Spatial epidemiology and pattern analysis of childhood cancers in Tehran, Iran

Hassan Amini¹, Morteza Seifi², Sadegh Niazi-Esfyani², Masud Yunesian²

¹ Kurdistan Environmental Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran
² Department of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

Abstract
Identification of cancer clusters may have an important value to the study of disease etiology in cancer surveillance. We aimed to determine the spatial pattern of childhood cancer cases (CCCs) from 2007 to 2009 in Tehran, Iran. Records of 176 childhood cancer counts (children younger than 15 years old) for 2007-2009 were obtained from Iran’s Ministry of Health and Medical Education. Thereafter, they were successfully geo-coded within a geographic information system based on their residence phone number or postal code. We used two distinct techniques, namely average nearest neighbor index (ANNI) and Quadrat analyses, to measure the spatial pattern of CCCs in Tehran. The count of childhood cancers for 2007-2009 in Tehran was 117.3 per 1,000,000 children. The ANNI analysis suggested that there was a clustered pattern for the CCCs in 2007-2009. There was less than 1% likelihood that this pattern could be the result of random chance (nearest neighbor ratio = 0.73; Z-score = -6.8 standard deviations; P < 0.01). In the Quadrat analysis, the largest absolute difference between the observed and expected cumulative proportions in the frequency table was 0.2778 while the critical value of Kolmogorov-Smirnov test was 0.1649. Therefore, the Quadrat analysis confirmed that the CCCs had clustered pattern in 2007-2009 in Tehran. Both used methods suggested that childhood cancers in Tehran had clustered pattern in 2007 and 2009. We believe further research is needed to study the etiological factors, especially environmental factors, which made this cluster.

Keywords: Childhood Cancers, Cluster Analysis, Geographic Information System, Iran, Medical Geography, Spatial Epidemiology, Tehran

Date of submission: 15 Sep 2013, Date of acceptance: 29 Nov 2013


Introduction
Cancer is one of the leading causes of mortality and disease burden, globally.¹-³ According to the latest global burden of disease (GBD) study,⁴ leukemia ranks 12th amongst the top 25 causes of death for children 5-14 years old (see the heat map from permanent online link).⁵ For Western Europe region of the GBD, brain cancer, leukemia, non-Hodgkin’s lymphoma, liver and kidney cancers rank, respectively, 2nd, 3rd, 10th, 22nd, and 25th amongst the top 25 causes of death for the same age group (see permanent online link).⁶ In the North Africa and Middle East region, leukemia and brain cancers rank 6th and 11th, respectively (see permanent online link).⁷ In addition, the GBD 2010 study reported that
leukemia and brain cancers rank 4th and 10th amongst the top 25 causes of death for children 5-14 years old in Iran, respectively (see the heat map from permanent online link).8

The etiology of childhood cancers is largely unknown.9 Although genetic factors play an important role in the development of childhood cancers, the role of environmental factors, infectious agents, maternal cigarette smoking and alcohol consumption, and socioeconomic status are also substantial.10-13 So far, few environmental risk factors, such as exposure to air pollutants and radiation have been suggested to be associated with the development of childhood cancers.13,15 However, these risk factors describe only a small proportion of childhood cancers. Nevertheless, environmental risk factors play an important role in the incidences of childhood cancers.13,14,16

In cancer surveillance, identification of areas with increased risk, mapping, and detection of cancer clusters may have an important value to the study of disease etiology.17-21 There have been numerous studies to date that reported geographic and spatiotemporal variations in variety of cancers, such as childhood cancers.22-29 A more recent study from Tehran, Iran, also reported clusters of childhood cancers within diverse administrative areas of this megacity.30

Currently, many methods have been developed and introduced to detect clusters in spatial point patterns.31,32 These include average nearest neighbor index (ANNI) analysis, Cuzick and Edwards’ method, gradient analysis, K function or Ripley’s function, kernel intensity function, Kulldorff’s scan statistic, mean center standard distance, Quadrat analysis, standard deviational ellipse, variance mean ratio test, etc.33-36 Each of these methods has its specific advantages and limitations, which are described elsewhere.18,21

In this study, the authors aimed to determine the spatial pattern and probable clusters of childhood cancer cases (CCCs) in the Middle Eastern city of Tehran, Iran from 2007 to 2009 using ANNI and Quadrat analyses.

**Materials and Methods**

Tehran is a large metropolis with a population of more than 8.2 million residents. It is noteworthy that based on latest census report (2011), more than 36% of Tehran’s population is less than 25-year old and the population of childhoods (children younger than 15 years old) is about 1.5 million children (19%). The city is located in a large area, roughly 613 km² and has diverse configuration. The altitude in the southern parts of the city is about 1000 m above the sea level and in the northern parts of the city is approximately 3800 m (Figure 1).37-40

Records of 176 childhood cancer counts (children younger than 15 years old) for 2007-2009 were obtained from Non- Communicable Disease Center of Iran’s Ministry of Health and Medical Education (MOHME). Thereafter, they were successfully georeferenced within Tehran megacity by Geographic Information System Bureau of Iran’s Post Office based on their postal code.

We used two methods, namely ANNI and Quadrat analyses within geographic information systems (GIS) to measure the spatial pattern and clustering of CCCs as discrete point features from 2007 to 2009 in Tehran.18,41

In the ANNI analysis, we calculated the distance between each CCC and the nearest CCC. Thereafter, we calculated the average distance between CCCs and compared with the distance of random cases within the study area. If the observed average distance of the CCCs was smaller than expected average distance of random distribution (nearest neighbor ratio), we considered the CCCs to have cluster pattern. Statistically, the null hypothesis says that the CCCs are randomly distributed. If the CCCs have a spatially random distribution, then the NNI would have a normal distribution. Thereafter, we can calculate the Z-score and compare the distributions. If the Z-score was positive, it indicates that the pattern is dispersed;
if it was negative, then the pattern is clustered. Meanwhile, at a confidence interval of 95%, the Z-score > 1.96 or < -1.96 would be considered statistically significant.42,43

In this analysis, we have overlaid the areas of equal size on the study area—known as quadrats or quads—and counted the number of CCCs in each square quadrat. One of the most important points in this analysis is the size of the used quadrats. Traditionally, the size of quadrats is identified as twice the size of the mean area per feature.44 Thus, we have applied the following equation to calculate the length of a side of each quadrat:

\[ L = \sqrt{\left\{ 2 \times \left( \frac{A}{N} \right) \right\}} \]

Where the \( L \) denotes the length of a side of each quadrat, \( A \) denotes the area of extend for CCCs, and \( N \) denoted the total observed number of CCCs.

Thereafter, we have calculated the expected counts of CCCs for a random distribution in the study area based on Poisson distribution. We first calculated the probability of the \( x \) number of CCCs occurring in any given quadrat or \( P(x) \) as the following equation:

\[ P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \]

where \( e \) is the Euler’s constant, \( X \) is the observed number of CCCs, and \( \lambda \) is the average number of CCCs per quadrat. In addition, the \( \lambda \) was calculated based on the following equation:

\[ \lambda = \frac{n}{k} \]

where \( n \) denotes the total number of CCCs, and \( k \) denotes the total number of quadrats.

We then multiplied the \( P(x) \) results by the total number of CCCs to get the number of quadrats expected to contain that number of CCCs. Therefore, we created two frequency tables; one for observed distribution, observed proportion, and observed cumulative proportion; and the other one for expected distribution, expected proportion, and expected cumulative proportion based on Poisson distribution. In quadrat analysis, if more square quadrats contain few or no CCCs than expected, and fewer square quadrats contain most of the CCCs than expected, the CCCs form a clustered
pattern. The most common statistical test that has been used to find out how much the two frequency tables and patterns differ is Kolmogorov-Smirnov test. In summary, we calculated the absolute difference between observed and expected cumulative proportions for each line in the frequency tables. Then, we found the largest absolute difference between these and compared this value with critical value of Kolmogorov-Smirnov test, which was calculated as the following equation:

\[ C = \frac{Z_\alpha}{\sqrt{n}} \]

where \( C \) is the critical value of Kolmogorov-Smirnov test, \( n \) is the number of quadrats, and \( Z_\alpha \) is the constant for a confidence interval. We have applied the confidence interval of 0.05 and therefore \( Z_\alpha = 1.36 \) as it is frequently used (Table 1).

We compared the largest absolute difference of the observed and expected cumulative proportions with the critical value of Kolmogorov-Smirnov test. If the largest absolute difference was greater than the critical value, we considered the difference statistically significant and rejected the null hypothesis that CCCs have random distribution over the study area.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>( Z_\alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2000</td>
<td>1.072749</td>
</tr>
<tr>
<td>0.1000</td>
<td>1.223848</td>
</tr>
<tr>
<td>0.0500</td>
<td>1.358099</td>
</tr>
<tr>
<td>0.0200</td>
<td>1.517427</td>
</tr>
<tr>
<td>0.0100</td>
<td>1.627624</td>
</tr>
<tr>
<td>0.0010</td>
<td>1.949475</td>
</tr>
<tr>
<td>0.0001</td>
<td>2.225251</td>
</tr>
</tbody>
</table>

Results and Discussion

The count of childhood cancers for 2007-2009 in Tehran was 117.3/1,000,000 children younger than 15 years old. Figure 2 illustrates the spatial distribution of CCCs for 2007-2009 in Tehran.

Results of ANNI analysis

There was a clustered pattern for the CCCs in 2007-2009. There was < 1% likelihood that this pattern be by random chance (nearest neighbor ratio = 0.73; Z-score = –6.8 standard deviations; \( P < 0.01 \)).
Results of Quadrat analysis
The length of the quads for 2007-2009 was 3396 m. For 2007-2009, the largest absolute difference between observed and expected cumulative proportions in the frequency table was 0.2778 while the critical value of Kolmogorov-Smirnov test was 0.1649.

In this study, the authors tried to map and analyze the spatial pattern of childhood cancer counts in Tehran, Iran, for a 3-year period from 2007 to 2009. We used two distinct techniques, namely ANNI and Quadrat analyses, to identify the probable clusters of CCCs, and to verify that the results are reliable.

Studies of the disease mapping especially cluster analysis of childhood cancers are not rare and there are numerous studies globally. However, the majority of studies on clustering of childhood cancers investigated the leukemia, as it represents about 25% of childhood cancers. Unfortunately, in case of our study, the MOHME provided us the data of childhood cancers without detail information on the type of registered cancers, but it seems that the majority of cases be leukemia as the prevalence of leukemia is more than other childhood cancers in Iran based on the GBD 2010 study. One of the famous studies related to the cluster analysis of childhood cancers is EUROCLUS project. The EUROCLUS is a European collaborative study to identify spatial clustering of cancer cases and to detect whether these clusters could be explained by environmental risk factors. This study analyzed spatial pattern of 13,351 childhood leukemia cases diagnosed between 1980 and 1989 in 17 countries. The EUROCLUS revealed that childhood leukemia cases had statistically significant clusters within small census areas. However, the study also indicated that the phenomenon of intense clusters is rare and needs careful surveys. Bellec et al. analyzed spatiotemporal pattern of childhood acute leukemia cases for two periods from 1990 to 1994 and from 1995 to 2000 for all the French territory. Indeed, they have found a statistically significant cluster pattern in the incidence of acute leukemia cases over 1990-1994, but neither over 1995-2000 nor over the whole period from 1990 to 2000. Gustafsson and Carstensen also reported clustering of 1020 childhood acute lymphatic leukemia and malignant non-Hodgkin’s lymphoma for Sweden from 1973 to 1996. They have found a statistically significant excess of case-pairs (25 observed, 14.9 expected, P = 0.01) in the 4-14 year age group for acute lymphatic leukemia. However, they have found no statistically significant clustering when the cases of leukemia, and the non-Hodgkin’s lymphomas were assembled. Noteworthy, we have found statistically significant clusters for combination of all CCCs in Tehran for 2007 and 2008, but not in 2009. Knox and Gilman (1996) analyzed data of (a) all childhood leukemia and non-Hodgkin lymphoma cases registered between 1966 and 1983 in England and Wales based on enumeration districts, and (b), all childhood leukemia and cancer deaths between 1953 and 1980, in England, Wales, and Scotland based on their post codes. They have found short range spatial clustering for both leukemia at a place of registration, and leukemia and cancer (separately and jointly) at both birth and death addresses. Wheeler used several cluster analyzes techniques in Ohio, USA, to identify clusters of childhood leukemia incidences from 1996 to 2003. Wheeler found some evidence of significant local clusters in childhood leukemia using kernel intensity function, but no significant overall clustering by other used methods. Alexander et al. analyzed data of childhood leukemia incidences during 1984-1990 for evidence of variation between small areas in Hong Kong and reported that there was evidence of spatial clustering for acute lymphoblastic leukemia at ages 0-4 year. On the other hand, Dockerty et al. analyzed spatial clustering of childhood leukemia and lymphoma cases in New Zealand and found little evidence for spatial clustering. In fact, they have reported
significant clustering in a specific subgroup of acute lymphoblastic leukemia aged 10-14 years, which believed may have been real or a chance association. Schmiedel et al. worked on spatial clustering of 1168 childhood leukemia cases diagnosed during 1980-2006 in Denmark and have found spatial clustering at time of diagnosis for children aged 2-6 years (observed/expected [95% confidence interval]: 1.35 [1.15-1.54]). Schmiedel et al. also worked on another study in Germany. They have obtained data of 1946 children with leukemia diagnosed during 1987-2007 from the German Childhood Cancer Registry and tried to identify probable clusters. As a result, they have declared no evidence of a tendency to clustering neither for the whole study population nor in the sub-groups.

In case of Tehran, we have found only one study that has tried to map and analyze the spatial pattern of childhood cancers during 1998-2002. In fact, Mosavi-Jarrahi et al. reported marginally statistically significant clustering (P = 0.056, RR = 1.3) for all childhood cancers in Tehran; however, they have found no statistically significant clusters for cancer categories.

One of the limitations of this study is that we have obtained the data of CCCs only from MOHME, while there may be some CCCs that have not been registered by MOHME. Meanwhile, the Quadrat analysis usually has been applied to evaluate point patterns for clustering based on density of features as it counts the number of features per unit area. In fact, it does not consider the arrangement of points and their proximity to each other. Therefore, we have tried to use another method—ANNI—that analyze cluster patterns based on the proximity of features. Both used methods are also biased by the fact that they assume the population is homogenous over the study area.

Conclusion

In our study, both methods suggested that childhood cancers in Tehran had clustered pattern. We believe further research in the future is needed to study the etiological factors, especially environmental factors, which made this clusters.

Conflict of Interests

Authors have no conflict of interests.

Acknowledgements

The authors extend their sincerest gratitude to two reviewers for their valuable constructive comments. We also thank the Non-Communicable Diseases Center of Iran’s Ministry of Health and Medical Education (MOHME) for providing us the data.

References


