Annals of Agricultural and Environmental Medicine 2015, Vol 22, No 1, 110–117
www.aaem.pl

ORIGINAL ARTICLE

The detection of areas in Poland with an increased prevalence of isolated cleft lip with or without cleft palate

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

[1] *University of Medical Sciences in Poznań, Chair and Department of Computer Science and Statistics, Poznań, Poland*
[2] *University of Medical Sciences in Poznań,Department of Medical Genetics, Poznań, Poland*
[3] *University of Medical Sciences in Poznań, Department of Preventive Medicine, Poznań, Poland*
[4] *Adam Mickiewicz University in Poznań, Institute of Socio-Economic Geography and Spatial Management, Poznań, Poland*

Więckowska B, Materna-Kiryluk A, Wiśniewska K, Kossowski T, Latos-Bieleńska A. The detection of areas in Poland with an increased prevalence of isolated cleft lip with or without cleft palate. Ann Agric Environ Med. 2015; 22(1): 110–117. doi: 10.5604/12321966.1141379

Abstract

Introduction and objectives. It is difficult to identify the environmental factors which together influence the occurrence of congenital malformations. It could be helpful to define the geographic location of the areas with an increased prevalence of such malformations. The aim of this study is to define if there are regions in Poland where the prevalence of isolated cleft lip, with or without a cleft palate (CL±P), is increased, and to present a method for searching for such areas.

Materials and methods. The analysis included the whole area of Poland monitored in 2007–2008 by the Polish Register of Congenital Malformations (PRCM). The area was divided into 3,045 census regions. The number of children with CL±P in those years was 514, and the size of the reference population (live births) was 802,372. Two methods were used for the detection of clusters with an increased prevalence of isolated CL±P: the LISA analysis and Kulldorff's scan statistic, and described in detail.

Results. The prevalence of isolated CL±P and the smoothed prevalence were calculated for every community. The results of the LISA and Kulldorff's analyses were consistent. Both methods located the sites with an increased prevalence of isolated CL±P. The lack of statistical significance of clusters indicated by Kulldorff's statistic, and the significance of clusters detected with the use of the LISA method, indicated the existence of clusters with an only slightly increased prevalence of isolated CL±P.

Conclusions. The study shows the usefulness of the LISA and Kulldorff's spatial analyses in epidemiological studies, including the etiology of congenital malformations. Because the two methods work in different ways, good results can be obtained when they are used together.

Key words

cluster analysis, cleft lip and/or palate, Scan Statistics, LISA method

INTRODUCTION

CL±P is a defect conditioned by several factors, and a combination of genetic and environmental factors comprises the etiology of that defect. A number of environmental factors which increase the risk of the defect have been identified. They include the economic factor, risks at work and at home: pesticides, living near dangerous waste disposal areas, environmental pollution, such as lead or sulfur pollution, or contamination of drinking water.

A review of the literature on the topic shows that the risk of CL±P is higher when the mother is exposed to tobacco smoke [1, 2, 3], takes medicines [4, 5], drinks alcohol [6, 7], is exposed to chemicals at work or at home [8,9], drinks contaminated water [10], lives in a lead-polluted [11] or sulfur-polluted area [12, 13], lives in an air-polluted area [14, 15] or near dangerous waste disposal areas [16], or has vitamin and folic acid deficiencies [17, 18]. Also, individual factors, such as mother's older age or a higher number of pregnancies, have a negative influence [19].

Identification of the environmental factors which together influence the occurrence of the defect is difficult. It could be helpful to define clusters of CL±P, that is the geographic location of areas with an increased prevalence of that malformation.

The identification of the geographic clusters, on the one hand, would allow a more precise analysis of the risk factors and, on the other hand, targeted prevention.

Although methods of spatial analysis are becoming increasingly popular in the epidemiology of congenital malformations, there is a significant obstacle to their use, namely, the relatively low prevalence of such malformations. Methods of spatial analysis have been used chiefly for data from large, long-running registers of congenital malformations, with full reportability of the malformations. This is why, to date, there have been few sufficiently detailed studies on the topic. The existing studies concerning isolated CL±P limit the analysis of the geographic factor to a dichotomous division of space into urban and rural areas [20, 21], or to a division of space into two, three or four regions, i.e. north, south, east, and west [22, 23]. Unfortunately, such a form of structuring the studies does not make it easier to identify environmental

Address for correspondence: Barbara Więckowska, 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

Received: 11 January 2013; Accepted: 28 February 2013

teratogens in a specific geographic location. A review of the literature shows that in only one study the defect is examined more thoroughly [24]. That study describes in broad terms the limitations of analyzing malformations with relatively small prevalences – especially cleft lip (n= 894 cases, 458,593 live births). The analyses presented concern the state of Utah In the USA in 1995–2004. In order to define clusters, the studied area of the state has been divided into 61 parts. The conducted statistical analyses with the use of, among other tools, Kulldorff's scan statistic, reveal clusters with an increased prevalence of CL±P, however, the result is on the verge of statistical significance.

The first studies on the spatial pattern of the prevalence of congenital malformations have also been conducted with the use of PRCM data and concern the Wielkopolskie Province of Poland in 1999–2006. In this case, too, a small differentiation of the prevalences of CL±P is shown [25].

Objectives. The aim of this study is to determine if there are regions In Poland with an increased prevalence of cleft lip and palate in the area covered by PRCM. Two methods were used for the identification of an increased prevalence of CL±P, namely: Anselin's LISA [26] and Kulldorff's scan statistic [27].

MATERIALS AND METHOD

The first stage of the study, i.e. the search for clusters, involved the proper preparation of the collected data so that they form a consistent Geographic Information System (GIS). The GIS database contained descriptions of objects located in a geographic space. In principle, such a database comprises two parts: 1) a set of attribute data characterizing the non-spatial features of objects; 2) a set of spatial data defining the location of an object with the help of coordinates in an established reference system. Such a base is complemented by a map representing the relative locations of the spatial objects.

The prevalence of CL±P was determined on the basis of the data obtained from PRCM for 2007–2008. In those years, PRCM covered the areas of all 16 provinces. The communities of residence of the mothers of children with diagnosed CL±P are the basis for the analysis.

Spatial data are the information about the location of the communities. A community (gmina) is the smallest administrative unit available for the determination of the mothers' places of residence. The geographic locations of communities are defined by their boundaries, i.e. by polygons. The obtained maps are provided in the form of polygons saved in the vector format in a scale of 1:250,000. Additionally, for the purpose of further analyses, the location of a community was also represented with the use of a centroid defined for every polygon. A centroid is a point within a polygon which represents a geometric calculation of the intuitive 'middle'. In terms of administrative divisions, communities are classified as: urban, rural, and urban-rural communities. For the purpose of this study, the urban and rural communities were left unchanged, but the urban-rural communities were divided into urban and rural parts. The communities comprise 3,045 separate areas within the area of the provinces covered by the analysis.

Descriptive data have been recorded in the form of a relational MySql database and constitute additional information about particular geographic objects. In the presented study, the descriptive data concern the studied group and the population data. The studied group consists of mothers of children with isolated CL±P who were born in 2007 and 2008 in the area covered by PRCM. The criteria of inclusion in the analysis were the completeness of the information concerning the community of a mother's residence and the year of birth. Those conditions were fulfilled by 514 mothers. The point of reference for the conducted analysis was a population of 802,372 live births in particular communities – data provided by the Central Statistical Office [28]. The basic descriptive statistics of the congenital malformation are the coefficients of the prevalences of that malformation in a given year.

Visualization of the spatial distribution of isolated CL±P. Statistical calculations and visualizations of the results on the surface of the map were conducted with the use of the PQStat 1.4.6 programme and (for Kulldorff's statistic) the StatScan v9.1.1 programme.

The suggested ways of presenting the studied phenomenon took into account the important epidemiological problem of personal data protection; the maps therefore had to be compiled in such a way that particular persons could not be identified via their data. For this reason, points were not used to mark the place of residence of particular persons; instead, the maps are coloured in accordance with the coefficients of the prevalence of the defect for aggregated individuals.

A cartogram was used to present the coefficients on the map surface**.** For better legibility of the presented maps, exact values of the coefficients have not been drawn for every community, but for classes of communities. There are many methods of classifying coefficients [29, 30, 31], and in this study the communities have been classified on the basis of their coefficient with the use of statistical formula of natural breaks, i.e. Jenks' Natural Breaks [32] which forms subgroups in an ordered dataset in such a way that the variance of values within each group is minimized.

Very often, especially in analyses of congenital malformations, datasets are encountered in which for many areas the studied group sizes are very small – many of them equal 0. Hence, the calculated coefficients are too sensitive (unstable) to small changes of data. A simple way of coping with such cases has been proposed by Cressi et al. [33]. It involves adding a small number to the size of a group of children with the defect in every community. Another method is smoothing, with the use of locally weighted averages, which reduces the range of raw coefficients and makes it closer to the average. The method is based on a calculated weighted mean for a given community and the neighboring communities. As a result of the addition of weights, the geographically nearest communities have a greatest influence on the calculated smoothed coefficient, and instability is reduced.

Excess risk maps are used to present an increased or decreased risk of the occurrence of a defect, in comparison with the average.

Neighbourhood structure. The neighborhood structure is usually represented by a spatial weight matrix. In the simplest case that is a symmetrical binary 0–1 matrix where one means that two units are neighbors and 0 means lack of a neighborhood. The relationship of neighborhood is defined in different ways, usually by contiguity (a common

boundary) of the spatial units, or in the meaning of distance, in a defined metric (most often, the Euclidean metric or on a sphere/geoid).

In the presented study, a row-standardized spatial weight matrix has been used for the visualization of maps, and in Moran's analysis. The matrix standardization balanced the influence of the communities with various numbers of neighbours on the obtained results. The matrix has been built on the basis of the distance criterion, with the use of the reverse Euclidean distance. It has been assumed that a distance of not more than 30 km between centroids constitutes a neighbourhood. Thus, the closer communities have a greater influence on the obtained result in the analyses, based on that matrix than the further ones, and the influence of the coefficients from communities outside the circle with a 30 km radius equals zero.

Kulldorff's analysis is not directly based on the neighbourhood matrix. For the results obtained with the help of local Moran's statistic and Kulldorff's statistic to be comparable, a similar assumption about the neighbourhood has been made in Kulldorff's analysis. The scanning window has been defined as an ellipse with the maximum length of the minor semi-axis equal to 30 km.

Methods of identifying spatial clusters. The adopted level of test significance for the statistical inference is α=0.05. In the case of multiple comparisons, the level of statistical significance has been corrected according to Bonferroni›s method.

In this study, two methods of identifying spatial clusters have been used: Anselin's LISA [26] and the second based on Kulldorff's scan statistic [27]. The aim of both methods is the same: to determine if a given area, understood as a set of spatial units, is a cluster, that is an area distinguishing itself from the surrounding because of an increased prevalence of the congenital malformation. Nevertheless, the two methods have different assumptions and origins. Anselin's method is derived from spatial econometrics and is used, in principle, for data aggregated into continuous spatial units. Still, on the level of the definition of spatial weight matrices there are no contraindications to use the method for point data. Kulldorff's method represents a geostatistical approach in which point data are most often used. The use and comparison of the two methods in this study is possible because the data have been aggregated to many small spatial units (communities, parts of communities). The data are characterized with relatively large homogeneousness in a territorial system, and can be transformed into a point representation with the use of a centroid, without a loss of generality.

The detection of a statistically significant cluster is the beginning of a long and costly procedure defined by cluster investigation protocols [34]. Each of the statistical methods is based on its own criteria; therefore, the obtained results are not always compatible. This is why it is important to confirm the existence of a cluster by more than one statistical method [35].

Anselin's LISA is based on the analysis of the so-called spatial autocorrelation. This method assumes that spatial autocorrelation is the result of the existence of spatial dependence in the located data. That means that the value of a given variable in a spatial unit depends on the value of that variable in other spatial units and, according to Tobler's law [36], the strongest influence is that of neighbouring observations. Spatial autocorrelation may not occur – in such a case, one speaks of spatial randomness. The obtained spatial distribution is as probable as any other distribution (Fig. 1a). When the neighbuoring values are similar to one another, one can speak about a positive autocorrelation (Fig. 1b). Negative autocorrelation occurs when the neighbouring areas differ more than could be explained by random distribution (Fig. 1c).

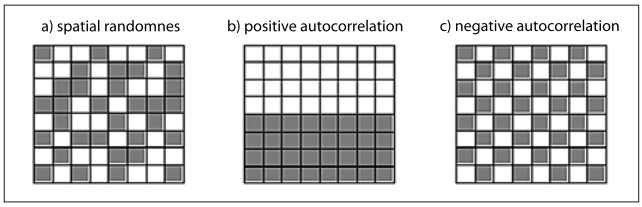

Figure 1. Sample spatial autocorrelation distributions

The most common coefficient in the studies of spatial autocorrelations is the global Moran's coefficient:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

where:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2$$

z_i, z_j – prevalence of isolated CL±P for particular communities;
$\bar{z}$ – value of the expected prevalence (average) of isolated CL±P for the whole analyzed area;
w_{ij} – an element of a matrix of spatial weights;
n – number of communities.

A positive value of the standardized Moran's ***I*** coefficient means the presence of clusters with similar values (hot spots, cold spots), a value near 0 means the lack of autocorrelation, and a negative value means the presence of the so-called outliers, that is decidedly different values in a neighbourhood.

Asumptions of the theory of Local Indicators of Spatial Autocorrelation. A disadvantage of Moran's statistic is that it only provides the estimate of an averaged global pattern of spatial autocorrelation in the studied area, whereas the presence of clusters is the result of the presence of local deviations from the global pattern of spatial autocorrelation which, in turn, are the result of the instability of the strength of spatial dependence, that is its heterogeneousness connected with the local non-stationarity of the analyzed variable. Local Indicators of Spatial Autocorrelation are used for identifying such areas. One of them is the local Moran's I_i statistic [26] which allows the search for clusters by checking if an area is surrounded by neighbours with similar or different values of the coefficient. The value of the statistic of this test is determined for every object (every community) and defined in analogy to its global counterpart. The interpretation is also similar. A positive value of the coefficient means that a community is surrounded by other communities with a similar value of the studied coefficient, whereas a negative value means that the community is surrounded by other communities with significantly different values. The null hypothesis in the local Moran's statistic assumes the lack of spatial autocorrelation (H_0:$I_i = 0$).

Because the local Moran's statistic can be correlated for neighbouring communities, the significance level is corrected by the average number of neighbours [26]. The tools for this are Bonferroni $\alpha_1 = \frac{\alpha}{k}$ or Šidák corrections $\alpha_1 = (1 - (1 - \alpha)^{1/k}$ where k is the average number of neighbours. In this study, the average number of neighbouring communities is 28; therefore, the significance level assumed for the LISA analysis is $\alpha_1 = 0.0018$.

Assumptions of Kulldorff's scan statistic. Kulldorff's scan statistic [37, 27] is another method which allows the identification clusters, and refers both to the popular Openshaw's GAM method [38] and Turnbull's procedure [39]. The general idea is to scan the studied set of spatial units with the help of a window of a predefined shape and maximum size. It was used for the first time in Kulldorff and Nagarwalla's study [36] to determine the clusters of leukemia cases in the state of New York. A review of the studies conducted on the basis of spatial scan statistic can be found in the book in *Spatial Analysis Epidemiology* [40]. Kulldorff also suggested and developed a time-spatial version of scan statistic [41, 42, 43].

The search for clusters with the use of Kulldorff's method is made with the help of a scan window:
- range – for temporal clusters,
- circle or an ellipsis – for spatial clusters,
- cylinder – for spatio-temporal clusters.

Various sizes of the scan window are used, determined with the use of the Euclidean metric. For every position and size of such a window, the observed sizes and the expected sizes, inside and outside the window, are calculated, and the space is scanned in the search for the most probable cluster. Seemingly, the scanning process can be defined as infinite, but it has been restricted to a finished number of steps by adding to the following conditions:
1) limitation of the number of attachment points of the scanning window;
2) limitation of the spatial range to be covered by the window.

These limitations make it possible to intervene in the size and placement of the scan windows, and thus to steer the scanning process at one's will (Fig. 2).

A study of the statistical significance of the scan window is defined by the likelihood ratio test. We test $H_0 : p = q$ where: p – probability of the occurrence of a defect inside the scan window Z, q – probability of the occurrence of a defect outside the scan window Z.

Construction of the test begins with the determination of the likelihood function $L(Z, p, q)$ for a given window Z. The gathered population data and data concerning the congenital malformation allow the use of Poisson's model as the basis for the likelihood function. The likelihood function for the Z window has the form:

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

where:
c – number of children with a congenital malformation inside the scan window;
C – total number of children;
$E(c)$ – expected frequency outside the scan window;
$C\text{-}E(c)$ – expected frequency inside the scan window;

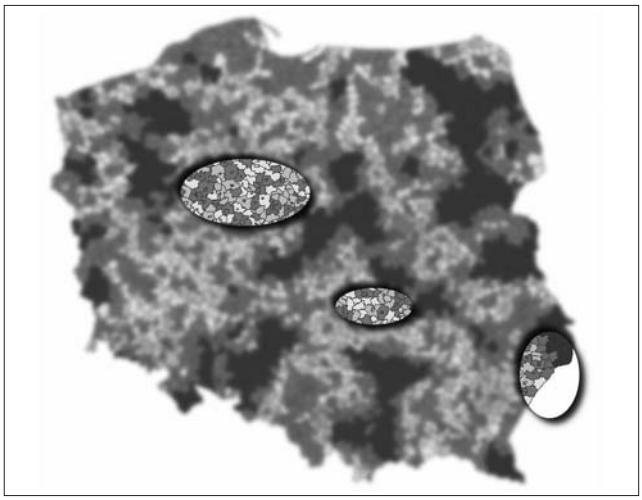

Figure 2. An example of scanning Poland using Kulldorff's scan windows

$I()$ – indicator function equals 1 when the number of cases inside the window is greater than the expected frequency, or equals 0 in the opposite case.

The procedure for determining the likelihood function is repeated for every position and size of the scan window. The most likely cluster $\hat{Z}$ is the scan window for which the maximum of the likelihood function is the highest. That is the cluster with the smallest likelihood of accidental appearance. Statistical significance is defined for that window with the use of the likelihood ratio test.

Having determined the most distinguished cluster, other clusters are searched for. The above procedure is repeated for that goal, with the exclusion of the previously detected cluster from the analyzed area.

Comparison of results obtained by the LISA and Kulldorff methods. The results obtained with the use of the LISA and Kulldorff methods, i.e. clusters with increased prevalence, were compared. The agreement of the assignment of communities to clusters was checked by Cohen's kappa agreement coefficient. The coefficients of the prevalence of the malformation was compared with the Mann-Whitney and Kruskal-Wallis tests, assuming the significance level to be α=0.05.

RESULTS

Visualization of spatial distribution of isolated CL±P. The basic descriptive statistics of the congenital malformation are the frequency coefficients of the prevalence of that malformation in a given area, in this case, in the area of communities.

When presenting the coefficients on a map with the use of a cartogram, 5 or 7 classes are usually distinguished. Here, the distinction was made by assigning communities to one of 5 Classes on the basis of their coefficients, with the use of Jenks natural breaks (Fig. 3).

Figure 4 shows the distinction between the areas previously described as 0, according to the Cressi method, by adding value 0.1 to the number of children with the malformation. This method allows the introduction of a distinction among the areas of small sizes, but the problem of instability remains.

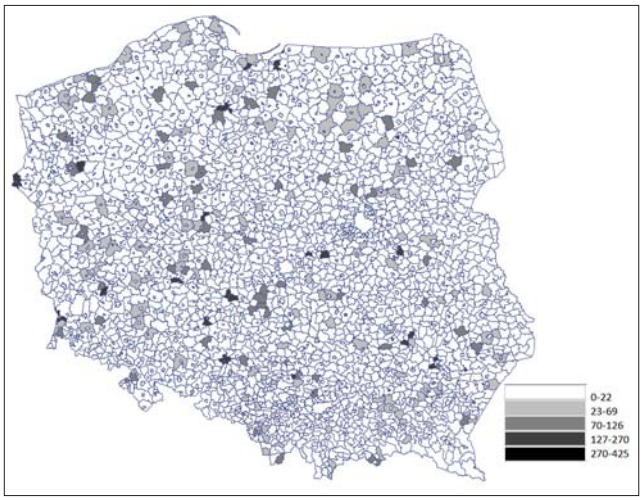

Figure 3. Raw coefficients of isolated CL±P in 2007–2008 in Poland, according to Jenks' natural breaks

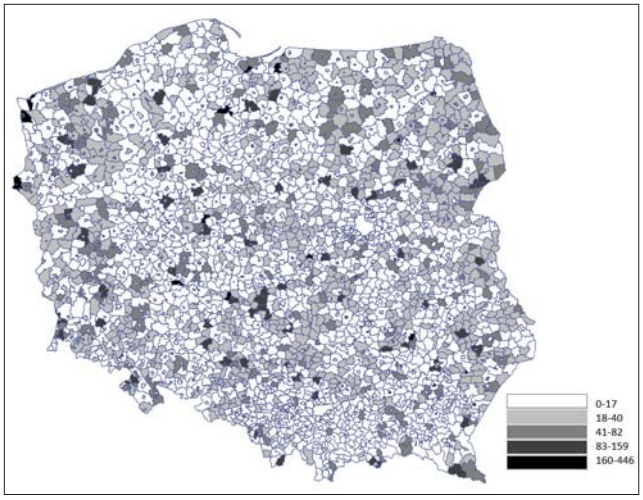

Figure 4. Raw coefficients of isolated CL±P in 2007–2008 in Poland, modified according to the Cressi method; presented according to Jenks' natural breaks

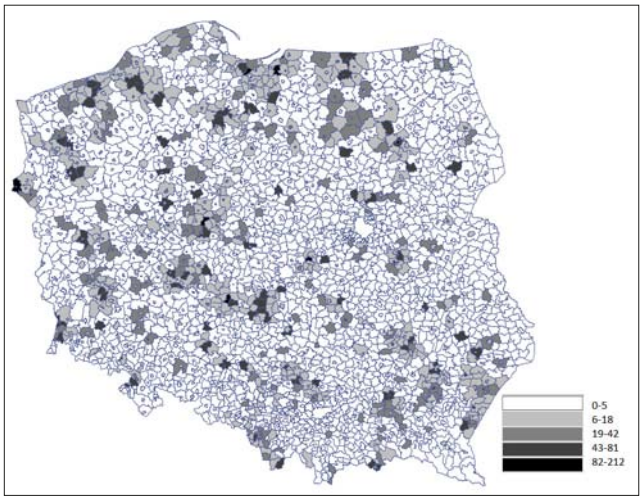

Figure 5. Smoothed coefficients of the prevalence of isolated CL±P in 2007–2008 in Poland, presented according to Jenks' natural breaks

Figure 5 presents the results of the use of smoothing with the use of the weighted averages method.

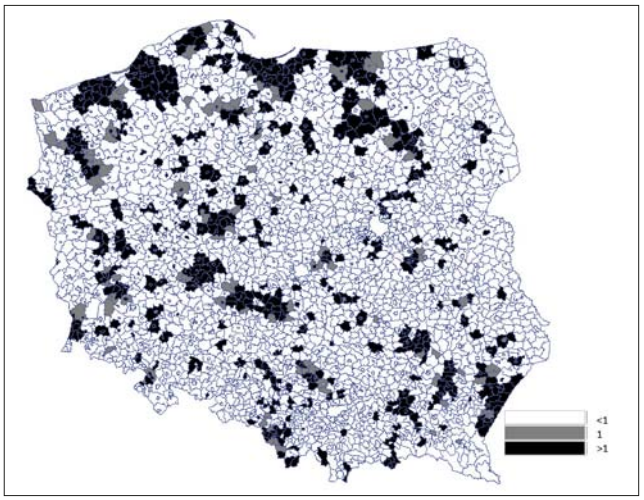

Figure 6. Excess risk map of the presence of isolated CL±P in 2007–2008 in Poland, presented according to Jenks' natural breaks

Excess risk map is presented in Figure 6. The grey colour represents communities in which the coefficient is close to the average coefficient calculated for the whole area. The black colour is for a coefficient higher than the average, and white is for a lower one.

Results of using the theories of Local Indicators of Spatial Autocorrelation (LISA). From the point of view of this study, the interesting clusters are those with a higher frequency of the prevalence of isolated CL±P. It has been checked if the communities with high CL±P prevalence neighbour the communities with high CL±P prevalence. Such communities are marked in black on the map (Fig. 7). When the corrected significance level of $\alpha_1 = 0.0018$ is assumed, the determined communities constitute the centres of clusters with a higher prevalence of isolated CL±P.

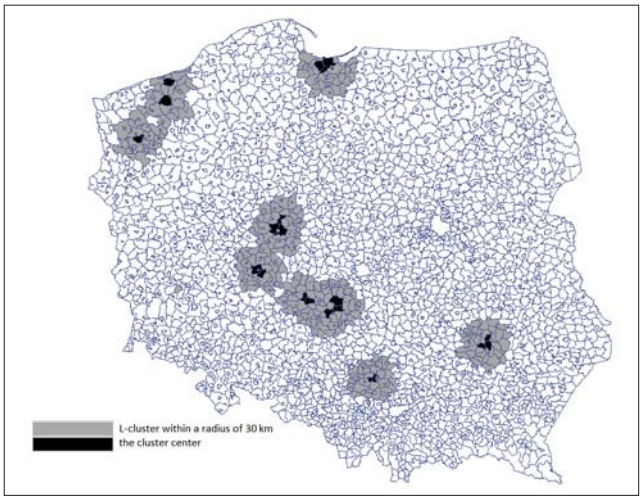

Figure 7. Results of the LISA analysis

24 communities marked in black on the map, determined on the basis of the values of the coefficients of the prevalence of isolated CL±P for those communities, and for the neighbouring communities within the radius of 30 km. This is why zones have been marked out around the 24 communities. The overlapping zones combine, forming homogeneous groups. In this way, 8 consistent, non-

Table 1. Results of Kulldorff's scan analysis

	K-clusters according to the order of their determination						
	K-cluster 1	K-cluster 2	K-cluster 3	K-cluster 4	K-cluster 5	K-cluster 6	K-cluster 7
Community at the centre of the ellipsis	Pyzdry	Bojadła	Stary Targ	Łaskarzew	Rymań	Widawa	Stara Dąbrowa
Province at the centre of the ellipsis	Wielkopolskie	Lubuskie	Pomorskie	Mazowieckie	Zachodniopomorskie	Łódzkie	Zachodniopomorskie
Semi-minor axis	22.6km	22.4km	29.8km	74.4km	23.9km	99.2km	80.8km
Semi-major axis	45.3km	50.7km	89.3km	11.2km	23.9km	19.8km	12.1km
Population	3387	5158	8048	281	1119	263	146
No. of cases	17	20	27	4	7	4	3
Expected frequency	4.34	6.61	10.31	0.36	1.43	0.34	0.19
p-value	0.118 (NS)	0.441 (NS)	0.652 (NS)	0.983 (NS)	0.993 (NS)	0.993 (NS)	0.998 (NS)

NS: lack of statistical significance - value p >= 0.05

overlapping areas with higher prevalence of the isolated CL±P were determined. Further in the study, these areas will be referred to as L-clusters.

Results of the use of the theory of Kulldorff's scan statistic. Looking for other clusters, Kulldorff's method excludes from the analysis those communities which were included in the previous clusters. Hence, the clusters determined with the use of Kulldorff's scan statistic do not overlap.

According to the result of Kulldorff's statistic, the determined clusters do not constitute statistically significant clusters, but only clusters with a somewhat higher prevalence of isolated CL±P (Tab. 1) – there are 7 of them. Further in the study, these places will be referred to as K-clusters.

Comparison of results obtained with the use of LISA and Kulldorff's methods. Comparative analysis of the local methods of searching for clusters includes the areas the only distinction of which is a somewhat higher coefficient of the prevalence of isolated CL±P. Table 2 presents the number of communities inside and outside the clusters determined with the use of the K (Kulldorff's) and L (LISA) methods.

Table 2. Number of communities inside and outside the K and L clusters

No. of communities	Outside the K-clusters	Inside the K-clusters	Total
Outside the L-clusters	2,604	70	2,674
Inside the L-clusters	2,72	99	371
Total	2,876	169	3,045

371 communities belonged to the clusters determined with the use of the LISA method, whereas only 168 communities belonged to the clusters determined with the use of Kulldorff's method. Because each technique is based on its own criteria, the two methods may not always select the same communities for clusters. However, the overlap of the results indicates that both methods illustrate various elements of the same clusters. The agreement of the assignment of communities to clusters was studied by determining Cohen's kappa coefficient on the basis of the results presented in Table 2. The coefficient was kappa (95%CI) =31.44% (26.12%, 36.76%) and statistically significant p <0.00001. This indicates agreement between the two methods.

A comparison of the geographic range of the zones shows that both methods determined a similar area (Fig. 8). Only two L-clusters and one K-cluster do not overlap.

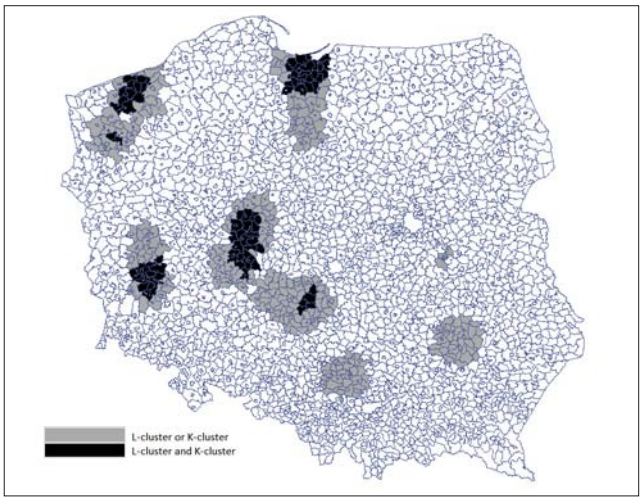

Figure 8. Assignment of communities to particular L-clusters and K-clusters

A comparison of the coefficient of the prevalence of the isolated CL±P among the areas covered by a cluster and the areas outside of the cluster indicates that both methods have shown the right placement of the clusters (Tab. 3). In both cases, the coefficient of the prevalence of the isolated CL±P in clusters is statistically significantly higher than outside the clusters (the mean of the ranks of the communities in the clusters is higher than outside the clusters).

The coefficient of the prevalence of the isolated CL±P among the areas outside the clusters, areas covered by only one cluster (L-cluster or K-cluster), and areas covered simultaneously by L-clusters and K-clusters (Fig. 9), were compared.

The Kruskall-Wallis test indicates the existence of a statistically significant difference between the values of the coefficients for the compared areas (p-value <0.00001). According to a *post-hoc* analysis made with the Dunn test,

Table 3. Results of the Mann-Whitney test corrected for rank ties

K-clusters	mean of the ranks = 1881.91	p<0.00001
Outside the K-clusters	mean of the ranks = 1501.90	
L-clusters	mean of the ranks = 1647.32	p<0.00001
Outside the L-clusters	mean of the ranks = 1505.75	

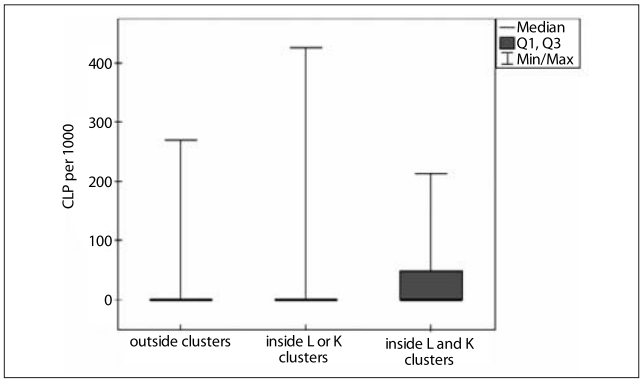

Figure 9. Coefficients of the prevalence of isolated CL±P for communities outside the clusters, in one of the clusters, or in two clusters

the difference is located only between the communities in both clusters and the remaining two groups of communities (Tab. 4).

Table 4. Results of the Dunn test

p-value	Outside clusters	Inside L or K clusters	Inside L and K clusters
outside clusters	–		
inside L or K clusters	0.24120	–	
inside L and K clusters	<0.00001	0.00081	–

DISCUSSION

The presented study represents a geostatistical and econometrical attitude toward the study of congenital malformations, on the example of isolated CL±P. It provides an insight into the possible ways of visualization on maps of frequency coefficients of the prevalence of congenital malformations. It discusses the ways of managing the instability resulting from small frequencies of congenital malformations for areas aggregated to small surfaces. Therefore, it facilitates the reception and understanding of maps by a reader who is neither a statistician nor a geographer.

The maps in the presented study show the areas aggregated to the level of the communities for which the analysis was made. Such a detailed division allows a more precise spatial analysis than the usually suggested analysis, i.e. an analysis limited to the distinction between the urban and the rural areas [20, 21], or to a distinction into other two to four regions, e.g. north, south, east, and west [22, 23].

These studies show that with the aggregation to the level of communities, the prevalence of CL±P in Poland is not completely independent from the location of the mother's place of residence. The indicated locations with an increased prevalence of the malformation are not, however, statistically significant in the light of the two implemented methods. That precludes the existence of a single, homogeneous and strong source of environmental exposure in the years 2007 and 2008 in Poland. Still, it is definitely recommended to systematically monitor the area covered by the Polish Register of Congenital Malformations. It is the more important as a similar tendency to non-homogeneous distribution of the prevalences has also been observed on the territory covered by other registers of congenital malformations. For example, similar, statistically non-significant clusters with higher prevalence of CL±P have been obtained for the state of Utah in the USA [24], and for the Norwegian population [22] where significant differences in the CL±P prevalences have been shown for different regions.

Although the presented study only found areas with higher prevalences of CL±P, it has been determined that the spatial location of the detected L-clusters and K-clusters is similar. It can therefore be assumed that the results obtained with the use of both methods are in agreement. The differences which appear in a more detailed analysis are due to the fact that in Kulldorff's method the manner of searching for clusters allows the finding of its optimum size, whereas in the LISA method that size is assumed in advance. The lack of elasticity of the L-cluster window in this analysis results from the fact that it also includes those communities which, due to a low coefficient, would not otherwise have to be placed in the cluster. This is why, in spite of the agreement as to the spatial location, the different range (size) of the zones determined by the compared methods makes the coefficient of the isolated CL±P of the communities inside the overlapping L-clusters and K-clusters much higher than that of the remaining areas. The coefficients of the communities which only belong to a cluster determined by one of the methods, do not significantly differ from the communities outside the clusters.

When looking for clusters, it is not usually possible to precisely define their size; greater the searched area, the more difficult the task, and the more windows of different sizes can be placed in it. It should to be remembered that each real cluster in the studied area can have a completely different size. It is therefore difficult to determine a size which would be appropriate for every cluster. In some cases, this might be impossible. It is much easier to determine a maximum, meaningful from the point of view of the studied phenomenon, radius to which the window can extend. This is why, from the practical point of view, Kulldorff's method is more universal and precise. Another advantage of the spatial scan statistic is great statistic power, needed to detect the most probable cluster [44, 45], especially when its shape is similar to an ellipsis.

CONCLUSIONS

The study shows that the method which has been comprehensively presented here is useful for analyzing epidemiological data collected by large, long-running medical registers. It gives much more precise information about the spatial pattern of the occurrence of a given malformation than the mere division of a region into urban and rural areas. The presented methods of searching for clusters (LISA and Kulldorff) on the example of the coefficient of the prevalence of the isolated CL±P, consistently pointed to the location of areas with a higher prevalence of CL±P for the area covered by PRCM. However, because the principles of the methods of spatial analysis are different, as are the possibilities of

dividing the analyzed area into, e.g. communities or counties, the results obtained with the use of particular methods with diverse initial settings can vary somewhat. This is why it is important to use more than one method at the same time.

Locating an area with an increased coefficient of isolated CL±P indicates the need for further monitoring by PRCM.

Acknowledgment
Project funded by the National Science Centre. Title: The analysis of congenital malformations prevalence in live born children on the territory of Poland – searching for clusters Number: UMO-2011/01/N/NZ7/02689

REFERENCES

1. Lieff S, Olshan AF, Werler M, Strauss RP, Smith J, Michell A. Maternal cigarette smoking during pregnancy and risk of oral clefts in newborns. Am J Epidemiol. 1999; 150: 683–694.
2. Radojičić J, Tanić T, Radojičić A. Smoking in pregnancy – The risk factor for the development of lip and palate clefts with fetus. Medicine and Biology 2006; 13(1): 44–48.
3. Hackshaw A, Rodeck Ch, Boniface S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. Hum Reprod Update. 2011; 17(5): 589–604.
4. Parkwyllie L, Mazotta P, Pastuszak A, Moretti ME, Beique L, Hunnisett L, et al. Birth defects after maternal exposure to corticosteroids: prospective cohort study and meta-analysis of epidemiological studies. Teratology 2000; 62(6): 385–392.
5. Hill DS, Wlodarczyk BJ, Palacios AM, Finnell RH. Teratogenic effects of antiepileptic drugs. Expert Rev Neurother. 2010; 10(6): 943–959.
6. Munger R, Romitti P, Daack-Hirsch S, Burns T, Murray J, Hanson J. Maternal alcohol use and risk of orofacial cleft birth defects. Teratology 1996; 54(1): 27–33.
7. Lorente C, Cordier S, Goujard J, Ayme S, Bianchi F, Calzolari E, et al. Tobacco and alcohol use during pregnancy and risk of oral clefts. Am J Public Health. 2000; 90(3): 415–419.
8. Laumon B, Martin JL, Collet P. Exposure to organic solvents during pregnancy and oral clefts: a case-control study. Reprod Toxicol. 1996; 10(1): 15–19.
9. Garcia AM, Fletcher T. Maternal occupation in the leather industry and selected congenital malformations. Occup Environ Med. 1998; 55(4): 284–286.
10. Bove FJ, Fulcomer MC, Klotz JB, Esmart J, Dufficy EM, Savrin JE, 1995. Public drinking water contamination and birth outcomes. Am J Epidemiol. 1995; 141(9): 850–862.
11. Vinceti M, Rovesti S, Bergomi M, Calzolari E, Candela S, Campagna A, et al. Risk of birth defects in a population exposed to environmental lead pollution. Sci Total Environ. 2001; 278(1–3): 23–30.
12. Hansen CA, Barnett AG, Jalaludin BB, Morgan GG. Ambient air pollution and birth defects in brisbane, Australia. PLoS One. 2009; 4(4):e5408.
13. Vrijheid M, Martinez D, Manzanares S, Dadvand P, Schembari A, Rankin J, et al. Ambient air pollution and risk of congenital anomalies: a systematic review and meta-analysis. Environ Health Perspect. 2011; 119(5): 598–606.
14. Ritz B, Yu F, Fruin S, Chapa G, Shaw GM, Harris JA. Ambient air pollution and risk of birth defects in Southern California. Am J Epidemiol. 2002; 155(1): 17–25.
15. Gilboa SM, Mendola P, Olshan AF, Langlois PH, Savitz DA, Loomis D, et al. Relation between ambient air quality and selected birth defects, seven county study, Texas, 1997–2000. Am J Epidemiol. 2005; 162(3): 238–252.
16. Dolk H, Vrijheid M, Armstrong B, Abramsky L, Bianchi F, Garne E, et al. Risk of congenital anomalies near hazardouswaste landfill sites in Europe: the EUROHAZCON study. Lancet. 1998; 352(9126): 423–427.
17. Van Rooij IA, Ocké MC, Straatman H, Zielhuis GA, Merkus HM, Steegers-Theunissen RP. Periconceptional folate intake by supplement and food reduces the risk of nonsyndromic cleft lip with or without cleft palate. Prev Med. 2004; 39(4): 689–694.
18. Wilcox AJ, Lie RT, Solvoll K, Taylor J, McConnaughey DR, Abyholm F, Vindenes H, et al. Folic acid supplements and risk of facial clefts: national population based case-control study. BMJ. 2007; 334(7591): 464–469.
19. Materna-Kiryluk A, Więckowska B, Wiśniewska K, Borszewska-Kornacka MK, Godula-Stuglik U, Limon J, et al. Maternal reproductive history and the risk of isolated congenital malformations. Paediatr Perinat Epidemiol. 2011; 25(2): 135–143.
20. Bing-Fang Hwang, Jouni JK Jaakkola. Ozone and Other Air Pollutans and the risk of Oral Cleft. Environ Health Perspect. 2008; 116(10): 1411–1415.
21. Messer LC, Luben TJ, Mendola P, Carozza SE, Horel SA, Langlois PH. Urban-rural residence and the occurrence of cleft lip and cleft palate in Texas, 1999–2003. Ann Epidemiol. 2010; 20(1): 32–39.
22. Harville EW, Wilcox AJ, Lie RT, Vindenes H, Abyholm F. Cleft Lip and Palate versus Cleft Lip Only: Are They Distinct Deffects? Am J Epidemiol. 2005; 162(5): 448–453.
23. Reddy SG, Reddy RR, Bronkhorst EM, Prasad R, Ettema AM, Sailer HF, et al. Incidence of cleft Lip and palate in the state of Andhra Pradesh, South India. Indian J Plast Surg. 2010; 43(2): 184–189.
24. Gebreab SY. Spatial Epidemiology of Birth Defects in the United States and the State of Utah Using Geographic Information Systems and Spatial Statistics. All GraduateTheses and Dissertations 2010; Paper 852.
25. Więckowska B, Materna-Kiryluk A, Kossowski T, Moczko J, Wiśniewska K, Latos-Bieleńska A. Location of cleft lip with or without cleft palate prevalence clusters using Kulldorff Scan Statistics. Comput Meth Sci Technol. 2012; 18(1): 53–62.
26. Anselin L. Local indicators of spatial association – LISA. Geogr Anal. 1995; 27(2): 93–115.
27. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods. 1997; 26(6): 1481–1496.
28. Bank Danych Lokalnych www.stat.gov.pl (access: 2012.05.12) (in Polish).
29. Lai PC, So FM, Chan KW. Spatial epidemiological approaches in disease mapping and analysis. CRC Press, 2009.
30. Anselin L, Kim YW, Syabri I. Web-based analytical tools for the exploration of spatial data. J Geogr Syst. 2004; 6: 197–218.
31. Boscoe FP, McLaughlin C, Schymura MJ, Kielb CL. Visualization of the spatial scan statistic using nested circles. Health Place 2003; 9(3): 273–277.
32. Jenks GF. The Data Model Concept in Statistical Mapping. International Yearbook of Cartography 1967; 7: 186–190.
33. Cressi NA. Statistics for Spatial Data. Revised edition. John Wiley, 1993.
34. Cluster Investigation Protocols. http://www.eurocat-network.eu/clustersandtrends/clusteradvisoryservice/clusterinvestigationprotocols (access: 2012.12.06).
35. Hanson CE, Wieczorek WF. Alcohol mortality: a compari- son of spatial clustering methods. Social Science and Medicine 2002; 55 (5): 791–802.
36. Tobler W. A computer movie simulating urban growth in the Detroit region. Econ Geogr. 1970; 46(2): 234–240.
37. Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and inference. Stat Med. 1995; 14(8): 799–819.
38. Openshaw S, Charlton ME, Wymer C, Craft A. A mark I geographical analysis machine for the automated analysis of point data sets. International Journal of Geographical Information System 1987; 1: 335–358.
39. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. Am J Epidemiol. 1990; 132(1): 136–143.
40. Stevenson M, Stevens KB, Rogers DJ, Clements ACA, Pfeiffer DU, Robinson TP. Spatial Analysis in Epidemiology. Oxford University Press, 2008.
41. Kulldorff M, Athas WF, Feuer EJ, Miller BA, Key CR. Evaluating cluster alarms. A space-time scan statistic and brain cancer in Los Alamos, New Mexico. Am J Public Health. 1998; 88: 1377–1380.
42. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. Journal of the Royal Statistical Society Series A 2001; 164(1): 61–72.
43. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med. 2005; 2(3): e59.
44. Kulldorff M, Tango T, Park P. Power comparisons for disease clustering tests. Comput Stat Data Anal. 2003; 42: 665–684.
45. Song C, Kulldorff M. Power evaluation of disease clustering tests. Int J Health Geogr. 2003; 2(1): 9.