

Geographic Disparities in Late-Stage Breast Cancer Diagnosis Rates and Their Persistence Over Time

Lee R. Mobley, PhD,¹ Florence K.L. Tangka, PhD,² Zahava Berkowitz, MSPH, MSc,²
Jacqueline Miller, MD, MPH,² Ingrid J. Hall, PhD,² Manxia Wu, MD, MPH,² and Susan A. Sabatino, MD, MPH²

Abstract

Background: Other than skin cancer, breast cancer is the most common cancer in the United States. Lower uptake of mammography screening is associated with higher rates of late-stage breast cancers. This study aims to show geographic patterns in the United States, where rates of late-stage breast cancer are high and persistent over time, and examines factors associated with these patterns.

Materials and Methods: We examined all primary breast cancers diagnosed among all counties in 43 U.S. states with available data. We used spatial cluster analysis to identify hot spots (*i.e.*, spatial clusters with above average late-stage diagnosis rates among counties). Demographic and socioeconomic characteristics were compared between persistent hot spots and those counties that were never hot spots.

Results: Of the 2,599 counties examined in 43 states, 219 were identified as persistent hot spots. Counties with persistent hot spots (compared with counties that were never hot spots) were located in more deprived areas with worse housing characteristics, lower socioeconomic status, lower levels of health insurance, worse access to mammography, more isolated American Indian/Alaska Native, Black, or Hispanic neighborhoods, and larger income disparity. In addition, persistent hot spots were significantly more likely to be observed among poor, rural, African American, or Hispanic communities, but not among poor, rural, White communities. This analysis includes a broader range of socioeconomic conditions than those included in previous literature.

Conclusion: We found geographic disparities in late-stage breast cancer diagnosis rates, with some communities experiencing persistent disparities over time. Our findings can guide public health efforts aimed at reducing disparities in stage of diagnosis for breast cancer.

Keywords: late-stage cancer diagnosis, geographic disparities, urban–rural disparities, spatial clusters

Introduction

OTHER THAN SKIN CANCER, breast cancer is the most common cancer in the United States. The 5-year relative survival rate among women with breast cancer has trended upward over time among all population groups.¹ Improvements in breast cancer survival reflect advances in treatment, as well as in early diagnosis of cancers as a result of public health initiatives for screening.

Breast cancer screening reduces breast cancer mortality by increasing detection of cancer at earlier stages when it is easier to treat.² Lower uptake of mammography screening is associated with higher rates of late-stage breast cancers and higher morbidity and mortality rates.³ Women diagnosed with breast cancer at regional stage have a 5-year survival rate of 82.3%. However, the 5-year survival rate among

women diagnosed with distant disease is only 24.5%.⁴ In a national study covering most breast cancer cases from 2004 to 2009, 31% of new cases were diagnosed at a late stage, defined as regional or distant disease, where the overall burdens of higher morbidity and mortality rates are worse.⁵ Rates of late-stage diagnoses for breast cancer in the United States vary across states,^{5,6} and those states with higher percentages (34%–36%) of late-stage diagnoses during 2004–2009 were Texas, Arkansas, Oklahoma, Louisiana, Mississippi, Alabama, and Utah.⁵

Understanding features of the geographic locations where rates of advanced stage diagnoses are high and persistent over time can help identify areas where increased breast cancer control efforts might be targeted. Previous geographic studies, which have ranked states in terms of late-stage breast cancer diagnosis rates, examined the geospatial predictors of

¹Mobley Consulting, LLC, Atlanta, Georgia, USA.

²Division of Cancer Prevention and Control, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

late-stage rates in multilevel models, and provided data about places where the predicted rates are higher (in the upper quantile) across the United States.^{5,7} We focus on late-stage (regional and distant stage combined) diagnosis rates and identify U.S. counties where rates of late-stage breast cancers are high and persistent over time during two periods (2004–2009 and 2010–2014). We used actual, not predicted, rates to define these areas across most of the United States. This research is unique in its methodology and comprehensive in its geographic scope given available data on official cancer statistics from the U.S. government. The results are descriptive, not confirmatory, and suggest regions where further study may be warranted to explain underlying causal mechanisms.

Materials and Methods

Data sources and measures

We examined breast cancer data from the United States Cancer Statistics (USCS) database during two intervals (2004–2009 and 2010–2014). The USCS database is a population-based surveillance system of cancer registries providing the official government statistics on cancer.⁸ The USCS database includes information on demographics (age, sex, race, and ethnicity), tumor characteristics, and geographic location (U.S. county) of residence at the time of diagnosis. We received access to county geocodes approved to researchers (for only 43 states) who perform their analyses at a National Center for Health Statistics Research Data Center.⁹ Seven states (Kansas, Missouri, Illinois, Michigan, Minnesota, Hawaii, and Alaska) and Washington, DC were excluded from the analysis because county-level data were not available.

All women who had a first primary diagnosis of breast cancer were included in the analysis. About 5% of cases with unknown cancer stage and <1% that were identified by death certificate and autopsy only were excluded. After case exclusions, the study included 1,073,595 cases during 2004–2009 and 957,348 cases during 2010–2014. These breast cancer cases were diagnosed in 2,599 counties within the 43 states.

The outcome variable of interest was the rate of late-stage breast cancer cases diagnosed in each county of the 43 states included in the study. County was the smallest geographic unit permissible to study with these restricted USCS data. Rate of late-stage diagnoses was defined by combining cases classified as regional or distant. The USCS database used the Collaborative Stage Data Collection System to derive the Surveillance, Epidemiology, and End Results (SEER) Summary Stage 2000 for diagnosis years 2004 and later,¹⁰ which we also used to assess our outcome. The late-stage breast cancer rates examined were the percentage of regional stage and distant stage disease combined out of all breast cancer cases in each county.

We added multiple, county-level, contextual (county-related socio-ecological conditions) factors derived from various sources to the USCS data and described them in Table 1. These factors were used to assess whether counties identified as high-rate clusters (hot spots) that persisted over time were different from counties not identified as persistent hot spots. Two variables reflect access to breast health care: the number of registered mammography facilities per capita

adult women and the percentage of the adult population aged 40–64 years without health insurance. We also included population density as a robust measure of the location of counties in relation to the urban–rural continuum.

The diversity index was included to reflect the extent of segregation into cultural enclaves within census tracts. This measure, as originally defined by Massey and Denton,¹¹ varies from 0 (perfectly even: all subunits/tracts have the same racial and ethnic residential composition as the larger area) to 1 (perfectly uneven: all subunits/tracts have only one population group). In addition, we included isolation indices, defined for each race or ethnicity according to Massey and Denton.¹¹ These measures reflect the likelihood that members of a specified race or ethnicity come into contact with other members of the same race or ethnicity, and range from 0 to 1. A “zero” would indicate no likelihood of contact with others of one’s own race or ethnicity, whereas a “one” would indicate highly likely contact, based on residential race or ethnic patterns (not actual contact information). This variable is continuous with the likelihood ranging from 0 to 1. We also examined housing characteristics and county typology codes describing different types of living and economic environments. Additional examined characteristics included area poverty, persistent poverty, and measures for uninsured and unemployed adults. Persistent poverty identifies counties that have remained poor for two decades and reflects a more pervasive measure than a simple poverty rate at a single point in time.

Analytic methods

For the analysis, we combined regional and distant stage to identify U.S. counties where rates of late-stage breast cancers are high and persistent over time, and examined factors associated with these patterns. We first used the Global Moran’s Index (Moran’s I) test to identify clusters of counties with higher than average rates of late-stage breast cancer diagnoses among areas of the 43 states. Positive spatial autocorrelation shows that closer places are more similar to one another than to places farther away. A positive Moran’s I test indicates that there is some clustering somewhere in the data.

A Moran’s I statistic indicating clustering is followed by using GeoDa software (version 1.12) to do an assessment of local spatial autocorrelation patterns with Local Indicators of Spatial Association (LISA) tests. These tests, described by Anselin,¹² were used to identify hot spots where the observed rate in each local area was higher than the expected rate. The LISA test statistics were calculated for each county and period to identify the locations of clusters of counties with either higher or lower late-stage breast cancer rates than the average of all the counties included in the analysis. We performed the cluster analyses by using a *p*-value of <0.01 and used a “queen” contiguity matrix to describe all adjacent counties with boundaries or vertices coinciding with the county in question.

GeoDa’s LISA test results identified the counties located at the center of each multicounty cluster only by using mapping and internally created categorical variables. The GeoDa map results traditionally show the high-rate (hot spot) and low-rate (cool spot) cluster centers (Maps in Fig. 1a, b). However, the full spatial clusters include both the center and the adjacent counties, defined by the Queen

TABLE 1. DEFINITIONS AND SOURCES OF ALL MEASURES AND FACTORS USED IN ANALYSES OF GEOGRAPHIC DISPARITIES IN LATE-STAGE BREAST CANCER DIAGNOSIS RATES

<i>Variable</i>	<i>Date of data</i>	<i>Source</i>
Proportion regional: proportion of BC cases diagnosed at regional stage	2004–2014	Developed by L.R.M. from the United States Cancer Statistics Database, https://www.cdc.gov/rdc/index.htm
Proportion distant: proportion of BC cases diagnosed at distant stage		
Proportion AI/AN: proportion of BC cases from AI/AN		
Proportion API: proportion of BC cases from API		
Proportion Black: proportion of BC cases from non-Hispanic African Americans		
Proportion Hispanic: proportion of BC cases from Hispanic Americans		
Proportion other: proportion of BC cases from all other races or ethnicities		
Proportion White: proportion of BC cases from non-Hispanic White Americans		
Diversity Index: Theil Index of residential diversity among all U.S. population	2000, 5-year aggregate 2007–2011	Developed by L.R.M. from the United States Decennial Census and American Community Survey, https://www.census.gov
% Graduate/professional: percent of U.S. population with an advanced degree		
% No high school or equivalent: percent of U.S. population with low education		
(GINI) index of income disparity: calculated by U.S. Census on U.S. population		
Isolation index American Indian: based on isolation index defined in literature		
Isolation index Asian: based on isolation index defined in literature		
Isolation Index African American: based on isolation index defined in literature		
Isolation index Hispanic: based on isolation index defined in literature		
Isolation index Pacific Islander: based on isolation index defined in literature		
Isolation index non-Hispanic White: based on isolation index defined in literature		
Proportion poor or no English language: proportion who do not speak English well		
% Unemployed: percent of U.S. population unemployed	2006, 2012	Developed by L.R.M. from SAHIE Program, https://www.census.gov/programs-surveys/sahie.html
% Uninsured aged 40–64: percent of U.S. population 40–64 years of age no health insurance		
% Uninsured: percent of U.S. population with no health insurance		

(continued)

TABLE 1. (CONTINUED)

<i>Variable</i>	<i>Date of data</i>	<i>Source</i>
Mammography facilities per 1,000: number of licensed mammography units per 1,000 women 25–79 years of age	2000–2005; 2005–2010	Developed from proprietary United States Food and Drug Administration data by Dr. Jan Eberth, https://sc.edu/study/colleges_schools/public_health/faculty-staff/eberth_jan.php
Proportion managed care penetration, calculated as the number of enrollees/number of eligible	2005, 2011	Developed by L.R.M. from Centers for Medicare and Medicaid Services, Geographic Service Area files, https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/HealthPlanRepFileData/Geo.html
Persistent poverty: proportion of counties that have been persistently poor for two decades	2000, 5-year aggregate 2007–2011	Developed by L.R.M. from United States Decennial Census and American Community Survey, https://www.census.gov
% Poor Black rural: percent of U.S. population who are poor, Black, and living rural		
% Poor Black urban: percent of U.S. population who are poor, Black, and living urban		
% Poor Hispanic rural: percent of U.S. population who are poor, Hispanic, and rural		
% Poor Hispanic urban: percent of U.S. population who are poor, Hispanic, and urban		
% Poor White rural: percent of U.S. population who are poor, non-Hispanic White, and rural		
% Poor White urban: percent of U.S. population who are poor, non-Hispanic White, urban		
Proportion population in poverty: percent of U.S. population who are poor		
Proportion vacant housing: percent of U.S. housing stock that is vacant		
Proportion heat with wood fire: percent of U.S. housing stock that is heated by wood		
Proportion living crowded: percent of U.S. population who live in crowded quarters		
Proportion no plumbing: percent of U.S. housing stock that has no indoor plumbing		
Emphasis farming: proportion counties with main economic emphasis being farming	2004, 2014	Developed by L.R.M. from United States Department of Agriculture Economic Research Service data County Typology Codes, https://www.ers.usda.gov/data-products/
Emphasis low education: proportion counties with high proportions of low educated people		
Emphasis manufacturing: proportion counties with main economic emphasis being manufacturing		
Emphasis mining: proportion counties with main economic emphasis being mining		
Emphasis population loss: proportion counties that consistently lose population to outmigration		

BC, breast cancer; AI, American Indians; AN, Alaskan Natives; API, Asians/Pacific Islander; SAHIE, Small Area Health Insurance Estimates.

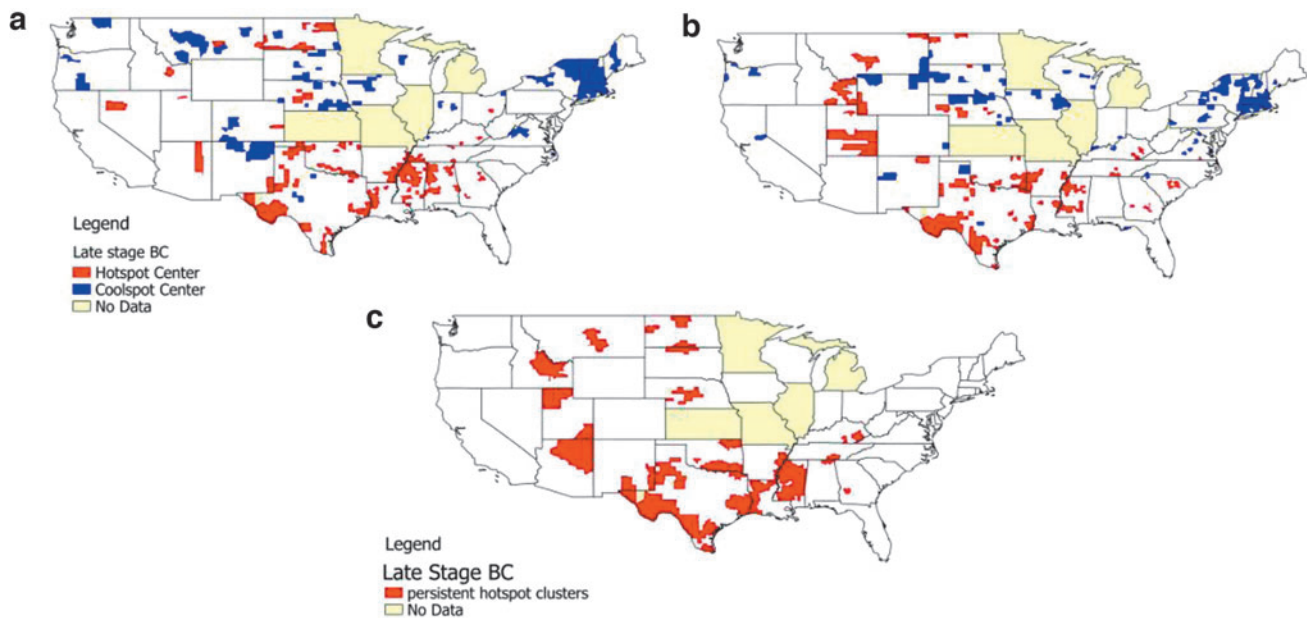


FIG. 1. Geographic disparities in clustering of late-stage.^a Breast cancer diagnosis rates and their persistence over time. (a) Late-stage cluster centers 2004–2009. (b) Late-stage cluster centers 2010–2014. (c) Persistent late-stage clusters. ^aLate stage, defined as regional or distant. Regional cluster centers—cluster of patients whose breast cancer involve only regional lymph nodes. Distant cluster centers—cluster of patients whose breast cancer have spread to distant parts of the body. Cluster centers (a, b) are centers of whole clusters (c), which contain center and contiguous neighbors. Hot spots are spatial clusters with above average late-stage diagnosis rates among counties. Cool spots are spatial clusters with below average late-stage diagnosis rates among counties.

contiguity-based weights matrix as spatial neighbors.¹² To capture the full cluster, we used the Queen weights (consistent with other studies)¹³ to create an indicator variable for every neighboring county around the cluster center. Although only cluster centers are given in Figure 1a and b, the full cluster is given in Figure 1c.

To examine temporal trends in the spatial clustering of the outcome rates, we compared the full set of counties identified as the local clusters during 2004–2009 with those during 2010–2014. Counties that were part of a hot spot during both periods were classified as being persistently high. Similarly, counties with no spatial cluster during both spans were classified as persistently nonclustered.

Persistent hot spots were compared with areas where there were no hot spots during either period (always vs. never a hot spot) by using the county-level contextual factors. To assess differences among local areas, we compared the means of the proportions of the contextual factors' characteristics among counties with persistent hot spots with those that were never hot spots. The contextual factors in these comparisons were performed only with data from the period 2010–2014 (e.g., using the later period Diversity Index rather than earlier period Index). Results using the earlier period for contextual factors were qualitatively the same. For each factor, differences in the proportions or percentages across the two area types for late-stage breast cancer were assessed with the simple *t*-test of two sample means assuming unequal variances.¹⁴ The *t*-statistics themselves are a standardized difference.¹⁵ Thus, the larger test statistics reflect greater standardized changes more so than smaller results.

Results

Hot spot maps

The Global Moran's *I* statistics were all statistically significant, indicating that there were pockets of local spatial autocorrelation (clustering) over the geographic area for late-stage breast cancer outcome and during both periods ($p < 0.01$). In other words, the significant Moran's *I* test indicates that there was spatial correlation among some neighboring counties but does not indicate where these counties are located. Over the geographic area, there was no spatial dependence (i.e., null hypothesis), which we rejected with a significance level of $< 1\%$. LISA statistics that were performed found many hot spots and cool spots in the outcome of late-stage breast cancer. Maps (Fig. 1a, b) for each period show abundant hot spots. The centers of the spatial clusters in these maps are depicted with counties colored red (hot spot center) and blue (cool spot center). Persistent hot spots (i.e., counties that were hot spot members during both periods) and their spatial neighbors (i.e., the entire cluster), are given in Figure 1c.

Following these results, we focused on the hot spots. We examined differences between “never” hot spots and “always” hot spot clusters by comparing means across contextual variables. We used a standard *t*-test statistic. The classification of clustered counties as “never” or “always” was based on whether they were present in the two intervals (2004–2009 and 2010–2014).

The difference between the mean values was calculated by subtracting the mean of the “always” group from the mean of the “never” group and performing a two-sided test for equal

means. A negative *t*-test statistic indicates that a “never clustered” counties mean was lower than an “always clustered” counties mean. A positive test statistic indicates that a “never clustered” counties mean was higher than an “always clustered” counties mean. No statistically significant difference between means of contextual variables is indicated with a shade of gray.

Results for late-stage breast cancer are given in Table 2. Counties that were persistent hot spots had higher concentrations of American Indian and Hispanic women and lower concentrations of non-Hispanic white women in the breast cancer sample than counties that were never hot spots.

Compared with counties that were never hot spots, counties in persistent hot spots had higher concentrations of adults with no high school or equivalent education and lower concentrations of people with graduate or professional degrees. These counties also had higher isolation indices for African Americans and Hispanics and lower isolation indices for non-Hispanic Whites, Asians, and Pacific Islanders. In addition, persistent hot spot counties had higher income disparity, higher concentrations of poor English language ability, and higher percentages of uninsured individuals. In addition, mean mammography facilities per capita and penetration of managed care insurance were lower in persistent hot spots.

TABLE 2. *T*-TESTS^a OF EQUAL MEANS OF COUNTY DESCRIPTIVE VARIABLES OVER TWO GROUPS OF COUNTIES THAT REFLECT DIFFERENT AREA TYPES AS REGARDS HOT SPOTS FOR LATE-STAGE BREAST CANCER

	<i>Group 1: Never a hot spot (N = 1,800)</i>		<i>Group 2: Always a hot spot (N = 269)</i>		<i>t-test</i>
	<i>Mean</i>	<i>SD</i>	<i>mean</i>	<i>SD</i>	
County descriptive variables					
Proportion American Indian	0.007	0.034	0.034	0.101	-4.429
Proportion Asian	0.008	0.020	0.003	0.007	7.031
Proportion non-Hispanic Black American	0.072	0.120	0.087	0.126	-1.851
Proportion Hispanic	0.033	0.084	0.104	0.197	-5.858
Proportion other	0.006	0.013	0.002	0.006	8.741
Proportion non-Hispanic White	0.875	0.150	0.769	0.204	8.182
Diversity index	0.121	0.077	0.106	0.096	2.376
% Graduate/professional	21.437	9.576	16.453	6.468	10.969
% No high school or equivalent	14.202	5.958	20.677	8.175	-12.503
GINI index of income disparity	0.436	0.033	0.457	0.038	-8.805
Isolation index American Indian	0.031	0.096	0.052	0.153	-2.141
Isolation index Asian	0.029	0.045	0.013	0.016	11.183
Isolation index Black	0.135	0.168	0.223	0.243	-5.746
Isolation index Hispanic	0.107	0.131	0.178	0.246	-4.654
Isolation index Pacific Islander	0.005	0.009	0.003	0.007	4.374
Isolation index non-Hispanic White	0.831	0.149	0.666	0.222	11.842
Proportion poor or no English language	0.013	0.036	0.062	0.142	-5.642
% Uninsured aged 40–64	16.348	5.163	22.084	5.256	-16.8732
% Uninsured	17.702	5.217	23.077	4.871	-16.722
% Unemployed	9.441	2.981	9.570	3.274	-0.606
Mammography facilities per 1,000	0.146	0.147	0.114	0.189	2.667
proportion managed care penetration	0.182	0.120	0.113	0.075	12.819
Proportion persistent poverty	0.062	0.242	0.424	0.495	-11.771
% Poor Black rural	1.471	3.297	5.353	7.831	-8.024
% Poor Black urban	0.676	1.833	1.157	3.414	-2.216
% Poor Hispanic rural	1.097	2.231	3.366	6.504	-5.671
% Poor Hispanic urban	0.561	1.413	1.021	3.534	-2.110
% Poor non-Hispanic White rural	7.934	5.363	8.017	6.238	0.205
% Poor non-Hispanic White urban	1.690	2.601	0.657	1.507	9.357
Proportion population in poverty	0.150	0.058	0.210	0.082	-11.633
Proportion vacant housing	0.227	0.257	0.263	0.224	-2.404
Proportion heat with wood fire	0.058	0.071	0.038	0.064	4.608
Proportion living crowded quarters	0.005	0.007	0.008	0.010	-4.653
Proportion no plumbing	0.007	0.007	0.010	0.014	-3.004
County typology codes					
Proportion emphasis farming	0.102	0.303	0.186	0.390	-3.371
Proportion emphasis low Education	0.109	0.312	0.383	0.487	-8.939
Proportion emphasis Manufacturing	0.169	0.375	0.108	0.311	2.922
Proportion emphasis mining	0.047	0.212	0.130	0.337	-3.919
Proportion emphasis population loss	0.121	0.326	0.234	0.424	-4.190

^aAll two-sided *t*-test statistics are significant at $p < 0.01$, unless bolded values. SD, standard deviation.

Persistent hot spots were also more likely to be persistently poor counties, with higher poverty levels and higher proportions of poor, rural African Americans or poor, rural Hispanics; to have lower proportions of poor, urban non-Hispanic Whites; to have proportionately more crowded housing; to have proportionately more emphasis on farming; and to have higher population loss to outmigration (a signal of decline).

Discussion

There were two key contributions of this study. (i) Of the 2,599 total counties examined, there were 269 counties identified as persistent hot spots and 1,800 counties identified as never being a hot spot for late-stage breast cancer diagnoses. (ii) Inhabitants of counties with persistent hot spots (compared with counties that were never hot spots) were located in significantly more deprived areas with worse housing characteristics, lower socioeconomic status, lower levels of health insurance, worse access to mammography, more isolated American Indian/Alaska Native, African American, or Hispanic neighborhoods, and larger income disparity.

This study examined a broader range of socioeconomic conditions than seen in previous literature, and results suggest that deprivation, lower levels of health insurance, and worse access to mammography are more prevalent in areas with higher rates of late-stage diagnoses. The findings are consistent with previous findings from existing studies. Existing studies have identified both personal and contextual risk factors related to late-stage breast cancer diagnosis, focusing on one or a few states. These factors included race or ethnicity, being unmarried, having fewer routine health care visits, having more comorbidities, being older, lower use of mammography screening, and being insured under Medicare fee-for-service as compared with Medicare managed care plans.^{16–25}

Area-level contextual factors in the literature have included lower household income and educational attainment.^{19–22} A study by Haas et al.²⁶ reported that residential isolation of racial or ethnic groups in combination with area-level income mediated the likelihood of a late-stage breast cancer diagnosis among older women across the SEER program populations. Specifically, researchers found that African American or Hispanic women who lived in low-isolation and low-income areas (*i.e.*, below 75th percentile isolation and median income of <200% of the federal poverty threshold, respectively) have a higher probability of a late-stage breast cancer diagnosis than non-Hispanic White women.

Previous studies used a combination of traditional multi-level analysis and mapping to show geographic areas that have higher risk of late-stage breast cancer.^{27–31} For example, Gumpertz et al.²⁸ found that, in Los Angeles County, advanced breast cancer diagnoses were more likely among areas that had a larger proportion of racial or ethnic groups or low-median household incomes. A study by MacKinnon et al.²⁹ used spatial cluster analysis to identify areas with higher than expected incidence of late-stage diagnosis. They then used logistic regression on the binary area type (clustered, not clustered) to examine associations between these variable and contextual factors. They found that areas with

severe or near poverty and lower mammography use were more likely to be located in the clusters of higher-than-expected incidence of late-stage breast cancer.

Similarly, several recent articles have used multilevel modeling with characteristics of residential isolation by race or ethnicity to predict counties with the largest late-stage breast cancer diagnosis rates.^{7,32,33} These rankings and underlying data can be used to identify places with greater need for additional study and potential for evidence-based intervention. Other studies focused on the effect of residential segregation,³⁴ variations across the states,⁵ and how both residential segregation and state-level regulations help predict late-stage breast cancer incidence.⁷ In general, the national-level effect estimates⁷ masked the underlying variability found when examining states separately for evidence of isolation effects within states.³⁴ The spatial analysis performed in this article also highlights the local variability.

There are limitations to the research conducted here related to the requirements and sufficiency of cancer data available to researchers. However, we use the most comprehensive, population-based, geographically explicit data available. There are restrictions on the use of the data and reporting of results, and the county is the smallest geographic level defined. Lists of counties identified or drawing of actual county boundaries in maps is prohibited. Using a smaller or larger geographic unit would produce different patterns of clusters with the LISA test method. Thus, results are sensitive to the geographic units available for analysis.

The main limitations of many previous studies of late-stage breast cancer diagnoses are related to their limited geographic scope. Our study limitations include the exclusion of seven states because of the lack of permissions to use county-level descriptors in their outcomes data. The LISA local spatial clustering statistics highlight clusters with high rates of late-stage breast cancer diagnosis relative to the complete set of counties in the data set. Thus, it is possible that adding some or all the missing states would change the observed patterns of spatial clusters.

In addition, this study is descriptive rather than confirmatory, and proposes no causal structures. However, the scope of this study is wider than most other studies conducted to date in terms of the factors studied. The proper interpretation of these factors is as a group of descriptors that characterize those places with persistent hot spots. Ranking of the likely effects of various factors is beyond the scope of the methods used here and would require multivariate modeling.

Given these limitations, the availability of population-based cancer data spanning decades and covering expansive geographies and geographic detail represents a valuable resource. We hope this article will encourage other researchers to use these data. Future work might explore drivers of higher rates in hotspot areas, including possible differences in screening and follow-up.

Conclusion

We found geographic disparities in late-stage breast cancer diagnosis rates, with some communities experiencing persistent disparities over time. Results from this study can be one of several tools to guide the design and implementation of plans by state and local cancer control programs to prevent and control late-stage breast cancer. This information can be

used in combination with other data for organized breast cancer screening programs, such as the National Breast and Cervical Cancer Early Detection Program, whose aim is to provide low-income, uninsured, and medically underserved women access to timely breast and cervical cancer screening and diagnostic services.³⁵ Identifying hot spots and persistent hot spots for late-stage breast cancer among certain communities can enhance public health initiatives to identify and diagnose breast cancer at earlier stages.

Acknowledgments

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the National Cancer Institute, the National Institutes of Health, Georgia State University, or the Centers for Disease Control and Prevention.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

Funding support for L.R.M. was provided by the Center for Disease Control and Prevention (Interpersonal Agreement No. 16IPA1604431). The National Program of Cancer Registries contributed funds to cover the standard Research Data Center fees for researchers conducting analyses under approved research projects. Funding support was also provided for data development by L.R.M. under a National Cancer Institute Grant (2R01CA126858).

References

- Howlander N, Noone AM, Krapcho M, et al., eds. SEER Cancer Statistics Review, 1975–2017, National Cancer Institute. Bethesda, MD. Available at: https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020. Accessed December 1, 2020.
- U.S. Preventive Services Task Force. September 2017. Available at: <https://www.uspreventiveservicestaskforce.org/uspstf/> Accessed December 1, 2020.
- Verdial FC, Etzioni R, Duggan C, Anderson BO. Demographic changes in breast cancer incidence, stage at diagnosis and age associated with population-based mammographic screening. *J Surg Oncol* 2017;115:517–522.
- Miller JW, Smith JL, Ryerson AB, Tucker TC, Allemani C. Disparities in breast cancer survival in the United States (2001–2009): Findings from the CONCORD-2 study. *Cancer* 2017;123:5100–5118.
- Mobley L, Kuo M, Scott L, Rutherford Y, Bose S. Modeling geospatial patterns of late-stage diagnosis of breast cancer in the US. *Int J Environ Res Public Health* 2017;14:484.
- Henley SJ, King JB, German RR, Richardson LC, Plescia M. Surveillance of screening-detected cancers (colon and rectum, breast, and cervix)—United States, 2004–2006. *MMWR Surveill Summ* 2010;59:1–25.
- Mobley L, Kuo T. Geographic and demographic disparities in late-stage breast and colorectal cancer diagnoses across the US. *AIMS Public Health* 2015;2:583–600.
- Centers for Disease Control and Prevention (CDC). U.S. cancer statistics. The official federal cancer statistics, 2018. Available at: <https://www.cdc.gov/cancer/uscs/> Accessed December 1, 2020.
- Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. Research Data Center (RDC). May 2020. Available at: <https://www.cdc.gov/rdc/> Accessed April 2021.
- USCS. United States cancer statistics restricted access data set data dictionary and data standards. 2013 November Data Submission, 2013. Available at: https://www.cdc.gov/rdc/data/b1/NPCR_CSS_RDC_DataDictionary_2013.pdf Accessed November 2019.
- Massey DS, Denton NA. The dimensions of residential segregation. *Social Forces* 1988;67:281–315.
- Anselin L. Local indicators of spatial association—LISA. *Geogr Anal* 1995;27:93–115.
- Schieb LJ, Mobley LR, George M, Casper M. Tracking stroke hospitalization clusters over time and associations with county-level socioeconomic and health care characteristics. *Stroke* 2013;44:146–152.
- Snedecor GW, Cochran WG. *Statistical methods*, 8th ed. Ames, IA: Iowa State University Press, 1989.
- Kuo T, Mobley L. How generalizable are the SEER registries to the cancer populations of the USA? *Cancer Causes Control* 2016;27:1117–1126.
- Lipscomb J, Fleming ST, Trentham-Dietz A, et al. What predicts an advanced-stage diagnosis of breast cancer? Sorting out the influence of method of detection, access to care, and biologic factors. *Cancer Epidemiol Biomarkers Prev* 2016;25:613–623.
- Badgwell BD, Giordano SH, Duan ZZ, et al. Mammography before diagnosis among women age 80 years and older with breast cancer. *J Clin Oncol* 2008;26:2482–2488.
- Galit W, Green MS, Lital K. Routine screening mammography in women older than 74 years: A review of the available data. *Maturitas* 2007;57:109–119.
- Keating NL, Landrum MS, Ayanian JZ, Winer EP, Guadagnoli E. The association of ambulatory care with BC stage at diagnosis among Medicare beneficiaries. *J Gen Intern Med* 2004;20:38–44.
- Kirsner RS, Ma F, Fleming L, et al. The effect of Medicare health care systems on women with breast and cervical cancer. *Obstet Gynecol* 2005;105:1381–1388.
- Lee-Feldstein A, Feldstein PJ, Buchmueller T, Katterhagen G. Breast cancer outcomes among older women: HMO, fee-for-service, and delivery system comparisons. *J Gen Intern Med* 2001;16:189–199.
- McCarthy EP, Burns RB, Coughlin SS, et al. Mammography use helps to explain differences in breast cancer stage at diagnosis between older black and white women. *Ann Intern Med* 1998;128:729–736.
- Randolph WM, Goodwin JS, Mahnken JD, Freeman JL. Regular mammography use is associated with elimination of age-related disparities in size and stage of breast cancer at diagnosis. *Ann Intern Med* 2002;137:783–790.
- Riley GF, Potosky AL, Klabunde CN, Warren JL, Ballard-Barbash R. Stage at diagnosis and treatment patterns among older women with breast cancer: An HMO and fee-for-service comparison. *J Am Med Assoc* 1999;281:720–726.
- Riley GF, Warren JL, Potosky AL, Klabunde CN, Harlan LC, Ossiwald MB. Comparison of cancer diagnosis and treatment in Medicare fee-for-service and managed care plans. *Med Care* 2008;46:1108–1115.

26. Haas JS, Earle CC, Orav JE, et al. Racial segregation and disparities in breast cancer care and mortality. *Cancer* 2008;113:2166–2172.
27. Goovaerts P. Visualizing and testing the impact of place on late-stage breast cancer incidence: A non-parametric geostatistical approach. *Health Place* 2010;16:321–330.
28. Gumpertz ML, Pickle LW, Miller BA, Bell BS. Geographic patterns of advanced breast cancer in Los Angeles: Associations with biological and sociodemographic factors (United States). *Cancer Causes Control* 2006;17:325–339.
29. MacKinnon J, Duncan R, Huang Y, et al. Detecting an association between socioeconomic status and late stage breast cancer using spatial analysis and area-based measure. *Cancer Epidemiol Biomarker Prev* 2007;16:756–762.
30. McElroy JA, Remington PL, Gangnon RE, Hariharan L, Andersen LD. Identifying geographic disparities in the early detection of breast cancer using a geographic information system. *Prev Chronic Dis* 2006;3:A10.
31. Roche LM, Skinner R, Weinstein RB. Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. *J Public Health Manag Pract* 2002;8:26–32.
32. Kuo T, Mobley L, Anselin L. Geographic disparities in late-stage breast cancer diagnosis in California. *Health Place* 2011;17:327–334.
33. Mobley L, Kuo T, Watson L, Brown G. Geographic disparities in late-stage cancer diagnosis: Multilevel factors and spatial interactions. *Health Place* 2012;18:978–990.
34. Mobley L, Kuo T. Demographic disparities in late-stage diagnosis of breast and colorectal cancers across the USA. *J Racial Ethn Health Disparities* 2017;4:201–212.
35. Centers for Disease Control and Prevention. National breast and cervical cancer early detection program (NBCCED). October 2020. Available at: <http://www.cdc.gov/cancer/nbccedp/about.htm> Accessed December 1, 2020.

Address correspondence to:

Florence K.L. Tangka, PhD

*Division of Cancer Prevention and Control
Centers for Disease Control and Prevention (CDC)*

4770 Buford Hwy, MS 107-4

Atlanta, GA 30341

USA

E-mail: ftangka@cdc.gov