

# Geographic-based Research and Applications at the National Cancer Institute

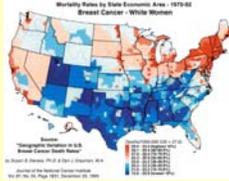
Linda Williams Pickle<sup>1</sup>, Ellen F. Heineman<sup>1</sup>, Mary H. Ward<sup>1</sup>, B. Sue Bell<sup>1</sup>, Dan J. Grauman<sup>1</sup>, Daniel B. Carr<sup>1</sup>, John R. Nuckols<sup>2</sup>, Marcia Gumpertz<sup>3</sup>

<sup>1</sup>National Cancer Institute, Bethesda, MD; <sup>2</sup>Department of Environmental Health, Colorado State University, Ft. Collins, CO; <sup>3</sup>Department of Statistics, North Carolina State University, Raleigh, NC

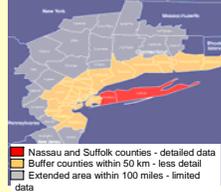
## I. GIS database development

Geographic Information System for Health (GIS-H)  
Long Island Breast Cancer Study Project ([www.healthgis-li.com](http://www.healthgis-li.com))  
A Resource for Researchers and the Public

High rates of breast cancer in the northeastern US extend over decades



Geographic Extent of GIS-H



Goals of the GIS-H for researchers:

- 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 researcher's addition of data)
- Address methodological issues
- Identify gaps in available information

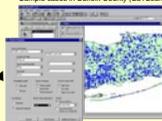
Disease rate calculator – calculating directly-Adjusted rate for selected census tracts



Researcher's toolbox

- ArcView, Spatial Analyst and 3D Analyst
- Extensions developed especially for GIS-H
  - Add Database Theme and Table Tools
  - Case File Formatter
  - Data query wizard
  - Disease Rate Calculator
  - Areal Interpolator
  - Cluster Analysis Tool (using SaTScan)
  - Empirical Bayes Tool
  - Geographic masking
- Online user's guide
- Additional ArcView extensions and software

Cluster analysis – checking for clusters of Sample cases in Suffolk County (SaTScan)



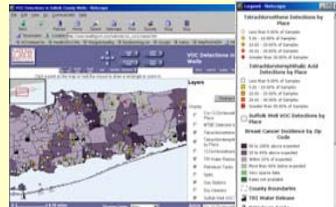
Areal interpolator – interpolating zip code Population from census tract population



Geographic masking – masking selected sample Cases (gold) using random-perturbation method



Example: interactive map of volatile organic compounds (VOC) in Suffolk County wells



Interactive maps for the public

- 16 interactive maps with up to 9 environmental exposure layers each
- Map topics and exposures reflect interests and concerns of community
- ArcExplorer allows creation of maps using publicly available data, plus interactive features and flexibility
- Offers reports and links to other sites

**Apply!** Access to researcher site is limited to researchers with approved protocols. (See Web site)

## II. Spatial data analysis

- A. Environmental exposure assessment
- B. Cluster identification
- C. Small area estimation
- D. Statistical modeling

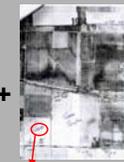
### A. Environmental exposure assessment

- GIS can provide information about potential environmental exposures that cannot be obtained through traditional epidemiologic methods
- Study in south central Nebraska demonstrated use of satellite imagery to reconstruct historical crop patterns  
*Ward et al. Env Health Perspectives, 2000*

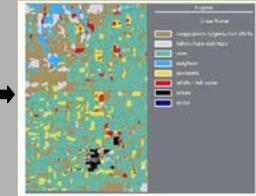
Landsat imagery (color infrared display of bands 4,2,1)



Farm Service Agency historical aerial photos with crops noted

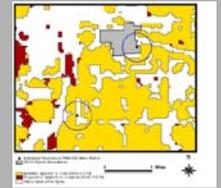


Classified land cover map



Example: epidemiologic study of non-Hodgkin's lymphoma (NHL)

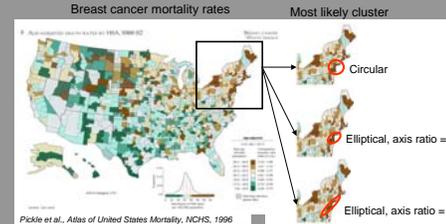
- Mapped residences, then assessed proximity of residences to specific crop
- Assigned probabilities of exposure based on available pesticide use data for each crop
- Demonstrated that zones of potential exposure to agricultural pesticides and proximity measures can be determined for residences



### B. Cluster identification

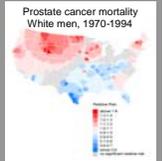
- Are apparent map clusters real or random noise?
- SaTScan software identifies most likely significant cluster over space, time or both
- Algorithm: spatial scan statistic for Poisson or Bernoulli event data, adjusts for population heterogeneity & covariates
- Originally identified circular clusters, new version scans for elliptical clusters, various shapes & angles
- Software: [srab.cancer.gov/satscan](http://srab.cancer.gov/satscan)

Developed by Martin Kulldorff: *Stat in Med, 1995, 1996; Communications in Statistics, 1997; Am J Epidemiology, 1997; Am J Public Health, 1998*



Nested clusters

- By simultaneously considering likelihood ratio and relative risk, Kulldorff's method can identify focused subclusters among broader regional patterns
- Application to prostate cancer data shows 72 distinct significantly high and low risk areas

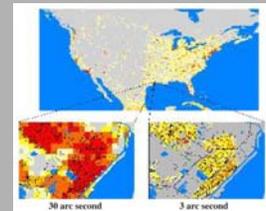


*Boscoe et al., Health and Place, in press*

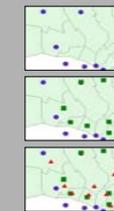
### C. Small area population centroids

- Are women who live farther away from a mammography facility more likely to be diagnosed with late-stage breast cancer in Los Angeles County, CA?
- Privacy: Only know census tract of residence for women diagnosed with breast cancer.
- LandScan USA provides a high-resolution population distribution model of residential (nighttime) & daytime population
- Oak Ridge National Laboratory ([www.ornl.gov/qist](http://www.ornl.gov/qist))  
Contact: *Budhendra Bhaduri*

LandScan USA 2000



Where do the women in LA live?



**In the water?? NO**

- Default tract centroids "in the water" because included 3-mile state coastal waters in the polygon.

**In the center of the tract?? Better**

- Calculated tract centroids using "dry land" polygons.

**Close to the beach?? Of Course!!**

- Tract centroids from LandScan USA<sup>1</sup> based on where people live.

<sup>1</sup>Centroids based on an interim release at a 15 arc second (~450 meter) resolution.

### D. Statistical modeling

- Cancer incidence prediction project goal is to project data from NCI cancer registries to all states
- Use hierarchical Poisson regression models to characterize associations between cancer incidence and mortality, sociodemographic, lifestyle factors
- These factors explain spatial variation well, so no spatial correlation is needed in the model
- Other modeling projects:
  - Time series analysis of cancer rates over time by state
  - Geographic patterns of % late stage breast cancer in Los Angeles county by census tract

Data used for cancer incidence prediction

NCI cancer registry data



CDC BRFSS lifestyle data



+ Other data by county

- Cancer mortality rate
- Medical facilities
- Census data
- Geographic Region

Smoothed predicted incidence rates

