Swedish Science Pioneers
Developing World Journal Series
Journal of Clinical Research & Governance
www.jcrg.sciencepub.se

Short Communication

A Review of Spatial Analysis Procedures: An Application for Congenital Anomalies Registries

Amin Bateni[a], Saeed Dastgiri[b], Mohammad Mosaferi[c], Mahammad Heidarzadeh[d]

a: Department of Community Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

b: Tabriz Health Services Management Research Centre, School of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

c: Tabriz Health Services Management Research Centre, Tabriz University of Medical Sciences, Tabriz, Iran

d: Pediatric Health Research Centre, Tabriz University of Medical Sciences, Tabriz, Iran

Correspondence

Professor Saeed Dastgiri
Tabriz Health Services Management Research Centre
Department of Community and Family Medicine
School of Medicine
Tabriz University of Medical Sciences
Tabriz, Post Code: 5166615739 Iran
Email: dastgiris@tbzmed.ac.ir

Keywords:

Spatial Analysis
Congenital Anomalies
Birth Defects

Received: 2013-03-31
Accepted: 2013-06-26

DOI:10.13183/jcrg.v2i2.23

Abstract

Purpose: Despite the potential role of spatial analysis in the investigation of the etiology of birth defects, there are relatively few published data on this field. The aim of this article is to discuss the main applications of spatial analysis in particular hot spot analysis in the registries of congenital anomalies.

Methods: Based on spatial characteristics of case locations which captured in congenital anomalies registries(point or regional level) and also type of determination of clustering(global or local clustering), there are different statistical procedures to detect hot spots. These procedures were reviewed briefly to determine which is suitable for hot spot analysis in TRoCA registry.

Results: The location of congenital anomalies cases are captured in regional level on 19 area in this province. Considering the data set and different procedures for hot spot analysis, it is not suitable to apply existing methods in TRoCA registry but choroplete map based on spatial variations can be depicted.

Conclusion: Large international registries of congenital anomalies can perform multi-level area analysis (i.e., city based, in hospitals, regions, etc) while small registries will have to improve their quality of data and expand the area and population for which the data collected enabling them for more reliable spatial analysis.

Introduction

The function of epidemiology is to evaluate the distribution of diseases in a defined population and to determine the etiology and predisposing factors influencing the occurrence of diseases. Three main characteristics of epidemiologic investigations include "Time", "Place" and "Person". Exploring the role of "Place" on the occurrence of diseases is the main aspect of "Spatial Analysis". The first application of spatial analysis began in 1854 when Jon Snow used spot maps as the locations of cholera cases in London for investigation of cholera epidemic. He concluded that the water sources, as a place factor, might have caused the cholera outbreak. The findings were then implemented in the control of the epidemic in London.

The application of spatial distinctive has been thought over by health policy makers because by pointing out spatial distribution, it is possible to prevent the disease dissemination.

In average 1 out of 33 infants bear birth defect related disabilities. Although many factors like genetic factors, infections, maternal nutritional status, socioeconomic factors and limited environmental factors are known as the risk factors or predisposing factors for occurrence of congenital anomalies but nearly 50% of all congenital anomalies etiological factors are obscure. In some cases the interaction between multiple factors has been notified [1].

Despite the potential role of spatial analysis in the investigation of the etiology of birth defects, there are relatively few published data on this field. The aim of this article is to discuss the main applications of spatial analysis in the registries of congenital anomalies. Advantages and limitations of spatial analysis will also be discussed using the data from Tabriz Registry of Congenital Anomalies (TRoCA).

Methods

Some softwares have been designed for spatial analysis including Arc GIS, cluster Py, R-analysis of spatial data, GeoDa and EpiMap. The first step in spatial investigation is the mapping of spatial data which visualizes the differences in the

study area. Then by using statistical procedures, it is possible to determine the disease clusters or hot spots in the field.

Mapping Spatial Data

Each software used in spatial analysis has its specific definition for the points taken as "Place" in mapping the diseases distribution. For instance, two main types of spatial data in Arc GIS are "vector" and "raster". Vector data consists of three types of features area (polygon), line and points. Raster data consists of cells for different values based on fuzzy theory [2]. The values may consist of any type of data including interval, ordinal and nominal. Some softwares create hazard maps based on the various data provided. Hazard maps display the distribution of risk/influencing factors together with the prevalence of diseases (i.e., congenital anomalies). However this does not necessarily indicate the relation between the occurrence of birth defects and environmental factors because of many confounding sources in between.

Point maps, distribution maps, proportional maps and choropletic maps are the most popular tools for visualizing the spatial data in public health [3,4]:

(a) ***Points (Dot or Location) Map:*** in this type of map, birth defect geographical locations are marked by different symbols in the study area. They are not however very accurate. Some softwares can determine the direction of the dissemination of disease using this map too.

(b) ***Distribution Map:*** the position of the study characteristics (i.e., location of the incident cases of congenital anomalies) is displayed by colors or shading of colors.

(c) ***Choropleth Map:*** in this map, different polygons depending on the exact quantities of the study characteristics are displayed with various colors. Choropleth maps are of most common ones used in spatial analysis.

(d) ***Proportional Map:*** in a proportional map, the data are displayed by symbols of different sizes to represent data associated with different areas or the severity of diseases, etc.

Spatial Clustering

European Surveillance of Congenital Anomalies (EUROCAT) defines a cluster as: “An aggregation of cases of congenital anomaly in time and/or space which appears to be unusual" [5]. Based on the input information in a registry, we may have “Point” or “Regional” data for the analysis of spatial. It is not always possible to get access to the exact “Point” (address) of the occurrence of a congenital anomaly because of the ethical and confidentially issues. We therefore have to rely on “Regional” data instead. There are two main types of cluster analysis for “Regional” data [6]:

(A)Global or Non-specific Clustering**:** the researcher tries to discover whether any cluster has occurred in the area, but the exact location of the cluster may not be detected.

(B)Local or Specific Clustering**:** in this approach, the exact location and extent of clusters are determined. This is also defined as “Cluster Detection Method” with two sub categories of “Focused” and “Non-focused” clusters depending on whether any pre-determined point exists.

As for some epidemiological studies, “Ecological Fallacy” must be considered in the analysis of regional clustering data as a source of confounding. The problem often occurs when the expected event is rare.

Statistical Tests

There are two main statistical tests for investigating the global clustering in regional data: **“**Geary’s C**”** and **“**Global Moran's I index**”.** The details of the application of these tests can be found elsewhere [7].

For investigation the local clustering in regional data, there are two tests available including **“**Getis and Ord's local Gi (d) Statistic**”** and **“**Local Moran Test**”** [8,9].

If “Point” data were available in the birth defects registry, the most common statistics used for spatial analysis of global clustering include:

(A)Cuzick and Edward's K-nearest Neighbor Test: this method is based on the location of cases and randomly selected control subjects for each defect [10].

(B)Ripley's k-function: using this, we can determine the distance at which clustering occurs.

These are statistical tests for local clustering in the analysis of spatial:

Openshaw's Geographical Analysis Machine

The GAM applies various radii with their centers based on each grid data to detect clusters. Openshaw used this method to investigate acute lymphoblast leukemia in Northern England [11].

Turnbull's cluster evaluation permutation procedure (CEPP)

This method can either locate disease cluster or either test the significance of clustering [12].

Besag and Newell's Method

In spite of GAM in this method the investigator can specifies GAM in this method the investigator can specifies, the expected cluster size to dissolve the different – sized populations problem [6].

Kulldorff’’s spatial scan statistics

Kulldorff developed a test that includes the both advantages of Openshaw's GAM and Turnbull's CEEP in which a series of circles with various radii are assembled, for each case [13].

Non-parametric Spatial Scan Statistics

Most of previous tests assume that disease clusters are circular. These methods were developed to detect various shaped disease clusters.

More statistical tests are available for different types of spatial analysis (i.e., space-time clustering, etc) [6,14].

Results

There has been no data available on the prevalence, etiology and preventive strategies of congenital anomalies in Iran until recent years. A research study was carried out in 2000 on the occurrence of birth defects aiming to document the epidemiologic features of congenital anomalies in the northwest of the country as the baseline information to set up a regional registry of birth defects in the region. It is now called Tabriz Registry of Congenital Anomalies (TRoCA). The principal aims of TRoCA are to establish a monitoring system of congenital anomalies in the area, and to implement control and preventive tasks in the region. More details on this program can be found in our website [15]. TRoCA is a hospital-based registry covering all births and children in a defined geographic region in the northwest of Iran. The program covers now 20 000-25 000 births (average per year) in the area with about 400 cases (average per year) of congenital anomalies born with one of the anomalies in this population. Congenital anomalies have been defined based on the standard coding system of the International Classification of Diseases (ICD) and British Pediatric Association (BPA) under one of the main headings according to the primary diagnosis of anomaly. A total of 247 695 births (245 063 live births, 2632 still births) with 4704 cases of congenital anomalies have now been registered between 2000 and 2011. Total prevalence of congenital anomalies was 1.9 per 100 births for this period of time. Genito-urinary tract and kidney defects, anomalies of nervous system, and limb anomalies accounted proportionally for more than 71% of anomalies in the region.

Because of confidentially restriction and ethical issues, we do not register the name and exact location (address) of the cases. The area (i.e., city, district, etc) are registered instead. Based on the 19 areas coded in the program, “Regional” spatial data

analysis can be performed in our registry data using statistical tests of Global Moran's I and Geary's C. The power of these tests depends on the level of data accumulation and population heterogeneity [4]. The relatively small number of the areas coded in the program and as there is no point data, descriptive features of spatial analysis will be more feasible for our data. Geo-coding system may improve the quality of registry for advance spatial analysis on the data in the future.

Discussion

Despite the importance of spatial analysis in surveillance of congenital anomalies, it has not been used broadly because of some limitations on the data derived from birth defects registries [16]. A major limitation of the registry data for spatial analysis is ethical issues and confidentiality restrictions of the data resulting in "Regional" data while "Point" data are technically preferred. ICD based coding system of the birth defects registries is another limitation of the application of spatial analysis for congenital anomalies. These codes are normally based on the body organs and systems whereas it is essential part of any spatial analysis to have etiologic pathways in the data too [17].

It is concluded that large international registries of congenital anomalies (i.e., European Surveillance of Congenital Anomalies and International Clearinghouse for Birth Defects, etc) may take the advantages of multi-level area analysis (i.e., city based, in hospitals, regions, etc) using the current softwares and advanced techniques. Small registries of congenital anomalies will have to improve their quality of data and expand the area and population for which the data collected enabling them for more reliable spatial analysis.

Acknowledgement: This study was financially supported by Tabriz Pediatric Health Research Centre (PHRC). The authors wish to thank Professor Mohammad Barzegar of PHRC for his comments and assistance.

Conflict of interests: The authors declare no conflict of interest.

References

1. *WHO fact sheet* on congenital anomalies[http://www.who.int/mediacenter/factsheets/fs370/]
2. Maffini G: Raster versus vector data encoding and handling: a commentary**.** *Photogrammetric Engineering and Remote Sensing* 1987, 53(10):1397-1398.
3. Rogers MY: Getting started with geographic information systems (GIS): a local health department perspective**.** *Journal of Public Health Management and Practice*1999, 5(4): 22-32.
4. Waller LA, Gotway CA. *Applied spatial statistics for public health data.* New Jersey*:* John Wiley & Sons Ltd; 2004.
5. EUROCAT Statistical Monitoring Protocol. *European Surveillance of Congenital Anomalies. 2009. Available at:[* http://www.ihe.be/eurocat. Accessed January 2013*.]*
6. Besag J, Newell J: The detection of clusters in rare diseases. *Journal of the Royal Statistical Society,* Series A (Statistics in Society)1991,154:143-155.
7. Goodchild MF: *Spatial autocorrelation.* Norwich: Geo Books; 1986.
8. Longly P, Batty M: *Spatial analysis: modelling in a GIS environment*. Canada: John Wiley & Sons Ltd;1996.
9. Anselin L: Local indicators of spatial association-LISA. *Geographical Analysis* 1995, 27(2): 93-115.
10. Cuzick J, Edwards R: Spatial clustering for inhomogeneous populations**.** *Journal of the Royal Statistical Society*. Series B (Methodological) 1990,73-104.
11. Openshaw S, Charlton M, Wymer C, Craft A: A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information System* 1987,1(4):335-358.
12. Turnbull BW, Iwano EJ, Burnett WS, Howe H, Clark LC: Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 1990, 132(supp1): 136-143.
13. Kulldorff M, Nagarwalla N: Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995, 14(8): 799-810.
14. Patil GP, Taillie C: Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics* 2004, 11(2): 183-197.
15. Tabriz Registry of Congenital Anomalies (TRoCA). 2013. Available at:[http://troca.tbzmed.ac.ir. Accessed January 2013.]
16. Siffel C, Strickland MJ, Gardner BR, Kirby RS, Correa A: Role of geographic information systems in birth defects surveillance and research. Birth Defects Research Part A: *Clin Mol Teratol* 2006, 76(11): 825-833.
17. Wellesley D, Boyd P, Dolk H, Pattenden S: An aetiological classification of birth defects for epidemiological research. *J Med Genet* 2005, 42(1): 54-57.