Geospatial analysis of multiple cancers in individuals in the US, 2004–2014

Lia C. Scott[1], **Tzy-Mey Kuo**[2], **Dora Il'yasova**[1], **Lee R. Mobley**[3]

[1]School of Public Health, Georgia State University, Atlanta, GA, USA

[2]Lineberger Cancer Center, University of North Carolina, Chapel Hill, NC, USA

[3]Mobley Consulting, LLC, Atlanta, GA, USA

Abstract

Background: There is a projected rapid increase in cancer survivors in the US population, from 15.5 million in 2016 to 26.1 million by 2040. Improvements in treatment and detection have led to increased survival, however, there is now a risk of developing new cancers as a result of environment toxins, behavioral risk factors, genetic predisposition, and late-term effects of radiation and chemotherapeutic treatments. This study takes a geospatial approach to examining the place of occurrence of multiple cancers originating in the population of four screenable cancers —female breast, colorectal, prostate, and cervical cancers—among the US population.

Methods: During 2004–2014, 6,523,532 primary cancer patients with one of these four screenable cancers were examined, and subsequent primary cancers (multiple cancers of any type) were noted. Individual level analyses estimated the odds of diagnosis with multiple cancers controlling for age, sex, and race-ethnicity. Change in effects on odds of multiple cancer diagnoses with age, sex, and race-ethnicity were evaluated controlling separately for late-stage diagnosis of the primary cancer or each primary cancer diagnosis type. County-level spatial cluster analysis was employed to identify and visualize higher than average multiple cancer rates.

Correspondence to: Lia C. Scott, PhD, MPH. Emory University School of Medicine, 550 Peachtree St NE, MOT-Suite 1075, Atlanta, GA 30308, USA. Lia.scott@emory.edu.

Results: Over half of the study population were female and almost 30% of the study population were diagnosed at late-stage for their first cancer. Multiple occurrences of all cancers increased during the time period for patients with initial breast or colorectal cancers. Among BC primary cancer cases, subsequent multiple cancers were mostly new breast cancers. By contrast, for CRC primary cancer cases, subsequent multiple cancers were about equally likely to be new CRC cases or other cancer types. Sex, age and race-ethnicity were all significantly associated with multiple cancers. In the model controlling for CRC as the primary type, the age and race-ethnicity effects were somewhat different than for all the other models. Thus, there was something distinctly different about the multiple cancer incidence among patients with CRC as their primary cancer as compared to patients with BC, CVC, or PC primaries. In subsequent analyses by county, there were distinct geospatial patterns in multiple cancer rates with most high-rate clusters occurring in the north- and mid-west US.

Conclusions: There were distinct individual level and geospatial disparities in multiple cancer diagnoses for the study population of all primary breast, colorectal, cervical, or prostate cancer patients during the decade studied. It is importance to emphasize continued screening for cancer survivors and research on personal and environmental drivers of multiple primary cancers.

Keywords

Spatial analysis; multiple primary neoplasms; cancer; National Program of Cancer Registries (NPCR)

Introduction

As of January 2019, there were an estimated 16.9 million cancer survivors in the United States (US), approximately 4.8% of the population, with a projected increase to 22.1 million by 2030 and to 26.1 million by 2040 (1,2). Factors driving the dramatic increase of the cancer survivor cohort include improvements in treatment and early detection, but also the aging of the population (2–4). One of the major concerns for clinicians, patients, and their families is the risk of developing new cancers, unrelated to the first (5,6). Therefore, the need for population- based monitoring and characterizing the distribution of multiple primary cancers has been widely recognized (2–7). There is a great need for up-to-date, population-based studies focused on characterizing risks of multiple primary cancers among individuals by age, sex, and cancer site of initial diagnoses. The small body of extant research has found that the most frequent diagnosis of secondary primary malignancies is among individuals initially diagnosed with prostate cancer (among men), breast cancer (among women) and colorectal cancer (among both sexes) (3). However, there is no information about the geospatial distribution of multiple primary cancer incidence or possible environmental predictors.

The National Cancer Institute defines a secondary cancer as cancer that has metastasized from the place where it first started and is the same type of cancer as the original, while a second metachronous primary (or multiple cancer) is a new primary cancer that occurs in a person who has had cancer in the past (8). The overall reported frequency of multiple primary cancers is between 2.4% and 17% (1,3), yet this topic remains important and not well studied (9,10). The issue of multiple primary cancers diagnosed in individuals over

time, or metachronously, is intimately connected to the increasing survival of cancer patients, as cancer treatment itself using chemotherapy or radiation can increase the risk of subsequent cancers, as can health behaviors such as diet, obesity, and smoking (11). In addition, genetic predisposition or environmental toxins may also increase the risk of multiple primary cancers (2,3,7). Information on these risk factors at the population level is limited. Geospatial analysis of clustering in these events is a first step towards identifying geographical disparities which may reflect underlying place-based conditions. Efforts to link information on patient reported outcomes with population-based cancer registry data to facilitate the surveillance of long-term and late effects are ongoing (12).

With such a large and growing population of cancer survivors in the US, there is increasing interest in improving cancer survivors' well-being and quality of life (11–13). Lifestyle factors such as diet, being overweight, physical activity, alcohol consumption, and tobacco use are important considerations for cancer survivors because they are known risk factors for cancer—including developing new primary cancers after cancer treatment (14). Of special concern is tobacco use, which varies among cancer survivors by cancer type with highest rates among bladder, lung, and ovarian survivors (15). Researcher found that 9.3% of cancer survivors continued to smoke up to 9 years after diagnosis (15). Younger cancer survivors in particular have been shown to have a higher prevalence of smoking after diagnosis than the general population. During 2008 to 2017, 31% of cancer survivors aged 18 to 44 years were current smokers, compared with 19% of the general population (16). Finally, the risk of developing a second primary cancer may be related to late-stage diagnosis of the primary cancer in Lymphoma patients (17). It is not known whether late-stage diagnosis of other cancers is significantly related to higher incidence of unrelated, multiple cancers but it provides a rationale for continued cancer screening among cancer survivors.

What can be done to reach the special needs of the cancer survivor population? This study focuses on a geospatial approach to examine occurrence of multiple cancers that manifested in the US population diagnosed with one of four primary and screenable cancers—female breast, prostate, cervix, or colon. From this population of more than 6 million cancer patients, about 1% developed multiple cancer(s) during a follow up period of 1–10 years. Prior to this study, it has been unknown whether the distribution of multiple cancer patients is spatially random, or clustered in some fashion that would enhance focused follow up and intervention policies to improve lifestyle choices.

Many years ago, Kerner and colleagues called for the use of spatial analysis to inform prevention strategy and policy implementation (18). While other methods look for epicenters of disease in a global pattern, the LISA methodology used here accounts for local spatial instabilities in overall patterns of global spatial association. This methodology is considered more reliable for inference in both the absence and presence of spatial autocorrelation, allowing for the identification of concentrations of both unusually high and low concentrated values (high and low clusters) and spatial outliers.

In this study, we examine a cohort of all primary cancer cases diagnosed during 2004–2014 as one of four 'screenable' types—female breast, cervix, colorectum, or prostate. These four cancer types represent about 34% of new cancer cases diagnosed in the U.S. in 2019 (14).

We examine the cohort during a period of follow up ranging from 1–10 years, accounting for all new primary cancers (multiple cancers) that manifested among survivors during 2004–2014. We then examine the national-level odds of diagnosis with multiple primaries by sex, age and race-ethnicity and conduct a geospatial analysis of the incidence rates of multiple cancers at the local county level. We present the following article in accordance with the STROBE reporting checklist (available at http://dx.doi.org/10.21037/ace-19-40).

Methods

We examined persons with primary cases of female breast, prostate, colorectal or cervical cancers diagnosed from 2004—2014 using the United States Cancer Statistics (USCS) database, which covered 99% of the U.S. population for this time period (19). While all states participate in the USCS database, five were excluded from the spatial cluster analysis because they did not provide county-level cancer data—Illinois, Kansas, Michigan, Minnesota, and Missouri—and Alaska and Hawaii were also excluded due to sparse cases. However, Illinois, Michigan, and Minnesota provided data without county indicators and included the entire cancer populations in their states. All states with complete cancer population data (n=46) were included in the logistic regression modeling. From this initial cohort of patients with one of the four primary cancers, multiple primaries were defined for those diagnosed with at least one other cancer of any type (as additional primary cancers) at any point in the subsequent time period, ending in 2014. Synchronous cancers were not included in the definition of multiple cancers. We used the protocols and definitions from a recent study to define our multiple cancer cases consistent with extant literature, which found that over all cancer survivors diagnosed during 1973–2005, about 8% had multiple cancer diagnoses (3).

Multiple cancers were then identified using 'Sequence Number Central' variable (I380_SeqNoCntrl). This code indicates the sequence of all reportable cancers (neoplasms) over the lifetime of a person and each primary case is assigned a different number. Progressions or recurrences of primary cancers are not assigned new numbers. Those coded with '00' indicate only one primary in the patient's time in the sample during 2004–2014. All subsequent numbers up to 59 identify which malignant primary the case report is, i.e., '02' is the second of two or more primaries. The combination of the two variables was used to identify multiple cancer occurrences for the study population.

At the county level, multiple cancer incidence rates were calculated as the count of cases with multiple cancers in each county divided by the adult county population in the year 2010, multiplied by 100,000. In calculating and visualizing the spatial patterns of the multiple cancer incidence rates, there were 2,438 counties from the 43 states included in the final analysis.

To indicate the primary cancer cases that were diagnosed at a late stage (defined as stage 3 or 4), we used the Collaborative Stage Data Collection System in the USCS database to derive SEER Summary Stage 2000 to describe our outcomes. To code for multiple cancers, we used the United States Cancer Statistics Restricted Access Dataset, Data Dictionary and Data Standards book (20). Primary cancer types were identified using the 'Primary Site'

variable (I400_Site), which is the code derived from the International Classification of Diseases for Oncology, Third Edition, or ICD-O-3.

Analyses were run at the individual level, producing national results based on 46 states, and at the county level, producing visualizations of local cancer incidence rate patterns across 43 states. At the individual level, unadjusted and adjusted logistic regression was employed to determine the odds of diagnosis with multiple cancers, controlling for age, sex, and race-ethnicity. An additional adjusted model was run to determine the odds of multiple cancer diagnosis, given late-stage diagnosis of the first primary cancer (pooling across all cancer types), while controlling for age, sex, and race-ethnicity. Separate adjusted models used primary cancer type as a control variable to evaluate whether any changes in the effects of age, race-ethnicity and sex were observed when controlling for each individual primary cancer diagnosis. Summary statistics were computed using SAS 9.4 (21).

Spatial analyses of county level multiple cancer rates were performed in GeoDa software (Open Source) (22) and results were mapped in QGIS (Open Source Geospatial Foundation Project) (23). Following previously established methods in applied policy studies (24–27), the global Moran's I and Local Indicators of Spatial Association (LISA) spatial clustering tests were performed using GeoDa software. The global Moran's I test determines if there is global clustering in the pattern of rates but cannot identify the location of the clusters, thus we use the LISA test to identify the local clusters. The LISA statistic is computed for each county, using all of the data across counties as a reference distribution to establish statistical significance of the relative size and degree of correlation with neighboring values for the county's observed cancer rate. The underlying assumption is of a Gaussian distribution of rates (approximately normal) which is appropriate for these rates. The resulting visualization using mapping is a description of the local patterns of association and is not predictive or confirmatory regarding why these patterns exist in their observed locations.

More specifically, the LISA statistic is computed using conditional permutation, or bootstrapping, that holds the value fixed for the county of interest and randomly permutes the remaining values to obtain a reference distribution for the correlation among the county's multiple cancer rate value and the group's (those assumed as comparators). The actual observed value with neighbors is then compared to the reference distribution to determine extremity of the value. Under the null hypothesis of spatial randomness, it would be unusual to observe a highly correlated value with the actual neighbors, which would then appear in the tail (rejection region) for the test. Thus, LISA statistics are relative to all of the observations of the variable of interest. In our analysis, clusters are determined statistically significant at the $P<0.05$ level. Using a series of LISA tests, one for each county, counties are classified as either non-significant or as falling into one of four categories (high-high, low-low, high-low, and low-high), relative to the mean multiple cancer incidence rate observed across all counties (28). For spatial clusters (high-high and low-low), the center of the cluster (index county used in the test) and neighbors are of interest (29). The maps show the center of the cluster in color (i.e., red for significant and higher than average surrounded by other counties higher than average), while the actual extent of the cluster includes the center and its surrounding neighbors. Neighbors are defined ex-ante by a queen weights matrix, then used in the cluster analysis to define the clusters. The neighbors are properly

included in the cluster, shown here as a grey buffer zone around the center. County lines are removed from all maps to meet data restriction standards.

Results

Sample statistics

This study included 6,523,532 persons with primary cases diagnosed during 2004–2014. Among the four cancer types, our cohort of primary cancer cases were distributed as follows: about 41% were primary breast cancer cases, 24% were colorectal cases, 2% were cervical and 34% were prostate. 54.5% of the study population were female, 35.9% were age 50–64, the largest of five age groups, and white (75.9%). About 29.7% were diagnosed at late stage, and <1% of the study population were diagnosed as synchronous multiple cancers (Table 1, Figure 1). These 65,008 individuals are mixed in among the primary cancer types (Table 1). Metachronous multiple cancers varied by primary cancer, ranging from 1.5% to 8.6% (Table 2), which is consistent with an earlier comprehensive study of cancer populations (3). The indicators for these metachronous cases define the dependent variable used in the logistic regressions.

When examined by cancer type and over time (Table 2), late-stage diagnoses were highest for colorectal and cervical cancers. Late-stage diagnoses increased over time for colorectal, cervical, and prostate cancers. Multiple occurrences of cancers in the cohort were highest for persons whose first cancer was breast, followed by colorectal cancers, which is consistent with earlier data over a longer horizon [for primary cancers of the breast or colon diagnosed and followed during 1973–2005, multiple cancers associated with an initial breast cancer had higher observed/expected ratios than multiple cancers associated with an initial colorectal cancer, across the age groups (1.17 *vs.* 1.06, respectively). However, the percentage of primary cancer cases with subsequent multiple cancers over this long horizon were similar (12.27% for breast, 12.44% for colorectal) (3)]. In our cohort using a much shorter but more recent timeframe (2004 to 2014), the percentage of multiple cancers increased over time for patients whose first cancer was breast or colorectal (Table 2). It has been noted that the reporting of multiple primary cancers may be influenced by significant changes in cancer risk factors and advances in diagnostic sensitivity and improved screening as time passes (1). Thus, the slight increases observed overall may be artifacts of improved screening and diagnostic practices over the decade. Individuals with primary BC were more likely to have BC diagnosed as a multiple cancer, while people with a primary CRC were about equally likely to have CRC or another cancer type as a multiple cancer (Table 2). Thus, there seems to be something different about the CRC survivors as compared to the BC survivors, suggesting a different etiology or exposures or behaviors.

Logistic regression

Table 3 shows results from adjusted multiple logistic regression models. The odds of multiple cancers given cancer type or late-stage diagnosis were examined, and effects of sex, age or race-ethnicity did not change substantially, with the exception of the CRC indicator model. In the late stage and colorectal regression models, females were less likely to be diagnosed with multiple primaries than males. Sex was not included in the female breast,

cervical and prostate models. Across all models, those age 65 and older were more likely to be diagnosed with multiple primaries compared to the referent group age 50–64, with age 75 and older having the highest odds. Those odds are attenuated in the colorectal cancer model to 1.32 times the odds of diagnoses compared to an average 2.43 times the odds in the late stage, female breast, cervical and prostate models. Those younger than 50 were less likely to be diagnosed with multiple primaries, with age less than 40 having the lowest odds. Those age 40–49 had slightly higher odds of diagnosis in the colorectal (OR 0.58) and prostate (OR 0.66) cancer models compared to the other models, averaging an OR of 0.25. Hispanic and Asian persons were less likely to be diagnosed with multiple primaries, while Black persons were more likely to be diagnosed compared to the White referent group. American Indian and Alaskan Native persons were more likely to be diagnosed with multiple primaries for all models except the colorectal model where they were less likely to be diagnosed with multiple cancers compared to White persons.

Spatial analysis

The average county-level rate of multiple cancers is a proportion (defined as the count of cases with multiple primaries/all cases) was 13.9 cases per 100,000 (SD 8.9). This rate ranged from 1.9 per 100,000 to 122.9 per 100,000. The Moran's I statistic for multiple cancer rates (I=0.19) was positive and statistically significant indicating evidence of clustering or significant spatial autocorrelation among the county values (P<0.0001). There were 113 high-rate (red-red) and 215 low-rate (blue blue) spatial cluster centers found for multiple cancer rates (Figure 2). Most high-rate spatial clusters were in the north- and mid-west Oregon, Montana, Wyoming, North and South Dakota, Nebraska, Iowa, with a couple of states in the northeast, West Virginia and Pennsylvania. The Southeast had mixtures of high-rate and low-rate clusters and the Northeast had the fewest coverage of clusters.

Figures 3,4 provide additional LISA cluster analyses to help gain insights. Figure 3 shows LISA clusters for incidence rates of each of the four cancers, mapped separately. Figure 4 provides LISA cluster analysis for the late-stage diagnosis rates by county, across all 4 cancer types, based on late-stage diagnosis of the primary cancers in the study cohort.

Conclusions

This population-based study found that sex, age, and race-ethnicity were significantly associated with subsequent diagnoses of multiple cancers. The oldest age group had the highest odds of diagnosis as expected since older age is a risk factor for cancer. Black and American Indian/Alaskan Native cases had consistently higher odds of diagnosis with multiple primaries.

These findings held even after adjusting for primary cancer type and late-stage diagnosis, with the exception of colorectal cancer. Initial diagnosis with colorectal cancer showed lower odds of diagnosis with multiple cancers in those aged 65 and older and in American Indian and Alaska Native groups compared to the effects in all other models This may warrant a need to further study the link between initial primary cancer type and its association with multiple primary cancer diagnoses

Further research is necessary to determine what may be driving observed disparities. The literature demonstrates that there are a multitude of factors that contribute to multiple primaries including genetic susceptibility and familial cancer syndromes, environmental and lifestyle exposures (e.g., tobacco, alcohol use), continued adherence to regular cancer screening, hormonal factors, immune deficiency and infection, carcinogenic effects of prior cancer treatments, and the interaction of these factors (1,30). Areas that have higher rates of multiple cancers may have a higher prevalence of risk factors that can be further explored. For example, North and South Dakota had high-rate spatial clusters, and those states also have tobacco use and excessive drinking rates higher than the national average (31).

Beyond individual level differences, there were distinct geospatial disparities in county level rates of multiple cancers. Based on Figure 2, high-rate (red-red) spatial clusters appeared to occur in more rural areas. There appear to be many high-rate (red-red) spatial clusters in the Midwest and Northwest. As a comparison, we examine LISA clusters in the simple cancer incidence rates for the four types of cancer, displayed in four maps in Figure 3. High-rate cluster areas of multiple cancers (Figure 2) seem to coincided roughly with places where high-rate clusters of colorectal and prostate cancer incidence are found (Figure 3), and places where there are high-rate clusters of late-stage diagnosis (Figure 4). Given that the occurrence of higher than average multiple cancer rates are not spatially random, this may suggest areas to target additional screening and cancer control planning and messaging to help prevent multiple cancer occurrences. Late-stage diagnosis can be addressed by encouraging screening and prevention messages of these four cancers.

Limitations

Multiple primary tumors can be classified as synchronous (2 or more primary tumors identified in the same patient and at the same time) or metachronous (new primary tumors unrelated to the patients' previous cancers) (Figure 1) (32,33). In our data there were <1% synchronous cases (Table 1) identified as individuals with more than one primary cancer case defined at the same time. We do not model these individuals with separate indicators and they are not counted among the metachronous cases used to define the dependent variable in the logistic regressions. There are more studies of metachronous than of synchronous primary cancers, as these metachronous cancers are thought to be more related to cancer treatment. Because synchronous primary cancers are diagnosed almost simultaneously, their rates may indicate a greater level of cancer predisposition in populations at certain locations. This predisposition can be either genetic or arise from gene-environment interaction. Although gene-environment interaction has been hypothesized to play a role in the etiology of multiple tumors (33), there are no data examining this hypothesis. While beyond the scope of the current paper, a distinction between these two types of multiple cancers may warrant further analyses. It is likely that the synchronous type may be more related to behavioral, genetic, or environmental carcinogens and exhibit different local spatial patterns than the metachronous type. It is beyond the scope of this study to distinguish between these two types of multiple cancers.

The initial adjusted model pooled observations and included all cancer types as separate indicators, omitting a reference group to avoid perfect collinearity—however the model

would not converge. We believe there is confounding caused by the 65,008 (<1% of cases, Table 1) synchronous multiple cancer cases which would cause overlap in the primary cancer type indicators where this phenomenon occurred. That is, no individuals were counted twice, but an individual with synchronous multiple cancer would be coded as having more than one primary cancer type. This would effectively create an interaction effect in the primary cancer indicators for a very small subset of the data and result in failure to converge in estimation. To keep these individuals in the sample, the models were then parsed out and run separately for each primary cancer type. This allowed us to observe if there were changes in effects of age, sex, and race-ethnicity when we controlled for each primary cancer type. The difference in the effects of age and race-ethnicity observed for the CRC model as compared to all other models (where effects were very consistently similar) suggests that the primary CRC cases are different than others. This is reflected in the profile of types of multiple cancers in Table 2, and is an interesting finding meriting further investigation.

This study is also limited in the sense that it only examines multiple cancers from a baseline group of persons with primary screenable cancers of four types, and is not based on all occurrences of cancer. Further, the reporting of multiple primary cancers may be influenced by significant changes in cancer risk factors and advances in diagnostic sensitivity and improved screening (1). Finally, this descriptive study examined univariate spatial clustering but did not examine geospatial associations with contextual variables that could perhaps further elucidate drivers of these geospatial disparities. Such context may be informative in distinguishing among synchronous or metachronous multiple cancer clusters, for example. Univariate spatial clustering was examined using crude rates rather than age-adjusted rates. Although age was considered in the multivariate modeling, the differences in age-distribution at the county level could potentially influence the clustering effect. While geographic differences could potentially reflect age structure differences rather than cancer rates, the stable rates across the individual analysis and multiple permutations of the spatial analysis may give us more robust results that are not influenced by age. Future studies could examine if there are differences in clustering give crude or age-adjusted rates for each individual outcome. Given the findings of this study, it is important that work continues to explore both personal and environmental drivers of multiple primary cancers, particularly among screenable cancers. Additionally, it is important to emphasize screening in cancer survivorship care or wellness plans to detect potential additional cancers after treatment.

Acknowledgments

References

1. Copur MS, Manapuram S. Multiple Primary Tumors Over a Lifetime. Oncology (Williston Park) 2019;33:629384. [PubMed: 31365752]

2. Bluethmann SM, Mariotto AB, Rowland JH. Anticipating the "Silver Tsunami": Prevalence Trajectories and Comorbidity Burden among Older Cancer Survivors in the United States. Cancer Epidemiol Biomarkers Prev 2016;25:1029–36. [PubMed: 27371756]

3. American Cancer Society. Cancer Facts & Figures 2009. Atlanta: American Cancer Society, 2009. Available online: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2009.html

4. Mariotto AB, Rowland JH, Ries LA, et al. Multiple cancer prevalence: a growing challenge in long-term survivorship. Cancer Epidemiol Biomarkers Prev 2007;16:566–71. [PubMed: 17372253]

5. Travis LB. The epidemiology of second primary cancers. Cancer Epidemiol Biomarkers Prev 2006;15:2020–6. [PubMed: 17057028]

6. Ng AK, Travis LB. Subsequent malignant neoplasms in cancer survivors. Cancer J 2008;14:429–34. [PubMed: 19060610]

7. Miller KD, Siegel RL, Lin CC, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016;66:271–89. [PubMed: 27253694]

8. National Cancer Institute. NCI Dictionary of Cancer Terms 2019. Available online: https://www.cancer.gov/publications/dictionaries/cancer-terms

9. Coyte A, Morrison DS, McLoone P. Second primary cancer risk - the impact of applying different definitions of multiple primaries: results from a retrospective population-based cancer registry study. BMC Cancer 2014;14:272. [PubMed: 24742063]

10. Weir HK, Johnson CJ, Thompson TD. The effect of multiple primary rules on population-based cancel survival. Cancer Causes Control 2013;24:1231–42. [PubMed: 23558444]

11. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019;69:363–85. [PubMed: 31184787]

12. Smith TG, Castro KM, Troeschel AN, et al. The rationale for patient-reported outcomes surveillance in cancer and a reproducible method for achieving it. Cancer 2016;122:344–51. [PubMed: 26619031]

13. Demark-Wahnefried W, Rogers LQ, Alfano CM, et al. Practical clinical interventions for diet, physical activity, and weight control in cancer survivors. CA Cancer J Clin 2015;65:167–89. [PubMed: 25683894]

14. American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society, 2019.

15. Westmaas JL, Alcaraz KI, Berg CJ, et al. Prevalence and correlates of smoking and cessation-related behavior among survivors of ten cancers: findings from a nationwide survey nine years after diagnosis. Cancer Epidemiol Biomarkers Prev 2014;23:1783–92. [PubMed: 25100826]

16. National Cancer Institute. Cancer Trends Progress Report. Bethesda, MD: National Cancer Institute, National Institutes of Health, Department of Health and Human Services, 2019. Available online: https://progressreport.cancer.gov/sites/default/files/archive/report2019.pdf

17. Major A, Smith DE, Ghosh D, et al. Risk and subtypes of secondary primary malignancies in diffuse large B-cell lymphoma survivors change over time based on stage at diagnosis. Cancer 2020;126:189–201. [PubMed: 31509235]

18. Kerner JF, Andrews H, Zauber A, et al. Geographically-based cancer control: methods for targeting and evaluating the impact of screening interventions on defined populations. J Clin Epidemiol 1988;41:543–53. [PubMed: 3385456]

19. National Program of Cancer Registries and Surveillance, Epidemiology, and End Results SEER*Stat Database. United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2018, based on November 2017 submissions. Available online: https://seer.cancer.gov/seerstat/

20. Centers for Disease Control and Prevention and National Cancer Institute. United States Cancer Statistics Restricted Access Data Set, Data Dictionary and Data Standards, 2015 November Data Submission 2016. Available online: https://www.cdc.gov/rdc/data/B1/USCS_RDC_DataDictionary_2016.pdf
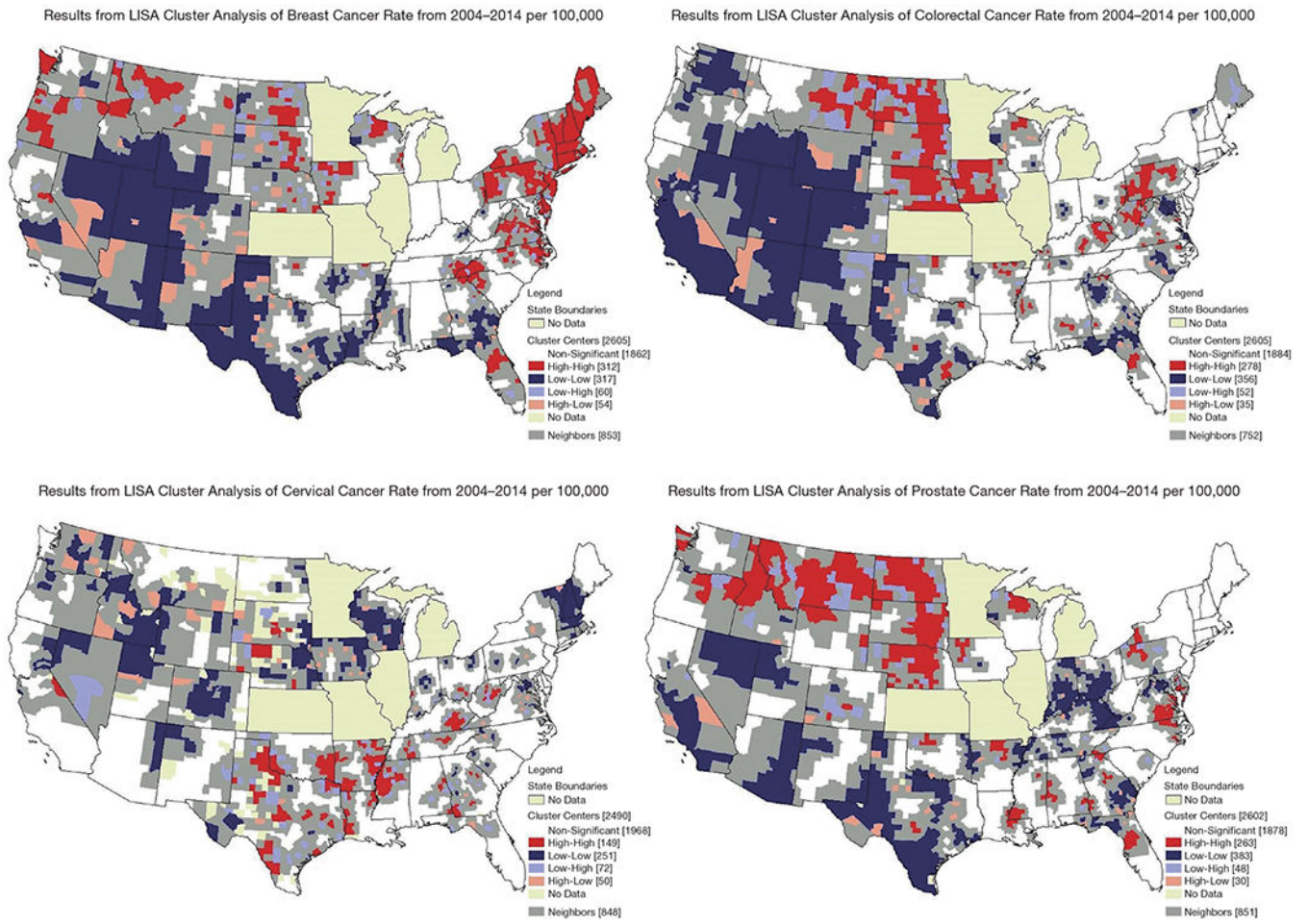
21. SAS Institute Inc., Cary, NC, USA. Available online: https://www.sas.com/

22. Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to Spatial Data Analysis. Geogr Anal 2016;38:5–22.

23. QGIS Development Team (2017). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available online: http://qgis.osgeo.org

24. Scott L, Mobley LR, Il'yasova D. Geospatial Analysis of Inflammatory Breast Cancer and Associated Community Characteristics in the United States. Int J Environ Res Public Health 2017;14:404.

25. Schieb LJ, Mobley LR, George M, et al. Tracking stroke hospitalization clusters over time and associations with county-level socioeconomic and healthcare characteristics. Stroke 2013;44:146–52. [PubMed: 23192758]

26. Martinez AN, Mobley LR, Lorvick J, et al. Spatial analysis of HIV positive injection drug users in San Francisco, 1987 to 2005. Int J Environ Res Public Health 2014;11:3937–55. [PubMed: 24722543]

27. Bray JW, Depro B, McMahon D, et al. Disconnected geography: a spatial analysis of disconnected youth in the United States. Int J Environ Res Public Health 2016;37:317–42.

28. Anselin L Local indicators of spatial association—LISA. Geogr Anal 1995;27:93–115.

29. Anselin L Exploring spatial data with GeoDaTM: a workbook. Center for spatially integrated social science. Available online: https://www.geos.ed.ac.uk/~gisteac/fspat/geodaworkbook.pdf

30. Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2018.

31. Vogt A, Schmid S, Heinimann K, et al. Multiple primary tumours: challenges and approaches, a review. ESMO Open 2017;2:e000172. [PubMed: 28761745]

32. Howe HL, editor A review of the definition for multiple primary cancers in the United States. Springfield: North American Association of Central Cancer Registries, 2002.

33. Travis LB, Wahnefried WD, Allan JM, et al. Aetiology, genetics and prevention of secondary neoplasms in adult cancer survivors. Nat Rev Clin Oncol 2013 ;10:289. [PubMed: 23529000]
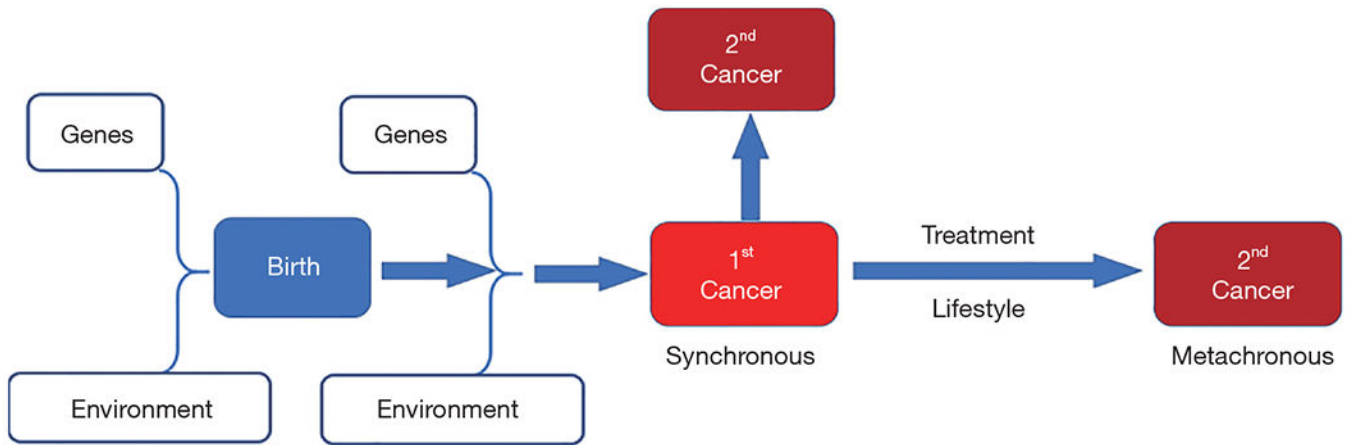
Figure 1.
Conceptual framework of multiple primary cancers etiology: hypothesized involvement of genetic, environmental, and lifestyle risk factors in etiology of synchronous cancers; and the role of treatment and lifestyle in etiology of metachronous cancers.
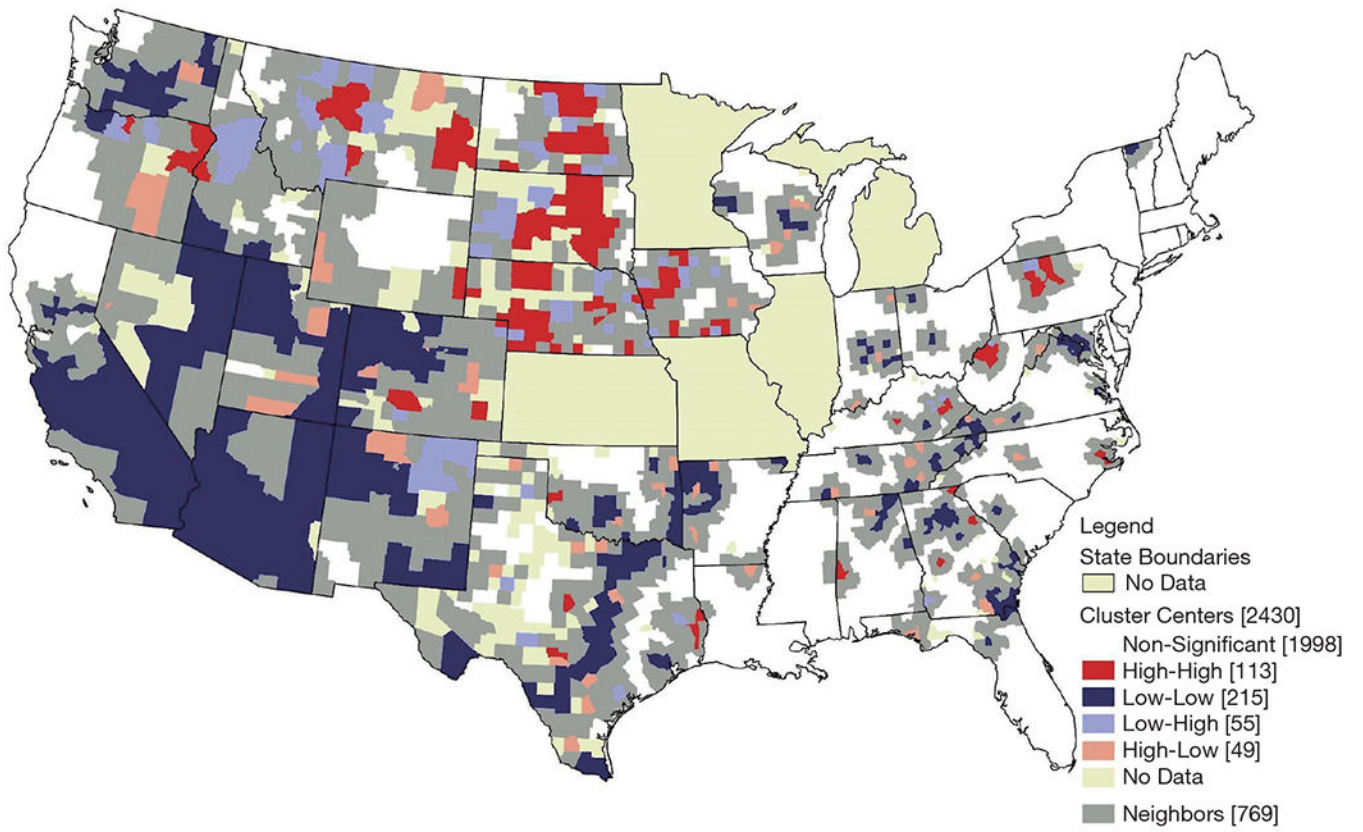
Figure 2.
LISA map of county-level multiple primary cancer rates per 100,000, 2004–2014.

Results from LISA Cluster Analysis of Breast Cancer Rate from 2004–2014 per 100,000

Results from LISA Cluster Analysis of Colorectal Cancer Rate from 2004–2014 per 100,000

Results from LISA Cluster Analysis of Cervical Cancer Rate from 2004–2014 per 100,000

Results from LISA Cluster Analysis of Prostate Cancer Rate from 2004–2014 per 100,000

Figure 3.
Map panel of LISA cluster results for female breast, cervical, colorectal, and prostate cancer incidence rates per 100,000 from 2004–2014.
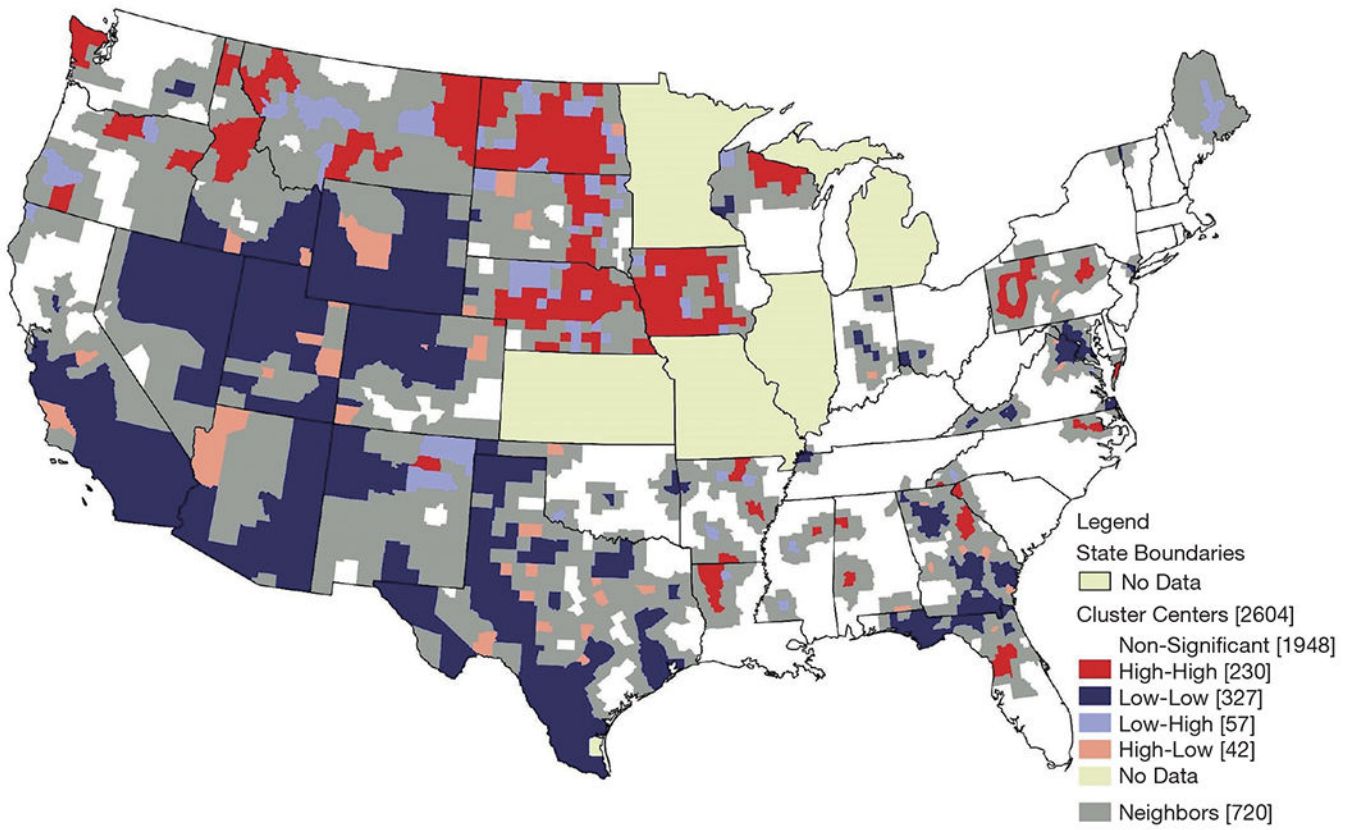
Figure 4.
LISA map of county-level late-stage cancer rates per 100,000, 2004–2014

Table 1

Descriptive statistics of the cohort of 4 primary, screenable cancers population, 2004–2014

Variable	Cases (n=6,523,532)	Percent (%)
Sex		
Female	3,552,841	54.46
Male	2,970,691	45.54
Age, years		
<40	193,157	2.96
40–49	660,668	10.13
50–64	2,342,475	35.91
65–74	1,789,766	27.44
75+	1,537,466	23.57
Race-ethnicity		
Hispanic	472,591	7.24
American Indian/Alaskan Native	29,030	0.45
Asian	174,393	2.67
Black	807,985	12.39
Other	91,604	1.4
White	4,947,929	75.85
Late stage		
No	4,235,499	64.93
Yes	1,935,981	29.68
Missing stage	352,052	5.4
Primary cancer type		
Female breast	2,703,357	41.44
Colorectal	1,543,108	23.65
Cervical	132,662	2.03
Prostate	2,209,413	33.87
Synchronous multiple primaries [*]	65,008	<1%

[*], the total number of primary cancer cases by type exceeds the total sample size by 65,008 individuals which appear in more than one of the primary cancer type tabulations.

Table 2

Percent of late-stage diagnoses by cancer type and time period for the study population, a cohort with one of four screenable cancers

Cancer type	Time period	Mean, late stage* (%)	Range, late stage (%)	Mean, distant stage** (%)	Range, distant stage (%)	Multiple primaries (%)	Multiple primaries same type (%)	Multiple primaries different type (%)
Breast	2004–2009	30	24–36	5	3–6	5.4	4.2	1.2
	2010–2014	29	22–33	5	3–6	8.0	7.0	1.0
Colorectal	2004–2009	53	48–59	19	17–21	7.2	3.4	3.8
	2010–2014	57	52–63	22	18–24	8.6	4.3	4.3
Cervical	2004–2009	52	40–56	13	6–19	2.8		
	2010–2014	54	33–59	16	10–23	2.7		
Prostate	2004–2009	15	10–21	4	3–6	1.8		
	2010–2014	18	13–25	6	4–8	1.5		

*
, late stage is defined as SEER summary stage 3 and 4;

**
, distant stage is defined as SEER summary stage 4.

Table 3

Results from adjusted multiple logistic regression models on multiple cancer study population (n=6,523,532)

Variable	Late-stage model			Female breast model			Colorectal model			Cervical model			Prostate model		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Sex															
Female	0.81	0.80, 0.82	<0.01	–	–	–	0.93	0.92, 0.95	<0.01	–	–	–	–	–	–
Male	REF						REF								
Age, years															
<40	0.27	0.24, 0.30	<0.01	0.25	0.22, 0.27	<0.01	0.25	0.23, 0.28	<0.01	0.17	0.16, 0.19	<0.01	0.36	0.32, 0.40	<0.01
40–49	0.54	0.52, 0.57	<0.01	0.48	0.46, 0.50	<0.01	0.58	0.56, 0.61	<0.01	0.47	0.45, 0.49	<0.01	0.66	0.63, 0.68	<0.01
65–74	1.83	1.79, 1.87	<0.01	1.88	1.84, 1.92	<0.01	1.70	1.66, 1.73	<0.01	1.87	1.83, 1.91	<0.01	1.70	1.67, 1.74	<0.01
75+	2.47	2.42, 2.52	<0.01	2.41	2.36, 2.46	<0.01	1.32	1.29, 1.35	<0.01	2.39	2.34, 2.44	<0.01	2.44	2.39, 2.49	<0.01
50–64	REF			REF			REF			REF			REF		
Race-ethnicity															
Hispanic	0.94	0.91, 0.97	<0.01	0.96	0.93, 1.00	<0.01	0.89	0.86, 0.92	<0.01	0.91	0.88, 0.94	<0.01	0.94	0.91, 0.97	<0.01
AI/AN	1.15	1.02, 1.29	<0.01	1.15	1.02, 1.29	<0.01	0.96	0.85, 1.08	<0.01	1.10	0.98, 1.24	<0.01	1.15	1.02, 1.29	<0.01
Asian	0.85	0.80, 0.90	<0.01	0.84	0.79, 0.89	<0.01	0.72	0.68, 0.76	<0.01	0.81	0.77, 0.86	<0.01	0.89	0.84, 0.94	<0.01
Black	1.34	1.31, 1.38	<0.01	1.40	1.37, 1.43	<0.01	1.38	1.35, 1.41	<0.01	1.33	1.30, 1.36	<0.01	1.29	1.27, 1.32	<0.01
Other	0.15	0.13, 0.19	<0.01	0.14	0.11, 0.16	<0.01	0.22	0.18, 0.26	<0.01	0.13	0.11, 0.15	<0.01	0.10	0.08, 0.12	<0.01
White	REF			REF			REF			REF			REF		
Cancer type															
Late stage	1.32	1.30, 1.35	<0.01	–	–	–	–	–	–	–	–	–	–	–	–