



PERGAMON

Health & Place 8 (2002) 191–199

HEALTH
& PLACE

www.elsevier.com/locate/healthplace

Data quality and the spatial analysis of disease rates: congenital malformations in New York State

Steven P. Forand^{a,*}, Thomas O. Talbot^a, Charlotte Druschel^b, Philip K. Cross^c

^a *Geographic Research and Analysis Section, Bureau of Environmental and Occupational Epidemiology, New York State Department of Health, 547 River Street, Room 200, Troy, NY 12180, USA*

^b *Congenital Malformations Research and Surveillance Section, Bureau of Environmental and Occupational Epidemiology, New York State Department of Health, USA*

^c *Congenital Malformations Registry, Bureau of Environmental and Occupational Epidemiology, New York State Department of Health, USA*

Received 1 January 2001; accepted 1 September 2001

Abstract

Spatial analyses of disease rates are increasing as the hardware and software used in disease surveillance and cluster investigations become more accessible and easier to use. The results of these analyses should be interpreted with caution since inconsistencies in health outcome reporting and population estimates may lead to erroneous conclusions. In this report we provide an example, using data on congenital malformations in New York State, to show how under-reporting of malformations by some New York City hospitals can lead to apparent clusters of malformations in other areas of the state where reporting is more complete. We illustrate how spatial analysis techniques can be used to locate under-reporting problems and determine the extent to which the problem exists. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Data quality; Spatial analysis; Geographic information systems; Congenital malformations; Birth defects

Introduction

Geographical information systems (GIS) are being used with increasing frequency to analyze the spatial distribution of disease (Richards et al., 1999a, 1999b; Higgs and Gould, 2001). As mapping and spatial analysis software become easier to use, public health professionals increasingly employ these tools to aid in disease surveillance and cluster investigations. However, caution must be taken when interpreting maps of disease and the results of spatial analyses. Errors in health outcome data, vital statistics data and population estimates used in calculating rates of disease, may lead health researchers to draw erroneous conclusions regarding the spatial distribution of a disease. Too often, these types of routinely collected data are taken at

face value with little attention paid to possible errors in the data. Researchers must be prepared to critically evaluate the quality of the data used in mapping rates of disease and determining areas of statistically significant excesses or deficits.

Health outcome data available to researchers come from a variety of sources including disease and vital statistics registries, managed care data, and hospital admission and discharge records. Under-reporting or misreporting of health outcome data often leads to inaccuracies in local disease rates. Other types of geographic bias can be introduced through the use of automated geocoding software, since the success of the software in locating an address is not equally distributed throughout a population (Gregorio et al., 1999). Health outcomes may also appear to cluster due to the presence of medical care facilities in an area (Goodman et al., 1997; Gregorio et al., 2000). For instance, proximity to hospitals has been shown to influence the rate of hospital utilization among nearby residents (Goodman

*Corresponding author. Tel.: +1-518-402-7950; fax: +1-518-402-7959.

E-mail address: spf02@health.state.ny.us (S.P. Forand).

et al., 1997). Furthermore, availability of sophisticated diagnostic equipment, physician judgment, and differences in screening rates, may also cause variations in local disease rates. Whether undertaking cluster examinations or conducting routine surveillance using spatially referenced data, these sources of potential bias should be taken into consideration.

Errors in underlying population or denominator estimates may also create problems when examining rates of disease. To calculate disease rates, health and population data are often aggregated into areal units such as postal zones, census areas or political subdivisions. The population counts for these areas are often calculated based on census counts or on intercensal population estimates. Since census counts are not taken every year, error can be introduced into rates due to inaccuracies of these intercensal population estimates. There is also a seasonal variation of populations in some communities that can lead to inaccurate estimation of disease rates (Boscoe and McLaughlin, 2000). In addition, systematic miscounts of certain populations can also result in bias when calculating rates. It has been well documented that the under count of inner city, minority populations is far greater than for the general population (Fein, 1990; Hahn, 1992; Raleigh and Balarajan, 1994).

While classic incidence and mortality analyses have focused on areas with elevated disease rates it is also important to examine areas with lower than expected rates. Deficiencies in reporting of disease in one area may lead, not only to a perception of lower rates in that area, but also to an incorrect inference of high rates in other geographic areas where reporting and case ascertainment are more complete. This paper provides an example of how the quality of data can have an impact on disease patterns. We evaluate congenital malformation data in New York State using spatial analysis to determine if the quality of the data varies across the state and how this affects local rates. We also show how GIS and spatial analysis techniques can be used to detect potential deficiencies in health outcome surveillance systems. This analysis was conducted as part of a New York State Department of Health (NYSDOH) initiative to enhance birth defect surveillance through data quality improvements and more comprehensive reviews of the spatial patterns of birth defects in New York State.

Methods

Source of data

The New York State Congenital Malformations Registry (CMR) served as the source for congenital malformation reports. The CMR is one of the largest

population-based birth defects registries in the United States. The registry receives reports on approximately 10,000 children per year with major congenital malformations. Major malformations are considered to be those that have an adverse effect on an individual's health, functioning, or social acceptability. Hospitals and physicians are required by law to report all children who are diagnosed by the age of 2 yr with a major structural or functional abnormality. Approximately, 99 percent of the reports are sent by hospitals and the remainder are sent by individual physicians.

We studied a subset of malformations reported to the CMR in children born in New York State between 1992 and 1995. The malformations included were those thought to be easily recognizable and consistently and accurately diagnosed by 1 yr of age by physicians (see Appendix A). Excluded from the analysis were malformations that required substantial judgment on the part of the attending physician and those that would not normally be identified in a standard physical examination (Holmes, 1999). The remaining subset consisted of approximately half of all malformations reported to the Registry, which we shall refer to as surveillance malformations. We included only singleton births in the analysis since multiple births are not independent events and are widely recognized to be at increased risk of malformations.

A total of 24,394 records were selected for analysis. US Postal Service delivery areas, which are identified as ZIP codes, served as the unit of analysis. Records with invalid or missing ZIP codes were removed from the analysis, which represented <0.03 percent of the surveillance malformation records. Since each record represents a single malformation, an infant with multiple malformations will have more than one record. Therefore, the subset was further restricted to include only one record for each infant, leaving a total of 16,378 infants born with one or more surveillance malformations. We then aggregated the records by the ZIP code of the infant's residence at birth. The CMR routinely matches infants' malformation reports to their birth certificates. Over 99.5 percent of these records had been successfully matched to their original birth certificates. Using the ZIP code of the infant's residence at birth assures consistency with the denominator data obtained from birth certificates. This also eliminates any discrepancies that might have resulted from reports received for children who had changed residence between the time of birth and time of the report.

Denominator data were obtained from the New York State Department of Health, Bureau of Vital Statistics' Birth Statistical File. This file contains information found on the birth certificates with the exception of personal identifiers such as names and addresses. ZIP codes are, however, included in the Statistical File. Records that had no ZIP code or had a ZIP code that

was not a valid New York State ZIP code between the years of 1992 and 1995 were removed from the analysis, leaving 1,072,191 live birth records. Only 0.3 percent of the births in New York were missing ZIP code information. The remaining records of singleton births from 1992–1995 inclusive were tabulated by ZIP code. The final dataset consisted of 1601 ZIP codes with the number of children having at least one surveillance malformation and the number of singleton births in each ZIP code.

Statistical analysis

We used the spatial scan statistic (Kulldorff and Nagarwalla, 1995; Kulldorff, 1997) to determine areas of the state, which either had statistically significant deficits or elevations of rates, after adjusting for the multiple testing inherent in the many possible locations and sizes of areas. This technique has been used to evaluate clustering of several other health outcomes including leukemia (Hjalmars et al., 1996; Kulldorff and Nagarwalla, 1995), brain cancer (Kulldorff et al., 1998), breast cancer (Kulldorff et al., 1997; New York State Department of Health, 2000), soft-tissue sarcoma and non-Hodgkin's lymphoma (Veil et al., 2000), systemic sclerosis (Walsh and Fenster, 1997), low birth weight (Talbot et al., 2000), and SIDS (Kulldorff, 1997).

This method uses overlapping circles of many different sizes and locations to identify areas with elevations or deficits in disease incidence. The population-weighted centroids of the ZIP codes served as the center for the overlapping circles beginning with a radius of zero and increasing until a user defined maximum population restriction is met. The test uses the likelihood ratio test statistic. For each circle, the likelihood of finding the observed number of malformations inside and outside the circle, respectively, is calculated. The circle with the rate least likely to have occurred by chance is then determined and secondary clusters with excesses or deficits are also located. Circles with statistically significant ($p < 0.05$) elevations or deficits in malformation rates are determined through Monte Carlo hypothesis testing. These circles, along with the ZIP codes in the least likely ($p < 0.05$) non-overlapping circles, were then displayed on a map. We restricted the number of births captured by any one circle to be no more than 2.5 percent of all births in New York State to enable us to better focus on geographic areas for further evaluation of the completeness of hospital reporting.

The spatial scan statistic is an extension of work done by Openshaw et al. (1987, 1988) and that of Turnbull et al. (1990). The advantage of spatial scan statistic is that the test is more readily able to control for multiple hypotheses testing caused by the generation of multiple circles. In addition, since there are theoretically an

infinite number of circles around each point, the test is not as limited in testing predefined areas or population sizes.

Evaluation of reporting

To evaluate the completeness of reporting of birth defects to the Registry by hospitals, we compared the number of malformations reported to the Registry to the number of malformations reported to Statewide Planning and Research Cooperative System (SPARCS). SPARCS is a comprehensive data system established by the NYSDOH to receive, process, store and analyze hospital inpatient and ambulatory surgery data. All hospital-based and freestanding ambulatory surgery facilities in New York State are required to submit data to SPARCS (New York State Department of Health, 1999). SPARCS data are considered reliable since they are based on the same data processing systems as hospital billing records. Information on number and type of hospital admissions for each facility in the state is available through SPARCS. Therefore, it was possible to tabulate the number of malformations reported to SPARCS for each hospital.

The subset of malformations examined from SPARCS was similar to the surveillance malformations described above except that multiple births were kept in the analysis. This was because in the SPARCS system there is no way to differentiate multiple births from singleton births. In addition, whereas previously multiple reports of the same individual were excluded, in this analysis it was only possible to exclude multiple reports of the same individual if they occurred in the same hospital. Thus, it would be possible for two or more reports relating to the same individual to be included if they came from different hospitals. This did not bias the results greatly as the same restrictions were applied to the CMR data and the results re-tabulated to match the SPARCS data for this part of the analysis.

Finally, we aggregated records of malformations by hospital for both the CMR and SPARCS datasets. Results were tabulated and the rate of malformations reported to the CMR was calculated for each hospital. The SPARCS data served as the denominator for calculating the reporting rates since the SPARCS data are thought to be approximately 99 percent complete (New York State Department of Health, 1999) whereas the reporting of birth defects to the New York State CMR was estimated to be approximately 86 percent complete (Honein and Paulozzi, 1999).

Results

The results of the spatial scan statistic are shown in Fig. 1. Two large areas with elevated surveillance

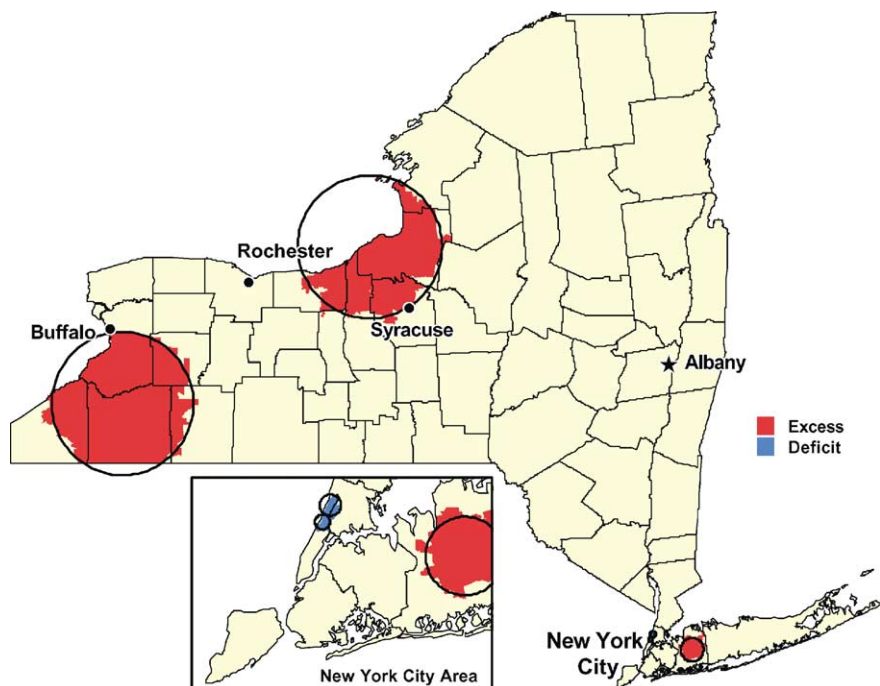


Fig. 1. Congenital malformation clusters identified using the scan statistic at $p < 0.05$ in New York State, 1992–1995.

malformation rates are apparent in upstate New York and one in Nassau County on Long Island, while two areas with rates significantly lower than expected are present in New York City. The two areas with lower than expected rates are directly adjacent to each other and cover much of the Harlem area in northern Manhattan as well as a portion of the Bronx.

Although the spatial scan test used in this study will locate only circular clusters, the actual shapes of the clusters identified are not exactly circular. This is because some ZIP codes whose centroids lie outside of the circular zone but whose boundaries cross into the circular zone would not be included while others whose centroids are included in the circular zone may actually extend outside the zone. In addition, two or more clusters may be located adjacent to one another, as is the case in the Harlem/Bronx area of New York City. While these represent two independent areas of deficit, they contain adjacent ZIP codes and thus appear as one elongated area of decreased prevalence of malformations.

To determine if these areas with lower than expected rates were due to under-reporting to the CMR by hospitals, the number of malformations reported to the Registry was compared to the number of malformations reported to SPARCS. Fig. 2 shows the location of hospitals in the metropolitan New York City area and the percent of SPARCS malformations that were

reported to the Registry. The number of malformations reported to SPARCS is also indicated by the size of the circle representing each hospital. Note the hospital in northern Manhattan, which had a large number of malformations reported to SPARCS combined with a low reporting rate to the CMR.

Since under-reporting was also apparent in several other New York City hospitals, the data were separated into two groups: one representing births to residents of New York City and one representing births to all other residents of the state. Table 1 shows a comparison of the percent of SPARCS malformations that were reported to the CMR for hospitals in the New York City area to those in the remainder of New York State. This shows the reporting rate of hospitals in New York City area to be approximately 13 percent lower than hospitals in the rest of the state (rate ratio (RR) 0.87; 95 percent confidence interval (CI) 0.86, 0.88).

The result of under-reporting by hospitals in the New York City area becomes evident when examining the malformation rate in this area compared to the rest of the state (Table 2). CMR malformation rates for births to New York City residents are shown to be approximately 20 percent lower than those in the remainder of the state (RR 0.80; CI 0.77, 0.82). This has a significant impact on the statewide rates as approximately 47 percent of all births to New York State residents occur in New York City Hospitals.

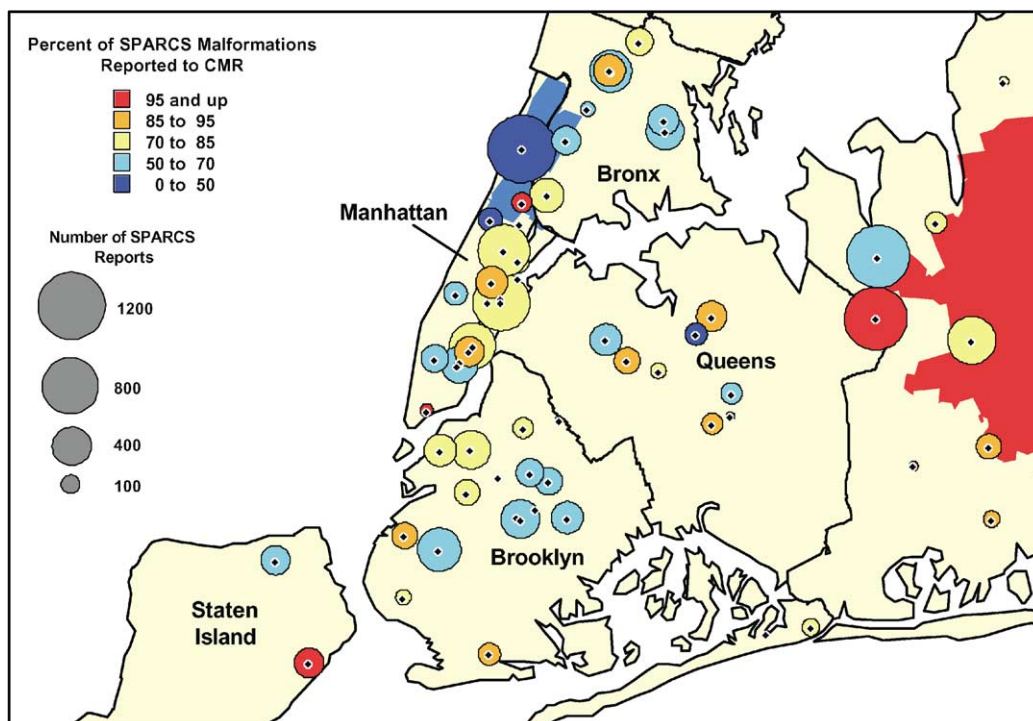


Fig. 2. Percent of SPARCS reports of congenital malformations reported to the New York State Congenital Malformations Registry by hospital in the New York City area.

Table 1

Surveillance malformations reported to the CMR and to SPARCS by hospital location in New York State, 1992–1995^a

	Location of hospital		
	New York City	Upstate/Long Island	Entire State
Malformations reported to CMR	9,311	12,205	21,516
Malformations reported to SPARCS	12,880	14,639	27,519
Percent reported to CMR	72.29%	83.37%	78.19%
Total births	516,021	577,948	1,093,969
CMR percent ^b	1.80%	2.11%	1.97%
SPARCS percent ^c	2.50%	2.53%	2.52%

^a The rates in this table are not meaningful when taken out of this context since there could be multiple malformation reports from different hospitals regarding a birth in this table. Multiple births are included in this table.

^b CMR percent = malformations reported to CMR/total births \times 100.

^c SPARCS percent = malformations reported to SPARCS/total births \times 100.

Table 2

Prevalence of surveillance malformations reported to the Congenital Malformations Registry by area of New York State, 1992–1995^a

	Area		
	New York City	Upstate/Long Island	Entire State
Births with malformations	6,720	9,658	16,378
Number of births	499,286	572,905	1,072,191
Malformation rate ^b	1.35%	1.69%	1.53%

^a Actual numbers of births and numbers of malformations differ between the two tables since different selection criteria were used for each. Multiple births are excluded in this table. In addition, there are slight differences in the way we tabulated births.

^b Malformation rate = surveillance malformations/total live births \times 100.

Actual numbers of births and numbers of malformations differ between the two tables since different selection criteria were used for each. The rates in Table 1 are not meaningful when taken out of this context since there could be multiple malformation reports from different hospitals regarding a birth in this table. This is not the case in Table 2, however. In addition, there are slight differences in the way we tabulated births in the two tables.

We compared the prevalence of infants with malformations reported to SPARCS for NYC with the rest of the state, since there is less under-reporting of children with malformation to SPARCS. From the data presented in Table 1, we see there is little difference in the rates of malformations reported to SPARCS for New York City compared to the remainder of the state (2.50 percent vs. 2.53 percent). Since substantial under-reporting of malformations was shown to exist in several large hospitals within New York City, malformations and births in all New York City ZIP codes were removed and the data reanalyzed using the spatial scan statistic. With New York City removed from the analysis, the large areas, which previously had statistically significant elevations in rates, now had rates within the expected range (Fig. 3). However, with NYC data removed from the analysis, two areas with statistically significant deficits emerged in the Upstate New York.

Discussion

Rate maps used in conjunction with statistical tests for clustering can be useful to health researchers for focusing attention on areas for further investigation. These types of maps are useful for showing general patterns or trends of disease, and to generate hypotheses regarding the role of environmental, genetic or lifestyle factors in the etiology of a disease. In addition, health administrators and planners may use maps to point to problems of health care delivery systems and to determine where underserved communities might exist. However, public health professionals should first examine the completeness and misclassification of the data. Errors in both health and population data may lead researchers or policy makers to draw erroneous conclusions regarding the spatial distribution of a disease.

Using malformation rates as an example we see how under-reporting in one area of the state affected not only local rates in that area but also resulted in apparent clusters of malformations in other areas of the state. If the CMR data were to be taken at face value one might conclude that large clusters of malformations exist near the cities of Syracuse, Buffalo and in Nassau County on Long Island. Not only could this lead to poor policy decisions, but it could also cause undue public alarm in these areas. When these areas are compared to the rates in New York State exclusive of New York City however



Fig. 3. Congenital malformation clusters identified using the scan statistic at $p < 0.05$ in New York State excluding New York City, 1992–1995.

as seen in Fig. 3, they fall within the expected range. This demonstrates that the apparent clusters of malformations in the Upstate and Long Island areas of the state may actually have been artifacts of under-reporting of malformations in New York City.

The two areas with statistically lower rates than expected were in adjacent areas of New York City, in Harlem and the Bronx. These areas have some of the highest poverty rates in the state (United States Bureau of the Census, 1992). Low socioeconomic status and poor nutrition have been associated with elevated prevalence of certain birth defects (Worthington-Roberts, 1997; Wasserman et al., 1998; Vrijheid et al., 2000). Thus, we might actually have expected an elevated rate of birth defects in this area. However, as Fig. 2 shows, there is one very large hospital serving much of the population in this area with a reporting rate to the CMR of <50 percent. The size of this hospital combined with the low reporting rate is responsible for the significantly decreased rates observed in this area.

Many congenital malformations are identified through prenatal screening. The prenatal identification of a serious malformation allows the choice to electively terminate the pregnancy. There may be spatial variation in screening rates, which in turn would impact the malformation rates. Rates of prenatal diagnosis of malformations vary among populations with blacks, Hispanics, and persons of lower socioeconomic status being less likely to undergo prenatal diagnosis of malformations (Kuppermann et al., 1996; Waller et al., 2000). In addition, even when screening results are known, women who are black or Hispanic and women of lower socioeconomic status are less likely to elect to terminate a pregnancy (Velie and Shaw, 1996). The exact effect of prenatal diagnosis and elective termination on malformation rates in the current study is unknown. However, the area identified with a significantly low malformation rate covers a large minority population with high rates of poverty. If prenatal diagnosis of malformations and elective termination were having an effect on the prevalence of children born with malformations in this area, one would expect an elevated malformation rate.

Based on SPARCS we found hospitals reported approximately 78 percent of the birth defects that occurred statewide (Table 1). These estimates are comparable to the 86 percent CMR reporting rate that has been reported previously (Honein and Paulozzi, 1999). The latter estimate is based on capture–recapture methods using birth certificate data for the years 1983–1986 to estimate the completeness of reporting to the CMR. While statewide reporting of all malformations to the Registry is incomplete, we found the completeness of reporting is not constant across the state. In the current evaluation, the reporting rates for hospitals located in New York City were approximately

13 percent lower than hospitals in the rest of the state. A previous analysis of the CMR data also indicated poorer reporting of malformations to the Registry among cases born in New York City during the years 1983–1986 (Olsen et al., 1996). The differences in the reporting of birth defects by hospitals to the CMR is responsible in part for the differences in malformation rates we see between New York City and the remainder of the state. This becomes apparent when we examine the SPARCS reports of malformations (Table 1) where little difference between the upstate and New York City rates of malformations was observed. If there were truly a lower prevalence of malformations in New York City one would expect this to be reflected in the SPARCS data as well as the CMR data.

One limitation of the spatial scan statistic is that it tests for unusual patterns within circles, though these patterns may not be circular in nature. We are not able with this test to determine the exact boundaries of the area where the unusual rates occurred. Other data can be displayed to gain a further understanding of the geographic pattern of the disease being studied. Talbot et al. (2000) have advocated showing smoothed health outcome rate maps in conjunction with the spatial scan statistic results.

While the scan statistic is useful in identifying areas of low as well as elevated incidence of disease, it cannot identify why a particular area is high or low. In the current study, data from a second data source was used to discover that the areas identified with low prevalence of birth defects were due to under-reporting rather than having a truly low prevalence rate. We show how thematically representing reporting rates of hospitals also provides useful information on how the quality of the data may vary spatially and impact statistical inference.

Spatial analysis can be a valuable tool in the quality control of disease and vital statistics registries. This analysis illustrates, through the use of maps, where deficiencies in hospital reporting of birth defects exist. These tools provide the registry with an additional way to monitor birth defect reporting statewide and to target hospitals for site visits and audits. By displaying the data on a map, hospital staff will recognize the value of providing complete and accurate data to the CMR. Once the completeness and accuracy of CMR data has been improved, spatial analysis techniques can be used as a proactive tool in the surveillance of birth defects.

There are bound to be some errors in all routinely collected data. However, the impact that these errors have on the final product can be minimized through the use of rigorous and thorough data cleaning and checking techniques. Traditional methods of quality control often focus on using descriptive statistics to identify outliers and data anomalies. Many times inconsistencies may remain hidden in the data until

they are viewed in a different fashion such as on a map. In the current study we use a combination of spatial techniques, both descriptive and analytical, to illustrate how spatial analysis can be used for data quality control. This is not only important in subsequent spatial analyses, but also for routine quality control of registry and surveillance data.

Appendix A

Malformations included in analysis are shown in Table 3.

Table 3

Malformation	ICD-9
Amniotic bands	658.8
Anencephalus	740.X ^a
Spina bifida with/without hydrocephalus	741.0X/741.9X
Encephalocele	742.0
Congenital hydrocephalus	742.3
An/microphthalmus	743.0X/.1X
Congenital cataract	743.3X
Coloboma of lens/iris	743.3X/.4X
Aniridia	743.45
Anomalies of ear causing impairment of hearing	744.0X
Common truncus	745.0
Transposition of great vessels	745.1X
Tetralogy of fallot	745.2
Common ventricle	745.3
Endocardial cushion defects	745.6X
Atresia/stenosis of pulmonary valve	746.01/.02
Insufficiency of pulmonary valve	746.09
Tricuspid atresia/stenosis/hypoplasia	746.1
Ebstein's anomaly	746.2
Congenital stenosis of aortic valve	746.3
Hypoplastic left heart syndrome	746.7
Coarctation of aorta	747.10
Interruption of aorta	747.11
Total/partial anomalous pulmonary venus connection	747.41/.42
Choanal atresia	748.0
Agenesis/hypoplasia, of lung	748.5
Oral clefts	749.0X/.1X/.2X
Tracheoesophageal fistula, etc.	750.3
Congenital hypertrophic pyloric stenosis	750.5
Atresia/stenosis of small intestine	751.1
Atresia and stenosis of rectum or anus	751.2
Hirschsprung's disease	751.3
Biliary atresia	751.61
Hypospadias/epispadias	752.6X
Indeterminate sex	752.7
Renal agenesis and dysgenesis	753.0
Cystic kidney disease	753.11–19
Obstructive defects renal pelvis and ureter	753.2753.4
Exstrophy of urinary bladder	753.5
Atresia/stenosis urethra and bladder neck	753.6
Reduction deformities of upper limb	755.2X

Table 3 (continued)

Malformation	ICD-9
Reduction deformities of lower limb	755.3X
Chondrodystrophy	756.4
Osteodystrophies	756.5X
Diaphragmatic hernia	756.6
Omphalocele/gastroschisis	756.7X
Down syndrome	758.0
Patau syndrome	758.1
Edwards syndrome	758.2
Gonadal dysgenesis	758.6
Klinefelter's syndrome	758.7
Fetal alcohol syndrome	760.71
Congenital rubella	771.0
Congenital cytomegalovirus infection	771.1
Other congenital infections	771.2

^aWhere X = 1–9, in these cases the 4 digit code will be sufficient.

References

- Boscoe, F.P., McLaughlin, C.C., 2000. Effects of seasonal migration on cancer incidence rates. North American Association of Central Cancer Registries Annual Meeting, New Orleans, LA, April 15–20.
- Fein, D.J., 1990. Racial and ethnic differences in US census omission rates. *Demography* 27 (2), 285–302.
- Goodman, D.C., Fisher, E., Stukel, T.A., Chang, C., 1997. The distance to community medical care and the likelihood of hospitalization: is closer always better? *American Journal of Public Health* 87 (7), 1144–1150.
- Gregorio, D.I., Cromley, E., Mrozinski, R., Walsh, S.J., 1999. Subject loss in spatial analysis of breast cancer. *Health Place* 5 (2), 173–177.
- Gregorio, D.I., Kulldorff, M., Barry, L., 2000. Cancer clusters resulting from census enumeration policy. International Health Geographics Conference 2000, Chevy Chase, MD, March 17–19.
- Hahn, R.A., 1992. The state of federal health statistics on racial and ethnic groups. *Journal of the American Medical Association* 267 (2), 268–271.
- Higgs, G., Gould, M., 2001. Is there a role for GIS in the 'new NHS'? *Health Place* 7 (3), 247–259.
- Hjalmar, U., Kulldorff, M., Gustafsson, G., Nagarwalla, N., 1996. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistical Medicine* 15 (7–9), 707–715.
- Holmes, L.B., 1999. Need for inclusion and exclusion criteria for the structural abnormalities recorded in children born from exposed pregnancies. *Teratology* 59, 1–2.
- Honein, M.A., Paulozzi, L.J., 1999. Birth defects surveillance: assessing the "gold standard". *American Journal of Public Health* 89 (8), 1238–1240.
- Kulldorff, M., 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26, 1481–1496.
- Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: detection and inference. *Statistical Medicine* 14 (8), 799–810.

- Kulldorff, M., Feuer, E.J., Miller, B.A., Freedman, L.S., 1997. Breast cancer clusters in the northeast United States: a geographic analysis. *American Journal of Epidemiology* 146 (2), 161–170.
- Kulldorff, M., Athas, W.F., Feurer, E.J., Miller, B.A., Key, C.R., 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health* 88 (9), 1377–1380.
- Kuppermann, M., Gates, E., Washington, A.E., 1996. Racial-ethnic differences in prenatal diagnostic test use and outcomes: preferences, socioeconomic, or patient knowledge? *Obstetrics and Gynecology* 87 (5 Pt 1), 675–682.
- New York State Department of Health, 1999. SPARCS Annual Report 1998, Vol. 1. New York State Department of Health.
- New York State Department of Health, 2000. Breast cancer in New York State, by ZIP Code, 1993–1997. New York State Department of Health. <http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>
- Olsen, C.L., Polan, A.K., Cross, P.K., 1996. Case ascertainment for state-based birth defects registries: characteristics of unreported infants ascertained through birth certificates and their impact on registry statistics in New York State. *Paediatric and Perinatal Epidemiology* 10, 161–174.
- Openshaw, S., Charlton, M., Wymer, C., Craft, A.W., 1987. A mark 1 analysis machine for the automated analysis of point data sets. *International Journal of Geographic Information Systems* 1 (4), 335–358.
- Openshaw, S., Craft, A.W., Charlton, M., Birch, J.M., 1988. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet* 1 (8580), 272–273.
- Raleigh, V.S., Balarajan, R., 1994. Public health and the 1991 census. *British Medical Journal* 309, 287–288.
- Richards, T.B., Croner, C.M., Novick, L.F., 1999a. Geographic information systems (GIS) for state and local public health practitioners, Part 1. *Journal of Public Health and Management Practice* 5 (2), 73–76.
- Richards, T.B., Croner, C.M., Novick, L.F., 1999b. Geographic information systems (GIS) for state and local public health practitioners, Part 2. *Journal of Public Health and Management Practice* 5 (4), 1–6.
- Talbot, T.O., Kulldorff, M., Forand, S.P., Haley, V.B., 2000. Evaluation of spatial filters to create smoothed maps of health data. *Statistical Medicine* 19 (17/18), 2399–2408.
- Turnbull, B.W., Iwano, E.J., Burnett, W.S., Howe, H.L., Clark, L.C., 1990. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 132 (Suppl. 1), S136–143.
- United States Bureau of the Census, 1992. Census of Population and Housing, 1990: summary tape file 3a on CD-ROM (New York State). Prepared by Bureau of the Census, Washington.
- Velie, E.M., Shaw, G.M., 1996. Impact of prenatal diagnosis and elective termination on prevalence and risk estimates of neural tube defects in California, 1989–1991. *American Journal of Epidemiology* 144 (5), 473–479.
- Viel, J.F., Arveux, P., Baverel, J., Cahn, J.Y., 2000. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology* 152 (1), 13–19.
- Vrijheid, M., Dolk, H., Stone, D., Abramsky, L., Alberman, E., Scott, J.E., 2000. Socioeconomic inequalities in risk of congenital anomaly. *Archives of Disease in Childhood* 2 (5), 349–352.
- Waller, D.K., Pujazon, M.A., Canfield, M.A., Scheuerle, A.E., Byrne, J.L., 2000. Frequency of prenatal diagnosis of birth defects in Houston, Galveston and the Lower Rio Grande Valley, Texas 1995. *Fetal Diagnosis Therapy* 15 (6), 348–354.
- Walsh, S.J., Fenster, J.R., 1997. Geographical clustering of mortality from systemic sclerosis in the Southeastern United States, 1981–90. *Journal of Rheumatology* 24 (12), 2348–2352.
- Wasserman, C.R., Shaw, G.M., Selvin, S., Gould, J.B., Syme, S.L., 1998. Socioeconomic status, neighborhood social conditions, and neural tube defects. *American Journal of Public Health* 88 (11), 1674–1680.
- Worthington-Roberts, B., 1997. The role of maternal nutrition in the prevention of birth defects. *Journal of American Dietetic Association* 97 (Suppl 2), S184–185.