ROLE OF GEOGRAPHIC INFORMATION SYSTEMS IN BIRTH DEFECTS SURVEILLANCE AND RESEARCH

Csaba Siffel,[1,2]* Matthew J. Strickland,[1,3] Bennett R. Gardner,[1] Russell S. Kirby[4] and Adolfo Correa[1]

Affiliations:

(1) Division of Birth Defects and Developmental Disabilities, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA

(2) Computer Sciences Corporation, Atlanta, GA

(3) Battelle Centers for Public Health Research and Evaluation, Atlanta, GA

(4) Department of Maternal and Child Health, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

*Correspondence and reprint requests to:

Csaba Siffel, MD, PhD

National Center on Birth Defects and Developmental Disabilities

Centers for Disease Control and Prevention

Mailstop E-86, 1600 Clifton Road, Atlanta, GA 30333

Phone: 404-498-3821
Fax: 404-498-3040
E-mail: csiffel@cdc.gov

CSC Papers

2008

ABSTRACT

BACKGROUND: With the significant advancement of geographic information systems (GIS), mapping and evaluating the spatial distribution of health events became easier. We examine the role of GIS in birth defects surveillance and research. **METHODS**: We briefly describe the geocoding process and potential problems in accuracy of the obtained geocodes, and some of the capabilities and limitations of GIS. We illustrate how GIS has been applied using the Metropolitan Atlanta Congenital Defects Program geocoded dataset. We provide some comments on potential data quality and confidentiality issues with birth defects in relation to GIS. **RESULTS**: It is desirable to geocode addresses using a multistrategy approach to achieve a high quality and accurate GIS dataset. Beyond the basic but important function of mapping, sophisticated statistical approaches and software are available to analyze the spatial or spatial-temporal occurrence of birth defects, alone or in association with environmental hazards, and to present this information without compromising the confidentiality of the subjects. **CONCLUSION**: We recommend a broad and systematic use of GIS in birth defects spatial surveillance and research.

Key words: GIS, birth defects, surveillance, geocoding, accuracy, spatial analysis, mapping, confidentiality

.

INTRODUCTION

Of the three basic epidemiologic characteristics (time, place, and person) used to describe the occurrence of birth defects in a population, place has been the least used because of difficulties defining spatial locations in a standard and meaningful manner. However, advances in information technology now make it possible to convert location information, such as addresses, into spatial coordinates (geocodes) and to analyze the geographic distribution of disease. One of these advances is the development of geographic information systems (GIS) (i.e., powerful tools combining geography, data, and computer mapping as defined by *Healthy People 2010* [U.S. Dept. HHS, 2000]), which allow for mapping health data, and evaluating spatial variation in disease (e.g., clusters or trends). Some of the most widely used and available mapping software include ArcView, MapInfo, EpiMap, Maptitude, Autodesk World, and Geomedia (Melnick, 2002). The Statistical Analysis Software (SAS Institute, Cary, NC) has also a GIS component.

Spatial features are stored in a coordinate system (i.e., latitude and longitude) that corresponds to locations on the earth's surface. Descriptive attributes in tabular form are associated with these spatial features. Spatial data and associated attributes in the same coordinate system can then be layered together for mapping and analysis. Traditional health data with addresses and increasing volumes of geo-referenced data such as environmental and sociodemographic data are now widely available for integration and GIS analysis. Some examples of the application of GIS in public health include malaria prevention (Martin et al., 2002), childhood lead poisoning prevention (CDC, 2004; Miranda et al., 2002), and studying infant mortality inequalities (Andes and Davis, 1995).

Spatial surveillance and research has been limited for birth defects. GIS has been used in studies of spatial analysis of birth detect rates (Rushton and Lolonis, 1996; Forand et al., 2002), analyses of associations between birth defects and exposures such as hazardous waste sites and air pollution (Orr et al., 2002; Gilboa et al., 2005), and analysis of socioeconomic status and neural tube defects (Wasserman et al., 1998). This paper describes the potential role of GIS in birth defects surveillance and research by providing: (1) brief overviews of the geocoding and disease mapping processes, including data requirements; (2) some of the capabilities and limitations of GIS in birth defects monitoring and research; (3) illustrations of potential GIS applications based on reviews of publications on reproductive outcomes and birth defects, and data from the Metropolitan Atlanta Congenital Defects Program (MACDP); and (4) comments on data quality and potential confidentially issues inherent in the use of GIS for analysis of birth defects data.

MATERIALS AND METHODS

GEOCODING PROCESS AND SOME OF ITS DIFFICULTIES

One important application of GIS techniques in public health data is using spatial attributes to identify target areas for public health interventions. Health events and populations across large administrative areas are grouped such as states, counties, and metropolitan areas, and do not use methods of spatial analysis on individual-level data. With the advent of GIS in the public health arena, it is now possible to use digital procedures to find latitude and longitude map coordinates that correspond to specific addresses or locations (a process termed “geocoding”), an important first step in the creation of datasets for spatial analysis. Geocodes for addresses can be obtained using a global positioning system (GPS). However, this approach could be time consuming and costly for large number of records. Instead, geocodes for residential addresses are usually generated using proprietary geocoding technology along with comprehensive street databases such as the U.S. Census Bureau‘s Topologically Integrated Geographic Encoding and Referencing (TIGER) files (Croner et al., 1996) or tax parcel databases (Cayo & Talbot, 2003). In this process, it is essential to provide and maintain complete address information in order to achieve the highest possible match rate (proportion of successfully geocoded addresses), either via probabilistic (mostly score-based) or deterministic (pattern- or rule-based) linkage methods. Probabilistic matching, which can be audited and validated, usually gives a better result when there is incomplete, inconsistent, or missing information in address records, which often is the case with public health datasets. Historically, geocoding has been measured by the match rate achieved. More recently, the accuracy of geocoded information has become an important issue, and is discussed in the Data Quality and Confidentiality Issues section.

Once map coordinates are obtained, census geographic levels such as state, county, city, census tract, census block group, or census block can be added to the records, and attributed data from the U.S. Census Bureau or other georeferenced datasets can be linked to them. When an address level match is not possible, alternatives for geocoding include mapping data to the centroid of areal units such as census tracts or postal Zone Improvement Plan (ZIP) Codes. ZIP Codes, especially ZIP+4 Codes, change frequently over time, whereas census tracts or blocks change every 10 years. Therefore, in analyses of temporal trends, it is usually preferable to conduct data aggregation and analysis on the census tract or block group level (Krieger et al., 2002). Furthermore, the population within census tracts or blocks tends to be more homogeneous with respect to socioeconomic characteristics relative to ZIP Codes, and can, therefore, allow for more sensitive evaluations of spatial variation (e.g., poverty level of spatial units). Problems can arise if ZIP Code values in administrative health data are incorrect or fail to reflect recent changes. Errors can also occur if denominator data are based on ZIP Code Tabulation Areas (ZCTAs), which are new statistical entities developed by the U.S. Census Bureau for tabulating summary statistics from Census 2000, because the ZCTAs do not correspond with complete congruence to ZIP Codes reported in health databases (Krieger et al., 2002).

MACDP has ascertained birth defects within the five counties (Clayton, Cobb, DeKalb, Fulton, and Gwinnett) of metropolitan Atlanta, Georgia, since 1967 (Correa-Villasenor et al., 2003). Following a feasibility study and validation of results, to date nearly forty thousand records with addresses have been geocoded by a commercial vendor using high-quality street reference dataset, and approximately 95% of the MACDP records have been geocoded by address (80% with batch and 15% with interactive and manual processes). MACDP maintains high quality address information but to achieve this high match rate, it was necessary to correct some incomplete or misspelled addresses; this process is known as address standardization. While manual labor is required for the correction process, few resources other than basic postal address files, maps, and local knowledge of the region under study are needed. These spatially referenced data are the basis for GIS analysis of birth defects. More recently, several state health departments have also built GIS data infrastructures (Ruiz and Remmert, 2004) and have geocoded maternal place of residence from birth and fetal death certificates (MacDorman and Gay, 1999). Birth defect datasets linked with these geocoded birth files could potentially provide a means for obtaining coordinates for birth defect cases. Although both birth defect and vital statistics systems usually intend to record the maternal residential address at the time of delivery, a linked record does not necessarily contain the same address for the same person. This raises further questions on how to handle these discrepant addresses. Older records are more difficult to geocode because street reference files usually reflect the current street network.

Geocoded birth defects data can be used for a variety of analyses, such as exploration of spatial or temporal-spatial patterns for particular defects or defect groups; analyzing infant mortality related to birth defects; and performing cluster analysis. Birth defects data can also be linked to sociodemographic and environmental databases. The U.S. Census Bureau data is a good source for underlying sociodemographic data but are not available at the individual level; instead, data are aggregated to various geographical levels. Because the infant population is underestimated by census data (U.S. Census Bureau, 2001), the ideal solution is to geocode the underlying birth population to the address level using the birth and fetal death certificate data. These geocodes are available in some states, including Georgia.

MAPPING OF CASES: KNOW YOUR MAP, THE SCALE, THE PROJECTION

Using GIS software, geocoded data can be plotted on a digital map. These points can be aggregated to various geographic levels (the process is also called spatial aggregation), and can be analyzed by statistical packages to address sociodemographic and environmental issues, and to identify spatial clusters.

Once health data are brought into a GIS database, important issues in data mapping remain. Different spatial databases need to have identical scale and map projection (Cromley and McLafferty, 2002); otherwise data can be distorted or cannot be mapped together. Map scale shows the relationship between a unit of length on a map and the corresponding length on the ground. The smaller the scale, the larger the area displayed on a map. Map projections are transformations of the three-dimensional surface of the earth onto a flat map using mathematical models. Some distortions of conformality (angle-preserving projection), distance, direction, scale, and area inevitably result from this process. Therefore, maps that focus on maintaining one feature (e.g., preserving distance) must distort other features (e.g., area and shape). Maps that accurately reflect area are called equal-area maps, whereas maps that correctly show the distance between points are called equidistant maps. A more complete review of map projections can be found in the book entitled “Map Projections: A Reference Manual” (Bugayevskiy and Snyder, 1995).

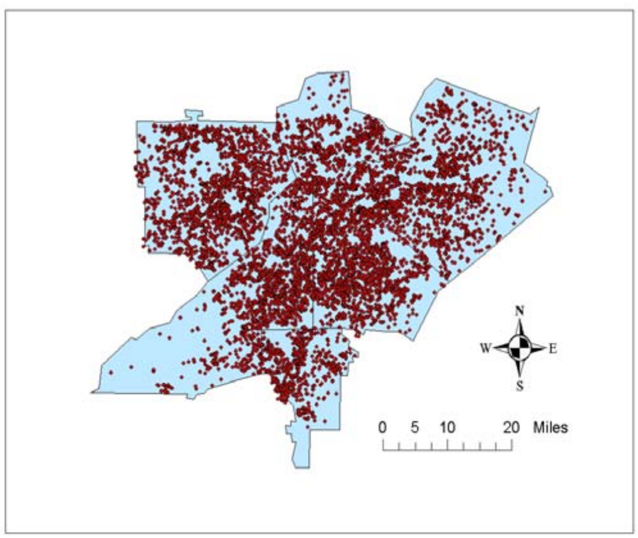
GIS databases are in either a vector or a raster map format, or a combination of the two formats. In vector maps, geographic features are represented by points (e.g., location of infants with birth defects), lines (e.g., streets), and polygons (e.g., census tracts) (Rogers, 1999). These features are based on latitude and longitude (x, y) coordinates of the various objects. The vector format is more frequently used in public health settings. In raster maps (e.g., orthophotos and scanned maps) data are stored as digital images (Vine et al., 1997). Usually grid cells are used to locate and represent a feature and these cells form a continuous surface. Smaller cells provide better resolution. Obtaining quality maps for a given geographical area for the time period of interest is crucial because maps are static but we live in a constantly changing environment.

DATA VISUALIZATION

Perhaps the two most frequently used maps in public health settings are the dot-density and choropleth maps (Rogers, 1999), both of which are vector-format maps.

Dot-density maps (Fig. 1) are a simple way to display events; dots or other symbols of different sizes, shapes, and colors represent the number of occurrences of a given data characteristic in a particular location (Thrall, 1999). Each dot can represent a single entity (one dot = one case) or a group (one dot = 1,000 people). Dot-density maps could be useful for area comparisons and show density of events in geographical areas of interest. If areas have different sizes, which many times is the case, the dots in a smaller area would be closer together than the dots in a larger area. However, dot-density maps need to be interpreted with caution regarding the “symbol-to-data characteristic” ratio (i.e., size and group representation of dots to data). It is also important to keep in mind that dots do not always indicate the exact location of the data; they could represent data occurring within a polygon such as a census tract, ZIP Code, or county. Dots or symbols on a dot-density map generated from aggregated data are randomly placed inside the areas of interest.

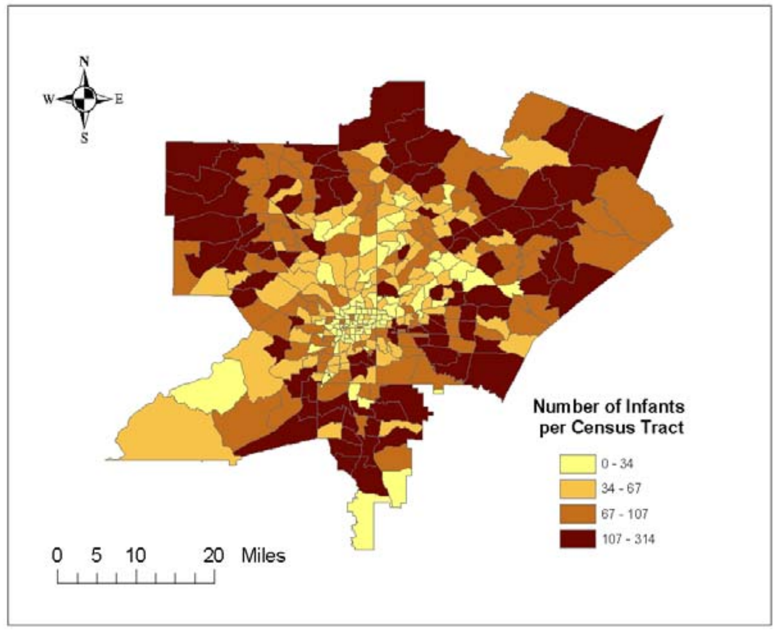
Figure 1. Example of a dot-density map. Birth defect cases in metropolitan Atlanta



Choropleth maps (Fig. 2) are area maps in which polygons (e.g., census tracts and counties) are shaded, colored, or patterned according to the extent to which each polygon is associated with a given attribute, such as population size or disease rate (Melnick, 2002).

Besides dot and choropleth maps, other two- and three-dimensional mapping techniques are available in GIS tools.

Figure 2. Example of a choropleth map. Number of infants per census tract in metropolitan Atlanta, 1990.



It is important to consider how map characteristics such as color, pattern, size, polygon shape, and class intervals influence map interpretation (Brewer, 2005). The three main types of color schemes are the following: 1) sequential, 2) diverging, and 3) qualitative. Sequential or single color maps with varying color intensity (shades) are usually a good way to present data (i.e., rates) without a critical midpoint value. To show data diverging from a critical midpoint value such as prevalence ratios of birth defects, two complementary color schemes that diverge from a common hue (the gradation of color within the visible spectrum of light) are applied. Qualitative color schemes are useful to present nominal differences (i.e., nonranked categorical data) by differences in hue. ColorBrewer is a useful Web tool for selecting color schemes for thematic maps (ColorBrewer, 2006). For black and white or grey-scale maps, applying patterns is a good idea. Proportions or rates can be displayed by different class intervals such as equal intervals (equal ranges of values) and quintiles (equal number of polygons falling into each class defined by dividing the range of values). The latter method is a good choice for presenting skewed data (McLafferty and Cromley, 1999). Cartographers recommend use of polygons of similar size, especially when presenting data, as the presence of a few polygons of larger size can dominate the map and potentially lead to misinterpretation. Large differences in quantitative values across a map may also result in erroneous impressions of spatial patterns. For example, the large census tracts with darker colors located on the outskirts of the metropolitan Atlanta area (Fig. 2) might be interpreted as areas with a high density of events (e.g., births). However, birth density is highest in central Atlanta. Larger areas do not necessarily mean that they have more population, either. The higher number of infants in the large census tracts in metro Atlanta is due to the rapid growth in the suburbs.

SPATIAL STATISTICAL CONSIDERATIONS

The analytical features of GIS include overlay, buffering, density estimation, interpolation, smoothing, modeling, and simulation (Kistemann et al., 2002). For instance, using the buffering feature of GIS tools, polygons can be created based on the distance from a specified location, which can be points, lines, or polygons. Buffers are particularly useful in identifying people at risk of exposure to environmental hazards. For example, the number of birth defect cases, live births, and fetal deaths can be captured within a buffer zone of interest, such as a buffer around a chemical factory or roadway.

Environmental databases include information on various environmental hazards such as air (ozone, nitrogen oxides), water (disinfection byproducts) or soil (pesticides). Environmental hazards can be emitted from point sources (i.e., hazardous waste sites or, coal-burning power plant emissions) or can be more diffuse (i.e., motor vehicle emissions or, drinking water contaminants). Environmental health tracking programs are particularly interested in these types of data (McGeehin et al., 2004). For studies of birth defects and environmental hazards, residential mobility during pregnancy is an important methodological issue and a potential source of bias. Traditionally, birth defects programs record maternal addresses at the time of delivery. However, based on earlier reports (Khoury et al., 1988; Shaw and Malcoe, 1992) and our recent analysis, approximately 20% of pregnant women in the United States change their residence between the dates of conception and delivery. Therefore, exposure misclassification can be introduced by using the mother's residence at delivery to extrapolate potential exposure information during earlier pregnancy periods (Schulman et al., 1993).

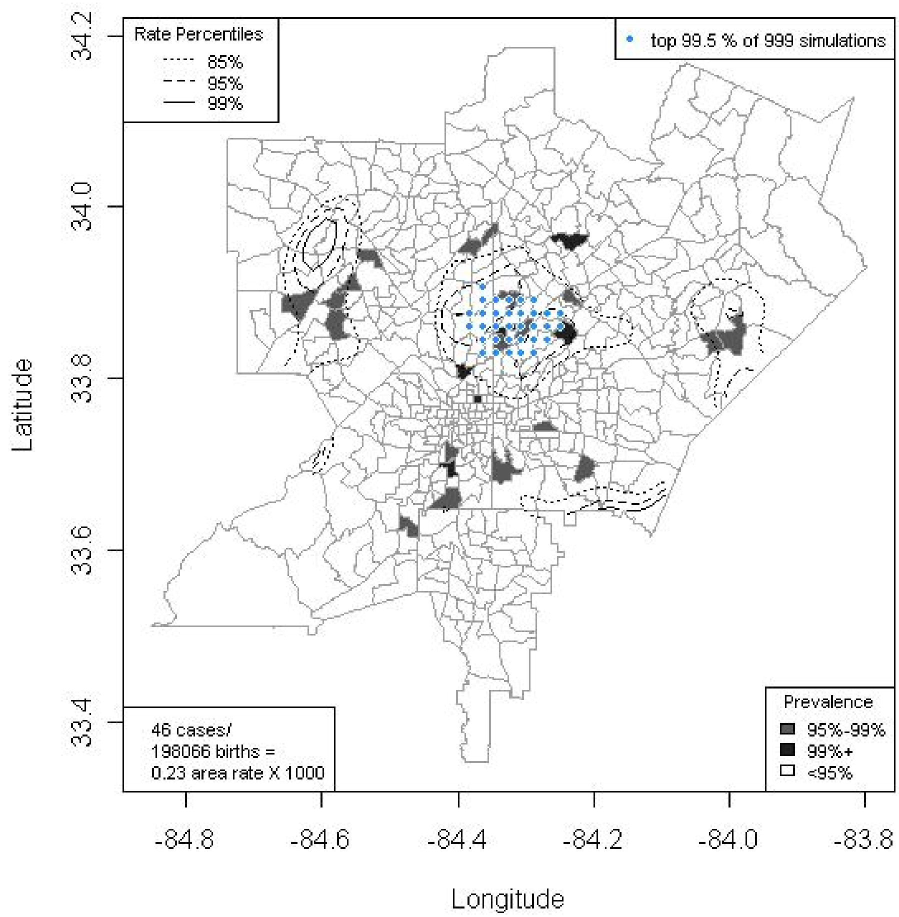
Spatial analysis of socioeconomic data and health data often requires aggregation into arbitrary areal units such as census tracts. Aggregated demographic datasets are associated with analytical problems due to the arbitrary nature of areal unit partitioning. Perhaps the most prominent of these issues is the modifiable areal unit problem (MAUP), defined as a situation in which modifying the boundaries and/or scale of data aggregation significantly affects the results of spatial data analysis (Openshaw 1984). Although MAUP remains unsolved, the use of well-chosen variables to adjust the area-level results can provide reliable estimates for underlying individual-level relationships. In particular, the potential for the "ecological fallacy" from drawing incorrect individual-level inferences from area-level analyses can be reduced (Holt et al. 1996).

Using a smoothing technique (Chung et al., 2004) to create smooth rates is a good alternative to overcome the variation (nonuniformity) of features in small areas on choropleth maps (Melnick, 2002), and stabilize the variance of small areal unit rates. While most widely used GIS applications now incorporate some smoothing procedures, users must make themselves aware of the assumptions on which these methods are based. Heterogeneity in population density across the areal units on a map can affect the results of some smoothing algorithms, such as kernel density or kriging methods. Other methods, for example head-banging (Pickle et al., 1999), hierarchical Bayesian methods (MacNab, 2003), and adaptive spatial filters (Talbot et al., 2000) have yet to be integrated into most commercially available GIS packages (Rushton, 2003).

In regards to spatial analysis and cluster detection methodologies such as ”hot spot” analysis, nearest neighbor analysis, density-equalized map projection, and close-pairs test statistics (Chung et al., 2004; Elliott et al., 2000, Shaw et al., 1988), a great variety of these methods are either built in as part of the GIS software or independently available, either freely or commercially. SaTScan is an example of a freely available and relatively easy to use statistical tool for detecting local clusters of disease (SaTScan, 2006). Either point data or rate data can be used. WinBUGS is a free program (WinBUGS, 2006) that can be used to model disease data using a Bayesian probability model estimated by Markov chain Monte Carlo methods (Lawson et al., 2003) and also to smooth rates for presentation. GeoDA is another GIS application freely downloaded from the Web (GeoDA, 2006). It is useful for conducting exploratory data analyses such as three-dimensional visualization and cartogram creation. GeoDA also supports spatial regression (Anselin, 2003).

Several birth defects surveillance programs, including MACDP, have considered monitoring their population on a regular basis for spatial patterns. GIS will play an important role in this endeavor, but researchers need tailored tools to look for patterns across several types of defects. We are developing a simple GIS tool, preliminarily titled Automated Spatial Surveillance Project (ASSP), to monitor spatial and temporal trends of birth defects within MACDP and elsewhere. ASSP is written for R, a language and environment for statistical computing (R Development Core Team, 2006). R is free software distributed under the GNU General Public License. We are designing ASSP to produce meaningful maps and tables while remaining user-friendly. Users import a comma-separated value (.csv) file of geocoded cases and controls, each tagged with a birth year, as well as shapefiles of the area under surveillance (e.g., a shapefile of census tracts in metropolitan Atlanta). ASSP creates one map for each birth defect category for each year. ASSP also creates several tables to be used in conjunction with these maps to track trends of a specific defect over time. An ASSP map (Fig. 3) depicts crude prevalence, smoothed prevalence contours, and potential clusters. ASSP calculates census tract prevalence by using a point-in-polygon routine, and uses a kernel smoothing routine (Rowlingson and Diggle, 1993) to create smoothed prevalence contours depicting areas of elevated risk. Kernel smoothing is a point-based method for spatial analysis (i.e., it uses geocodes rather than small-area rates). This kernel smoothing routine is also used in Monte Carlo simulations to identify potential clusters (Kelsall and Diggle, 1995), which are depicted by the blue asterisks on the map. Although we hope ASSP proves helpful in our surveillance activities, these maps have some important limitations. One limitation is the possibility for type 1 error, which is likely to occur if maps are generated for large numbers of birth defect categories across multiple time points. Furthermore, birth defects are rare events, and ASSP maps are sensitive to the inclusion and exclusion of a small number of cases. For example, if a birth defects case has a missing geocode (or has been geocoded incorrectly) this can impact which areas on the map appear to be at elevated risk. Accurate case definition of birth defects cases is also important; if a cluster is detected, the user should examine the surveillance records to ensure that cases are correctly classified.

Figure 3. Automated Spatial Surveillance Project (ASSP) map of crude prevalence, smoothed prevalence contours, and potential clusters of birth defects in metropolitan Atlanta.



In Table 1, we list selected examples published in the last 15 years demonstrating how GIS contributed to analyses and conclusions in studies on reproductive outcomes and birth defects in the United States. Of these papers, six included geocoded data on birth defects and eight other papers used various reproductive outcome data, such as fetal death, low birth weight, preterm birth, and infant mortality geocoded data. In these studies GIS played a role in many ways: from basic functions such as geocoding and spatial linkage to more sophisticated spatial modeling. The spatial unit of analyses varied from point locations (address-level geocodes) to ZIP Codes.

DATA QUALITY AND CONFIDENTIALITY ISSUES

An important methodological issue in analyses of spatial data relates to data quality. The accuracy or overall quality of a GIS dataset can be characterized by (1) the positional accuracy, or how close the coordinates fall compared with their actual location; and (2) the attribute accuracy, or how well the features are described in the dataset. Logical consistency and completeness of the dataset also contribute to accuracy measure. Underascertainment of birth defects in portions of the region under surveillance may lead to apparent clusters of birth defects in areas where ascertainment is more complete (Forand et al., 2002). Estimates for areas on the border of the surveillance region can be especially problematical. GIS analysis can be a valuable tool in the quality control of birth defects surveillance systems.

Knowing the quality and validity of geocodes is also important, for instance, in cluster analyses, or when distance from an exposure source is a key factor. There is inherent uncertainty between the true residential location and the coordinates generated via the geocoding process. The accuracy of geocodes depends on the completeness and correctness of the submitted addresses, as well as the quality of the street databases used to generate the geocodes. High quality street databases (1) contain all roads and address ranges within a geographical area; (2) are free of spelling errors; (3) contain the correct attribute information; (4) correspond with the submitted addresses with respect to time; and (5) are spatially accurate (Karimi et al., 2004). If geocodes are used to assign exposure to environmental hazards, geocode location (positional) errors can result in exposure misclassification. In Atlanta, we observed that for a sample of 599 MACDP addresses the median positional error was less than 100 meters. However, for a small proportion of addresses (<1%), positional error was greater than 1 kilometer. To assess the distribution of positional errors, we calculated the distance between the “true” location of addresses, which were determined by using geocoded tax parcel data and high-resolution orthoimages, and MACDP geocodes. It has been reported that the magnitude of positional error is typically smaller in urban areas than in rural areas (Cayo & Talbot, 2003). Inaccurate coordinates can result in the address being aggregated into the incorrect census tract (Ratcliffe, 2001), which can lead to spurious findings.

To improve results, some recommend using a multistage strategy for geocoding (McElroy et al., 2003; Yang et al., 2004). Although commercially available GIS tools can significantly improve geocoding accuracy, accurate address collection remains the most critical first step in the process (Yang et al., 2004). In epidemiologic research, the acceptable accuracy needs to be carefully considered before studying associations between exposure and health outcomes in a spatial context (i.e., ecologic, etiologic and geographic correlation studies). Depending on the specifics of the study design, inaccuracy might be more acceptable for measuring association between ambient air pollution and health outcome than measuring association between a point-source exposure (e.g., radiation from a nuclear power plant) and living within a certain distance from the exposure source.

Another issue involves protecting the confidentiality of the birth defects surveillance records. Because residential address is a personal identifier, maps showing point locations (indirect identification of people) or even aggregate data in a small geographic area (group disclosure) might reveal the identity of individuals or could lead to certain unintended and undesired consequences (Brownstein et al., 2005; Melnick, 2002). For example, exposure information derived from environmental studies could decrease property values affecting the financial situation of people living in an exposed community (Cox, 1996). Disclosure of the residential locations of families of children with birth defects has potential social consequences for both the family and the affected child. . Therefore, birth defects surveillance systems and other agencies might wish to avoid disseminating data in a disaggregated format. In Figure 1 we chose to show point locations of cases with birth defects on a small-scale map without any other identifiable information such as year of birth, sex, etc. Although GIS methods have been developed to protect privacy and limit disclosure of information by geographically masking individual records (Armstrong et al., 1999), using masked data in small-area analysis limits the ability to detect clusters (Kamel Boulos et al., 2006), and there is a trade-off between the extent of geographic masking and the accuracy of subsequent spatial analyses (Kwan et al., 2004). Careful choice of geographical units and mapping aggregate data (rates) on birth defects helps ensure confidentiality when presenting maps with small areas. The increasing number of promising web-based applications (Kistemann et al., 2002), which offer new ways to interactively generate maps, geocode addresses on the fly, and build space-related disease surveillance systems, might result in opportunities for individuals to be identified. Appropriate safeguards that take local laws and geographical characteristics into consideration will help ensure confidentiality.

CONCLUSIONS AND RECOMMENDATIONS

Spatial surveillance and research has not been systematically conducted for birth defects. Geographic information systems provide a set of tools to address difficulties in processing, analyzing, and visualizing spatial data. With GIS it is possible to manage, manipulate, and integrate large volumes of information such as birth defects, birth and death certificates, and census and environmental data in an efficient way. Using the spatial analysis and modeling functions, including overlays, buffering, and basic statistics of GIS, it is possible to explore and show relationships between disease rates and environmental hazards, demonstrate changes in rates over time, and identify high-risk areas and clusters. Mapping different types of birth defects can provide new insights into both etiology and interventions. We recommend the application of GIS in the birth defects field, but moving beyond basic GIS techniques to in-depth data analysis and implementation of interventions is a challenge (Caley, 2004). Because many public health professionals do not have extensive training in GIS, we recommend that experts from various backgrounds work together to avoid the caveats of this multidisciplinary approach (Kirby, 1996).

REFERENCES

Andes N, Davis JE. 1995. Linking public health data using geographic information system techniques: Alaskan community characteristics and infant mortality. Stat Med 14:481-490.

Anselin L. 2003. GeoDa™ 0.9 User's Guide. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, Urbana, IL. Available at https://www.geoda.uiuc.edu/pdf/geoda093.pdf, last accessed on August 8, 2006.

Armstrong MP, Rushton G, Zimmerman DL. 1999. Geographically masking health data to preserve confidentiality. Stat Med 18:497-525.

Banerjee S, Wall MM, Carlin BP. 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. Biostatistics 4:123-142.

Blake BJ, Bentov L. 2001. Geographical mapping of unmarried teen births and selected sociodemographic variables. Public Health Nurs 18:33-39.

Brewer CA. 2005. Designing Better Maps: A Guide for GIS Users. Redlands: ESRI Press.

Brownstein JS, Cassa CA, Kohane IS, Mandl KD. 2005. Reverse geocoding: concerns about patient confidentiality in the display of geospatial health data. AMIA Annu Symp Proc 905.

Bugayevskiy LM, Snyder JP. 1995. Map Projections: A Reference Manual. London: Taylor and Francis Ltd.

Caley LM. 2004. Using geographic information systems to design population-based interventions. Public Health Nurs 21:547-554.

Cayo MR, Talbot TO. 2003. Positional error in automated geocoding of residential addresses. Int J Health Geogr 2:10.

Centers for Disease Control and Prevention (CDC). 2004. Using GIS to Assess and Direct Childhood Lead Poisoning Prevention: Guidance for State and Local Childhood Lead Poisoning Prevention Programs. Atlanta: CDC.

Chung K, Yang DH, Bell R. 2004. Health and GIS: toward spatial statistical analyses. J Med Syst 28:349-360.

ColorBrewer. 2006. Available at http://www.ColorBrewer.org, last accessed on August 8, 2006.

Cox LH. 1996. Protecting confidentiality in small population health and environmental statistics. Stat Med 15:1895-1905.

Correa-Villasenor A, Cragan J, Kucik J, O'Leary L, Siffel C, Williams L. 2003. The Metropolitan Atlanta Congenital Defects Program: 35 years of birth defects surveillance at the Centers for Disease Control and Prevention. Birth Defects Res Part A Clin Mol Teratol 67:617-624.

Cromley EK, McLafferty SL. 2002. GIS and public health. New York: The Guilford Press. .

Croner CM, Sperling J, Broome FR. 1996. Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships. Stat Med 15:1961-1977.

Elliott P, Wakefield J, Best N, Briggs D. 2000. Spatial Epidemiology: Methods and Applications. New York: Oxford University Press Inc.

Forand SP, Talbot TO, Druschel C, Cross PK. 2002. Data quality and the spatial analysis of disease rates: congenital malformations in New York State. Health Place 8:191-9.

GeoDA. 2006. Available at https://www.geoda.uiuc.edu, last accessed on August 8, 2006.

Gilboa SM, Mendola P, Olshan AF, Langlois PH, Savitz DA, Loomis D, Herring AH, Fixler DE. 2005. Relation between ambient air quality and selected birth defects, seven county study, Texas, 1997-2000. Am J Epidemiol 162:238-252.

Holt D, Steel DG, Tranmer M, Wrigley N. 1996. Aggegation and ecological effects in geographical based data. J Geogr Analysis 28:244-62.

Ihrig MM, Shalat SL, Baynes C. 1998. A hospital-based case-control study of stillbirths and environmental exposure to arsenic using an atmospheric dispersion model linked to a geographical information system. Epidemiology 9:290-294.

Kamel Boulos MN, Cai Q, Padget JA, Rushton G. 2006. Using software agents to preserve individual health data confidentiality in micro-scale geographical analyses. J Biomed Inform 39:160-170.

Karimi HA, Durcik M, Rasdorf W. 2004. Evaluation of uncertainties associated with geocoding techniques. Comput-Aided Civ Inf Eng19:170-185.

Kelsall JE, Diggle PJ. 1995. Non-parametric estimation of spatial variation in relative risk. Stat Med 14:2335-2342.

Khoury MJ, Stewart W, Weinstein A, Panny S, Lindsay P, Eisenberg M. 1988. Residential mobility during pregnancy: implications for environmental teratogenesis. J Clin Epidemiol 41:15-20.

Kirby RS. 1996. Toward congruence between theory and practice in small area analysis and local public health data. Stat Med 15:1859-1866.

Kistemann T, Dangendorf F, Schweikart J. 2002. New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. Int J Hyg Environ Health 205:169-181.

Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. 2003. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). J Epidemiol Community Health 57:186-199.

Krieger N, Waterman P, Chen JT, Soobader MJ, Subramanian SV, Carson R. 2002. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas - the Public Health Disparities Geocoding Project. Am J Public Health 92:1100-1102.

Kwan MP, Casas I, Schmitz BC. 2004. Protection of geoprivacy and accuracy of spatial information: how effective are geographical masks? Cartographica 39:15-28.

Lawson AB, Browne WJ, Vidal Roderio CL. 2003. Disease Mapping with WinBUGS and MLwiN. Chichester: John Wiley & Sons Ltd.

MacDorman MF, Gay GA. 1999. State initiatives in geocoding vital statistics data. J Public Health Manag Pract 5:91-93.

MacNab YC. 2003. Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease. Stat Med 22:1761-73.

Martin C, Curtis B, Fraser C, Sharp B. 2002. The use of a GIS-based malaria information system for malaria research and control in South Africa. Health Place 8:227-236.

McElroy JA, Remington PL, Trentham-Dietz A, Robert SA, Newcomb PA. 2003. Geocoding addresses from a large population-based study: lessons learned. Epidemiology 14:399-407.

McGeehin MA, Qualters JR, Niskar AS. 2004. National environmental public health tracking program: bridging the information gap. Environ Health Perspect 112:1409-1413.

McLafferty S, Cromley E. 1999. Your first mapping project on your own: from A to Z. J Public Health Manag Pract 5:76-82.

Melnick AL. 2002. Introduction to geographic information systems in public health.Gaithersburg: Aspen Publishers, Inc.

Miranda ML, Dolinoy DC, Overstreet MA. 2002. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. Environ Health Perspect 110:947-953.

Openshaw, S. 1984. The modifiable areal unit problem. Concepts and techniques in modern geography 38. Nowich: Geo Books.

Orr M, Bove F, Kaye W, Stone M. 2002. Elevated birth defects in racial or ethnic minority children of women living near hazardous waste sites. Int J Hyg Environ Health 205:19-27.

Pickle LW, Mungiole M, Jones GK, White AA. 1999. Exploring spatial patterns of mortality: the new atlas of United States mortality. Stat Med 18:3211-20.

Ratcliffe JH. 2001. On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. Int J Geographical Information Science 15:473-485.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org, last accessed on August 8, 2006.

Rogers MY. 1999. Getting started with Geographic Information Systems (GIS): a local health department perspective. J Public Health Manag Pract 5:22-33.

Rowlingson BS, Diggle PJ. 1993. SPLANCS: spatial point pattern analysis code in S-Plus. Computers and Geosciences 19:627-655.

Ruiz MO, Remmert D. 2004. A local department of public health and the geospatial data infrastructure. J Med Syst 28:385-395.

Rushton G. 2003. Public health, GIS, and spatial analytic tools. Annu Rev Public Health 24:43-56.

Rushton G, Krishnamurthy R, Krishnamurti D, Lolonis P, Song H. 1996. The spatial relationship between infant mortality and birth defect rates in a U.S. city. Stat Med 15:1907-1919.

Rushton G, Lolonis P. 1996. Exploratory spatial analysis of birth defect rates in an urban population. Stat Med 15:717-726.

SaTScan. 2006. Available at http://www.satscan.org, last accessed on August 8, 2006.

Schulman J, Selvin S, Shaw GM, Malcoe LH. 1993. Exposure misclassification due to residential mobility during pregnancy in epidemiologic investigations of congenital malformations. Arch Environ Health 48:114-119.

Shaw GM, Malcoe LH. 1992. Residential mobility during pregnancy for mothers of infants with or without congenital cardiac anomalies: a reprint. Arch Environ Health 47:236-238.

Shaw GM, Selvin S, Swan SH, Merrill D, Schulman J. 1988. An examination of three spatial disease clustering methodologies. Int J Epidemiol 17:913-919.

Talbot TO, Kulldorff M, Forand SP, Haley VB. 2000. Evaluation of spatial filters to create smoothed maps of health data. Stat Med 19:2399-408.

Tatham LM, Bove FJ, Kaye WE, Spengler RF. 2002. Population exposures to I-131 releases from Hanford Nuclear Reservation and preterm birth, infant mortality, and fetal deaths. Int J Hyg Environ Health 205:41-48.

Thrall SE. 1999. Geographic information system (GIS) hardware and software. J Public Health Manag Pract 5:82-90.

U.S. Census Bureau. 2001. Census 2000 Summary File 2 - Technical documentation, prepared by the U.S. Census Bureau, Washington, DC: Department of Commerce.

U.S. Department of Health and Human Services. 2000. Healthy People 2010 (conference ed., 2 vols.), Washington, DC: USDHHS.

Vine MF, Degnan D, Hanchette C. 1997. Geographic information systems: their use in environmental epidemiologic research. Environ Health Perspect 105:598-605.

WinBUGS. 2006. Available at http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml, last accessed on August 8, 2006.

Wasserman CR, Shaw GM, Selvin S, Gould JB, Syme SL. 1998. Socioeconomic status, neighborhood social conditions, and neural tube defects. Am J Public Health 88:1674-80.

Xiang H, Nuckols JR, Stallones L. 2000. A geographic information assessment of birth weight and crop production patterns around mother's residence. Environ Res 82:160-167.

Yang DH, Bilaver LM, Hayes O, Goerge R. 2004. Improving geocoding practices: evaluation of geocoding tools. J Med Syst 28:361-370.

Table 1. Publications on reproductive outcomes and birth defects when GIS was used in the United States

Author(s) and Year of Publication	GIS Application (Role) in the Study	Association / Outcome Examined
Andes and Davis, 1995	Identification of spatially homogeneous regions; assessing spatial compatibility; allocating geographical units across boundaries	Community characteristics and infant mortality
Banerjee et al., 2003	Geostatistical modeling; mapping hazard rates	Infant mortality
Blake and Bentov, 2001	Mapping spatial distribution by ZIP Code; linkage with Census data	Teen births and sociodemographic variables
Forand et al., 2002	Mapping elevated / lower birth defect rates based on spatial analysis using the scan statistic	Birth defect rates
Gilboa et al., 2005	Geocoding of residence; distance calculation from air pollution monitor and exposure assignment	Air pollution and birth defects
Ihrig et al., 1998	Link exposure data to study subjects with geocoded addresses, and link with census data	Arsenic exposure and stillbirths
Krieger et al., 2003	Geocode cases to and compare by different geographical level (ZIP Code, census tract, census block group)	Socioeconomic inequalities among children with low birth weight and childhood lead poisoning
Orr et al., 2002	Geocode waste sites and residential addresses to census tract level, and assign exposure categories	Hazardous waste sites and birth defects
Rushton and Lolonis, 1996	Geocode addresses of cases with birth defects; grid and spatial filter were used to measure rates	Birth defect rates
Rushton et al., 1996	Geocode addresses of cases with birth defects and infant deaths; grid and spatial filter were used to measure mortality	Birth defects and infant mortality
Talbot et al., 2000	Evaluate spatial filters to create smoothed maps by buffering, overlaying birth data geocoded to ZIP Code centroids	Low birth weight rates
Tatham et al., 2002	Grids with assigned exposure category were linked to geocoded addresses	Radiation exposure and preterm birth, fetal death, infant mortality
Wasserman et al., 1998	Link geocoded addresses of cases with neural tube defects and controls to census data to obtain neighborhood socioeconomic indicators	Socioeconomic status and neural tube defects
Xiang et al., 2000	Satellite images were used to determine crop type and exposure; buffering around residences	Crop production relative to maternal residence and birth weight

Reference it as:

Csaba Siffel, Matthew J Strickland, Bennett R Gardner, Russell S Kirby, Adolfo Correa: Role of Geographic Information Systems in Birth Defects Surveillance and Research.

Birth Defects Research Part A: Clinical and Molecular Teratology, 2006, 76(11):825-833.