spatial aggregation for the stomach or colorectal cancers (Figure 3).

Figure 4 shows the results of the analysis for the breast cancer (in women) and prostate cancer. In both, cases and controls the spatial density distributions are similar. The same figure shows the results for the haematological tumours in which there is a certain aggregation of cases but it does not exceed the envelope of tolerance (K-functions difference). The comparison of spatial densities designates an area of possible aggregation that partly coincides with the one shown by tumours in Figure 2. Finally, the detected

area in the Southwest as seen in the map of Figure 1, which corresponds to an area called the *Catholic Monarchs*, which has a low socioeconomic status according to the data. Figure 5 shows the average socioeconomic status for each of the census tracts in the study region. The information about this socioeconomic status is a combination of variables about occupation, activity and professional status of the household.

Discussion

The achievement of this work shows the feasibility of designing and carrying out cancer case-control studies from the MBDS. These data originated in the National Health System, not designed for this purpose but the idea is transferable to other data of public administration. Moreover, the main result of the study includes the
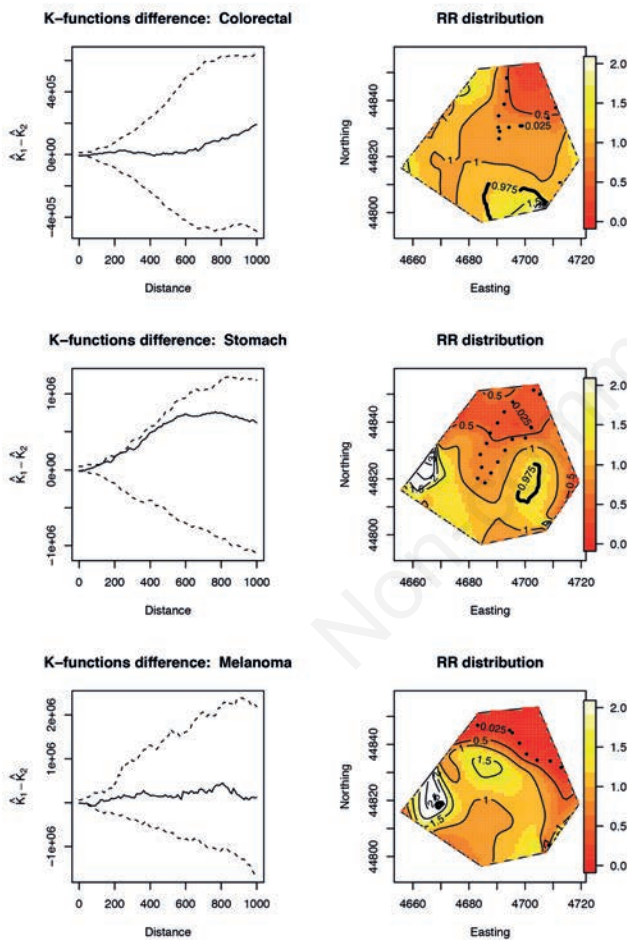

Figure 3. Colorectal, bladder and kidney cancer.
On the left side: estimation of the difference of K-functions (k_1-k_2), *i.e.* controls subtracted from cases; On the right side: map surfaces of tolerance and intensity, *i.e.* kernel ratios of relative risk (RR); bold, solid lines indicate possible case cluster areas; dotted lines indicate the least RR; distance in the difference of K-functions plots in meters and the RR distribution maps axes in hectometers.

Figure 4. Breast (women) and prostate cancers and haematological tumours.
On the left side: estimation of the difference of K-functions (k_1-k_2), *i.e.* controls subtracted from cases; on the right side: map surfaces of tolerance and intensity, *i.e.* kernel ratios of relative risk (RR); bold, solid lines indicate possible case cluster areas; dotted lines indicate the least RR; distance in the difference of K-functions plots in meters and the RR distribution maps axes in hectometers.
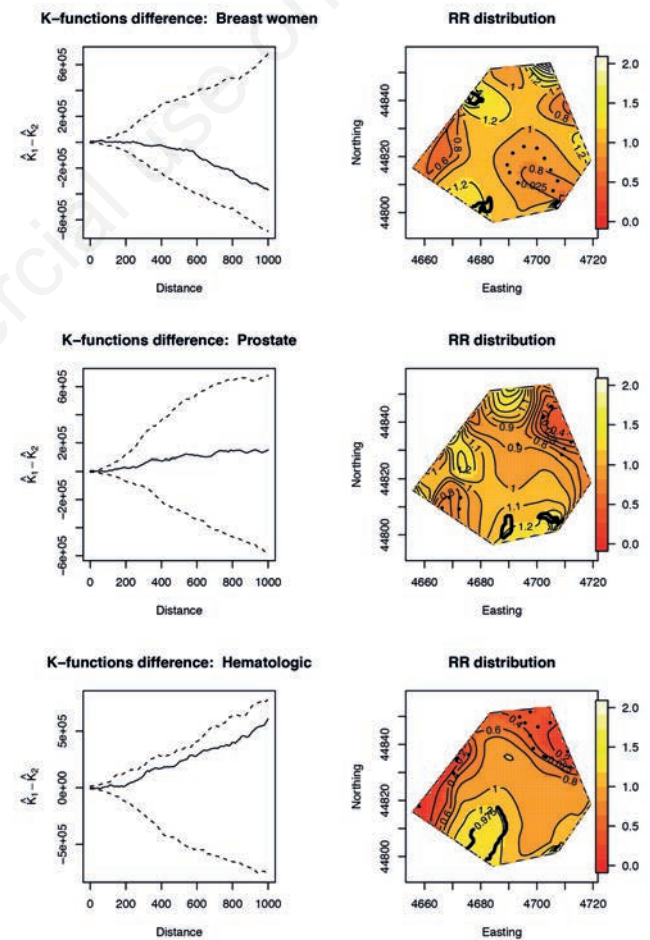
existence of a clear and similar pattern of aggregation of cases for cancers of the lung, bladder and kidney in a particular region of the urban area of study. These cancers have a common risk factor such as the use of tobacco (Adami *et al.*, 2008).

Internationally, smoking rates are particularly high among people who are socioeconomically disadvantaged (Hiscock *et al.*, 2012). This fact, could explain the aggregation of cases with tumours closely related to smoking found in the area with a low socioeconomic status depicted in Figure 5. Possibly there is a high rate of smoking in this area.

Minimum Basic Data Set as a tool in epidemiological research on cancer

This work shows the feasibility of using MBDS as an exploratory tool in epidemiological cancer research, providing useful information for cancer monitoring in a region and for identifying potential risk factors. The suitability of the use of the MBDS for any of the functions described, depends on the sensitivity of this registry to identify cancer cases. In this way, a previous study (Fernández-Navarro *et al.*, 2016) of our group has shown that it is 74% (72-76; 95% CI) for all cancer sites, although it varies depending on the type of tumour, reaching the highest values for bladder cancer (96% (92-98; 95% CI)) and their specificity and VPN were very high for all types of cancer studied, always on top of 95%. All these results suggest that, except for certain tumour sites, the MBDS can be a valid source for information to be used in epidemiological studies. Although the study that was carried out using the MBDS in a specific location, *i.e.* the region of Alcalá, it could serve as a pilot study for other regions of Spain, except for big cities like Madrid in central Spain and Barcelona in the northeastern part of the country. In these cities, health care is based on more complex models involved several public and private hospitals and health centres. The design could also be applied to hospital tumour registers where they exist. In our case, knowing that not all hospitals have such records, we have selected the MBDS as a source of cases looking for some portability design.

Figure 5. Average socioeconomic status by census tracts in Alcala de Henares.
2001 Spanish Census data from the National Statistics Institute.

ASE Quartiles
- ■ [0.79,0.95]
- ▦ (0.95,0.99]
- ▨ (0.99,1.08]
- ☐ (1.08,1.29]

Cancer *vs* other diseases

Although the study shown here focused on cancer, a pathology with a long induction period (Schottenfeld and Fraumeni, 2006), there are other situations where the spatial techniques used could have an important role. In this sense, the performance of studies that attempt to elucidate the environmental origin of a series of cases that cluster in time and space, like the one shown here in the manuscript, may be higher in the case of acute conditions such as communicable diseases and/or associated with emission sources of biological agents or toxic substances (*e.g.* outbreaks of *Legionella* infections or respiratory diseases such as asthma), where the induction periods are very short. These studies, apart from detecting cluster of disease, could help to find the emission sources and/or issuing alarms or recommendations to the population.

Weaknesses

The main limitation of this study is the possible existence of unidentified cases, due to problems related to geocoding and sensitivity of the MBDS registry. In relation to the first potential problem, the geocoding using the geocoded municipal register where the addresses where geolocated by professionals of the Community of Madrid, the 19% of addresses of all cancer cases were not found in this registry. These values do not vary too much depending on the type of tumour analyzed in this study (range 13-25%). This limitation does not seem to be too relevant in the context of the exploratory objectives of the study. Furthermore, the sensitivity of the MBDS to detect cancer cases described above is quite high. In that way, cancer cases included in the analysis seems to fit very well to the reality.

Another possible limitation is the non-inclusion of covariates in the detection of spatial clusters of cases. The inclusion of covariates in the analysis could also allow a statistical evaluation of the influence of certain factors in the spatial aggregations but this option exceeded the objectives of this work. Finally, the kernel smoothing techniques requires the specification of parameters related to the window size that could modify the obtained images. In our analysis we have not studied the differences that may exist by gender, or by the failure to take into account variables such as ethnicity or other intrinsic characteristics of census tracts given the exploratory nature of the study. An added difficulty to control contextual variables such as socioeconomic status is the absence of such data stratified by sex in the census from the INE. We should get this kind of information from other places.

Advantages and strengths

The main advantages of using the MBDS as an exploratory tool for spatial cancer research as shown in this article, are the speed and low cost of the process. The geocoding process by identifying patients in municipal records is also very efficient and allows selecting control groups with ease. Additionally, the design of the study shown here, that involves a close collaboration between different public institutions, could form the basis of an automatic warning system surveillance. These systems that should be pending updates of the MBDS, may emit different types of warnings managed by epidemiologists. All these processes could be extrapolated to primary care to provide a comprehensive alert system. Moreover, the study has other strengths derived from the validity of intrinsic inference of case studies and controls, and the data quality both in terms of coverage (Table 1) and detection of cancer cases in the region. Moreover, the design and the model

applied could be exported to many other areas. Finally, in the context of health surveillance, it should be noted that due to the clear exploratory nature of this study, all the results should be taken into account with caution. However, the inclusion of these techniques in the routine exploitation of hospital administrative data could be an easy first way to explore health problems and not only in hospitals because the primary health through its information systems could be included. Furthermore, the results from the study of clusters of disease could generate environment exposure hypothesis that could be affecting the health of people giving the opportunity to be verified by the competent authorities.

Conclusions

The MBDS can be a useful tool for epidemiological cancer research, both at the level of surveillance and identification of risk factors, through case-control spatial analysis based on point processes.

References

Abellán JJ, Martínez-Beneito MA, Zurriaga O, Jorques G, Ferrándiz J, López-Quílez A, 2002. Point processes as a tool for analyzing possible sources of contamination. Gac Sanit SESPAS 16, 445-9.

Adami HO, Hunter DJ, Trichopoulos D, 2008. Textbook of cancer epidemiology. Oxford, UK: Oxford University Press.

Diggle P, 2003. Statistical analysis of spatial point patterns. London, UK: Arnold.

Diggle PJ, Chetwynd AG, 1991. Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics 47:1155-63.

Elliott P, Wakefield JC, Best N, Briggs S, 2000. Spatial epidemiology. New York, NY: Oxford University Press.

Fernández-Navarro P, López-Abente G, Salido-Campos C, Sanz-Anquela JM, 2016. The Minimum Basic Data Set (MBDS) as a tool for cancer epidemiological surveillance. Eur J Intern Med 34, 94-7.

Forman D, Bray F, Brewster D, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher M, Swaminathan R, Ferlay J, 2013. Cancer incidence in five continents vol. X. Lyon: IARC Scientific Publications.

García-García L, Ariza Megía MC, Alvaro A, Gil de Miguel A, Gil-Prieto R, 2010. Epidemiology of hospitalizations due to syphilis in large urban areas in Spain between 1997 and 2006. Sex Reprod Healthc Off J Swed Assoc Midwives 1:123-7.

Gatrell A, Loytonen M, 1998. GIS and Health, GISDATA. London, UK: Taylor & Francis.

Gil A, San-Martín M, Gil R, Hernández V, Ribes J, González A, 2007. Burden of hospital admissions for cervical cancer in Spain during 1999-2002. Hum Vaccin 3:276-80.

Gil Prieto R, Alejandre CG, Meca AA, Barrera VH, de Miguel AG, 2009. Epidemiology of hospital-treated Salmonella infection; data from a national cohort over a ten-year period. J Infect 58:175-81.

Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M, 2012. Socioeconomic status and smoking: a review. Ann NY Acad Sci 1248:107-23.

Kelsall JE, Diggle PJ, 1995. Non-parametric estimation of spatial variation in relative risk. Stat Med 14:2335-42.

López-Abente G, Ibáñez C, 2002. Aplicación de técnicas de análisis espacial a la mortalidad por cáncer en Madrid, Documentos Técnicos de Salud Pública. Comunidad de Madrid, Madrid.

Márquez Cid M, Valera Niñirola I, Chirlaque López MD, Tortosa Martínez J, Párraga Sánchez E, Navarro Sánchez C, 2006. Validation of colorectal cancer diagnostic codes in a hospital administration data set. Gac Sanit SESPAS 20:266-72.

Mayoral Cortes JM, Aragonés Sanz N, Godoy P, Sierra Moros MJ, Cano Portero R, González Moran F, Pousa Ortega Á, Grupo de Vigilancia Epidemiológica de la Sociedad Española de Epidemiología, 2016. Chronic diseases as a priority for the public health surveillance system. Spain Gac Sanit SESPAS 30:154-7.

R Development Core Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schottenfeld D, Fraumeni JFJr, 2006. Cancer Epidemiology and Prevention. 3rd ed. New York, NY: Oxford University Press.