Geographic Information System for Breast Cancer Studies on Long Island (LI GIS)

Shannon Lynch, M.P.H.[1]; Linda Anderson, M.P.A.[1]; Todd Barr[2]; Linda Pickle, Ph.D.[1]; and Deborah Winn, Ph.D.[1]

[1]Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute; [2]L-3 Communications (Titan Corp.)

Abstract

The Geographic Information System for Breast Cancer Studies on Long Island (LI GIS) is an enterprise geographic information system combining data, ESRI ArcGIS, and statistical and spatial software and extensions. The research tool is designed to study and visualize the potential relationships between environmental exposures and breast cancer in Nassau and Suffolk Counties (Long Island), New York. It also is available to researchers for studying other diseases and will become available, on a modified level, to the general public. The LI GIS warehouse contains more than 80 data sets covering geographic, demographic, health outcome, and environmental data. Researchers can apply to use the entire LI GIS and/or the LI GIS statistical software and spatial extensions. In addition, four custom extensions developed for the LI GIS are available to download freely from the LI GIS Web site. Researchers with approved protocols can access the LI GIS remotely or work in its laboratory in Reston, VA. There usually is no fee to use the LI GIS, but funding for research is not provided. Maps from the LI GIS will be available soon for the general public at the LI GIS Web site. The LI GIS is funded by the National Cancer Institute (NCI), which is part of the National Institutes of Health (NIH), U.S. Department of Health and Human Services (HHS). It was developed and is maintained under a contract with Titan Corp. Learn more about the LI GIS and apply to use it at *www.healthgis-li.com.*

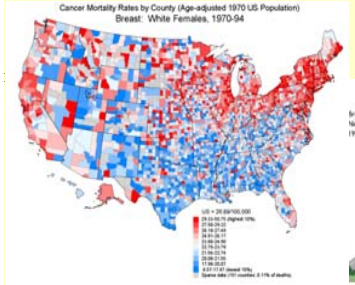
Introduction

Maps of the geographic distribution of breast cancer mortality for the United States have shown a pattern of elevated rates among white females in the Mid-Atlantic, Northeast, and North Central regions and some areas of the West. This pattern has persisted for more than four decades and also is seen in Long Island, NY. Geographic variations in the distribution of breast cancer mortality rates for black women were not as pronounced.

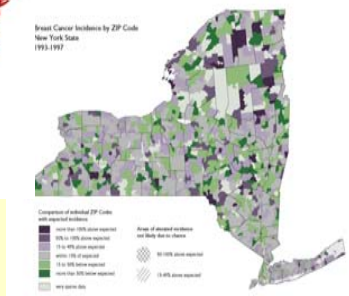
Map 1 (below) shows the geographic distribution of mortality rates for breast cancer for white females in the United States from 1970-1994 (Devesa SS, NCI, 1999). The dark red areas have the highest mortality rates, and the dark blue areas have the lowest mortality rates. See also NCI Cancer Mortality Maps & Graphs: *www3.cancer.gov/atlasplus.*

Map 2 is from the New York State Department of Health Cancer Surveillance Improvement Initiative (NYSDOH CSII). It shows whether the breast cancer incidence for each ZIP Code in New York State is higher (purple color), lower (green color), or about the same as expected (gray color) for 1992-1997. The map also shows areas where the breast cancer incidence is higher than expected compared to the state of New York as a whole, and the elevation likely is not to be the result of chance (represented by dashed patterns). See also: *www.health.state.ny.us/healthaz.*

Map 1.



Map 2.



In the early 1990s, the U.S. Congress mandated an investigation into the high rates of breast cancer in Nassau and Suffolk Counties (Long Island), New York, and the development of a geographic information system for use in exploring possible environmental causes of breast cancer. The Geographic Information System for Breast Cancer Studies on Long Island (LI GIS) is one of a series of activities undertaken as part of the Long Island Breast Cancer Study Project (LIBCSP). The LI GIS is funded by NCI and managed by the Epidemiology and Genetics Research Program and Surveillance Research Program, Division of Cancer Control and Population Sciences. Titan Corp. (now L-3 Communications) developed and maintains the LI GIS under a contract.

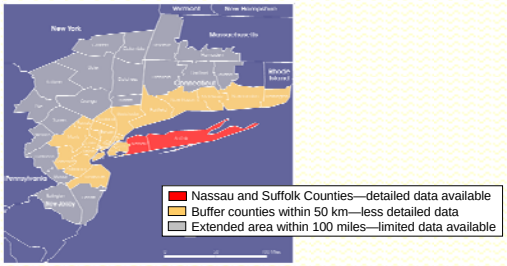
The LI GIS: A Research Tool

The LI GIS is a unique tool for both scientists and the general public. Researchers can use the LI GIS to study relationships between environmental exposures and risk for breast cancer, as well as to estimate exposure to environmental contamination. Access to the LI GIS is limited to researchers with approved protocols and is at no expense to the user, although special requests may entail costs to the researcher. Funding to support research is not provided. Researchers may apply to use the entire LI GIS and/or its statistical and spatial extensions. Also, four custom extensions are available freely and may be downloaded from the Web site.

A modified version of the LI GIS is under development to be used by the general public. Using information from the LI GIS, public maps are being developed to provide users with an insight into GISs and how they work using breast cancer and environmental data from Long Island, NY. The public maps and information on the LIBCSP, LI GIS, GISs in general, and cancer and the environment are accessible freely from the LI GIS Web site at *www.healthgis-li.com.*

LI GIS Overview

Geographic Extent of LI GIS



Data Warehouse

The ***LI GIS warehouse*** contains more than 80 data sets covering:

- Geographic attributes, including location of roads, water features, parks, and landmarks
- Demographic data, such as age, race, sex, and income of the population
- Health outcome data, including relative breast cancer incidence
- Environmental data for Nassau and Suffolk Counties, including land use; land cover; railroads; traffic; water use; potential sources of water pollution; releases of chemicals into water, air, and soil; information on toxic chemicals and hazardous and municipal waste; radiation; and, to a lesser extent, environmental data for surrounding counties.

Sources of the data include:

- State Health Departments
- U.S. Geological Survey
- U.S. Postal Service
- U.S. Census Bureau
- U.S. Environmental Protection Agency
- U.S. Department of Agriculture

The ***Metadata Browser*** provides more detailed information on the data sources and data sets and can be viewed on the public Web site: *www.healthgis-li.com.*

LI GIS Maps for the Public

Public maps from the LI GIS data are being developed to provide the public a window into how a GIS works. The public will be able to:

- Visualize and explore environmental exposure information and breast cancer incidence rates simultaneously
- Query the maps for more information on the location and type of environmental exposure
- Learn more about GISs in general and their uses in health research
  - Users also will be linked to Google Earth, where they can download and visualize data contained in the LI GIS using the Google GIS tool.
- Learn more about breast cancer and the environment.

Three Interactive Maps will become available on the public Web site:

- Breast Cancer Incidence Rates Map—provides breast cancer rates from 1993-1997 by ZIP Code for Long Island (data provided by NYSDOH CSII).
- Hazardous Waste Sites on Long Island Map—displays a variety of hazardous waste and materials sites on Long Island, in addition to breast cancer rates.
- Pesticide Detections on Long Island Map—displays pesticide detection data in Suffolk County drinking wells aggregated to the level of place (as defined by the U.S. Census Bureau), in addition to breast cancer rates.

LI GIS for Researchers

Possible Research Uses

The LI GIS enables researchers to:

- Explore and synthesize available information on potential exposures
- Generate hypotheses
- Identify spatial and temporal clusters of disease
- Evaluate risk factors for breast cancer and other health outcomes (with your addition of data)
- Address methodological issues
- Identify gaps in available information.

Researcher's Toolbox—*Sophisticated, Time-Saving Tools!*

The LI GIS has a full suite of GIS software and extensions related to the study of breast cancer:

- ESRI ArcGIS software suite
  - ArcView and ArcInfo
  - Spatial Analyst and 3D Analyst
- Extensions for epidemiological studies
  - Case File Formatter
  - Cluster Analysis Tool (using SaTScan)* (see Fig. 1, below)
  - Disease Rate Calculator* (see Fig. 2, below)
  - Areal Interpolator*
  - Empirical Bayes Tool*
  - Geographic masking
- EpiAnalyst
- S-Plus with Spatial Stats
- SAS
- Oracle 9i
- Additional ArcView extensions and software
- Online User's Guide

* These four custom extensions are available publicly and may be downloaded freely from the LI GIS Web site.

Download Free Custom Extensions from LI GIS Web Site

Four custom extensions have been developed for the LI GIS and are available freely from its public Web site.

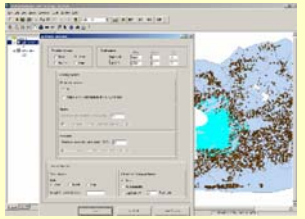


Figure 1. The Cluster Analysis Tool is an interface to the cluster analysis software application, SaTScan, developed by Martin Kulldorff, Ph.D., while at NCI. SaTScan Web site: *www.satscan.org.*

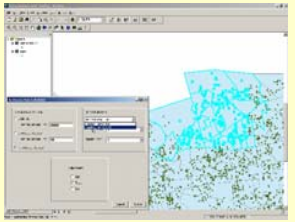


Figure 2. The Disease Rate Calculator calculates adjusted disease rates.

In addition to the extensions illustrated above, two other custom extensions are available:

- Empirical Bayes (EB) Tool applies the EB method, which is a statistical method useful for small areas and/or rare diseases. An EB rate is a weighted average of local rates and an overall rate, weighted by relative population sizes of the areas.
- Areal Interpolator, which uses simple areal interpolation to calculate the number of cases, controls, and the population for a given area, assuming they are distributed uniformly over the area.

Researchers may access the LI GIS remotely or work in its laboratory located in Reston, VA. To learn about the LI GIS and apply:
www.healthgis-li.com.

LI GIS Lab: 703-434-4386
E-mail: administrator@healthgis-li.com