

RESEARCH ARTICLE

# Spatial Analysis of Childhood Cancer: A Case/Control Study

Rebeca Ramis<sup>1,2\*</sup>, Diana Gómez-Barroso<sup>2</sup>, Ibon Tamayo<sup>2,3</sup>, Javier García-Pérez<sup>1,2</sup>, Antonio Morales<sup>4,5</sup>, Elena Pardo Romaguera<sup>6</sup>, Gonzalo López-Abente<sup>1,2</sup>

**1** Environmental Epidemiology and Cancer Unit, National Centre for Epidemiology, Instituto de Salud Carlos III - ISCIII, Madrid, Spain, **2** Consortium for Biomedical Research in Epidemiology & Public Health (*CIBER en Epidemiología y Salud Pública-CIBERESP*), Madrid, Spain, **3** Public Health Division of Gipuzkoa, BIODonostia Research Institute, Department of Health of the regional Government of the Basque Country, Donostia, Spain, **4** Rare Disease Research Institute (Instituto de Investigación de Enfermedades Raras-IIER), Instituto de Salud Carlos III - ISCIII, Madrid, Spain, **5** Consortium for Biomedical Research in Rare Diseases (Centro de Investigación Biomédica en Red de Enfermedades Raras-CIBERER), Madrid, Spain, **6** Registro Español de Tumores Infantiles (RETI-SEHOP), Universidad de Valencia, Valencia, Spain

\* [rramis@isciii.es](mailto:rramis@isciii.es)



OPEN ACCESS

**Citation:** Ramis R, Gómez-Barroso D, Tamayo I, García-Pérez J, Morales A, Pardo Romaguera E, et al. (2015) Spatial Analysis of Childhood Cancer: A Case/Control Study. PLoS ONE 10(5): e0127273. doi:10.1371/journal.pone.0127273

**Academic Editor:** David O. Carpenter, Institute for Health & the Environment, UNITED STATES

**Received:** January 21, 2015

**Accepted:** April 14, 2015

**Published:** May 20, 2015

**Copyright:** © 2015 Ramis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data are the geographic coordinates of the address of cases and controls. The authors can not provide individual coordinates because these are under protection by the Spanish law LOPD 15/1999 [25]. Privacy, confidentiality and rights of the cases and controls were ensured by changing the last digits of every coordinate (X and Y) by a random number. However, the authors have included in the paper some maps with this information.

**Funding:** This study was funded by Spain's Health Research Fund (Fondo de Investigación Sanitaria - FIS 12/01416) <http://www.isciii.es/ISCIII/es/>

## Abstract

### Background

Childhood cancer was the leading cause of death among children aged 1-14 years for 2012 in Spain. Leukemia has the highest incidence, followed by tumors of the central nervous system (CNS) and lymphomas (Hodgkin lymphoma, HL, and Non-Hodgkin's lymphoma, NHL). Spatial distribution of childhood cancer cases has been under concern with the aim of identifying potential risk factors.

### Objective

The two objectives are to study overall spatial clustering and cluster detection of cases of the three main childhood cancer causes, looking to increase etiological knowledge.

### Methods

We ran a case-control study. The cases were children aged 0 to 14 diagnosed with leukemia, lymphomas (HL and NHL) or CNS neoplasm in five Spanish regions for the period 1996-2011. As a control group, we used a sample from the Birth Registry matching every case by year of birth, autonomous region of residence and sex with six controls. We geocoded and validated the address of the cases and controls. For our two objectives we used two different methodologies. For the first, for overall spatial clustering detection, we used the differences of K functions from the spatial point patterns perspective proposed by Diggle and Chetwynd and the second, for cluster detection, we used the spatial scan statistic proposed by Kulldorff with a level for statistical significance of 0.05.

[contenidos/fd-investigacion/fd-planificacion-2/accion-estrategica-salud.shtml](https://doi.org/10.1371/journal.pone.0127273).

**Competing Interests:** The authors have declared that no competing interests exist.

## Results

We had 1062 cases of leukemia, 714 cases of CNS, 92 of HL and 246 of NHL. Accordingly we had 6 times the number of controls, 6372 controls for leukemia, 4284 controls for CNS, 552 controls for HL and 1476 controls for NHL. We found variations in the estimated empirical  $D(s)$  for the different regions and cancers, including some overall spatial clustering for specific regions and distances. We did not find statistically significant clusters.

## Conclusions

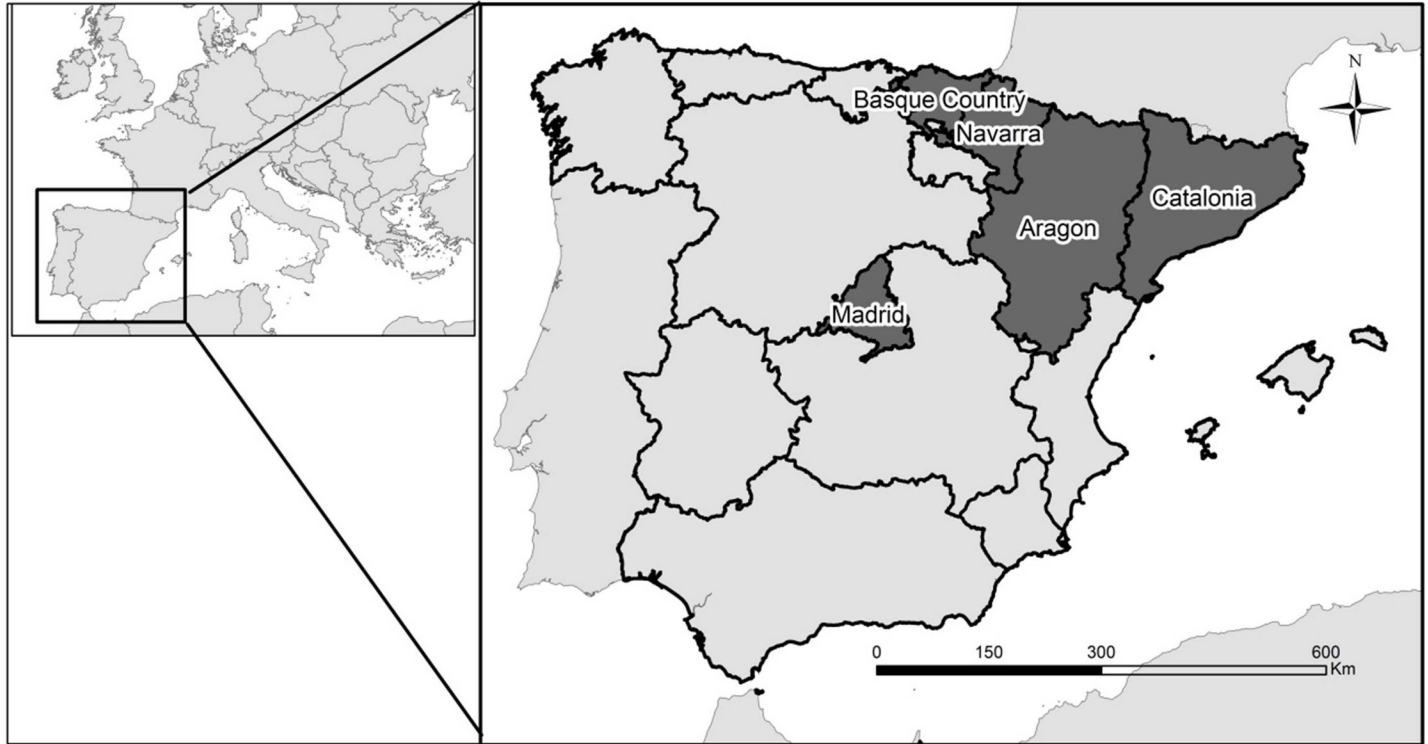
The variations in the estimated empirical  $D(s)$  for the different regions and cancers could be partially explained by the differences in the spatial distribution of the population; however, according to the literature, we cannot discard environmental hazards or infections agents in the etiology of these cancers.

## Introduction

Childhood cancer was the leading cause of death among children aged 1–14 years for 2012 in Spain [1]. Among the 12 major groups of childhood cancer in the International Classification of Childhood Cancer third edition (ICCC-3) [2], leukemia has the highest incidence (aged adjusted rates per million children 0–14 years of age): Europe 44.0, Spain 47.0; followed by tumors of the central nervous system (CNS): Europe 29.9, Spain 33.2; and lymphomas: Europe 15.2, Spain 19.4 [3,4]. Causes for childhood cancer are mainly unknown with the exception of a small percentage of cases attributable to hereditary cancer syndromes (familial retinoblastoma) or genetic syndromes and to exposure to ionizing radiation [5,6]. Early life exposure to environmental contaminants is suspected to be responsible for initial anomalies occurring in utero and leading to cancer [7]. Regarding leukemia many studies have addressed the hypothesis of infectious agents, but the association it is still not clear [8,9].

Spatial distribution of childhood cancer cases has been under concern in the last few decades [10–14]. In the nineties the EUROCLUS project for childhood leukemia analyzed the spatial distribution of 13351 cases diagnosed between 1980 and 1989 in 17 countries with the idea that the study of cluster and clustering could help to identify etiological factors. Their results indicated statistically significant evidence of clustering, but the magnitude was small [10,11,15]. A number of more recent studies have been conducted with this idea. A case-control study in California area showed no evidence of a non-random spatial pattern of childhood leukemia cases, although they have only 112 cases [14], two French studies with cases from the French National Register did not find statistically significant evidence of global heterogeneity of acute leukemia at small area level [12,16]; however, one study from the UK with data from the National Centre of Cancer Tumours found spatial clustering of leukemia in children aged 0–14 [17] and a second study found evidence of overall space-time clustering of childhood central nervous system tumors [13].

The study of the spatial distribution of cases can have two different purposes: one is overall spatial clustering analysis, which examines if the cases are closer to each other than the reference population; and the second purpose is cluster detection, the detection of a number of cases greater than expected in a specific geographical area. The objectives of this paper match these two purposes mentioned above. We studied overall spatial clustering and clusters of cases of the three main childhood cancer causes, looking to increase etiological knowledge.



**Fig 1. Map of Spain by regions.** The regions included in the study are highlighted.

doi:10.1371/journal.pone.0127273.g001

## Materials and Methods

### Cases

The Spanish Childhood Cancer Registry (RETI-SEHOP) collects information from all the pediatric oncology units in Spain and has the collaboration of the regional cancer registries. The completeness of the national coverage of childhood cancer by this registry is estimated at over 90% and 100% for the following five regions: Catalonia, Aragon, Navarre, the Basque Country and Region of Madrid [4]. The data used for the present study were children aged 0 to 14 diagnosed with a leukemia, lymphomas or CNS neoplasm, diagnostic groups I, II and III defined according to the 12 main diagnostic groups of the ICCC-3 [2]. For our analysis we separated the lymphomas into two groups, Hodgkin's lymphomas (HL) and non-Hodgkin's lymphomas (NHL). We included incidence cases from the five mentioned regions, four of them spatially contiguous located in the north-east part of Spain (North-East regions: Catalonia, Aragon, Navarre and the Basque Country) and an isolated one located in the center of Spain (Madrid). The studied period was 1996 to 2011 for all regions but Madrid, where the studied period was 2000 to 2011. (Fig 1 shows a map with the location of the regions).

Standard variables for each case included basic demographic data, such as date of birth, sex, province of residence and address at diagnosis. Also information on the diagnosis such as date, basis of diagnosis and morphology were included. We geocoded and validated the addresses of the cases by a geocoding strategy that is described later. We successfully validated 87% of the addresses. The remaining 13% of cases were fairly uniformly distributed along the different regions and therefore we did not think the data were biased in this sense.

## Controls

As a control group we used a sample from the population at risk extracted from the Birth Registry of the National Statistics Institute (Instituto Nacional de Estadística, INE). To select the controls we used a sampling strategy matching every case (with validated coordinates) by year of birth, region of residence and sex with six controls. Then we geocoded the addresses of the controls and we validated the coordinates. Only 2% of the controls did not have valid coordinates. Having had a small number of failures we decided to select more controls to replace this 2%, and we geocoded and validated this last group to end up with 6 controls with valid coordinates for every case.

## Geocoding strategy

For the purposes of this study, we conducted a retrospective geocoding (association of geographic coordinates from an input addresses) using the Google Map Javascript API v3. The obtained latitude and longitude data were projected into the ETRS89/UTM zone 30N (EPSG:25830) using QGIS software [18]. We then validated the coordinates and we kept those where the address and the coordinates matched. For the validation we performed the inverse process, we got the addresses of the obtained coordinates and we compared these new addresses to the original addresses. We compared town or city name, street name and street number.

## Methods

For our two objectives, overall spatial clustering analysis and cluster detection, we used two different methodologies. The first, for overall spatial clustering detection, was the differences of K functions from the spatial point patterns perspective proposed by Diggle and Chetwynd [19]. The second, for cluster detection, was the spatial scan statistic proposed by Kulldorff [20].

**Overall spatial clustering analysis.** Point patterns theory studies the spatial distribution of events occurring in a study region. The intensity of the phenomenon is the average density of points and it measures the ‘abundance’ or ‘frequency’ of the events recorded by the points. The intensity may be constant (‘uniform’ or ‘homogeneous’) or may vary from location to location (‘non-uniform’ or ‘inhomogeneous’). There are several methods to measure the intensity, one of them the K function proposed by Ripley [21]. The K function measures the aggregation of events at distance  $s$  and it is defined as:

$$K(s) = \frac{\text{mean of events at distance } \leq s \text{ from other event}}{\lambda}$$

where  $\lambda$  is the density for the whole region.

Generally, the distribution of the population in space is inhomogeneous, especially the distribution of the population at risk when we study health events. To assess if the events (cases) are somehow spatially aggregated we need to compare their spatial distribution with the spatial distribution of the population at risk (controls). One way to do this is to compare the intensity of the cases and the controls by the comparison of their K functions. The method was proposed by Rowlingson and Diggle [22] and it is implemented in the Splanacs library of R [21]. They defined  $D(s)$  as the difference between  $K(s)$  for the cases and  $K(s)$  for the controls.

$$D(s) = K_{cases}(s) - K_{controls}(s)$$

The null hypothesis is  $D(s) = 0$ , not differences between the distributions. The distribution of  $D(s)$  under the null hypothesis is computed by a Monte Carlo simulation using random labeling. An envelope with the limits of the  $D(s)$  under the null hypothesis is also computed at the same time. While the empirical  $D(s)$  is between the limits of the envelope there is no

evidence against the null hypothesis, only when  $D(s)$  is outside the envelope we can say that the spatial distribution of the cases is different to the spatial distribution of the controls. If the empirical  $D(s)$  is above the upper limit, we can say the cases are more aggregated than the controls and if the empirical  $D(s)$  is under the lower limit, controls are more aggregated than cases. We define as a maximum distance  $s$  equal to 8 km and we used the R software for the analysis [23].

**Cluster detection.** The spatial scan statistic is a test for spatial randomness based on likelihoods [20]. A cylindrical window that continuously changed its center and radius scanned the studied region seeking potential clusters. More precisely, the window moved from the address of a case to the address of another case. For each location the radius varied continuously from zero to a maximum distance (for our specific study we set a maximum of 5 km). Therefore, for each case the circular windows included different sets of neighboring cases and controls. For each location and size of scanning window the null hypothesis was that the risk was constant in space, the risk inside the window was the same that the risk outside. The alternative hypothesis was that the risk was higher inside than outside the window. During the process many different circular windows sides were evaluated in order to find the most likely cluster. Likelihood functions were calculated and maximized. The most likely cluster was the one with the maximum likelihood corresponding to a given location and specific radius. Its P value was obtained through Monte Carlo hypothesis testing (9999 replications), with a 95% confidence interval. Under the binomial assumption, the likelihood function for a specific window is proportional to:

$$LR_i = \left(\frac{n_i}{N_i}\right)^{n_i} \left(\frac{m_i}{M_i}\right)^{m_i} I\left(\frac{n_i}{N_i} > \frac{m_i}{M_i}\right)$$

For each potential cluster  $i$ ,  $n_i$  is the number of cases inside the potential cluster and  $m_i$  the number of cases outside, and  $N_i$  and  $M_i$  are the numbers at risk (cases and controls) inside and outside, respectively.  $I()$  is an indicator function that equals 1 when the risk inside the window is greater than the risk outside and 0 otherwise. We defined the level used for statistical significance as 0.05. Statistical analysis was performed with SaTScan 9.0.1 developed by Kulldorff [24]

**Ethical considerations.** Data used in this study are under protection by the Spanish law LOPD 15/1999 [25]. Privacy, confidentiality and rights of the cases and controls were ensured by changing the last digits of every coordinate (X and Y) by a random number.

## Results

After the geocoding and validation we had 1062 cases of leukemia, 714 cases of CNS, 92 of HL and 246 of NHL. Accordingly we had 6 times the number of controls, 6372 controls for leukemia, 4284 controls for CNS, 552 controls for HL and 1476 controls for NHL. For the analysis we separated the 4 Regions and Madrid, as we can see in Table 1 which shows the number of

**Table 1. Number of cases and controls by cause and region.**

Cancer site	Cases (Controls)North-East regions	Cases (Controls) Madrid
Leukemia	638 (3828)	424 (2544)
CNS	513 (3078)	201 (1206)
HL	53 (318)	39 (234)
NHL	144 (864)	102 (612)

doi:10.1371/journal.pone.0127273.t001

**Table 2. Number of cases by cause, region and sex.**

Region	Sex	Leukemia	CNS	HL	NHL
North-East regions	M	369	273	38	106
	F	269	240	15	38
Basque Country	M	76	42	6	11
	F	43	41	3	6
Navarra	M	23	23	1	7
	F	18	12	2	1
Aragon	M	34	30	4	13
	F	26	22	2	6
Catalonia	M	236	178	27	75
	F	182	165	8	25
Madrid	M	247	111	30	71
	F	177	90	9	31

doi:10.1371/journal.pone.0127273.t002

cases and control by cause and region. [Table 2](#) shows a disaggregation of cases by cause, administrative region and sex.

### Overall spatial clustering

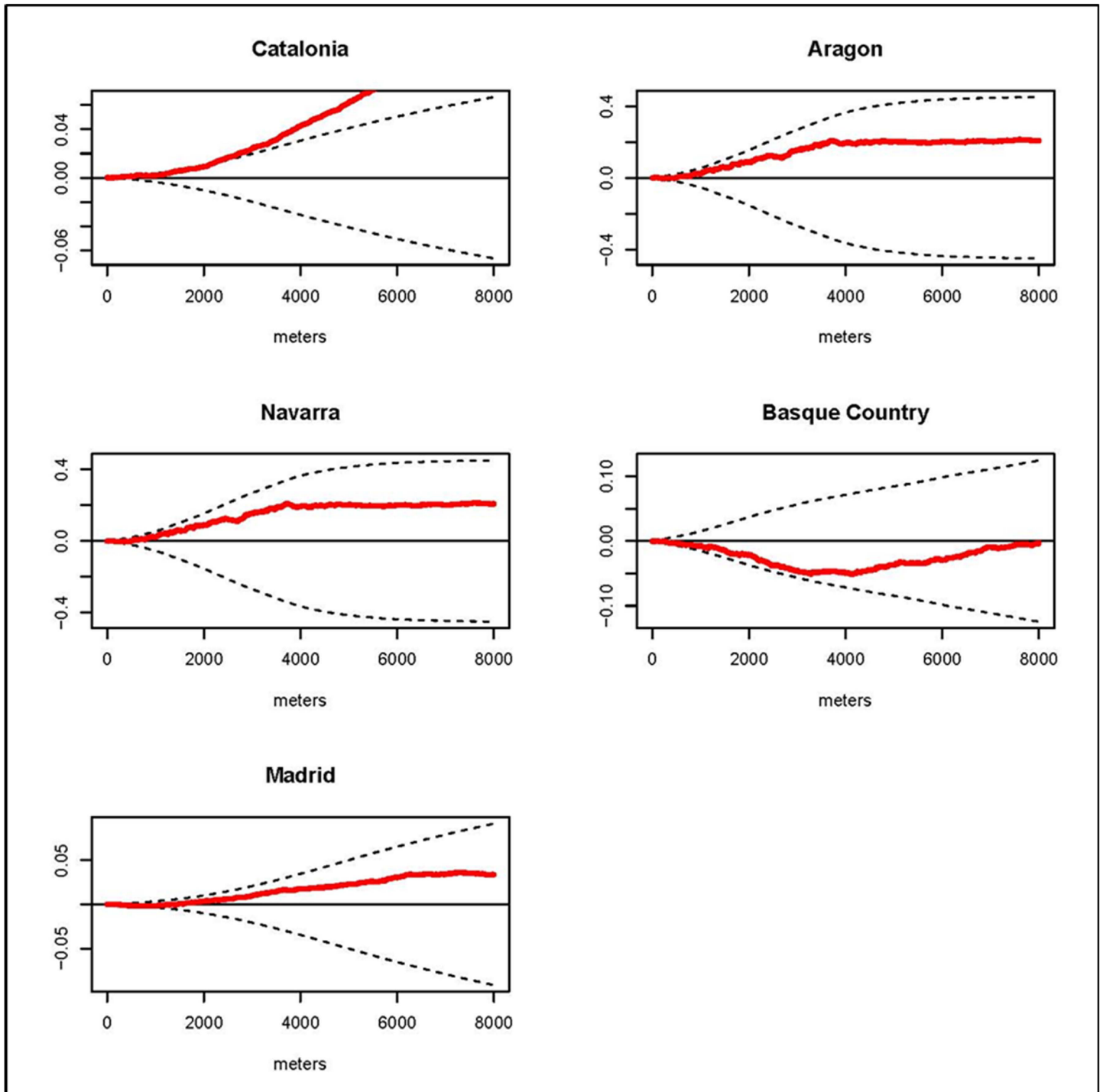
We estimated the  $D(s)$  statistic by region. The results for the K function by regions are shown in the graphs of the  $D(s)$  statistic included in [Figs 2, 3, 4](#) and [5](#). These graphs show the evolution of the  $D(s)$  statistic versus the distance from 0 to 8000 meters. Generally, the empirical  $D(s)$  statistics are inside the envelopes for most of the causes, regions and distances  $s$ . However, for leukemia the empirical  $D(s)$  shows fairly regional variation and even for Catalonia it exceeds the upper limit from distance 3km and for the Basque Country the empirical  $D(s)$  statistic is closer to the lower limit. For CNS in the Basque Country the empirical  $D(s)$  slightly exceeds the upper limit for distances smaller than 1 km, and in Madrid it is very close to the lower limit. For HL the empirical  $D(s)$  shows obvious variations between the different regions: exceeding the upper limit from distance 0 to distance 2 km in Catalonia, from distance 1km to 2km Aragon, and for distances from 3km in Navarra. And for NHL the empirical  $D(s)$  exceeds the upper limit from distance 2 km to distance 6 km for the Basque Country, and in Madrid it is very close to the lower limit.

### Cluster

The results from the spatial scan statistics are in [table 3](#). We present the details for the primary cluster of every cause and region. There was no a statistically significant cluster detected. The lowest p-value (0.063) was for an aggregation of four cases of NHL in Madrid, and the second lowest (0.074) was for an aggregation of 5 cases of leukemia in Barcelona. [Figs 6](#) and [7](#) show maps with the cluster of leukemia in Barcelona and the cluster of NHL in Madrid.

### Discussion

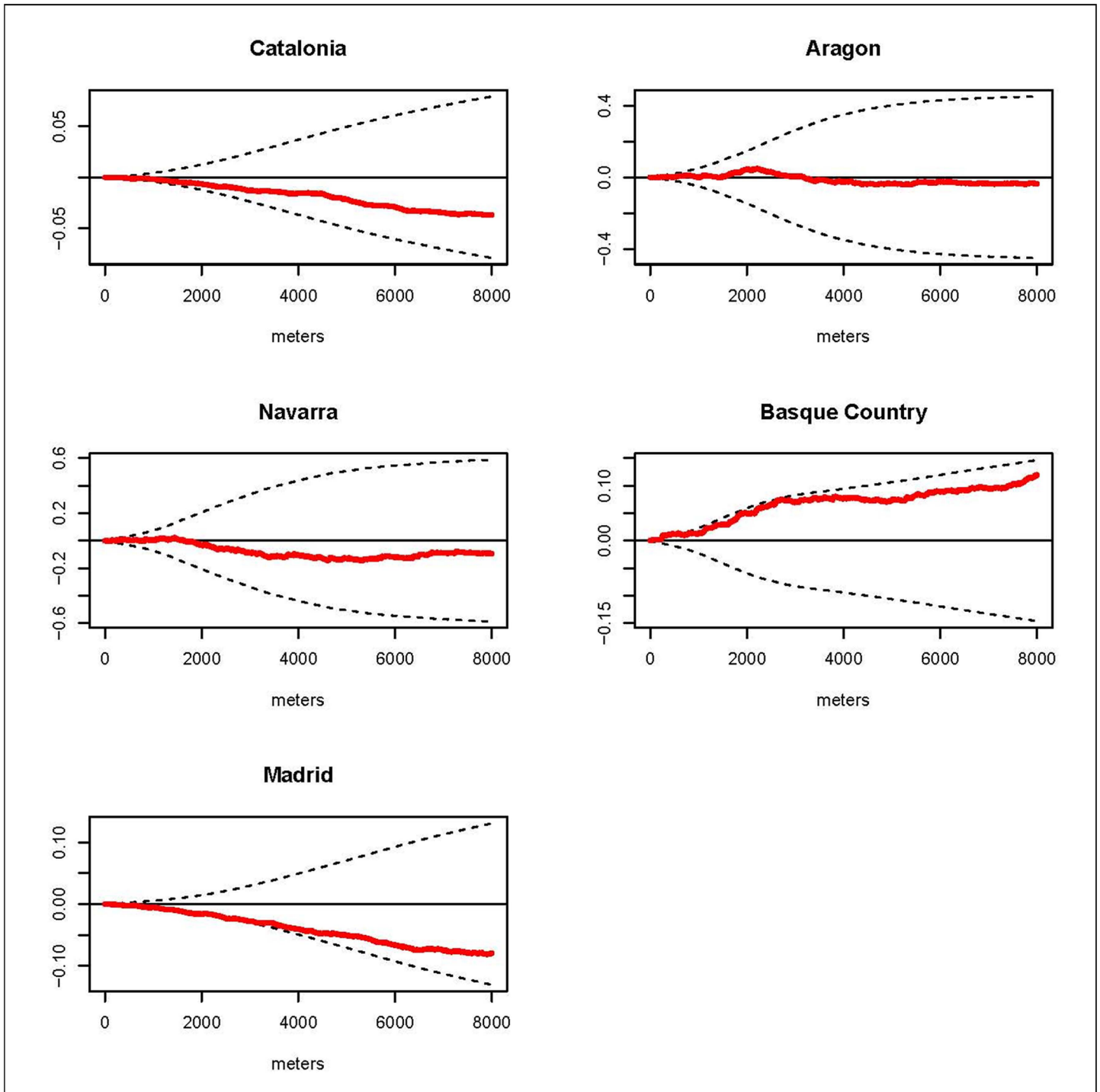
This case-control study analyzes the incidence geographical patterns of the main childhood cancers, looking for overall spatial clustering and clusters of cases. For overall spatial clustering analysis generally there were no statistically significant differences in the spatial distribution of the cases and controls for leukemia, CNS tumors, HL and NHL in the studied regions. Nevertheless, we found clustering for leukemia cases in Catalonia at distances superior 3 km; for



**Fig 2. Leukemia.** Graphs of the  $D(s)$  statistic function (red line) and the envelopes (dotted lines) from distance 0 to 8000 meters by region.

doi:10.1371/journal.pone.0127273.g002

CNS in the Basque Country at distances smaller than 1 km; for HL in Catalonia up to distance 2 km, in Aragon from distance 1km to 2km and in Navarra for distances from 3km; and for NHL in the Basque Country from distance 2 km to distance 6 km. Generally there were variations in the estimated empirical  $D(s)$  for the different regions and cancers. Regarding cluster

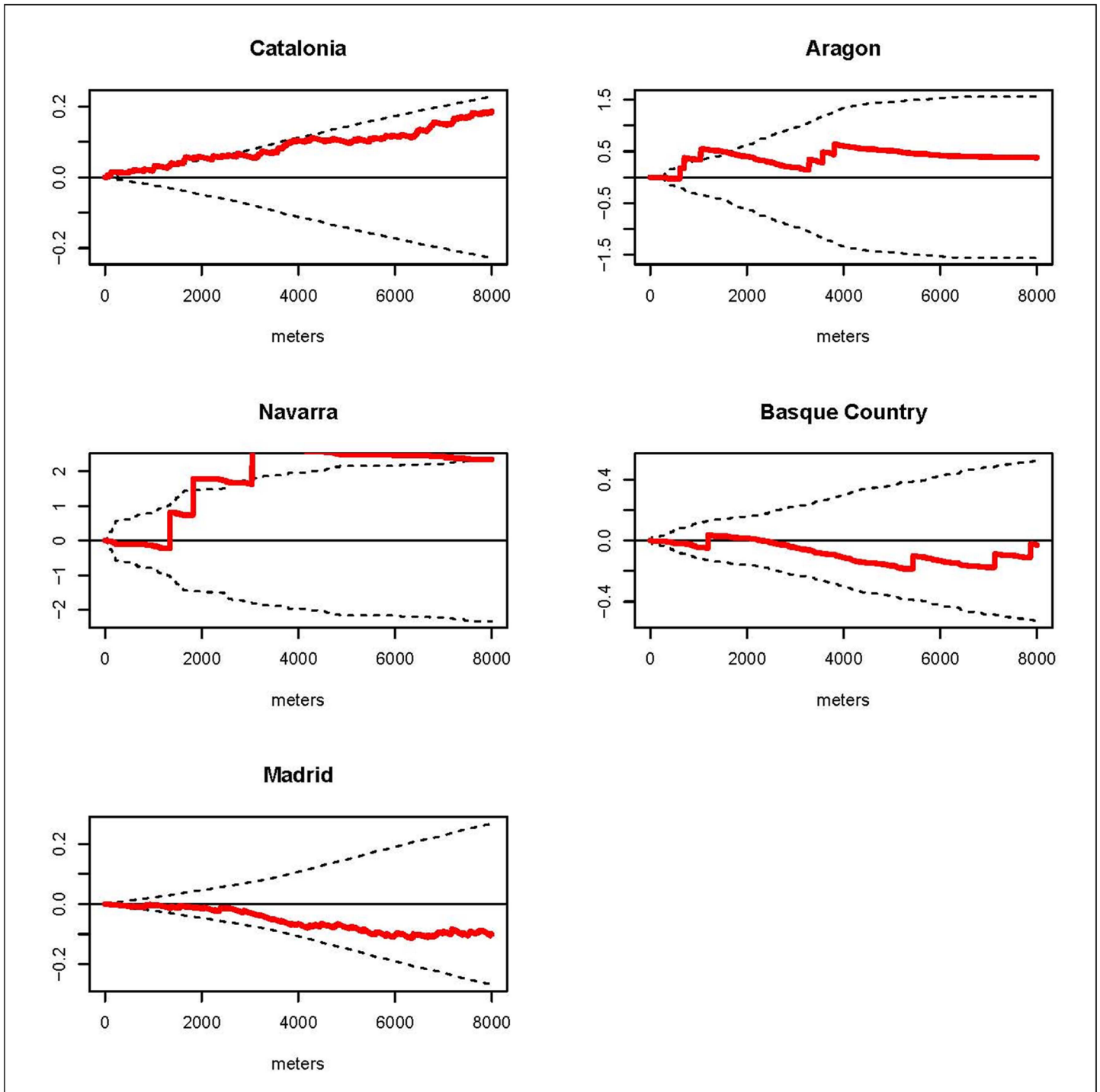


**Fig 3. CNS.** Graphs of the  $D(s)$  statistic function (red line) and the envelopes (dotted lines) from distance 0 to 8000 meters by region.

doi:10.1371/journal.pone.0127273.g003

detection, we did not find statistically significant clusters. Nonetheless the results of the spatial scan statistic identified two aggregations of cases for leukemia and NHL with p-values close to 0.05. The locations of these detected potential clusters did not match with distances at which the empirical  $D(s)$  statistics were out of the envelopes.

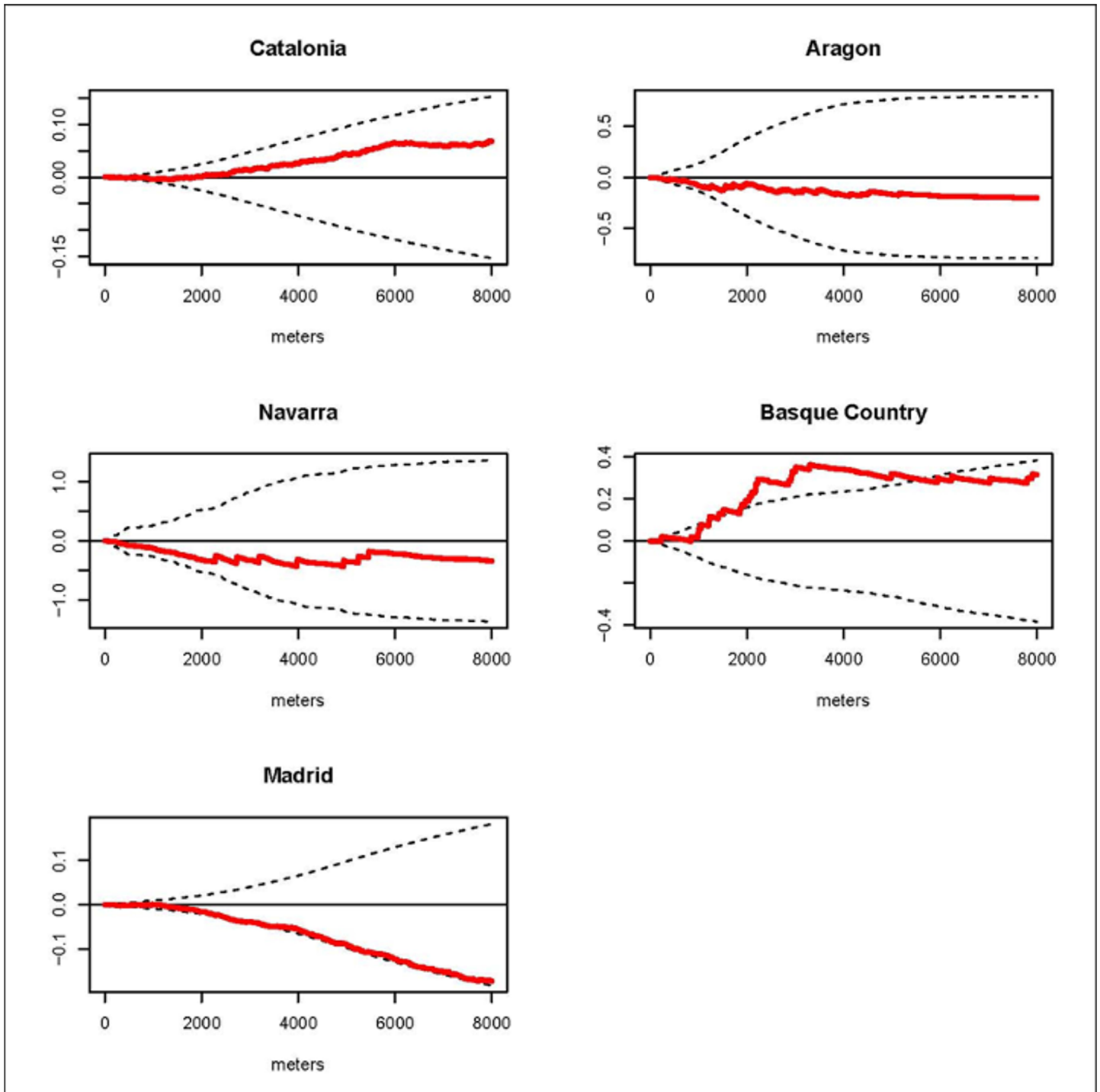




**Fig 4. HL.** Graphs of the  $D(s)$  statistic function (red line) and the envelopes (dotted lines) from distance 0 to 8000 meters by region.

doi:10.1371/journal.pone.0127273.g004

Leukemias are the most frequent type of tumor in children and make up about the 30% of the cases [3,4]. According to Peris et al. the Spanish age standardized rate (ARS) (World standard population) for the period 1983–2002 was 45.93 cases per million [26]. The few well established risk factors are inherited cancer-predisposition and exposure to ionizing radiation



**Fig 5. NHL.** Graphs of the  $D(s)$  statistic function (red line) and the envelopes (dotted lines) from distance 0 to 8000 meters by region.

doi:10.1371/journal.pone.0127273.g005

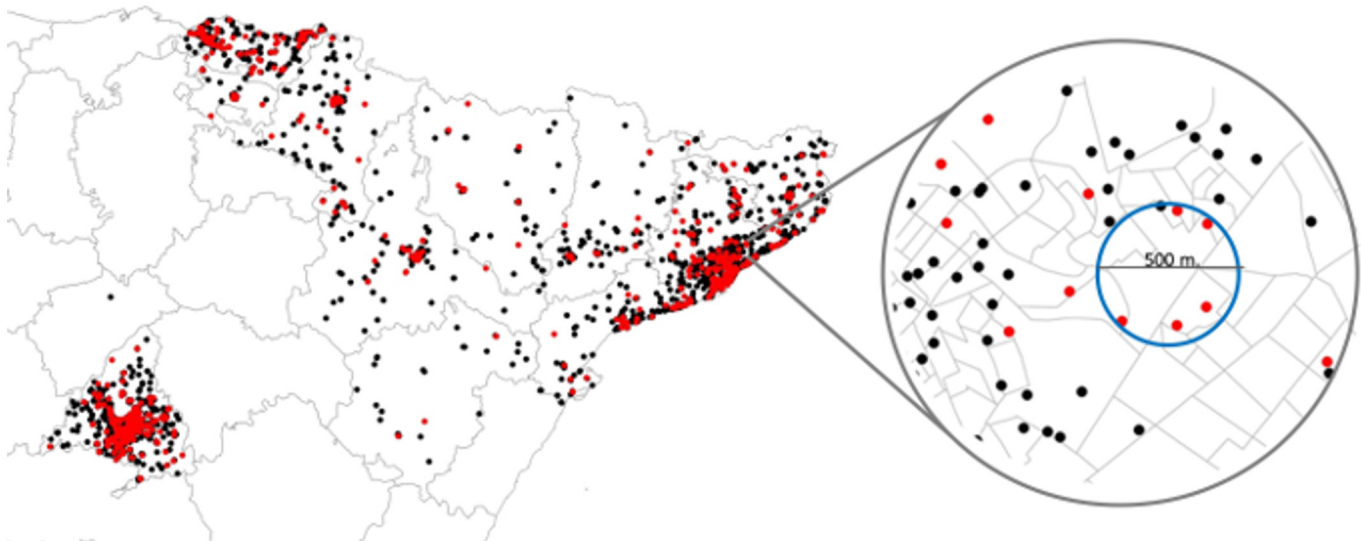
but these factors account for only a few cases [27]. In our study leukemia is the tumor with more cases, 1062 in total. The results for the overall spatial clustering analysis did not show more aggregation in the cases than in the controls at any distance in any region but Catalonia for distances superior to 3 km. Nevertheless, the empirical  $D(s)$  showed fairly regional variation

**Table 3. SatScan results.** Aggregation of cases with the lowest p-value by cause and region.

Cause	Region	Population	Cases	Expected cases	P-value	Radius (meters)	Location (Province)
Leukemia	NE regions	5	5	0.7	0.074	460	Badalona (Barcelona)
	Madrid	8	6	1.14	0.63	666	Colmenar Viejo (Madrid)
CNS	NE regions	20	10	2.86	0.92	886	Badalona (Barcelona)
	Madrid	9	6	1.29	0.54	1451	Madrid. Vallecas (Madrid)
HL	NE regions	3	3	0.43	0.122	124	Hospitalet de Llobregat (Barcelona)
	Madrid	2	2	0.29	0.54	573	Rivas Vaciamadrid (Madrid)
NHL	NE regions	3	3	0.43	0.48	619	Hospitalet de Llobregat (Barcelona)
	Madrid	4	4	0.57	0.063	643	Parla (Madrid)

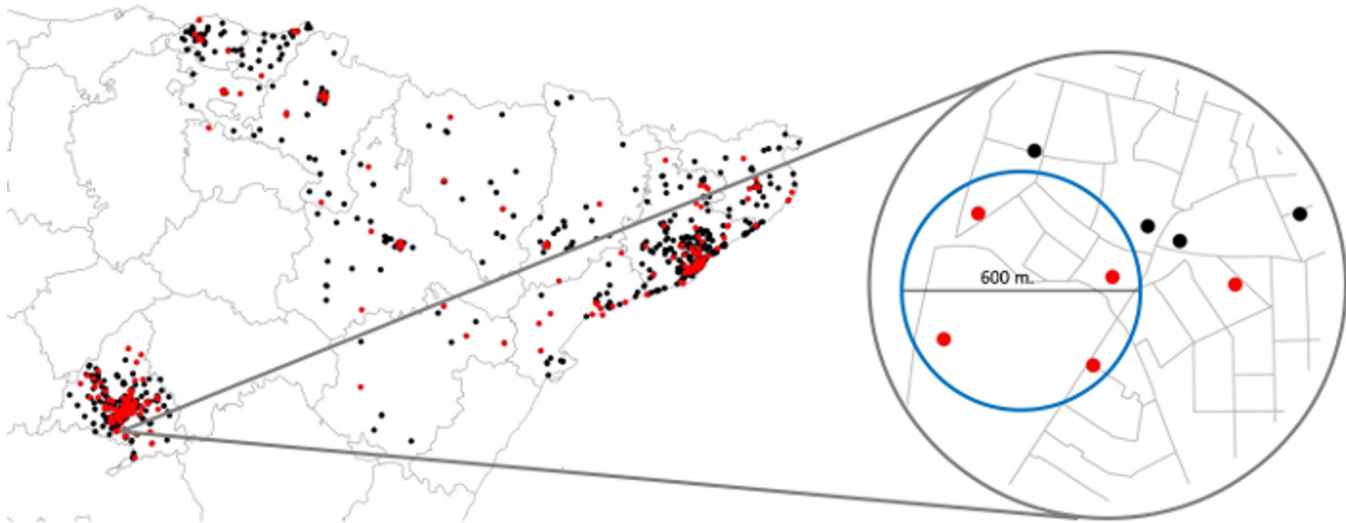
doi:10.1371/journal.pone.0127273.t003

especially in the Basque Country and Catalonia in comparison to other 3 regions. The spatial scan statistics showed an aggregation of 5 cases with a p-value of 0.074 in Barcelona. The hypothesis of clustering in children with leukemia has been studied in the past. In the late nineties, the project EUROCLUS's primary objective was to determine if the residence locations of cases at diagnosis showed a tendency towards spatial clustering; however, the results of the project could not confirm the hypothesis [11]. A case-control study in the San Francisco Bay area of the USA with 112 cases and 221 birth controls did not find evidence of non-random spatial patterns in the residences of the cases [14]. A French study of aggregated cases at area level (1916 areas for mainland France) did not find evidence of spatial heterogeneity either [12]. More studies have investigated links between exposure to environmental hazards and leukemia, a link that could show heterogeneity in the spatial pattern of the cases like those associated with exposure to pesticides [7]. Another etiological hypothesis that has been studied on many occasions is the exposure to infectious agents but its relation with leukemia is still unclear [8]. Our results did not show any spatial heterogeneity linked to environmental hazards, though this study was not designed with that aim in mind. Nonetheless, we cannot discard environmental hazards or infectious agents in the etiology of leukemia in children based on our results, a spatio-temporal analysis could be more conclusive in this sense.



**Fig 6. Map of leukemia cases and controls.** Cases in red dots and controls in black dots. The right hand side shows a zoom to the primary cluster suggested by the scan statistic.

doi:10.1371/journal.pone.0127273.g006



**Fig 7. Map of NHL cases and controls.** Cases in red dots and controls in black dots. The right hand side shows a zoom to the primary cluster suggested by the scan statistic.

doi:10.1371/journal.pone.0127273.g007

CNS tumors are the second most common cancer in children accounting for about 20% of the cases [3,4]. In Spain the estimated ASR for the period 1983–2002 was 32.83 cases per million [26]. For this study we had 714 cases of CNS tumors. Very little is known about the etiology of primary CNS and brain tumors. An estimated 5% of the cases may be explained by genetic predisposition, and the only established environmental risk factor is a high dose of ionizing radiation [28]. In our study CNS is the second most common tumor with 714 cases. For this cancer type neither of the two methods used suggested the presence of a pattern of overall spatial clustering or cluster, though the empirical  $D(s)$  for the Basque Country shows a tendency of clustering close to statistical significance. Results regarding the Basque Country and Navarra have special interest because these regions have showed higher incidence and mortality for CNS tumors at all ages [29]. On the other hand, several previous studies have found spatial patterns. A British study showed evidence of overall spatio-temporal clustering among cases of primitive neuroectodermal tumors [13]. Another British study found evidence of space-time clustering in cases of astrocytoma and ependymoma [30].

Lymphomas are the third most common cancer in children and make up around the 10–12% of the cases [3,4]. The estimated ASR for the period 1983–2002 for Spain was 18.48. For this study we separated the lymphomas into two groups: HL, with 92 cases, and NHL, with 247. HL in children is associated with the Epstein Barr virus [27]. For this lymphoma none of the two methods used suggested the presence of a pattern of spatial clustering or clusters, though the empirical  $D(s)$  is different for each of the 5 regions this could be due to the small number of cases. Little is known about the etiology of the NHL, but the associated genetic factors include congenital immunodeficiency syndromes [27]. Our results suggest no overall spatial clustering but the empirical  $D(s)$  is different for each of the 5 regions, and, again, this could be due to the small number of cases. There was a cluster of four cases of NHL in Madrid (p-value = 0.063).

One of the main strengths of our study is the large control group. Most studies of this type have one or two controls per case [14,31,32] in our study we have 6 controls per case and that gives a much more realistic image of the spatial distribution of the population at risk. Initially, we selected 6 controls as it is recommended by Rothman [33], since the georeference process was good and we got almost all the controls' addresses' coordinates, we decided to keep all of them and to replace the few that were missing with new controls.

The controls were randomly selected from birth certificates. This implies the possibility of having cases included in the control group, as excluding the cases as controls could bias the results [34]. The control group should give a clear view of the spatial distribution of the population at risk and should have the same risk of exposure as the cases. We matched the controls by sex, year of birth and region of residence to account for the temporal and regional variation in the child population. For the temporal trend there has been a moderate variation in the birth rate during studied period, starting with a birth rate of 1.15 for the year 1996 and reaching a maximum of 1.46 for year 2008 [1]. Regarding the regional variation, there are big differences between the regions; this study includes regions such as Madrid with a total population close to 6,500,000 inhabitants and Navarra with a total population of around 650,000 inhabitants for year 2012 [1]. This regional variation could be causing, to some extent, the variations observed in empirical  $D(s)$  for the different regions and causes.

It should be noted that we have the home address of the cases at the moment of diagnosis and the home address of the mother at birth for the controls. This difference could introduce bias in the analysis but according to official data, in Spain, only around 1% of the child population change their residence to a different province [1]. Therefore we considered that the home address at diagnosis is the same as the home address at birth for most of the cases.

We have used well-established methods for this analysis. For the overall spatial clustering analysis we used the method proposed by Diggle and Chetwynd for case-control studies. This method is based on the comparison of the second-order property of the observed point processes as a function of distance, a comparison made by the difference between the  $k$  functions [35]. The method has been used in different spatial analyses of epidemiological data [36,37] and it has an advantage above other methods in that the results show the specific distance at which the clustering occurs. The second method, Spatial scan statistics is a very popular method for cluster detection that has been used in many studies, primarily due to its availability in the free software package SatScan [24,35,38], its main flaw comes from the use of a regularly shaped window for the buffer [39].

A limitation for the study is that we performed the analysis across two disconnected areas, on one hand the north-east regions and on the other hand the Madrid region. However, this fact does not affect the results and conclusions. For overall spatial clustering we estimated the empirical  $D(s)$  for every region and cancer cause separately and then for cluster detection we defined 8 km as a maximum window. Consequently, we considered that the disconnection of the areas was not an issue. Another limitation is the lack of more information about potential exposure to risks factors for the children and their parents. The inclusion of these data in the analysis could provide more conclusive results.

## Conclusion

This study analyses the geographical patterns, overall spatial clustering and clusters, of individual incidence cases of childhood leukemia, CNS tumors, HL and NHL in Spain. We found spatial variation in the incidence of the main childhood cancers between the different regions and cancers, variation that could be partially explained by the differences in the spatial distribution of the population. Still according to the literature, we cannot discard the participation of environmental hazards or infectious agents in the etiology of these cancers.

## Acknowledgments

This study was funded by Spain's Health Research Fund (Fondo de Investigación Sanitaria—FIS 12/01416). We also want to thank Iñigo Tamayo for his technical support in the geocoding of the data.

## Author Contributions

Conceived and designed the experiments: RR DGB GLA. Performed the experiments: RR IT DGB JGP EPR. Analyzed the data: RR DGB. Contributed reagents/materials/analysis tools: RR IT DGB JGP EPR. Wrote the paper: RR. Critically reviewed the manuscript and approved the final version for submission: DGB IT JGP AM EPR GLA.

## References

1. Instituto Nacional de Estadística. (National Statistics Institute). Available: <http://www.ine.es/jaxi/tabla.do?path=/t15/p417/a2012/10/&file=01001.px&type=pcaxis&L=0>. Accessed 29 Sep 2014.
2. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer*. 2005; 103: 1457–1467. doi: [10.1002/cncr.20910](https://doi.org/10.1002/cncr.20910) PMID: [15712273](https://pubmed.ncbi.nlm.nih.gov/15712273/)
3. Stiller CA, Marcos-Gragera R, Ardanaz E, Pannelli F, Almar Marqués E, Cañada Martínez A, et al. Geographical patterns of childhood cancer incidence in Europe, 1988–1997. Report from the Automated Childhood Cancer Information System project. *Eur J Cancer Oxf Engl* 1990. 2006; 42: 1952–1960. doi: [10.1016/j.ejca.2006.05.017](https://doi.org/10.1016/j.ejca.2006.05.017)
4. Peris-Bonet R, Ruiz Martínez N, Felipe García S, Pardo Romaguera E, Valero Poveda S. Cáncer infantil en España. Estadísticas 1980–2012. Registro Nacional de Tumores Infantiles (RNTI-SEHOP). Valencia: Universitat de València; 2013.
5. Ross JA, Swensen AR. Prenatal epidemiology of pediatric tumors. *Curr Oncol Rep*. 2000; 2: 234–241. PMID: [11122848](https://pubmed.ncbi.nlm.nih.gov/11122848/)
6. Robison LL, Buckley JD, Bunin G. Assessment of environmental and genetic factors in the etiology of childhood cancers: the Childrens Cancer Group epidemiology program. *Environ Health Perspect*. 1995; 103 Suppl 6: 111–116. PMID: [8549456](https://pubmed.ncbi.nlm.nih.gov/8549456/)
7. Vinson F, Merhi M, Baldi I, Raynal H, Gamet-Payrastre L. Exposure to pesticides and risk of childhood cancer: a meta-analysis of recent epidemiological studies. *Occup Environ Med*. 2011; 68: 694–702. doi: [10.1136/oemed-2011-100082](https://doi.org/10.1136/oemed-2011-100082) PMID: [21606468](https://pubmed.ncbi.nlm.nih.gov/21606468/)
8. McNally RJQ, Eden TOB. An infectious aetiology for childhood acute leukaemia: a review of the evidence. *Br J Haematol*. 2004; 127: 243–263. doi: [10.1111/j.1365-2141.2004.05166.x](https://doi.org/10.1111/j.1365-2141.2004.05166.x) PMID: [15491284](https://pubmed.ncbi.nlm.nih.gov/15491284/)
9. Alexander FE. Viruses, clusters and clustering of childhood leukaemia: a new perspective? *Eur J Cancer Oxf Engl* 1990. 1993; 29A: 1424–1443.
10. Alexander FE, Boyle P, Carli PM, Coebergh JW, Ekbohm A, Levi F, et al. Population density and childhood leukaemia: results of the EUROCLUS Study. *Eur J Cancer Oxf Engl* 1990. 1999; 35: 439–444.
11. Alexander FE, Boyle P, Carli PM, Coebergh JW, Draper GJ, Ekbohm A, et al. Spatial clustering of childhood leukaemia: summary results from the EUROCLUS project. *Br J Cancer*. 1998; 77: 818–824. PMID: [9514064](https://pubmed.ncbi.nlm.nih.gov/9514064/)
12. Demoury C, Goujon-Bellec S, Guyot-Goubin A, Hémon D, Clavel J. Spatial variations of childhood acute leukaemia in France, 1990–2006: global spatial heterogeneity and cluster detection at “living-zone” level. *Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP*. 2012; 21: 367–374. doi: [10.1097/CEJ.0b013e32834e31d8](https://doi.org/10.1097/CEJ.0b013e32834e31d8)
13. McNally RJ, James PW, Picton SV, McKinney PA, Laar van M, Feltbower RG. Space-time clustering of childhood central nervous system tumours in Yorkshire, UK. *BMC Cancer*. 2012; 12: 13. doi: [10.1186/1471-2407-12-13](https://doi.org/10.1186/1471-2407-12-13) PMID: [22244018](https://pubmed.ncbi.nlm.nih.gov/22244018/)
14. Selvin S, Ragland KE, Chien EY-L, Buffler PA. Spatial analysis of childhood leukemia in a case/control study. *Int J Hyg Environ Health*. 2004; 207: 555–562. doi: [10.1078/1438-4639-00327](https://doi.org/10.1078/1438-4639-00327) PMID: [15729836](https://pubmed.ncbi.nlm.nih.gov/15729836/)
15. Alexander F. Clustering of childhood acute leukaemia: The EUROCLUS Project. *Radiat Environ Biophys*. 1998; 37: 71–74. PMID: [9728737](https://pubmed.ncbi.nlm.nih.gov/9728737/)
16. Faure C, Mollié A, Bellec S, Guyot-Goubin A, Clavel J, Hémon D. Geographical variations in the incidence of childhood acute leukaemia in France over the period 1990–2004. *Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP*. 2009; 18: 267–279. doi: [10.1097/CEJ.0b013e32832bf43a](https://doi.org/10.1097/CEJ.0b013e32832bf43a) PMID: [19444126](https://pubmed.ncbi.nlm.nih.gov/19444126/)
17. McNally RJQ, Alexander FE, Vincent TJ, Murphy MFG. Spatial clustering of childhood cancer in Great Britain during the period 1969–1993. *Int J Cancer*. 2009; 124: 932–936. doi: [10.1002/ijc.23965](https://doi.org/10.1002/ijc.23965) PMID: [19035447](https://pubmed.ncbi.nlm.nih.gov/19035447/)
18. Quantum GIS Development Team (2014). Quantum GIS Geographic Information System. Available: <http://qgis.osgeo.org/es/site/>. Accessed 17 Oct 2014.
19. Diggle PJ, Chetwynd AG. Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*. 1991; 47: 1155–1163. PMID: [1742435](https://pubmed.ncbi.nlm.nih.gov/1742435/)

20. Kulldorff M. A spatial scan statistic. *Commun Stat—Theory Methods*. 1997; 26: 1481–1496. doi: [10.1080/03610929708831995](https://doi.org/10.1080/03610929708831995)
21. Dixon PM. Ripley's K function. *Encycl Environmetrics*. 2002; Available: doi: [10.1002/9780470057339.var046.pub2/full](https://doi.org/10.1002/9780470057339.var046.pub2/full)
22. Rowlingson BS, Diggle PJ. Splanx: spatial point pattern analysis code in S-Plus. *Comput Geosci*. 1993; 19: 627–655.
23. The Comprehensive R Archive Network [Internet]. [cited 23 Sep 2014]. Available: <http://cran.r-project.org/>
24. SaTScan—Software for the spatial, temporal, and space-time scan statistics. Available: <http://www.satscan.org/>. Accessed 23 Sep 2014.
25. <http://www.ximindex.com> XO source semantic web C for X-. Agencia Española de Protección de Datos. Available: <https://www.agpd.es/portalwebAGPD/canaldocumentacion/legislacion/estatal/index-ides-idphp.php>. Accessed 9 Apr 2015.
26. Peris-Bonet R, Salmerón D, Martínez-Beneito MA, Galceran J, Marcos-Gragera R, Felipe S, et al. Childhood cancer incidence and survival in Spain. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 2010; 21 Suppl 3: iii103–110. doi: [10.1093/annonc/mdq092](https://doi.org/10.1093/annonc/mdq092) PMID: [20427353](https://pubmed.ncbi.nlm.nih.gov/20427353/)
27. Schottenfeld D, Fraumeni JF, editors. *Cancer Epidemiology and Prevention* [Internet]. Oxford University Press; 2006. Available: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195149616.001.0001/acprof-9780195149616>.
28. McKinney PA. Central nervous system tumours in children: epidemiology and risk factors. *Bioelectromagnetics*. 2005; Suppl 7: : S60–68. doi: [10.1002/bem.20149](https://doi.org/10.1002/bem.20149) PMID: [16142778](https://pubmed.ncbi.nlm.nih.gov/16142778/)
29. López-Abente G, Pollán M, Ardanaz E, Errezola M. Geographical pattern of brain cancer incidence in the Navarre and Basque Country regions of Spain. *Occup Environ Med*. 2003; 60: 504–508. PMID: [12819284](https://pubmed.ncbi.nlm.nih.gov/12819284/)
30. McNally RJQ, Cairns DP, Eden OB, Alexander FE, Taylor GM, Kelsey AM, et al. An infectious aetiology for childhood brain tumours? Evidence from space-time clustering and seasonality analyses. *Br J Cancer*. 2002; 86: 1070–1077. doi: [10.1038/sj.bjc.6600228](https://doi.org/10.1038/sj.bjc.6600228) PMID: [11953851](https://pubmed.ncbi.nlm.nih.gov/11953851/)
31. Pedersen C, Raaschou-Nielsen O, Rod NH, Frei P, Poulsen AH, Johansen C, et al. Distance from residence to power line and risk of childhood leukemia: a population-based case-control study in Denmark. *Cancer Causes Control CCC*. 2014; 25: 171–177. doi: [10.1007/s10552-013-0319-5](https://doi.org/10.1007/s10552-013-0319-5) PMID: [24197706](https://pubmed.ncbi.nlm.nih.gov/24197706/)
32. Badaloni C, Ranucci A, Cesaroni G, Zanini G, Vienneau D, Al-Aidrous F, et al. Air pollution and childhood leukaemia: a nationwide case-control study in Italy. *Occup Environ Med*. 2013; 70: 876–883. doi: [10.1136/oemed-2013-101604](https://doi.org/10.1136/oemed-2013-101604) PMID: [24142970](https://pubmed.ncbi.nlm.nih.gov/24142970/)
33. Rothman KJ, Professor TLLA, Greenland S. *Modern Epidemiology*. Third, Mid-cycle revision edition. Philadelphia: LWW; 2012.
34. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. *Lancet*. 2005; 365: 1429–1433. doi: [10.1016/S0140-6736\(05\)66379-9](https://doi.org/10.1016/S0140-6736(05)66379-9) PMID: [15836892](https://pubmed.ncbi.nlm.nih.gov/15836892/)
35. Waller Lance. *Point Process Models and Methods in Spatial Epidemiology*. Handbook of Spatial Statistics. CRC Press; 2010. pp. 403–423. Available: doi: [10.1201/9781420072884-c22](https://doi.org/10.1201/9781420072884-c22)
36. McNally RJQ, James PW, Ducker S, Norman PD, James OFW. No rise in incidence but geographical heterogeneity in the occurrence of primary biliary cirrhosis in North East England. *Am J Epidemiol*. 2014; 179: 492–498. doi: [10.1093/aje/kwt308](https://doi.org/10.1093/aje/kwt308) PMID: [24401563](https://pubmed.ncbi.nlm.nih.gov/24401563/)
37. Luquero FJ, Banga CN, Remartínez D, Palma PP, Baron E, Grais RF. Cholera epidemic in Guinea-Bissau (2008): the importance of “place.” *PloS One*. 2011; 6: e19005. doi: [10.1371/journal.pone.0019005](https://doi.org/10.1371/journal.pone.0019005) PMID: [21572530](https://pubmed.ncbi.nlm.nih.gov/21572530/)
38. Torabi M, Rosychuk RJ. An examination of five spatial disease clustering methodologies for the identification of childhood cancer clusters in Alberta, Canada. *Spat Spatio-Temporal Epidemiol*. 2011; 2: 321–330. doi: [10.1016/j.sste.2011.10.003](https://doi.org/10.1016/j.sste.2011.10.003) PMID: [22748230](https://pubmed.ncbi.nlm.nih.gov/22748230/)
39. Ramis R, Gómez-Barroso D, López-Abente G. Cluster detection of diseases in heterogeneous populations: an alternative to scan methods. *Geospatial Health*. 2014; 8: 517–526. PMID: [24893029](https://pubmed.ncbi.nlm.nih.gov/24893029/)