Appendix 11.2 Use of Geographic Information Systems (GIS) to Map Data

The application of Geographic Information Systems (GIS) methods has become an integral component of aggregating, analyzing, and evaluating health data. The current practical applications of GIS in epidemiologic studies range from descriptive statistics (i.e., plotting data on a map) to evaluation of spatial relations between environmental exposures and health outcomes.

Several definitions exist for geographic information systems. One of the most recent, as found in *Healthy People 2010*, defines GIS as "powerful tools combining geography, data and computer mapping" (U.S. Department of Health and Human Services, 2000). Software packages available today, such as ArcMap and MapInfo, integrate many GIS functions. These include (1) database management, (2) data manipulation and analysis, and (3) data presentation (i.e., displaying data on a map). To be included in GIS, the data should have some kind of geographical or spatial component that can be translated into digit maps.

Digital Map Formats

GIS applications use either a vector or a raster map format, or a combination of the two. In *vector* maps (Figure A11.2-1) 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 coordinates of the different objects. The vector format is the most commonly used in public health. In *raster* maps the data are stored as digital images (e.g., orthophotos, scanned maps) (Vine et al., 1997). Usually a grid cell is used to represent a feature, and these cells can be connected. As such, smaller cells provide a more detailed resolution. Obtaining quality maps for a given geographical area for the time period of interest is crucial as maps are static while environments change.

Bringing Health Data into GIS

Ultimately, the application of GIS to birth defects data requires the transformation, as accurately as possible, of health records containing addresses or location information into geographic objects. This process is called *geocoding*, also known as address matching. During geocoding, latitude and longitude coordinates are assigned by the GIS software to each address by matching against an *address-range* (i.e., street segment) in a street reference map such as the Census Topologically Integrated Geographic Encoding and Referencing (TIGER) files (Croner et al., 1996). Interpolation is used to estimate the actual address location within the given range (Rushton, 1999). The address match rate depends on several factors, including the completeness of addresses in health records and the accuracy of reference maps (McElroy et al., 2003). In case of incorrect or missing house numbers and/or street names, coordinates are usually assigned to a centroid of a larger geographical entity, such as a census tract or a ZIP code. If available, other reference files such as tax parcel databases can also be used for geocoding purposes. Alternatively, in areas where latitude and longitude coordinates have not been predetermined (e.g., rural communities), a global positioning system (GPS) device can be used, although this may prove time and resource demanding. Figure A11.2-2 shows an example of how a point is placed within an address range.

Figure A11.2-1 An Example of GIS Data Layers

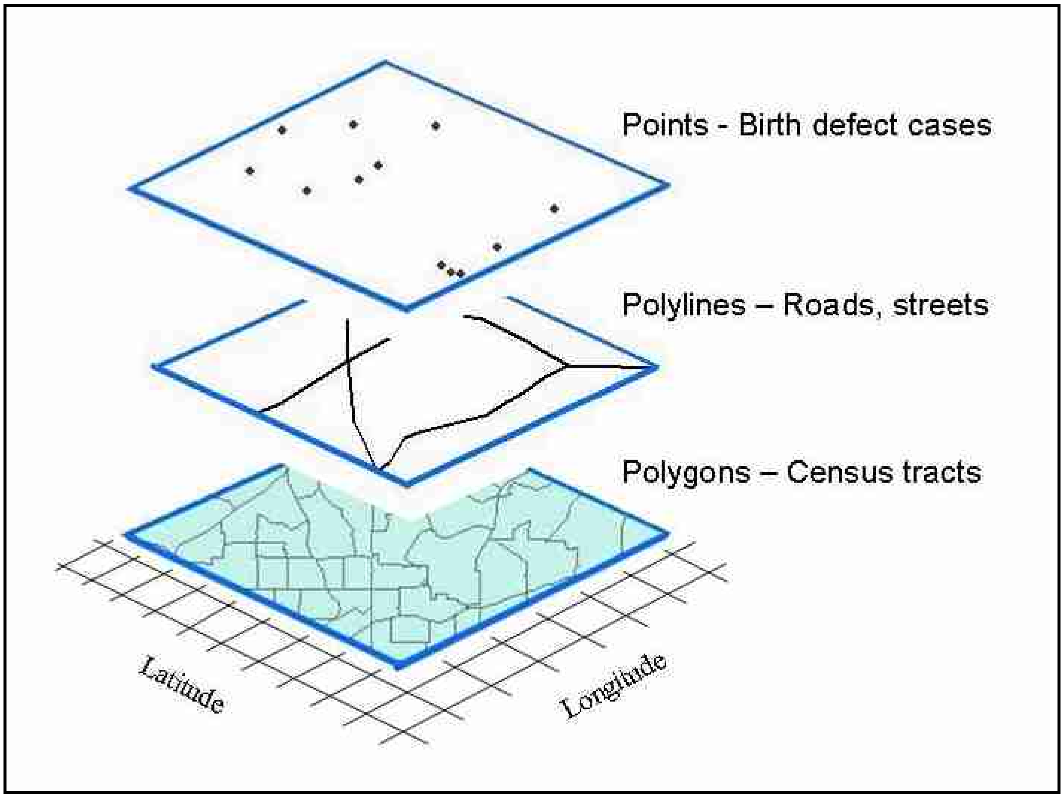
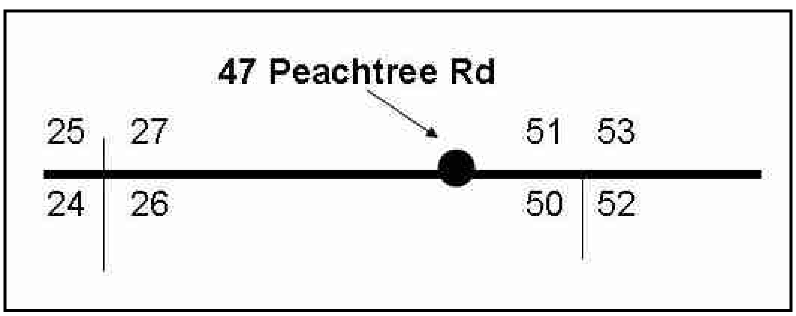


Figure A11.2-2 A Specific Location Within an Address Range



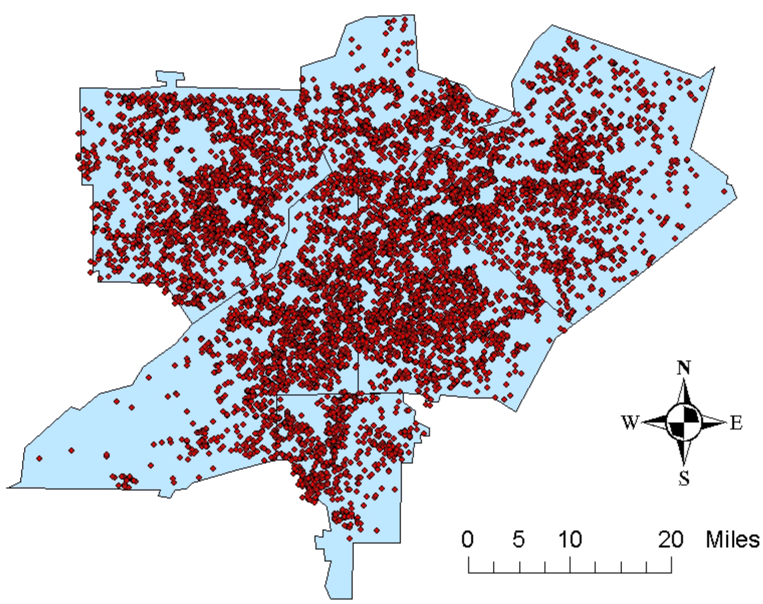
Mapping Data

Once health data are brought into a GIS database, users need to be aware of several important issues in data mapping. For example, different spatial databases must have the same scale and projection (McLafferty and Cromley, 1999). Otherwise data will 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. It is also an expression of how much the area represented has been reduced on the map. The smaller the scale, the larger the area displayed on a map. *Map projections* are attempts to portray/transform the surface of the three-dimensional earth or a portion of the earth on a flat map using a mathematical model. Some distortions of conformality, distance, direction, scale, and area always result from this process. Maps that focus on maintaining one feature (e.g., preserving distance) must distort other features (e.g., area, shape). Maps that accurately reflect area are called *equal-area maps*, while maps that correctly show the distance between points are called *equidistant maps*.

Two types of maps frequently used in public health research are dot-density and choropleth maps (Rogers, 1999).

Dot-density mapping. *Dot-density maps* are the simplest way to display events. These maps use dots or other symbols to represent the number of occurrences of a given data characteristic (Thrall, 1999). Each dot or symbol used on the map may represent a single entity (one dot = one case) or a group (one dot = 1,000 people). Dot-density maps are useful for area comparisons. However, dot-density maps need to be interpreted with caution regarding the "symbol to data characteristic" ratio. It is also important to keep in mind that dots do not always indicate the exact location of the data. An example of a dot-density map for metropolitan Atlanta is presented in Figure A11.2-3 (Source: Siffel et al., 2006, Figure 1, p. 828).

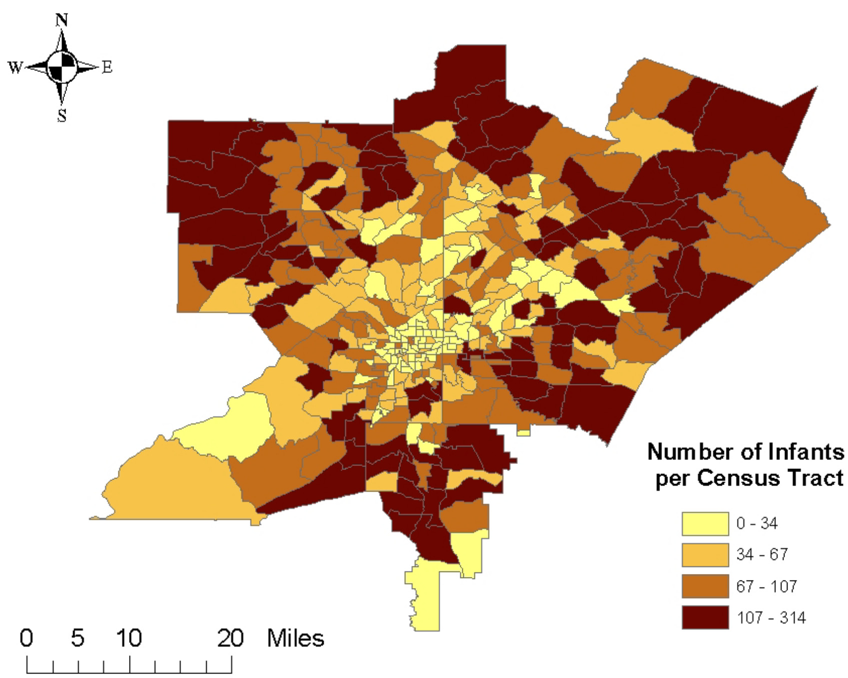
Figure A11.2-3 A Dot-Density Map of Metropolitan Atlanta



Choropleth mapping. *Choropleth maps* are area maps in which polygons (e.g., census tracts, counties) are shaded, colored, or patterned according to the extent to which a given attribute (such as population size or disease rate) is associated with each polygon. Choropleth maps are also called *thematic maps* or *shaded maps*. An example of a choropleth map for metropolitan Atlanta is presented in Figure A11.2-4 (Source: Siffel et al., 2006, Figure 2, p. 828).

It is important to choose the right characteristics for map presentations as the choice of color, pattern, size, polygon shape, and class intervals can impact how one interprets the information presented in a map. Single-color maps with varying color intensity (shades) are often an effective means of presenting data, but the use of differing patterns can help a black-and-white or grey-scale map. Similar-size polygons are recommended to the extent possible, as a few large polygons can dominate a map, leading to misinterpretation of information. Proportions or rates can be displayed by different class interval schemes, such as equal intervals (equal ranges of values) or quintiles (equal number of polygons falling into each class defined by dividing the range of values). The latter method is particularly useful for presenting skewed data. These methods are standard in GIS software.

Figure A11.2-4 Choropleth Map of Infants per Census Tract in Metropolitan Atlanta, 1990



Additional Technical Details

Maps showing point locations or even aggregate data in a small geographic area have the potential to reveal the identity of individuals (Cox, 1996). Therefore, as noted elsewhere in this chapter (see Appendix 11.1 on Data Suppression), one must generally limit the presentation of disaggregated birth defects information. While GIS methods and techniques exist for protecting privacy and limiting disclosure of information by geographically masking individual records (Armstrong et al., 1999), the use of masked data in small-area analysis can limit one's ability to detect clusters of cases (Kamel Boulos et al., 2005). As such, careful choice of geographical units and data aggregation are vital.

Below we present several practical suggestions for preparing and presenting maps above and beyond those already mentioned.

- The use of the same scale, colors, class intervals, and legends when presenting a series of maps.
- The inclusion of a scale bar and a "North" arrow.
- The use of patterns when printing in black and white. Color maps produced on a black-and-white printer usually do not provide as good results as grey scale.
- Avoid the use of red and green on the same map.
- Be wary of font-related problems. If symbols, which are special GIS fonts, are used on a map, do not export the map as an MS Windows meta file (.wmf). This type of file requires access to the GIS fonts. Similarly, do not include such files in presentations being made on an unfamiliar computer. If the GIS fonts are not available, other fonts will be substituted for symbols in the image. Instead, export your maps as JPEG files.

References on Geographic Information Systems

CITED REFERENCES

Armstrong MP, Rushton G, Zimmerman DL. Geographically masking health data to preserve confidentiality. *Stat Med.* 1999;18(5):497-525.

Cox LH. Protecting confidentiality in small population health and environmental statistics. *Stat Med.* 1996;15(17–18):1895-1905.

Croner CM, Sperling J, Broome FR. Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships. *Stat Med.* 1996;15(17-18):1961-1977.

Kamel Boulos MN, Cai Q, Padget JA, Rushton G. Using software agents to preserve individual health data confidentiality in micro-scale geographical analyses. *J Biomed Inform.* Aug 10, 2005. (Epub ahead of print)

McElroy JA, Remington PL, Trentham-Dietz A, Robert SA, Newcomb PA. Geocoding addresses from a large population-based study: lessons learned. *Epidemiology.* 2003;14(4):399-407.

McLafferty S, Cromley E. Your first mapping project on your own: from A to Z. *J Public Health Manag Pract.* 1999;5(2):76-82.

Rogers MY. Getting started with Geographic Information Systems (GIS): a local health department perspective. *J Public Health Manag Pract.* 1999;5(4):22-33.

Rushton G. Methods to evaluate geographic access to health services. *J Public Health Manag Pract.* March 1999;5(2):93-100.

Siffel C, Strickland MJ, Gardner BR, Kirby RS, Correa A. The role of GIS in birth defects surveillance and research. *Birth Defects Research (Part A).* November 2006;76(11):825-833.

Thrall SE. Geographic information system (GIS) hardware and software. *J Public Health Manag Pract.* 1999;5(2):82-90.

U.S. Department of Health and Human Services. *Healthy People 2010* (conference edition, 2 volumes). Washington, DC: US DHHS; 2000.

Vine MF, Degnan D, Hanchetter C. Geographic Information Systems: Their use in environmental epidemiologic research. *Environmental Health Perspectives.* 1997;105(6):598-605.

SUGGESTED READINGS/REFERENCES

Cromley EK, Cromley RG. An analysis of alternative classification schemes for medical atlas mapping. *Eur J Cancer.* 1996;32A(9):1551-1559.

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

Melnick AL. *Introduction to Geographic Information Systems in Public Health.* Gaithersburg, MD: Aspen Publishers, Inc.; 2002.

Melnick AL, Fleming DW. Modern geographic information systems: promise and pitfalls. *J Public Health Manag Pract.* 1999;5(2):viii-x.

Peters A, MacDonald H. *Unlocking the Census with GIS.* Redlands, CA: ESRI Press; 2004.

Thrall GI. The future of GIS in public health management and practice. *J Public Health Manag Pract.* 1999;5(4):75-82.

Tufte ER. *The Visual Display of Quantitative Information.* 2nd edition. Cheshire, CT: Graphic Press; 2004.