

Spatial Clusters of Breast Cancer Mortality and Incidence in the Contiguous USA: 2000–2014

Raid W. Amin, BS, MS, Ph.D.¹, Bridget A. Fritsch, BS, MS¹, and John E. Retzlaff, DO²

¹Department of Mathematics and Statistics, University of West Florida, Pensacola, FL, USA; ², Sacred Heart Health System, Pensacola, FL, USA.

BACKGROUND: Clusters of breast cancer with varied incidence or mortality are known to exist. No national scale of analysis of geographical variation in breast cancer incidence has been published before for the contiguous USA.

METHODS: This was a spatial cluster analysis of incidence and mortality data on breast cancer in the contiguous USA at the county resolution. Data for the years 2000–2014 were downloaded and analyzed with the software SaTScan with the goal to identify significant spatial clusters of breast cancer. Regression analysis was used to then adjust breast cancer incidence and mortality for several key risk factors such as age, smoking, particulate matter air pollution, physical inactivity, urban living, education level, and race.

RESULTS: Spatial clusters of counties for higher than expected breast cancer incidence and also for breast cancer mortality were identified. All identified clusters have $p < 0.05$. The mortality clusters show the mean breast cancer rates inside the cluster, while the incidence clusters show the relative risk inside each cluster. This is the first study of the contiguous USA for breast cancer mortality and incidence together. The clustering for mortality is quite different from the clustering for incidence. Using the software JOINPOINT, it is shown that the annual US downward trend for breast cancer mortality slowed down in recent years.

CONCLUSIONS: There exist several significant clusters in the contiguous USA, both for breast cancer incidence and for breast cancer mortality. Some of the clusters persisted even after adjusting for several key risk factors. These geographic areas warrant further investigation to potentially identify additional local concerns or needs to further address female breast cancer in those specific sites.

KEY WORDS: female breast cancer; incidence and mortality; spatial clusters; epidemiology.

J Gen Intern Med 34(3):412–9

DOI: 10.1007/s11606-018-4824-9

© Society of General Internal Medicine 2019

INTRODUCTION

The National Cancer Institute (NCI) states that the rate of new cases of female breast cancer was 124.9 per 100,000 women per year, while the rate of deaths was 21.2 per 100,000 women

per year for 2010–2014.¹ An extensive study was done on trends and patterns of disparities in cancer mortality among US counties.² In 2014, there were an estimated 3,327,552 women living with female breast cancer in the USA. In an article on cancer mortality trends for 1980–2014 in which 29 types of cancers were studied at the county level, it was found that breast cancer mortality rates statistically declined over that time period, which is consistent with the NCI's findings.³ Multiple studies have identified several risk factors for female breast cancer including screening,¹ smoking,⁴ diabetes,⁵ obesity,^{5,6} physical inactivity,⁷ educational attainment,⁸ race,⁸ poverty,⁹ and environmental factors such as air pollution.^{10–12} Additional risk factors, such as age at menarche, age at menopause, breast feeding, parity, and percent foreign born, could also be considered if such data can be obtained for all counties in the contiguous USA, but it may be very difficult or impossible to obtain data on all risk factors.¹³

The use of heat maps for breast cancer rates can be quite useful to show the geographic differences descriptively, but it is even more useful and important to add a cluster analysis based on modern disease surveillance algorithms by which some groups of adjacent counties with high cancer rates can be identified as a “clusters.” Such a cluster analysis also tested for the statistical significance of each identified cluster in order to weed out random groupings of counties. Identifying significant cancer clusters allows researchers to focus on geographic areas that need to be studied further to better understand why a certain grouping of adjacent counties display unusually high cancer incidence or mortality or both. Not all counties that are located within a high cluster may have high cancer rates. A cluster identified with the software SaTScan™ can identify a “disease outbreak” and it can warn Health Departments about such a possibility.^{14–16} A follow-up cluster analysis using only counties in a (large sized) cluster can identify small-sized hotspots, if desired.

The main goal of this project is to study any significant spatial clustering and space-time of female breast cancer mortality and spatial clustering of incidence of female breast cancer mortality between 2000 and 2014 in the contiguous USA. Any possible association of breast cancer clustering to county-level risk factors including age, smoking, race, physical inactivity, carcinogenic air pollution, education, and urban living is investigated. Clusters of counties with high breast cancer burdens are identified. Specifically, we aim to answer the following five research questions:

Received August 31, 2018

Revised October 22, 2018

Accepted December 19, 2018

Published online January 16, 2019

1. Are there geographical areas where the age-adjusted breast cancer mortality rate is significantly higher than that in the rest of the USA?
2. Are there geographical areas where the age-adjusted breast cancer incidence rate is significantly higher than that in the rest of the USA?
3. Is there a significant space-time interaction for breast cancer mortality in the contiguous USA?
4. Is there a significant association between the breast cancer mortality rate and any of the studied covariates?
5. Is there a significant association between breast cancer incidence rate and any of the studied covariates?

DATA AND METHODS

Data for age-adjusted mortality used in this study were downloaded from the IHME, which used small area estimation for any missing data.¹⁷ No mortality data was missing for any county for the time period 2000–2014. The total sample size was $n = 46,620$, which is based on 3108 counties for 15 years. The IHME data are based on spatially explicit Bayesian mixed effects regression models for cancer mortality.¹⁸ The breast cancer mortality data includes ICD9 (174–175.9, 217–217.8, 233.0, 238.3, 610–610.9) and ICD10 (C50–C50.929, D05–D05.92, D24–D24.9, D48.6–D48.62, D49.3, N60–N60.99). The incidence data from the Centers for Disease Control (CDC) is based on the SEER Site Recode ICD-0-3 codes C500–C509 definitions. Since breast cancer is not a rare type of cancer, no data was suppressed by the CDC. There were some missing data for some counties (around 3%), and we estimated the missing counts with the corresponding state rates so that the cluster analysis would not be much affected by such estimation.

The data set that we obtained from IHME had no missing mortality data, but the estimated mortality counts were non-integers. While the normal model in SaTScan™ could be used for any continuous data, counties with small populations would have the same weight as counties with large populations.¹⁹ Therefore, a weighted normal model was used with the disease surveillance software SaTScan™. The weighted normal model assigns different weights based on the county population.²⁰ The incidence counts were gathered from the CDC (2017), and were analyzed with a Poisson model in SaTScan™ after age adjusting with the Statistical Analysis System (SAS™), using Poisson Regression in proc glimmix. The GIS-based software ArcMap™ was then used to create the cluster maps.

Data on several covariates (smoking, diabetes, obesity, PM 2.5 air pollution, drug poisoning deaths, inactivity, poverty, urban living, race, and educational attainment) at the county level were obtained from the CDC,^{21–23} National Center for Health Statistics,¹⁷ United States Census Bureau,^{24–26} and United States Department of Agriculture.²⁷ Such risk factors

are not independent of each other, and we trimmed down the number of used covariates based on the association with breast cancer and we deleted from the list of covariates that were confounded with each other, such as drugs poisoning, obesity, and poverty. Using a correlation analysis with the software SAS™, the covariates were analyzed, checked for potential confounding.

In addition to age adjusting, mortality was then adjusted for smoking, particulate matter PM 2.5 air pollution, inactivity, and race using linear regression in SAS™. The residuals obtained from the fit of mortality on the covariates (age, smoking, physical inactivity, PM 2.5 air pollution, race) are basically the mortality adjusted for these covariates. Linear regression was used due to the IHME mortality data estimates not being integers. We attempted to round up/down the county mortality estimates to be able to use the Poisson model, but the rounding would greatly distort the resulting heat maps and also the cluster analysis. As for incidence counts, these were adjusted for smoking, air pollution, urban living, and education using Poisson regression (proc glimmix) in SAS™.

The resulting residuals from the regression are used in SaTScan™ for a spatial cluster analysis adjusted for the above selected covariates. The residuals are analyzed with a purely spatial cluster analysis, followed by using ArcMap™ to create cluster maps. Basically, “residuals” represent the information left on the cancer cases after modeling out all information on the specific covariate used in the regression model from the cancer variable. We identify clusters with information that is beyond what the covariate can account for. In the case of cancer incidence, the Poisson regression predicts the counts. This is then followed by analyzing the predicted covariate-adjusted incidence counts with SaTScan. www.satscan.org explains this methodology for the Poisson model. The raw incidence counts are used in the case file while the predicted counts from the regression analysis are used in the population file. The resulting clusters are clusters based on covariate-adjusted incidence. Using the data collected on mortality, a graph using the trend analysis software JOINPOINT™ was then created to check for any changes in the slope for female breast cancer mortality since 2000. The more sophisticated space-time analysis was also done with SaTScan for the breast cancer mortality. Since the CDC averages incidence data for 5 years at a time, we could not do a space-time analysis for breast cancer incidence data.

RESULTS

After running the correlation analysis on the available covariates, four covariates (in addition to age) for both mortality rate and incidence rate were chosen. Figure 1 gives a purely spatial map of age-adjusted breast cancer mortality rates where the four significant clusters are shown with (red) circles based on the SaTScan™-weighted normal model results. Each county is colored based on the age-adjusted breast cancer mortality

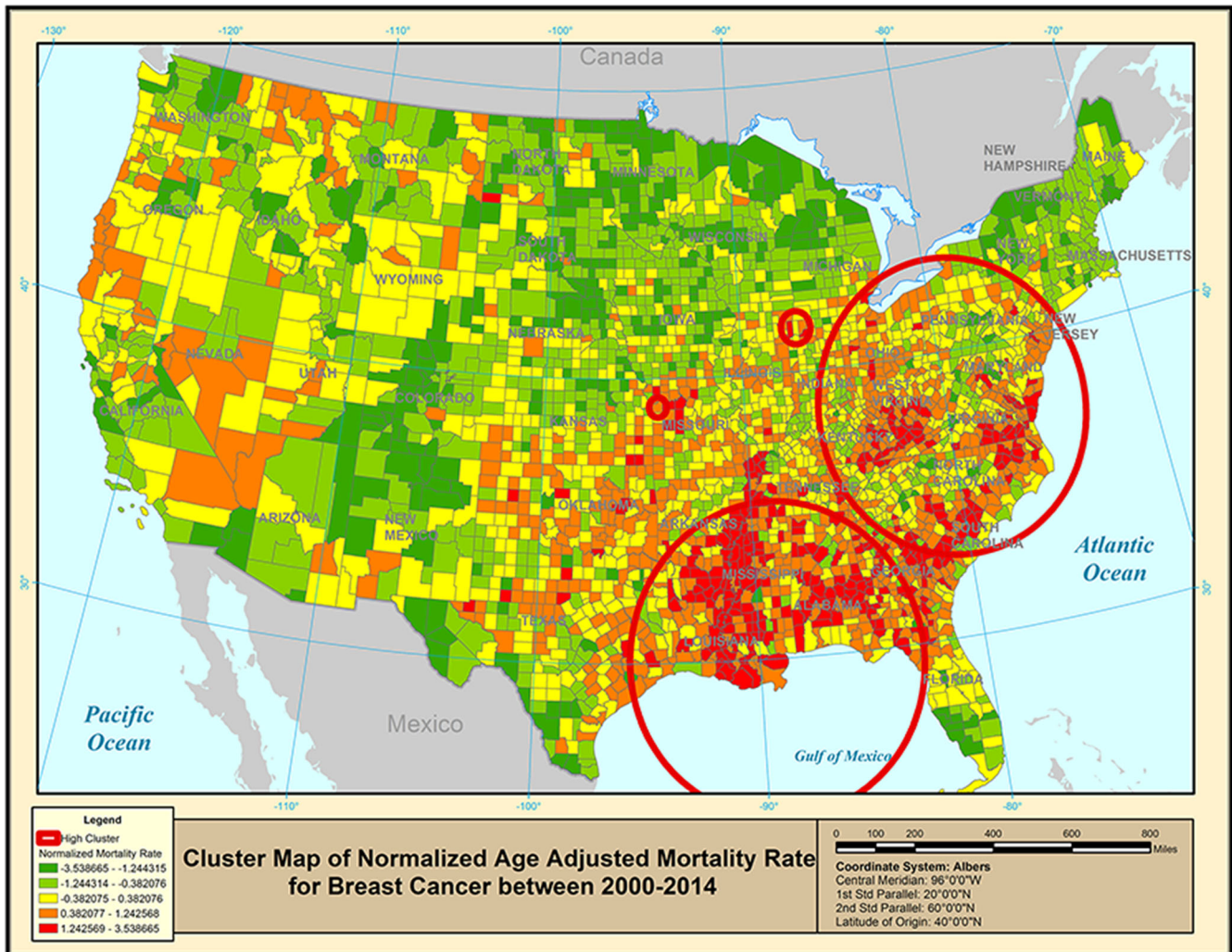


Figure 1 Purely spatial clusters of age-adjusted breast cancer mortality rate for 2000–2014.

rates, in addition to identifying (with red circles) the significant clusters. The Southeast cluster also shows a clustering around the Mississippi River, starting from New Orleans. The second cluster covers part of the East Coast, plus West Virginia. Table 1 lists some additional information on these spatial clusters.

The purely spatial cluster analysis was followed with a space-time analysis to address the question whether there exist geographical areas where there exists a breast cancer mortality space-time interaction. As shown in Table 1, there are two significant space-time clusters that resemble closely the two purely spatial clusters shown in Figure 1. The NE space-time cluster identifies the years 2000–2006 as the time period for the increase in breast cancer mortality in this geographic location. The normalized mortality rate inside this cluster is 0.67 standard deviations above the normalized national average (of zero). The SE space-time cluster was identified for the time period 2008–2014, with a mean of 0.94 inside the cluster.

The covariate-adjusted breast cancer mortality rates are shown in Figure 2, where the orange counties were present in a high age-adjusted breast cancer cluster when no other

covariates were used. Such clusters are not associated with any of the covariates used (smoking, air pollution, inactivity, and race) and yet they persisted and stayed in place inside the original clusters shown in Figure 1. There seem to exist other factors or covariates not studied here that may have led to such high breast cancer mortality rates. Such counties would require health agencies to look deeper into the high cancer rates. The red counties used to exist inside a high cluster in Figure 1, but these counties disappeared from the cluster after adjusting for the selected covariates. This means that in these red counties, we identify high age-adjusted breast cancer mortality that is associated with the selected covariates. In other words, the cancer rates are “explained” by the covariates. On the other hand, the yellow counties are counties that display high cancer rates only after adjusting the original rates for the selected covariates. Such counties appeared in Figure 1 as if their breast cancer mortality rates were not too high, but once adjusted for the covariates, these counties now display elevated breast cancer mortality rates. From Figure 2, the cluster in Missouri with three counties was explained by the covariates. Both the cluster in Indiana-Illinois and near the

Table 1 Breast Cancer Cluster Details

1(a) Purely spatial age-adjusted female breast cancer mortality clusters						
Cluster	Type	Location(s)	Weighted mean inside	Weighted mean outside	<i>P</i> -value	
1	High	MidEast Coast = D.C., DE, GA, IL, IN, KY, MD, MI, NC, NJ, NY, OH, PA, SC, TN, VA, WV	0.60	- 0.088	0.001	
2	High	South USA = AL, AR, FL, GA, LA, MS, TN, TX	0.92	0.015	0.001	
3	High	IL, IN, MI	1.14	0.070	0.001	
4	High	3 counties in KS, MO	1.19	0.090	0.043	
1(b) Space-time age-adjusted female breast cancer mortality clusters						
Cluster	Type	Years	Location(s)	Weighted mean inside	Weighted mean outside	<i>P</i> -value
1	High	2000–2006	Eastern Great Lakes = CT, D.C., DE, IL, IN, KY, MA, MD, MI, NJ, NY, OH, PA, VA, VT, WI, WV	0.67	- 0.033	0.001
2	High	2008–2014	Southern USA = AL, AR, FL, GA, LA, MS, TN, TX	0.94	0.057	0.001
1(c) Purely spatial age-adjusted female breast cancer incidence clusters						
Cluster	Type	Location(s)	Observed/expected	Relative risk		
1	High	North East Coast = CT, DE, D.C., MA, MD, ME, NC, NH, NJ, NY, OH, PA, RI, VA, VT, WV	1.06	1.08		
2	High	East of Orlando, FL	1.41	1.41		
3	High	MidWest Coast = CA, NV, OR	1.15	1.15		
4	High	South of Orlando, FL	1.17	1.18		
5	High	Northern Great Plains = IA, IL, KS, MI, MN, MO, ND, NE, SD, WI	1.06	1.07		
6	High	GA, NC, SC, TN	1.06	1.06		
7	High	Upper MI	1.23	1.23		
8	High	ID, MT, WA	1.13	1.13		

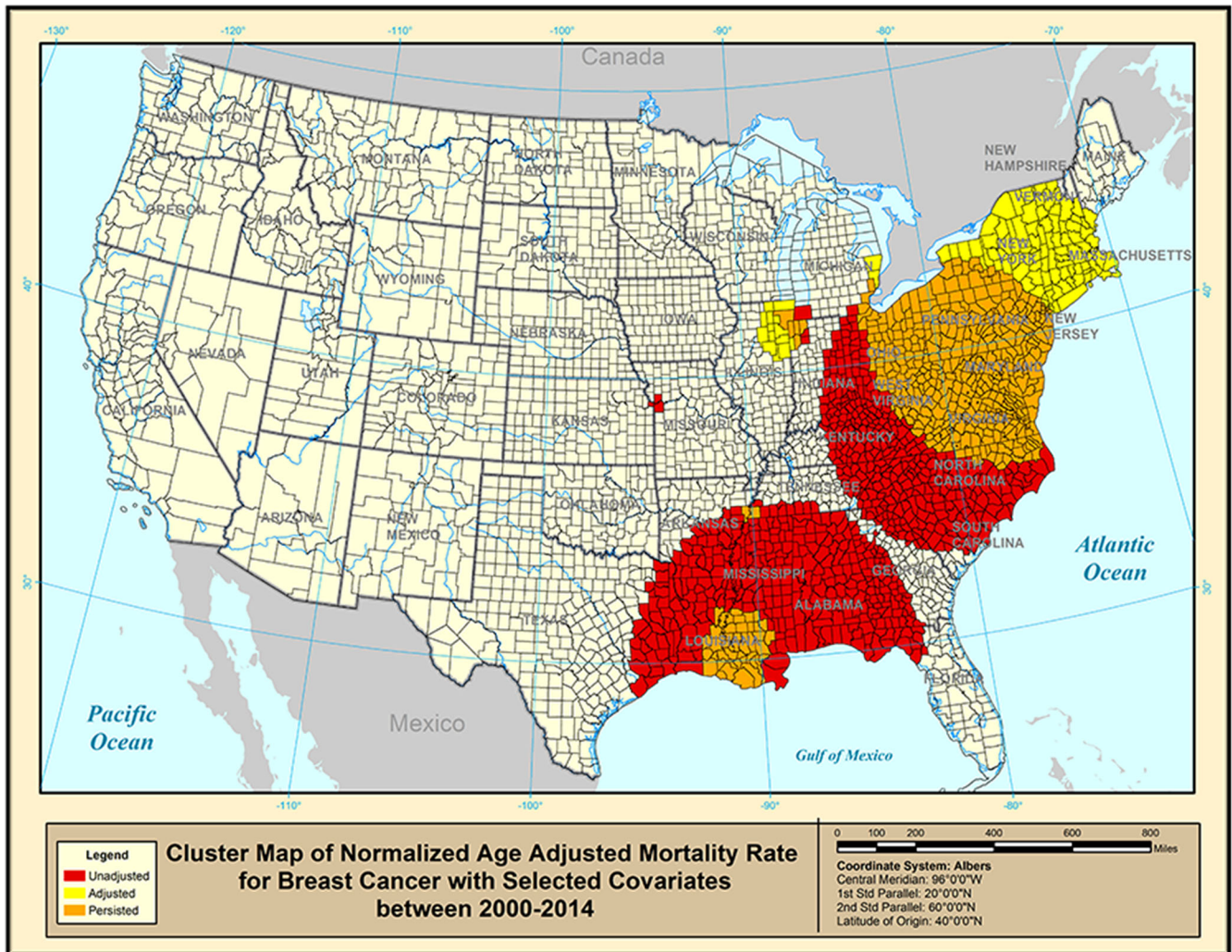


Figure 2 Purely spatial clusters of covariate-adjusted breast cancer mortality rate for 2000–2014.

mid-Atlantic coast shifted while the southern cluster decreased dramatically.

As for the incidence of breast cancer, a Poisson regression model is used to predict incidence counts (in SAS™). The resulting predicted counts are used in SaTScan™ Poisson model. Figure 3 gives a cluster map of breast cancer incidence for the USA. While some incidence clusters seem to be located in similar locations as some of the mortality clusters, the incidence map (Fig. 3) appears to be very different looking from the corresponding mortality map (Fig. 1). The mortality map had its clusters along the eastern part of the USA, while the incidence map shows clusters along the eastern coast and all along the northern border of the USA. It is useful to compare cluster maps based on breast cancer mortality with the corresponding maps for incidence since each type of cancer data can be associated with a different population SES. Table 1 gives details on the clusters.

Figure 4 shows the role of the selected covariates (smoking, educational attainment, urban living, and air

pollution) in adjusting the breast cancer incidence rates based on the results from the SaTScan™ Poisson model using the incidence data that has been adjusted for the selected covariates. Only the red-colored counties have high breast cancer incidence rates that can be associated with the chosen covariates. The clusters in the northern Midwest seemed to have merged together as well as the two clusters near the Pacific Northwest. The two clusters along the eastern coast seemed to have shifted and merged as well. The two clusters in Florida were not affected by the chosen covariates. The orange-colored counties have high breast cancer incidence rates that are not associated with the chosen covariates. There may exist some (unknown) factors that are associated with the high cancer incidence in such counties. The yellow counties were initially not inside any cluster before adjusting for the covariates.

In addition to the purely spatial analysis, we also analyzed the trend of the mortality rate data to address Research Question 5. Using JOINPOINT™ software and the age-adjusted mortality data, a “joinpoint” is identified in 2008.²⁸ The rate

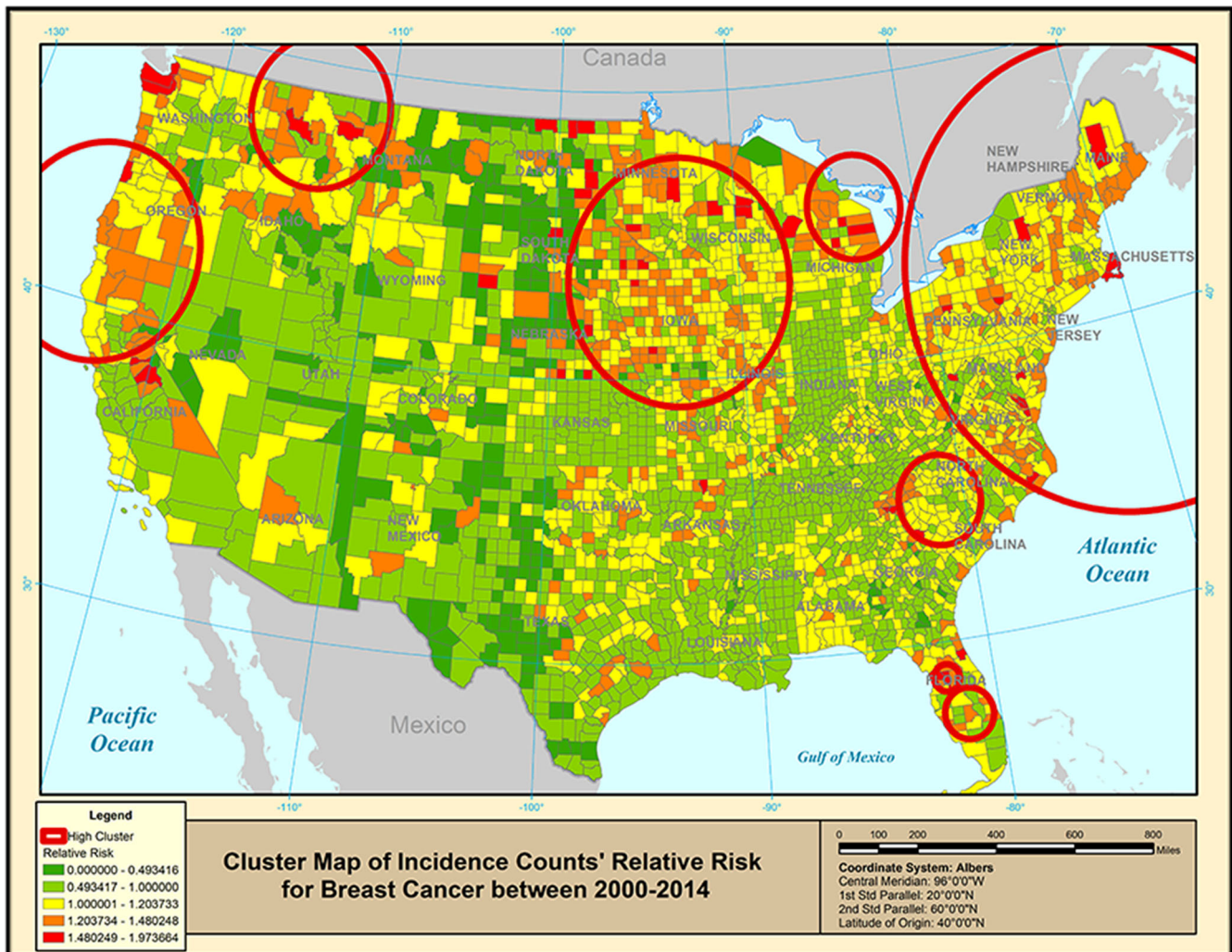


Figure 3 Purely spatial clusters of age-adjusted breast cancer incidence rates for 2000–2014.

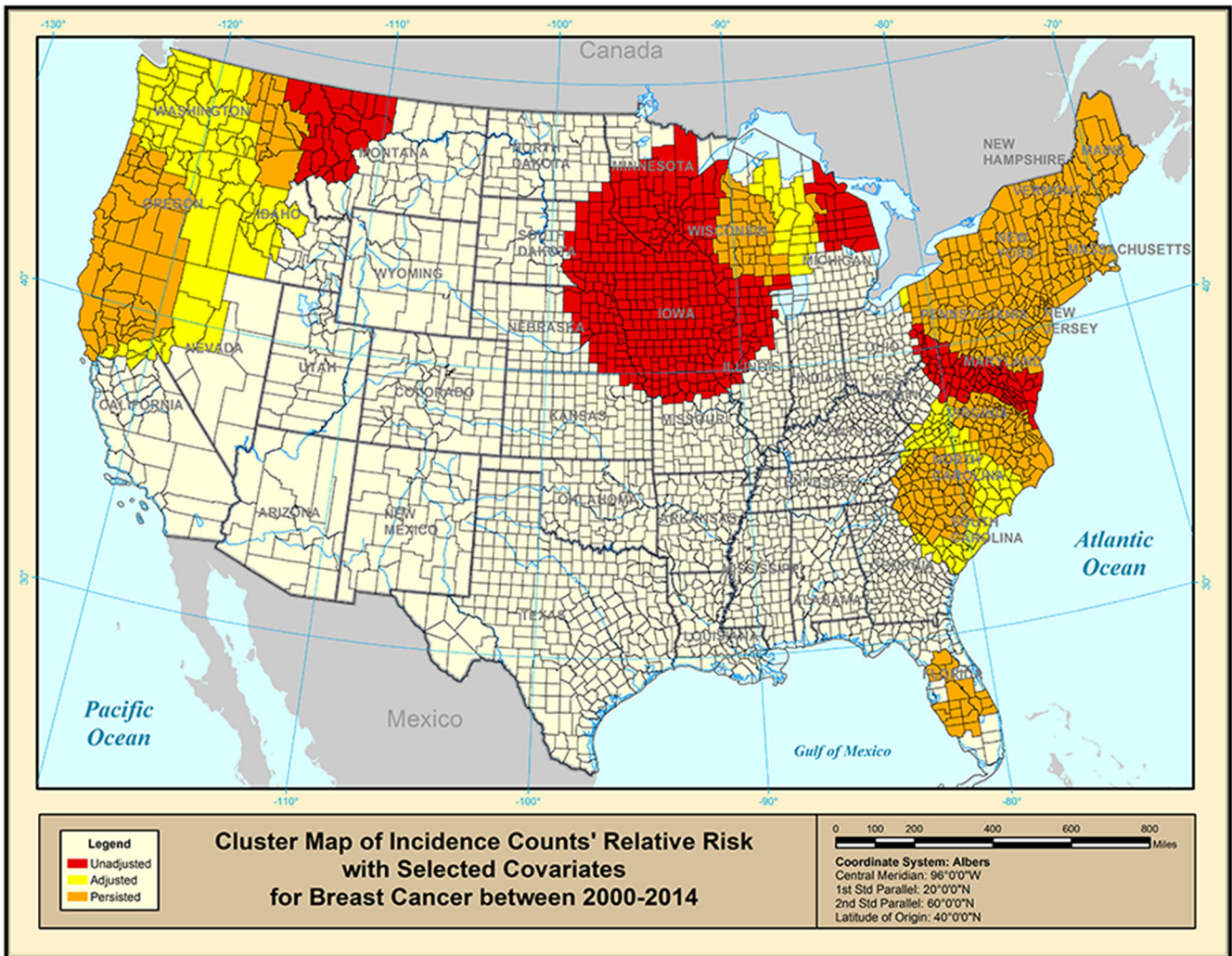


Figure 4 Purely spatial clusters of covariate-adjusted breast cancer incidence rate for 2000–2014.

of the decreasing mean normalized mortality rate changed. The breast cancer mortality rate used to be decreasing by 0.36 (per 100,000) each year for 2000–2008, changing to only a decline of 0.05 (per 100,000) for 2008–2014.

DISCUSSION

This is the first study of purely spatial breast cancer clusters in the contiguous USA for incidence in addition to including a purely spatial and a space-time cluster analysis of breast cancer mortality. This study provides an opportunity to contrast mortality clusters with incidence clusters, while carefully taking into account the possibility of several confounded effects when trying to shed light on why some clusters appeared on the map. Using SaTScan™, clusters of breast cancer mortality and incidence are identified for the contiguous USA, showing the role of the used covariates on such rates. While Figures 1 and 2 show mortality clusters to be located only in one part of the USA, incidence clusters are more spread out in several parts of the

contiguous USA (Figs. 3 and 4). Even after adjusting for known covariates and risk factors, several clusters were present, which could be attributed to factors outside the scope of this project and this should be investigated further. For example, we did not directly study breast cancer screening. It is well known that the frequency of mammograms in counties with low SES is lower than that in counties with high SES, and there will be a confounded effect with county-level poverty, say. Counties with low SES have lower rates of screening for breast cancer, resulting in lower incidence rates and higher mortality rates.^{29,30} Some other possible risk factors that were not studied here include parity, percent foreign born, age at menopause, and breast feeding. As our study focused at the county level, localized effects will not be found using data at this geographic resolution. Using smaller geographic units such as census tracts, data may be more effective in finding impacts at a smaller geographic level if needed. While the incidence data from the CDC was reported as averages of 5 years of counts, it is possible that 1 year may have had an unusually high or low rate for one particular

county. Such spikes will not be apparent in this incidence data. Further analysis using trends in mortality rate shows the year 2008 being a pivot year where the rate of the decreasing mortality rate slowed. This shift in mortality rate should be investigated further. The space-time cluster analysis of mortality rates reveals that the NE cluster has an increase in breast cancer mortality rates for the years 2000–2006, while the SE cluster has an increase for the years 2008–2014. Note that the space-time interaction for the SE cluster coincides with the JOINPOINT in 2008 that is based on national breast cancer mortality rates. The increase in the SE of the USA may have been the main reason why from 2008 the decrease in the national breast cancer mortality rates slowed down. It is worthwhile to look deeper into the identified differences in the geographic locations of the significant breast cancer mortality clusters versus the corresponding incidence clusters. Why are these clusters so different? Some possible explanations are (i) geographical variation in breast cancer screening: counties with higher SES may have a higher health insurance rate, which results in higher rate of visits to the physician for mammograms. Such screening will result in a higher cancer incidence but not necessarily in a higher mortality, (ii) geographical variation in the breast cancer type: a higher proportion of more aggressive cancers may lead to clusters of mortality but not necessarily to clusters of incidence, (iii) geographical variation in breast cancer treatment: it is possible that better treatment leads to lower mortality without affecting incidence, (iv) geographical variation in the concentration of women from ethnical groups with a higher probability of having a genetic predisposition for breast cancer, and (v) geographical variation in parity or hormone use.

In answer to our research questions, we conclude that the female breast cancer incidence and mortality data analysis in this study supports the association with certain covariates. However, there are certain geographic areas that appear to have incidence and mortality rates beyond the simple association with the studied covariates, as identified as significant clusters. These geographic areas warrant further investigation to potentially identify additional local concerns or needs to further address female breast cancer in those specific sites. In general, counties with high rates of smoking, blacks, physically inactive, and PM 2.5 air pollution also have high breast cancer mortality. Counties with high rates of smoking and college education have high incidence rates, but counties with high rates of urban living and PM 2.5 air pollution have low incidence rates.

Limitations of the Study. (i) While all counties in the contiguous USA had data on cancer mortality, 3% of the counties had missing incidence data which had to be estimated. The accuracy of the data may be a limitation factor. (ii) Not all possible risk factors were used. For

example, the data did not include information on stage of diagnosis. (iii) US counties were the smallest geographical units used in this study.

Acknowledgements: We would like to thank each of the three reviewers and all involved Editors for their suggestions and recommendations which have made our manuscript a much better document for publication.

Corresponding Author: Raid W. Amin, BS, MS, Ph.D.; Department of Mathematics and Statistics, University of West Florida, Pensacola, FL 32514, USA (e-mail: ramin@uwf.edu).

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. National Cancer Institute. SEER cancer stat facts: female breast cancer. National Cancer Institute. n.d.. Retrieved from <http://seer.cancer.gov/statfacts/html/breast.html>. Accessed 9 Nov 2018.
2. Pickle LW, Mungiole M, Jones GK, et al. Atlas of United States mortality. Hyattsville: National Center for Health Statistics; 1996.
3. Mokdad AH, Dwyer-Lindgren L, Fitzmaurice C, et al. Trends and patterns of disparities in cancer mortality among US counties, 1980–2014. *JAMA*. 2017;317(4):388–406.
4. Terry PD, Rohan TD. Cigarette Smoking and the Risk of Breast Cancer in Women. *Cancer Epidemiol Biomark Prev*. 2002;11(10):953–971.
5. Danaei G, Hoorn SV, Lopez AD, Murray CJ, Ezzati M. Causes of cancer in the world: Comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*. 2005;366(9499):1784–1793.
6. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*. 2003;348:1625–1638.
7. Thune I, Brenn T, Lund E, Gaard M. Physical activity and the risk of breast cancer. *N Engl J Med*. 1997;336:1269–1275.
8. Albano JD, Ward E, Jemal A, et al. Cancer Mortality in the United States by Education Level and Race. *J Natl Cancer Inst*. 2007;99(18):1384–1394.
9. Freeman HP. Poverty, culture, and social injustice: Determinants of cancer disparities. *CA Cancer J Clin*. 2004;54(2):72–77.
10. Crouse DL, Goldberg MS, Ross NA, Chen H, Labrèche F. Postmenopausal breast cancer is associated with exposure to traffic-related air pollution in Montreal, Canada: A case-control study. *Environ Health Perspect*. 2010;118(11):1578–1583.
11. Hystad P, Villeneuve PJ, Goldberg MS, Crouse DL, Johnson K. Exposure to traffic-related air pollution and the risk of developing breast cancer among women in eight Canadian provinces: A case-control study. *Environ Int*. 2015;74:240–248.
12. Sheehan TJ, DeChello LM, Kuldorff M, Gregorio DI, Gershman S, Mroszczyk M. The geographic distribution of breast cancer incidence in Massachusetts 1988–1997, adjusted for covariates. *Int J Health Geogr*. 2004;3:17.
13. Chen Y, Yi G, Mao Y. Cluster of liver cancer and immigration: a geographic analysis of incidence data for Ontario 1998–2002. *Int J Health Geogr*. 2008;7:28.
14. Kuldorff M. A spatial scan statistic. *Commun Stat Theory Methods*. 1997;26:1481–1496.
15. Kuldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: A geographic analysis. *Am J Epidemiol*. 1997;146(2):161–170.
16. Kuldorff M, Information Management Services, Inc. SaTScanTM v8.0: Software for the spatial and space-time scan statistics. 2009. <http://www.satscan.org/>. Accessed 9 Nov 2018.
17. Institute for Health Metrics and Evaluation (IHME). Data from: United States Cancer Mortality Rates by County 1980–2014 [Dataset]. Institute

- for Health Metrics and Evaluation (IHME). 2017. Retrieved from <http://ghdx.healthdata.org/record/united-states-cancer-mortality-rates-county-1980-2014>. Accessed 9 Nov 2018.
18. **Srebotnjak T, Mokdad AH, Murray CJL.** A novel framework for validating and applying standardized small area measurement strategies. *Popul Health Metrics.* 2010;8:26.
 19. **Kulldorff M, Huang L, Konty K.** A scan statistic for continuous data based on the normal probability model. *Int J Health Geogr.* 2009;8:58.
 20. **Huang L, Huang L, Tiwari R, Zuo J, Kulldorff M, Feuer E.** Weighted normal spatial scan statistic for heterogeneous population data. *J Am Stat Assoc.* 2009;104:886–898.
 21. Centers for Disease Control and Prevention (CDC). Data from: Environmental Public Health Tracking Network [Dataset]. 2017. Retrieved from www.cdc.gov/ephtracking. Accessed 9 Nov 2018.
 22. Centers for Disease Control and Prevention (CDC). Data from: Diabetes Home: County Data Indicators [Dataset]. Centers for Disease Control and Prevention (CDC). 2016. Retrieved from <https://www.cdc.gov/diabetes/data/countydata/countydataindicators.html>. Accessed 9 Nov 2018.
 23. **Rossen LM, Bastian B, Warner M, Khan D, Chong Y.** Data from: Drug Poisoning Mortality: United States, 1999–2016 [Dataset]. *Natl Center Health Stat.* 2017. Retrieved from <https://www.cdc.gov/nchs/data-visualization/drug-poisoning-mortality>. Accessed 9 Nov 2018.
 24. United States Census Bureau. Data from: Historical County Level Poverty Estimates Tool [Dataset]. U. S. Census Bureau. 2018. Retrieved from <https://www.census.gov/library/visualizations/time-series/demo/census-poverty-tool.html>. Accessed 9 Nov 2018.
 25. United States Census Bureau. Data from: Programs and Surveys [Dataset]. U. S. Census Bureau. 2016. Retrieved from <https://www2.census.gov/programs-surveys/pepest/datasets/>. Accessed 9 Nov 2018.
 26. United States Census Bureau. Data from: American FactFinder [Dataset]. U. S. Census Bureau. 2010. Retrieved from <https://factfinder.census.gov>. Accessed 9 Nov 2018.
 27. **Kim HJ, Fay MP, Feuer EJ, Midthune DN.** Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 2000;19:335–51.
 28. United States Department of Agriculture. Data from: Economic Research Service: County-level Data Sets [Dataset]. United States Department of Agriculture. 2017. Retrieved from <https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>. Accessed 9 Nov 2018.
 29. **Gregorio DI, Kulldorff M, Barry L, Samociuk H.** Geographic differences in invasive and in situ breast cancer incidence according to precise geographic coordinates, Connecticut, 1991–95. *Int J Cancer.* 2002;100:194–198.
 30. **MacKinnon JA, Duncan RC, Huang Y, Lee DJ, Fleming LE, Voti L, Rudolph M, Wilkinson JD.** Detecting an Association between Socioeconomic Status and Late Stage Breast Cancer using Spatial Analysis and Area-Based Measures. *Cancer Epidemiol Biomark Prev.* 2007;16(4):756–62.