International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 4– No.7, December 2012 – www.ijais.org

A Case Study of Geographic Distribution of Breast Cancer in New York State

Arwa Al Madan
Eastern Michigan University
109 Sill Hall
Ypsilanti MI, 48197

Mohammadjafar Esmaeili
Eastern Michigan University
109 Sill Hall
Ypsilanti MI, 48197

ABSTRACT

Breast cancer is one of the most serious and deadly diseases spread among women in the world. The American Cancer Society expected an estimated 211,240 new cases of breast cancer and 40,410 deaths (representing 15% of total deaths) to occur among women in 2005. Cancer is caused by both external factors (such as tobacco, chemicals, radiation and infectious organisms) and internal factors (such as inherited mutations, hormones, immune conditions and mutations that occur from metabolism). These causal factors may act together or in sequence to initiate or promote carcinogenesis. This study is utilizing Geographic information system (GIS) in public health and is aiming to investigate the geographic distribution of breast cancer in New York State in order to identify any possible high-risk areas. ArcGIS spatial statistics tools, Local Gi* statistics, Local Morans I, and Geoda Excess risk map were used in this Study. The success of this research may increase female awareness and help researchers to focus more on areas with high risk of breast cancer and to find any geographical or environmental factors that may cause cancer.

General Terms

Information System Application, public health.

Keywords

Geographic Information System (GIS), breast cancer, ArcGIS spatial statistics, Data Mining, Data Clustering

1. INTRODUCTION

Breast cancer is one of the deadliest diseases threatening women in United States of America. According to the American Cancer Society, 211,240 new breast cancer cases were expected to occur among women in 2005 and approximately 40,410 deaths representing a15% death rate in breast cancer new cases [1].

Moreover, the American Cancer Society in 2007 and 2008 estimates 178,480 new breast cancer cases among other kind of cancers representing 26% of total new cancer cases, as well as about 40,460 deaths from breast cancer representing 15% of the total deaths from cancer. Comparing the breast cancer death rate estimation in 2005, 2007, and 2008 reveals no change in death rate, which is 15% [2] and [3].

Moreover, according to Brody et al. (2007), breast cancer is the leading cause of death in US women between their late 30s and early-50s. A 45- year-old woman is 35% more likely to die of breast cancer than of ischemic heart disease and 2.5 times more likely to die of breast cancer than in a motor vehicle accident. The authors state that the research in recent years shows evidence of increased risk of breast cancer associated with exposure to polychlorinated byphenols or PCBs (banned chemicals previously used in electrical equipment and other products) in women genetically susceptible to the effects of these chemicals. Several recent studies have also found that breast cancer risk increases with exposure to polycyclic aromatic hydrocarbons (PAHs), ubiquitous air pollutants from vehicle exhaust and combustion. [4]

Davis et al,(1998) state that women with breast cancer include inherited genetic defects, menarche before age 12, menopause after age 55, having either no pregnancy or late childbearing, no lactation, early or repeated exposures to radiation, prolonged use of hormone replacement therapy, increased breast density, higher socioeconomic status, and postmenopausal obesity. Most of these risk factors can be linked with increased lifetime exposure to estrogen, other hormones, and higher exposures early in life. Most of these risk factors are not easily modifiable. Breast development depends on the complex interplay of estrogen, progesterone, and other growth factors. Therefore, the earlier in life that regular menses begins and the later that it ceases the greater exposure to hormones that affect breast growth in a woman's life (p. 523) [5].

According to Lewis-Michl et al (1996) most geographic variations in risk have not been explained by prior research, which suggests that environmental factors may be implicated (p.256) [6].

Han et al, (2004) conclude that there is speculation that environmental factors may explain geographic variation in breast cancer rates not explained by known risk factors. They were using residence as a proxy measure for environmental exposures in their study focused on Erie and Niagara counties in NY State. They investigated whether there was any evidence of geographic clustering of adult breast cancer cases associated with their residences in early life. Clustering analyses have often been used to provide clues for the unknown etiology of disease. They looked at the geographic clustering of residence at early critical time points: at birth, at menarche, and at the woman's first birth. By comparing differences in clustering patterns between case and control residences, they were interested in identifying time periods critical to potential environmental exposures and subsequent breast cancer risk [7].

For these reasons, it is very important to recognize the factors that impact breast cancer. Cancer can be caused by both external factors such as tobacco, chemicals, radiation and infectious organisms and internal factors such as inherited mutations, hormones, immune conditions and mutations that occur from metabolism . These causal factors may act together or in sequence to initiate or promote carcinogenesis. One methodology that has been utilized to identify the external factors is experimental research. This research attempts to facilitate the process of selecting external environmental factors by identifying the high risk areas in

New York State. In another words, this study will utilize GIS to identify the high risk area, which will help researchers to focus on the external factors in high breast cancer risk areas.

2. GEOGRAPHIC INFORMATION SYSTEM (GIS)

According to Padilla 2008, "GIS is a collection of computer hardware, software, and geographic data for capturing, managing, analyzing, and displaying all forms of geographically referenced information"(p.33)[8]. Also, Haque in 2001 provides another definition of GIS as following; "GIS is analytical and decision-making tools that organize, compare, and analyze disparate types of information into one organized system" (p.259)[9]. In another words, GIS can clarify connections among data or statistics that are not readable and represent them in a map. In general there are two types of data in GIS: spatial data which is related to location, and attribute data which is related to statistical information such as population and disease. GIS could be utilized to map the identified cases of breast cancer which will assist with recognizing the high risk areas [10] and [11].

3. RESEARCH PROBLEM

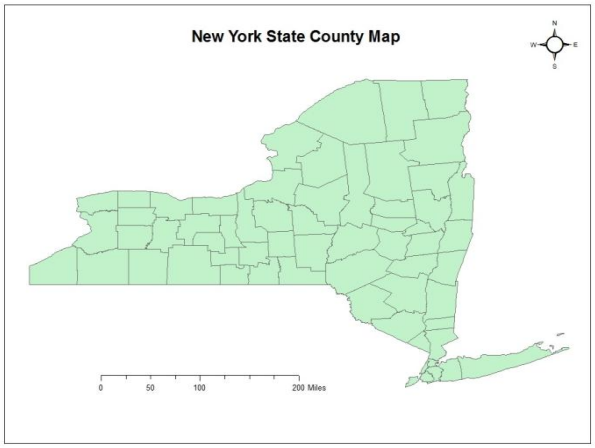
The area most susceptible to breast cancer in the state of New York during the period of 2001-2005 is not recognized. For this reason, this study attempts to apply GIS to a study of breast cancer in New York State and to point out the focus of geographic clustering of breast cancer. Moreover, this research attempts to answer the following research questions:

What are the high risk counties in New York State to breast cancer and where are they located?

Is there a huge gap of geographic distribution of breast cancer between annual average cases death rate and annual average death rate in NY State during 2001-2005?

4. NEW YORK CASE STUDY

The study focuses on New York State which is located in the northeastern US. New York State is the largest of the three Middle Atlantic States and ranks 30th in size among the 50 states. New York State has 62 counties as illustrated in Map 1.



Map 1: New York State County map

The data which is used for analysis in this study ranges from 2001 to 2005 and includes the average annual death rate and average annual cases rate. Analysis has been adopted primarily on average annual cases of breast cancer, rate of annual cases per 100,000, average annual deaths from breast cancer, and rate of annual deaths per 100,000. See Table 1 below. The data used in this study is from New York State Department of Health website.

Table 1: Attributes table of NY state ship file.

Attributes of cty036Copy

FID	Shape *	STATE	COUNTY	NAME	LSAD_TRANS	Avg_A_Case	Ra_100_000	Avg_A_Deat	DR_100_000	9
0	Polygon	36	089	St. Lawrence	County	77.6	125.8	16.2	24.7	
1	Polygon	36	019	Clinton	County	51.4	117.6	8	17.4	
2	Polygon	36	033	Franklin	County	35.4	130.3	6	20.5	
3	Polygon	36	031	Essex	County	26	106.9	4.2	15.2	
4	Polygon	36	045	Jefferson	County	65.2	118.3	14.4	24	
5	Polygon	36	049	Lewis	County	15.6	100.8	3.4	20	
6	Polygon	36	041	Hamilton	County	6.2	147	1	24.9	
7	Polygon	36	043	Herkimer	County	47.4	115.7	8	18.9	
8	Polygon	36	115	Washington	County	46.4	128.3	8.8	23.2	
9	Polygon	36	113	Warren	County	52.4	126.4	13.8	31.8	
10	Polygon	36	075	Oswego	County	83.4	128.8	14.8	22.6	
11	Polygon	36	065	Oneida	County	177.8	121.2	39.8	23.3	
12	Polygon	36	011	Cayuga	County	59.2	123.8	7.8	14.7	
13	Polygon	36	091	Saratoga	County	161.8	139.9	27.2	23.2	
14	Polygon	36	073	Orleans	County	30.2	126.4	5.6	23	
15	Polygon	36	063	Niagara	County	180.8	132.9	37	25.5	
16	Polygon	36	055	Monroe	County	588.4	138.4	108.4	23.9	
17	Polygon	36	117	Wayne	County	67.6	129.2	13.4	24.4	
18	Polygon	36	035	Fulton	County	40.2	119.3	11.4	30.7	
19	Polygon	36	067	Onondaga	County	354.8	132.6	73.8	26.1	
20	Polygon	36	053	Madison	County	50.2	129.6	11.8	29	
21	Polygon	36	037	Genesee	County	53.8	151.2	8	20.9	
22	Polygon	36	029	Erie	County	783.8	129.9	183.6	27.8	
23	Polygon	36	057	Montgomery	County	44.4	137.7	9.2	23.5	
24	Polygon	36	069	Ontario	County	84	138.7	14.6	23.2	
25	Polygon	36	099	Seneca	County	29.2	144	4.8	22.2	

Record: 0 Show: All Selected Records (0 out of 67 Selected) Options

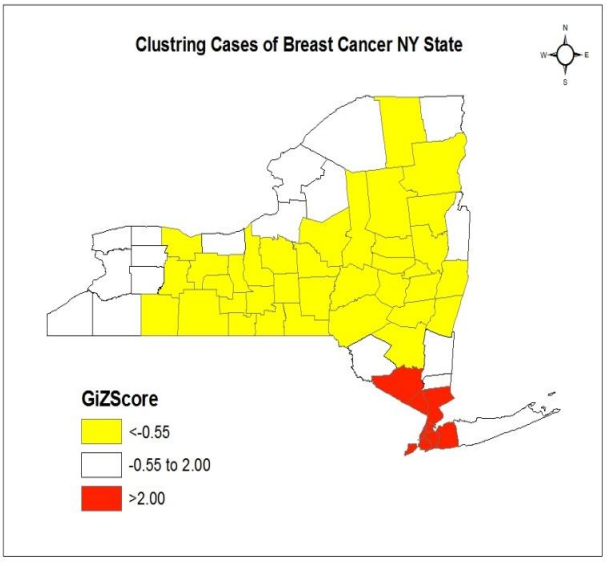
5. METHODOLOGY AND ANALYSIS

The study which generated this data was conducted at New York State from 2001 to 2005 to evaluate the breast cancer situation. The data was obtained from the New York State Department of Health (Available at www.health.state.ny.us).

The data were organized and associated with the map of NY state attribute table. ArcGis software 9.3 was used to analyze the average annual cases, average annual deaths and death rate. Geoda software was used to create an excess risk map.

There were two kinds of GIS maps utilized in this study: a Raw rate map (simple death rate map or Local Morans I) and a Hot spot map (local Gi* statistics). Also, Geoda software was used to create an excess risk map.

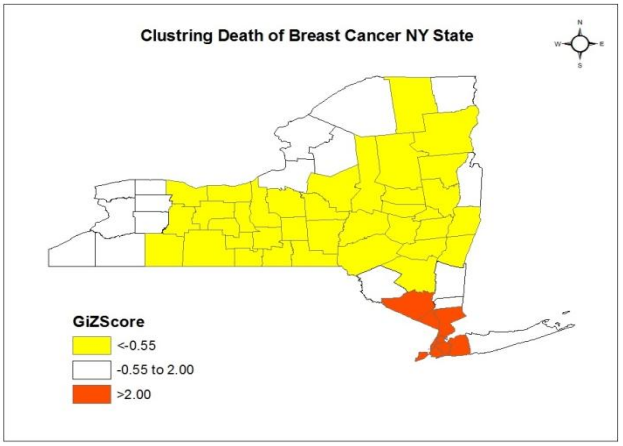
The raw rate or Crude Mortality rates map was created based on the death rates per 100,000. The death rate in NY State was concentrated on the state boundaries with a few counties in the middle. Most of the death concentrated on the east site of the state with range of 24.00 – 31.80. There were 26 counties with a high death rate (24.00 – 31.80), only two counties with zero mortality (New York City and Suffolk County), and the rest of the counties with a low death rate (0.01 – 24.00) (Map 2).



Map 2: Death Rate of breast cancer per 100,000 NY State 2001-2005

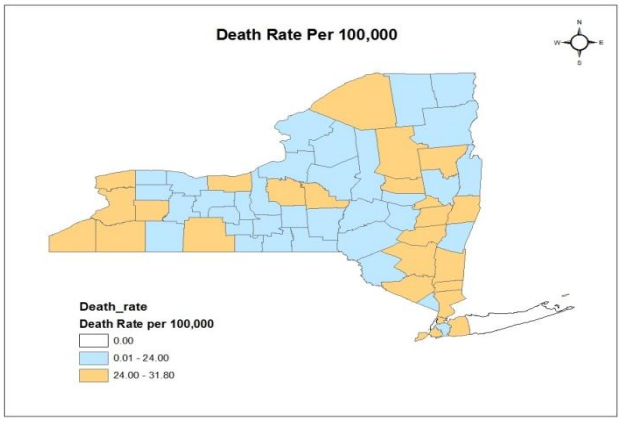
The annual average death of breast cancer in NY State was more than 200,000 female for the year 2001 – 2005. This high number of deaths was located at the southeast of the State (Map 3). Map 3 shows that there are eight counties that have a high value of breast cancer counts: Orange, Bronx, Queens, Nassau, Richmond, Kings, Rockland, and Westchester. 36 counties have a low significance value and the remaining counties have no significance value.

There were 35 counties with low significance value of annual cases excluding Wayne County (Map 4). This county appears in annual average death but not in annual average case Hotspot (Map 3 and 4).



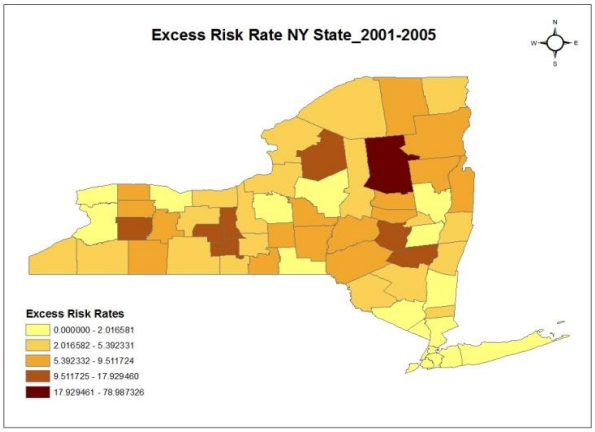
Map 3: Average Annual Death of breast cancer NY State 2001-2005

According to the Hotspot map (Map 4) concerning the annual average case there are 8 counties have high values of breast cancer counts: Orange, Bronx, Queens, Nassau, Richmond, Kings, Rockland, and Westchester. The annual death of the disease (Map 3) is the same as the annual cases except for Wayne County (Map 4).

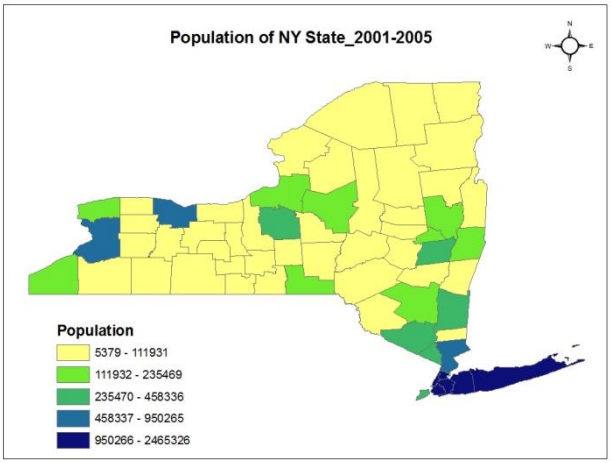


Map 4: Average Annual Case of breast cancer NY State 2001-2005

The Excess risk map shows that the Hamilton County has the highest risk death rate (24.9) with the lowest population density (5379) compared to the other counties (Map 5 and 6).



Map 5: Excess Risk Rate Map NY state 2001-2005



Map 6: population of New York State 2001-2005

6. DISCUSSION AND CONCLUSION

Breast cancer is the most common cancer among women in the United States. This study attempts to investigate the geographic distribution of breast cancer in New York State in order to answer the following research questions: how many counties in New York State are located in a high risk area for breast cancer? Is there a huge gap of geographic distribution of breast cancer between the annual average cases rate and the annual average death rate in NY State during 2001-2005? Moreover, the study attempts to identify possible high risk areas. ArcGIS spatial statistics tools, Local Gi* statistics, Local Morans I, and Geoda Excess risk map were used in this study.

Raw rate maps show that the clustering death and clustering causes of breast cancer are located in the southeast of NY State where the population is concentrated. This finding agrees with the finding of Brody et al (2007) [4]. In his report he mentioned that there are proportions of direct correlation between the disease and the population. Lewis-Michl et al. (1996) in a study of 21 New Jersey counties conducted in 1985 found a significant correlation between mortality rates for breast cancer and geographic density. [6]

The results show that the distribution of breast cancer in NY State is located on the state boundaries with a few counties in the middle. The interactions between NY State and the boundary states may be considered as one factor increasing the disease incidence. On the other hand, Prehn et al, (1998) have shown that geographic variation in breast cancer rates in the San Francisco Bay Area may be explained by geographic variation in risk factors, though they did not directly address environmental causes of breast cancer. [11]

Moreover, the analysis of the data demonstrates that Hamilton County has the highest risk death rate among the New York counties as shown in the Excess Risk Rate Map. The Hot spot maps show that 35 counties have low significance values excluding Wayne County which appears with no significance value counties in annual average death. Moreover, since this research points out the areas highly susceptible to breast cancer in New York State, future research needs to be done to identify the possible environmental factors that lead to these high risk areas.

7. REFERENCES

[1] American Cancer Society, Cancer Fact and Figures (2005). Retrieved April 24,2009, http://www.cancer.org/downloads/STT/CAFF2005f4PWSecured.pdf.

[2] American Cancer Society, Cancer Fact and Figures (2007). Retrieved April 24, 2009, http://www.cancer.org/downloads/STT/CAFF2007PWSecured.pdf.

[3] American Cancer Society ,*Cancer Fact and Figures* (2008). Retrieved April 24, 2009, http://www.cancer.org/downloads/STT/2008CAFFfinalsecured.pdf.

[4] Brody, Julia, Ruthann Rudel, Karin Michels, Kirsten Moysich et al.. "Environmental Pollutants, Diet, Physical Activity, Body Size, and Breast Cancer." American Cancer Society 109.12 (2007): 2627-34.

[5] Davis, Devra, Deborah Axelrod, Lisa Bailey, Mitchell Gaynor et al.. "Rethinking Breast Cancer Risk and the Environment: The Case for the Precautionary Principle." Environmental Health Perspectives 106.9 (1998): 523-29.

[6] Lewis-Michl, Elizabeth L., et al. "Breast Cancer Risk and Residence Near Industry or Traffic in Nassau and Suffolk Counties, Long Island, New York." Archives of Environmental Healthg, 51.4 (1996): 255-265.

[7] Han, Daikwon, Peter Rogerson, Jing Nie, Matthew Bonner et al.. "Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer (United States)." Cancer Causes & Control 15.9 (2004): 921-29.

[8] Padilla, C. (2008). Historical GIS, Mapping the Past to Understand The Future. ABI/ INFORM Global, *32*(2), 32-35.

[9] Haque, A. (2001). GIS, Public Service, and the Issue of Democratic Governance Public Administration Review, *61*(3), 259-265.

[10] Olendorf, D., jeryan, C., & Boyden, K. (1999). The Gale Encyclopedia of Medicine (Vol.2, , pp. 567-569). Farmington Hills, MI: Gale Research.

[11] Prehn, Angela, and Dee West. "Evaluating Local Differences in Breast Cancer Incidence Rates: A Census-Based Methodology (United States)." Cancer Causes & Control 9.5 (1998): 511-17