Mapping the Environment

Maps have been a key part of public health decision making since 1854, when John Snow mapped the association between cholera and London's Broad Street station water pump. Today, geographic information system (GIS) technology is used to compile and present the epidemiological data required for making public health risk assessments. The primary purpose of a GIS is to associate data—sometimes huge volumes of it—with a point, line, or area on a map. A relational database program within the GIS endows it with analytic capabilities, setting it apart from thematic mapping programs such as The Cancer Atlas. The data entered into a GIS database can come from a wide variety of sources, and these sources may be of varying resolution (representing anything from a census block to a town, county, or state) as the software segregates each source's data set into a separate data "layer." Once they are layered atop one another, the software clarifies the geographical relationships between the various sets of data, and acts as a tool for analyzing and exploring their spatial and temporal relationships. A GIS can be used in environmental health research to aggregate many sources of data to promote understanding of complex, multidimensional relationships between pollution and disease. Sources of data in an environmental health GIS may include demographic data from the U.S. Census, exposure databases such as the EPA's Toxic Release Inventory, and disease registries for cancer and birth defects.

For years scientists have used pens or pins to mark U.S. Geological Survey (USGS) maps in order to visualize the geographic component of disease incidence and hazardous sites. The idea of using computers as a means to make maps first occurred in the 1960s, when professional geographers became intrigued by the possibility of associating data with specific geographic locations. By the mid-1970s, U.S. government researchers involved in resource management and land-use planning had discovered GISs' merits. When the EPA jumped on the bandwagon in the mid-1980s, nearly every government agency involved with environmental subjects was using the technology and building databases of information.

The potential utility of these databases for epidemiological analysis was realized in the early 1990s when, for the first time, the data from the 1990 U.S. Census was distributed in GIS form. Called TIGER (for Topographical Integrated Geographic Encoding and Referencing system), the census database includes much more than just who lives where. It shows where they live, as well as most of the structures that support their lifestyles: roads, railroads, hydrography, power lines, pipelines, even the locations of schools and churches. Much of the infrastructure data came from USGS digital line graphs, which were incorporated into the census data. Armed with this enhanced census, the Centers for Disease Control and Prevention, as well as many individual state and local health departments, began using a GIS to maintain their registries of disease incidence. By 1995, the *Epidemiology Monitor* pronounced the emerging use of GIS technology one of the ten most notable developments in the field of epidemiology.

The most powerful GIS software, including ARC/INFO (manufactured by the Environmental Science Research Institute (ESRI) in Redlands, California) and GIS Office (manufactured by Intergraph in Huntsville, Alabama), must run on large, powerful computers such as engineering workstations; learning to use this software demands either a great deal of dedication or a part-time programmer. The programs that can be used with desktop computers, notably ESRI's ArcView software and MapInfo (manufactured by MapInfo in Troy, New York), are considerably easier to use. Almost all GIS software is now capable of incorporating "raster" photographic data that are harvested from satellites. However, the vast majority of the GIS data that interest people in public health take the form of vector data—data associated with geometrical constructs like points, lines, and areas, rather than a photographic image.

A Blessing and a Curse

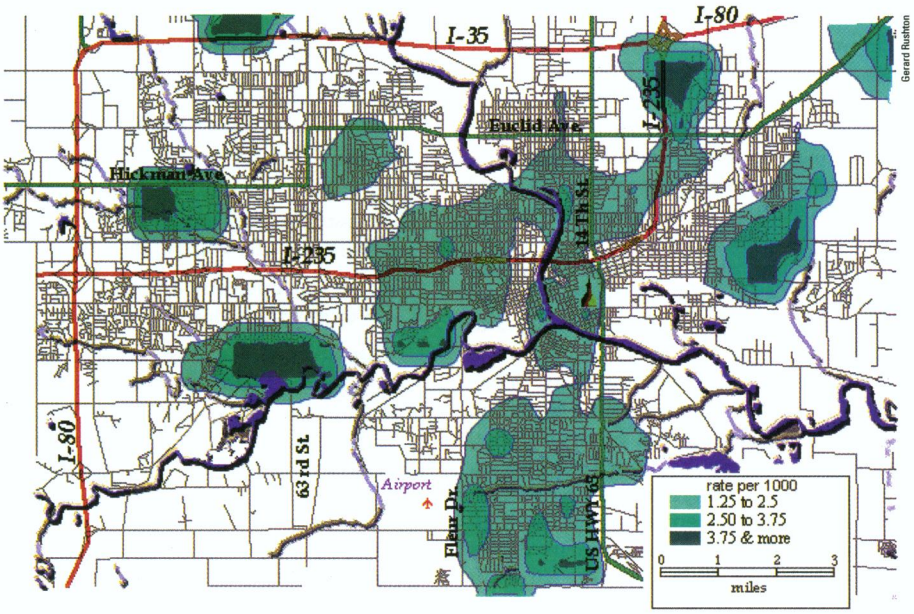
A picture may be worth a thousand words, but only if it accurately represents those words' meaning. GISs' facility for digesting vast amounts of data from a variety of sources has inspired a growing number of public health pioneers to embrace it. A profusion of government data, from census tracts to geographic features to disease registries, is available in GIS form, and the maps that the software draws are a universally accessible means for presenting risk information to the public. Yet even the

technology's most ardent enthusiasts qualify their praise, because the pictures a GIS paints can appear to convey more than they actually do.

"It looks like it's the solution to all your problems, but there's a great deal that we don't know how to do, and a lot of tools have to be developed to exploit this potential," says David Ozonoff, chairman of the environmental health department at the Boston University School of Public Health. "We don't really know whether it's going to pay off or not, though it seems like it should." The most promising harbinger of its potential for environmental risk assessment is the success of a GIS model that predicts the conditions under which people are most likely to contract Lyme disease, documented by G.E. Glass and colleagues in the July 1995 issue of the *American Journal of Public Health.*

The power of a GIS to seamlessly link together data gleaned from a wide variety of sources and present them in such an accessible and professional fashion has proven both a blessing and a curse. Everyone agrees that the software greatly expands the ability of researchers to consider the spatial component of exposure. "Because of the difficulty dealing with the spatial component of things, that extremely important information about where somebody was located was often crudely used or thrown away," recalls Ozonoff. With a GIS, "old fashioned epidemiology becomes easier and more productive, because you get to use information that before was lost to you."

The problem, ironically, is that once epidemiological and exposure data are pulled into a GIS, the polished presentation afforded by the software can make the data look more conclusive than they truly are. Unfortunately, this has resulted in some poorly executed studies, in addition to a volume of more rigorous work. The situation "is similar to what we went through in the '70s with computer-generated information. It was a given that if a computer gave you a result, no one really questioned it," says J.R. Nuckols, associate professor of environmental health at Colorado State University in Fort Collins. "We need to be very, very judicious about how information [is] derived. There needs to be 'meta-data,' or 'meta-information,' that goes along with every piece of information that's produced to tell how it was derived. In order to mesh with the standards of reporting risk established by epidemiologists, we in the exposure assessment community need to develop methods to bring along our calculations of uncertainty that are associated with making the predictions of exposure."



GIS in action. Researchers at the University of Iowa in Iowa City used a GIS to match the rate of birth defects in surrounding Polk County from 1983–1990 with U.S. Census data to determine geographical areas where such defects turned up more consistently.

Still, there is universal consensus that a GIS can be a useful aid at the beginning of an environmental epidemiology or risk assessment study. "It's a big tool for exploratory data analyses, generating hypotheses, and modeling exposures," says Paul English, a research specialist at Impact Assessment, Inc., in La Jolla, California, which contracts for the California Department of Health Services. A year and a half into an NIEHS-funded study, English is using a GIS to correlate sentinel health events with environmental changes at the California–Mexico border.

English's group is using TIGER census data for two counties, together with other data layers. These layers include details about ambient air quality (calculated based on data from eight air quality monitoring stations), pesticide use, traffic density, and contaminated wells. Thus armed, the researchers are looking at the incidence of childhood asthma, childhood cancer, birth defects, and infectious diseases that could be correlated with environmental changes. "The methodologies we're using require a GIS," English explains. The air quality model he's using takes the data from the monitoring stations and calculates the spread of emissions over the entire area to estimate exposure to contaminants. "These types of interpolation methods are a lot easier to do with a GIS," he says. "Geocoding [linking individual addresses with their locations on a map] thousands of cases by hand would be very tedious, if not impossible. Linking all the data by geographic coordinates in one database is really impossible to do without a GIS."

Refining Exposure and Reducing Misclassification

The ability to use GIS-based technology to predict, quantify, and locate contaminants as they disperse through the ambient environment should reduce the number of misclassification errors commonly made when scientists use less-exacting methods of assessing exposure, according to Nuckols. "The GIS allows you to have units of analysis at a refined enough scale so that you can, in relatively small areas, do things like simulation modeling and data analysis that will allow you to differentiate degrees of exposure," he says, citing a pilot study he conducted on the exposure of a farming population in an agricultural area to pesticide residues, which was published in the proceedings of the Summer 1996 meeting of the National Institute for Farm Safety. "Because we can use technology such as remote sensing to identify crop species, and we know information on pesticides used in farming areas for certain crops, we can go down to a unit of analysis of less than 100 acres, the field level. By having information at this scale for analysis of data such as crop [type], soil, geology, water resources or supplies, and hydrology, we can start stacking data into a GIS and feed that information into a simulation model to predict where applied farm chemicals will end up in the environment." Nuckols says that he is finding a common theme in the studies he's been conducting. By refining the level of

resolution for exposure classification in environmental epidemiology studies, the resultant determinations of relative risk are significantly different—and usually greater—than those calculated using more traditional exposure classification methods. "GIS-based technology has greatly increased our capabilities in making such refinements," Nuckols stresses.

A GIS can also help support decision making, as demonstrated by Gerard Rushton, a professor of geography and adjunct professor of preventive medicine at the University of Iowa in Iowa City. When the *Des Moines Register* ran an article in July 1993 alleging that environmental exposures such as abandoned coal mines, toxic chemical dumps, and a nearby military base were responsible for the high numbers of birth defects and infant deaths in surrounding Polk County, Rushton used a GIS to evaluate the assertion. Under his direction, the Iowa Department of Health employed the then-just-released TIGER data to look at the changing rates of incidence, using data smoothing methods to plot the incidence rate on a grid. Rushton and a student of his wrote their own software program using Monte Carlo simulation methods, which show how rates of birth defects and infant mortality would vary across the area if each child had the county's average chance of having a birth defect or dying in infancy. "We were able to make a map that showed areas in Des Moines where the rates [of birth defects and infant death] were highly likely to be significantly higher than the null hypothesis," he says. "The areas were different than the areas the newspaper had counted." After Rushton's data were publicized, the allegations were dropped.

Supplemental tools like Rushton's statistical software program are crucial to making a GIS a truly useful tool for studying the epidemiology of environmentally induced disease. Though there are quite a few modeling tools available—thanks to the efforts of earlier communities of GIS users such as land-use planners—statistical tools are in somewhat short supply. The main GIS programs include only a very limited selection of statistical routines. At present, the only direct connections researchers can make use data networked between a few of the better known GIS and statistical software packages. Other statistics programs do not work directly with GIS data. For example, SAS Institute in Cary, North Carolina, maker of SAS statistical software, is currently offering an add-on module with GIS capabilities of its own. "The people who know the GIS very well are mostly interested in adding in the statistics. And the people who know the statistics want to add the GIS capabilities to them," observes Lance Waller, an assistant professor of biostatistics at the University of Minnesota in Minneapolis. "We're sort of at this awkward stage of these things coming together. It really needs to be an interdisciplinary kind of development."

GIS technology will still face significant hurdles even after more statistical tools are developed and standard methodologies for using them are established. "I think the challenge is for epidemiologists to try to demonstrate that these spatial tools will provide better insight and more informative analyses in order to bring this approach into the mainstream of environmental epidemiology," says Daniel Wartenberg, an associate professor in the Department of Environmental and Community Medicine at the Environmental and Occupational Health Sciences Institute, which is affiliated with the Robert Wood Johnson Medical School and Rutgers University in Piscataway, New Jersey. "We need to show why that kind of approach really gives a more accurate answer than simply ignoring location."

Regardless of whether everyone agrees that GIS technology provides better insights, it appears to be here to stay. "I don't think there's a choice," says Nuckols. "One of the most interesting things to me is that this has really been a grassroots movement. It's in the field offices where people down in the trenches say this technology has merit . . . The field [scientists] will find a way to apply the technology. The role of the research community is to work to refine methods by which they can most successfully improve this application, and to evaluate its public and ecological health benefits."

Kellyn S. Betts

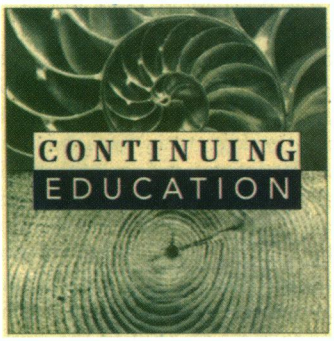
Suggested Reading

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

Nuckols JR, Berry JK, Stallones L. Defining populations potentially exposed to chemical waste mixtures using computer-aided mapping analysis. In: Toxicology of Chemical Mixtures (Yang RSH, ed). Orlando, FL:Academic Press, 1994; 473–504.

Waller LA. Epidemiologic uses of geographic information systems. Stat Epidemiol Rep 7(Spring/Summer):1,4–7 (1996).

Wartenberg D. Use of geographic information systems for risk screening and epidemiology. In: Hazardous Waste and Public Health: International Congress on the Health Effects of Hazardous Waste (Andrews JS, Frumkin H, Johnson BL, Mehlman MA, Xintaras C, Bucsela JA, eds). Princeton, NJ:Princeton Scientific Publishing Company, Inc., 1994, 853–859.



Duke

Nicholas School of the Environment
Center for Environmental Education

Intensive Courses and Workshops in Natural Resource and Environmental Management

For more information call or write:
Continuing Education, Nicholas School of the Environment, Center for Environmental Education
Box 90328, Durham, NC 27708-0328.
Telephone (919) 613-8082; Fax (919) 684-8741; e-mail: britt@duke.edu