

Article

Exploratory Spatial Data Analysis of Congenital Malformations (CM) in Israel, 2000–2006

Keren Agay-Shay ^{1,*}, Yona Amitai ², Chava Peretz ³, Shai Linn ^{1,4}, Michael Friger ⁵ and Ammatzia Peled ⁶

¹ Faculty of Social Welfare and Health Sciences, School of Public Health, University of Haifa, Mount Carmel, Haifa 31905, Israel

² Department of Management, Bar Ilan University, Ramat Gan 52900, Israel; E-Mail: yonaamitai89@gmail.com

³ Department of Epidemiology, Faculty of Medicine, Tel Aviv University, P.O. Box 39040, Ramat Aviv 69978, Israel; E-Mail: cperetz@post.tau.ac.il

⁴ Faculty of Welfare and Health Sciences, University of Haifa, and Unit of Clinical Epidemiology, Rambam Medical Center, P.O. Box 9602, Haifa 31096, Israel; E-Mail: slinn@univ.haifa.ac.il

⁵ Department of Epidemiology and Health Services Evaluation, Faculty of Health Sciences, Ben-Gurion University of the Negev, P.O. Box 653, Beersheba 84105, Israel; E-Mail: friger@bgu.ac.il

⁶ Department of Geography and Environmental Studies, Faculty of Social Sciences, University of Haifa, Haifa 31905, Israel; E-Mail: peled@geo.haifa.ac.il

* Author to whom correspondence should be addressed; E-Mail: kagayshay@gmail.com; Tel.: +972-4-678-9821.

Received: 10 January 2013; in revised form: 11 March 2013 / Accepted: 13 March 2013 /

Published: 19 March 2013

Abstract: Congenital Malformations (CM) impose a heavy burden on families and society. Identification of spatial patterns of CM is useful for understanding the epidemiology of this public health issue. In Israel, about 1,000,000 births and 25,000 CM cases at 37 groups were geocoded during 2000–2006. These were geo-analyzed using global-Moran's-I statistics. Eight groups demonstrated geospatial heterogeneity and were further analyzed at both the census tract (Local Indicator of Spatial Association (LISA) and hot spot analyses) and street levels (spatial scan statistics with two population threshold sizes). The positional definition of results is further discussed in relevance to possible exposure to teratogenic sources in the region. Limitations of data and methods used are presented as well.

Keywords: birth defects; congenital malformations; epidemiology; spatial clusters/statistics; LISA; hot spot; SaTScan

1. Introduction

Congenital Malformations (CM), as defined by the March of Dimes Birth Defects Foundation, refer to any malformation, functional or structural, that presents in infancy or later in life and is induced by events preceding birth, whether inherited or acquired. Varying from minor cosmetic irregularities to life-threatening disorders, CM are the major causes of infant mortality in the world and a leading cause of disability [1]. CM impose a heavy burden on families and society both in terms of direct and indirect costs [2]. Birth and birth defects registrations enable the detection of spatial variation in the incidence of CM in the population.

The application of Exploratory Spatial Data Analysis (ESDA) is one of the more commonly used spatial techniques in initial studies involving health-related issues [3,4]. The utility of ESDA tools allows researchers to detect spatial patterns, identify local variability in health data, and assess the efficiency of spatial models [5,6]. In particular, the ESDA includes cluster analyses, which enable researchers to identify geographically correlated group occurrences that are unlikely to have occurred by chance [7]. Although the identification of clusters does not provide a causal explanation for the spatial pattern and spatial clustering of specific health-related events, working hypotheses are often more easily generated and can be subsequently tested in future research efforts [8]. Associations between environmental hazards and the occurrence of CM may be detected by seeking evidence of non-random occurrence of cases (spatial clusters). In general, there are two types of spatial clustering methods: (i) Spatial global clustering methods and (ii) local cluster detection methods [9]. The former method tests for the presence of spatial clustering in the whole study area but does not provide any information as to location. The latter identifies a spatial cluster with excess disease risk without previous knowledge of either how many or where they are located. Among the most used method for spatial clustering and cluster analyses at the area level is Moran's *I* statistic [10]. Methods for the identification of local clusters are local indicators of spatial association (LISA) [3] and Getis and Ord's local *G_i* (d) statistic (Hot Spot analysis) [11]. These are widely used to identify local clusters by measuring spatial autocorrelation and by comparing disease risks between neighboring areas respectively, while the spatial scan statistics [12] are widely applied in the detection of local clusters of disease without any pre-selection bias.

Table 1 lists published studies demonstrating spatial analysis of CM, applying various geo-statistical methods, geocoding approaches, different malformation groups and population size. In Israel, research on spatial distributions of health outcomes is limited and it is not routinely reported, with the exception of sub-districts based analysis [13] and census tract based analysis [14] of the Israel National Cancer Registry. Only one study investigated CM spatial distribution in Israel in an ecological study with regards to proximity to the National Toxic Waste Site, and major CM rate ratios based on settlements' name and wind directions [15].

Table 1. List of published studies using different geospatial approaches demonstrating spatial analysis of congenital malformation (CM).

Reference Name	CM Subtypes	Statistical Method	Cluster Analysis Unit	N Cases	Covariates/Complimentary Data to Live Births
Armstrong <i>et al.</i> 2007 [16]	Combined chromosomal (n = 3) and non-chromosomal (n = 27) anomalies A total of 31 classifications as “subtypes”	Spatial scan statistics. Negative binomial model adjustment.	Residential addresses of mothers classified into areas at 4 hierarchical levels using their postcode of address at delivery. From the smallest, these were census enumeration districts, electoral ward, hospital catchment area, and register	10,722 cases	Maternal age, Carstairs index for deprivation
Chi <i>et al.</i> 2008 [17]	Total CM	Spatial scan statistic, Poisson-based model	Centroid of the village	127 cases	-
Forand <i>et al.</i> 2002 [18]	Surveillance malformations (Total CM well diagnosed)	Spatial scanning analysis SaTScan, Poisson-based model	Centroids of the ZIP codes	24,394 cases included only singleton births	-
Kuehl and Loffredo, 2006 [19]	Hypoplastic left heart syndrome	Spatial scanning analysis SaTScan, Poisson-based model	A census block group address- the address of the mother conception was the address used	77 Hypoplastic left heart syndrome cases	Maternal diabetes, medications, number of previous pregnancies, pre-pregnancy weight, weight gain, infant birth weight and gestational age, family income and parental education, Maternal/Paternal solvent exposure
Liao <i>et al.</i> 2011 [20]	Neural tube defects	Hierarchical Bayesian model, and Comparison between methods spatial filtering method	The location of 326 villages.	97 Neural tube defects cases	Not specified
Moris <i>et al.</i> 1998 [21]	Down Syndrome	Knox Methods	Non spatial statistical analysis. Regional Health Authority level	8,000 cases	Maternal age, prenatal care, Abortions
Poletta <i>et al.</i> 2007 [22]	Cleft lip with/without, and Cleft palate only	Spatial scanning analysis SaTScan	Hospital, country.	5,128 Cleft lip with/without, 1,745 Cleft palate only	Maternal, Paternal age, maternal and paternal education, gravidity, parental consanguinity
Root <i>et al.</i> 2007 [23]	Gastroschisis	Spatial scan statistic	Residential address at birth-aggregation by blocks, block groups and tracts.	242 cases- excluded multiple births	Maternal age, Race, Parity, Medicaid status, maternal smoking
Rushton and Lolonis 1996 [24]	CM, all births in the area	Spatial filtering Spatial scan, Punctual Kriging	Specific Street and number address Control address in a town address.	2,406 cases	Maternal age
Siffel <i>et al.</i> 2006 [25]	CM	Automated Spatial Surveillance Program, kernel smoothing	Specific address	49 cases	Not specified

The goal of this study is to describe and analyze the spatial patterns and clusters of CM cases in Israel during a period of seven years (2000–2006) and to compare the analysis results achieved by using different geo-statistical approaches.

2. Results and Discussion

2.1. Global Moran I

The results of the “global Moran I ” analyses for all the 37 CM categories mentioned above are summarized in Table S1, supplementary material. This was the first step of the analyses aimed to detect those CM categories that should be further analyzed. This resulted in eight CM subtypes that depicted heterogeneity and thus were further investigated. These eight CM subtypes are presented in Table 2 that depicts subtypes of CM with statistical significant global Moran I results ($p < 0.05$). These subtypes are: (1) four congenital heart defect types; (2) two male-genital malformations; (3) polydactyly and (4) Down syndrome. In addition, the whole group of congenital heart defects yielded a statistical significant global Moran I result.

The circulatory system was the anatomic system with most of the CM subtypes demonstrating geospatial heterogeneity (four subtypes). The highest mean rate ratio was observed for ventricular septal defect (VSD) not included in the conotruncal category (146 VSD cases for 10,000 births). The highest Moran’s I value (0.16) was observed for atrial septal defects (ASD), followed by cryptorchidism (0.03) (See Table 2).

2.2. Local Clustering Analyses

Three different methods were used to detect local cluster and one of them included two different thresholds. This resulted with many analyses products demonstrating areas with higher than expected CM cases. Figure 1 demonstrates the Israeli fifteen administrative sub-districts. Only the main results will be discussed, and all the results are presented in Figures 2–4. Figure 2 and 3 illustrate the spatial analyses results based on census tract level for the eight CM subtypes. Figure 4 illustrates the spatial analyses results based on street level for CM subtypes with 1% and 10% birth population size thresholds.

In general, congenital heart defects analyses resulted with the highest rates comparing to the national average, in the Southern Region of the country near Beersheba, around Ramat-Hovav and in the Tel Aviv and the Central sub-districts. Specifically, ASD and VSD malformation subtypes, depicted high ratio clusters at the Central and Southern Regions (Figure 5(a,b)). In addition, VSD also demonstrated a cluster in the northern region (Figure 6(b)). Two major congenital heart defects, tricuspid anomalies and hypoplastic left heart syndrome (HLHS), depicted a spatial cluster of higher incidence rate, around Ramat-Hovav, in the Southern Region (Figure 5(c,d)). Our observations for heart malformations were consistent with two previous works. CM total rates analyzed for districts heterogeneity, demonstrated highest rate of congenital heart defects in the Southern District at the Beer Sheba and Ashkelon sub-districts (in comparison to the national average) during 2001 and 2005 [26].

Table 2. Street level and census tract level descriptive statistics and global Moran *I* autocorrelation of unadjusted rates, Israel, 2000–2006.

Anatomic System	CM Included	ICD-10	Street Based Address-Total Counts	Census Tract Based Address			Global Moran <i>I</i>			
				Total Counts	Mean Unadjusted Rate Ratio (* 10,000 Births)	Standard Division Unadjusted Rate Ratio (* 10,000 Births)	Maximum Unadjusted Rate Ratio (* 10,000 Births)	Moran's Index	Z Score	<i>p</i> -value
Circulatory system	Tricuspid anomalies	Q22.8, Q22.4, Q22.9, Q22.5	365	275	4	20	303	0.01	6.41	0
	Atrial and atrial septal defects	Q21.1, Q24.2, Q26.2	4,651	3,334	48	110	3,333	0.16	74.53	0
	Hypoplastic left heart syndrome	Q23.4	136	98	1	7	2,200	0.01	2.81	0.005
	Ventricular septal defect not included in the conotruncal category	Q21, Q21.0, Q21.2	5,266	3,824	57	146	3,333	0.02	10.79	0
	All Heart CM	Q20–Q25	17,294	12,929	187	301	6,667	0.09	41.52	0
Male-genital organs	Cryptorchidism	Q53, Q53.0, Q53.1, Q53.2, Q53.9	2,485	1,905	25	65	1,052	0.03	11.99	0
	Hypospadias	Q54, Q54.0, Q54.1, Q54.2, Q54.3, Q54.8, Q54.9	2,593	1,937	26	71	2,000	0.01	3.05	0.002
Chromosomal CM	Down syndrome	Q90.0, Q90.2, Q90.9	1,048	726	9	67	2,500	0.01	2.93	0.003
Musculo-skeletal system	Polydactyly	Q69.0, Q69.1, Q69.2, Q69.9,	760	578	6.5	29.7	526.3	0.01	2.51	0.012
TOTAL BIRTHS		1,000,470	740,899	283	596	9,090				

* CM—congenital malformations.

All other CM types were not previously studied with regards to spatial distribution or heterogeneity. Therefore, we could not compare our results with previous results.

Clusters of male genital malformations (cryptorchidism and hypospadias) were found in Jezreel and Acre sub-districts, of the Northern District of Israel. Also, this was observed at the Rehovot sub-districts, of the Central District (Figures 2, 3(e–f) and 4(d,e)).

Clusters of Down syndrome (Figures 2, 3(g) and 4(f)) were found at the Jewish orthodox religious communities in Jerusalem and Bnei Brak (Figures 5(g) and 6(f)). A third cluster of the Down syndrome was detected at the vicinity of Safed in the Northern part of Israel (Figure 6(g)). This third cluster around Jewish religious communities was detected also by the Hot spot and LISA, at the census tract level.

A cluster of Polydactyly (Figures 2, 3(f) and 4(g)) was observed in the Southern District of Israel, at the kibbutzim near the Gaza strip (Figure 5(h)). In addition, clusters were found in the Acre and Safed sub-district, in the Northern District of Israel (Figure 6(h)).

Figure 1. Israeli fifteen sub-districts as index map.



Figure 2. Census tract level analysis Local Indicator of Spatial Association (LISA) unadjusted rates, Israel, 1999–2006 for: (a) atrial septal defects; (b) ventricular septal defect; (c) hypoplastic left heart syndrome; (d) tricuspid anomalies; (e) hypospadias; (f) cryptorchidism; (g) Down syndrome; (h) polydactyly.

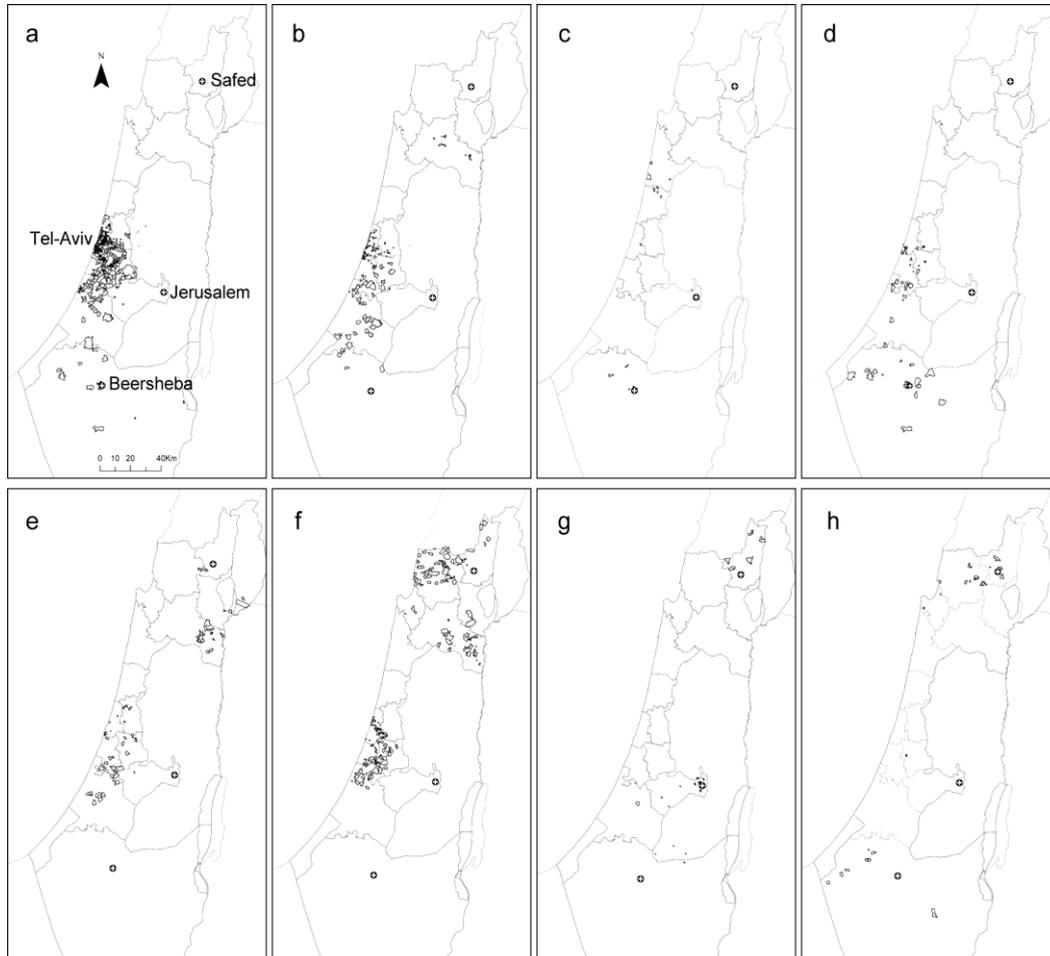


Figure 3. Census tract level analyses Hot Spot unadjusted rates, Israel, 1999–2006 for: (a) atrial septal defects; (b) ventricular septal defect; (c) hypoplastic left heart syndrome; (d) tricuspid anomalies; (e) hypospadias; (f) cryptorchidism; (g) Down syndrome; (h) polydactyly.



Figure 3. Cont.

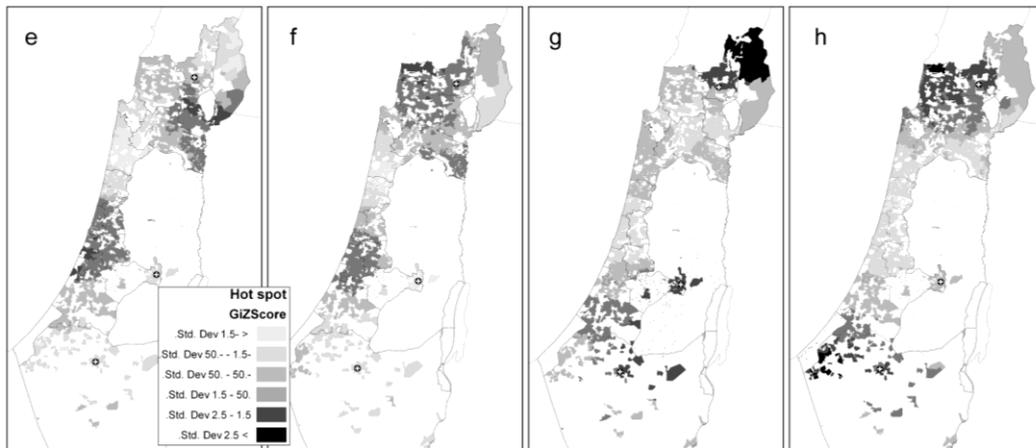


Figure 4. Street level analyses using SaTScan methods, 1% and 10% population size, Israel, 1999–2006 for: (a) atrial septal defects; (b) ventricular septal defect; (c) tricuspid anomalies; (d) hypospadias; (e) cryptorchidism; (f) Down syndrome; (g) polydactyly.

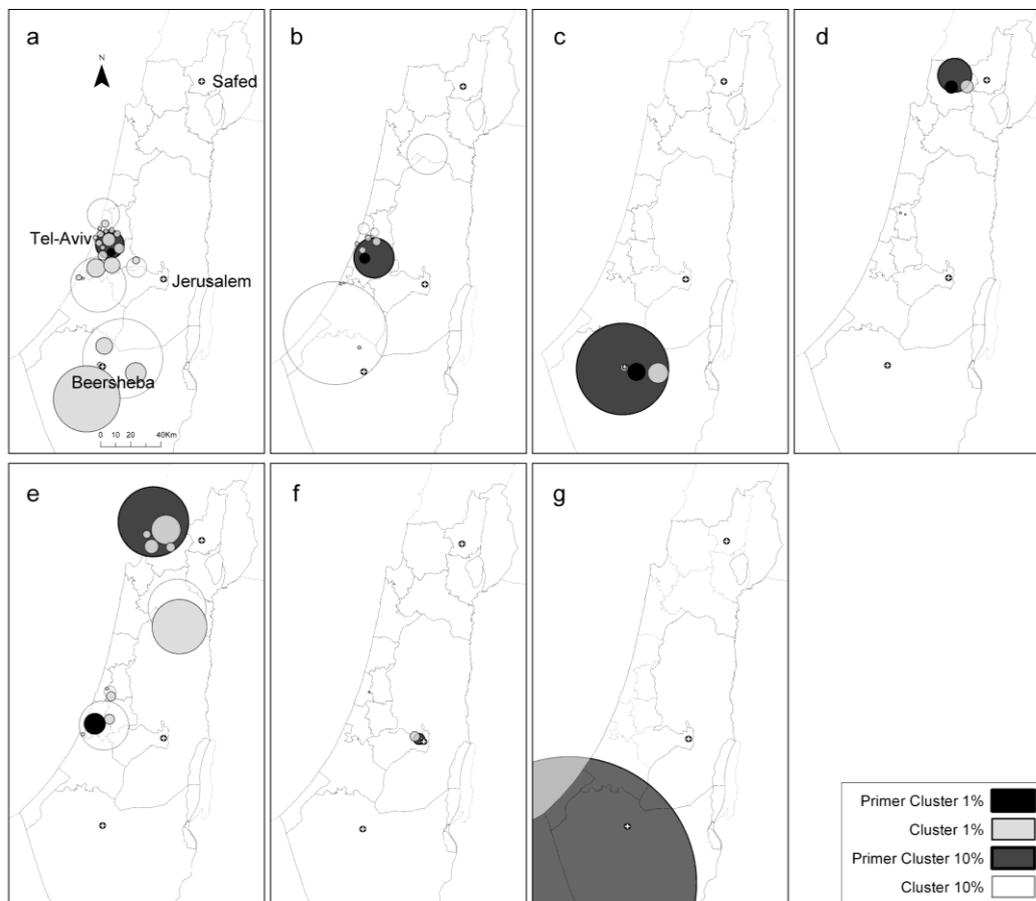


Figure 5. Comparison between methods, clusters demonstrating consistency between methods, Israel, 1999–2006 for: (a) atrial septal defects; (b) ventricular septal defect; (c) tricuspid anomalies; (d) hypospadias, (e,f) cryptorchidism; (g) Down syndrome; (h) polydactyly.

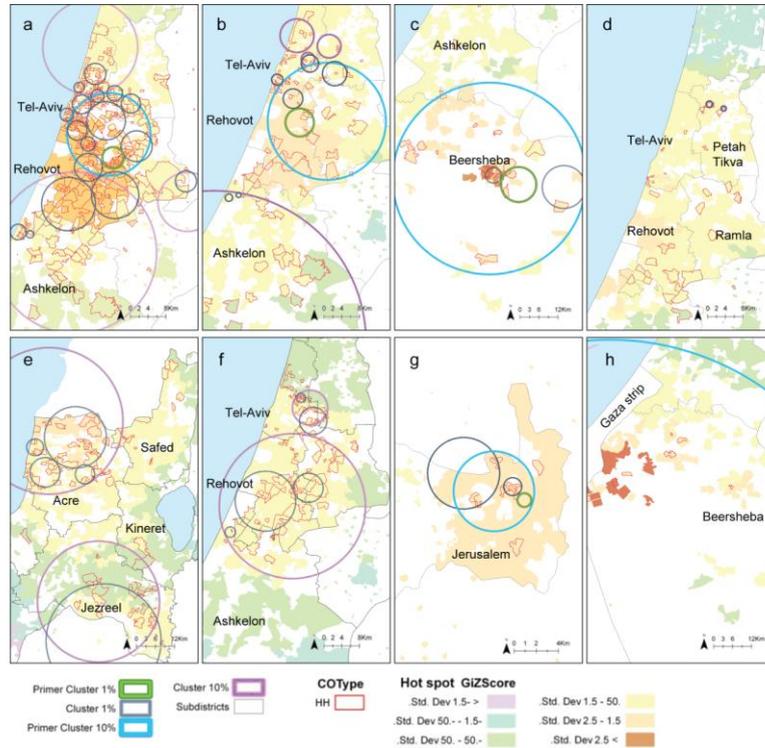
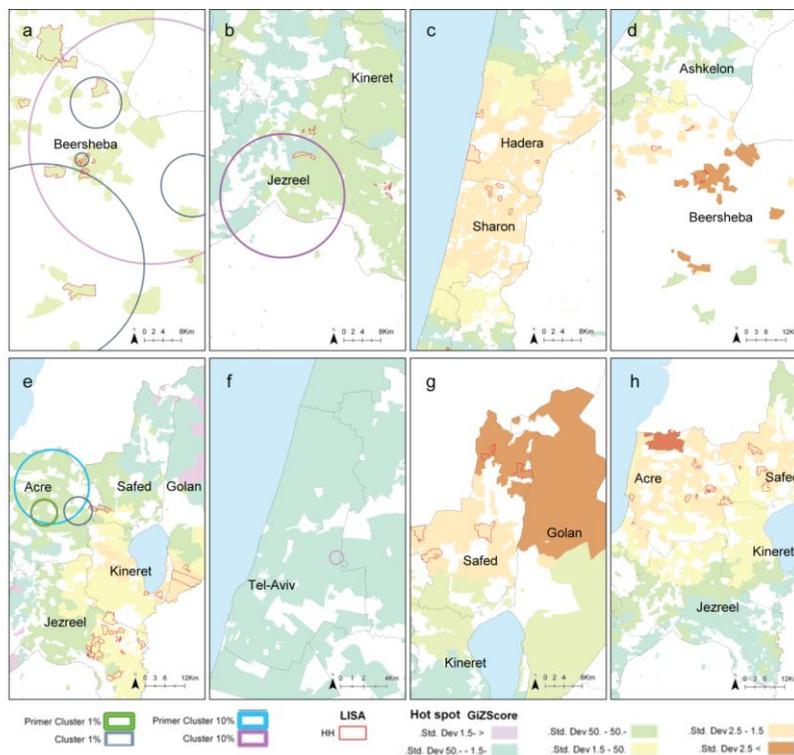


Figure 6. Comparison between methods and clusters demonstrating inconsistency between methods, Israel, 1999–2006 for: (a) atrial septal defects; (b) ventricular septal defect; (c,d) hypoplastic left heart syndrome; (e) hypospadias; (f,g) Down syndrome; (h) polydactyly.



2.3. Comparison between the Local Clustering Methods

Comparison between the different methods revealed mainly that they all defined the same geographic locations and depicted the same geospatial distribution of CM clusters. In addition, it was observed that the SaTScan clusters of 1% population threshold size were nested within the cluster of 10% threshold size. However, some of the CM subtypes presented variation in the CM cluster positioning related to the different methods used. Except for hypospadias, all regions demonstrating consistency between the different methods were those areas where the primary cluster was detected by the SaTScan method.

The ASD and VSD congenital heart malformations, demonstrated a cluster in the Rehovot area at the Central District. This was true for all methods (Figure 5(a,b)). However, the ASD cluster at the Southern District, detected by all other methods, was not detected by the Hot spot methods (Figures 6(a)). In addition, the VSD Northern cluster was only detected by the SaTScan method with 10% population threshold size and LISA (Figure 6(b)). The LISA and Hot Spot, the two census tract level methods, demonstrated consistency in positioning the HLHS clusters (Figure 6(c,d)). However, the street level analyses, using SaTScan did not reveal any comparable statistically significant results. Tricuspid anomalies presented a cluster in Beer-Sheba sub-district at the Southern District, regardless of the method used (Figure 5(c)).

Male genital malformations demonstrated similar spatial distribution as defined by all methods. However, the results for cryptorchidism were more consistent comparing the different methods. Hypospadias demonstrated consistency in the Central District (Figure 5(d)). However, inconsistency was observed between the methods based on a census tract level and those on street level results (Figure 6(e)). All methods demonstrated consistency in defining cryptorchidism geospatial clusters. The results overlapped each other in all three main areas (Figure 5(e)).

All methods demonstrated consistency in positioning the Down syndrome cluster in the city of Jerusalem (Figures 4(f) and 5(g)). Yet, this consistency was not demonstrated at the city of Bnei Brak where the cluster was detected only by SaTScan method (Figure 6(f)) and another cluster in the Golan area, detected only by the Hot spot analysis (Figure 6(g)).

Consistency between the different methods was demonstrated in positioning the polydactyly cluster in the area near the Gaza Strip (Figure 5(h)). In contrary, the polydactyly cluster in the Northern District was detected only by census tract level analysis methods (Figure 6(h)).

2.4. Discussion

Four different spatial analysis methods were used to investigate and illustrate the spatial variability of CM prevalence in Israel, at both the census tract and the street levels. In addition, the significant positive spatial autocorrelation identified by Moran's I statistics in our birth cohorts, as well as significant local clusters identified by LISA and hot spot analysis, confirm the presence of spatial heterogeneity in unadjusted CM rates [3]. Eight subtypes of CM, with non-random spatial heterogeneity were observed in Israel during the years 2000–2006. Of these, four were congenital heart defect types, two were male-genital malformations; and in addition polydactyly and Down Syndrome.

The application of ESDA to identify spatial patterns and clustering of events is important. This is not routinely conducted in Israel. ESDA provides important information for the development and refinement of geographically and population specific prevention programs to reduce CM risks. In addition, this information is useful to public health planners because many current policies and health initiatives are principally based on the assumption of spatial homogeneity. Furthermore, public health administrators and planners may use maps to point out problems of health care delivery systems and to determine where underserved communities might exist.

Statistical tests for cluster detection can be useful for health researchers in order to focus their attention on specific areas in terms of geographic locations for further investigation. Geo-spatial clusters are also useful for presenting general patterns and trends in the disease distribution. It has a tremendous importance and a proved potential for generating hypotheses regarding the role of environmental, genetic or lifestyle factors in the etiology of a disease. For example, congenital heart defect (CHD) clusters detected in the Tel Aviv District which have the highest levels of automobile ambient air pollution, were further analyzed to examine the association between personal maternal exposure, air pollution and CHD risk [27].

Some of the clusters that have been detected by using different methods might be explained by sociological differences in the spatial non-homogenous distribution of the Israeli populations. For example, hypospadias and cryptorchidism clusters that were detected in the Northern District at the Akko sub-district. This area is populated with Druze and Muslim villages, which demonstrate higher rates of cryptorchidism and hypospadias in the national level. This higher proportion of male genital malformations are due to unexplained higher secondary sex ratio (male to female proportion at births) for Druze and Muslim that differ from the Jewish population [27], in addition to a high rate of consanguinity. Clusters of Down syndrome were also observed in the ultra-orthodox Jewish population centers in Israel (Safed, Jerusalem and Bnei-Brak). Currently, the Israeli CM National Registry includes only live and still births and not early termination of pregnancy. Many CM are identified through prenatal screening. The prenatal identification of a serious CM allows the choice to electively terminate the pregnancy. Rates of prenatal diagnosis of malformations vary among populations [28,29]. In addition, even when prenatal screening results are known, the ultra-orthodox Jewish women are less likely to elect termination of pregnancy.

Polydactyly is in general a “genetic” malformation with some well known mutations that cause this abnormality [30]. In addition there is very little published literature with regards to associations between environmental exposure and polydactyly [31]. Two regions with higher rates of polydactyly were detected in Israel. The Northern cluster might be explained by higher consanguineous marriage rates within the Druze and Muslim populations. However, the cluster observed in the southern part of the country cannot be explained by these “traditional” behaviors.

Other spatial clusters might be explained by possible differences in reporting rate. Under-reporting of CM by some Israeli hospitals may lead to apparent clusters in other areas of the country where reporting is more complete, as has been previously demonstrated in New York city [18]. VSD and cryptorchidism are two CM types that are not mandatory to be reported by law. Therefore, the cluster in the central region of the country, mainly in the Tel-Aviv District, might be explained by this information bias.

Some of the clusters that have been reported above might be associated with environmental exposures. ASD, tricuspid anomalies and HLHS demonstrated spatial clusters in the Southern District, around Ramat-Hovav. Ramat-Hovav is a regional industrial park hosting 17 chemical plants and serves as the National Industrial Toxic Waste dump site. Previous work examined the possible association between major CM and residence near this location. It was reported that during 1995–2000 there was a statistically significant increased risk for major CM even after stratification for at-risk populations [15]. There are some published reports demonstrating higher incidence rate of CHD and residence near landfill sites [32–34].

The Northern District and especially the more rural areas of Jezreel, Kinneret and Golan sub-districts have been previously reported with most intensive pesticides use in a survey conducted in 1998. VSD, hypospadias and cryptorchidism demonstrated clusters in these areas as well. The associations between pesticides exposure and increased risk for hypospadias and cryptorchidism are well reported [35]. This is not the case for VSD.

Tricuspid, VSD, ASD and cryptorchidism, all demonstrated clusters at the Rehovot sub-district. This area is known to have many research centers (biological, chemical and nuclear energy). This observation, a result of the geo-spatial analyses made in this research was not reported previously. The nuclear energy research center at Soreq, which began operating in 1960, will be replaced by 2017 or 2018 by a particle accelerator [36]. It will be of great interest to learn whether the rate of CM will decrease in this area.

The Israeli Birth Defect Registry is unique in a sense that only three people are involved in coding the birth defects registrations and during the study period, it was done by a single person. This minimizes errors that might occur due to differences in coding pattern. Spatial locations of CM clusters may vary depending on the spatial scale and units used in the analyses. This is commonly known as the modifiable areal unit problem (MAUP) or “ecological fallacy” [37]. One of the limitations in this study is the fact that a “point-in-polygon” approach was used to allocate street level geocode to a census tract level. This might cause possible errors in allocation that result from inaccurate geocoding results and might introduces a potential bias. In Israel, small settlements in rural areas are labeled by one unique census tract. A high percentage of the streets in Israel cities are relatively short and they all belong to one census tract. Only in the rare cases of streets extended over two or three such tracts, the geocoding of full addresses (both street name and house number) allowed an accurate point to point calculations and point to polygon unique relations to the local census tract. Only those addresses with the street name, but without the number were related to the “central” census tract related to the streets central point. In the small settlements, this brought no major spatial bias, as the distances are short. In big cities, the percentage of such inaccuracies was very small. Thus, this limitation had very little influence as well. To partly address the “ecological fallacy” problem in the study, spatial analyses were conducted at both the census tracts level and at the streets level. Additional bias filtering and screening were conducted by using different thresholds as the maximal boundary values of the population size in the cluster analyses. This practice led to the detection of nested clusters, as previously demonstrated for Autism research [38]. The differences between the clusters that have been detected, by using the different methods, might be due to the pre-assumption about the aggregation units and the statistical distribution of CM.

CM types, which demonstrated clusters in only one method, were those CM types that presented a lower p-value in the global Moran *I* test; these demonstrated less cases than those in which overlap between the methods was full (Table 2). This means that when a very heterogeneous spatial distribution with more cases is present, all methods show the finding of a cluster in the same place. However, when the CM are rare, the use of different approaches and various methods allows for “fishing out” the undetectable clusters.

3. Experimental Section

3.1. Data Source and Study Population

Data on 1,001,525 births population during 2000–2006 were extracted from the National Birth and Birth Defect Registry of the Department of Mother Child & Adolescent Health in the Public Health Service of the Israeli Ministry of Health. The data set includes all CM identified and reported live births, stillbirths and terminations of pregnancy following prenatal diagnosis. Spontaneous abortions and fetal deaths prior to 23 weeks’ gestational age are absent. In addition, identification of CM correlates with fetal death or terminated pregnancy with CM was impossible. CM diagnoses are reported by hospital physicians and are coded by staff members at the Departments of Mother Child & Adolescent Health and Community Genetics using standardized coding (ICD-10 codes).

Only CM from a defined, recorded list were included in the research dataset (see Table S1, supplementary material). Also included were minor malformations, which are not systematically recorded as these were reported in previous works to be associated with environmental exposures (for example cryptorchidism [39]). The classification used for these analyses comprised 33 classifications, referred as “subtypes” of CM, and 4 more aggregated categories. In total, all the CM cases were grouped in 37 categories. The total population of CM cases was 26,427 (26.4 per 1,000 livebirths).

3.2. Ethics in Publishing

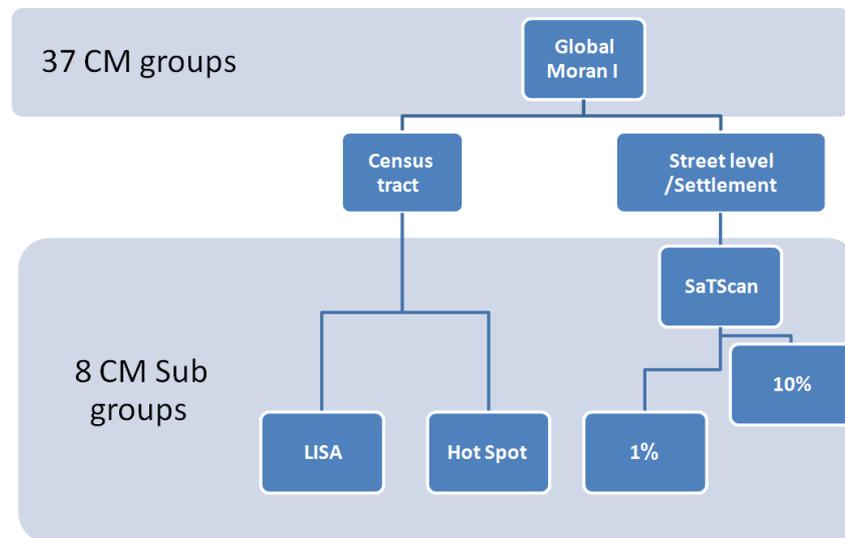
The study was approved by the Ethics Committee of the University of Haifa. Further approvals from the director of Public Health Services and from the legal advisor in the MOH were also obtained. The investigators were provided with the CM data set only after deletion of all personal identifiers (e.g., ID numbers, name of mother).

3.3. Study Design

The first step of the study was to geocode all births addresses, this discussed in details at Section 3.5. To describe and analyze the spatial patterns and clusters of CM cases in Israel, a two-step analytical process was used (Figure 7). The first step involved a global Moran *I* statistics by which all 37 CM categories were analyzed (33 subtypes and 4 aggregated groups). Those CM subtypes demonstrating geospatial heterogeneity (four congenital heart defect types, two male-genital malformations, polydactyly and Down syndrome) were further analyzed to define the location of this spatial heterogeneous and high risk clusters at both the census tract and street levels. Two methods were used at the census tract level: the Local Indicator of Spatial Association (LISA) statistics together with Hot

Spot analyses. The Spatial scan statistics (SaTScan) was used at the street level with two different threshold values of 1% and 10% of the total births population size.

Figure 7. Course of study.



3.4. Geospatial Analyses

3.4.1. Global Moran *I*

The Moran's *I* statistics was used as a measure of the global heterogeneous and assessed by testing the null hypothesis that the spatial patterns of CM cases ratio (CM cases divided by the number of live and stillbirths) were random [3]. Therefore, rejection of the null hypothesis implies a non-random spatial pattern that is also referred to as spatial autocorrelation. More specifically, positive spatial autocorrelation indicates similar values that occur at adjacent locations; whereas negative autocorrelation implies that high values appear next to low values. Moran's *I* ranges approximately from +1 (for positive spatial autocorrelation) to -1 (negative autocorrelation), and its expected value in the absence of autocorrelation approximates zero. Global Moran *I* was conceptualized by Inverse distance and Euclidean distance methods band was used as the distance method. A multiple testing problem arises that usually requires prescription of a more stringent significance level [40], *i.e.*, p -value < 0.05. Nevertheless, as these tests are used only as a pre-filter for elimination of implausible cases, it was decided to use p -value < 0.05 and thus to unduly elevate the chances of detecting spatial autocorrelations and increase the probability of a type II error. CM subtypes, demonstrating geospatial heterogeneity were further analyzed at both the census tract and street levels.

3.4.2. Local Indicator of Spatial Association (LISA)

Anselin's Local Indicator of Spatial Autocorrelation (LISA) was used as an indicator of local spatial association that measures whether or not the CM rate for a particular spatial polygonal unit (e.g., a census tract) is closer to the values of a neighboring unit or to the average of the study area [3]. LISA was conceptualized by Inverse distance and Euclidean distance methods band was used as the

distance method. The results of these analyses yielded five categories of spatial units at the census tract level. These categories were defined as “high-high (HH)”, “low-low (LL)”, “high-low (HL)”, “low-high (LH)” and “not significant (NS)”. The HH category indicates clustering of high values of unadjusted CM rates, while the LL category indicates clustering of low values of unadjusted CM rates. These are equivalent to a positive spatial autocorrelation. In addition, the HL category indicate units with high values that are adjacent to polygonal units with low values of the unadjusted CM rates, while the LH category indicates that units with low values are adjacent to units with high values of unadjusted rates of CM. These are equivalent to a negative spatial autocorrelation. Lastly, the NS category indicates that there is no statistically significant spatial autocorrelation.

3.4.3. Hot Spot Analysis

Hot spot analysis was also conducted producing cluster maps illustrating the local spatial patterns. Cluster detection techniques such as the Getis-Ord G_i^* statistics [11] provides a statistically robust way to identify areas that share similar characteristics by identifying clusters of census tracts with magnitude values higher than might be expected by random chance. The local $G_i^*(d)$ statistic (local G-statistic) is used to test the statistical significance of local clusters, and to determine the spatial extent of these clusters [41]. Hot spot was conceptualized by fixed distance and Euclidean distance methods band was used as the distance method. The local G-statistic is useful for identifying individual members of local clusters by determining the spatial dependence and relative magnitude between an observation and neighboring observations [42].

3.4.4. Spatial Scan Statistics (SaTScan)

Kulldorff's Spatial Scan Statistic [12] reduces the problem of cluster detection to a problem of maximum likelihood estimation over geographical space. Through a single hypothesis test over space, the method was used to determine areas (all over the country), which had statistically significant high CM counts (adjusted for total births) in comparison to the national average. By carrying out just one hypothesis test over the entire geographical space, this method involves no problem of multiple hypothesis testing. This technique has been used previously (see Table 1). The method uses overlapping circles of various different sizes and locations to identify areas with high CM rate in comparison to the expected rate (based on the national average). The population-weighted centroids of the streets or settlements served as the centers for the overlapping circles starting with a radius of zero in length and increasing until a user defined maximum population restriction is met. The test uses the maximum likelihood ratio test statistics. For each circle, the likelihood of finding the observed number of CM inside and outside the circle, respectively, is calculated. The likelihood formula assumes that CM cases and total births are distributed as a Poisson random variable and the likelihood ratio is compared to simulated likelihood ratios generated by Monte Carlo randomizations of the data to assess statistical significance. The circle with the rate least likely to have occurred by chance is then determined and secondary clusters with high rate values are also located. Circles with high CM rates were determined through the Monte Carlo hypothesis testing and the statistical significance was adjusted for multiplicity. These circles were then embedded in a GIS layer to be later visually displayed. The choice of an appropriate maximum population restriction size is important. The

maximum allowable population size was limited to 1% and 10% of the Israel birth population. A 1% upper bound on cluster size provides location certainty without compromising statistical power, it is comparable to census tract (in terms of the population size) and it allows zooming in on specific neighborhoods. Larger clusters are more robust, but provide lesser geographic information.

3.5. Research Process

Due to privacy and confidentiality concerns, maternal residential address was provided at the street name level only in settlements with more than 5,000 residents. In settlements smaller than 5,000 residents, the address is based on settlement names, only. These were positioned at the street or settlement center points. The geocoding algorithm was embedded into newly developed software and was integrated in the generic ArcView and ArcInfo GIS software programs (ESRI, Redlands, CA, USA). In the first step of automatic address matching, 62% of the CM cases and 64% of the total population were positioned successfully. The addresses that were not matched automatically were then gradually positioned in interactive sessions of discrete steps using different geocoding services including: Google Earth and Microsoft Bing Maps. While this may not be the gold standard for geocoding, it did utilize a combination of high-quality proprietary street files and satellite imagery to locate addresses and are available at no cost. The addresses were matched to the centroid of the street or the neighborhood level (number of streets and settlements = 21,320). After all these steps, the remaining unmatched addresses were geocoded to the settlement level address. These automatic and interactive processes resulted with a database that included 26,010 CM cases of the original 26,427 CM cases. This achievement of 98.4% success in the CM cases address matching was repeated in a 99.8% success in the total births addresses matching (1,000,470 in total of the 1,001,525 births).

The global Moran's I , local Moran's I and hot spot geospatial analyses are based on single value polygonal schema (number of polygons = 3,203). For these purposes, all the cohort population addresses that were related only to single census tracts were matched in a point to polygon GIS process. Each census tract in Israel was assigned with relevant values of CM cases, non-cases and the ratio. This point-to-polygon process was conducted separately for each CM categories, resulting with 37 layers of different GIS CM ratio. About 70% of the CM cases and total population that were geocoded on the street level were referred to census tract level by point-to-polygon process. This resulted in 24,538 matched addresses of CM cases and 740,899 of the non-cases population.

4. Conclusions

This comparative study for CM spatial distribution in Israel is the first one to examine an approximately one million births cohort during seven years of the study. In summary, we found evidence of significant local clusters in eight subtypes of CM. The use of three different local methods in two aggregation approaches allowed detecting excess areas of interest, which in some cases was not detected using only a single method. The use of spatial scan statistics in more refined geographical scale (street vs. census tract) allowed for detecting significant clusters even though they were sometimes different in precise location. Using two definitions of the maximum population threshold size also revealed further insights regarding the locations of clusters.

Some of these clusters may be explained by spatial sociological heterogeneity of the Israeli population. Others might be explained by differences in detecting or reporting rate of CM to the Israeli Birth Defect Registry. A potentially interesting and relevant future comparison would be between the results presented here with those obtained from using statistical scanning approach (such as SaTScan) with poissonic regressions adjusted for all possible socio-demographic risk factors.

Nevertheless, some of the clusters that have been detected may not be explained by intrinsic differences in the population and might be due to environmental etiology. Such as the CHD cluster detected at the main metropolitan center of Israel; the male genital malformation detected at areas with highest pesticides use; and the clusters in Rehovot that might be associated with the cluster of research centers. Some of these observations, such as the spatial clusters at Rehovot region for four different CM subtypes, were never reported before. Additional future work could systematically investigate this by using personal exposure approaches to understand the spatial associations reported in this study.

It is important to note that these methods of spatial analysis need to be conducted at real time as part of regular health data management.

Acknowledgments

The authors would like to acknowledge the dedicated assistance and contribution of Bashir Haj-yehia to this work. We thank Noga Yoselavich for her help in producing the figures. We also would like to thank Joel Zlotogora for his devoted work and exceptional efforts on the national database of congenital malformations in Israel and to Ziona Haklai for the management of the computerized CM database. Also we would like to thank to all employees of the Department of Mother Child and Adolescent Health at the Ministry of Health for their daily work of data collection.

This work was supported by the Israeli Ministry of Environmental Protection (research grant number-7-2-7) and by the Environment and Health fund (Ph.D. Fellowship program).

References

1. Christianson, A.; Howson, C.P.; Modell, B. *March of Dimes—Global Report on Birth Defects*; White Plains: New York, NY, USA, 2006.
2. Davies, K. Economic costs of childhood diseases and disabilities attributable to environmental contaminants in Washington State, USA. *Ecohealth* **2006**, *3*, 86–94.
3. Anselin, L. Local indicators of spatial association—Lisa. *Geogr. Anal.* **1995**, *27*, 93–115.
4. Crighton, E.J.; Elliott, S.J.; Moineddin, R.; Kanaroglou, P.; Upshur, R.E.G. An exploratory spatial analysis of pneumonia and influenza hospitalizations in Ontario by age and gender. *Epidemiol. Infect.* **2007**, *135*, 253–261.
5. Kieffer, E.; Alexander, G.; Lewis, N.; Mor, J. Geographic patterns of low-birth-weight in Hawaii. *Soc. Sci. Med.* **1993**, *36*, 557–564.
6. Poulou, T.; Elliott, S.J. An exploratory spatial analysis of overweight and obesity in Canada. *Prev. Med.* **2009**, *48*, 362–367.
7. Auchincloss, A.H.; Gebreab, S.Y.; Mair, C.; Diez Roux, A.V. Review of spatial methods in epidemiology, 2000–2010. *Annu. Rev. Public Health* **2012**, doi: 10.1146/annurev-publhealth-031811-124655.

8. Curtis, A.; Leitner, M. *Geographic Information Systems and Public Health: Eliminating Perinatal Disparity*; IRM Press: Hershey, PA, USA, 2006.
9. Besag, J.; Newell, J. The detection of clusters in rare diseases. *J. Roy. Statist. Soc. Ser. A Stat.* **1991**, *154*, 143–155.
10. Moran, P.A. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23.
11. Getis, A.; Ord, J. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **1992**, *24*, 189–206.
12. Kulldorff, M. A spatial scan statistic. *Commun. Stat. Theor. Method.* **1997**, *26*, 1481–1496.
13. Barchana, M.; Liptshitz, I.; Fishler, Y.; Green, M. *Geographical Mapping of Malignant Diseases in Israel 2001–2005*; Israel National Cancer Registry: Jerusalem, Israel, 2007.
14. Eitan, O.; Yuval; Barchana, M.; Dubnov, J.; Linn, S.; Carmel, Y.; Broday, D.M. Spatial analysis of air pollution and cancer incidence rates in Haifa Bay, Israel. *Sci. Total Environ.* **2010**, *408*, 4429–4439.
15. Bentov, Y.; Kordysh, E.; Hershkovitz, R.; Belmaker, I.; Polyakov, M.; Bilenko, N.; Sarov, B. Major congenital malformations and residential proximity to a regional industrial park including a national toxic waste site: An ecological study. *Environ. Health* **2006**, doi:10.1186/1476-069X-5-8.
16. Armstrong, B.G.; Dolk, H.; Pattenden, S.; Vrijheid, M.; Loane, M.; Rankin, J.; Dunn, C.E.; Grundy, C.; Abramsky, L.; Boyd, P.A.; *et al.* Geographic variation and localised clustering of congenital anomalies in Great Britain. *Emerg. Themes Epidemiol.* **2007**, doi:10.1186/1742-7622-4-14.
17. Chi, W.; Wang, J.; Li, X.; Zheng, X.; Liao, Y. Analysis of geographical clustering of birth defects in Heshun county, Shanxi province. *Int. J. Environ. Health Res.* **2008**, *18*, 243–252.
18. Forand, S.P.; Talbot, T.O.; Druschel, C.; Cross, P.K. Data quality and the spatial analysis of disease rates: Congenital malformations in New York State. *Health Place* **2002**, *8*, 191–199.
19. Kuehl, K.S.; Loffredo, C.A. A cluster of hypoplastic left heart malformation in Baltimore, Maryland. *Pediatr. Cardiol.* **2006**, *27*, 25–31.
20. Liao, Y.; Wang, J.; Wu, J.; Wang, J.; Zheng, X. A comparison of methods for spatial relative risk mapping of human neural tube defects. *Stoch. Environ. Res. Risk Assess.* **2011**, *25*, 99–106.
21. Morris, J.K.; Alberman, E.; Mutton, D. Is there evidence of clustering in Down syndrome? *Int. J. Epidemiol.* **1998**, *27*, 495–498.
22. Poletta, F.A.; Castilla, E.E.; Orioli, I.M.; Lopez-Camelo, J.S. Regional analysis on the occurrence of oral clefts in South America. *Amer. J. Med. Genet. A* **2007**, *143A*, 3216–3227.
23. Root, E.D.; Meyer, R.E.; Emch, M.E. Evidence of localized clustering of gastroschisis births in North Carolina, 1999–2004. *Soc. Sci. Med.* **2009**, *68*, 1361–1367.
24. Rushton, G.; Lolonis, P. Exploratory spatial analysis of birth defect rates in an urban population. *Stat. Med.* **1996**, *15*, 717–726.
25. Siffel, C.; Strickland, M.J.; Gardner, B.R.; Kirby, R.S.; Correa, A. Role of geographic information systems in birth defects surveillance and research. *Birth Defects Res. A. Clin. Mol. Teratol.* **2006**, *76*, 825–833.
26. Canaku, D.; Tulchinsky, T.H.; Neumark, Y. Regional Analysis of Congenital Malformation in Israel, 2001 and 2005. Master Thesis, Hebrew University, Jerusalem, Israel, 2007.
27. Agay-Shay, K.; Peretz, C.; Friger, M.; Peled, A.; Amitai, Y.; Linn, S. Congenital Malformation and Air Pollution in Israel, 2000–2006. Ph.D. Thesis, Haifa University, Haifa, Israel, 2013.

28. Zlotogora, J.; Haklai, Z.; Rotem, N.; Georgi, M.; Rubin, L. The impact of prenatal diagnosis and termination of pregnancy on the relative incidence of malformations at birth among Jews and Muslim Arabs in Israel. *Isr. Med. Assoc. J.* **2010**, *12*, 539–542.
29. Apple, A. *Pre-Natal Diagnostic of CM and Pregnancy Termination among Orthodox Women*; Knesset Research and Information Center: Jerusalem, Israel, 2003.
30. Gurrieri, F.; Kjaer, K.; Sangiorgi, E.; Neri, G. Limb anomalies: Developmental and evolutionary aspects. *Am. J. Med. Genet.* **2002**, *115*, 231–244.
31. Engel, L.S.; O'Meara, E.S.; Schwartz, S.M. Maternal occupation in agriculture and risk of limb defects in Washington State, 1980–1993. *Scand. J. Work Environ. Health* **2000**, *26*, 193–198.
32. Elliott, P.; Richardson, S.; Abellan, J.J.; Thomson, A.; de Hoogh, C.; Jarup, L.; Briggs, D.J. Geographic density of landfill sites and risk of congenital anomalies in England. *Occup. Environ. Med.* **2009**, *66*, 81–89.
33. Dolk, H.; Vrijheid, M.; Armstrong, B.; Abramsky, L.; Bianchi, F.; Garne, E.; Nelen, V.; Robert, E.; Scott, J.E.; Stone, D.; *et al.* Risk of congenital anomalies near hazardous-waste landfill sites in Europe: The EUROHAZCON study. *Lancet* **1998**, *352*, 423–427.
34. Dummer, T.J.; Dickinson, H.O.; Parker, L. Adverse pregnancy outcomes near landfill sites in Cumbria, northwest England, 1950–1993. *Arch. Environ. Health* **2003**, *58*, 692–698.
35. Yiee, J.H.; Baskin, L.S. Environmental factors in genitourinary development. *J. Urol.* **2010**, *184*, 34–41.
36. Reuters Israel to phase out civilian atomic reactor by 2018. *Ynetnews* 20/3/2012. Available online: <http://www.ynetnews.com/articles/0,7340,L-4205587,00.html> (accessed on 16 July 2012).
37. Elliott, P. *Spatial Epidemiology: Methods and Applications*; Oxford University Press: Oxford, UK, 2000.
38. Mazumdar, S.; King, M.; Liu, K.; Zerubavel, N.; Bearman, P. The spatial structure of autism in California, 1993–2001. *Health Place* **2010**, *16*, 539–546.
39. Main, K.M.; Skakkebaek, N.E.; Virtanen, H.E.; Toppari, J. Genital anomalies in boys and the environment. *Best Pract. Res. Clin. Endoc. Metab.* **2010**, *24*, 279–289.
40. Garcia, L. Escaping the Bonferroni iron claw in ecological studies. *Oikos* **2004**, *105*, 657–663.
41. Ord, J.K.; Getis, A. Local spatial autocorrelation statistics—Distributional issues and an application. *Geogr. Anal.* **1995**, *27*, 286–306.
42. Getis, A.; Morrison, A.; Gray, K.; Scott, T. Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru. *Amer. J. Trop. Med. Hyg.* **2003**, *69*, 494–505.