COMPUTATIONAL METHODS IN SCIENCE AND TECHNOLOGY **18(1)**, 53-62 (2012)

Location of Cleft Lip with or without Cleft Palate Prevalence Clusters using Kulldorff Scan Statistics

Barbara Więckowska[1], Anna Materna-Kiryluk[2], Tomasz Kossowski[3], Jerzy Moczko[1], Katarzyna Wiśniewska[4], Anna Latos-Bieleńska[2]

[1]*University of Medical Sciences in Poznań, Chair and Department of Computer Science and Statistics*
Dąbrowskiego 79, 60-529 Poznań, Poland
e-mail: basia@ump.edu.pl
[2]*University of Medical Sciences in Poznań,Department of Medical Genetics*
Grunwaldzka 55/15, 60-352 Poznań, Poland
e-mail: akiryluk@ump.edu.pl
[3]*Adam Mickiewicz University in Poznań, Institute of Socio-Economic Geography and Spatial Management*
Dzięgielowa 27, 61-680 Poznań, Poland
e-mail: tkoss@amu.edu.pl
[4]*University of Medical Sciences in Poznań, Department of Preventive Medicine*
Smoluchowskiego 11, 60-179 Poznań, Poland
e-mail: kwisniewska@ump.edu.pl

(Received: 09 March 2011; revised: 20 March 2012; accepted: 31 March 2012; published online: 27 April 2012)

Abstract: The prevalence of clusters with the increased morbidity rate is the area of interest among epidemiologists. Not only does the identification of clusters require collecting precise epidemiological data but it also requires the application of reliable spatial statistics techniques. The identification of atypical clusters in this article is performed using data from the Polish Registry of Congenital Malformations (PRWWR) on children with isolated cleft lip with or without cleft palate; the study was carried out in the Wielkopolska Region (Greater Poland). For this purpose, Kulldorff Scan Statistics and the LISA method were used. Since each technique used in the study focuses on a slightly different aspect of spatial structure, the obtained clusters do not always completely overlap. This study presents and compares the efficiency and accuracy of these two non-standard methods of geo-static analysis in children living in the Greater Poland counties. The study has identified 5 agglomerations with an increased prevalence rate of the examined malformation, no statistically significant cluster has been detected. On the basis of the agglomerations, it was possible to compare the applicability of two statistical methods used in the study. Despite the fact that the located clusters do not always completely overlap, the study has proved similarity in qualifying particular counties for given clusters and areas outside the clusters. Taking into account its applicability and monitoring the process of spatial scanning, the Kulldorff method has occurred more universal and accurate in examining the children with congenital malformations.

Key words: clusters, Kulldorff Scan Statistics, LISA, cleft lip, cleft palate

I. INTRODUCTION

According to the definition by the Eurocat Working Group[1], a cluster of congenital malformation is an aggregation of cases of congenital anomaly in time and/or space which appears to be unusual. The incidence of a such cluster may be caused by local spatial relations. Fotheringham (1997 [1], 2000 [2]) distinguishes four main groups of methods used for examining clusters: 1) Geographical Analysis Machine (GAM), 2) Local Indicators of Spatial Association (LISA), 3) Geographically Weighted Regression, and 4) Mathematical Modelling of Flows. While the

[1] EUROCAT is an European network of population-based registries for the epidemiologic surveillance of congenital anomalies, started in 1979. The EUROCAT Working Group on the Management of Clusters and Environmental Exposure Incidents has developed cluster definition in 2003.

first two methods are useful in identifying clusters, the third and the fourth methods are used for establishing local spatial relations. GAM is a set of methods applied in the analysis of point data, while LISA is applied in the analysis of aggregate data. The current work will demonstrate the use of Kulldorff Scan Statistics which is not only related to GAM but can also be useful in investigating the aggregate data on certain conditions. The clusters identified by means of Kulldorff Scan Statistics will be compared with the results obtained by the LISA method.

The geographical Analysis Machine (GAM) was proposed by Oppenshawn (1987) [3] and has been used for analyzing point data. This method forms sets of different sized circles with the centers located in the points serving as knots of a regular grid (usually made of squares) that covers the study area. Next, the true number of cases with a given malformation in each circle is compared to the score expected when the number of prevalence cases has Poisson distribution. Turnbull et al. (1990) [4] have developed a test that allows for examining both the location and significance of clusters. The test consists in forming circular windows in the shape of a circle; the windows contain a fixed number of units at risk that are estimated a priori. In turn, Besag and Newell have expanded the GAM method with the possibility of setting letter k as an expected size of cluster. Letter k usually varies between 2 and 10. The method was described by Besag & Newell (1991) [5] and Alexander & Cuzick (1992) [6]. Kulldorff Spatial Scan Statistics (Kulldorff and Negarwall 1995 [7], 1997 [8]) relates both to the Geographical Analysis Machine (GAM) by Oppenshaw et al. (1987) and the procedure proposed by Turnbull et al. (1990). Not only does it allow for identifying circular clusters but it can also determine elliptical ones. It was used for the first time by Kulldorff and Nagarwall (1995) to identify the leukemia prevalence clusters in the state of New York. A review of studies based on Spatial Scan Statistics has been included in Spatial Analysis in Epidemiology [9]. In addition, Kulldorff has elaborated and proposed a space-time version of Scan Statistics.

The I_i Moran and C_i Geary statistics (Anselin 1995) [13] which form Local Indicators of Spatial Association (LISA) are treated as local variants of global spatial statistics and operate on aggregate data. They allow for the identification of local spatial relation (autocorrelation) that manifests itself in the form of clusters with similar values of an examined variable and outliers that stay in contrast with their surrounding. The cluster identification has been carried out on the basis of Local Moran Statistics.

The LISA metod was used by Jacquez and Greiling (2003) [14] to identify spatial clusters with increased prevalence cases of breast, lung, and large intestine tumor in the population of Long Island, USA. The results which they obtained coincided with the results that Kulldorff (1997) had obtained before. The areas identified by these two methods differed from each other to some extent. It seems important to stress that the process of detecting statistically significant clusters requires long and expensive alarm procedures proposed by the Eurocat Working Group[2]. Jacquez suggests that the identified clusters be examined with the use of more than one method as it will allow for more accurate and reliable examination. A comparative analysis of the LISA method and Kulldorff Scan Statistics was carried out by Hanson and Wieczorek (2005) [15] who investigated the death rate among alcohol abusers in New York, USA. Their study has proved Kulldorff Scan Statistics more accurate in the cluster identification than the LISA method. The obtained results also show that, as each technique entails the use of its own criteria, different counties can be chosen when different techniques are applied. However, when the results are compared we can see that both methods illustrate different elements of the same clusters. Consequently, we can achieve the best results by combining the methods instead of using them separately. Both methods were proven to be successful in the Discovery of clusters of Hemorrhagic fever with renal syndrome (HFRS) which was common in China in the years 1994-98 (L. Fang et al. (2006) [16]). Clusters were located in the eastern and northwestern parts of China.

Although spatial analysis methods are more and more popular in epidemiology, analyses of congenital malformations are difficult due to their relatively small prevalence rate. The analysis can only be made when a large register of congenital malformations is available in which the reporting of defects is full. That is why there are still few sufficiently detailed works on that issue. Usually the investigation of cleft lip with or without cleft palate has limited the analysis of geographical factors to a town-country distinction (B-H Hwang et al. 2008 [17], L.C. Messer et al. 2010 [18]), or the distinction between 2-3 regions, e.g. North-South (E.W. Harville 2005 [19]). We have only found one work in which the malformation was studied in more detail: S.Y. Gebreab 2010 [20]. The work describes the limitations resulting from analyzing diseases with relatively small prevalence, on the example of congenital malformations, especially oral clefts (n = 894 cases, 458593 live births). The analyses presented in the work pertain to the state of Utah (USA) in the years 1995-2004. To analyze the studied area with the view to finding clusters the state was divided into 61 sections. The

[2] Cluster investigation protocol developed by the Eurocat Working Group can be found on the websites of EUROCAT: http://www.eurocat.ulster.ac.uk/clusterinvprot.html

statistical analyses, including Kulldorff's scan statistics, located clusters with an increased prevalence rate of congenital malformations; however, the results are on the border of statistical significance.

II. DATA

The research is based on the data concerning 229 liveborn children at the age of 0-2 with isolated cleft lip with or without cleft palate from Greater Poland (divided into 315 counties), recorded by the Polish Registry of Congenital Malformations (PRWWR) in the period of 1999-2006. This malformation has been chosen mainly due to its recognizibility, recordability, and the contribution of environmental teratogenic factors to the etiology of this defect. It seems vital to stress that there are no reports on where exactly in Poland the clusters with this defect can be found.

The present analysis was carried out on a county level. Therefore, the information about the place of residence has become a vital criterion for being qualified for the analysis.

Population data regarding 277136 liveborn children in individual county of Greater Poland in 1999-2006 period come from the Central Statistical Office of Poland.

III. METHODS

The basic descriptive statistics of congenital malformations are manifested by coefficients of frequency with which the defects occur. A frequency coefficient that deals with the prevalence of congenital malformations in liveborn children is defined as the proportion of liveborn children with congenital malformations to the total number of liveborn children at a given time and space. In the case of isolated cleft lip with or without cleft palate in a given area, the prevalence is 8.3 in 10 thousand livebirths.

What takes place very often, especially in epidemiology, is the occurrence of data bases in which the number of examined participants is very small, and some groups are even empty. Therefore, the examined coefficients appear sensitive towards small data variations. In this case, the use of local smoothing (by means of Spatial Empirical Bayes) will allow for determining smoothed values of prevalence, which will help to retrieve stability and provide distinction between the regions considered before as 0. Mathematical details of the smoothing coefficients process with the use of the Bayes' method can be found in the works of Anselin et al. (2004) [21] and Boscoe F.P. et al. (2003) [22]. Smoothed coefficients for isolated cleft lip with or without cleft palate in particular counties are presented below (Fig. 1).

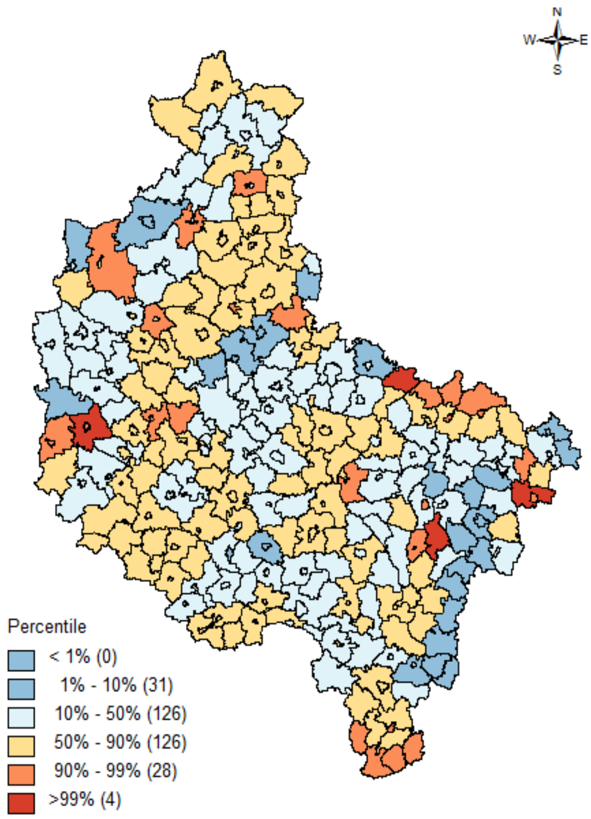

Fig. 1. Map of Greater Poland. Smoothed prevalence coefficients (SEB) of isolated CL +/- P in 1999-2006

Smoothing coefficients of frequency with which cleft lip with or without cleft palate takes place, drawing maps, and the LISA analysis were carried out on the basis of the OpenGeoDa v0.9.8.14, 2009 program (http://geodacenter.asu.edu). For the need of this analysis, counties are defined as polygons. In turn, Kulldorff in his analysis defines counties as centroids and applies the SatScan v8.2.1, 2010 (http://www.satscan.org) software to perform this analysis. The comparison of results obtained from the two methods was made using StatXact v9.0.0, 2010 (http://www.cytel.com).

Assumptions of the Kulldorff Scan Statistics Theory

Identifying the location of clusters by means of the Kulldorff (1995, 1997) method requires using a scanning window. The scanning window can be realized as: 1) a range – for time clusters, 2) a circle or ellipse – for space clusters, 3) a cylinder – for space-time clusters. The scanning window with different sizes is applied after it is identified by Euclidean matrix. Both the observed and expected number of cases for every position and size of the scanning window are estimated inside and outside the window. In addition, the study area is scanned in order to find the most likely cluster. Theoretically, the process of scanning can be infinite; however, it may turn finite when narrowed to two conditions, namely: 1) limiting the number of starting points of the scanning window in which every centroid or another limited set of points in space at a specified time dimension can serve as the center for building the window, and 2) limiting the size of the scanning window by determining the maximum size of population that can be located within the window or in the space range that covers the window, the range being determined by the distance from the window center. Similar restrictions are imposed on a time unit in space-time clusters, which allows for manipulating the window size and position; it is then possible to control the scanning process. Moreover, the restrictions allow for identifying unfocused clusters as well as examining the cluster range located around a previously identified point of an increased risk, known as *focus.*

Statistical significance of the scanning window is determined by the likelihood ratio test in which $H_0 : p = q$ is tested against $H_1 : p > q$, with p standing for the likelihood of malformations to occur inside scanning window *Z*, and q acting as the likelihood of malformations to occur outside scanning window *Z*. In the starting phase of the test, the likelihood function $L(Z, p, q)$ for a given window Z is determined. The collected data on population and those concerning congenital malformations enable to base a likelihood function on the Poisson model.

The likelihood function for window *Z* is:

$$L(Z,p,q) = \left(\frac{c}{E(c)}\right)^{c} \left(\frac{C-c}{C-E(c)}\right)^{C-c} I(),$$

where:
c – number of cases (children with congenital malformation) inside the scanning window,
C – total number of cases,
$E(c)$ – expected number of cases inside the scanning window,
$C - E(c)$ – expected number of cases outside the scanning window,
$I()$ – an indicator function equals 1 when the number of cases inside the window is bigger than the expected number of cases, or 0 in the other case.

The procedure of estimating the likelihood function is performed for each position and size of the scanning window.

Next, the maximum of the function is being determined:

$$L(Z) \overset{def}{=} \sup_{p>q} L(Z,p,q).$$

The scanning window for which

$$\hat{Z} = \{Z : L(Z) > L(Z') \quad \forall Z'\}$$

is the most likely cluster $\hat{Z}$.

This cluster shows the least likelihood of accidental prevalence.

When defining L_0 as:

$$L_0 \overset{def}{=} \sup_{p=q} L(Z,p,q),$$

then the equation below (the so-called likelihood ratio test) will act as the likelihood ratio for the most likely cluster.

$$LR = \frac{L(\hat{Z})}{L_0}.$$

The distribution of *LR* test statistics appears difficult to obtain in an analytical form, therefore it is obtained by means of the Monte Carlo method. By this method, random sets of data are being generated; next, the maximum of a likelihood function for the true data is being compared with the maximum of a likelihood function for different sets of random data. For 9999 repetitions (generated sets), the test is significant on level 0.05 if a statistical value for a true set of data is located between 500 highest values of test statistics coming from these generated sets.

After having identified the most likely cluster, other clusters are being located, that is why the whole procedure is repeated with the exclusion of a previously located cluster.

Assumptions of the Local Indicators of Spatial Association (LISA) Theory

The analysis of LISA was carried out using Local Moran Statistics which requires the identification of the so-called spatial weights. The weights represent spatial relations recorded by the matrix or graph. In order to identify the weights, a neighborhood matrix in the form of a binary table is designed, where the neighborhood of two counties is defined as 1, and its lack as 0. Next, the matrix of weights is obtained by standardizing the neighborhood matrix by rows to 1. In the presented method, the neighborhood matrix is usually determined by the common boundary between counties, in turn, the scope of cluster is determined by the neighborhood matrix. However, in the present study, to compare the results gained by the Kulldorff method, the neighborhood matrix is based on Euclidean distance.

Local Moran I_i Statistics, also known as a Moran coefficient, is used to investigate whether the region is surrounded by the neighbors with a similar or different value of the following variable:

$$I_i = \frac{(z_i - \bar{z})\sum_{i=1}^{n} w_{ij}(z_j - \bar{z})}{\frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{n}}$$

where:

z_i, z_j – prevalence values for examined counties,

$\bar{z}$ – the average value of prevalence for the whole voivodship,

w_{ij} – elements of spatial matrix of weights,

n – the number of counties.

The null hypothesis assumes that $I_i = 0$, and its statistical distribution is made by Monte Carlo simulation. It is suggested that the significance level be adjusted in regard to the average number of neighboring areas. Therefore, Bonferroni's correction $\alpha_1 = \alpha/k$ or Sidak's $\alpha_1 = 1-(1-\alpha)^{1/k}$ are used, where k stands for the average number of neighboring areas.

IV. RESULTS

Results of Kulldorff Scan Statistics

Spatial clusters

The scanning window is defined as a circle. With the use of the window, the whole area of Greater Poland is being scanned (Fig. 2).

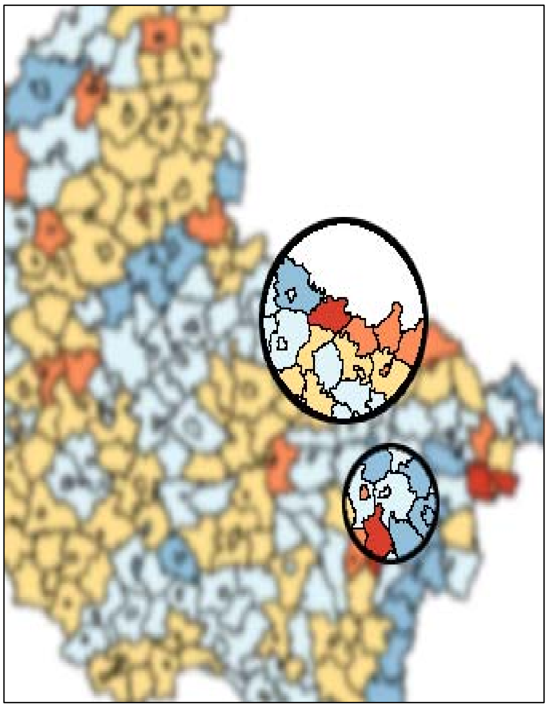

Fig. 2. An example of the space scanned with the use of a circle

The choice of a starting point in the scanning window and the manner in which the window will move while scanning was the basic and fundamental issue in detecting spatial clusters. As a result, the centroids of counties, serving as geometrical confirmation of the expected center, became the starting point for the described process (Fig. 3).

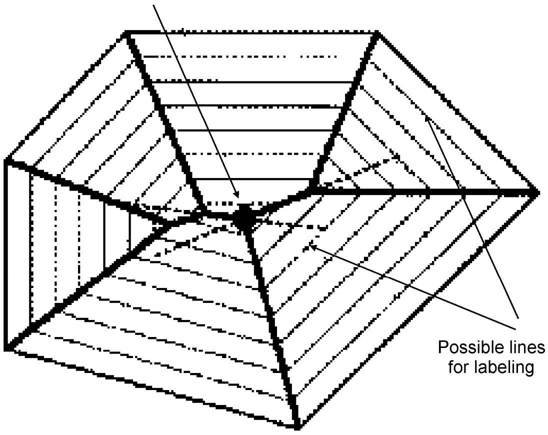

Fig. 3. An example of centroid estimated on the basis of the poligon

Source: Geographic Information Analysis, David O'Sullivan, David J. Unwin 2010

The process of scanning was limited by determining the maximum size which the radius of scanning window can reach. As it was assumed, the scanning window in each starting point continued to make the radius bigger (every 1 meter) until it reached the size which would allow for locating over 50% of liveborn children from Greater Poland inside the window.

As the results show, the present study has failed to identify clusters that would exhibit a statistically significant clusters for the prevalence of cleft lip with or without cleft palate; however, the study has managed to determine few agglomerations with a higher likelihood of the malformation to occur (Table 1, Fig. 4).

In those agglomerations the presented differences between the expected and the observed number of cases are quite big but based on small sizes, which makes them statistically insignificant.

Space-time clusters

The scanning window was defined as a cylinder with circular base. The starting point for the window in space dimension was a centroid, whereas in time dimension it

Table 1. Results of Kulldorff Space Scan Statistics

Scanning window: circle	The most likely cluster	Other likely clusters			
	circle 1	circle 2	circle 3	circle 4	circle 5
County of the circle center	Skulsk	Mycielin	Zbąszyń – town	Buk – Town	Łęka Opatowska
Circle center – X-coordinate	454450	447680	289740	331150	438500
Circle center – Y-coordinate	512210	454850	491930	501380	372290
Radius length (in metres)	14720.41	0	16293.02	13376.67	8866.89
Population size	438	52	569	573	204
Observed number of cases	9	3	10	10	5
Expected number of cases	2.89	0.34	3.76	3.79	1.35
Prevalence (in 1000)	205.4794521	576.9230769	175.7469244	174.5200698	245.0980392
Relative risk	3.2	8.9	2.73	2.72	3.77
Log (LR) statistics	4.192513	3.878769	3.625298	3.586343	2.93216
P value	0.778 (NS)	0.855 (NS)	0.912 (NS)	0.924 (NS)	0.994 (NS)

NS: no statistic significance – p value $\geq$ 0.05

Table 2. Results of Kulldorff Space-time Scan Statistics

Scanning window: cylinder Cylinder base: a circle on the space surface Height of cylinder: time in years	The most likely cluster	Other likely clusters			
	circle 1	circle 2	circle 3	circle 4	circle 5
Municipality of the circle centre	Wierzbinek	Łęka Opatowska	Buk – rural area	Pyzdry – rural area	Powidz
Years	2006	2004-2005	2002-2004	2000	2006
Circle center – X-coordinate	465560	438500	332670	413450	427600
Circle center – Y-coordinate	508500	372290	500710	475980	507840
Radius length	5244.44	8866.89	14425.67	0	10932.43
Population size	138	204	882	51	266
Observed number of cases	3	4	9	2	3
Expected number of cases	0.12	0.33	2.14	0.052	0.2
Prevalence (in 1000)	217.3913043	196.0784314	102.0408163	392.1568627	112.7819549
Relative risk	25.05	12.32	4.33	39.14	15.16
Log (LR) statistics	6.764217	6.341233	6.166035	5.3772	5.336653
p value	0.791	0.888	0.922	0.997	0.997

NS: no statistic significance – p value $\geq$ 0.05

Table 3. Results of LISA spatial analysis

	Likely clusters				
	circle 1	circle 2	circle 3	circle 4	circle 5
County of the cluster center (high-high)	Krzywiń rural area	Osieczna town	Buk rural area	Wierzbinek	Wronki-town
Population size	1548	7334	4539	2369	2488
Observed number of cases	2	10	9	6	4
Prevalence (in 1000)	12.91989664	13.63512408	19.82815598	25.32714225	16.07717042
Moran coefficient	0.737909	3.14348	0.638811	1.47087	0.158063
p value	0.016 (NS)	0.028 (NS)	0.036 (NS)	0.044 (NS)	0.046 (NS)

Correction of Bonferrony/Sidak's level of significance: $k = 4.75$, $\alpha_1 = 0.011$; NS: no significance, p value $\geq$ 0.011

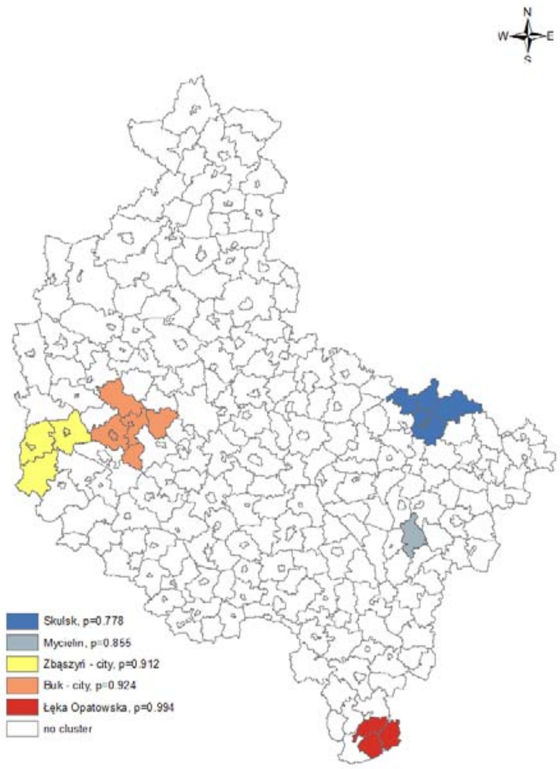

Fig. 4. Map of Greater Poland. Results of Kulldorff Space Scan Statistics

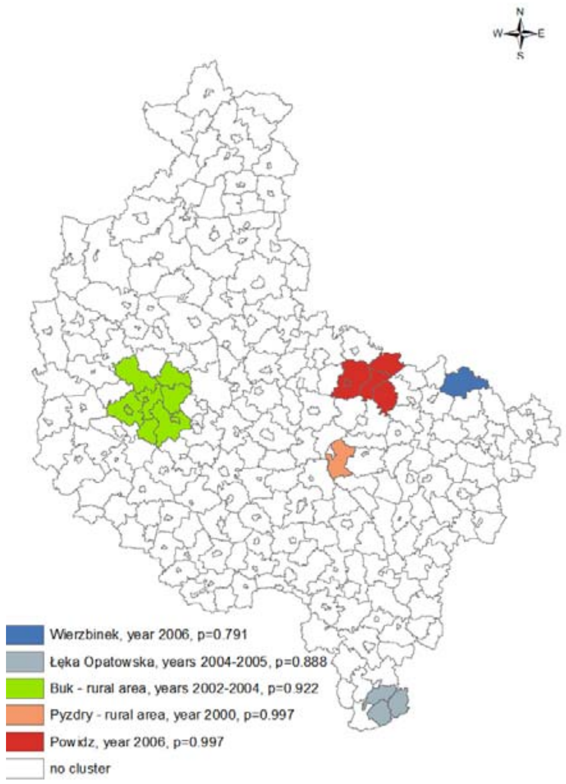

Fig. 5. Map of Greater Poland. Results of Kulldorff Space-Time Scan Statistics

was each year. The window size was limited so as not to fit more than 50% of population examined and cover less than half of the period examined, namely not more than four years. In this case, substantial differences were also observed between the expected and observed number of cases; however, due to small sizes the discovered agglomerates shown in the table (Table 2) and on the map (Fig. 5) are statistically insignificant.

Results of LISA analysis

For the sake of Local Moran Statistics, a given county is accepted as neighbouring if its centroid is located at a distance of less than 14 km (about 9 miles) from a particular county, since the average size of the agglomeration identified by Kulldorff is of similar size. As a result, no statistically significant agglomerations have been found (Table 3, Fig. 6).

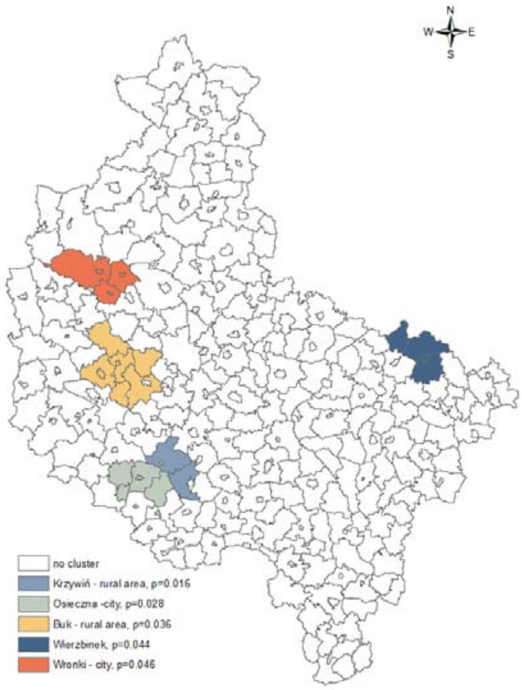

Fig. 6. Map of Greater Poland. Results of LISA spatial analysis

V. COMPARISON OF KULLDORFF SCAN STATISTICS RESULTS WITH THE FINDINGS OBTAINED BY THE LISA METHOD

In the present study the range agreement of agglomerations identified by means of Kulldorff Scan Statistics has beencompared with agglomerations obtained using the LISA method. Since LISA is not equipped with its own equivalent which would include several-year time dimension, the comparison was made only among spatial agglomerations. Table 4 presents the arrangement of counties related to particular clusters.

Table 4. The comparison of results obtained by means of Kulldorff Scan Statistics with those obtained through the LISA method

	Counties located outside the agglomerations of Kulldorff Statistics	Counties located inside the agglomerations of Kulldorff Statistics	Total
Counties located outside the agglomerations of LISA statistics	276	13	289
Counties located inside the agglomerations of LISA Statistics	17	9	26
Total	293	22	315

The agreement percent of a county related to a given cluster (on the main diagonal) is 90.48%. A Kappa coefficient, correct the size by chance on agreement, is 32.38% and 95% confidence interval for Kappa is: 14.04%, 50.73%. The result of the agreement analysis, so the significance of the Kappa's coefficient ($p < 0.000000$) proves that the two methods assign counties both to the areas belonging to the agglomerations with an increased risk of the cleft lip ± cleft palate prevalence and to the ones that do not belong to the agglomerations. At the same time, only 40.91% of Kulldorff's agglomerations overlap with agglomerations identified by the LISA method, and only 34.62% of agglomerations identified by the LISA method overlap with Kulldorff's agglomerations.

VI. CONCLUSIONS

Congenital malformations, being influenced by both environmental and genetic factors, appear relatively seldom; therefore, patterns on the map remain rather random and irregular. Because of their low number of cases, particular agglomerations are created randomly and not on the basis of statistically significant relations, despite the differences between the observed and expected number of cases of malformations. Similar statistically insignificant differences were obtained between the observed and the expected number of cases in S.Y. Gebreab's work from 2010 [20] for the state of Utah (USA).

Techniques used for investigating the clusters are not common as they are based on specially designed databases that are related to the maps. In Poland, the Polish Registry of Congenital Malformations (PRWWR) was established in 1997 and since then it has collected data on the frequency and types of congenital malformations in Polish society. Reports on the children with congenital malformations and information about livebirths provided by the Central Statistical Office of Poland databases in connection with a detailed map of Greater Poland with all counties highlighted have contributed to the establishment of databases needed for detailed spatial analysis.

Since each technique used in the present study focuses on different aspects of information presented by the data, slightly different areas have been considered as likely but statistically insignificant agglomerations. It may come from the fact that the way in which cluster is sought in the Kulldorff method allows for determining the optimal size of the cluster, whereas in the LISA method the size is determined a priori by estimating the manner of neighboring. In identifying clusters, the researcher is usually unable to pinpoint their exact size. It becomes even more difficult when the area of investigation is bigger and contains more windows than usual. It is important to stress that each true cluster, being located in the area of inquiry, can have different sizes. Therefore, the process of identifying clusters will become easier if the researcher is provided with information about starting points of the window and the maximum of radius to which the window can spread. In addition, not only is the Kulldorff method able to identify spatial clusters but also space-time ones, which seems useful in epidemiological studies. Therefore, considering the comfort of application and use, the Kulldorff method has proved more precise and universal than the LISA method.

On the basis of the study results, thesis on the lack of clusters of isolated cleft lip with or without cleft palate in Greater Poland in the period of 1999-2006, has not been rejected. The results were obtained due to two different methods of geostatistical analysis. Kulldorff Scan Statistics has identified five agglomerations with higher risk of congenital malformation prevalence, which has been confirmed by the LISA method. The result may suggest slight influence of environmental teratogenic factors on the development of congenital malformations. Until now, malformations have not been extensively analysed (apart from S.Y. Gebreab's work, 2010 [20]), therefore today it appears difficult to find detailed and thorough information about them in the subject literature. Early analyses of isolated cleft lip with or without cleft palate usually limited the investigation of geographical factors to a simple distinction into town-country (village) or the division of a particular area into two or three regions (Bing-Fang Hwang et al. 2008 [17], Messer L.C. et al. 2010 [18], Harville E.W. et al. 2005 [19]). As a result, the subject of residence was not entirely investigated and included in the study. Therefore, a variety of tools for investigating clusters used in the present study may contribute to the development of methods useful in lowering the risk of delivering a child with congenital malformation. Moreover, the clusters detected at specific locations may occur helpful in determining environmental teratogenic factors being at the same time responsible for the development of clusters. Narrowing the area of invesgation in quest of teratogenic factors will lower effort and time needed to identify them. Then, after identification, teratogenic factors will allow for reducing the costs (financial costs connected with the necessity to spend money on treatment of diseases caused to some extent by environmental factors) and effort put by parents whose children suffer from congenital malformations.

References

[1] A.S. Fotherigham, *Trends in Quantitative Methods I: Stressing the Local*, Progress in Human Geography 21, 88-96 (1997).

[2] A.S. Fotheringham, *Context-dependent spatial analysis: A role for GIS?* Journal of Geographical Systems 2, 71-76 (2000).

[3] S. Openshaw, M. Charlton, C. Wymer, A. Craft, *A mark I geographical analysis machine for the automated analysis of point data sets*, International Journal of Geographical Information System 1, 335-358 (1987).

[4] B.W. Turnbull, E.J. Iwano, W.S. Burnett, H.L. Howe, L.C. Clark, *Monitoring for clusters of disease: application to leukemia incidence in upstate New York*, American Journal of Epidemiology, Jul 132(1 Suppl), 136-143 (1990).

[5] J. Besag, J. Nowell, *The detection of clusters in rare diseases*, Journal of the Royal Statistical Society, Series A 154, 143-155 (1991).

[6] F.E. Alexander, J. Cuzick, *Methods for the assessment of disease cluster*, in: Elliott P., Cuzick J., English D., Stern R. (eds), *Geographical end Environmental Epidemiology: Methods for Small – Area Studies*, Oxford University Press, Oxford, p. 238-250 (1992).

[7] M. Kulldorff, N. Nagarwalla, *Spatial disease clusters: Detection and inference*, Statistics in Medicine 14, 799-819 (1995).

[8] M. Kulldorff, *A spatial scan statistic*, Communications in Statistics: Theory and Methods 2, 1481-1496 (1997)

[9] D.U. Pfeiffer, T.P. Robinson, M. Stevenson, K.B. Steven, D.J. Rogers, A.C.A. Clements, *Spatial Analysis in Epidemiology*, Oxford University Press, Oxford (2008).

[10] M. Kulldorff, W.F. Athas, E.J. Feuer, B.A. Miller, C.R. Key, *Evaluating cluster alarms. A space-time scan statistic and brain cancer in Los Alamos*, New Mexico, American Journal of Public Health 88, 1377-1380 (1998a).

[11] M. Kulldorff, *Prospective time periodic geographical disease surveillance using a scan statistic*, Journal of the Royal Statistical Society, Series A 164, 61-72 (2001).

[12] M. Kulldorff, R. Heffernan, J. Hartman, R. Assuncao, F. Mostashari, *A space-time permutation scan statistic for disease outbreak detection*, PLoS Medicine 2, e59 (2005).

[13] L. Anselin, *Local indicators of spatial association – LISA*, Geographical Analysis 27, 93-115 (1995).

[14] G.M. Jacquez, D.A. Greiling, *Local clustering in breast, lung and colorectal cancer in Long Island*, New York, International Journal of Health Geographics 2: 3, (2003)

[15] C.E. Hanson, W.F. Wieczorek, *Alcohol mortality: a comparison of spatial clustering methods*, Social Science and Medicine 55, 791-802 (2002).

[16] L. Fang, L. Yan, S. Liang, S.J. de Vlas, D. Feng, X. Han, W. Zhao, B. Xu, L. Bian, H. Yang, P. Gong, J.H. Richardus, W. Cao, *Spatial analysis of hemorrhagic fever with renal syndrome in China*, BMC Infectious Diseases, 6, 77 (2006).

[17] B.F. Hwang, Jouni J.K. Jaakkola, *Ozone and Other Pollutans and the risk of Oral Cleft*, Environ Health Perspectives 116(10), 1411-1415 (2008).

[18] L.C. Messer, T.J. Luben, P. Mendola, S.E. Carozza, S.A. Horel, P.H. Langlois, *Urban-Rural Residence the Occurrence of Cleft Lip and Cleft Palate in Texas, 1999-2000*, Annals of Epidemiology 20, 32-39 (2010).

[19] E.W. Harville, A.J. Wilcox, R.T. Lie, H. Vindenes, F. Åbyholm, *Cleft Lip and Palate versus Cleft Lip Only: Are They Distinct Deffects?* American Journal of Epidemiology 162 (5), 448-453 (2005).

[20] S.Y. Gebreab, *Spatial Epidemiology of Birth Defects in the United States and the State of Utah Using Geographic Information Systems and Spatial Statistics.* All Graduate Theses and Dissertations. Paper 852 (2010).

[21] L. Anselin, Y.W. Kim, I. Syabri, *Web-based analytical tools for the exploration of spatial data*, Journal of Geographical Systems 6, 197-218 (2004b).

[22] F.P. Boscoe, C. McLaughlin, M.J. Schymura, C.L. Kielb, *Visualization of the spatial scan statistic using nested circles*, Health and Place 9, 273-277 (2003).

Barbara Więckowska received her MSc degree in Applied Mathematics from Adam Mickiewicz University in Poznań. Since 2004 she has been working at the Chair and Department of Computer Science and Statistics at Poznań University of Medical Sciences. There she received her PhD in Medical Sciences. She develops methods of multivariate statistics used in epidemiology of congenital malformations. The results are published in many Polish and international scientific journals.

ANNA MATERNA-KIRYLUK – Doctor of medicine, pediatrician, specialist in clinical genetics in Department of Medical Genetics, Poznan University of Medical Sciences, Poland. Coordinator of the Polish Registry of Congenital Malformations (PRCM). Degree: PhD (1993) Poznan University of Medical Sciences. Research interests: congenital malformations, epidemiology and prevention of congenital malformations. Participant in the EU Projects: EUROCAT, EUROmediCAT, PROTECT, CHERISH.

TOMASZ KOSSOWSKI works at the Institute of Socio-Economic Geography and Spatial Management of the Adam Mickiewicz University in Poznań. He received his MSc degree in socio-economic geography, MSc degree in mathematics, and the PhD in earth sciences from Adam Mickiewicz University. His main areas of interest are: methods and models of spatial econometrics, spatial statistics, mathematical modeling of spatial structures and processes and some aspects of random field theory. In last years he focused his research on neo-classical growth models and spatial convergence.

ANNA LATOS-BIELEŃSKA – Professor of medicine, physician, specialist in clinical genetics and laboratory medical genetics. Head of Department of Medical Genetics, Poznan University of Medical Sciences, Poland. Degrees: PhD (1979) and Habilitation (1992), Poznan University of Medical Sciences, 2002 – full professor. Post-doc position in Department of Clinical Genetics, University of Ulm, Germany, fellowship of Boehringer Ingelheim-Stiftung and Alexander von Humboldt-Stiftung (1985, 1986, 1988-89). Research interests: molecular basis of mental retardation and congenital malformations, epidemiology and prevention of congenital malformations, identification of new genes for rare genetic diseases. Author of more than 200 scientific publications, many times awarded, supervisor of 16 PhD students. Member of the Committees of the Polish Academy of Sciences: Human Development Committee (deputy president of the Committee) and Genetics and Molecular Pathology Committee. Partner in the EU Projects: EURO-MRX, CHERISH, EUROCAT, EUROmediCAT, PROTECT, project leader of a number of international and national projects. Creator and Director of the Polish Registry of Congenital Malformations (PRCM). PRCM monitors all births in Poland, is member of the EUROCAT and has information on more than 125,000 children with congenital malformations in the data base.

KATARZYNA WISNIEWSKA is Master of Public Health. She works in the Department of Preventive Medicine at the Faculty of Health Sciences, Karol Marcinkowski University of Medical Sciences in Poznan. Her main areas of interest are the epidemiology, etiology and prevention of congenital malformations.

JERZY A. MOCZKO – Full Professor and Head of the Chair and Department of Computer Science and Statistics at Poznań University of Medical Sciences in Poznań. Author of more than 250 scientific publications, supervisor of 9 PhD students. Member of the Committees of the Polish Academy of Sciences, Polish Academy of Arts and Sciences (Committee of Technical Sciences), Polish Society of Biomedical Engineering, Polish Society of Medical Physics, Polish Statistical Association, International Society of Perinatal Obstetrics. His research interests concern, amongs other topics, methods of digital processing of biomedical signals and methods of medical diagnosis support based on statistical, artificial intelligence and data mining techniques.