Clustering of childhood cancer in the inner city of Tehran metropolitan area: A GIS-based analysis

Alireza Mosavi-Jarrahia,c,*, Mehdi Moinib, Mohammad-Aali Mohagheghic, Mardawich Alebouyehd, Bahareh Yazdizadehc, Akbar Shahabianb, Azin Nahvijo, Reza Alizadehb

Abstract

The aim of this study was both to map the childhood cancer incidence in the districts of Tehran metropolitan area and to explore possible clustering of cancer cases in the diverse environments of this area. All incidence cases of childhood cancers (age group under 15 years) belonging to the 22 districts of Tehran metropolis and occurring during the period of 1998 till 2002 were ascertained from three sources. Each case’s place of residency was geo-referenced. The scan statistics cluster detecting technique was used to evaluate clustering of cases throughout Tehran. The overall incidence rate (IR) of childhood cancer was 176.3/1,000,000 children under 15 years of age. The lowest IR among both boys and girls was observed in district 22 (69.4/1,000,000) and the highest was observed in district 6 (242.09/1,000,000). The detection of clusters was performed for all cancer sites. All the cancer sites combined category showed clustering in the districts 7, 13, 8, 6, 3, 14, 12, 11, and 4. For this category, the clustering likelihood was marginally statistically significant (p-value = 0.056), with an overall relative risk of 1.30. No statistically significant patterns of clustering were detected for other categories.

Keywords: Iran; Childhood cancer; Cluster; Tehran

Background

Cancer is one of the major causes of childhood death in the developed and developing countries. It accounts for 4% of the death of children under 5 years of age and 13% for children 5–15 years of age in the Iranian population; contributing to 15% of total loss of life in the under 15-year age group (Naghavi, 1999).
from a small percentage of childhood cancers that are genetically determined, environmental factors also play an important role in the aetiology of childhood cancers (Little, 1999). It is anticipated that environmental hazards cause childhood cancer through direct exposure or parental exposure prenatally or perinatally (McKenzie, et al., 1998; Petruzelli et al., 1998). The search for clusters of cancer cases is one approach when studying the epidemiology of childhood cancer. Studies of cases in highly urbanized areas such as Liverpool, Greater London, Buffalo City, San Francisco and Metropolitan Atlanta have suggested clustering of childhood cancer in both space and time (Alexander et al., 1998; Little, 1999). Recent studies of space–time clustering have suggested clustering for lymphoblastic leukaemia (McNally et al., 2002) (precursor B-cell sub-type), Wilms’ tumour, and soft tissue sarcoma (McNally et al., 2003; Parodi, et al., 2003).

Tehran is a large metropolis with a population of 6,758,840 (based on 1995 census report) living in an area of 1500 km². The population is young (more than 50% of the population is under 25 years of age) and no incidence of childhood cancers have been estimated or reported for this population yet. The residing population is exposed to different environmental hazards ranging from a neighbouring dye factory to air polluted with lead and benzene (Halek et al., 2004). A recent study of school-aged children throughout the Tehran metropolitan area showed that the abnormal pattern of immunity index (measured as CD4/CD8 ratio) significantly correlates with the concentration of benzene in the air in the different residential parts of Tehran (Azari et al., 2005). The aim of this study was both to map the childhood cancer incidence in the municipality districts of Tehran metropolitan area and also to explore the possible clustering of cancer cases in the diverse environments of Tehran.

Methods

All incidence cases of cancer for the age group under 15 years belonging to the population residing in 22 districts (the Tehran municipality is dived into 22 districts administratively and a serial number of 1–22 has been assigned as district number for each district) of the Tehran municipality from 1998 till 2002 were included in the study. The cases were ascertained from three sources to assure inclusion of all cases: (1) the Tehran Population-Based Cancer Registry (TPBCR) database which includes all cases of cancer referred to the treatment and diagnostic facilities throughout the Tehran metropolitan from 1998 to 2002, (2) the database of the Mahak organization (a charity organization providing treatment and social support for children with cancer and maintains databases of served patients since 1980), and (3) the mortality database of Tehran municipality for the period of 1998–2002. For the purpose of this study, an eligible case was defined as a case in which cancer had been verified by histopathology, or had been registered in the TPBCR (an eligible case to the registry), or the cause of death was reported as cancer and the patient or parents of the patient had lived in the Tehran metropolitan area for a period of at least 5 years prior to diagnosis.

Case identifying procedure

The three databases were used to ascertain the cases. A sensitive strategy for linkage, developed by the TPBCR Office (Mosavi-Jarrahi et al., 2001), was used to eliminate duplicated and repeatedly reported cases.

Procedure to verify the residency of the cases

Two of the databases (the TPBCR database and the Mahak database) used in the study did not include residency addresses, but it did have a contact address or telephone number. The databases also included cases that did not belong to the Tehran metropolis’ population but had a contact address in Tehran. To exclude such cases from the study, the residential status of all cases identified through the identifying procedure were verified by one or combination of the following procedures:

1. Telephone calls: this was applicable to all the cases that had a telephone number in the databases. Two telephone call attempts were made (one call during the day and the other if there was no answer; during the weekends). If both attempts failed, the second procedure was used. In the telephone calls, each case’s complete addresses and duration of residency was verified.

2. Mailing a questionnaire to the patient: for patients who had no telephone number available or if the calls were not successful, a letter asking for the complete address and duration of residency were sent to the patient. Thus, the contact address and the residency status were verified.

3. Sending a dispatcher: in such cases when the telephone call and mailing procedure were not successful or the post office returned the mail, a dispatcher was sent to the patient’s contact address or to the office of the physician treating the patient to verify the residency address from his records.
Procedure to locate the cases in the Tehran municipality map

The verified residency address for each case was sent to the Tehran Geographic Information System (TGIS) office and two trained persons were assigned to locate and develop the coordinate of the location of each case on the Tehran map.

Data analysis

Data analysis was done in two levels (1) descriptive and (2) analytical. On the descriptive level, sex-specific incidence rates (IRs) were calculated for all 22 districts. For the purpose of estimating incidences, all cases mapped to each district were considered as a numerator and the number of at-risk population (age group less than 15 years of age) for each district during the study period (1998–2002), were considered as denominator.

The number of at-risk population for each district was derived from the census report of the National Bureau of Statistics, Iran. The incidence was estimated for cases in the age group less than 15 years. On the analytical level, the possibility of clustering of cases was tested using the “scan statistic cluster detecting” technique. This technique was developed by Turnbull et al. (Kulldorff, et al., 1997) in 1999 and further modified by Kulldorff and Nagarwalla (1995). The technique can control the covariate affecting the distribution of cases.

Analysis was done using the Sat Scan™ software (Kulldorff, 2004), a public domain software written by Martin Kulldorff, used in cluster analysis of cancer by the National Cancer Institute, USA (Kulldorff and Nagarwalla, 1995). Briefly, Sat Scan™ imposes a circular scan window centred on each of several possible centroids positioned throughout the study region. For each scan window, a likelihood ratio test is calculated to test the hypothesis that there is an elevated rate of incidence in the scan window compared to the distribution outside the window. The software needs three electronic files to run cluster detection analysis: (1) the case file that consists of the information about the cases and their location – this file was developed through case identifying procedures, (2) the population file consisting of the population distribution (based on sex and age) of the 22 districts of Tehran metropolis – this file was constructed using population information obtained from the National Bureau of Statistics, and (3) the coordinate file containing the geographic coordinate information of the districts in the Tehran municipality map – this file was generated by the TGIS office and was customized to the need of the SatScan™ software.

For the purpose of cluster analysis, cancer sites were grouped into six categories: (1) all cancer sites combined, (2) lymphomas and other reticuloendothelial neoplasms, (3) CNS and intracranial/intraspinal neoplasms and sympathetic and allied nervous system tumours, (4) retinoblastoma (5) bone and soft tissue sarcoma, and (6) leukaemias. The possibility of clustering was tested for all categories.

Results

Eligible cases in the database

The cancer registry database had registered 5947 cases under 20 years of age (including mortality from cancer). The Mahak data bank included 7000 cases of childhood cancer from the beginning of its activity (1980) till 2003. The Mahak database provided 1049 cases as eligible cases, other cases either did not belong to the Tehran area, or the incidence date was before 1998. The pool of the two databases provided 6996 cases for analysis. Further analysis of the data eliminated 3980 cases as repeatedly reported cases, leaving 3016 cases for further verification of residency and geo-referencing.

Address verification and geo-referencing

Out of 3016 patients, 2412 had either an address or a telephone number written in their record. All attempted calls (a total of 660 successful calls were made), mailing enquiry or sending a dispatcher provided 2343 eligible cases belonging to the 22 districts of the Tehran municipality. Out of these, 874 cases were in the age category over 15 years and 369 cases were not geo-referenced in the Tehran municipality office due to the fact that details of their addresses were not in the GIS address layer or it was located out of the legitimate boundary of the 22 districts of the Tehran municipality, leaving a total of 1133 cases for further analysis.

Results for the descriptive analysis

The yearly average IR of childhood cancer in the whole Tehran was 176.3 cases/1,000,000 children under 15 years of age in the study period (1998–2002). The lowest IR among both male and female were observed in district 21 (69.4/1,000,000) and the highest was observed in district 6 (242.09/1,000,000; see Table 1). The geographic distribution of cases differed based on gender for different districts (in boys, the highest IR was mapped to district 12 and in girls, the highest IR was mapped to district 7). Figs. 1 and 2 show the spatial distribution of cancer cases based on gender for different districts of Tehran municipality.
Results for the analytical analysis

The detection of clusters was performed for five categories of cancer. All cancer sites combined category showed clustering in the districts 7, 13, 8, 6, 3, 14, 12, 11, and 4. For this category, the likelihood of clustering was marginally statistically significant (p-value $< 0.056$) with an overall relative risk of 1.30 (see Table 2). Other categories of cancer did not show a statistically significant pattern of clustering. Table 2 shows the result of cluster detection for each category of cancer sites. In terms of the magnitude of the relative risk of accommodating in the cluster; retinoblastoma with a relative risk of almost 5 and leukaemia with a relative risk of 1.7, showed a likelihood of clustering at the districts mainly located in the inner part of the city, however, the p-value was not significant (Table 2).

Discussion

The descriptive and analytical epidemiology of childhood cancer has been extensively studied in developed countries and less extensively in developing countries. This is despite the fact that in developing countries that have more of younger population, there is a higher proportion of the population at risk of childhood cancer.
resulting in higher burdens for these countries. The estimated incidence of childhood cancer over the different geographic boundaries of the world shows wide variations in magnitude. Our study demonstrated that childhood cancer in Iran have a moderate incidence (176 cases/1,000,000 children under 15 years of age). In countries where reliable registries have collected data, the incidence of childhood cancer ranges between 64.7/1,000,000 (India, Karunagappally registry – 1991–1992) to about 300 cases/1000,000 (Canada, Yukon Registry – 1983–1992) in children under 15 years of age (Parkin et al., 1997). Along with the geographic variations seen, a recent study in Europe has demonstrated a rise of 1% per year in the IR of childhood cancer (Steliarova-Foucher et al., 2004).

The wide variations encountered in the incidence of childhood cancer are attributed to the genetic background and environmental exposures of parents and children to different carcinogens and environmental hazards. Clustering of cases and localized excesses of proximity to environmental sources of hazards have provided strong means to develop hypotheses in addressing the aetiology of childhood cancer. Spatial clustering of leukaemia has been behind the hypothesis that an infectious agent may cause leukaemia (Kinlen, 1998, 2004). One of the first clusters of childhood cancer (leukaemia) was seen in villages around the nuclear re-processing plant at Sellafield, in the north of England. Further epidemiologic investigation indicated that the excess might be related to occupational exposure of
fathers before the conception of their children (Gardner et al., 1990). Most of the works on clustering of childhood cancer are specifically related to leukaemia. Reviews of published cluster studies of leukaemia (Little, 1999) have shown that out of 24 studies, 15 studies were interpreted by the authors as showing some evidence of clustering. Similarly, our study demonstrated statistical evidence of clustering childhood cancer in certain districts of Tehran. The districts located in the observed cluster are located in the inner city areas of Tehran where the older part of the city with higher population density is located and environmental hazards (from traffic exhaust to air pollution) are more rampant compared to other districts located off the central part of the city. In addition, there are more commercial activities (small workshops such as furniture, shoes, and leather processing factories as well as offices such as banks and other administrative bodies) in the districts included in the cluster. One of the main contributing factors to environmental pollution in Tehran is exhaust gases released into the air from the automobiles (Masjedi et al., 2003). Environmental studies done in the Tehran population show disturbances in the immunity indexes of school children that are correlated with pollution distribution in the city (Azari et al., 2005) and an increase in the number of nucleated red blood cells of neonates born in the central part of the city (Ziaei et al., 2005), where environmental hazards are more rampant. The urban environmental pollution has already been related to DNA damages in other populations (Whyatt et al., 1998). From an ecological prospect, the distribution of area-related environmental hazards could contribute to the differences in IRs seen in our analysis as well as clustering of cases in the inner part of the city.

In our study, other categories of cancer sites did not show statistical evidence of clustering. However, in certain categories such as leukaemia with a relative risk of 1.7 and retinoblastoma with a relative risk of almost 5, clustering of cases was observed. Again, accommodation of certain districts such as district 6, 12, and 11 (districts in the inner part of the city) in the retinoblastoma and leukaemia clusters, may be indicative of a stronger contribution of environmental factors to the aetiology of childhood cancer in this population. A lack of statistical significance for the clustering of leukaemia and especially retinoblastoma cases shows both the low powers of our study as well as the need for utilizing other methods than mere scan statistics techniques of cluster investigation. To address the deficiencies in different cluster detecting techniques, it has been recommended that a battery of spatial pattern methods be employed to better describe the different aspects of the geographical patterns in cancer incidence (Jacquez and Greiling, 2003).

The result of our study has to be interpreted under three aspects: (1) completeness of case ascertainment, (2) quality of data available for the study, and (3) inherent limitation of ecological and GIS studies. The completeness of ascertained cases were strengthened by the fact that we collected data from three major sources. While using different databases may not guarantee completeness, it is, however, one of the legitimate means of data quality control in disease registries (Parkin et al., 1998). The quality of data that had been made available to our study has been evaluated in different contexts (Mohagheghi and Mosavi-Jarrahi, 2006). Childhood cancer is, in fact, a rare disease and ascertaining a large numbers of cases needs a long period of time and a large population. The small numbers of cases in our study limits the explanatory capacity of our results. The inherent limitation of ecological studies and especially the modelling of geographical distribution of diseases have been addressed intensively (Jacquez, 2004). In our study, we used spatial scan statistic methods because it offered advantages not provided by all of the other methods: identifying clusters of any size located anywhere within the study area while controlling for multiple hypotheses. The scan statistic cluster detection technique uses a circle as the basic shape for detecting clusters. However, a cluster may follow the presence of hazards in the environment, and that needs not necessarily be in circular shape. For exploratory analysis (as in our study) instead of a confirmatory hypothesis testing, a circular shape may be more efficient when compared to a non-circular shape. However, the fact that the districts studied are irregularly shaped may be in contrary to the circular shape argument and can be considered a limitation for this study. Another limitation of our study is the fact that only 22 polygons (districts) were analyzed. The use of a limited number of polygons has increased the chances of an aggregation bias (bias that resulted from the arbitrary aggregation of cases at districts as geographical units) in our study. The limitations on the number of polygons and the irregularity of district shape may have contributed to the lack of statistical power to detect clusters for leukaemia and retinoblastoma in this study.

Conclusions

Our study showed some evidences of clustering of childhood cancer in the inner city of Tehran and that further studies are needed to address the aetiological factors contributing to this clustering.
Acknowledgements

This investigation received financial support from WHO/EMRO (TSA no. 03/79).

The authors need to thank Mr. Mohammad Golmahi, Fatemeh Golmahi, Somayeh Hajizadeh, and other people from the Cancer Research Centre of the Cancer Institute of Iran, because without their help the study would have not been able to achieve its goals.

References


Naghavi, M., 1999. Mortality in Iran, reports of four provinces, Ministry of Health and Medical Education, Office of Research and Development, Tehran, Iran.


