Estatística Espacial Aplicada aos Estudos de Incidência de Câncer.

Spatial Statistic Applied to Incidence Studies of Cancer.

Estadística espacial aplicada a los estudios de incidencia de cáncer.

Jane Kelly Oliveira Friestino¹
Denisa Mendonça
Pedro Oliveira²
Carla M. Oliveira³
Rosemeire de Olanda Ferraz⁴
Djalma de Carvalho Moreira Filho⁵

RESUMO: O Sistema de Informação Geográfica (SIG) como ferramenta em estudos epidemiológicos tem se tornado algo frequente na área da Saúde Coletiva. Em estudos sobre incidência de câncer, o uso da análise espacial ainda é incipiente no país, entretanto possui grande potencial para auxílio nos modelos de planejamento em saúde. O objetivo deste trabalho é apresentar um método empírico de análise de padrões espaciais para os casos de incidência de cânceres considerados raros que acometem a população com idades entre 0 e 19 anos, discutindo a aplicabilidade da estatística espacial e as possíveis utilizações da interpretação da disposição espacial dos eventos. Utilizou-se o SIG para georreferenciar as taxas de incidências de casos de cânceres de alguns grupos da Classificação Internacional de Câncer Infantil (CICI), ajustadas por idade e sexo, segundo os Centros de Saúde de suas residências. Foram estimadas as razões padronizadas de incidência (Standardized Morbidity Ratio-SMR). Para controlar a instabilidade das taxas devido ao pequeno número de casos, foi feita uma suavização pelo método Empirical Bayes. Comparando as taxas anuais, suavizadas pelo referido método Bayesiano, constatou-se diferenças entre elas quanto à distribuição espacial, com concentração destas nas regiões Sudoeste e Noroeste do município de Campinas. Conclui-se que o uso de estatística espacial como ferramenta de apoio

¹ School of Medical Sciences. University of Campinas – UNICAMP. Federal University of Fronteira Sul, Chapecó. Brasil. Email: jane.friestino@uffs.edu.br
² ISPUP-EPIUnit – Institute of Public Health; ICBAS - Abel Salazar Biomedical Sciences Institute. University of Porto, Porto, Portugal. Email: pnoliveira@icbas.up.pt
³ i3S - Institute for Research and Innovation in Health; INEB – Institute of Biomedical Engineering. University of Porto. ESTSP - School of Health Technology of Porto. Polytechnic Institute of Porto, Porto. Portugal. Email: carlaoliveir@gmail.com
⁴ School of Medical Sciences. University of Campinas – UNICAMP. Brasil. Email: rose01@fcm.unicamp.br
⁵ School of medical sciences. University of Campinas – UNICAMP. Brasil. Email: djalmore@unicamp.br

DOI: http://dx.doi.org/10.18569/tempus.v10i2.1772
à decisão, permite identificar a distribuição de casos de cânceres do grupo da CICI, contribuindo para o planejamento em saúde.


**ABSTRACT:** The Geographic Information System (GIS) as a tool in epidemiological studies has become something common in the area of Public Health. In studies on cancer incidence, analysis spatial is still in its infancy in the country; however possess great potential to aid in health planning models. The aim of this paper is to present an empirical method of spatial patterns analysis in cases of incidence of rare considered cancers that affect the population aged 0 to 19 years, discussing the applicability of spatial statistics and the possible uses of interpretation of the spatial arrangement of events. We used the GIS to georeference the incidence rates of cases of cancers of some groups of the International Classification of Childhood Cancer (ICCC), adjusted for age and gender, per health centers area from residences. Standardized Morbidity Ratios - SMR were estimated. To control the instability of rates due to the small number of cases, a smoothing the Empirical Bayes method was applied. Comparing the annual rates softened by Empirical Bayes it was found differences between them as the spatial distribution, with concentration of these regions in Southwest and Northwest of the city of Campinas, SP. In conclusion, the use of spatial statistics as a decision support tool can be identify the distribution of cases of ICCC group of cancers, contributing to the health planning.

Keywords: Spatial Analysis, Incidence, Neoplasms, Geographic Mapping

**RESUMEN:** El uso del Sistema de Información Geográfica (SIG), como herramienta en estudios epidemiológicos, se ha convertido en algo usual en Salud Pública. Aunque en estudios sobre el cáncer el análisis espacial es poco utilizado en el país, posee un gran potencial para ayudar en modelos de planificación de salud. El objetivo de este trabajo es presentar un método empírico de análisis de patrones espaciales en casos de incidencia de cánceres considerados raros que afectan a la población de 0 a 19 años, discutiendo la aplicabilidad de la estadística espacial y los posibles usos de la interpretación de la disposición espacial de los eventos. Se utilizó el GIS para georreferenciar las tasas de incidencia de casos de cánceres de algunos grupos de la Clasificación Internacional de Cáncer Infantil (CICI), según edad y sexo, y por centros de salud de sus hogares. Se estimaron la morbilidad estandarizada (SMR). Para controlar la inestabilidad, debido al pequeño número de casos, se realizó un suavizado a través del método Bayesiano Empírico. En la comparación de las tasas anuales suavizadas se encontraron diferencias entre ellas, como la distribución espacial, la concentración de estas regiones en el suroeste y noroeste de la ciudad de Campinas, SP. Se concluyó que el uso de la estadística espacial como herramienta de apoyo a la decisión permite identificar la distribución de casos de cánceres del grupo de cánceres del CICI, lo que contribuye a la planificación de la salud.

Palabras clave: análisis espacial, incidencia, neoplasias, mapeo geográfico.
INTRODUCTION

Chronic non-communicable diseases (NCDs) represent a public health problem, especially in underdeveloped and developing countries, as they exhibited rapid growth in recent years, with the presence of 30% of premature deaths recorded\(^1\). In Brazil, the NCDs represent about 77.2% of the total burden of diseases, especially cardiovascular disease, which are noted for years as being the main cause of death and disability in developing countries\(^2,3\).

Inequalities in health are noted as one of the determinants for the worse health condition of the population, since the poorest layers of society are the hardest hit by the lack of access to effective healthcare, with a later diagnosis and the delay at the beginning of the treatment for the disease\(^4,5\). In developing countries such as Brazil, regional differences due to the existence of more developed regions than others also imply differences in the epidemiological profiles with typical diseases in each region\(^1\).

In this perspective, the increased occurrence of NCDs in Brazil has awakened the need of health plans in the country covering this issue, putting on the agenda the need to establish actions to confront them over the next 10 years\(^4\).

In order to promote the development and implementation of effective public policies for prevention and control of NCDs has listed the strengthening of direct action in the four main groups of diseases account for approximately 80% of all NCDs - the circulatory system, cancer, diseases chronic respiratory and diabetes and its risk factors: smoking, alcohol intake, physical inactivity, unhealthy diet and obesity\(^1,4\).

In the current context of NCDs, cancer deserves a different approach for its high prevalence, for expensive treatments and especially the high mortality rates that still have significant results with children and adolescents. In Brazil, in the period from 1996 to 2005, the Neoplasms were the third cause of death, behind only the “circulatory ailments and symptoms and signs and abnormal findings in clinical and laboratory examinations”; and for less than 19 years old, Neoplasms were the second most common cause of death, followed by: “Some infectious and parasitic diseases”\(^6\).

Despite the deaths, cancers in children under 19 years of age, represent on average 2.5%, high lethality requires attention and care on the part of health authorities for appropriate and effective actions in the mapping and identification of risk factors, as well as the monitoring of socio-economic conditions and geographic context present in the daily life of individuals affected by these diseases\(^6,7\).

In this aspect, many authors \(^8\) have incorporated the use of geographical space as a category of analysis in conjuncture with the social and environmental context with the determinants that influence human health. Thus, it is possible to identify morbidity and mortality characteristics, adding knowledge about events such as NCDs, which may change in the pattern of occurrences
According to the local reality experienced by the population in the different territories\textsuperscript{11}.

Currently we live a time of intense access to data from different information systems, including: IBGE, DATASUS, GIS etc. In particular, the Geographic Information System (GIS) is a technology that generates physical-spatial information which allow the geoprocessing: geo-referenced information analysis\textsuperscript{6}, which in the area of health, constitute as important decision support tools\textsuperscript{8, 12-13}.

Although there are these tools, of great potential for the health field, the actions planned together that impact health are still hampered by a limited use of GIS sometimes the lack of integration and interface with other computer languages, quality and presentation of different information systems in the country\textsuperscript{9,14}.

In the literature, some studies have emphasized the use and improvement of GIS in risk studies, especially for cases of communicable disease control\textsuperscript{10,15}, however the possibility of using the called Spatial Epidemiology advance, moreover, to provide knowledge and monitoring of other diseases. Spatial epidemiology is used to some of the main methods of spatial analysis in epidemiology, such as: disease mapping, ecological studies, environmental exposure and spatial cluster analysis\textsuperscript{7}.

In recent NCD studies, analysis with the use of GIS is a useful way of detecting spatial and temporal patterns of mortality and incidences of cancer and other diseases\textsuperscript{16-18}.

In relation to cancer studies, the incidence in the country is identified by Population-based Cancer Registries (PBCR), considered as fundamental in the epidemiological surveillance of cancer incidences, and indispensable for the development of epidemiological and clinical research, as well as for the planning and evaluation of cancer control\textsuperscript{19}.

In Brazil there are more than 20 PBCR (in the Federal District and in the state capitals, except in six: Boa Vista, Rio Branco, Porto Velho, Macapá, Florianópolis and São Luis) that implement systematized and unified information in a database, of the National Cancer Institute José Alencar Gomes da Silva (INCA) which allows the generation annual profile reports of cancers in the country. These records provide information on an ongoing basis regarding the number of new cases of a given locale, which detects the population segments most affected by cancer. The records are evaluated for coverage and quality of its data by indicators such as Diagnostic Histopathology Percentage (% VH) and Percentage Only by Death Certificate (ODC%). The PBCR allows to evaluate the impact of cancer in a given population\textsuperscript{19}.

As the incidence monitoring of cancer cases is done by PBCR, and records exist only in some cities in the country, there is no knowledge of the real-incidence and consequently the magnitude

\textsuperscript{6} Geoprocessing represents a set of technologies able to collect and treat georeferenced information, with the purpose of developing spatial analysis and surface modeling. Available at: <www.dsr.inpe.br#sthash.5V1aGK3M.dpu>. Accessed March 2016.
of neoplasms in all of the locations. With this, the estimates used are those performed by the INCA, the organ responsible for uniting all the systematized information produced by different PBCR\textsuperscript{20}.

Beyond the little knowledge about cancer incidence at the national level, when it comes to Pediatric Oncology (population between 0 and 19 years of age), this information becomes more difficult. This is because the cancers that affect children and adolescents are considered rare when compared to tumors in adults because about 2 to 3% of all malignant tumors occur in children under 19 years of age\textsuperscript{20}. This low incidence complicates even more the information monitoring and follow-up of the magnitude of the disease in the population\textsuperscript{20}. In addition, epidemiological studies on children and adolescents diagnosed with cancer are few referred to in Brazilian literature, although they are important studies to: monitor the disease, evaluate the availability of service network and assist the planning of services to meet the real needs of users\textsuperscript{20,21}.

The aim of this paper is to present an empirical analysis method of spatial patterns to the incidence of cases considered rare cancers that affect the population between 0 and 19 years of age, discussing the applicability of spatial statistics and the possible uses of the interpretation of the spatial layout of events.

**Area of study**

The municipality of Campinas is located in the Central-West region of São Paulo State, occupying an area of 795.70 km\textsuperscript{2} and is located northwest of the capital, about 96 km away. According to the 2010 population census, the population of Campinas was of 1,080,113 inhabitants with a population density of 1,372.31 inhabitants/km\textsuperscript{2}.

The population of Campinas has a life expectancy of 72.2 years, with a human development index (HDI) of longevity of 0.87, being above the national average which is 0.6. During the study period, there was an increase of 8.34 percent in population with less than 19 years of age\textsuperscript{22}. Per capita income was R$ 1,390.83, with the percentage of poor 3.16% and Gini index equal to 0.56\textsuperscript{23}.

In 2000, Campinas was divided by the Municipal Health in 47 coverage areas for health centers area, which are the geographic units used in the spatial analysis.

The main health care for children and adolescents with diseases such as cancer occurs in a high complexity unit exclusively for Pediatric Oncology. The service is located north of the city, in an area of 100,000 square meters, being made up of the following units: Hospital; Radiation therapy; Nuclear Medicine and imaging; Rehabilitation Center; Institute of Pediatrics; Central funding; Home of the child and family Research Institute, and although it is a philanthropic unit, has an agreement and technical cooperation with a public University of the State of São Paulo. That service represents around 85% of incident cases of cancer in children and adolescents in the municipality of Campinas and covers both patients who use exclusively the Unified Health System (SUS), as those who have private health plans or group health insurance\textsuperscript{24}.
Definition of the studied cases

The cases refer to children and adolescents living in Campinas-SP, between 0 and 19 years of age, diagnosed with cancer between 01 January 1996 to 31 December 2005. To define a case using the diagnosis date, from medical records and the International Classification of Diseases for Oncology, ICD-O, third revision

To prevent the inclusion of cases that do not belong to the population, was made a manual verification of the location of addresses according to the geographical area of residence of individuals at the time of diagnosis. The records of children with benign tumors were excluded. Only the four groups were considered as the most incidents in the municipality: Group I- Leukemias, myeloproliferative diseases, Group II- Lymphomas and reticuloendothelial neoplasms, Group III-CNS and miscellaneous intracranial and intraspinal neoplasms and Group IX- Soft tissue and other extraosseous sarcomas.

All addresses of residences of recorded cases were georeferenced, with the aid of the google Earth software, then identified which health center area belonged each of the individuals. Taking into account the geo-referencing of the residences of the cases, it can be observed through their spatialization, and geographical location in the municipality of Campinas, in accordance with the scope of the health centers area (Figure 1).

Figure 1. Geographic location of residence of cases of cancer in children under 19 years of Groups I, II, III and IX identified, according to scope health centers area of the municipality of Campinas SP, 1996 to 2005.

Location of the identified residence cases according to the health centers in Campinas, SP.
Incidence of cancer in children and adolescents

The crude incidence rates, per 1,000,000 inhabitants, to the municipality of Campinas-SP and the age-specific for ages of 0-14 years and 15-19 years were estimated and stratified according to diagnostic Group. Population denominators used to calculate rates were based on the population census in 2000 and the other years were used estimates of population projection inter-census of the Brazilian Institute of Geography and Statistics- IBGE22.

The Adjusted Incidence Rate (AIR) by age and gender was calculated, for 1,000,000 inhabitants, in order to control for the differences that may exist in the age and gender structure among the regions. Considering that \( y_{ijk} \) represents the number of cancer cases in the area of health \( ii \) \((i = 1, ..., I)(j = 1, ..., J)\) and age group \( kk \) \((k = 1, ..., K)\) and \( N_{ijk} \) the number of individuals at risk in the area, gender and age group, the AIR was estimated by the method of direct standardization:

\[
TIA_i = \frac{1}{\sum_{j=1}^{J} \sum_{k=1}^{K} N_{ijk}} \times e_{i} TIA_i = \frac{1}{\sum_{j=1}^{J} \sum_{k=1}^{K} N_{ijk}} \times e_{i} \quad , \quad i = 1, ..., I
\]

in that, \( N_{ijk} \) represents the number of the population pattern in the gender \( jj \) and age group \( kk \), \( e_{i} = \sum_{j=1}^{J} \sum_{k=1}^{K} (N_{jk} \times TIB_{ijk})\) represents the number of cases expected in each region, \( N_{j} \) represents the number of individuals in the population gender \( jj \) and age group \( kk \) for the municipality and \( TIB_{ijk} \) represents the gross incidence rate in the area, gender and age group, i.e., \( TIB_{ijk} = y_{ijk} / N_{ijk} \).

As a measure of risk was estimated standardized morbidity ratio (Standardized Morbidity Ratio, SMR) in each area, and \( y_{ijk} \) represents the number of observed cases and \( e_{ijk} \) represents the number of expected cases estimated by indirect method, i.e. \( e_{ijk} = \sum_{j=1}^{J} \sum_{k=1}^{K} e_{ijk} = \sum_{j=1}^{J} \sum_{k=1}^{K} e_{ijk} \) in that \( e_{ijk} = TIB_{ijk} \times N_{ijk} \times TIB_{ijk} = TIB_{ijk} \times N_{ijk} \) and \( TIB_{ijk} \) represents the gross incidence rate in gender \( jj \) and age group \( kk \) for the municipality and \( N_{ijk} \) the number of individuals in the population of the area, in the gender \( jj \) and the age group \( kk \).

Use of Empirical Bayesian Method

In order to control the instability of rates due to the small number in each region, an empirical Bayesian method was applied to smooth the rates. This methodology takes into account a weighted average rate of the surrounding areas and depends on the size of the population and the variability of these rates. Thus, the Empirical Bayesian estimator to the adjusted rate of i-th area is given by:

\[
EB_{i} = \theta_{i} TIA_{i} + (1 - \theta_{i}) TIA_{i}^{ \text{vis} } = \theta_{i} TIA_{i} + (1 - \theta_{i}) TIA_{i}^{ \text{vis} }\]

where:
\[
\theta_i = \frac{v_i}{v_i + \frac{TIA_i}{N_i}} \\
v_i = \sum_{i \in A_i} N_h (TIA_h - TIA_i) ^2 \quad \frac{TIA_i}{N_i} - \frac{TIA_i}{N_i} 
\]

Being that \(\theta_i\) represents the weighting factor, \(v_i\) represents the variance of observed rate around the Middle, and \(N_iN_i\) represents the number of individuals in the population in each region \(ii\). \(A_iA_i\) represents the set of neighbors of the region \(ii\) (excluding the region \(ii\)). \(N_i^v = \sum_{i \in A_i} N_h\) is the number of individuals in the population of neighbors and \(N_i^v = \frac{N_i^v}{n_i^v}\) represents the average of the population of the region \(ii\), where \(n_i^v = A_i\), \(n_i^v = A_i\) represents the number of neighbors of the region \(ii\) (excluding the region \(i\)). \(TIA_i, TIA_i\) represents the rate adjusted for the age in the region \(ii\) and \(TIA_i^v, TIA_i^v\) represents the rate adjusted for the age of the neighbors in the region \(ii\). Similarly, the smoothing was used to the SMR. By convention, if \(v_i < 0 v_i < 0\) so \(v_i = 0 v_i = 0\), then \(\theta_i = 0 \theta_i = 0\) and \(EB_i = TIA_i^vEB_i = TIA_i^v\). The \(\theta_i\) ranges between 0 and 1, and depends on the size of the population and of the variance \(v_i\) \(v_i\), that is, \(\theta_i\) approaches 1 in areas with a large population and 0 in areas with little population.

The analyses were carried out on the basis of the results obtained by means of the software: SPSS 22 and R version 2.14.1. The study was approved by Ethics Committee of Children’s Center of Boldrini, process number 006-11.

**Description of the territory findings**

A total of 180 cases were diagnosed, and 154 (86%) of children aged between 0 and 14 years and 26 (14%) children between the ages of 15 and 19 years. There is greater predominance of males in all cancer Groups (except Group III), and this difference between the sexes is more pronounced in Group II (Lymphomas and reticuloendothelial neoplasms). In children younger than 15 years of age, the most common group corresponds to the Leukemias, representing approximately 52.0% of cases in this age group. For the children older than 15 years of age, Group II was the most frequent with 42.2% of the cases. The crude incidence rate (per million) was 54.2, distributed in the age group of 0-14: 64.2 and 15-19: 28.2 per million (Table 1).
Table 1. Incidence of cancer in children and adolescents in the municipality of Campinas, São Paulo, according to age group and gender and CICI Group, during the period from 1996 to 2005.

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Age Group</th>
<th>Gross Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M/F* reason</td>
<td>0 - 14 n(%)</td>
</tr>
<tr>
<td>I. Leukemias, myeloproliferative diseases and myelodysplastic diseases</td>
<td>1.3</td>
<td>80(52.0)</td>
</tr>
<tr>
<td>II. Lymphomas and reticuloendothelial neoplasms</td>
<td>2.9</td>
<td>28(18.1)</td>
</tr>
<tr>
<td>III. CNS and miscellaneous intracranial and intraspinal neoplasms</td>
<td>0.7</td>
<td>35(22.6)</td>
</tr>
<tr>
<td>IX. Soft tissue sarcomas and other extraosseous sarcomas</td>
<td>2.3</td>
<td>11(7.1)</td>
</tr>
<tr>
<td>Overall Incidence</td>
<td>1.4</td>
<td>154(100)</td>
</tr>
</tbody>
</table>

*M/F: Male/Female

In relation to the spatial distribution of the incidence rates found with the use of Empirical Bayesian approach, there is a less fragmented identification of possible spatial aggregates for the occurrence of cancer in childhood. The identified areas were where they concentrate the health centers area with higher detection rates: region South and West of the municipality. However, the rates estimated by Empirical Bayesian method, indicate a higher concentration in the central region of the municipality (Figure 2).

Currently, the Bayesian models are currently the most widely used method of analysis for correction or for smoothing of estimates, and feature an increasingly relevant role in epidemiological studies. The approach according to the Bayesian empirical method was used, rather than the purely Bayesian methods, because it is a relatively simple, easy integration in geoprocessing environments and present, in many cases, very similar results to those purely Bayesian. This method has the advantage of being able to be easily used by health professionals and need a low computational resource in terms of installation capacity and in terms of the computational time.
Figure 2. Estimates by Empiric Bayesian method of average annual incidence rates of cancer among children and adolescents standardized according to age, in each health center area (Campinas-SP) in the 1996-2005 period.

Comparing figures 2 and 3, it is observed that there is a spatial distribution different from the average annual rates observed in relation to the rates estimated by the Empirical Bayesian method. From the SMR analysis, it can be seen that the highest SMR values are concentrated in C. S. equivalent to the regions Southwest and Northwest municipality (Figure 3).

Figure 3. Standardized Morbidity Ratio (SMR) for gender and age of children and adolescents, in accordance with the health center area (Campinas-SP), for period 1996-2005.
The study was based on an approach that considers the space and time, and found that the spatial pattern of cancer cases presented focus on regions corresponding to the Southwest and Northwest health Districts of the municipality.

The dynamic processes present on the territory help us to think that the knowledge produced through the Cartographic method are in constant motion, subject to continuous changes that should be explored in healthcare. For NCDs, particularly cancer in children and adolescents, the use of aggregate data for areas aimed at smoothing the random fluctuation inherent in small cases, allows for pointing out seldom used ecological development parameters for this type of disease.

The ecological type of study models has the potential to respond to existing needs in the country to scale the magnitude of pediatric cancer and may identify risk differentials. This information is useful for planning preventive measures or to scale the supply of professionals and services to meet the different existing demands.

In the present scenario of statistical advances and the evolution of the spatial epidemiology it is possible to perform information combinations both individually and collectively. However, faced with the limitations of their own aggregate data estimates for ecological studies, commonly described as the ecological fallacy, which are inferred to individual’s information obtained based on collective observations. However, some authors use this recent methodology in order to incorporate the individual effects, additional information about the corresponding areas, even though these are obtained from small samples. The combination of both sources of information, and individual, will mitigate the ecological bias in the estimated effect, and thus make a contribution to the implementation of strategies for the prevention and control of oncological disease.

In this study we only used methods involving a spatial component, but the evolution over time can be an important feature to be considered, and thus for the next steps could be used spatiotemporal modeling to further research initiated here.

FINAL CONSIDERATIONS

With the use of spatial statistics, it was possible to identify the distribution of cases of cancer diagnosed according to the groups of ICCC, thus allowing the basic units and even their own referral hospital had an adequate knowledge of this reality.

In this proposal, the study brought the application of spatial statistical techniques to some local results, which can be used in information from other locations. Thus, characteristic features of other sites and portraying other communities can be mapped. With this, a greater number of people involved with the issue, there will be wide dissemination of knowledge of the applicability of these techniques, including favoring the approach of tumors outside the level of high complexity of
healthcare.

In conclusion, incidences of major diagnostic groups of cancers in children and teenagers in London were similar to those of previous reports from other cancer records16,20,35. Studies on cancer rates and spatial patterns effects serve as reference to the development of benchmarks to formulate of health care network, the provision of health services and will contribute to future improvements and health planning and decision support.

REFERENCES


