RESEARCH ARTICLE

REVISED Clustering of non-leukemia childhood cancer in Colombia: a nationwide study [version 2; peer review: 3 approved, 1 approved with reservations]

Edgar F. Manrique-Hernández [ID][1], Marcela Pilar Rojas Díaz[2], Laura Andrea Rodriguez-Villamizar [ID][1]

[1]Public Health Department, School of Medicine, Universidad Industrial de Santander. Bucaramanga, Santander, Colombia
[2]Instituto Nacional de Salud, Bogotá D.C., Colombia

Abstract

Background: Childhood cancer is considered one the most important causes of death in children and adolescents, despite having a low incidence in this population. Spatial analysis has been previously used for the study of childhood cancer to study the geographical distribution of leukemias. This study aimed to identify the presence of space-time clusters of childhood of cancer excluding leukemia in Colombia between 2014 and 2017.

Methods: All incident cancer cases (excluding leukemia) in children under the age of 15 years that had been confirmed by the National Surveillance System of Childhood Cancer between 2014 and 2017 were included. Kulldorf's circular scan test was used to identify clusters using the municipality of residence as the spatial unit of analysis and the year of diagnosis as the temporal unit of analysis. A sensitivity analysis was conducted with different upper limit parameters for the at-risk population in the clusters.

Results: A total of 2006 cases of non-leukemia childhood cancer were analyzed, distributed in 432 out of 1,122 municipalities with a mean annual incidence rate of 44 cases per million children under the age of 15. Central nervous system (CNS) tumors were the most frequent type. Two space-time clusters were identified in the central and southwest regions of the country. In the analysis for CNS tumors, a spatial cluster was identified in the central region of the country.

Conclusions: The distribution of non-leukemia childhood cancer seems to have a clustered distribution in some Colombian regions that may suggest infectious or environmental factors associated with its incidence although heterogeneity in access to diagnosis cannot be discarded.

Keywords
Cluster Analysis, Neoplasms, Childhood, Epidemiology, Colombia

Open Peer Review

Reviewer Status ✔ ✔ ? ✔

	Invited Reviewers			
	1	**2**	**3**	**4**
version 2 (revision) 25 Jun 2021	✔ report	✔ report		✔ report
	↑	↑		↑
version 1 09 Feb 2021	? report	? report	? report	✘ report

1. **Lisandro Agost** [ID], Centro de Ecología y Recursos Naturales Renovables (CERNAR) – IIByT CONICET-UNC, Córdoba, Argentina

2. **Smita Asthana**, Institute of Cytology and Preventive Oncology, NOIDA, New Delhi, India

3. **Stéphanie Goujon**, UMR 1153 Centre of Research in Epidemiology and StatisticS (CRESS), Villejuif, France

4. **Richard J.Q. McNally** [ID], Newcastle University, Newcastle, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Laura Andrea Rodriguez-Villamizar (laurovi@uis.edu.co)

Author roles: Manrique-Hernández EF: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Rojas Díaz MP**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Rodriguez-Villamizar LA**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Manrique-Hernández EF, Rojas Díaz MP and Rodriguez-Villamizar LA. **Clustering of non-leukemia childhood cancer in Colombia: a nationwide study [version 2; peer review: 3 approved, 1 approved with reservations]** F1000Research 2021, **10**:86 https://doi.org/10.12688/f1000research.27766.2

First published: 09 Feb 2021, **10**:86 https://doi.org/10.12688/f1000research.27766.1

Introduction

Childhood cancer (CC) is considered one the most important causes of death in children and adolescents, despite having a low incidence. The World Health Organization (WHO) estimates that nearly 400,000 new cases of CC are diagnosed every year in children between 0 and 19 years of age[1]. The mean annual worldwide incidence of CC was estimated at 140.6 cases per million children between the age 0–14 years in the period of 2001 to 2010[1]. In the Americas it has been estimated that every year there are approximately 27,000 new cases of cancer in children under the age of 14 years, with an estimated mortality rate of 10,000 deaths/year[2]. The majority of the incident cases in the Americas belong to the Latin American and Caribbean region making up nearly 65% of the diagnosed cases[2].

CC is a set of diseases that does not have a clear etiology yet. There are several conditions that have been identified as risk factors which include genetic (some genetic syndromes and polymorphisms) and non-genetic factors being high-dose radiation, prior chemotherapy, and some viruses the most consistent in literature[3]. Other potential non-genetic risk factors include, some infectious diseases, exposure to pesticides, benzene and radiation, alcohol consumption during pregnancy, smoking, and the socioeconomic condition of the family[4–6]. Some of these factors are more specific than others, as was found with Burkitt´s and Hodgkin´s lymphoma, where the Epstein-Barr virus plays a relevant role. However, there are still controversies surrounding the etiology of these diseases[5].

Spatial analysis allows the identification of geographical patterns of health and disease related events that point out variations between populations contributing to the generation of hypotheses about possible etiologies[7]. Spatial analysis has been previously used for the study of CC, mainly for studying the geographical distribution of leukemias[4,8], This type of analysis allows for the identification of space and time variations in a geographical area and clustering detection[4]. Clusters of

acute childhood leukemia have been identified in Colombia[9], but analyses for CC other than leukemia are scarce[5,10,11]. The objective of this study was to perform an exploratory study with space-time aggrupation to identify clusters of incident cases of non-leukemia CC other than leukemia in Colombian municipalities between 2014 and 2017.

Methods

Population

Colombia is a country located in the north corner of South The Colombian population for 2018 was approximately 48 million people[12]. Women make up 51.2%, and children under 15 years represent 22.5% (12% female and 11.5% male) of the total population. Most of the Colombian population live in urban areas (77.1%)[12].

Cancer and population data sources

All incident cases of non-leukemia CC diagnosed in children under 15 years of age between 2014 and 2017 were included. The data source was the National Surveillance System for Public Health (SIVIGILA, for its name in Spanish), which registers the newly confirmed and probable cases of CC in a systematic and mandatory manner. All cases included in the study were confirmed cases. Surveillance for CC started in Colombia in 2008 with the registry of childhood leukemia cases and starting in 2013 the system registers all types of CC[13]. SIVIGILA verifies the confirmation of reported cases according to the results of diagnostic tests such as myelograms, immunotyping, histopathology or cytogenetic tests; adjusting the real number of confirmed cases and the diagnosis date. De-identified non-leukemia CC data were provided by the National Health Institute (INS for its name in Spanish), allowing access to the following variables: municipality of residence, date of birth, diagnosis date and type of CC according to the International Classification of Childhood Cancer, Third Edition (ICCC-3)[14]. Cases were assigned a consecutive number which cannot be used to identify cases. SIVIGILA is the most complete registry of CC in Colombia, taking into account that it has a nationwide coverage and the reports are updated weekly[9].

Data from CNS and miscellaneous intracranial and intraspinal neoplasms (Group III) cases according to the ICCC-3[14] was extracted for a sub-analysis. CNS tumors include malignant and non-malignant cases. This group is the second with the highest incidence after leukemias[5,15].

Data for the at-risk population in the 1122 municipalities of Colombia was provided by the National Department of Statistics (DANE for its name in Spanish)[10] which performed its last national census in 2018. For the calculation of the population between the years 2014 and 2017 the dynamics of DANE projections of population was used, and an interpolation of the population was conducted for each one of the municipalities for previous years[16]. The childhood population under 15 years varies widely across municipalities with a mid-period population with a mean of 9,880 children, median of 3,336. The minimum childhood populations is located in

La Guadalupe municipality of Guainía (149 children) and maximum in Bogotá, the capital district (1,381,081 children). The calculation of the coordinates (longitude and latitude) of the centroid of each municipality was done in QGIS version 3.16.3 using free cartographic information from the DANE Geoportal[17].

Statistical analysis

We performed a descriptive analysis calculating frequencies with percentages and incidence rates. The incidence of CC was calculated for each municipality and a direct standardization by age and sex of the incidence rates was conducted using as reference the structure of children population for Colombia in 2017. Standardized rates and their respective confidence interval were calculated in STATA® version 14. The global Moran index was calculated to estimate the spatial autocorrelation. The analysis considered neighboring based on the distance between the municipality´s centroids calculated as the Euclidean distance measured between two centroids of municipalities (with no threshold specification). Choroplethic maps were built in order to visualize the standardized rates using the WGS84 projection for Colombia and the cartographic archives available for each municipality in the DANE cartography website[17]. Moran´s index and maps were obtained site using ArcGIS version 10.3 and QGIS version 3.16.3.

Kuldorff's circular scan test was used to identify spatial and spatio-temporal clusters[18], using the SaTScan® software version 9.6. This is a spatial hypothesis test that runs consecutive scans in the study area with different circumference radii that increase in size; the null hypothesis of the test is that the risk of the event (in this case risk of non-leukemia CC) within the circle is the same as outside the scanned area. Space and space-time exploratory analysis were run using a Poisson distribution and scanning for high rates; the space analysis unit was the municipality of residence and the time analysis unit was the year of diagnosis. The selection of the most likely cluster was selected based on the p-value of the log likelihood ratio ($p>0,05$ was considered statistically significant) and 999 replications were used in the simulation to evaluate the significance of the inference. We used an upper limit of the population at risk of 25% and for a sensitivity analysis we assess the results using upper limits of 50% and 10% to identify consistency of clustering results across different upper limits. We used Kulldorff's spatio-temporal scan statistics because it is commonly used to detect spatial and/or temporal disease clusters in epidemiological studies and are appropriated for detecting regularly shaped clusters which we expect to find if clusters are related to localized environmental exposures at municipality level; this method have very good performance to detect large compact clusters of rare diseases in large territories compared to other scan methods[19], and it has a open software to implement the analysis which make it highly reproducible[20]. As part of the sensitive analysis, we also used the Besag and Newell´s (BN) statistic as an additional method for assessing spatial clustering which tests each geographic area separately and combines them to obtain a specific cluster size (k)[21]. We ran the BN statistic with different cluster sizes ($k=10,20,30,50$) using the DCluster package in R software.

Ethical approval

This research received ethical approval from the ethics committee of scientific research at the Universidad Industrial de Santander (CEINCI UIS), on October 27, 2017 (approval number 24-2017).

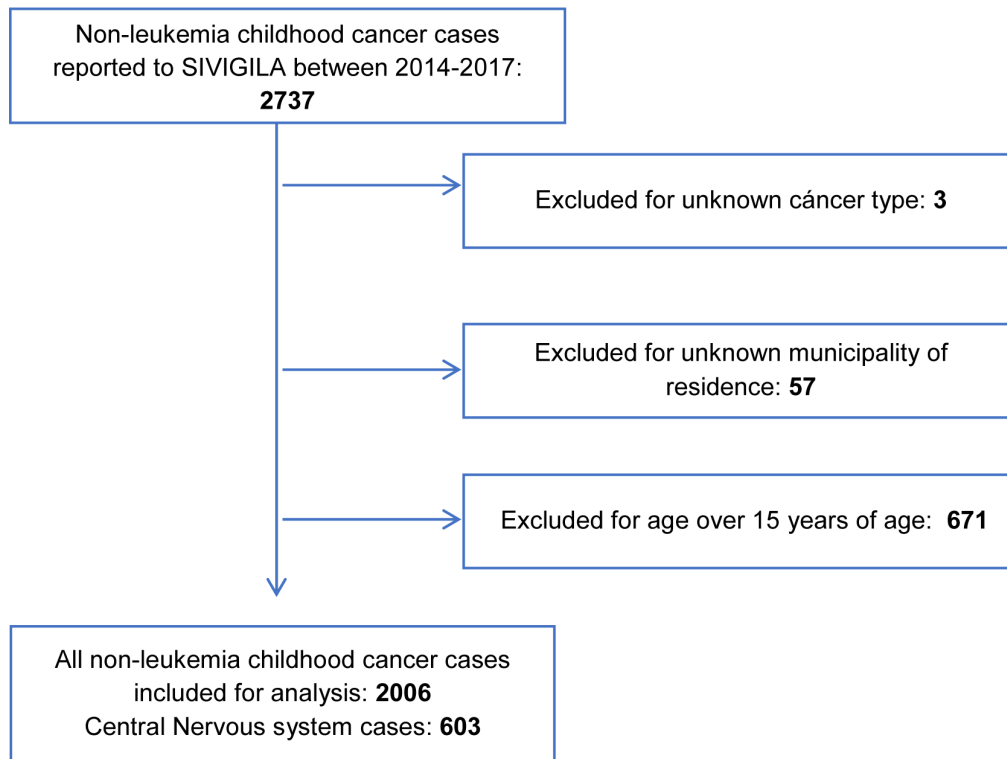# Results
Study population

SIVIGILA reported 2737 cases of non-leukemia CC between January 1st 2014 and December 31st 2017. A total of 731 cases were excluded for different reasons (Figure 1). A total of 57 cases were reported with codification for department with no specification of municipality. The 57 cases belonged to 20 departments distributed across the country (Atlántico, Magdalena, Meta, Cesar, La Guajira, Valle, Tolima, Antioquia, Cundinamarca, Huila, Caquetá, Casanare, Amazonas, Chocó, Putumayo, Cauca, Santander, Bolívar, Norte de Santander y Córdoba). Therefore, a total of 2006 cases were included for the analysis, which were reported in 432 out of the 1122 municipalities of Colombia (38.5%). The analysis of CNS tumors included 603 cases reported in 201 municipalities (17.9%). The distribution of cases by sex, age group, year of diagnosis, and department of residence are presented in Table 1.

A slight majority of reported cases corresponded to males (54.74%) and 39.03% were reported in children under five years of age (0–4 years 33.5%, 5–9 years 36.99%, 10–14 years 29.51%). The mean annual incidence rate of non-leukemia CC was of 44 cases per million children under 15 years of age between 2014 and 2017 in Colombia. The highest incidence rates were reported in Meta (Villavicencio), Bogota D.C., Santander (Bucaramanga, Floridablanca), Bolivar (Cartagena), Valle del Cauca (Cali), Antioquia (Medellin), Cundinamarca (Soacha), Nariño (Pasto). The standardized rates by age and sex varied between 0 and 198 cases per million inhabitants under 15 years of age (Figure 2). The Moran index was of 0.0023 (p=0.211) which indicates a low spatial autocorrelation of the incidence rates across Colombian municipalities.

For CNS tumors, again the slight majority of cases were reported in the male population (55.39%) and the 39.47% of the cases were reported in children between five and nine years of age. The departments with the highest number of cases were Bogota D.C, Valle del Cauca (Cali and Palmira), Antioquia (Medellin), Bolivar (Cartagena), Meta (Villavicencio), Santander (Bucaramanga), Cundinamarca (Soacha) and Nariño (Pasto).

Clustering results

We identified four clusters in the spatial analysis for non-leukemia CC with overlap in the three clusters located in the center of the country which made up a large single cluster.

```
┌─────────────────────────────────────┐
│ Non-leukemia childhood cancer cases  │
│ reported to SIVIGILA between 2014-2017:│
│              2737                     │
└─────────────────────────────────────┘
         │
         │         ┌──────────────────────────────────────────┐
         ├────────▶│ Excluded for unknown cáncer type: 3       │
         │         └──────────────────────────────────────────┘
         │
         │         ┌──────────────────────────────────────────┐
         ├────────▶│ Excluded for unknown municipality of      │
         │         │ residence: 57                             │
         │         └──────────────────────────────────────────┘
         │
         │         ┌──────────────────────────────────────────┐
         ├────────▶│ Excluded for age over 15 years of age: 671│
         │         └──────────────────────────────────────────┘
         ▼
┌─────────────────────────────────────┐
│ All non-leukemia childhood cancer cases│
│ included for analysis: 2006           │
│ Central Nervous system cases: 603     │
└─────────────────────────────────────┘
```

Figure 1. Study population selection flow chart.

Table 1. Characteristics of the study population.

Variable	All non-leukemia childhood cancer (n=2006)			Central nervous system tumors (n=603)		
	No. Cases	%	Cumulative %	No. Cases	%	Cumulative %
Sex						
Female	908	45.26	45.26	269	44.61	44.61
Male	1.098	54.74	100.00	334	55.39	100.00
Age (years)						
0 – 4 years	783	39.03	39.03	207	34.33	34.33
5 –9 years	663	33.05	72.08	238	39.47	73.80
10–14 years	560	27.92	100.00	158	26.20	100.00
Year of diagnosis	No. Cases	%	IR per million	No. Cases	%	IR per million
2014	283	14.11	25.52	78	12.94	7.03
2015	544	27.12	48.54	179	29.68	15.97
2016	590	29.41	52.06	165	27.36	14.56
2017	589	29.36	51.39	181	30.02	15.79

Department	All non-leukemia childhood cancer (n=2006)			Central nervous system tumors (n=603)		
	No. Cases	%	MIR per million	No. Cases	%	MIR per million
Amazonas	2	0.10	69.58	1	0.17	34.79
Antioquia	206	10.27	149.95	62	10.28	45.13
Arauca	11	0.55	160.58	3	0.50	43.80
Atlántico	26	1,35	42.72	5	0.83	8.21
Bogotá. D.C.	423	21.09	299.81	148	24.54	104.90
Bolívar	111	5.53	202.30	26	4.31	47.39
Boyacá	53	2.64	187.07	17	2.82	60.00
Caldas	56	2.79	288.44	14	2.32	72.11
Caquetá	21	1.05	168.63	5	0.83	40.15
Casanare	22	1.10	192.40	6	1.00	52.47
Cauca	57	2.84	153.98	17	2.82	45.92
Cesar	23	1.15	66.85	10	1.66	29.07
Choco	10	0.50	53.96	1	0.17	5.40
Cordoba	23	1.15	47.88	7	1.16	14.57
Cundinamarca	175	8.72	290.14	54	8.96	89.53
Guainía	3	0.15	147.90	2	0.33	98.60
Guajira	14	0.70	48.58	3	0.50	10.41
Guaviare	4	0.20	151.56	0	0	0.00
Huila	45	2.24	146.86	8	1.33	26.11
Magdalena	22	1.10	57.92	6	1.00	15.80
Meta	61	3.04	226.23	27	4.48	100.14
Nariño	56	2.79	142.00	17	2.82	43.11
Norte Santander	37	1.84	105.31	9	1.49	25.61
Putumayo	15	0.75	150.58	5	0.83	50.19
Quindío	23	1.15	236.91	5	0.83	51.50
Risaralda	32	1.60	167.41	10	1.66	52.32
Santander	140	6.98	295.97	38	6.30	80.34
Sucre	23	1.15	95.77	3	0.50	12.49
Tolima	63	3.14	203.36	15	2.49	48.42
Valle del Cauca	243	12.11	227.73	79	13.10	74.04
Vaupés	3	0.15	157.96	0	0	0.00
Vichada	2	0.10	48.09	0	0	0.00

IR: Incidence Rate; MIR: Mean annual Incidence Rate 2014–2017
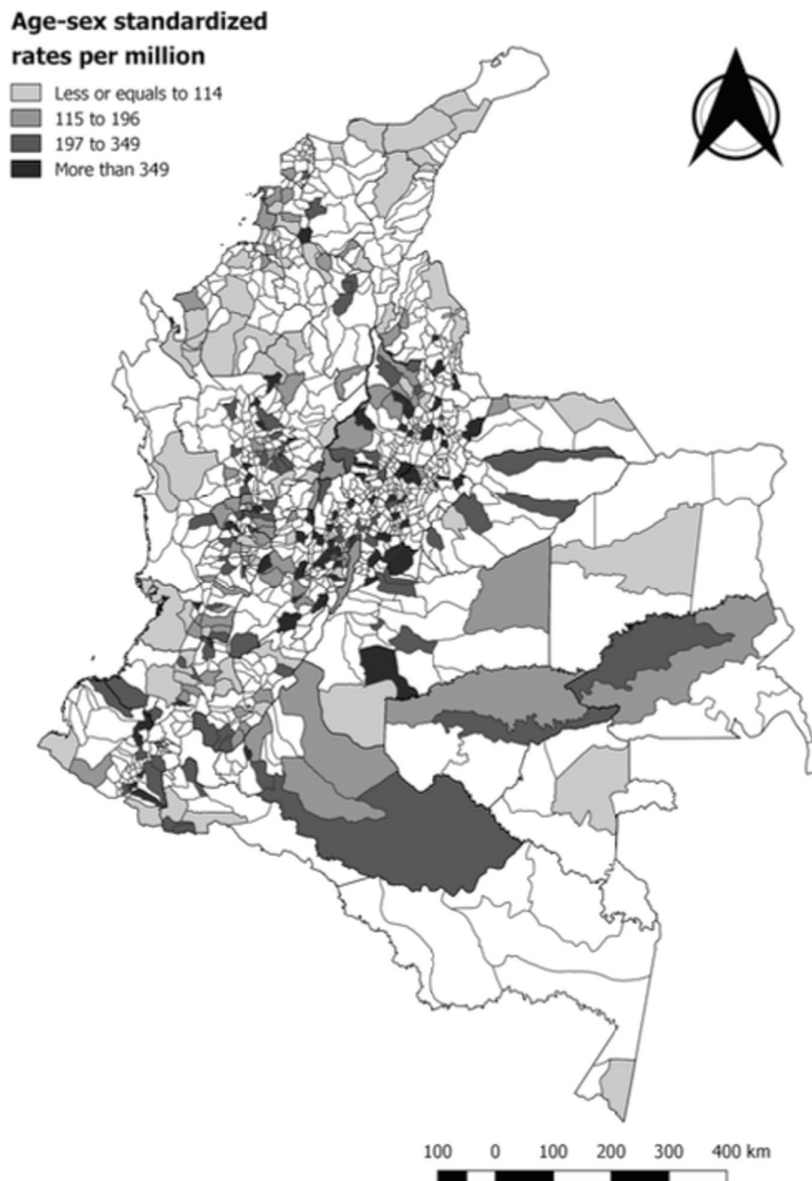
Figure 2. Standardized rates of non-leukemia childhood cancer by municipality, Colombia 2014–2017.

The first cluster in the central region of the country included 327 municipalities distributed in the following departments: Cundinamarca (95), Meta (6), Boyaca (122), Santander (76), Antioquia (7), Caldas (4), Casanare (13), Tolima (3) and Bogota D.C

This cluster has a radius of 172.11 Km and 798 cases, with an expected number of cases of 497.88 with a relative risk (RR)=2 and p value <0.0001. The second cluster was identified in the departments of Cundinamarca (57), Meta (7) and Bogota D.C. with a radius of 72.04 km, and superposition with the first cluster; in this cluster 623 cases were identified for

a total of 358.29 expected cases with RR = 2.07 and p value <0.0001. The third cluster was identified in the southwest region of Colombia, corresponding to the city of Cali (Valle del Cauca) and without superposition with any other cluster. In this third cluster 152 cases were identified where 87.95 cases were expected, with a RR = 1.79 and a p value <0.0001. The fourth cluster was identified in 23 municipalities of the Santander (Northeast region) department, with a 58.50 km radius, super positioning with some municipalities from the first cluster; in this cluster there were 106 cases with 55.15 as the expected number of cases, obtaining a RR = 1.97 and a p value <0.0001. When conducting the clustering analysis with

no overlap we confirm the presence of two spatial clusters being the first and larger located at the center of the country and the second in the southwest region.

The same two clusters were identified in the space-time analysis for non-leukemia CC. The first cluster was located in the central region of the country corresponding to the following departments: Boyaca (122), Santander (76), Cundinamarca (95), Casanare (13), Meta (6), Caldas (4), Antioquia (7), Tolima (3), Bogota D.C. This cluster was identified between 2015 and 2016 with a radius of 172.11 Km, 491 cases reported with 249 expected obtaining a RR = 2.29 and p value <0.0001. The second cluster was identified in the city of Cali between 2016 and 2017, with 97 cases reported and 43.7 expected obtaining a RR = 2.28 and p value <0.0001.

The spatial analysis for CNS tumors identified one cluster located in the center of the country in the following departments: Meta (27), Cundinamarca (86), Casanare (8), Huila (1),

Tolima (8) and Bogota D.C. This cluster has a radius of 177.43 km, with 237 cases reported and 122 expected for a RR = 2.55 and a p value <0.0001. The space-time analysis for CNS identified the same cluster in the central region corresponding to the following municipalities: Meta (16), Cundinamarca (95), Boyaca (30), Casanare (3), Tolima (1) and Bogota D.C. This cluster was identified between 2015 and 2016 with a radius of 112.53 km, with 143 cases reported and 58.9 expected obtaining a RR = 2.87 and p values <0.0001

In the sensitivity analysis for non-leukemia CC circular scan tests were run using values of the at-risk population of 10% and 50%. There were 304 identified municipalities in the central region of the country that showed consistency in the three analysis (using 10%, 25% and 50% upper limit of at-risk population) and represent the two clusters identified in the center and southwest regions of Colombia (Figure 3). The sensitivity analysis of spatial clustering results using the

Figure 3. Municipalities consistently identified within spatial clusters of non-leukemia childhood cancer in Colombia, 2014–2017.

Besag-Newell statistic with k=30 also identified three clusters of municipalities located predominantly in the center of the country and one at the southwest of the country. The test was statistically significant for 18, 61 and 98 municipalities for cluster sizes of 10, 20, and 30 cases, respectively.

Discussion

This study identified the presence of non-leukemia CC clusters between 2014 and 2017 in Colombia, using information with nationwide coverage available in SIVIGILA. To our knowledge, this is the first nation-wide study in South America using spatial analysis to describe the distribution and clustering of non-leukemia CC.

Spatial and spatio-temporal analysis have been previously used in CC, mainly in the study of the geographic pattern of leukemias[8,22]. A recent systematic review of space-time analysis identified 70 studies published up to 2016 of which 47 reported results for leukemias, 26 for lymphomas, 13 for CNS tumors and 12 for other types of tumors[23]. All 32 analyses used for the meta-analysis were from Europe and United States; and the analysis showed evidence of leukemia clustering in children between 0 and 5 years of age. However, the evidence was not conclusive for lymphomas and CNS tumors. This study, however focuses on space-time clustering not on cluster detection which is aimed to detect localized excesses of CC cases.

Studies of clustering for non-leukemia CC have been conducted in different continents showing some heterogeneity in their results. In Europe, Ortega *et al.*[24] used elliptic analysis to identify clusters of CC in children under 15 years of age in Murcia, Spain between 1998 and 2009. This analysis identified a space-time cluster of lymphomas between 2011 and 2013. Also in Spain, a spatial case-control analysis conducted between 1985 and 2015 including data from five autonomous regions explored the clustering of non-leukemia CC by site of residence and date of diagnosis. The authors found spatial clusters for all CC combined and for lymphomas at date of diagnosis, and for CNS embryonal tumors clustering at birth and diagnosis. The results, however, did not reach statistical significance for evidence of clustering when adjusted for multiple testing[5]. In France, a study of clustering of CC between 2000 and 2014 used different spatial scan methods and different geographical scales and found spatial heterogeneity and two large clusters for SNC tumors (glioma) and non-Hodgkin lymphoma[11].

In the Asian continent, a study in Palestine performed an analysis of CC clusters between 1998 and 2007 using the circular scan method; a greater clustering effect was found in metropolitan districts and one cluster of lymphomas was identified in an agricultural city between 1998 and 2002[25]. In Canada, Torabi and Rosychuk explored the presence of clusters of CC between 1983 and 2004 in the province of Alberta, Canada, using five different methods to analyze clustering, including circular scan tests. The study showed

evidence of clustering to the south of the province but did not showed results by type of cancer[26]. Then, a specific analysis of leukemia and lymphoma did not find specific spatiotemporal clustering[27]. In Florida, United States, a study assessed the clustering of CC (0–19 years) between 2000–2010 using spatial scan methods applied to zip code areas and found evidence of clustering for CNS, leukemia and lymphoma[10]. In South America, in the province of Cordoba, Argentina, Agost reported one of the first studies in the region using the circular scan test to detect clusters of CC. Spatial clusters were found for leukemias, lymphoid neoplasms, CNS tumors and in the space-time analysis clusters of neuroblastoma and other peripheral tumors were also identified[28].

Overall, most European studies tend to report lack of evidence for CC clusters, whereas other continents (such as in this study) tend to show some clustering evidence. The heterogeneous nature of the findings could be related to different factors, primarily environmental conditions and the methods used. In classic epidemiology, the consistency of the results of association between exposure and events is core when assessing causality[29]. Nonetheless, in the spatial analysis the focus is on the description of the patterns and not the causality; this is why the heterogeneity of the results is important in these exploratory studies, since it can reflect conditions or exposures that may vary between and within populations.

The results of the studies can also differ due to the diversity of methods used. The spatial studies based on the analysis of areas (ecological approach) such as this study, and the studies in Argentina, Canada, France and in Palestine[11,19,25,26,28], seem to identify more often clusters compared to the results of the studies based on point analysis (case-control studies) conducted in Europe[5,15]. However, differences might be explained for difference in distribution of CC cases in countries. We used an ecological approach for this first exploratory study because of the quality of information available in the country at municipality level and the absence of official data sources for selecting comparable controls. Kulldorf's circular scan tests was chosen because it is optimal to detect clusters in a regular way, it has excellent performance detecting rare diseases in large populations such as CC[19], and for its easy use through specific software that makes it standardized and reproducible.

Non-leukemia CC clusters identified in Colombia are located mainly in the central region of the country. One cluster for childhood leukemia was also identified in the center of the country in a previous study[9]. Clusters for both leukemia and non-leukemia cases might be related to each other, however the non-leukemia cluster is larger (327 municipalities compared with 109 identified in the leukemia cluster), more expanded to the North of the leukemia cluster, and with higher incidence rates located in municipalities with predominant rural areas. The cluster location corresponds to a large area in the mountain ranges that blend with large zones of agriculture and mining operations. These combined zones can generate special

environments that allow the interaction of infectious agents, environmental, and occupational conditions that may have a space and time effect in the incidence of events such as CC. There is evidence that exposure to arsenic[30] and pesticides[31–33] is related to a greater risk of developing CC, especially leukemias, lymphomas and CNS tumors. The large area covered by the central cluster and its high relative risk is of concern and suggest the presence of an infectious or environmental factor strongly associated to the risk of CC, mainly non-leukemia, that is highly prevalent in this area compared to the remaining areas of the country. Further studies using ecological and individual approaches should be conducted addressing the relationship of CC cases with specific infectious, environmental, and occupational exposures in Colombia.

The spatial heterogeneity in this type of ecological spatial analysis can be also observed due to diagnosis or reporting heterogeneity. For this study we selected as source of cancer cases the report to the national surveillance system for childhood cancer (NSSCC) from national public health surveillance system (SIVIGILA) because this is the strongest and more complete health information system that is operating in all 1,122 municipalities in Colombia. Unfortunately, the cancer population-based registries in Colombia are limited to four regions in the country which are representative of specific urban areas but do not represent the full spectrum of municipalities and regions in Colombia[34]. The SIVIGILA is operated by the National Institute of Health (INS for Spanish) as a mandatory, systematic, and continuous registry with standardized protocols. The system operates permanently in all municipalities based on immediate report for selected health events and weekly report for all events, including childhood cancer. The CC surveillance began in 2008 when acute leukemia was included as a mandatory health notification event. In 2013 the system was extended to all types of childhood cancer. The system preserved the core formats and software for reporting acute childhood leukemia and therefore the extension to other cancer types had a shorter learning curve for the surveillance system´s personnel in municipalities. During the study period, notification of non-leukemia cancer were reported for 432 municipalities in almost all departments and districts (including municipalities with predominantly rural remote areas), which support the wide coverage of the surveillance system.

For a previous study, we conducted a comparison between national high-cost account registry and SIVIGILA report during 2016. We identified 1394 incident cases of CC and 1206 (86.5%) of them were reported to SIVIGILA, indicating that the systems captured 83% of all incident cases of CC in Colombia, which included non-leukemia cases[9]. The 188 cases missing in SIVIGILA corresponded to different cancer diagnosis and municipalities distributed in 28 departments across the country. Therefore, we assumed that the presence of underreporting it is not concentrated in specific areas of the country and underreporting although is present, might not be the main explanation for the spatial heterogeneity in our results. However, health care access to cancer diagnosis is limited to specific regions in the country located mainly in the main capital cities and therefore delay in diagnosis (and derived delayed in reporting) might be present in remote semirural and rural municipalities[35]. This study analyzed CC data for 2014–2017 and databases were consolidated in 2019, therefore cases with delayed diagnosis had the opportunity to be included in the last two years as SIVIGILA required the reporting of incident and prevalent cases since 2014 for non-leukemia cases. However, the cases with missing diagnosis due to limitations in access to health care might be still present in the study but cannot be quantified.

One limitation of this study is that the cluster frontier for the main central cluster was difficult to delineate in the overlapping analysis. This is a known limitation of the cluster detection methods and for the Kulldorf´s scan test used it is added to its limitation to detect clusters with irregular shapes. We conducted a sensitivity analysis with different population at-risk proportions and to delineate better the cluster and Besag-Newell statistic as alternative spatial clustering method with similar results, however further studies might include alternative clustering methods . Other important limitation of our study is that we assessed clusters based on place of residence at the time of diagnosis but we were not able to compare with clusters based on place of residence at birth or during gestation as this information was not available in SIVIGILA. Additionally, the limited number of reported cases for group IV and subsequent groups of the ICCC-3 did not allow for the analysis of other groups different to group III (CNS).

Conclusion

The spatial distribution of non-leukemia CC seem to have clustered patterns in some regions of the country that suggest possible infectious, environmental or occupational factors related to its incidence. Future studies should assess the effect of these factors related to non-leukemia CC.

Data availability

Source data

We declare that we have permission for the free use of this data.

Zenodo: Clustering of childhood cancer in Colombia: a nationwide study. http://doi.org/10.5281/zenodo.4781513[36].

This project contains the following source data:

- Database SIVIGILA by municipality (database of cases by type of cancer and municipality (geographic location) taken from SIVIGILA data)

- Database DANE population (database of populations by municipality (geographic location) taken from DANE data)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

1. Steliarova-Foucher E, Colombet M, Ries LAG, *et al.*: **International incidence of childhood cancer, 2001-10: a population-based registry study.** *Lancet Oncol.* 2017; **18**(6): 719–31.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Bravo CP, Cuadrado C, Gónzales I: **Cáncer Infantil en Latinoamérica: Un Análisis Comparativo de la Respuesta de los Sistemas de Salud.** *Medwave.* 2019; **19**(Suppl 1): SP95.
   Publisher Full Text

3. Spector LG, Pankratz N, Marcotte EL: **Genetic and nongenetic risk factors for childhood cancer.** *Pediatr Clin North Am.* 2015; **62**(1): 11–25.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Wheeler DC: **A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996-2003.** *Int J Health Geogr.* 2007; **6**: 13.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Konstantinoudis G, Kreis C, Ammann RA, *et al.*: **Spatial clustering of childhood cancers in Switzerland: a nationwide study.** *Cancer Causes Control.* 2018; **29**(3): 353–62.
   PubMed Abstract | **Publisher Full Text**

6. Schüz J, Erdmann F: **Environmental Exposure and Risk of Childhood Leukemia: An Overview.** *Arch Med Res.* 2016; **47**(8): 607–614.
   PubMed Abstract | **Publisher Full Text**

7. Valbuena-García AM, Rodríguez-Villamizar LA: **Análisis espacial en epidemiología: revisión de métodos.** *Rev Uni Ind Santander Salud.* 2018; **50**(4): 358–65.
   Publisher Full Text

8. Knox EG, Gilman EA: **Spatial clustering of childhood cancers in Great Britain.** *J Epidemiol Community Health.* 1996; **50**(3): 313–9.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Rodriguez-Villamizar LA, Díaz MPR, Merchán LAA, *et al.*: **Space-time clustering of childhood leukemia in Colombia: A nationwide study.** *BMC Cancer.* 2020; **20**(1): 48.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Amin R, Hendryx M, Shull M, *et al.*: **A Cluster Analysis of Pediatric Cancer Incidence Rates in Florida: 2000–2010.** *Statistics and Public Policy.* 2014; **1**(1): 69–77.
    Publisher Full Text

11. Goujon S, Kyrimi E, Faure L: **Spatial and temporal variations of childhood cancers: Literature review and contribution of the French national registry.** *Cancer Med.* 2018; **7**(10): 5299–5314.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Departamento Administrativo Nacional de Estadistica (DANE): **Censo Nacional de Población y Vivienda**. (accessed on 22 November 2019).
    Reference Source

13. Ministerio de Salud y Protección Social de Colombia: **Protocolo de vigilancia en salud pública- cáncer en menores de 18 años**. 1–21. [cited 2020 Jun 11].
    Reference Source

14. Steliarova-Foucher E, Stiller C, Lacour B, *et al.*: **International classification of childhood cancer, third edition.** *Cancer.* 2005; **103**(7): 1457–67.
    PubMed Abstract | **Publisher Full Text**

15. Ramis R, Gómez-Barroso D, Tamayo I, *et al.*: **Spatial analysis of childhood cancer: A case/control study.** *PLoS One.* 2015; **10**(5): e0127273.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Departamento Administrativo Nacional de Estadistica (DANE): **Proyecciones de Poblacion Colombia**. (accessed on 24 January 2020).
    Reference Source

17. Departamento Administrativo Nacional de Estadistica (DANE): **Geoportal DANE**. (accessed on 24 January 2020).

18. Kulldorff M: **A spatial scan statistic.** *Commun Stat - Theory Methods.* 1997; **26**(6): 1481–96.
    Publisher Full Text

19. Goujon-Bellec S, Demoury C, Guyot-Goubin A, *et al.*: **Detection of clusters of a rare disease over a large territory: Performance of cluster detection methods.** *Int J Health Geogr.* 2011; **10**: 53.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Mathes RW, Lall R, Levin-Rector A, *et al.*: **Evaluating and implementing temporal, spatial, and spatio-temporal methods for outbreak detection in a local syndromic surveillance system.** *PLoS One.* 2017; **12**(9): e0184419.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Besag J, Newell J: **The detection of clusters in rare diseases.** *Journal of the Royal Statistical Society A.* 1991; **154**(1): 143–155.
    Publisher Full Text

22. Kreis C, Grotzer M, Hengartner H, *et al.*: **Space-time clustering of childhood cancers in Switzerland: A nationwide study.** *Int J Cancer.* 2016; **138**(9): 2127–35.
    PubMed Abstract | **Publisher Full Text**

23. Kreis C, Doessegger E, Lupatsch JE, *et al.*: **Space-time clustering of childhood cancers: a systematic review and pooled analysis.** *Eur J Epidemiol.* 2019; **34**(1): 9–21.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Ortega-García JA, López-Hernández FA, Cárceles-Álvarez A: **Analysis of small areas of pediatric cancer in the municipality of Murcia.** *An Pediatr (Barc).* 2016; **84**(3): 154–62.
    PubMed Abstract | **Publisher Full Text**

25. Bailony MR, Hararah MK, Salhab AR, *et al.*: **Cancer registration and healthcare access in West Bank, Palestine: A GIS analysis of childhood cancer, 1998-2007.** *Int J Cancer.* 2011; **129**(5): 1180–9.
    PubMed Abstract | **Publisher Full Text**

26. Torabi M, Rosychuk RJ: **An examination of five spatial disease clustering methodologies for the identification of childhood cancer clusters in Alberta, Canada.** *Spat Spatiotemporal Epidemiol.* 2011; **2**(4): 321–30.
    PubMed Abstract | **Publisher Full Text**

27. Kulkarni K, Stobart K, Witol A, *et al.*: **Leukemia and lymphoma incidence in children in Alberta, Canada: a population-based 22-year retrospective study.** *Pediatr Hematol Oncol.* 2011; **28**(8): 649–60.
    PubMed Abstract | **Publisher Full Text**

28. Agost L: **Analysis of spatial-temporal clusters of childhood cancer incidence in the province of Córdoba, Argentina (2004-2013).** *Arch Argent Pediatr.* 2016; **114**(6): 534–543.
    PubMed Abstract | **Publisher Full Text**

29. Hill AB: **The Environment and Disease: Association or causation?** *Proc R Soc Med.* 1965; **58**(5): 295–300.
    PubMed Abstract | **Free Full Text**

30. Engel A, Lamm SH: **Arsenic exposure and childhood cancer--A systematic review of the literature.** *J Environ Health.* 2008; **71**(3): 12–6.
    PubMed Abstract

31. Chen M, Chang CH, Tao L, *et al.*: **Residential exposure to pesticide during childhood and childhood cancers: A meta-analysis.** *Pediatrics.* 2015; **136**(4): 719–29.
    PubMed Abstract | **Publisher Full Text**

32. van Maele-Fabry G, Lantin AC, Hoet P, *et al.*: **Residential exposure to pesticides and childhood leukaemia: A systematic review and meta-analysis.** *Environ Int.* 2011; **37**(1): 280–91.
    PubMed Abstract | **Publisher Full Text**

33. Carles C, Bouvier G, Esquirol Y, *et al.*: **Residential proximity to agricultural land and risk of brain tumor in the general population.** *Environ Res.* 2017; **159**: 321–30.
    PubMed Abstract | **Publisher Full Text**

34. Cendales R, Pardo C, Uribe C, *et al.*: **Calidad de los datos en los registros de cáncer de base de población en Colombia.** *Biomedica.* 2012; **32**(4). [cited 2020 Jun 11].
    Publisher Full Text

35. Cuenta de Alto Costo Fondo Colombiano de Enfermedades de Alto Costo: **Situación del cáncer en la población pediátrica atendida en el SGSSS de Colombia.** Bogotá: Cuenta de Alto Costo Fondo Colombiano de Enfermedades de Alto Costo, 2018; 2019.

36. Manrique Hernández EF, Rojas Diaz MP, Rodriguez-Villamizar LA: **Clustering of childhood cancer in Colombia: a nationwide study**. 2021.
    http://www.doi.org/10.5281/zenodo.4488080

Open Peer Review

Current Peer Review Status: ✓ ✓ ? ✓

- -

Version 2

Reviewer Report 22 July 2021

https://doi.org/10.5256/f1000research.57186.r88297

✓   **Lisandro Agost** iD

Centro de Ecología y Recursos Naturales Renovables (CERNAR) – IIByT CONICET-UNC, Córdoba, Argentina

The authors have incorporated and responded to all suggestions satisfactorily. I believe the article is now ready for indexing.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Ecology - Epidemiology - Childhood Cancer - Pesticide Risk Indicators

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 09 July 2021

https://doi.org/10.5256/f1000research.57186.r88298

✓   **Smita Asthana**

Division of Epidemiology and Biostatistics, Institute of Cytology and Preventive Oncology, NOIDA, New Delhi, Delhi, India

All the suggestions are incorporated by the authors. Changes made are satisfactory and I recommend the article for indexing.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 06 July 2021

https://doi.org/10.5256/f1000research.57186.r88296

✔ **Richard J.Q. McNally** iD

Institute of Health and Society, Newcastle University, Newcastle, UK

I am happy with the revisions that have been made.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

- -

Version 1

Reviewer Report 07 May 2021

https://doi.org/10.5256/f1000research.30704.r81746

✘ **Richard J.Q. McNally** iD

Institute of Health and Society, Newcastle University, Newcastle, UK

This is an interesting paper reporting clustering of childhood cancer in Colombia. However, there are a number of specific issues and these are listed below:

1. There needs to be more justification for choice of methods used.

2. Is the interest only in the identification of specific spatial or space-time clusters, rather than more generalised spatial or space-time clustering?

3. The presence of specific clusters is more consistent with some localised environmental

sources of exposure, rather than general exposures (such as infections). Could these clusters be linked with data on more specific localised exposures?

4. I suggest that the authors also consider other methods for looking at generalised spatial clustering of space-time clustering. These include methods of Cuzick, Besag, Knox, Jacquez and Diggle.

5. The English language needs improving throughout.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 24 May 2021
Edgar F. Manrique-Hernández, Universidad Industrial de Santander. Bucaramanga., Bucaramanga, Colombia

Reviewer #4
Richard J.Q. McNally, Institute of Health and Society, Newcastle University, Newcastle, UK

This is an interesting paper reporting clustering of childhood cancer in Colombia.

Author's response to reviewer:
Thank you for your comment.

However, there are a number of specific issues and these are listed below:

1. There needs to be more justification for choice of methods used.

Author's response to reviewer:
Thank you for your comment. We have added a comment on methods section for the selection of the circular scan test.
We used Kulldorff's spatio-temporal scan statistics because it is commonly used to detect spatial and/or temporal disease clusters in epidemiological studies and are appropriated for detecting regularly shaped clusters which we expect to find if clusters are related to localized environmental exposures at municipality level; this method have very good performance to detect large compact clusters of rare diseases in large territories compared to other scan methods, and it has a open software to implement the analysis which make it highly reproducible.

1. Is the interest only in the identification of specific spatial or space-time clusters, rather than more generalised spatial or space-time clustering?

Author's response to reviewer: Thank you for your comment. We were interested in overall clustering but specifically in the identification of localized spatial or space-time clusters. That is the reason we used a scan clustering method.

1. The presence of specific clusters is more consistent with some localised environmental sources of exposure, rather than general exposures (such as infections). Could these clusters be linked with data on more specific localised exposures?

Author's response to reviewer: Thank you for your comment. We conducted this exploratory study to assess the presence of localized clusters and as commented in the discussion sections, the localized central cluster open the door to further studies assessing specific localized exposures mainly related to environmental and occupational exposures related to pesticide's uses and mining operations in the localized area identified.

1. I suggest that the authors also consider other methods for looking at generalised spatial clustering of space-time clustering. These include methods of Cuzick, Besag, Knox, Jacquez and Diggle.

Author's response to reviewer: Thank you for your suggestion. We have added the Besag and Newell´s statistic as an additional method for assessing spatial clustering and we found similar results at different cluster sizes.

1. The English language needs improving throughout.

Author's response to reviewer: Thank you for your comment. We have reviewed and improved the language style.

Competing Interests: No competing interests were disclosed.

Reviewer Report 21 April 2021

? **Stéphanie Goujon**

Childhood and adolescent cancer team (EPICEA), UMR 1153 Centre of Research in Epidemiology and StatisticS (CRESS), Villejuif, France

The nationwide study conducted by Manrique-Hernández and colleagues described the space-time distribution of non-leukemia childhood cancers in Colombia. Based on data from the national health surveillance system, the study aimed at detecting spatial and spatio-temporal localised excesses of cases over the period 2014-2017, on the municipality scale. Describing spatial and temporal variations is of a great importance for childhood cancer surveillance. The authors used registry based data and the scan method developed by M. Kulldorff, which is appropriate for such a study. In a purely spatial analysis, they detected four widespread overlapping clusters of non-leukemia cases in two different areas, in which the number of observed cases were about twice the numbers expected under the hypothesis of homogeneous incidence rates over the whole study period. Space-time analyses identified clusters in the same areas, with excesses observed during shorter time periods.

Major comments:
- ○ There is a major issue concerning the interpretation of the results.

The authors concluded that the detected clusters "may suggest infectious or environmental factors associated with its incidence" (abstract and main text).
The presence of localized clusters might actually suggest that a risk factor is present with a higher prevalence in these places than elsewhere in Colombia, but spatial heterogeneity might also be due to differences in case registration. I consider this point is of a primary importance and should be discussed further in the paper and in the abstract (it was discussed rapidly at the end of the discussion section).

To discuss that point, it would be useful to describe (at least briefly) in the paper the surveillance system in Colombia and to provide information on how the cases are identified nationwide. Are the data provided automatically to the NHS by hospital centers? Or do SIVIGILA members visit the hospital centers to collect data actively? Which hospital centers are visited or contacted? What are the main care pathways in Colombia? The authors cited several interesting papers that were written in Spanish and therefore not easily understandable.

Spatial differences in case registration might be observed because of under-diagnoses or difficulties for the registry to identify cases in some regions. In the paper by Rodriguez-Villamizaron *et al.* on childhood leukemia (2020), the authors indicated that childhood cancer became a priority in Colombia since 2010. Do the authors consider that the new regulation was adopted homogeneously in Colombia since that date, or is it possible that access to diagnosis was different over the study period 2014-2017 depending on the place of residence? Besides, were all medical reports available nationwide during the study period?

In Rodriguez-Villamizaron *et al.* 2020, under-registration was also discussed and it was estimated that 17% of AL cases were not captured by the registry in 2016 ("During this year [2016] 1394 incident cases of childhood cancer were identified and 1206 (86.5%) of them were reported to the NSSCC [...] the NSSCC captured and registered 83% of all incident cases of childhood cancer in Colombia during 2016"). If the missing cases were not homogeneously distributed in the municipalities, this could lead to clusters in areas where registration was more exhaustive. Do the authors have similar information for non-leukemia cases?

A large cluster of childhood leukemia was detected in the center of Colombia. Can it be related to the large cluster of non-leukemia cases reported in the present study?
- The four spatial clusters detected in the study were located in two distinct regions, as three clusters overlapped. The space-time clusters corresponded to the same spatial areas. In all, as a conclusion, I would say that only two clusters were detected (maybe over 2 years only) and I suggest to discuss the fact that the cluster frontiers were difficult to delineate (a well-known limit of the cluster detection methods). The size of the largest detected cluster, in the central region, and the magnitude o the relative risk should also be discussed. A relative risk of about 2 in such a large area is quite surprising and unexpected. If an infectious or environmental factor was responsible for such an excess, it would have to be strongly associated to the risk of non-leukemia cancer and highly prevalent in the cluster area (in comparison to the remaining part of Colombia).
- The statistical methods used in this study, Moran's test for spatial autocorrelation and the Kulldorff's scan method for cluster detection, were appropriate. However, some details on the methodology and the parameters used could be added. In particular, the authors could add a reference for the Moran's test for spatial autocorrelation, and explain how the neighborhood was defined in the study. Were two municipalities considered neighboring areas if the distance between their centroids was below a given threshold or if they shared a common border?

It would also be useful to provide further details on the scan method (estimation of a likelihood ratio for each window and selection of the most likely cluster, i.e. the window associated to the maximum ratio), and the simulations that were conducted to evaluate the significance thresholds (how many simulations were done?).

- Several other studies should be referred to in the introduction and discussion sections to provide more accurate information on childhood cancer etiology and the literature on childhood cancer clusters.
(Suggested publications):

Etiology:
- 'Genetic and nongenetic risk factors for childhood cancer' by Spector *et al.* (2015[1]).
- 'Environmental Exposure and Risk of Childhood Leukemia: An Overview' by Schüz *et al.* (2016[2]).

Clustering:
- 'Childhood cancer trends in a western Canadian province: a population-based 22-year retrospective study' by Rosychuk *et al.* (2010[3]).
- 'Spatial clustering and space-time clusters of leukemia among children in Germany, 1987-2007' by Schmiedel *et al.* (2007[4]).

- ○ 'Leukemia and lymphoma incidence in children in Alberta, Canada: a population-based 22-year retrospective study by Kulkarni *et al.* (2011[5]).
- ○ 'A cluster analysis of Pediatric Cancer Incidence Rates in Florida: 2000-2010' by Amin *et al.* (2014[6]).
- ○ 'Spatial and temporal variations of childhood cancers: Literature review and contribution of the French national registry' by Goujon *et al.* (2018[7]).

Minor comments:

Abstract:
- ○ "A sensitivity analysis was conducted with different upper limit parameters for the at-risk population." The upper limit was for the at-risk population <u>included in the cluster.</u>

- ○ There are 1122 municipalities nationwide. Several municipalities had no observed cases, so that the cases were actually distributed in 432 different municipalities. However, all the municipalities were included in the analyses (if the at-risk population was not null). I suggest therefore to report that "2006 cases were distributed in 1122 municipalities".

- ○ Regarding CNS tumors, it would be informative to specify whether non malignant cases were included.

- ○ Conclusion: differences in case registration should also be considered as a possible explanation.

Introduction:
- ○ Based on reference 1, "The mean annual incidence of CC was estimated at 140.6 cases per million children". This incidence rate was estimated worldwide, which could be specified.

- ○ "There are several conditions that have been identified as risk factors". The factors cited in this sentence were associated with childhood cancer with different degree of evidence. Some factors are considered as known (high dose ionising radiation, chemotherapy, certain genetic syndromes and some genetic polymorphisms, some viruses in lymphomas) or highly suspected (domestic and occupational parental exposure, socioeconomic conditions, infections and immune system stimulation for leukemia, birth weight, benzene exposure, air pollution) risk factors, while for other factors the literature is more heterogeneous and no firm conclusion can be drawn to date (tobacco and alcohol consumption). It is important to consider that point when discussing the etiology.

- ○ The following sentence is quite long and difficult to understand: "contributing to the generation of hypotheses about possible etiologies. Spatial analysis has been previously used for the study of CC, mainly for studying the geographical distribution of leukemias, since this type of analysis allows for the identification of space and time variations in a geographical area that generate clusters that indicate an increase in the tendency of the cases".

- ○ At the end of the introduction, several studies on childhood cancer cluster detection are cited, but only one reference (ref 5) is provided for non-leukemia cancers. Maybe some of the references I reported above could be added here.

Methods:

○ CC cases were identified by the National Surveillance System for Public Health, which registers "the newly confirmed and probable cases of CC". It is unclear to me what "probable cases" means? It seems that CC diagnoses are confirmed on the basis of diagnostic exams and coded according to the ICCC-3, so could the authors explain what are the probable cases that are registered (if not confirmed why are they registered?).

○ Are non-malignant CNS tumors registered in SIVIGILIA, and included in the study? Based on the number of CNS tumor cases reported in the result section (17.9% of non-leukemia cases), I assume that only malignant CNS tumors were included in the study. This information should be added in the main text and the abstract.

Statistical analysis:

○ "We performed a descriptive analysis calculating frequencies and central tendency measurements." I don't understand what "central tendency measurements" refers to?

○ I understand that standardized rates were considered to account for potential differences in the age distribution of the pediatric population between municipalities. Were those potential differences also accounted for in the SaTScan analyses?

Results:

○ 731 cases were excluded from the analyses, of which 57 had an unknown municipality of residence. It would be interesting to describe those cases in terms of type of diagnosis, year of diagnosis. Could the authors get other geographic information related to the area of residence? Were those cases grouped in a particular region?

○ It would be useful to describe the distribution of the pediatric population in the 1122 municipalities to illustrate the potential heterogeneity.

○ The annual incidence rate for non leukemia cancer (44 cases/million) seems to be quite low compared to the overall incidence rate reported in Stealiarova-Foucher *et al.* 2017 for South America (133.9 cases/million, table S3), even if leukemia were excluded, and compared to the range reported in figure 2 legend ("more than 349 cases/million in the highest category"). This point needs clarification.

○ The scan method selects the most likely cluster on the basis of a likelihood ratio that is calculated for each spatial window (from one municipality to the maximum size defined by the user). Likelihood ratios are very similar between two consecutive windows (as only one municipality or a small number of municipalities is added to the window at each step) that's why overlapping significant clusters can be detected (as in this study). Wouldn't it be interesting to run the scan method to detect non-overlapping cluster (this is an option in SaTScan)?

○ Figure 2 is not as clear as figure 5. The four clusters can't be identified precisely. Presenting one map for each cluster may be useful (or just 3 overlapping circles around the detected cluster areas in the central area and another circle centered on Cali).

Discussion:

○ The study cited in reference 18 focused on space-time clustering not on cluster detection. The authors reviewed the studies which tested for a space-time interaction, i.e. a general tendency of childhood cancer cases to occur more closely in space and time than expected under independent spatial and temporal patterns. This issue is really different from the question of detecting localized excesses of cases. The difference between space-time interaction and cluster detection (and even spatial clustering) should be clearly stated when presenting the results from reference 18.

○ Regarding the heterogeneity of the previous study results on CC cluster detection, I agree with the authors that considering count data in geographical units or individual point data to detect localized clusters may lead to different conclusions, and may explained some differences between study results (end of page 7). However, I wouldn't say that the ecological approach is more sensitive than the point analysis on the basis of a small number of studies.
It may be that some excesses actually existed in some countries in some particular time periods (not necessarily related to an environmental factor), while cases were more homogeneously distributed in other countries. Sensitivity refers to situations of true excesses.

○ At the end of the discussion, the authors noted "that in the SNCCC could exist some level of sub-registry caused by the limitation of the access to the health care services, especially in rural and isolated areas. "Again, this point is really important for the interpretation of the results (already noted as a major comment).

References

1. Spector LG, Pankratz N, Marcotte EL: Genetic and nongenetic risk factors for childhood cancer. *Pediatr Clin North Am*. 2015; **62** (1): 11-25 PubMed Abstract | Publisher Full Text
2. Schüz J, Erdmann F: Environmental Exposure and Risk of Childhood Leukemia: An Overview.*Arch Med Res*. **47** (8): 607-614 PubMed Abstract | Publisher Full Text
3. Rosychuk RJ, Witol A, Stobart K: Childhood cancer trends in a western Canadian province: a population-based 22-year retrospective study.*Pediatr Blood Cancer*. 2010; **55** (7): 1348-55 PubMed Abstract | Publisher Full Text
4. Schmiedel S, Blettner M, Kaatsch P, Schüz J: Spatial clustering and space-time clusters of leukemia among children in Germany, 1987-2007.*Eur J Epidemiol*. 2010; **25** (9): 627-33 PubMed Abstract | Publisher Full Text
5. Kulkarni K, Stobart K, Witol A, Rosychuk RJ: Leukemia and lymphoma incidence in children in Alberta, Canada: a population-based 22-year retrospective study.*Pediatr Hematol Oncol*. 2011; **28** (8): 649-60 PubMed Abstract | Publisher Full Text
6. Amin R, Hendryx M, Shull M, Bohnert A: A Cluster Analysis of Pediatric Cancer Incidence Rates in Florida: 2000–2010. *Statistics and Public Policy*. 2014; **1** (1): 69-77 Publisher Full Text
7. Goujon S, Kyrimi E, Faure L, Guissou S, et al.: Spatial and temporal variations of childhood cancers: Literature review and contribution of the French national registry.*Cancer Med*. **7** (10): 5299-5314 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: epidemiology; biostatistics; childhood cancer; space-time variations; environmental epidemiology;

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 May 2021
Edgar F. Manrique-Hernández, Universidad Industrial de Santander. Bucaramanga., Bucaramanga, Colombia

Clustering of childhood cancer in Colombia: a nationwide study

Reviewer #3

Stéphanie Goujon, Childhood and adolescent cancer team (EPICEA), UMR 1153 Centre of Research in Epidemiology and StatisticS (CRESS), Villejuif, France
21 Apr 2021 | for Version 1

The nationwide study conducted by Manrique-Hernández and colleagues described the spacetime distribution of non-leukemia childhood cancers in Colombia. Based on data from the national health surveillance system, the study aimed at detecting spatial and spatio-temporal localised excesses of cases over the period 2014-2017, on the municipality scale. Describing spatial and temporal variations is of a great importance for childhood cancer surveillance. The authors used registry based data and the scan method developed by M. Kulldorff, which is appropriate for such a study. In a purely spatial analysis, they detected four widespread overlapping clusters of nonleukemia cases in two different areas, in which

the number of observed cases were about twice the numbers expected under the hypothesis of homogeneous incidence rates over the whole study period. Space-time analyses identified clusters in the same areas, with excesses observed during shorter time periods.

Major comments:
○ There is a major issue concerning the interpretation of the results.

The authors concluded that the detected clusters "may suggest infectious or environmental factors associated with its incidence" (abstract and main text). The presence of localized clusters might actually suggest that a risk factor is present with a higher prevalence in these places than elsewhere in Colombia, but spatial heterogeneity might also be due to differences in case registration. I consider this point is of a primary importance and should be discussed further in the paper and in the abstract (it was discussed rapidly at the end of the discussion section).

To discuss that point, it would be useful to describe (at least briefly) in the paper the surveillance system in Colombia and to provide information on how the cases are identified nationwide. Are the data provided automatically to the NHS by hospital centers? Or do SIVIGILA members visit the hospital centers to collect data actively? Which hospital centers are visited or contacted? What are the main care pathways in Colombia? The authors cited several interesting papers that were written in Spanish and therefore not easily understandable.

Spatial differences in case registration might be observed because of under-diagnoses or difficulties for the registry to identify cases in some regions. In the paper by RodriguezVillamizaron et al. on childhood leukemia (2020), the authors indicated that childhood cancer became a priority in Colombia since 2010. Do the authors consider that the new regulation was adopted homogeneously in Colombia since that date, or is it possible that access to diagnosis was different over the study period 2014-2017 depending on the place of residence? Besides, were all medical reports available nationwide during the study period?

In Rodriguez-Villamizaron et al. 2020, under-registration was also discussed and it was estimated that 17% of AL cases were not captured by the registry in 2016 ("During this year [2016] 1394 incident cases of childhood cancer were identified and 1206 (86.5%) of them were reported to the NSSCC [...] the NSSCC captured and registered 83% of all incident cases of childhood cancer in Colombia during 2016"). If the missing cases were not homogeneously distributed in the municipalities, this could lead to clusters in areas where registration was more exhaustive. Do the authors have similar information for non-leukemia cases?

Author's response to reviewer:
Thank you for your detailed review and comments. Certainly, spatial heterogeneity in this type of ecological spatial analysis can be observed due to diagnosis or reporting heterogeneity. For this study we selected as source of cancer cases the report to the national surveillance system for childhood cancer (NSSCC) from SIVIGILA because this is the

strongest and more complete health information system that is operating in all 1,122 municipalities in Colombia. Unfortunately, the cancer population-based registries in Colombia are limited to four regions in the country which are representative of specific urban areas but do not represent the full spectrum of municipalities and regions in Colombia. The national surveillance system (SIVIGILA) is operated by the National Institute of Health (INS for Spanish) as a mandatory, systematic, and continuous registry with standardized protocols for more than 100 events of interest in public health. The system operates permanently in all municipalities based on immediate report for selected health events and weekly report for all events, including childhood cancer. The system is administrated and regulated by INS and operational support and training for municipalities is provided by the health secretary of each state (departments in Colombia). Childhood cancer surveillance began in 2008 when acute leukemia was included as a mandatory health notification event. In 2013 the system was extended to all types of childhood cancer. The system preserved the core formats and software for reporting acute childhood leukemia and therefore the extension to other cancer types had a shorter learning curve for the surveillance system´s personnel in municipalities. During the study period, notification of non-leukemia cancer were reported for 432 municipalities in almost all departments and districts (including municipalities with predominantly rural remote areas), which support the wide coverage of the surveillance system.

For a previous study, we conducted a comparison between national high-cost account registry and SIVIGILA report during 2016 and found that 1394 incident cases of childhood cancer were identified and 1206 (86.5%) of them were reported to SIVIGILA, indicating that the systems captured 83% of all incident cases of childhood cancer in Colombia, which included non-leukemia cases. The 188 cases missing in SIVIGILA corresponded to different cancer diagnosis and municipalities distributed in 28 departments across the country. Therefore, we assumed that the presence of underreporting it is not concentrated in specific areas of the country and underreporting although is present, might not be the main explanation for the spatial heterogeneity in our results. However, health care access to cancer diagnosis is limited to specific regions in the country located in the main capital cities in Colombia and therefore delay in diagnosis (and derived delayed in reporting) might be present in remote semirural and rural municipalities. The analysis was conducted for 2014-2017 and databases were consolidated in 2019, therefore cases with delayed diagnosis had the opportunity to be included in the last two years as SIVIGILA required the reporting of incident and prevalent cases since 2014 for non-leukemia cases. However, the cases with missing diagnosis due to limitations in access to health care might be still present in the study but cannot be quantified.

We have summarized and added these points or potential underreporting and underdiagnosis in the discussion section.

- ○ A large cluster of childhood leukemia was detected in the center of Colombia. Can it be related to the large cluster of non-leukemia cases reported in the present study?

Author's response to reviewer:
Thank you for your comment. One cluster for childhood leukemia was also identified in the center of the country. Clusters for both leukemia and non-leukemia cases might be related

to each other, however the non-leukemia cluster is larger (327 municipalities compared with 109 identified in the leukemia cluster), more expanded to the North of the leukemia cluster, and with higher incidence rates located in municipalities with predominant rural areas. We have added this comment in the discussion section.

- ○ The four spatial clusters detected in the study were located in two distinct regions, as three clusters overlapped. The space-time clusters corresponded to the same spatial areas. In all, as a conclusion, I would say that only two clusters were detected (maybe over 2 years only) and I suggest to discuss the fact that the cluster frontiers were difficult to delineate (a wellknown limit of the cluster detection methods). The size of the largest detected cluster, in the central region, and the magnitude o the relative risk should also be discussed. A relative risk of about 2 in such a large area is quite surprising and unexpected. If an infectious or environmental factor was responsible for such an excess, it would have to be strongly associated to the risk of non-leukemia cancer and highly prevalent in the cluster area (in comparison to the remaining part of Colombia).

Author's response to reviewer:
Thank you for your comment. We agree with your comment about the overlapping of clusters and concentration of cases in two clusters (rather than four) with the first expanded in a large are in the central region with a high relative risk. We have corrected this aspect in the abstract and conclusion and added a more detailed comment on the limitation to delimited clusters in the central region in the discussion section.

○ The statistical methods used in this study, Moran's test for spatial autocorrelation and the Kulldorff's scan method for cluster detection, were appropriate. However, some details on the methodology and the paramete,rs used could be added. In particular, the authors could add a reference for the Moran's test for spatial autocorrelation, and explain how the neighborhood was defined in the study. Were two municipalities considered neighboring areas if the distance between their centroids was below a given threshold or if they shared a common border?

Author's response to reviewer:
Thank you for your comment. The analysis considered neighboring based on the distance between the municipality´s centroids based on the Euclidean distance measured between two centroids of municipalities (with no threshold specification). We have added this comment in the methods section.

○ It would also be useful to provide further details on the scan method (estimation of a likelihood ratio for each window and selection of the most likely cluster, i.e. the window associated to the maximum ratio), and the simulations that were conducted to evaluate the significance thresholds (how many simulations were done?).

Author's response to reviewer:
Thank you for your comment. The selection of the most likely cluster was selected based on the p-value of the log likelihood ratio (p>0,05 was considered statistically significant) and 999 replications were used in the simulation to evaluate the significance of the inference. We have added this comment in the methods section.

    ◦ Several other studies should be referred to in the introduction and discussion sections to provide more accurate information on childhood cancer etiology and the literature on childhood cancer clusters.

○ (Suggested publications):

Etiology:
'Genetic and nongenetic risk factors for childhood cancer' by Spector et al. (20151 ○).
'Environmental Exposure and Risk of Childhood Leukemia: An Overview' by Schüz et al. (20162).

○ Clustering:
'Childhood cancer trends in a western Canadian province: a population-based 22-year retrospective study' by Rosychuk et al. (20103).
○ 'Spatial clustering and space-time clusters of leukemia among children in Germany, 1987-2007' by Schmiedel et al. (20074).
○ 'Leukemia and lymphoma incidence in children in Alberta, Canada: a population-based 22- year retrospective study by Kulkarni et al. (20115).
 ○ 'A cluster analysis of Pediatric Cancer Incidence Rates in Florida: 2000-2010' by Amin et al. (20146).
○ 'Spatial and temporal variations of childhood cancers: Literature review and contribution of the French national registry' by Goujon et al. (20187).

Author's response to reviewer:
Thank you for your suggestions. We have added the suggested references in the manuscript.

<u>Minor comments:</u>

Abstract:
"A sensitivity analysis was conducted with different upper limit parameters for the at-risk population." The upper limit was for the at-risk population included in the cluster.

Author's response to reviewer:
Thank you for your correction. We have edited the sentence.

○ There are 1122 municipalities nationwide. Several municipalities had no observed cases, so that the cases were actually distributed in 432 different municipalities. However, all the municipalities were included in the analyses (if the at-risk population was not null). I suggest therefore to report that "2006 cases were distributed in 1122 municipalities".

Author's response to reviewer:
Thank you for your comment. We have edited the sentence specifying the 1122 municipalities in the methods and the identification of cases in 432.

○ Regarding CNS tumors, it would be informative to specify whether non malignant cases were included.

Author's response to reviewer:
Thank you for your comment. According to SIVIGILA protocol, CNS tumors include malignant and non-malignant cases. We have added this specification in the methods section.

◯ Conclusion: differences in case registration should also be considered as a possible explanation.
Author's response to reviewer:
Thank you for your suggestion. We have added that differences in case diagnosis in remote rural areas should also be considered as a possible explanation.

Introduction:

Based on reference 1, "The mean annual incidence of CC was estimated at 140.6 cases per million children". This incidence rate was estimated worldwide, which could be specified.
Author's response to reviewer:
Thank you for your suggestion. We have edited the sentence.

◯ "There are several conditions that have been identified as risk factors". The factors cited in this sentence were associated with childhood cancer with different degree of evidence. Some factors are considered as known (high dose ionising radiation, chemotherapy, certain genetic syndromes and some genetic polymorphisms, some viruses in lymphomas) or highly suspected (domestic and occupational parental exposure, socioeconomic conditions, infections and immune system stimulation for leukemia, birth weight, benzene exposure, air pollution) risk factors, while for other factors the literature is more heterogeneous and no firm conclusion can be drawn to date (tobacco and alcohol consumption). It is important to consider that point when discussing the etiology.
Author's response to reviewer:
Thank you for your suggestion. We have added this comment in the introduction section for possible etiology for clusters.

The following sentence is quite long and difficult to understand: "contributing to the generation of hypotheses about possible etiologies. Spatial analysis has been previously used for the study of CC, mainly for studying the geographical distribution of leukemias, since this type of analysis allows for the identification of space and time variations in a geographical area that generate clusters that indicate an increase in the tendency of the cases".

Author's response to reviewer:
Thank you for your comment. We have edited the sentence.

◯ At the end of the introduction, several studies on childhood cancer cluster detection are cited, but only one reference (ref 5) is provided for non-leukemia cancers. Maybe some of the references I reported above could be added here.

Author's response to reviewer:

Thank you for your comment and suggested references. We have added some of the suggested references for non-leukemia cancers

Methods:

CC cases were identified by the National Surveillance System for Public Health, which registers "the newly confirmed and probable cases of CC". It is unclear to me what "probable cases" means? It seems that CC diagnoses are confirmed on the basis of diagnostic exams and coded according to the ICCC-3, so could the authors explain what are the probable cases that are registered (if not confirmed why are they registered?).

Author's response to reviewer:
The SIVIGILA protocol for childhood cancer includes the case definition of "probable" when first clinical diagnosis is given and the confirmation of cases should be reported in a maximum of four weeks. All included cases in the analysis are confirmed cases. We have added a clarification in this sentence.

◯ Are non-malignant CNS tumors registered in SIVIGILIA, and included in the study? Based on the number of CNS tumor cases reported in the result section (17.9% of non-leukemia cases), I assume that only malignant CNS tumors were included in the study. This information should be added in the main text and the abstract

Author's response to reviewer:
Thank you for your comment. According to SIVIGILA protocol, CNS tumors include malignant and non-malignant cases. We have added this specification in this section.

Statistical analysis:

"We performed a descriptive analysis calculating frequencies and central tendency measurements." I don't understand what "central tendency measurements" refers to?

Author's response to reviewer:
Thank you for your comment. We refer to summary and dispersion measurements (i.e. mean and standard deviation), however in reported results we are providing only percentages so we agree on eliminating the word. We have edited the sentence specifying frequencies and percentages.

◯ I understand that standardized rates were considered to account for potential differences in the age distribution of the pediatric population between municipalities. Were those potential differences also accounted for in the SaTScan analyses?

Author's response to reviewer: No, the cluster analyses were conducted with total population by municipality and year.

Results:

731 cases were excluded from the analyses, of which 57 had an unknown municipality of residence. It would be interesting to describe those cases in terms of type of diagnosis, year of diagnosis. Could the authors get other geographic information related to the area of residence? Were those cases grouped in a particular region?

Author's response to reviewer:
Thank you for your comment. The cases were reported with codification for department with no specification of municipality. The 57 cases belonged to 20 departments distributed across the country (Atlántico, Magdalena, Meta, Cesar, La Guajira, Valle, Tolima, Antioquia, Cundinamarca, Huila, Caquetá, Casanare, Amazonas, Chocó, Putumayo, Cauca, Santander, Bolívar, Norte de Santander y Córdoba).
We have added this information in the results section.

◯ It would be useful to describe the distribution of the pediatric population in the 1122 municipalities to illustrate the potential heterogeneity.

Author's response to reviewer:
Thank you for your comment. We have added a sentence describing the childhood population by municipaliy. (mean 9,880, median 3,336, minimum 149 in La Guadalupe municipality of Guainía and maximum 1,381,081 in Bogotá, the capital district)

◯ The annual incidence rate for non leukemia cancer (44 cases/million) seems to be quite low compared to the overall incidence rate reported in Stealiarova-Foucher et al. 2017 for South America (133.9 cases/million, table S3), even if leukemia were excluded, and compared to the range reported in figure 2 legend ("more than 349 cases/million in the highest category"). This point needs clarification

◯ The scan method selects the most likely cluster on the basis of a likelihood ratio that is calculated for each spatial window (from one municipality to the maximum size defined by the user). Likelihood ratios are very similar between two consecutive windows (as only one municipality or a small number of municipalities is added to the window at each step) that's why overlapping significant clusters can be detected (as in this study). Wouldn't it be interesting to run the scan method to detect non-overlapping cluster (this is an option in SaTScan)?

Author's response to reviewer:
Thank you for your comment. We have run the analysis with no overlap and found two spatial clusters. Therefore, results are presented for both analysis concluding the presence of two clusters.

Figure 2 is not as clear as figure 5. The four clusters can't be identified precisely. Presenting one map for each cluster may be useful (or just 3 overlapping circles around the detected cluster areas in the central area and another circle centered on Cali).

Author's response to reviewer:
Thank you for your comment. Figure 2 is showing rates by municipality and therefore the clusters are not identified and they are shown in figure 3 (previous figure 5).

Discussion:

The study cited in reference 18 focused on space-time clustering not on cluster detection. The authors reviewed the studies which tested for a space-time interaction, i.e. a general tendency of childhood cancer cases to occur more closely in space and time than expected under independent spatial and temporal patterns. This issue is really different from the question of detecting localized excesses of cases. The difference between space-time interaction and cluster detection (and even spatial clustering) should be clearly stated when presenting the results from reference 18.

Author's response to reviewer:
Thank you for your comment. We have complemented the paragraph and made this clarification.

○ Regarding the heterogeneity of the previous study results on CC cluster detection, I agree with the authors that considering count data in geographical units or individual point data to detect localized clusters may lead to different conclusions, and may explained some differences between study results (end of page 7). However, I wouldn't say that the ecological approach is more sensitive than the point analysis on the basis of a small number of studies. It may be that some excesses actually existed in some countries in some particular time periods (not necessarily related to an environmental factor), while cases were more homogeneously distributed in other countries. Sensitivity refers to situations of true excesses.

Author's response to reviewer:
Thank you. We agree with your comment. We have edited the sentence.

○ At the end of the discussion, the authors noted "that in the SNCCC could exist some level of sub-registry caused by the limitation of the access to the health care services, especially in rural and isolated areas. "Again, this point is really important for the interpretation of the results (already noted as a major comment).

Author's response to reviewer:
Thank you. We have added this aspect in abstract and conclusions and it is better explained in the discussion section.

References
1. Spector LG, Pankratz N, Marcotte EL: Genetic and nongenetic risk factors for childhood cancer. Pediatr Clin North Am. 2015; 62 (1): 11-25 PubMed Abstract | Publisher Full Text
2. Schüz J, Erdmann F: Environmental Exposure and Risk of Childhood Leukemia: An Overview.Arch Med Res. 47 (8): 607-614 PubMed Abstract | Publisher Full Text
3. Rosychuk RJ, Witol A, Stobart K: Childhood cancer trends in a western Canadian province: a population-based 22-year retrospective study.Pediatr Blood Cancer. 2010; 55 (7): 1348-55 PubMed Abstract | Publisher Full Text
4. Schmiedel S, Blettner M, Kaatsch P, Schüz J: Spatial clustering and space-time clusters of leukemia among children in Germany, 1987-2007.Eur J Epidemiol. 2010; 25 (9): 627-33

PubMed Abstract | Publisher Full Text
5. Kulkarni K, Stobart K, Witol A, Rosychuk RJ: Leukemia and lymphoma incidence in children in Alberta, Canada: a population-based 22-year retrospective study.Pediatr Hematol Oncol. 2011; 28 (8): 649-60 PubMed Abstract | Publisher Full Text
6. Amin R, Hendryx M, Shull M, Bohnert A: A Cluster Analysis of Pediatric Cancer Incidence Rates in Florida: 2000–2010. Statistics and Public Policy. 2014; 1 (1): 69-77 Publisher Full Text
7. Goujon S, Kyrimi E, Faure L, Guissou S, et al.: Spatial and temporal variations of childhood cancers: Literature review and contribution of the French national registry.Cancer Med. 7 (10): 5299-5314 PubMed Abstract | Publisher Full Text.

Competing Interests: No competing interests were disclosed.

Reviewer Report 15 April 2021

https://doi.org/10.5256/f1000research.30704.r81741

? **Smita Asthana**

Division of Epidemiology and Biostatistics, Institute of Cytology and Preventive Oncology, NOIDA, New Delhi, Delhi, India

It's a well written paper, a nationwide summarization of non-leukemia childhood cancer cases clustering in Colombia from 2014 to 2017. The topic is important. I recommend indexing with minor changes. My suggestions as given below.

1. Since it's a study on non-leukemia cases this should be reflected in title of the manuscript also. Instead of 'childhood cancer' add 'non-leukemia childhood cancer'.

2. Introduction section second line, first paragraph "Childhood cancer (CC) .... in this population." Define the population instead of 'this population'.

3. Reference 2 can be given in English.

4. Introduction section, paragraph 1 line 5, for reference 2 please check the statement "The world health organization (WHO) estimates that nearly 300,000 new cases of CC are diagnosed every year in children between 0 and 19 years of age". The article that the authors are referring to is giving the incidence as 400,000 and not 300,000, and age group is also given different in this WHO fact sheet.

5. Method section - Its better to give reference for International childhood cancer third edition.

6. It would be better if the authors discuss in a few lines the implications of the study.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 May 2021

Edgar F. Manrique-Hernández, Universidad Industrial de Santander. Bucaramanga., Bucaramanga, Colombia

Clustering of childhood cancer in Colombia: a nationwide study

Reviewer #2: Smita Asthana. Division of Epidemiology and Biostatistics, Institute of Cytology and Preventive Oncology, NOIDA, New Delhi, Delhi, India

1. Since it's a study on non-leukemia cases this should be reflected in title of the manuscript lso. Instead of 'childhood cancer' add 'non-leukemia childhood cancer'.

Author's response to reviewer: Thank you for your detailed review and comments. We have edited the title
   - Clustering of non-leukemia childhood cancer in Colombia: a nationwide study.

1. Introduction section second line, first paragraph "Childhood cancer (CC) …….. in this population." Define the population instead of 'this population'.

Author's response to reviewer: Thank you for your detailed review and comments. This sentence uses the term "this population" since we previously referred to "children and adolescents" in the first line. We consider removing "in this population" from the sentence to avoid confusion for readers.

1. Reference 2 can be given in English

Author's response to reviewer: Thank you for your detailed review and comments. We have eliminated reference 2 (WHO fact sheet) as the source of data is the study reference 1.

1. Introduction section, paragraph 1 line 5, for reference 2 please check the statement "The world health organization (WHO) estimates that nearly 300,000 new cases of CC are diagnosed every year in children between 0 and 19 years of age". The article that the authors are referring to is giving the incidence as 400,000 and not 300,000, and age group is also given different in this WHO fact sheet.

Author's response to reviewer: Thank you for your detailed review. We have corrected the number (400,000 new cases) and checked the age range is correct according to the WHO fact sheet and reference 1.

1. Method section - Its better to give reference for International childhood cancer third edition.

Author's response to reviewer: Thank you for your comment. We have added the suggested reference in the manuscript.

1. It would be better if the authors discuss in a few lines the implications of the study.

Author's response to reviewer: Thank you for your comment. We have added the implications of the study and limitations of data and methods in the discussion section.

Competing Interests: No competing interests were disclosed.

Reviewer Report 04 March 2021

https://doi.org/10.5256/f1000research.30704.r79234

? **Lisandro Agost** iD

Centro de Ecología y Recursos Naturales Renovables (CERNAR) – IIByT CONICET-UNC, Córdoba, Argentina

The article attempts to identify non-leukemia childhood cancer clusters (population under 15 years of age), across Colombia between 2014 and 2017. It uses data from official government sources, both disease and population, at the municipal level.

Overall the article is clear and concise, perhaps too concise on some points. It is important research but I would consider the following changes and suggestions before indexing:

1. The title is omitting information as it is not a study of all types of cancers. It should clarify that it is about non-leukemia childhood cancer.

Introduction section:

1. *The mean annual incidence of CC was estimated at 140.6 cases per million children between the age 0–14 years in the period of 2001 to 2010.* A geographical reference is missing - worldwide

or what?

2. *The world health organization (WHO) estimates that nearly 300,000 new cases of CC are diagnosed every year in children between 0 and 19 years of age.* Although the data is important, perhaps it would be better to obtain it for the 0-15 age group in order to be more coherent in the text.

Methods section:

1. *Colombia is a country located in the north of South America...* Description is too long, make a map of the study area or use the standardized rate map (Figure 2).

2. *Women make up 51.2% of the population, and children under the age of 15 years make up 22.6% compared to adults over the age of 65 years which represent 9.1%.* It is not clear why the data is divided in this way. Why don't the authors put the total female - male population and then only the population under 15 (divided into female and male if want)?

3. *The global Moran index was calculated to estimate the spatial autocorrelation.* It is not clear whether they do this statistical analysis in STATA or with what software.

4. *We used an upper limit of the population at risk of 25% and for a sensitivity analysis we assess the results using upper limits of 50% and 10%.* I would explain in a little more detail what this analysis is for and why it is done.

Results section:

1. It would be interesting to show in this section (or in the annexes) a table with the distribution of cases per year, percentage of each group, crude and standardized rate, divided by ICCC 3 group (including leukaemias).
   Another interesting table, if it is not too big, would be to show the results of the distribution of cases per department, per year.

2. In general, none of the graphs or maps show the data sources. I don't know if this is a journal rule or if the authors omitted them.

3. Figures 3 and 4 are not useful. They should be redrawn showing more clearly the units of analysis (municipalities), and the clusters individually. For figure 3, for example, I suggest making four enlarged maps (from the general map of Colombia) showing the departments with the number of municipalities that include each cluster. Or simply show the municipalities included in the cluster. Why not use a map similar to the one in figure 2 or 5?

Discussion section:

1. In general, I feel that the discussion of the results is not enough. The authors should expand it, taking into account that they found numerous clusters, with important associated indicators.
   At no point is there an in-depth discussion of the results of this research in relation to the findings of previous research on leukaemia clusters (Space-time clustering of childhood leukemia in Colombia: A nationwide study, 2020).
   It would be very interesting to complement these studies and discuss what they have in common.

2. *The spatial studies based on the analysis of areas (ecological approach) such as this study, and the studies in Canada and in Palestine*[20⁻22],*...*
   There seems to be an error in the text or in the citation.

3. It should be made explicit or discussed why the authors used only this cluster analysis method (Kulldorf's) and not others. Also, what other statistical methods the authors believe could complement, in future research, the results obtained.

4. *Non-leukemia CC clusters identified in Colombia are located mainly in the central region of the country near the mountain ranges...* In this paragraph, I believe that the discussion of possible risk factors or associated risk factors for this disease should be expanded a little further and supported by literature from other research, reports or reviews. While the research does not set out to study causality, it can provide new lines of research by collecting and discussing possible risk factors.
   The bibliography of the risk factors mentioned above should be expanded.

5. I did not find the definition of SNCCC.

6. The discussion of limitations is poor. They should be expanded, better explaining the limitations and scope of the methodologies used and the results obtained.

Conclusion section:
1. Since the conclusion focuses on the possible factors associated with the results obtained, it seems important to me to develop these topics further in the Discussion section.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Ecology - Epidemiology - Childhood Cancer - Pesticide Risk Indicators

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 May 2021

Edgar F. Manrique-Hernández, Universidad Industrial de Santander. Bucaramanga., Bucaramanga, Colombia

Clustering of childhood cancer in Colombia: a nationwide study

Reviewer #1: Lisandro Agost. Centro de Ecología y Recursos Naturales Renovables (CERNAR) – IIByT CONICET-UNC, Córdoba, Argentina.

The article attempts to identify non-leukemia childhood cancer clusters (population under 15 years of age), across Colombia between 2014 and 2017. It uses data from official government sources, both disease and population, at the municipal level.

Overall the article is clear and concise, perhaps too concise on some points. It is important research but I would consider the following changes and suggestions before indexing:

1. The title is omitting information as it is not a study of all types of cancers. It should clarify that it is about non-leukemia childhood cancer.

Author's response to reviewer: Thank you for your detailed review and comments. We have edited the title.
   ○ Clustering of non-leukemia childhood cancer in Colombia: a nationwide study.

Introduction section:
1. The mean annual incidence of CC was estimated at 140.6 cases per million children between the age 0–14 years in the period of 2001 to 2010. A geographical reference is missing - worldwide or what?

Author's response to reviewer: Thank you for your detailed review and comments. We have added the clarification to worldwide reference in the manuscript.
1. The world health organization (WHO) estimates that nearly 300,000 new cases of CC are diagnosed every year in children between 0 and 19 years of age. Although the data is important, perhaps it would be better to obtain it for the 0-15 age group in order to be more coherent in the text.

Author's response to reviewer: Thank you for suggestion. The data is given by WHO for that range of age in the reference material. However, attending your comment on age range we modified the order of this sentence and it is now presented first and following information is related only to age range 0-14 years.

Methods section:
1. Colombia is a country located in the north of South America... Description is too long, make a map of the study area or use the standardized rate map (Figure 2).

Author's response to reviewer: Thank you for your suggestion. We have edited the population paragraph eliminating the details of geographical limits.
1. Women make up 51.2% of the population, and children under the age of 15 years

make up 22.6% compared to adults over the age of 65 years which represent 9.1%. It is not clear why the data is divided in this way. Why don't the authors put the total female - male population and then only the population under 15 (divided into female and male if want)?

Author's response to reviewer: Thank you for your detailed review and comments. We have edited the sentence in the manuscript.

1. The global Moran index was calculated to estimate the spatial autocorrelation. It is not clear whether they do this statistical analysis in STATA or with what software.

Author's response to reviewer: Thank you for your detailed review and comments. We have added the clarification or ArcGIS use in the manuscript.

1. We used an upper limit of the population at risk of 25% and for a sensitivity analysis we assess the results using upper limits of 50% and 10%. I would explain in a little more detail what this analysis is for and why it is done.

Author's response to reviewer: Thank you for your suggestion. We have edited the sentence clarifying the it was conducted to identify consistency of clustering results across different upper limits.

Results section:

1. It would be interesting to show in this section (or in the annexes) a table with the distribution of cases per year, percentage of each group, crude and standardized rate, divided by ICCC 3 group (including leukaemias). Another interesting table, if it is not too big, would be to show the results of the distribution of cases per department, per year.
2. In general, none of the graphs or maps show the data sources. I don't know if this is a journal rule or if the authors omitted them.

Author's response to reviewer: Thank you for your comments and suggestions. We have added Table 1 with the distribution of cases by sex, age group, year of diagnosis, and department or residence. Data source for childhood cancer is mentioned in the methods section (National Surveillance System for Public Health SIVIGILA, for its name in Spanish) but usually not included in maps as they are part of the manuscript.

1. Figures 3 and 4 are not useful. They should be redrawn showing more clearly the units of analysis (municipalities), and the clusters individually. For figure 3, for example, I suggest making four enlarged maps (from the general map of Colombia) showing the departments with the number of municipalities that include each cluster. Or simply show the municipalities included in the cluster. Why not use a map similar to the one in figure 2 or 5?

Author's response to reviewer: Thank you for your comment. We agree with your comment and have eliminated figures 3 and 4 as they show similar information in figure 5.

Discussion section:

1. In general, I feel that the discussion of the results is not enough. The authors should expand it, taking into account that they found numerous clusters, with important associated indicators. At no point is there an in-depth discussion of the results of this research in relation to the findings of previous research on leukaemia clusters (Space-time clustering of childhood leukemia in Colombia: A nationwide study, 2020). It

would be very interesting to complement these studies and discuss what they have in common.

Author's response to reviewer: Thank you for your comment. We have expanded the discussion section pointing out the potential meaning of the results and their relation with findings of clusters of leukemia in Colombia.

1. The spatial studies based on the analysis of areas (ecological approach) such as this study, and the studies in Canada and in Palestine[20–22],... There seems to be an error in the text or in the citation.

Author's response to reviewer: Thank you for your detailed review. We have edited the sentence in the manuscript.

1. It should be made explicit or discussed why the authors used only this cluster analysis method (Kulldorf's) and not others. Also, what other statistical methods the authors believe could complement, in future research, the results obtained.

Author's response to reviewer: Thank you for your comment. We have added a mention about the selection of method for cluster detection, its limitations and future analysis to complement data.

1. Non-leukemia CC clusters identified in Colombia are located mainly in the central region of the country near the mountain ranges... In this paragraph, I believe that the discussion of possible risk factors or associated risk factors for this disease should be expanded a little further and supported by literature from other research, reports or reviews. While the research does not set out to study causality, it can provide new lines of research by collecting and discussing possible risk factors. The bibliography of the risk factors mentioned above should be expanded.

Author's response to reviewer: Thank you for your comment. We have expanded the discussion of the potential factors that might be related to the clustering of non-leukemia childhood cancers in the center of the country, including infections, pesticides use, and other occupational and environmental risk factors for parents.

1. I did not find the definition of SNCCC.

Author's response to reviewer: Thank you for your detailed review and comments. We have corrected the acronym for SIVIGILA in the manuscript.

1. The discussion of limitations is poor. They should be expanded, better explaining the limitations and scope of the methodologies used and the results obtained.

Author's response to reviewer: Thank you for your comment. We have expanded the discussion of the limitations of the data sources and methodologies and the potential effects on the reported results.

Conclusion section:
Since the conclusion focuses on the possible factors associated with the results obtained, it seems important to me to develop these topics further in the Discussion section.

Author's response to reviewer: Thank you for your comment. The topics mentioned in the conclusion section are now better explained in the discussion section.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com