SPECIAL EDITION

The Ribbon

BCERF

Volume 8 • Number 1 • Early Spring 2003

A Newsletter of the Cornell University Program on Breast Cancer and Environmental Risk Factors in the New York State (BCERF)

Mapping Disease: Deciphering Geographic Patterns From Cholera to Breast Cancer

Introduction to this issue by Zev Ross

Zev Ross is a graduate student in the Department of Natural Resources at Cornell University. He spent last summer assisting California's Environmental Health Investigation Branch with a spatial analysis of breast and testicular cancer data.

The more than 200-year history of disease mapping is filled with examples of maps that helped provide etiological clues to diseases from cholera to lung cancer. Geographic patterns in disease, such as those discussed in this issue of The Ribbon, can help provide insight into disease incidence and mortality or help identify environmental sources of risk. While geographic patterns can be an important research tool, these patterns can also pose statistical challenges. Patterned data violates the fundamental statistical assumption of independence and ignoring this violation can lead to distorted statistical results. In 2002, as part of a larger project to look at geographic variations in breast cancer rates, we initiated a project to look at geographic patterns in breast cancer and, more specifically, to evaluate their affect on statistical models. We expect to have results from our analysis in May 2003.

Maps to Solve Medical Mysteries

It was the terrifying epidemic diseases of the 18th and 19th centuries that first prompted researchers to develop maps of disease. In contrast to often dreary statistical tables, maps can (and did) sharpen otherwise obscured relationships between disease and possible environmental influences. With maps, researchers could look at the location of cholera or yellow fever deaths in relation to water wells, garbage dumps, or outhouses.

Dr. John Snow's map of cholera deaths in relation to London's water pumps, for example, was one of the first, and perhaps the most celebrated, disease maps. With the help of his famous map, Snow was not only able to track the source of what he called "the most terrible outbreak of cholera which ever occurred in this kingdom," but he was able to convince authorities to take action against the disease (2). His map (next page) demonstrated for future epidemiologists the value of maps as both a research *and* a communication tool.

But clearly disease mapping is not limited to epidemic diseases in the 1800s. Modern researchers have also successfully used geographic patterns to identify etiological clues. Examples include:

- Mapping of cancer mortality rates in the United States 1950-69 revealed "exceptionally" high rates of lung cancer along the eastern seaboard, particularly in parts of Georgia, Florida and Louisiana (3, 4). These patterns led to further investigation. Several studies found that exposure to asbestos among employees at shipyards accounted for a significant part of the excess mortality from lung cancer (3, 5, 6).

"Whether they are stalking bioterrorists or tracing the origin of a hepatitis outbreak, today's epidemiologists stand on the shoulders of John Snow (1813-1858), the English anesthesiologist whose methodical investigations showed that cholera spreads through polluted water."
www.ph.ucla.edu/epi/snow.html

index

An excerpt from John Snow's famous map of cholera deaths (denoted by black bars). People who died drank water from the Broad Street Pump (center). Source:UCLA Department of Epidemiology, John Snow Site, http://www.ph.ucla.edu/epi/snow.html.

- Researchers in the 1950s used the geographic patterns in malignant melanoma to identify causes. In his influential 1956 paper, H.O. Lancaster noted that rates of malignant melanoma were higher in the people of Southern Africa and Australia than in the countries from which they originated. He also noted that incidence increased with proximity to the equator in the United States, Australia and New Zealand and correctly proposed that sunlight may be a cause of the increased incidence. Due to the fact that tumors tend to be more common on non sun-exposed sites (the trunk in men and legs on women) sun had not been considered a risk factor prior to analyzing the geographic patterns (7, 8).

Significant Innovations Since the Sixties

In the time since Dr. Snow's cholera map, and more specifically in the past forty years, we've seen enormous innovations in mapping and techniques to analyze spatial patterns. Geographic Information Systems (GIS), for example, the topic of this issue of *The Ribbon*, were developed in the 1960s and have become widespread only in the past 10 years. Similarly, while traditional statistical techniques (such as regression) are hundreds of years old, statistics specific to spatial data evolved out of research in the 1960s. These innovations come alongside an exponential increase in the volume and quality of data on disease and environmental pollutants.

These innovations are particularly significant in light of the challenges presented by modern diseases. In contrast to diseases such as cholera or yellow fever, whose sources can be traced to a single bacterium or virus, diseases such as breast cancer appear to have more complex causes.

Breast Cancer's Enigmatic Geographic Patterns and Statistical Considerations

Breast cancer has posed one of the greatest challenges to researchers investigating geographic patterns. Age-adjusted breast cancer incidence and mortality exhibits a strong geographic pattern. For decades, researchers in the United States have noted substantial regional variation in breast cancer mortality rates, most notably a "regional excess" of breast cancer in the Northeast United States (9-13). Nationally, even after controlling for variables that, themselves, have a geographic pattern (e.g., race, socioeconomics), breast cancer mortality still exhibits a (slight) regional pattern. The source of the regional variation is hotly debated and has led to significant research on the issue.

"The cartography of disease owes its genesis to the abrupt, terrifying challenge which epidemic outbreaks presented, whereas endemic disease, more or less constantly active, offered no comparable stimulus to cartographic creativity. Plague, yellow fever, and cholera – all exotic – accomplished what tuberculosis could not." Jarcho (1969)(1).

In conjunction with Dr. Peggy Reynolds and collaborators at the California Department of Health Services and Dr. Patrick Sullivan, a biostatistician and Professor at Cornell University, I am looking at the statistical implications of geographic patterns in breast cancer incidence in California. Specifically, we are in the process of evaluating the affect of autocorrelation – the tendency for near things to be more similar than distant things – on parameters, predictions and confidence intervals.

Our analysis of geographic patterns relies on geostatistics, a statistical method that describes how data (in this case rates) are related with distance and direction. Unlike the more traditional nearest-neighbor analyses that rely on an arbitrarily defined neighborhood (e.g., adjacent neighbors), geostatistics can analyze the strength of relationships at the full range of distances.

Our preliminary results suggest that even after accounting for race, age, socioeconomics and urbanization, breast cancer incidence rates in California still exhibit autocorrelation. The known risk factors, however, account for a significant amount of the autocorrelation (and the variation) seen in unadjusted rates and the "left-over" autocorrelation appears to be minimal. The statistical and etiologic implications of the autocorrelation are still under investigation.

In the tradition of Dr. John Snow, the process of analyzing the geographic patterns will hopefully lead us to insights on breast cancer's origins and influences. This issue of *The Ribbon* describes several ongoing studies pursuing these questions in California, Massachusetts and New York State.

1. Jarcho S. The contribution of Heinrich and Hermann Berghaus to medical cartography. Journal of the History of Medicine and Allied Sciences 25:131-142(1969).
2. Snow J. On the Mode of Communication of Cholera. London:John Churchill, 1855.
3. Blot WJ, Harrington JM, Toledo A, Hoover R, Heath CW, Jr., Fraumeni JF, Jr. Lung cancer after employment in shipyards during World War II. N Engl J Med 299:620-4(1978).
4. Devesa SS, Grauman DJ, Blot WJ, Pennello GA, Hoover RN, Fraumeni JF. Atlas of Cancer Mortality in the United States: 1950-1994 NIH Publication No. 99-4564. Washington: National Cancer Institute, 1999.
5. Blot WJ, Davies JE, Brown LM, Nordwall CW, Buiatti E, Ng A, Fraumeni JF, Jr. Occupation and the high risk of lung cancer in Northeast Florida. Cancer 50:364-71(1982).
6. Blot WJ, Morris LE, Stroube R, Tagnon I, Fraumeni JF, Jr. Lung and laryngeal cancers in relation to shipyard employment in coastal Virginia. J Natl Cancer Inst 65:571-5(1980).
7. Hutt M, Burkitt D. The Geography of Non-Infectious Disease. Oxford:Oxford University Press, 1986.
8. Lancaser H. Some geographical aspects of the mortality from melanoma in Europeans. Medical Journal of Australia 1:1082-7(1956).
9. Blot WJ, Fraumeni JF, Jr., Stone BJ. Geographic patterns of breast cancer in the United States. J Natl Cancer Inst 59:1407-11(1977).
10. Blot WJ, McLaughlin JK. Geographic patterns of breast cancer among American women. J Natl Cancer Inst 87:1819-20(1995).
11. Sturgeon SR, Schairer C, Gail M, McAdams M, Brinton LA, Hoover RN. Geographic variation in mortality from breast cancer among white women in the United States. J Natl Cancer Inst 87:1846-53(1995).
12. Canto MT, Anderson WF, Brawley O. Geographic variation in breast cancer mortality for white and black women: 1986-1995. CA Cancer J Clin 51:367-70(2001).
13. Devesa SS, Grauman DJ, Blot WJ, Pennello GA, Hoover RN, Fraumeni JF. Atlas of Cancer Mortality in the United States: 1950-94:National Cancer Institute, 1999.

Mapping the Environment and Breast Cancer on Cape Cod, MA.

Theresa C. Kennedy and Julia G. Brody, Silent Spring Institute

EVER SINCE JOHN SNOW'S FAMOUS WORK, THEMATIC MAPS HAVE EMERGED AS VALUABLE TOOLS IN UNDERSTANDING THE RELATIONSHIP BETWEEN OUR ENVIRONMENT AND THE OCCURRENCE OF DISEASE.

Among breast cancer activists and researchers, geographic patterns have prompted concern in high incidence communities and spawned hypotheses about environmental factors. One of the most widely known environmental epidemiology studies, the Long Island breast cancer study, came about in part because women, concerned at the number of their neighbors with breast cancer, began pinpointing their neighbors' homes on maps laid out on their kitchen tables.

Today, Geographic Information Systems, or GIS, replace paper maps and transparent overlays with a sophisticated mix of computer hardware, software and expertise, all focused on the visualization and analysis of spatial data. Beyond simple mapping, GIS allows the researcher to integrate layer upon layer of different kinds of spatial information, and describe or quantify the spatial and temporal relationships between them. Gaining insight into these relationships can help form causal hypotheses, which in turn moves us closer to the goal of disease prevention.

The power and versatility of GIS is apparent in the many applications seen thus far in health studies. GIS is used in disease surveillance to monitor incidence by geographic units, such as zip code or county, and in environmental surveillance, for example, to track toxic releases or map air and water quality. By putting disease and environmental data together, GIS is also used in ecological analysis, to analyze 'clusters' of disease with respect to the sociological or environmental aspects of the affected population and place. Lastly, GIS is beginning to be used to assess environmental exposures to individuals in health studies.

Researchers approach this last use of GIS – individual exposure assessment for health studies – with caution because limitations in the quality of spatial data can be formidable. Differences in scale, resolution, accuracy, and completeness of the datasets are a constant hindrance, particularly for studies of diseases involving long latencies, like breast cancer, where exposures of interest may have occurred decades in the past. Collecting new data can be time intensive, expensive, and, for historical exposures, sometimes impossible.

Despite these challenges, the Cape Cod Breast Cancer and Environment study (Cape Cod Study), set out to confront the limitations and develop new techniques to reconstruct historical exposures to various contaminants on Cape Cod, MA, in an effort to see if certain exposures are related to breast cancer risk in the region. The study is conducted by Silent Spring Institute, a non-profit scientific research organization, in collaboration with researchers at Applied Geographics Inc., Boston University, Harvard, and Tufts.

In the Cape Cod Breast Cancer and Environment Study, Silent Spring Institute unites GIS, health, and environmental data to assess the exposures of 2,100 individual women to multiple environmental pollutants throughout the past 40 years.

The use of GIS in the Cape Cod Study was motivated by the gaps in other exposure assessment methods that limit our ability to identify environmental factors that may one day lead to prevention. Most of what we know about breast cancer risk relates to factors women can report in interviews: age at menarche or a first pregnancy, use of oral contraceptives and hormone replacement therapy, history of exercise and alcohol use, and so on. Biological and environmental sampling hold promise for the future, but tell us little about the past, and sampling is expensive at the scale needed in breast cancer epidemiology to uncover risks that are probably moderate in scale. Geographic data can add another dimension – answering questions about their environment that women cannot answer for themselves, and allowing estimation of environmental influences on a large scale. While the data developed in the study are specifically relevant to Cape Cod, the methods are widely applicable for environmental health studies and planning.

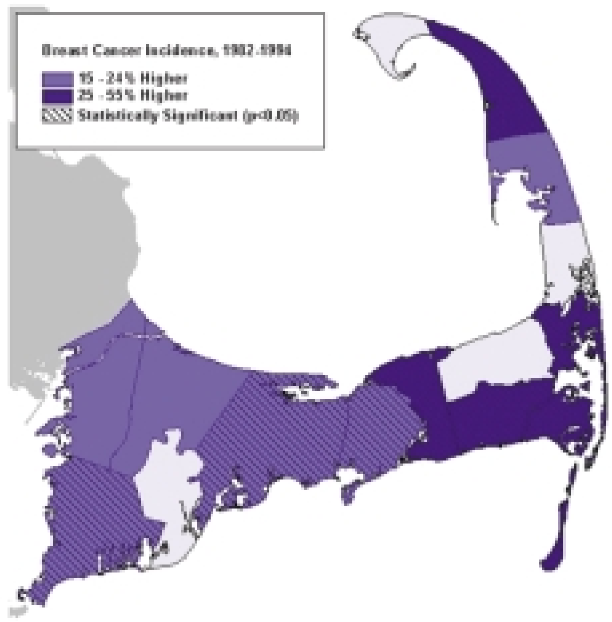


Figure 1 - Breast Cancer Incidence 1982 – 1994, compared to state average

Breast Cancer on Cape Cod, MA

The study was instigated in 1993 when the Massachusetts Cancer Registry (MCR) reported elevated incidence in a majority of Cape Cod towns compared with the rest of the state (1). Using the GIS to better estimate how many breast cancer cases would be expected in each census block group on the Cape over the 14-year history of the MCR, detailed disease surveillance revealed approximately 20% higher age-adjusted breast cancer incidence on Cape Cod for the period 1982 through 1994. (See Figure 1.) Using data from the Collaborative Breast Cancer Study, we learned that incidence was elevated in comparison with other areas of Massachusetts, even after statistically controlling for a long list of established and hypothesized risk factors for breast cancer, including family and reproductive history, physical exercise, alcohol, tobacco, and certain aspects of diet (2).

Cape Cod's history of pesticide use in support of tourism, cranberry cultivation and other agriculture is also quite distinctive (3). Forests were repeatedly sprayed for gypsy moth and other tree pests, and wetlands were sprayed for mosquito control. Other wide area uses include applications to manage golf courses and rights of way. The Cape's sandy soils allow pollutants to travel quickly to groundwater, which is the sole drinking water source and Cape Cod residences have been developed in or adjacent to pesticide use areas or on land where pesticides were previously applied. Persistent organochlorine chemicals including DDT and dieldrin were widely used on the Cape from the late 1940s to the mid 1970s, and less persistent compounds including carbaryl, malathion, and carbamates have been applied in more recent years (3).

In assessing which environmental exposures may be important to measure on Cape Cod, the study team started with what they knew about the already well-established risk factors. The most promising candidates were substances that mimic estrogen — a known breast cancer risk factor — or that have been shown to cause mammary tumors in animals. These substances include pesticides (insecticides, herbicides and fungicides) and other chemicals found in detergents, plastics and personal products which have been widely used on Cape Cod in the past, and may lead to exposure from their use and disposal in conjunction with certain characteristics of the Cape Cod environment.

The Cape Cod Breast Cancer and Environment Study

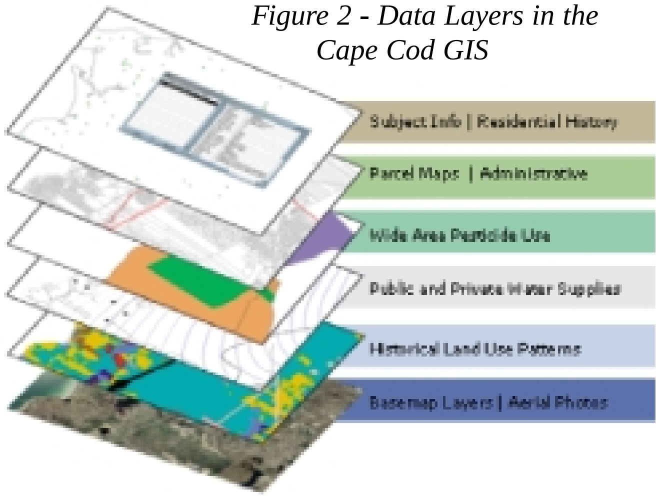
The Cape Cod study involves 2100 women who lived on Cape Cod between 1988 and 1995. Women diagnosed with breast cancer during those years were compared with women of similar age who had not been diagnosed. Interviewers collected information on where the women had lived on Cape Cod and on established risk factors for breast cancer, including family history of breast cancer, menstrual and reproductive history, height and weight, and education, an indicator of socioeconomic status. Interviews also included topics of interest in recent studies as possible breast cancer risk factors: alcohol and tobacco use, physical activity, and pharmaceutical hormone use. Other questions assessed use of home pesticides, tap water, and certain consumer products.

To measure historical exposures from pesticides and other chemicals that could not be ascertained by interview we designed and developed special GIS tools to model the historical Cape environment. One of these tools, the Spatial Proximity Tool, is described in more detail later.

Thematic data layers which served as input to the models (Figure 2) were gathered from federal and state sources, such as the United States Geological Survey (USGS), state agencies, and the Cape Cod Commission and were integrated into the GIS.

Land use maps from four periods beginning in 1951, supplemented with local information and other federal

Figure 2 - Data Layers in the Cape Cod GIS



datasets were used to assess the historical locations of forested areas, cranberry bogs, and other land use types known to have been regularly sprayed with pesticides. Information on local spraying activities was exhaustively researched, compiled, and mapped by Institute staff. Data on public water supply systems and private wells was gathered, and most importantly, the information collected at interview identifying our women's Cape Cod addresses was used to 'geocode' each woman's residential history during the study period,1948 - 1990, when the target chemicals were used.

A new GIS Tool to Measure Environmental Exposures: The Spatial Proximity Tool

The Spatial Proximity Tool was developed to relate the women's residential address history with the historical environmental data like pesticide spraying areas (Figure 3). The tool is based on modeling of spray drift and deposition, and takes into account the distance of a residence from the pesticide use area, the size of the area, and the wind direction typical during the early morning hours in spring when pesticides were applied. The goal was to reconstruct spatial, temporal, and intensity or "dose" information. Given the expected limitations of historical records, the focus was on assessing relative intensity – i.e., on correctly ranking higher and lower exposures and differentiating exposed from unexposed residences.

While these measures do not approach the ideal of historical biological exposure assessment, which, indeed, cannot be attained retrospectively, they offer insight into historical exposure patterns.

Mapping Residential History over a 40 year period

One of the significant challenges in the study was the 'geocoding' of residential history data gathered at interview. Geocoding is the process of taking a street address, such as '1 Main St. Hyannis', and translating it into a latitude and longitude so that it can be placed on a map. Early automated methods of geocoding women in our study involved using town parcel maps to place them in the center of the land parcel corresponding to their street address. The parcel maps however, like many publicly available datasets, had been created independently by each town on the Cape, using different source data, and at varying resolution and scale. The net result is that while the parcel maps are useful basemaps for each individual town, they do not represent a *uniform* basemap across the Cape, and are subject to local variation in accuracy and completeness. The environmental datasets, originating from uniform state- or nationwide basemap data, did not therefore 'fit' with the parcel maps in some areas, introducing the possibility of underestimating (or overestimating) a woman's exposure to a particular source. These problems of exposure misclassification are common in environmental epidemiology and remain serious barriers to accurately identifying health impacts of pollutants.

In order to overcome the limitations of the parcel maps, we used high resolution aerial photographs to move the women directly onto their house rooftops, using a process of on-screen editing and digitization. The aerial photographs, available from MassGIS (4), constitute the state basemap for Massachusetts, and represent a consistent cape-wide reference on which to map other information. This process, undertaken by Institute researchers and Applied Geographics Inc. (AGI), maximizes the quality of the residential history data, and has the additional benefit of reducing error associated with the significant number of large parcels found on the Cape, in which the center of a parcel may be some distance away from the actual residence.

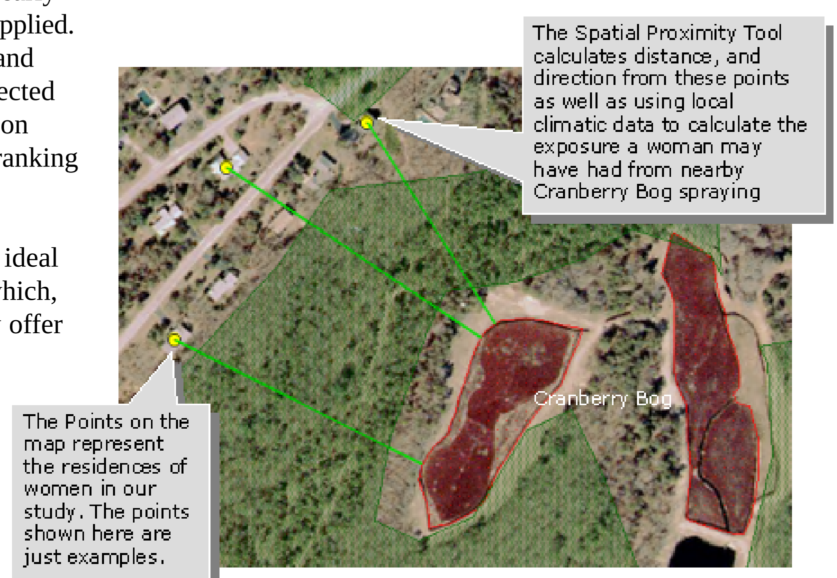


Figure 3 - The Spatial Proximity Tool

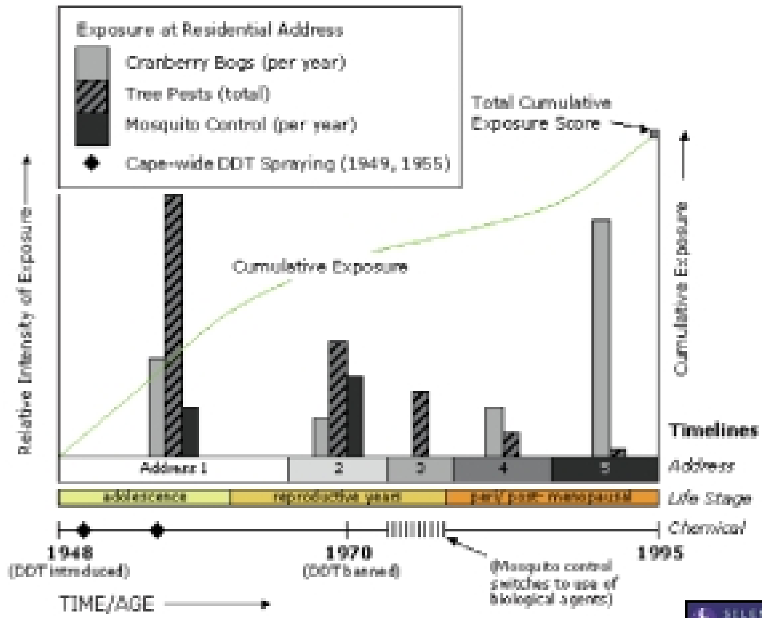


Figure 4 - Personal Exposure Profile (example only)

Building an Historical Exposure Profile for each woman

The Spatial Proximity Tool calculates relative exposure intensities for each source (e.g. an actively used Cranberry Bog) at each residential address during a particular time period in the study. A woman's total exposure score for each source, or group of like-sources (e.g. all aerially sprayed pesticides) was calculated by adding together all appropriate exposures for each year at each address over a woman's residential history (Figure 4). Cumulative exposure to residual pesticides, stemming from the previous application of persistent chemicals to residential areas or adjacent land, was also calculated.

Results of the analysis comparing aggregated exposure scores with breast cancer risk are in preparation.

Conclusions and Recommendations

Given that limitations due to disparate datasets and missing exposure information (such as we found in this study) are typical in retrospective environmental health studies, GIS may be seen as a tool for creatively, but judiciously, constructing proxies and developing methods for estimation and interpolation (3). The promise of GIS for this purpose is held back by the lack of statewide data, both present day and historical, and also by gaps in our understanding of the dynamics of chemicals in our environment. Standardized data collection, consistency across geographical boundaries, and long-term monitoring, are critical to the success of initiatives such as Health Track (5), in which the integration of many kinds of health and environmental data offer hope for studying the long term effects of the environment on health.

Despite the challenges, the data and tools developed during the Cape Cod study represent a rich resource for studying a wide range of health and environmental questions. The Spatial Proximity Tool in particular, while demonstrated here for historical analysis, can also be used to model future hypothetical events. With applications limited only by imagination and available data, GIS represents a promising tool for environmental stewardship, and protecting our health.

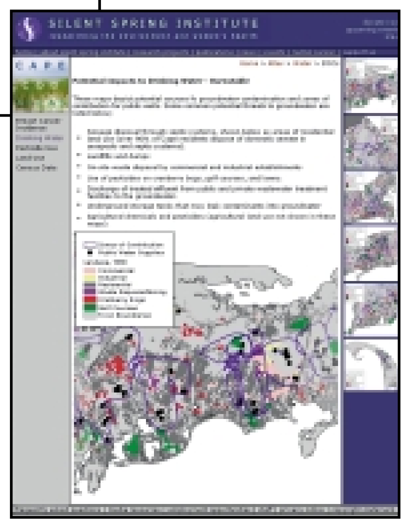


Figure 5

More information about Silent Spring Institute and the Cape Cod Breast Cancer and Environment Study, including an environmental data atlas (Figure 5), can be found at **http://www.silentspring.org**.

1. Brody, J.G.; Rudel, R.A.; Maxwell, N.I.; Swedis, S.R. 1996. Mapping out a search for environmental causes of breast cancer. Public Health Reports, 6(3): 494-507.
2. Silent Spring Institute, 1997. Cape Cod Breast Cancer and Environment Study: Final report, December 8, 1997, Newton, MA.
3. Brody, J.G.; Vorhees, D.J.; Melly, S.J.; Swedis, S.R.; Drivas, P.J.; Rudel, R.A. 2002. Using GIS and historical records to reconstruct residential exposure to large-scale pesticide application. Journal of Exposure Analysis and Environmental Epidemiology, 12: 64-80.
4. MassGIS. Massachusetts Geographic Information System. Boston, MA: MassGIS, Executive Office of Environmental Affairs. accessed on January 31, 2003 [accessed at: http://www.state.ma.us/mgis/massgis.htm].
5. Trust for America's Health. Trust for America's Health Reports. Washington, DC. Available at: http://healthyamericans.org/resources/reports/; accessed February 2003..

GIS Approaches to Studying Variations in Breast Cancer Incidence in California

Peggy Reynolds and Susan Hurley
California Department of Health Services

OBSERVED REGIONAL VARIATIONS IN BREAST CANCER INCIDENCE HAVE BEEN A SOURCE OF PUBLIC HEALTH CONCERN AS WELL AS, TRADITIONALLY, A SOURCE OF HYPOTHESIS GENERATION FOR FACTORS WHICH MIGHT CONTRIBUTE TO DIFFERENCES IN RISK.

The rates of breast cancer among women in the urban San Francisco Bay and Los Angeles areas of California have been historically higher than those in many other areas of the U.S. and the world (1). Rates in these areas have also been notably higher than rates in other areas of California (2, 3). Recently, a dramatic increase in rates for residents of Marin County, a small affluent area north of San Francisco, has been a topic much discussed in both the popular press and scientific literature (4).

The reasons for these variations are not well understood. Public concern over possible environmental contributors to such differences has prompted a series of studies undertaken by the California Department of Health Services (CDHS) designed to assess the influence of sociodemographic and environmental factors. These studies are built on the availability of a high-quality, population-based cancer surveillance system, improving information on environmental toxicants, a large well-defined cohort study and more widely available geographic information system (GIS) tools. Complementary funding from the National Cancer Institute and National Institute for Environmental Health Sciences (Grants No. U01-CA81789 and R01-CA77398), and from the California Breast Cancer Research Program (Grant No. 6JB-0111), support a multidisciplinary team of researchers with expertise in epidemiology, environmental health, statistics, geographic information systems and survey research to conduct these studies.

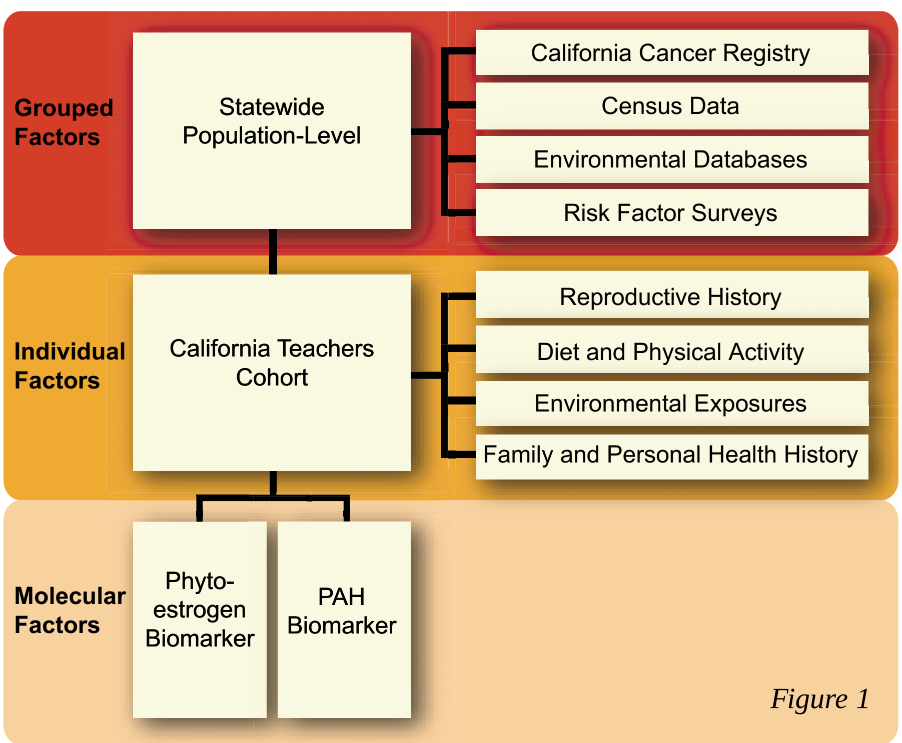
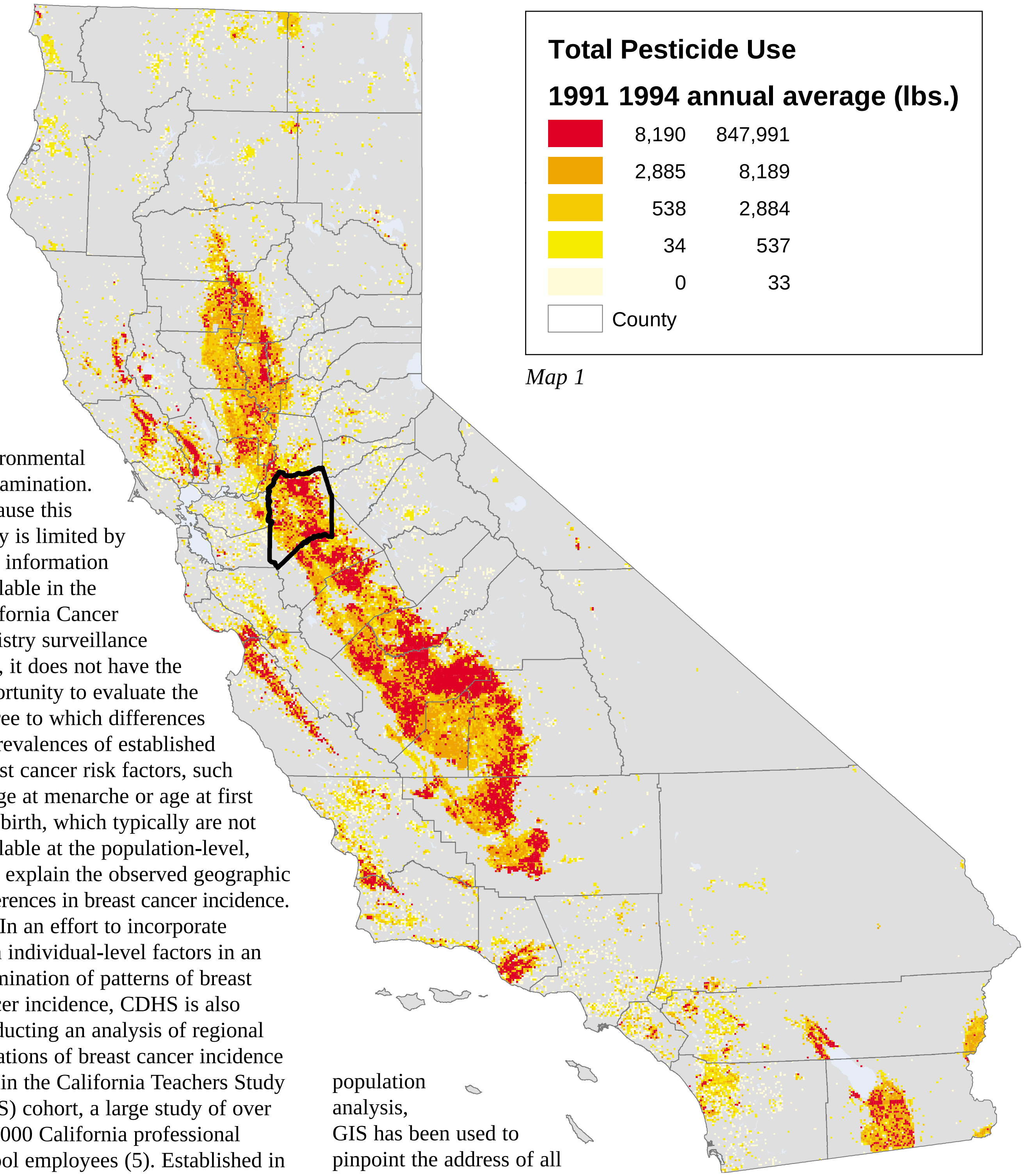


Figure 1

Designed to utilize the tools of GIS, researchers at the CDHS are currently investigating regional variations in breast cancer incidence at three different levels of analysis. They are designed to incorporate both large-scale, population-based patterns and individual level exposure potential into a comprehensive assessment of risk relationships and are illustrated in Figure 1. The largest of these studies is focused on investigating statewide patterns in breast cancer incidence using GIS to unite cancer surveillance data, demographic information from the census and data on potential environmental exposures from various statewide environmental datasets. For this study, more than 180,000 cases of invasive breast cancer have been identified in California during 1988 to 1997 and the address at diagnosis for all cases has been geocoded to a census blockgroup. GIS is being used to overlay sociodemographic and environmental characteristics and statistical models are being constructed to evaluate whether areas with high breast cancer incidence are associated with area measures of sociodemographic characteristics

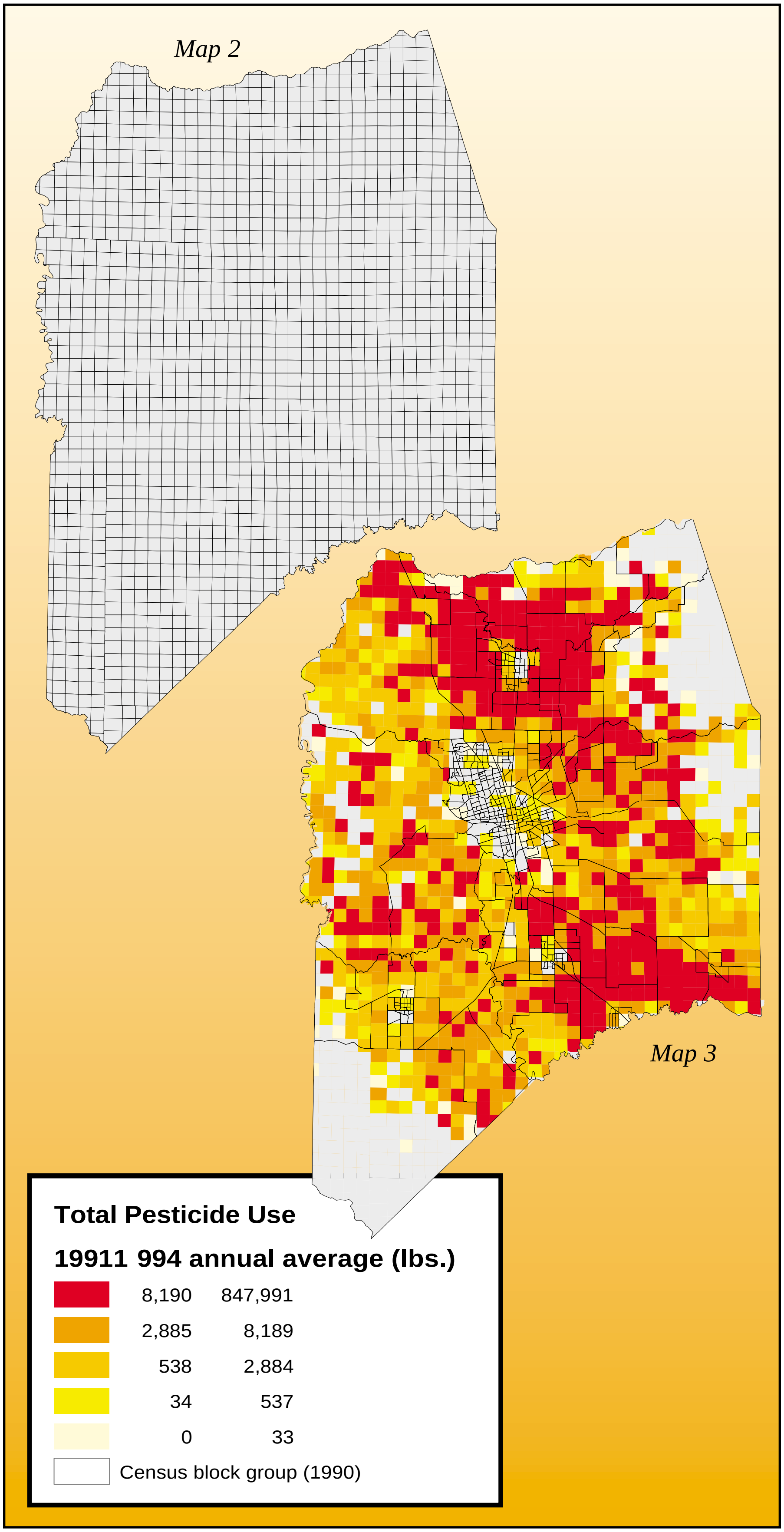


Map 1

and environmental contamination. Because this study is limited by case information available in the California Cancer Registry surveillance data, it does not have the opportunity to evaluate the degree to which differences in prevalences of established breast cancer risk factors, such as age at menarche or age at first live birth, which typically are not available at the population-level, may explain the observed geographic differences in breast cancer incidence.

In an effort to incorporate such individual-level factors in an examination of patterns of breast cancer incidence, CDHS is also conducting an analysis of regional variations of breast cancer incidence within the California Teachers Study (CTS) cohort, a large study of over 133,000 California professional school employees (5). Established in 1995, the CTS gathers extensive information on breast cancer risk factors and has followed participants for breast cancer incidence since 1995. Designed to build on the statewide population analysis, this study has identified over 1,500 cases of breast cancer occurring in the cohort in the first four years since its inception. Similar to the statewide population analysis, GIS has been used to pinpoint the address of all study members living in California at the time they joined the cohort and statewide census and environmental data are being used to characterize sociodemographic and potential environmental contamination. With regional patterns of breast cancer incidence similar to what is observed in the statewide population, the CTS provides a unique opportunity to incorporate both individual-level and population-level risk factor information in an examination of the factors underpinning the observed geographic patterns in breast cancer rates.

Finally, the last level of analysis and the smallest of the three studies,



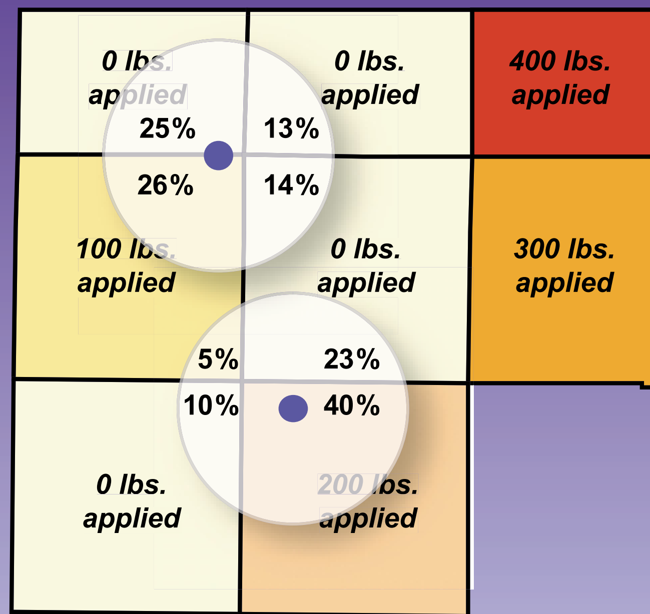
focuses on molecular markers of exposure as they might relate to potential urban-based sources of environmental exposures. Nested within the CTS cohort, this pilot study is designed to evaluate a number of questions regarding exposure assessment in these types of studies. Approximately 150 urban-based and 150 rural-based cohort members are enrolled in this study in which participants provided a 24-hour urine sample and answered a number of questions regarding diet, environmental exposures and residential history. Its objectives are to: (1) evaluate whether biomarkers of selected exposures of emerging interest in breast cancer etiology (e.g., traffic, pesticides) differ in urban versus rural women; (2) evaluate the degree to which these biomarkers correlate with GIS-derived measures of these exposures; (3) evaluate the degree to which self-reported exposures correlate with levels of urinary biomarkers; (4) evaluate the ability of CTS members to provide a life-time residential history; and (5) evaluate the impact that residential mobility may have on studies examining regional patterns of breast cancer that rely on address at diagnosis.

Initial analyses have focused on examining relationships with the pattern of factors influencing regional differences in breast cancer rates in California and on environ-mental data for agricultural pesticide use and air pollution. The question of whether breast cancer rates are higher in areas of agricultural pesticide use, as one example, has been of particular interest to many Californians. As the largest agribusiness state in the U.S., pesticide use is quite intensive in some areas but not uniformly so (see Map 1). California's Pesticide Use Reporting (PUR) system is quite unique, with detailed information on *all* agricultural pesticide applications by date of application, by method, by crop, by chemical and by amount. These, as illustrated in a sample map for San Joaquin County (Map 2), are reported at the level of approximately one square mile (based on the Public Land Service System designation of Township, Range and Section). This information provides us with an opportunity to evaluate pesticide use density by block group (see Map 3) for studies of population rates, or within a specified distance for individuals in the CTS cohort study (see Figure 2). Our environmental scientists have reviewed the more than 850 different chemicals reported by the PUR and grouped them for these analyses into agents thought to be probable or possible carcinogens, agents thought to be endocrine disruptors and agents with evidence of producing mammary tumors in laboratory studies. We prioritized the pesticides for study based on a variety of factors including cancer classification and potency, environmental fate and persistence, amount used, the distribution of potentially exposed populations and correlations among chemicals. Validation studies are also under way to assess exposure potential for people living near treated fields.

For these kinds of efforts, new and evolving GIS tools combined with traditional epidemiologic study designs offer the opportunity to better assess environmental influences for a number of disease outcomes. As this work is currently in development, it will not represent definitive evidence for environmental influences on health but it should provide us with some initial steps to better formulate our study questions and to better address the kinds of questions we get from the public.

When results from this study become available they will be posted on our website at www.ehib.org.

Figure 2



Colored squares represent Public Land Survey System (PLSS) township-range-sections (TRSs), with average annual pounds of pesticides applied in each TRS, for areas near case residences. White circles represent half-mile buffer area around cases (blue dots). Percentages represent percent of buffer area within each TRS.

Sample Pesticide Exposure Estimation:

- Estimated pesticides applied in top buffer area is 26 lbs./mi^2 (26% of 100 lbs.)
- Estimated pesticides applied in bottom buffer area is 85 lbs./mi^2 (5% of 100 lbs. + 40% of 200 lbs.)

1. Parkin DM, Muir CS, Whelan SL, et al, eds. Cancer Incidence in Five Continents. Vol VI. Lyon, France: World Health Organization, International Agency for Research on Cancer, 1992 (IARC Publication No. 120).
2. Perkins C, Morris C, Wright W, Young J. Cancer incidence and mortality in California by detailed race/ethnicity, 1988-1992. Sacramento, CA: California Department of Health Services, Cancer Surveillance Section, 1995.
3. Robbins AS, Brescianini S, Kelsey JF. Regional differences in known risk factors and the higher incidence of breast cancer in San Francisco. J Natl Cancer Inst 1997; 89: 960-5.
4. Clarke CA, Glaser SL, West DW et al. Breast cancer incidence and mortality trends in an affluent population: Marin County, California, USA, 1990-1999. Breast Cancer Research 2002; 4: R13.
5. Bernstein L, Allen M, Anton-Culver H et al. High breast cancer incidence rates among California teachers: results from the California Teachers Study (United States). Cancer Causes Control 2002; 13: 625-35.

A New Approach Using GIS and Spatial Analysis

Elizabeth L. Lewis-Michl, Center for Environmental Health
New York State Department of Health

New York State Department of Health's Cancer Mapping Project (officially the Cancer Surveillance Improvement Initiative) uses geographic information system tools and spatial statistical tools to provide information about the incidence of specific types of cancer for counties and ZIP codes. This project, which began in 1998, is an ongoing effort to give New Yorkers more information about cancer in their communities. It is designed to help guide cancer prevention programs and future research on the causes of cancer.

Information about cancer incidence in each of New York's 62 counties, including New York City's five boroughs, already was available in data tables provided annually. In 1999 and 2000, the Cancer Mapping Project presented the data for the first time in maps. County maps showed relative cancer incidence for 11 categories of cancer — lung and bronchus, prostate, breast, colon and rectum, kidney and renal pelvis, bladder, liver and bile duct, thyroid, leukemia, lymphoma, and brain and other nervous system.

In 2000 and 2001, the Cancer Mapping Project developed maps and data appendices of relative cancer incidence by ZIP Codes for the most frequently diagnosed types of cancer, breast, prostate, lung and colorectal. These maps and additional information about the New York State Cancer Registry and cancer risk factors are available at www.health.state.ny.us or by calling the New York State Department of Health at 1-800-458-1158.

What Information do the Maps Provide?

Both the county and the ZIP Code maps show geographic areas shaded with a range of colors to indicate relative cancer incidence in relation to statewide incidence. For the county maps, the relative shading is based on the **incidence rate** — the number of newly diagnosed cancer cases per population for five years. The county rates are adjusted to make them comparable even if there are differences in age of the populations.

For the ZIP Code maps, relative incidence is shown for each geographic area using a **standardized incidence ratio (SIR)**. This ratio indicates whether the number of new cancer diagnoses is higher, lower, or about the same as "expected" for a five-year period in each geographic unit. The SIR is calculated by dividing the actual observed number of cases by the expected number of cases in each county or ZIP Code. **Expected incidence** is the number of new cancer diagnoses that would be expected in that geographic area if the rate per population of cancer in this area were the same as in New York State as a whole.

Protecting the confidentiality of each person diagnosed with cancer limits how information can be presented in maps available to the public. Data grouped by ZIP Code are shown only for the most frequently diagnosed types of cancer (breast, prostate, lung and colorectal). However, confidentiality issues still occur when providing information at the ZIP Code level for some ZIP Codes with very small populations. For this reason, some ZIP Codes were combined to provide information for a larger number of cases.

Maps Create Visual Images that can be Misleading

In a map that is shaded according to relative cancer incidence, large geographic areas with small populations can contribute to a misleading visual impression. While the shading for high or low incidence for the large, rural areas can dominate the map visually, the incidence estimates for these sparsely populated areas are based on small numbers of cases. On the other hand, elevated or lowered cancer rates in large populations in densely populated areas may be almost invisible on the map if the areas are small in physical size, such as the boroughs of New York City.

In addition, chance variations in numbers in small populations with very small numbers of cases expected can create the appearance of large differences in comparative cancer incidence rates. For example, consider a small population where the expected number of cases of cancer to be diagnosed over a five-year period of time, is five. If six cases were diagnosed, the SIR would be 6/5, a 20% excess of cancer incidence. For comparison, in a larger population where 100 cases of cancer were expected, an additional case would result in a SIR of 101/100, a 1% excess.

One way to help prevent misinterpreting a map is to use additional information to estimate the accuracy, or margin of error, associated with each cancer incidence rate. In the county maps, the estimated cancer rate is accompanied by the actual number of cases upon which it is based. The county map pamphlets also show a bar graph of each county's relative cancer incidence rate.

On the State Health Department's web site there is a visual aid showing a margin of error, called a **confidence interval**, around each of the rates to indicate that the rate likely falls somewhere within this range. If the range includes the statewide rate, this indicates that the difference between this County's cancer incidence and the statewide cancer incidence is more likely due to chance.

Spatial Statistics Help with Interpretation of ZIP Code Maps

New York State researchers used a statistical technique to evaluate elevations of cancer incidence that would not be expected to appear just by chance in individual ZIP Codes and groups of ZIP Codes. This was done to help offset potentially misleading visual impressions and to help with interpretation of the ZIP Code maps. This technique uses a program called SaTScan, developed by Martin Kulldorff, Ph.D. and distributed by the National Cancer Institute.

A computer program randomly distributes the state's total number of cases of cancer among each of the state's ZIP Codes and possible groups of adjoining ZIP Codes (using only the age distribution and population size for each ZIP Code.) This random assignment is repeated 9,999 times, calculating SIRs for each ZIP Code and possible group of ZIP Codes in each of these simulations. For each simulated SIR, a value called a "likelihood statistic" is calculated. The likelihood statistic takes into account the number of observed and expected cases and the SIR. The results of the computer simulations are evaluated by identifying the most extremely unlikely elevation (highest value of the likelihood statistic) in each simulation.

The actual SIRs are compared with these simulated results to identify cancer incidence in any ZIP Code(s) that are extremely unlikely. ZIP Code(s) are considered to have a statistically significant elevation in cancer incidence if the likelihood statistic for that area is higher than 95% of the maximum likelihood statistics from the simulations. In this way, the actual cancer incidence estimates for ZIP Codes and groups of adjoining ZIP Codes are compared to the simulated values to decide what level of excess of cancer incidence is truly unusual, or not expected simply due to random or chance variation.

Cancer Booklets, including color versions of county-level cancer maps and graphs for 1992-1996, are available in portable document format (PDF) at http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii2.htm

Follow-up and Next Steps

The ZIP Code maps show hatching (diagonal lines) and cross hatching to identify areas of unusual cancer elevation not expected by chance. The State Health Department developed a protocol to investigate unusual disease patterns and is using the protocol to investigate a five ZIP Code area in Suffolk County where breast cancer incidence is estimated at about 50% above the average statewide incidence.

The first step in the Coram/Mt. Sinai/Port Jefferson Station investigation defined the geographic boundaries for follow-up. Next, researchers looked at some factors that might explain the higher rate of breast cancer, such as incorrect population estimates, or unusually high screening rates, detecting more cases of the disease in its early stages. These did not appear to be important factors for explaining the area's elevated incidence.

Researchers also began reviewing existing environmental data, examining information about the area's air, water and soil.

This past summer, area residents in the affected communities attended an information session to hear about the status of the investigation and other aspects of the State's Cancer Mapping Project. The State Health Department collected additional information from residents about historical sources of exposure that may have been a concern. Researchers are factoring this information into the ongoing investigation.

Currently State Health researchers are evaluating the literature on the biology of breast cancer and other factors that may contribute to disease incidence. They are reviewing the toxicology of environmental agents, especially those that are known or suspected to increase the risk of breast cancer. Researchers will recommend future activities after this review is completed and the environmental exposure information from the community is evaluated.

The follow-up work being done in the Coram/Mt. Sinai/Port Jefferson Station investigation has moved beyond the scope of tools such as geographical information system and spatial statistics. However, these kinds of tools can be valuable in helping to identify areas where further research into disease incidence may be needed.

Geographic Information Systems and Breast Cancer in Western New York

By Matthew R. Bonner, MPH

Environmental factors may play an important role in the etiology of breast cancer.

Here at the University at Buffalo's Center for Preventive Medicine, under the direction of Dr. Jo L. Freudenheim, we are looking at several environmental pollutants which may play a role in the development of breast cancer. These pollutants include benzene and polycyclic aromatic hydrocarbons (PAHs), which are formed from the combustion of organic material. The compounds are commonly released from tobacco smoke, automobile exhaust, and from the combustion of organic material. To facilitate these investigations, we are relying heavily on Geographic Information System (GIS) technology to reconstruct historical exposure to these environmental pollutants and to examine spatial and temporal clustering of breast cancer cases in western New York.

Our goal is to examine participants' lifetime exposure to these compounds and the subsequent risk of developing breast cancer focusing primarily on early life exposures. Early life exposures may be more important than recent exposure because breast epithelium may be more susceptible to carcinogenic insults during development. Evidence from atomic bomb survivors indicates that females who were less than 20 years old when they were exposed to ionizing radiation from the atomic bombing of Hiroshima and Nagasaki have a greater increase in risk of breast cancer than women exposed at older ages. Consequently, we are reconstructing historical exposures to these compounds in several ways and at several time periods when breast tissue may be most susceptible to environmental carcinogens.

First, we are using a recently completed case-control study where participants were asked to complete a residential history. These histories are being used to establish each participant's address at several time periods: 1) birth, 2) menarche, 3) time of their first pregnancy, 4) 20 years before interview, and 5) 10 years before interview. This has allowed us to geographically locate each subject's historical address in a GIS for each time period.

Second, we are using the GIS to estimate the proximity of each subject to known sources of benzene and PAHs. A database is being created containing addresses and other information on industries in Erie and Niagara counties including steel mills, chemical factories, foundries, and petroleum refineries dating back to the 1920s. These industries will be geographically located in a GIS with study participants' addresses to estimate proximity to these sites as a surrogate for exposure to benzene and PAHs generated by these facilities.

Our goal is to examine participants' lifetime exposure to these compounds and the subsequent risk of developing breast cancer focusing primarily on early life exposures.

Third, we are examining exposure to total suspended particulates, a surrogate measure of ambient air pollution. Total suspended particulates has been measured since the late 1950s in Erie County and we are using these measured concentrations at various locations throughout Erie and Niagara counties to estimate each participant's residential exposure to total suspended particulates. Again, GIS techniques have an integral role in the exposure assessment. Specifically, we are using GIS to interpolate the concentrations of total suspended particulates at each participant's residence for all five time periods based on the monitoring locations total suspended particulates levels.

Finally, we are examining exposure to PAHs from automobile exhaust. In this phase of the study, historical records of traffic counts will be used to model exposure to automobile exhaust at each participant's residence. These investigations will help us determine: 1) if these compounds are associated with breast cancer and 2) in what time period these exposures are most relevant. The analyses of these data are ongoing and we should have reportable results within the next year.

The Ribbon is published by the Cornell Program on Breast Cancer and Environmental Risk Factors in New York State. Comments are welcome; contact the Editor.

Editor
Carmi Orenstein, M.P.H., *Assistant Director*

Associate Editor
Carin Rundle, *Administrative/Outreach Coordinator*

Special Edition Design
West Hill Graphics, Ithaca, NY

Cornell University
Program on Breast Cancer and Environmental Risk Factors in New York State
112 Rice Hall, Cornell University
Ithaca, NY 14853-5601

Phone: (607) 254-2893
Fax: (607) 254-8207
Email: breastcancer@cornell.edu

Cornell Cooperative Extension