Spatial Patterns of Breast Cancer Incidence and Uninsured Women of Mammography Screening Age

Linda Highfield, PhD, MS*

*St. Luke's Episcopal Health Charities, Houston, Texas; †University of Texas, School of Public Health, Division of Management, Policy and Community Health Practice, College Station, Texas

Abstract: Breast cancer is the most common cancer and second leading cause of cancer mortality in women in the United States. Women who lack insurance have mammography screening rates that are suboptimal. Our objective was to spatially correlate incidence rates of breast cancer and uninsured women aged 40–64 years and identify outliers—areas where women may be underscreened due to poor access. The eight-county consolidated metropolitan statistical area centered on Harris County, Texas was selected as the study region. Breast cancer incidence data from 1995 to 2004 were acquired from the State of Texas Cancer Registry as individual case data geocoded at the census tract level. A bivariate local indicator of spatial autocorrelation was used to evaluate the spatial pattern of breast cancer incidence and uninsured. Statistically significant negative spatial autocorrelation was observed between breast cancer incidence and uninsured status in women aged 40–64 (Moran’s I = -0.2065, p < 0.001), indicating that as breast cancer incidence increased, uninsured rates decreased globally. Statistically significant local clusters of low breast cancer incidence and high incidence of uninsured were found. Future research is needed to assess mammography screening behaviors and barriers to screening at the local level.

Key Words: breast cancer, Moran’s I, spatial analysis, uninsured

Approximately one in eight American women will develop breast cancer during her lifetime (1,2). Breast cancer is the most common cancer, and second leading cause of cancer mortality in women in the United States (3,4). Breast cancer incidence rates historically have been the highest in Caucasian women—largely attributed to higher screening rates (5). Breast cancer incidence has also been associated with socioeconomic status, with increasing incidence in higher socioeconomic groups (6). However, women of lower socioeconomic status, regardless of race, tend to access mammography screening at lower rates and have lower adherence to recommended follow-up appointments after abnormal screening results (7). Lack of insurance has been shown to be an important factor in accessing mammography screening and follow-up care (8,9). Women who lack insurance have screening rates that are suboptimal compared with women with insurance (10–12). Rates of uninsured adults in Texas are the highest in the U.S., averaging 25% across the state (13). The current total estimate of uninsured in Texas is slightly more than 1.5 times higher than the national average (13). While lack of insurance can be linked empirically to lack of mammography screening and follow-up care, an exploratory spatial data analysis (ESDA) of the patterns of breast cancer incidence and uninsured women of screening age at a subcounty resolution in Texas has not been conducted. ESDA is an important tool to help visualize and reveal patterns, clusters and outliers in geographically referenced data. Applying ESDA techniques to these data aids our understanding of the patterns of cancer in populations and highlights areas where further research is needed and outreach programs can be targeted.

COMPARATIVE STATISTICAL TOOLS

Several methods exist for ESDA. A classical test for measuring spatial autocorrelation is the Moran’s I statistic (14). The global Moran’s I statistic provides a measure of overall spatial autocorrelation (15). Spatial autocorrelation is a measure of the degree of dependency of data in space (similar to a Pearson correlation
coefficient) (14–16). The local indicator of spatial association (LISA) is the localized equivalent of the global Moran’s I (15,16). For each location on a map, the LISA statistic evaluates and statistically tests the similarity of the data at the given location (e.g., incidence at the source census tract) with the values of its local neighbors (surrounding census tracts). A bivariate LISA (BiLISA) provides a measure of the degree of linear association (positive or negative) between the value for variable \( y \) (e.g., cancer incidence) at a given location (census tract \( i \)) and the average of variable \( x \) (e.g., uninsured) at neighboring locations (15). Similar to the LISA statistic, the BiLISA yields spatial clusters (positive autocorrelation) falling into two categories (High-High and Low-Low) and two classes of outliers (negative autocorrelation) (High-Low and Low-High). Inference for all Moran’s statistics is based on permutation testing (e.g., Monte Carlo methods), where a reference distribution is calculated for spatial randomness and compared with the observed data over multiple iterations.

The objective with this research was to locate areas in the greater Houston region with low incidence rates of breast cancer and high rates of uninsured women of screening age and to intentionally identify outliers—areas where women may be underscreened due to poor access as a result of lack of health insurance. Using ESDA to identify these areas is a first step in addressing potential lack of access. ESDA can be a tool for targeting areas at the highest need for a limited resource.

**METHODS**

**Study Location and Data Source**

The eight-county consolidated metropolitan statistical area (CMSA) centered on Harris County, Texas, was selected as the study region. The CMSA is home to a large, diverse population and a large number of uninsured women, making it the ideal location to study breast health disparities. The CMSA is comprised of Brazoria, Liberty, Harris, Montgomery, Waller, Chambers, Fort Bend and Galveston Counties (see Fig. 1).

The CMSA population at the time of the study was approximately 5.5 million people (13), of which an estimated 27% of women under age 65 were uninsured (13). Breast cancer incidence data from 1995 to 2004 were acquired from the State of Texas Cancer Registry. The number of uninsured women of screening age, 40–64 years, in the study area was estimated using data from the 2004 to 2006 Current Population Survey, Annual Social and Economic Supplement (CPS ASEC) (13). Census tract-level small area estimates were developed as described in the data analysis. Data on the population of women of screening age (aged 40–64 years) were taken from the 2000 Census for the CMSA. Due to concerns with confidentiality, the cancer incidence data was de-identified case-level data, pregeocoded to a census tract level geography.

**Data Analysis**

**Estimation of Proportion of Uninsured Women Aged 40–64**

A population-based synthetic estimation model was used to estimate the numbers and proportions of uninsured for census tracts in the Houston CMSA. The model was population based because the variables used to calculate the proportions and numbers of uninsured were demographic (i.e., race/ethnicity, age, and sex). It was synthetic because it is a method that makes use of information available at higher levels of geography and applies them, without change, to lower levels (17). In order to have sufficiently large sample sizes, 3 years of CPS ASEC data was aggregated prior to calculating rates of uninsured. Data used for this article were from the 2004 to 2006 surveys. Because the survey asked about insurance status in the previous year, the data reflected self-reported conditions for 2003 to 2005.

Current Population Survey, Annual Social and Economic Supplement national public use files (PUF) were downloaded from the Bureau of Labor Statistics and compiled in SPSS 17.0. Household and family data set elements were combined with person-level data to construct an individual case file. The data files contained a variety of Federal Information Processing Standards (FIPS) geographic identifiers, which were used to reduce the amount of data to only those cases used in the estimation. Cases coded for Texas (FIPS Code 48, where “gestfips” = 48) were selected. Race/ethnicity, age group, and health insurance status recoding took place in this file. After the initial recodes and aggregations were completed, the revised file was tested against state results published by the Census Bureau. The state-level results matched, and the Houston metropolitan area (FIPS Code 26420, where “gtcbsa” = 26420) sample cases were identified and isolated for analysis.
Rates of uninsured for all selected population groups were developed after applying person-level population weights. In the CPS ASEC instrument, health insurance status was determined through a series of questions that asked about specific types of coverage—e.g., private, work-related, Medicare, Medicaid, and others. For the estimation data set, the responses to these questions were collapsed and recoded into a single variable. Any single positive response to one of the coverage questions was recoded as “insured.” The population groups for which rates were developed were comprised all possible combinations ($n = 32$) of the following:

1 Race/Ethnicity: White non-Hispanic, Hispanic, Black non-Hispanic, and Asian/other race (non-Hispanic).

2 Age: 0–17, 18–39, 40–64, and 65 plus.

3 Gender: Male and Female.

Rates of uninsured were calculated for the listed categories and applied to population numbers, found in tabulated Census Bureau data, for equivalent groups at the census tract level of geography. The source for the population data was the 2000 Census Summary File 1 (SF1). Uninsured proportions for women aged 40–64 years were extracted from the working data file for ESDA.

**Calculation of Breast Cancer Incidence Rates**

Breast cancer incidence rate was calculated as the cumulative number of breast cancer cases in females aged 40–64 years divided by the population of females aged 40–64 years as reported by Census 2000 at the census tract level.
tract level. The resulting number for each census tract across the 1995–2004 time period was divided by 10 to obtain an average annual incidence rate. The raw incidence rate was compared with smoothed rates prior to ESDA to evaluate variance instability in GeoDa (15).

**Bivariate Local Moran’s I Statistic** The Formula for the Bivariate Local Moran’s I Statistic is given Below:

$$I_{kl} = z_k^i \sum_j w_{ij} z_l^j$$

Where $k$ and $l$ are breast cancer incidence rate and uninsured proportion for census tract $i$ and the neighboring tract $j$, respectively. $Z_k$ and $Z_l$ are the standardized $z$-scores of variables $k$ and $l$, respectively. The standardized $Z$-score for each variable is computed as the observed rate (e.g., cancer incidence) at location $i$ minus the mean rate for the neighbors $j$ (e.g., average cancer incidence for all neighbors) divided by the standard deviation. $W_{ij}$ is the spatial weights matrix which is defined as a binary contiguity matrix. This provides the spatial structure for the locations included in the calculation of the Local Moran’s I statistic. All observations that share a common border have weights $= 1$, otherwise the weights $= 0$. Using queen first-order contiguity neighbors with a common border or vertex are considered $= 1$. Results from the BiLISA statistic are based on permutation. A randomization of 999 permutations was used prior to interpretation of the results.

**Descriptive Statistics** Race-stratified breast cancer incidence rates per 100,000 for women aged 40–64 years were calculated for both the CMSA and the State of Texas. Descriptive statistics on census tracts identified as clusters of low breast cancer incidence and high uninsured were calculated based on location within Super Neighborhood including: average proportion of uninsured women aged 40–64 years, average household income, average educational attainment and average proportion Caucasian, African American, Hispanic and Asian. Super Neighborhoods are designations created by the City of Houston to encourage communities to work together in identifying, prioritizing, and addressing community needs (18). The boundaries of super neighborhoods rely on major physical features (e.g., bayous, freeways) to create contiguous communities that share physical characteristics, identity and/or infrastructure (18). Super Neighborhood boundaries were used as a way to compare the demographics of these Census tracts to the City of Houston. These designations are commonly used in the City of Houston and are the basis for many community-based projects focused on health disparities in the study area.

**RESULTS**

Average annual incidence rates per 100,000 women aged 40–64 years were lower for all races in the CMSA as compared with Texas, except in Caucasian and African American women (296 versus 276 and 238 versus 233, respectively) (Table 1). The spatial distribution of average annual breast cancer incidence rates in women 40–64 years per 100,000 by Census tract for the CMSA are shown in Figure 1. The highest proportions of uninsured women in the study area were found in Harris County (Fig. 2) with most of those falling within Houston city limits.

Significant negative spatial autocorrelation was observed between breast cancer incidence (Fig. 1) and uninsured status in women (Fig. 2) aged 40–64 years across the study area (Moran’s $I$ $= -0.2065$, $p < 0.001$), indicating that as breast cancer incidence increased, the average neighboring uninsured rates decreased globally. Locally, several statistically significant clusters of low breast cancer incidence and high uninsured status were found using the BiLISA, within Harris County (Fig. 3). A total of 79 individual Census tracts were within the low breast cancer incidence and high uninsured clusters. The average annual breast cancer incidence rate for women aged 40–64 years in these tracts was 121 per 100,000. The average proportion of uninsured women aged 40–64 years within these tracts was 35%.

Seventy-one of the 79 low-high Census tracts fell within 26 of the 88 City of Houston Super Neighbor-

| Race                | Rate per 100,000 women aged 40–64 | Rate per 100,000 women aged 40–64 |
|--------------------|----------------------------------|----------------------------------|
| Caucasian          | 296.51                           | 276.72                           |
| African American   | 238.85                           | 233.9                            |
| Hispanic           | 162.36                           | 174.6                            |
| Asian              | 121.59                           | 132.58                           |

CMSA, consolidated metropolitan statistical area.
hoods (Fig. 4). The highest average population proportion within these Super Neighborhoods was Hispanic (55%) (Table 2). The average median income was $30,056 (in 1999 dollars) (Table 2). The data used for income in this analysis were from the Census 2000 and was not inflation adjusted, meaning it is not directly comparable with income estimates from other years. The average median income for Census tracts within the Super Neighborhoods fell at or below 135% of the federal poverty line (based on a family of four) (Table 2). The average proportion with a High School diploma within the Super Neighborhoods was 13% (Table 2). The Super Neighborhoods in the low-high clusters had 22% lower average median income as compared with the City of Houston (Table 2). The Super Neighborhoods had 50% lower proportions of Caucasian residents than the City of Houston and had 48% higher proportions of Hispanic residents (Table 2).

**DISCUSSION**

Exploratory spatial data analysis revealed a statistically significant negative association between breast cancer incidence and uninsured status. As breast cancer incidence increased, the proportion of uninsured women decreased overall. Dissecting this association with a local geographical level revealed several areas of “low-high” clusters. This study is exploratory in nature; it assesses correlation, not causation. It is intended to be a first step in assessing the relationship between breast cancer and potential health disparities, such as
lack of health insurance. A geographically based measure of spatial correlation (BiLISA) was used; however, other methods exist for assessing spatial correlation. The rationale for selecting BiLISA was based on the areal nature of the cancer data. The BiLISA method can account for areal data in the assessment of spatial autocorrelation using a contiguity matrix for assessing neighboring values, whereas many other available autocorrelation methods work off of pseudo-spatial approaches, such as distance (e.g., using centroids) (19,20).

The majority (71/79) of the low-high cluster Census tracts fell within City of Houston Super Neighborhoods. On the average, income in these areas was lower than the City ($30,056 versus $36,616) and was home to a much higher proportion of Hispanics (48%). Hispanics were overrepresented in the low-high cluster Census tracts and this finding is consistent with the pattern of uninsured in the greater Houston region. Uninsured rates exceeding 50% have been observed in the Hispanic population in other studies (13). Not all of the Census tracts aligned with these boundaries, nor were many of these neighborhoods completely covered by Census tracts. In other words, this is an approximate representation of the population demographics for these areas.

Breast cancer incidence was used as a proxy for mammography screening behaviors in this study. The rationale for using incidence as a proxy for screening was the hypothesis that there may be a relationship between low incidence and high uninsured areas indicating lower screening compliance. Previous research has confirmed poor screening behaviors in uninsured women (8–12). Based on the ESDA, it cannot be definitively stated that the observed low incidence and high uninsured clusters were a direct result of low mammography screening compliance. Additional data on mammography screening behaviors are needed to further elucidate this relationship in the local population. However, previous research in this population (on mammography screening) suggested low screening compliance among this population (L. Highfield, unpublished data). In addition, those unpublished data show availability of free and low-cost screening to be far below the actual need in the population (L. Highfield, unpublished data). Previously published studies conducted with the target population have also indicated poor screening compliance in uninsured women (21).

An average annual breast cancer incidence rate and a snapshot uninsured proportion was used in this study. Estimates of the uninsured were only recently available and the data are available only for certain time periods. The data used in this study were the most recent available. The rationale for selecting an average annual breast cancer incidence rate was based on the time period of analysis, sample size, and available data for the denominator. Due to a small sample size for incidence in any given year, it was necessary to aggregate cases over time to generate disease rates. A limitation of aggregating the incidence data is that hormone replacement therapy (HRT) behaviors changed in 2001, when women were no longer recommended to take HRT. The cessation of HRT led to decreased breast cancer incidence starting in 2003. The assumption is that this change did not differentially affect the population by subgroup (e.g., race) or geographically. While this is not likely to be the case, for the purposes of looking at “on the average” incidence rates globally in this population, this was not likely to cause a major issue with the ESDA findings.

Denominator selection is a difficult issue when dealing with public health data and obtaining an accurate denominator over time can be challenging. Census 2000 population numbers for women aged 40–64 years was used as the denominator and it was assumed that this represented the “centered” population estimate for the study time period. For the objective of ESDA between breast cancer incidence...
and uninsured women, using a centered population estimate and a snapshot evaluation of the spatial pattern of breast cancer incidence and uninsured was warranted. The primary aim was to explore the “on the average” geographical pattern in the study area rather than looking retrospectively for any particular space–time association. The focus on women aged 40–64 years was based on their need for regular mammography screening and because women 65 and over qualify for Medicare coverage.

**CONCLUSIONS**

To the author’s knowledge, this is the first study to look at the subcounty geographical pattern of breast cancer incidence and uninsured status in women of mammography screening age in the Houston CMSA. Future studies are needed to further evaluate specific mammography behaviors in these identified subcounty areas. Furthermore, surveys of women in these local areas could aid insight into mammography screening behaviors and barriers to screening.

**Acknowledgments**

The author would like to acknowledge St. Luke’s Episcopal Health Charities for funding this research. The author would also like to thank Ning Zhang and Tom Reynolds for their contribution to the manuscript. Marlynn May and Karen Williams provided internal review.
Table 2. Sociodemographics of Low-Incidence High Uninsured Clusters Falling within Houston Super Neighborhoods

| Super neighborhood name | Total Pop* | Prop African American | Prop Asian | Prop Caucasian | Prop Hispanic | Average median household income* | Prop high school graduates* |
|-------------------------|------------|----------------------|------------|----------------|---------------|---------------------------------|--------------------------|
| Alief                   | 97,889     | 0.28                 | 0.20       | 0.17           | 0.31          | $37,047                        | 0.13                     |
| Braeburn                | 33,809     | 0.36                 | 0.04       | 0.25           | 0.33          | $36,030                        | 0.13                     |
| East Little York/Homestead | 22,140   | 0.83                 | 0.02       | 0.02           | 0.14          | $28,495                        | 0.21                     |
| Eastex/Jensen area      | 28,196     | 0.22                 | 0.04       | 0.10           | 0.67          | $25,236                        | 0.13                     |
| Golfcrest/Villaggio     | 26,054     | 0.21                 | 0.09       | 0.11           | 0.67          | $30,893                        | 0.14                     |
| Greater Fondren S.W.    | 49,436     | 0.53                 | 0.06       | 0.15           | 0.25          | $36,122                        | 0.12                     |
| Greater Greenspoint     | 40,671     | 0.34                 | 0.02       | 0.10           | 0.53          | $27,240                        | 0.14                     |
| Greater heights         | 41,486     | 0.04                 | 0.00       | 0.42           | 0.53          | $41,576                        | 0.12                     |
| Greater hobby area      | 41,198     | 0.25                 | 0.06       | 0.17           | 0.51          | $32,601                        | 0.14                     |
| Gulftown/Pine Valley    | 12,905     | 0.03                 | 0.01       | 0.09           | 0.87          | $29,430                        | 0.09                     |
| Guiltton                | 46,369     | 0.09                 | 0.05       | 0.11           | 0.74          | $25,056                        | 0.08                     |
| Harrisburg/Manchester   | 3,768      | 0.06                 | 0.05       | 0.11           | 0.74          | $26,898                        | 0.12                     |
| Lawndale/Wayside        | 14,132     | 0.01                 | 0.04       | 0.12           | 0.86          | $29,537                        | 0.15                     |
| MacGregor               | 13,997     | 0.80                 | 0.04       | 0.09           | 0.67          | $39,615                        | 0.14                     |
| Magnolia Park           | 21,302     | 0.006                | 0.02       | 0.03           | 0.96          | $14,875                        | 0.07                     |
| Meadowbrook/Allendale   | 22,929     | 0.03                 | 0.09       | 0.21           | 0.75          | $34,823                        | 0.14                     |
| Midtown                 | 5,311      | 0.18                 | 0.06       | 0.46           | 0.28          | $40,383                        | 0.10                     |
| NorthShore              | 27,350     | 0.17                 | 0.08       | 0.22           | 0.59          | $33,899                        | 0.15                     |
| Northside Northline     | 54,676     | 0.07                 | 0.06       | 0.18           | 0.74          | $27,773                        | 0.13                     |
| Park Place              | 9,902      | 0.03                 | 0.03       | 0.11           | 0.74          | $28,956                        | 0.11                     |
| Pecan Park              | 19,230     | 0.03                 | 0.01       | 0.05           | 0.90          | $27,214                        | 0.10                     |
| Second Ward             | 14,836     | 0.08                 | 0.00       | 0.05           | 0.86          | $23,473                        | 0.09                     |
| Sharpstown              | 77,085     | 0.17                 | 0.16       | 0.20           | 0.46          | $31,377                        | 0.12                     |
| South Park              | 22,282     | 0.81                 | 0.03       | 0.014          | 0.16          | $27,060                        | 0.20                     |
| Sunny side              | 18,629     | 0.93                 | 0.01       | 0.01           | 0.04          | $20,383                        | 0.19                     |
| Willowbrook             | 2,741      | 0.19                 | 0.07       | 0.56           | 0.15          | $32,366                        | 0.17                     |
| Super neighborhood average | n/a    | 0.26                 | 0.03       | 0.15           | 0.55          | $30,056                        | 0.13                     |
| City of Houston average | n/a       | 0.25                 | 0.05       | 0.31           | 0.37          | $36,616                        | 0.13                     |

*Data based on 2000 Census. Available at http://www.houstontx.gov/planning/suprnbhds/recognized_sn.htm.

COMPETING INTERESTS

The author declares that there are no competing interests.

REFERENCES

1. American Cancer Society. Cancer Facts & Figures 2008. Atlanta: American Cancer Society, 2008.
2. Ries LAG, Harkins D, Krapcho M, et al., eds. SEER Cancer Statistics Review, 1975–2003. Bethesda, MD: National Cancer Institute, 2006.
3. Schueler KM, Chu Philip W, Smith-Bindman R. Factors associated with mammography utilization: a systematic quantitative review of the literature. J Womens Health 2008;17:1477–98.
4. Williams KP, Sheppard VB, Todem D, Mabiso A, Wulu JT, Hines RD. Family matters in mammography screening among African American women. Cancer Causes Control 2001;12:327–36.
5. National Cancer Institute. Cancer Health Disparities (2010). Available at: http://www.cancer.gov/cancertopics/factsheet/cancer-health-disparities/disparities (accessed December 15, 2010).
6. Kobayashi KA, Chang SM, Tyczewski L, et al. Socioeconomic risk factors for breast cancer: distinguishing individual- and community-level effects. Epidemiology 2004;15:442–50.
7. Yabroff KR, Washington KS, Leader A, Nelesen E, Mandelblatt J. Is the promise of cancer-screening programs being compromised? Quality of follow-up care after abnormal screening results. Med Care Res Rev 2003;60:294–313.
8. Christie RE, Benard VB, Blackman D, et al. Breast cancer screening among low-income or uninsured women: results from the National Breast and Cervical Cancer Early Detection Program, July 1995–March 2002. Cancer Cause Control 2006;17:29–38.
9. Swan J, Breen N, Coates RJ, Rimer BK, Lee NC. Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. Cancer 2003;97:1528–40.
10. Mickey RM, Durski J, Worden JK, Danigelis NL. Breast cancer screening and associated factors for low-income African-American women. Prev Med 1995;24:467–76.
11. Adams KE, Florence CS, Thorpe KE, Becker ER, Joski Pj. Preventive care: female cancer screening, 1996–2000. Am J Prev Med 2003;25:301–7.
12. Barrett K, Legg J. Demographic and health factors associated with mammography utilization. Am J Health Promot 2005;19:401–5.
13. U.S. Census Bureau, Current Population Survey, 2009 Annual Social and Economic Supplement. – Table HI06, Coverage Status by State for All People: 2008.
14. Moran P. Notes on continuous stochastic phenomena. Biometrika 1950;37:17–23.
15. Anselin L, Syabri I, Smirnov O Visualizing multivariate spatial correlation with dynamically linked windows. In: Anselin L, Rey S, eds. New Tools for Spatial Data Analysis: Proceedings of the Specialist Meeting. Santa Barbara: CD-ROM: Center for Spatially Integrated Social Science (CSISS), University of California, 2002.
16. Anselin L. Computing environments for spatial data analysis. J Geogr Sys 2000;2:201–20.
17. Cohen SB. Small area estimation, AHCPR Pub. No. 98-R073. Reprinted from, Encyclopedia of Biostatistics, vol. 5, Armitge P, Colton T, eds. New York: John Wiley, 1998.
18. City of Houston. Super Neighborhood Guidelines 2010. Available at: http://www.houstontx.gov/planning/suprnhd/guidelines.htm (accessed December 15, 2010).
19. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43:671–81.

20. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. Appl Stat 1991;40:283–94.
21. Peranteau J, Highfield L, Parekh J, et al. Fort Bend County Breast Health Assessment. Prepared on behalf of the Susan G. Komen Foundation, 2010. Available at: www.slehc.org (accessed July 5, 2011).