Investigation Of Geographical Disparities: The Use of An Interpolation Method For Cancer Registry Data

Janis E. Campbell
The University of Oklahoma Health Sciences Center

Ami Elizabeth Sedani (ami-sedani@ouhsc.edu)
The University of Oklahoma Health Sciences Center https://orcid.org/0000-0001-8758-0826

Hanh Dung N. Dao
The University of Oklahoma Health Sciences Center

Ayesha B Sambo
The University of Oklahoma Stephenson Cancer Center

Mark P. Doescher
The University of Oklahoma Stephenson Cancer Center

Amanda E. Janitz
The University of Oklahoma Health Sciences Center

Research Article

Keywords: cancer, public health, ZIP Code Tabulation Areas

Posted Date: June 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-592167/v1

Abstract

Purpose: This study aimed to demonstrate a method to systematically illustrate cancer incidence in an accurate and visually attractive, yet straightforward, GIS-based method using ZIP Code level cancer registry data for all cancers combined and four major types of cancer. This analysis will better inform public health researchers by enabling them to empirically assess and describe geographic disparities in cancer for their own areas.

Methods: We describe a process of creating smoothed maps with an appropriate, well-described, simple, replicable method. We calculated SIRs for each ZIP Code Tabulation Areas (ZCTAs) for each cancer type. For the IDW, we used the ArcGIS 10.8.1 to create smoothed maps of Oklahoma ZCTA SIR.

Conclusion: This study demonstrates a method that public health practitioners can duplicate with minimal skills, no or low-cost applications, and limited data to assist health care professionals and the community in interpreting cancer in their state.

Background

Despite decreasing cancer death rates over the past decade, cancer remains the second most common cause of death in the United States (US) as of 2019, following only heart disease.[1] However, cancer mortality surpasses heart disease as the leading cause of death for multiple states, certain racial/ethnic groups, and in adults less than 80 years of age, thus demonstrating that multidimensional health disparities in cancer remain.[1, 2, 3]

The uneven distribution of resources and risks by place is foundational in epidemiology, including disease surveillance, access to care, understanding disparities in health and environmental risk, and health outcomes.[4–18] Cancer, with its sophisticated worldwide disease surveillance systems, provides many examples of the importance of geography, including breast cancer and income,[19–22] lung cancer and radon,[23] lung cancer and petrochemical,[24] lung cancer and ambient air,[25] environment and childhood cancer,[26] prostate cancer and ambient air concentrations,[27] colorectal cancer and segregation,[28] and geographic access to gynecological cancer care.[29–32] Therefore, a geographic information system (GIS) can play a major role in helping understand the spatial distribution of diseases such as cancer, and subsequently, informing policymakers for the allocation of scarce resources.[16, 17, 33]

While immensely beneficial, findings can be biased because of the variability in spatial methods in health research. Maps are often used to display geographic disparities without a theoretical underpinning to their development. Spatial autocorrelation is the presence of systematic variation in a feature or attribute where those closer together are more likely to have similar values. The tendency for cancer to cluster geographically has been recognized for centuries.[18, 34] The identification of such hot and cold spots has often resulted in heightened levels of fear, both publicly and among healthcare providers. The vast majority of cancer clusters are related to lifestyle choices and the distribution of socio-economic

characteristics with geography, such as breast cancer clusters in the high-income areas of Nassau and Suffolk counties in New York[35] and Marin County in California.[36] There have been several high-profile suspected, but unproven, geographic clusters, such as Camp Lejeune in North Carolina.[37–39] Most high-profile geographic clusters of cancer have been occupational, such as vermiculite mining in Libby, Montana.[40, 41]

Clusters of cancer have been represented with maps that show differing incidence or mortality rates based on administrative districts, such as state, county, or census tract in the form of choropleth maps. However, choropleth maps can be misleading with potentially serious consequences since these use political and administrative boundaries that may not represent true risk.[42–44] The creation of artificially imposed boundaries for administrative purposes can exclude geographic neighbors from analyses because they depend on the values that exclude neighbors.[44] Population size and land area may vary within the geographic units;[44] thus, choropleth maps are subject to small number problems, particularly in rural areas, and these maps typically do not include error estimates.[45] Moreover, choropleth map classification systems, such as natural breaks, equal intervals, or quantiles, can relay differing messages depending on the system used.[46–48] Smoothed maps, however, created with interpolation methods maintain the accuracy of significant high and low cluster locations better than choropleth maps while allowing clusters to be displayed visually. To our knowledge, well-described, easily accessible methods for displaying geographic disparities in cancer have not been published.

This study aimed to demonstrate a method to systematically illustrate cancer incidence in an accurate and visually attractive, yet straightforward, GIS-based method using ZIP Code level cancer registry data for all cancers combined and four major types of cancer (e.g. colorectal, lung, female breast, and prostate). This analysis will better inform public health researchers by enabling them to empirically assess and describe geographic disparities in cancer for their own areas.

Methods

Study Population and Data Sources

Cancer incidence data were obtained from the Oklahoma Central Cancer Registry (OCCR), Oklahoma's statewide cancer registry system, through a data-sharing agreement. Cancers were grouped by all cancers combined, colorectal cancer (ICD-0-3 18.0-18.9, 19.9, 20.9), lung cancer (C34. 0–34. 9), female breast cancer (C50. 0–50. 9), and prostate cancer (C61.9). Those whose histology codes specified mesotheliomas, Kaposi sarcomas, lymphomas (9050–9055, 9140, 9590, and 9989), and males with breast cancer were excluded. For each cancer type and all cancers combined, we received the number of cases for each diagnosis, the number and percentage of cases diagnosed at a late stage, age group (0–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and 85 or older), sex, and ZIP code of residence at diagnosis for the latest five years available (2013–2017).

Geographic Unit of Analysis

For this study, we aimed to use data from the smallest geographical unit available, which is the USPS ZIP Code level. Limitations related to the utilization of USPS ZIP code data in public health research, compared to census blocks or tracts, is well known.[49, 50] In response to this concern the US Census Bureau created ZIP Code Tabulation Areas (ZCTAs). While ZIP codes represent a collection of mail delivery routes established for use by the US Postal Service, ZCTAs are generalized areal representations.[50–52] It is important to note there are strengths and limitations to both of these geographic areas.[49, 53, 54] For purposes of this paper ZCTAs were considered adequate and are easily attainable; therefore, we used 2015 ZCTAs for mapping purposes.[55] ZCTAs may cross county lines and sometimes also cross state lines. For the purposes of this study of interpolation methods, ZCTA are sufficient to illustrate generalized areas.

Interpolation Methods

Using inverse distance weighting (IDW), a method where unknown points are calculated using the weighted average of the values of nearby known points, we created maps displaying the incidence rates of colorectal (male and female), female breast, lung (male and female), prostate, and all cancers combined in Oklahoma for 2013–2017. We divided cancer incidence into 13 classes, which were determined using geometrical interval breaks. This large number of classes was used to present a smooth transition between categories.[56] The geometrical interval classification method (or smart quantiles) is particularly good for visualizing continuous data; it lessens within-class variance and works well with “heavily skewed and duplicate values” introduced by the use of a Standardized Incidence Ratio (SIR).[57] An SIR is the observed number of cases divided by the expected number of cases of, in the present study, cancer. The expected number of cases is the number of cases that would have occurred if a standard was applied throughout the area; in the present study, the incidence rate of the state of Oklahoma from 2013–2017 was used as the standard. SIR is typically used when the occurrence of cancer in a relatively small population is disparate or a small number of observed cases occur, such as in ZIP Codes. This study had a heavily skewed SIR with this dataset having a skewness of, for example, 24.16 for all males and 3.10 for all females.

We used ZCTA polygon data for the US Census. ZIP codes were matched to their respective ZCTA; however, nine (n = 648) cancer case ZIP Codes did not match a ZCTA. For ZIP Codes that did not match, the ZIP code was geocoded (using ESRI® ArcGIS ready-to-use geocoding tool), and the resulting location was used to place the ZIP Code data within an appropriate ZCTA. There were 90 ZIP Codes that were either recently created or merged with another ZIP Code. These were placed on the map in the corrected area. We, then, used an incorporated places shapefile from the US Census Bureau for Oklahoma that included county, ZCTA, and incorporated places. We joined the population count to the incorporated places shapefile to determine the largest population center in the ZCTA. For those ZCTAs without incorporated places, the ZCTA centroid was used. We used a color scheme that transitioned from red to blue to indicate hot (red) and cold (blue) spots, which experienced higher and lower rates than the statewide rate, respectively.

Statistical Analysis

Figure 1 shows the final data workflow. We calculated SIRs for each ZCTA for each cancer type using SAS 9.4 (SAS Institute Inc. 2013, Cary, NC). For the IDW, we used the ArcGIS 10.8.1 to create smoothed maps of Oklahoma ZCTA SIR.

Indirect Age-Sex Standardization

The number of expected cancer cases for each ZIP code was determined by using indirect age-sex standardization and Oklahoma ZCTA population data with the following age groups, by years: 0–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and 85 or older. Indirect standardization was used rather than direct standardization because it applies the stable statewide rate to local populations, instead of applying local disease rates, which are unstable for small areas, to standard population weights.[56] An SIR is was used to determine whether the occurrence of cancer in a relatively small population was high or low.

For each ZCTA, to calculate the SIR for incidence (or proportion of late-stage diagnosis cases), the expected number of cancer cases (or the number of late-stage diagnosis cases) of each cancer was computed by applying the statewide rates to the numbers of people (or cases) in each age-sex group in the ZIP Code.

Hot Spot Analysis

To confirm that the interpolations were reasonable, we also created a choropleth map of the ZIP Code data to determine areas of high rates as a validation step. We then performed a Getis Ord Gi* to determine low (cold) or high (hot) areas or spots. Hot spot and cold spot analysis using the Getis-Ord Gi* statistic uses fixed distance band in ArcGIS software. The subsequent Z score identified ZIP Code centroids having high or low values of clustering spatially. Positive Z scores indicate the clustering of high values, or hot spots. Negative Z scores indicate clustering of low values, or cold spots. A Z score near zero indicates no apparent spatial clustering. The Getis-Ord Gi* statistic works by examining each feature within the context of adjacent features.[58]

Smoothed Maps Using Inverse Distance Weighting

Interpolation is used to estimate the values of intermediate and extended pixels (or point on a map) by applying a mathematical function to available data. Inverse distance weighted (IDW) interpolation determines pixel values using a linearly weighted combination of sample points with the weight as a function of inverse distance. The interpolated surface should be a geographically dependent variable, such as cancer incidence or late-stage cancer. IDW is often used to show interpolation for rainfall or elevation. IDW is represented by the following formula where Z_i is the value of known point, d_{ij} is the distance to the known point, Z_j is the unknown point, and n is a user-selected exponent.[59]

$$Z_j = \frac{\sum_i \frac{Z_i}{d_{ij}^n}}{\sum_i \frac{1}{d_{ij}^n}}$$

Map Production

The USA Contiguous Albers Equal Area Conic Projected projection was used for all maps. This study was approved by the IRB at the University of Oklahoma Health Sciences Center and the Oklahoma State Department of Health.

Results

Overall, there were 648 ZCTAs located in Oklahoma. Among those ZCTAs, there were 53,123 males and 53,486 females diagnosed with cancer from 2013–2017. The range of cancers diagnosed within each ZCTAs in Oklahoma was 0 to 775 (mean: 82.0) for males and 0 to 919 (mean: 82.5) for females. Six maps were created for all cancer types combined (Figs. 2 and 3). These maps show ZIP Code (Figs. 2a and 3a), hot and cold spots (Figs. 2b and 3b), and IDW Interpolated maps (Figs. 2c and 3c) for males and females for all cancers combined. Only the interpolated maps are shown for specific cancers and late-stage cancer.

Overall Cancer

When reviewing the three maps together (Fig. 2a-c and Fig. 3a-c), we see that a standard ZCTA choropleth map using five classifications with natural breaks does not depict a clear smoothed picture of cancer patterns (Fig. 2a and Fig. 3a). While the hot spot analysis (Getis-Ord Gi*) clearly shows areas of hot spots, this analysis leaves the impression that these spots are very precisely located (Fig. 2b and Fig. 3b). The IDW maps (Fig. 2c and Fig. 3c) show clearer and intuitive results. Hot spots for overall males diagnosed with cancer include areas throughout central Oklahoma, one high SIR (hot spot) in southwestern Oklahoma, and a few random hot spots in northern and northeastern parts of Oklahoma (Fig. 2c). Hot spots for overall females diagnosed with cancer were mainly in the northeastern portion of Oklahoma (Fig. 3c). No cold spots were observed.

Lung Cancer

Lung cancer in Oklahoma is pervasive (Fig. 3a-b).[60] Using SIR revealed that for males, there are large hot spots in the eastern and southern parts of the state, with a larger cold spot area in northwestern Oklahoma (Fig. 3a), compared with the high rates in Oklahoma overall. For males, the Oklahoma Metropolitan Area (Central Oklahoma) and the eastern part of the area have high rates than the north, west, and even southern areas of Central Oklahoma. For males, the Tulsa area shows hot spots (small) in the northwestern part of the county (Fig. 3a). Throughout Oklahoma, there were small hot spots for females (Fig. 3b). For females, large areas of northwest and western Oklahoma showed cold spots (Fig. 3b) compared with the overall Oklahoma rate. Finally, there was a large swath of cold spots from the southeastern to the northeastern parts of the state (Fig. 3b).

Colorectal Cancer

Hot spots for male colorectal cancer were located primarily in southeastern, northwestern, and southwestern Oklahoma, with a large hot spot in central Oklahoma county and northern Oklahoma (Fig. 4a). For females, the SIR showed only one hot spot in southeastern Oklahoma county (Fig. 4b).

For late-stage colorectal cancer, there are hot spots throughout the state, primarily in rural areas. There are hot and cold spots in both urban areas (Tulsa and Oklahoma counties); the hot spots are located in southwestern and northwestern Oklahoma county and northwestern, central, and southern Tulsa County. These geographically smaller, but highly populated, areas are not as visually obvious (Fig. 4c) as are the rural areas in Oklahoma.

Female Breast Cancer

Hot spots for female breast cancer include an urban areas with a higher SIR from southwest Oklahoma to northeast Oklahoma (Fig. 5a). There are also hot spots in southern Oklahoma and the panhandle (Fig. 5a). Late-stage breast cancer mapping suggests that rural areas have a higher concentration of hot spots than urban areas, although there are two large urban hot spots in the southern and northwestern Oklahoma City Metropolitan Area and the northwest Tulsa Metropolitan Area (Fig. 5b).

Prostate Cancer

For men in Oklahoma diagnosed with prostate cancer, the SIR showed hot spots in southwestern Oklahoma, in south central, north and south Tulsa, and in central Oklahoma (Fig. 6).

Discussion

GIS can play a major role in epidemiology, helping understand the spatial distribution of diseases, and thus informing allocation of resources. However, map making methodologies directly impact the subsequent visual output. The output can be misleading and thus lead to potentially serious consequences. Currently there is a lack of well described easily accessible methods for displaying geographic disparities in cancer. This study described and demonstrated a method to systematically describe cancer incidence that is accurate and visually attractive, yet simple to make, GIS-based method using state cancer registry data.

Choropleth maps (maps made from administrative districts such as counties) are the mainstay of spatial epidemiology. Choropleth maps, however, are often difficult to interpret. Interpolated maps are much easier for resource planners and the public to interpret. These smoothed maps allow researchers and community members to understand the geographic areas of interest for future resource planning. Working with the community, researchers and planners can then identify why some of these areas have high (hot) or low (cold) SIRs.

The present study used data from a high-quality data set (e.g., OCCR) and strong methods to create tools for understanding geographic cancer disparities in Oklahoma. Moreover, this study used an accepted

method to show the hot and cold spots using an indirect age standardization method. Because we know that cancer rates do not change at administrative borders, interpolation proposes a more realistic picture of cancer rates across a geographic area. Finally, this methodology is achievable using a well-documented industry-standard simple software program (ArcGIS), but can be completed in other GIS packages (QGIS or R).

Despite the strengths of this study, there are still limitations. First, spatial resolution may not be consistent since the maps are based on points with different densities (based typically on population).[56, 61] Another limitation may be the small sample size, particularly in the rural areas. Even combining five years of Oklahoma data, there were geographic areas based on small numbers. Also, aggregate estimates of cancer incidence across large geographic areas often mask differences within the area. While ZIP Codes are not typically large geographic areas, they can still mask differences, particularly in geographically large ZIP Codes, such as those in rural areas. Besides the overall ZIP Code size issue, the ZCTAs were used to represent ZIP Codes; thus, there are likely areas of geographic inconsistency. Although IDW does not smooth as well as some other methods (e.g., Empirical Bayesian Kriging), this project had the goals of producing maps that are accurate, smoothed, and easy to understand. While we considered adaptive spatial filters to create high-quality accurate maps,[62] for geographic areas with widely varying population density, we determined that IDW was as effective at producing maps that were accurate and legible. Moreover, IDW does not require specialized software, and there are many options, including ArcGIS, QGIS, and R, the latter two being open source, no-cost software. We believe that the methods we present in this paper provide an effective compromise that allows the pragmatic pinpointing of hot and cold spots.

Understanding the relationships between health and place is foundational in epidemiology and public health. Geographical areas can show areas in need of screening or preventive services. It can show areas that are doing well in screening or prevention efforts leading to improved public health activities. With the emergence of COVID-19 and the efforts of the Johns Hopkins University Coronavirus resources center maps (https://coronavirus.jhu.edu/us-map) the significance of GIS in public health has become even more apparent. This study demonstrates a method that public health practitioners can duplicate with minimal skills, no or low-cost applications, and limited data to assist health care professionals and the community in interpreting cancer in their state.

Declarations

Funding: JC was partially funded by National Institute of General Medical Sciences, Grant/Award Number: U5GM104938 and in part by the National Cancer Institute Cancer Center Support Grant P30CA225520 awarded to the University of Oklahoma Stephenson Cancer Center for use of the Biostatistics and Research Design Shared Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest: The authors declare no potential conflicts of interest.

Authors' contribution statements: J.E.C. and M.D. conceived of the presented idea. A.S. helped J.E.C. with data analysis and interpretation. J.E.C. took the lead in writing the manuscript with consultation and critical review from A.E.S. H.D.N.D validated methodology. A.E.J supervised the project. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements: We would like to acknowledge Alexandra Feld and Raffaella Espinoza for their assistance in acquiring the OCCR data.

Ethics Statement: This study was approved by the institutional review boards at the University of Oklahoma Health Sciences Center and the Oklahoma State Department of Health. I certify that the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Availability of data and material: The data that support the findings of this study are available from Oklahoma State Department of Health, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Oklahoma State Department of Health.

ACKNOWLEDGEMENTS

JC was partially funded by National Institute of General Medical Sciences, Grant/Award Number: U5GM104938 and in part by the National Cancer Institute Cancer Center Support Grant P30CA225520 awarded to the University of Oklahoma Stephenson Cancer Center for use of the Biostatistics and Research Design Shared Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We would like to acknowledge Alexandra Feld and Raffaella Espinoza for their assistance in acquiring the OCCR data.

References

1. Centers for Disease Control and Prevention, Statistics NCfH. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, released in 2020. Center for Disease Control and Prevention. Accessed 02/04/2021, 2021. http://wonder.cdc.gov/ucd-icd10.html

2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin*. Jan 2021;71(1):7-33. doi:10.3322/caac.21654

3. Murphy SL, Xu J, Kochanek KD, Arias E, Tejada-Vera B. Deaths: Final Data for 2018. *Natl Vital Stat Rep*. Jan 2021;69(13):1-83.

4. Gordis L. *Epidemiology*. Fifth edition. ed. Elsevier/Saunders; 2014:xv, 392 pages.

5. Lawson A, Banerjee S, Haining RP, Ugarte MaD. *Handbook of spatial epidemiology*. Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press/Taylor & Francis; 2016:xviii, 684 pages.

6. Krieger N. Place, space, and health: GIS and epidemiology. *Epidemiology*. Jul 2003;14(4):384-5. doi:10.1097/01.ede.0000071473.69307.8a

7. Cromley EK, McLafferty S. *GIS and public health*. 2nd ed. The Guilford Press; 2012:xxiv, 503 p.

8. Krieger N. Follow the North Star: Why Space, Place, and Power Matter for Geospatial Approaches to Cancer Control and Health Equity. *Cancer Epidemiol Biomarkers Prev*. Apr 2017;26(4):476-479. doi:10.1158/1055-9965.EPI-16-1018

9. Prehn AW, West DW. Evaluating local differences in breast cancer incidence rates: a census-based methodology (United States). *Cancer Causes Control*. Oct 1998;9(5):511-7. doi:10.1023/a:1008809819218

10. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Publ Health*. 1997;18:341-78. doi:10.1146/annurev.publhealth.18.1.341

11. Baade PD, Yu XQ, Smith DP, Dunn J, Chambers SK. Geographic disparities in prostate cancer outcomes–review of international patterns. *Asian Pac J Cancer Prev*. 2015;16(3):1259-75. doi:10.7314/apjcp.2015.16.3.1259

12. Wan N. Pesticides exposure modeling based on GIS and remote sensing land use data. *Appl Geogr*. Jan 2015;56:99-106. doi:10.1016/j.apgeog.2014.11.012

13. Hines R, Markossian T, Johnson A, Dong F, Bayakly R. Geographic residency status and census tract socioeconomic status as determinants of colorectal cancer outcomes. *Am J Public Health*. Mar 2014;104(3):e63-71. doi:10.2105/AJPH.2013.301572

14. Cooper RA, Cooper MA, McGinley EL, Fan X, Rosenthal JT. Poverty, wealth, and health care utilization: a geographic assessment. *J Urban Health*. Oct 2012;89(5):828-47. doi:10.1007/s11524-012-9689-3

15. 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*. Apr 2006;17(3):325-39. doi:10.1007/s10552-005-0513-1

16. Graves BA. Integrative literature review: a review of literature related to geographical information systems, healthcare access, and health outcomes. *Perspect Health Inf Manag*. Jul 29 2008;5:11.

17. Nykiforuk CI, Flaman LM. Geographic information systems (GIS) for Health Promotion and Public Health: a review. *Health Promot Pract*. Jan 2011;12(1):63-73. doi:10.1177/1524839909334624

18. Aghajani J, Farnia P, Velayati AA. Impact of geographical information system on public health sciences. *Biomedical and Biotechnology Research Journal (BBRJ)*. 2017;1(2):94.

19. Kuo TM, Mobley LR, Anselin L. Geographic disparities in late-stage breast cancer diagnosis in California. *Health Place*. Jan 2011;17(1):327-34. doi:10.1016/j.healthplace.2010.11.007

20. Meliker JR, Jacquez GM, Goovaerts P, Copeland G, Yassine M. Spatial cluster analysis of early stage breast cancer: a method for public health practice using cancer registry data. *Cancer Cause Control*. Sep 2009;20(7):1061-1069. doi:10.1007/s10552-009-9312-4

21. Grann V, Troxel AB, Zojwalla N, Hershman D, Glied SA, Jacobson JS. Regional and racial disparities in breast cancer-specific mortality. *Soc Sci Med*. Jan 2006;62(2):337-47. doi:10.1016/j.socscimed.2005.06.038

22. Sturgeon SR, Schairer C, Gail M, Mcadams M, Brinton LA, Hoover RN. Geographic-Variation in Mortality from Breast-Cancer among White Women in the United-States. *J Natl Cancer I*. Dec 20 1995;87(24):1846-1853. doi:DOI 10.1093/jnci/87.24.1846

23. Ou JY, Fowler B, Ding Q, et al. A statewide investigation of geographic lung cancer incidence patterns and radon exposure in a low-smoking population. *Bmc Cancer*. Jan 31 2018;18(1):115. doi:10.1186/s12885-018-4002-9

24. Lin CK, Hsu YT, Christiani DC, Hung HY, Lin RT. Risks and burden of lung cancer incidence for residential petrochemical industrial complexes: A meta-analysis and application. *Environ Int*. Dec 2018;121(Pt 1):404-414. doi:10.1016/j.envint.2018.09.018

25. Guo YM, Zeng HM, Zheng RS, et al. The association between lung cancer incidence and ambient air pollution in China: A spatiotemporal analysis. *Environ Res*. Jan 2016;144:60-65. doi:10.1016/j.envres.2015.11.004

26. Farazi PA, Watanabe-Galloway S, Westman L, et al. Temporal and geospatial trends of pediatric cancer incidence in Nebraska over a 24-year period. *Cancer Epidemiol*. Feb 2018;52:83-90. doi:10.1016/j.canep.2017.12.006

27. Weichenthal S, Lavigne E, Valois MF, et al. Spatial variations in ambient ultrafine particle concentrations and the risk of incident prostate cancer: A case-control study. *Environ Res*. Jul 2017;156:374-380. doi:10.1016/j.envres.2017.03.035

28. Mobley LR, Scott L, Rutherford Y, Kuo TM. Using residential segregation to predict colorectal cancer stage at diagnosis: two different approaches. *Ann Epidemiol*. Jan 2017;27(1):10-19. doi:10.1016/j.annepidem.2016.11.008

29. Shalowitz DI, Vinograd AM, Giuntoli RL, 2nd. Geographic access to gynecologic cancer care in the United States. *Gynecol Oncol*. Jul 2015;138(1):115-20. doi:10.1016/j.ygyno.2015.04.025

30. Najafabadi AT, Pourhassan M. Integrating the geographic information system into cancer research. *Indian J Cancer*. Jan-Mar 2011;48(1):105-109. doi:10.4103/0019-509x.75834

31. Siegel RL, Jemal A, Wender RC, Gansler T, Ma J, Brawley OW. An assessment of progress in cancer control. *CA Cancer J Clin*. Sep 2018;68(5):329-339. doi:10.3322/caac.21460

32. Cramb SM, Moraga P, Mengersen KL, Baade PD. Spatial variation in cancer incidence and survival over time across Queensland, Australia. *Spat Spatiotemporal Epidemiol*. Nov 2017;23:59-67. doi:10.1016/j.sste.2017.09.002

33. Sahar L, Foster SL, Sherman RL, et al. GIScience and cancer: State of the art and trends for cancer surveillance and epidemiology. *Cancer*. Aug 1 2019;125(15):2544-2560. doi:10.1002/cncr.32052

34. Koch T. *Cartographies of disease : maps, mapping, and medicine*. 1st ed. ESRI Press; 2005:xvii, 389 p.

35. Gammon MD, Neugut AI, Santella RM, et al. The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. *Breast Cancer Res Treat*. Jun 2002;74(3):235-54. doi:10.1023/a:1016387020854

36. Clarke CA, Glaser SL, West DW, et al. Breast cancer incidence and mortality trends in an affluent population: Marin County, California, USA, 1990-1999. *Breast Cancer Res*. 2002;4(6):R13. doi:10.1186/bcr458

37. Bove FJ, Ruckart PZ, Maslia M, Larson TC. Mortality study of civilian employees exposed to contaminated drinking water at USMC Base Camp Lejeune: a retrospective cohort study. *Environ Health*. Aug 13 2014;13:68. doi:10.1186/1476-069X-13-68

38. Bove FJ, Ruckart PZ, Maslia M, Larson TC. Evaluation of mortality among marines and navy personnel exposed to contaminated drinking water at USMC base Camp Lejeune: a retrospective cohort study. *Environ Health*. Feb 19 2014;13(1):10. doi:10.1186/1476-069X-13-10

39. Ruckart PZ, Bove FJ, Maslia M. Evaluation of exposure to contaminated drinking water and specific birth defects and childhood cancers at Marine Corps Base Camp Lejeune, North Carolina: a case-control study. *Environ Health*. Dec 4 2013;12:104. doi:10.1186/1476-069X-12-104

40. Antao VC, Larson TC, Horton DK. Libby vermiculite exposure and risk of developing asbestos-related lung and pleural diseases. *Curr Opin Pulm Med*. Mar 2012;18(2):161-7. doi:10.1097/MCP.0b013e32834e897d

41. McDonald JC, McDonald AD, Armstrong B, Sebastien P. Cohort study of mortality of vermiculite miners exposed to tremolite. *Br J Ind Med*. Jul 1986;43(7):436-44. doi:10.1136/oem.43.7.436

42. Krieger N, Zierler S, Hogan JW, et al. Geocoding and measurement of neighborhood socioeconomic position: a US perspective. *Neighborhoods and health*. 2003:147-178.

43. Schiewe J. Empirical studies on the visual perception of spatial patterns in choropleth maps. *KN-Journal of Cartography and Geographic Information*. 2019;69(3):217-228.

44. Richards TB, Berkowitz Z, Thomas CC, et al. Choropleth map design for cancer incidence, part 2. *Prev Chronic Dis*. Jan 2010;7(1):A24.

45. Kennedy S. The small number problem and the accuracy of spatial databases. *Accuracy of spatial databases*. 1989:187-196.

46. Brewer CA, Pickle L. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Ann Assoc Am Geogr*. 2002;92(4):662-681.

47. Cromley RG, Ye Y. Ogive-based legends for choropleth mapping. *Cartography and Geographic Information Science*. 2006;33(4):257-268.

48. Richards TB, Berkowitz Z, Thomas CC, et al. Choropleth map design for cancer incidence, part 1. *Prev Chronic Dis*. Jan 2010;7(1):A23.

49. Krieger N, Waterman P, Chen JT, Soobader M-J, Subramanian SV, Carson R. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and us census–defined geographic areas—the public health disparities geocoding project. *American journal of public health*. 2002;92(7):1100-1102.

50. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *Int J Health Geogr*. Dec 13 2006;5:58. doi:10.1186/1476-072X-5-58

51. Zhu L, Waller LA, Ma J. Spatial-temporal disease mapping of illicit drug abuse or dependence in the presence of misaligned ZIP codes. *GeoJournal*. Jun 1 2013;78(3):463-474. doi:10.1007/s10708-011-9429-3

52. Golden R, Schell JD. Using ZIP code and GIS studies to assess disease risk. *Environ Health Perspect*. Jan 2008;116(1):A18; author reply A18-9. doi:10.1289/ehp.10840

53. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International journal of health geographics*. 2006;5(1):1-15.

54. Rushton G, Armstrong MP, Gittler J, et al. *Geocoding health data: the use of geographic codes in cancer prevention and control, research and practice*. CRC Press; 2007.

55. Bureau USC. ZIP Code Tabulation Areas (ZCTAs). United State Census Bureau,. Updated August 26, 2020. Accessed January 22, 2021, 2021. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html

56. Beyer KM, Rushton G. Mapping cancer for community engagement. Research Support, N.I.H., Extramural

Research Support, Non-U.S. Gov't

Research Support, U.S. Gov't, P.H.S. *Prev Chronic Dis*. Jan 2009;6(1):A03.

57. ESRI. ESIR. Accessed 11/2/2020, 2020. https://www.esri.com/arcgis-blog/products/product/mapping/about-the-geometrical-interval-classification-method/

58. ESRI. How Hot Spot Analysis (Getis-Ord Gi*) works. ESRI. Accessed January 22, 2021, 2021. https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm

59. Gunarathna M, Kumari M, Nirmanee K. Evaluation of interpolation methods for mapping pH of groundwater. *International journal of latest technology in engineering, management & applied science*. 2016;3:1-5.

60. Duong Q, Hill CL, Jr., Janitz AE, Campbell JE. Trends in Lung and Bronchus, Prostate, Female Breast, and Colon and Rectum Cancers Incidence and Mortality in Oklahoma and the United States from 1999 to 2012. *J Okla State Med Assoc*. Jul-Aug 2016;109(7-8):347-353.

61. Rushton G. Public health, GIS, and spatial analytic tools. *Annu Rev Publ Health*. 2003;24:43-56. doi:10.1146/annurev.publhealth.24.012902.140843

62. Talbot TO, Kulldorff M, Forand SP, Haley VB. Evaluation of spatial filters to create smoothed maps of health data. *Stat Med*. Sep 15-30 2000;19(17-18):2399-408. doi:10.1002/1097-0258(20000915/30)19:17/18<2399::aid-sim577>3.0.co;2-r

Figures

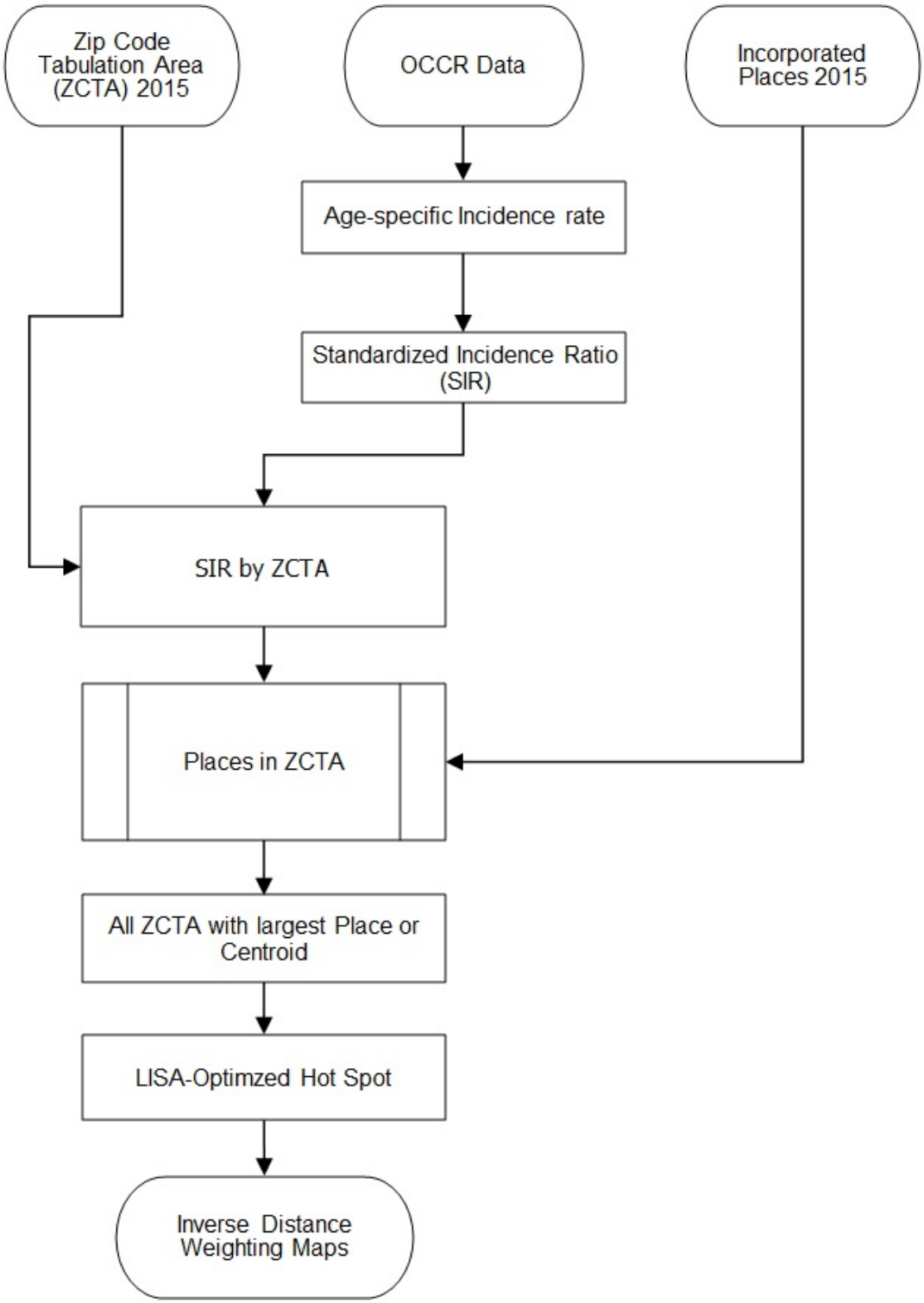

Figure 1

Data Flow Diagram for Registry Data.

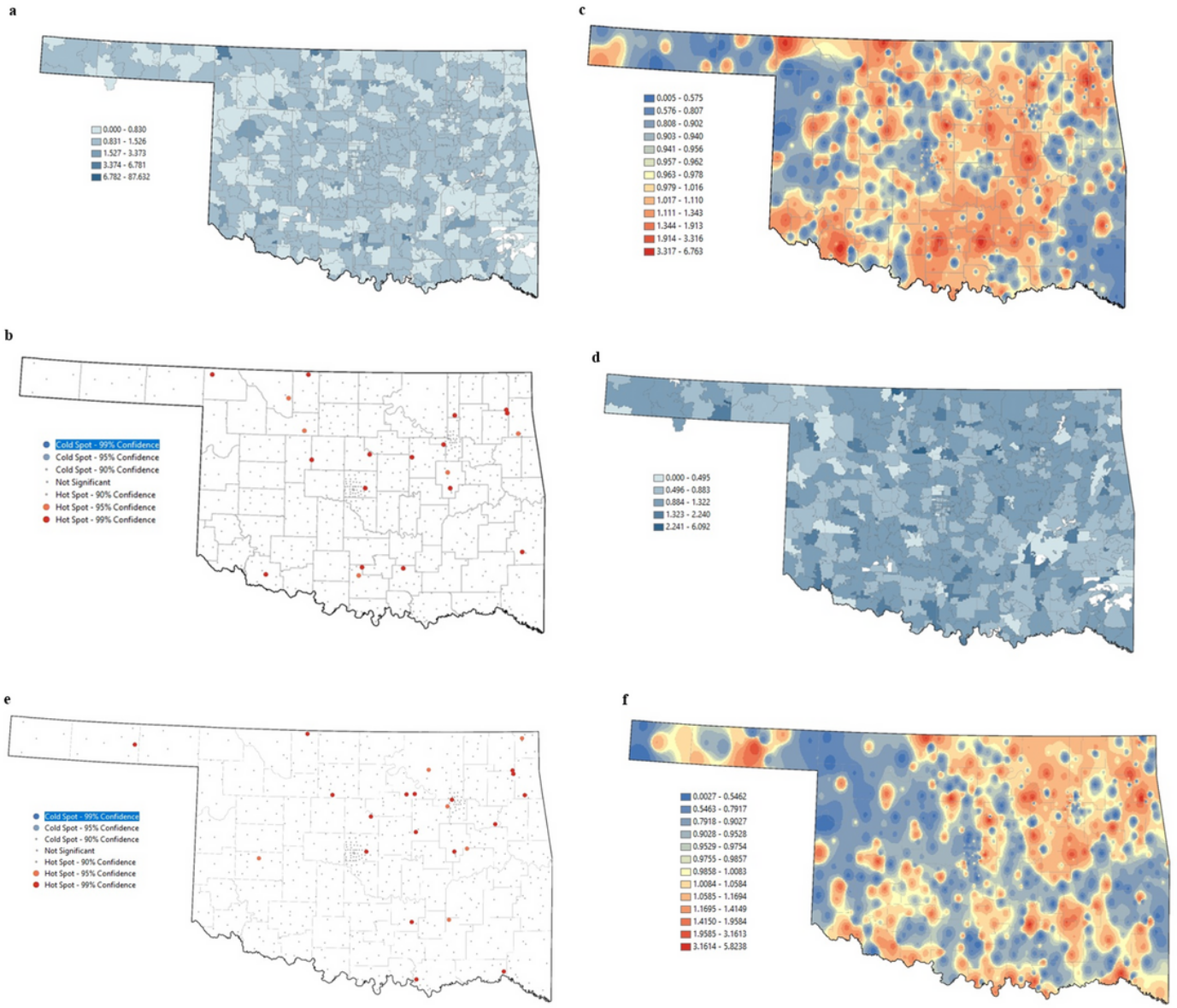

Figure 2

All cancers standardized incidence ratio for males a) by zip code tabulation areas , b) Getis-Ord Gi* hot and cold spots, c) inverse distance weighting (geometrical intervals) interpolated map; for females d) by zip code tabulation areas , e) Getis-Ord Gi* hot and cold spots, f) inverse distance weighting (geometrical intervals) interpolated map for females by zip code, Oklahoma 2013-2017. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

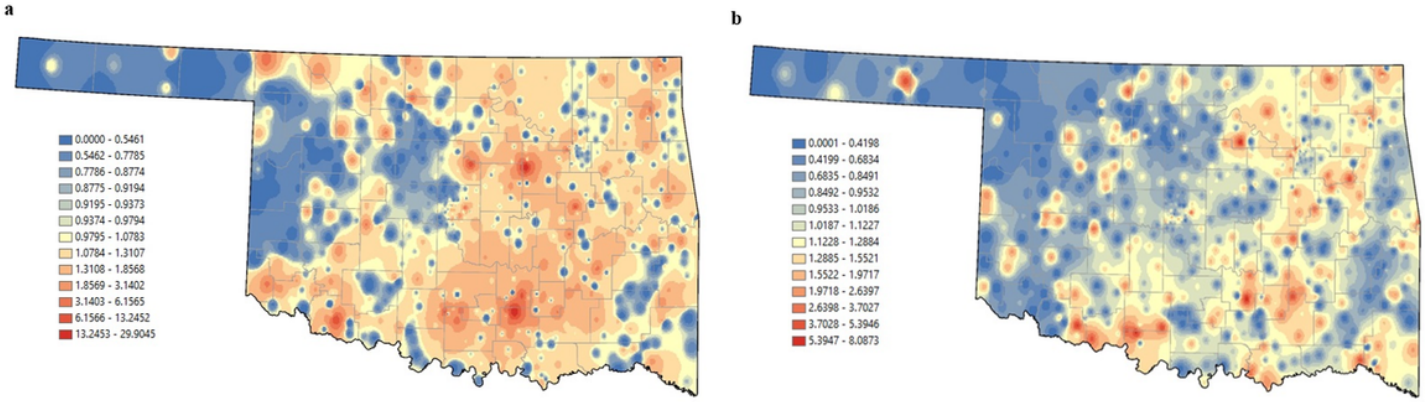

Figure 3

Lung cancers standardized incidence ratio Inverse distance weighting (geometrical intervals) interpolated map a) for males and b) for females Oklahoma 2013-2017. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

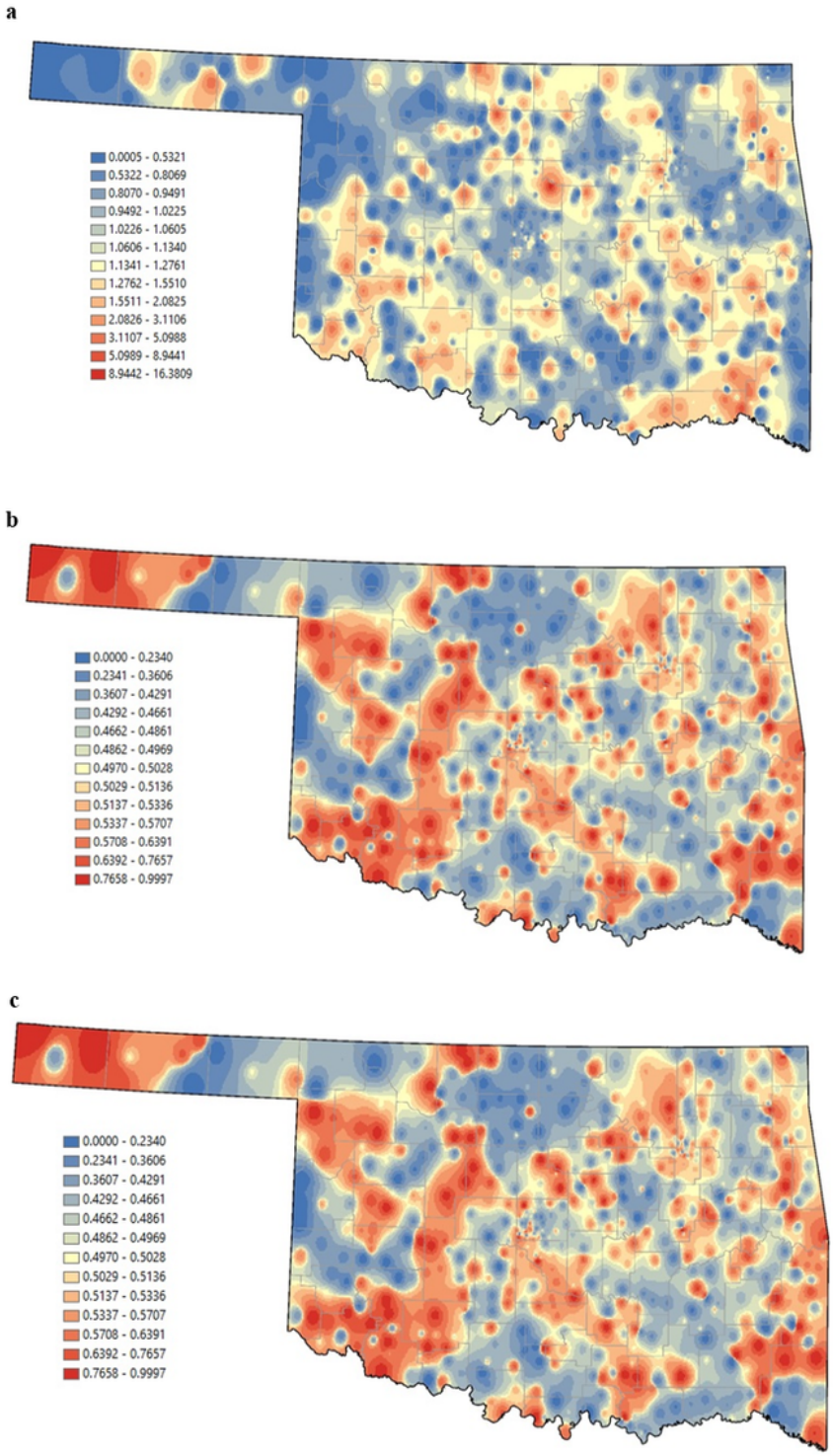

Figure 4

Colorectal cancers standardized incidence ratio Inverse distance weighting (geometrical intervals) interpolated map a) for males, b) for females, and c) male and female late stage Oklahoma 2013-2017. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any

country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

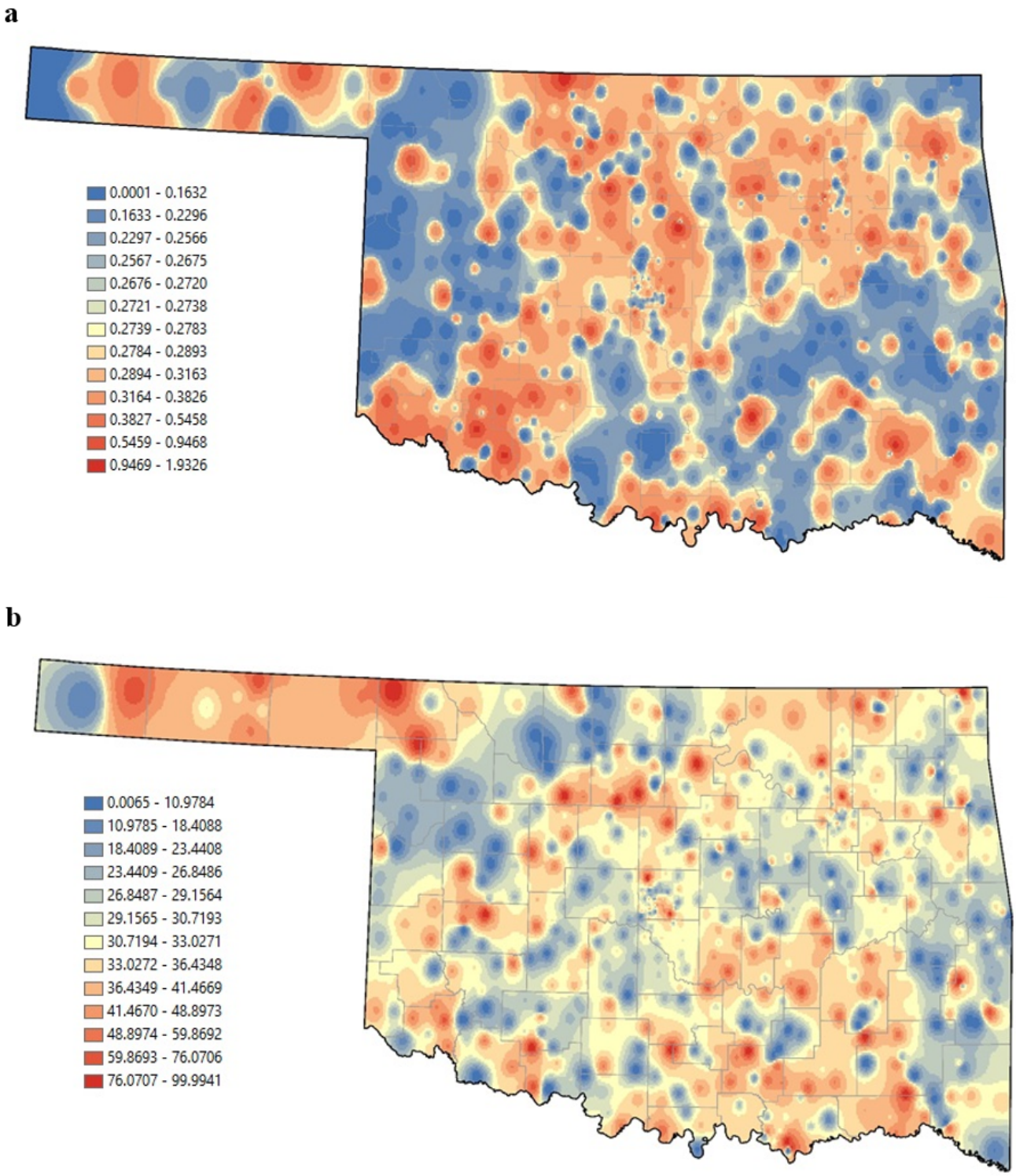

Figure 5

Female breast cancers standardized incidence ratio Inverse distance weighting (geometrical intervals) a) interpolated map and b) percent of late-stage standardized incidence ratio Inverse distance weighting (geometrical intervals) interpolated map by zip code, Oklahoma 2013-2017. Note: The designations

employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

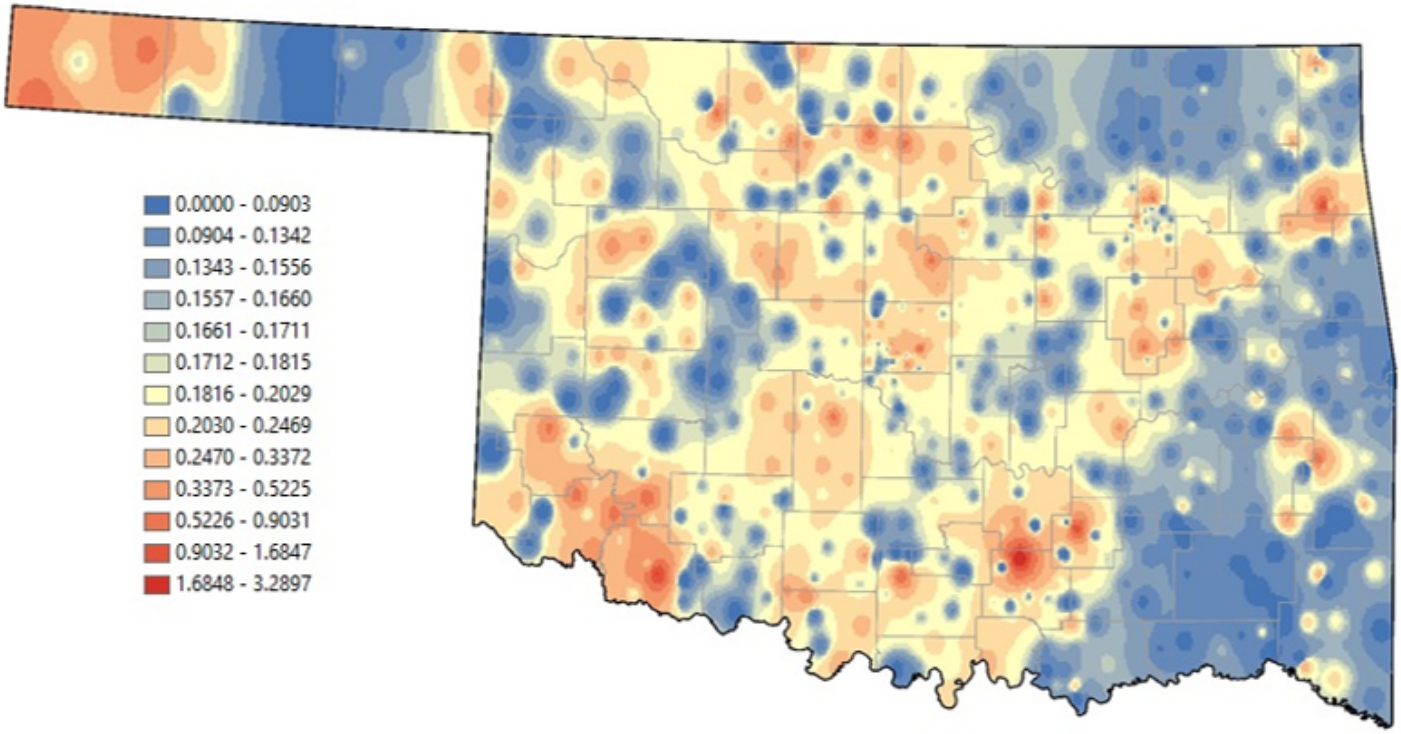

Figure 6

Prostate cancers standardized incidence ratio Inverse distance weighting (geometrical intervals) interpolated map for males by zip code, Oklahoma 2013-2017. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.