

# Community Health Behaviors and Geographic Variation in Early-Onset Colorectal Cancer Survival Among Women

Andreana N. Holowatyj, PhD, MS<sup>1,2,3</sup>, Marvin E. Langston, PhD, MPH<sup>4</sup>, Yunan Han, MD<sup>5</sup>, Richard Viskochil, MS, PhD<sup>3,6</sup>, Jose Perea, MD, PhD<sup>7</sup>, Yin Cao, MPH, ScD<sup>5,8</sup>, Charles R. Rogers, PhD, MPH, MS<sup>9</sup>, Christopher H. Lieu, MD<sup>10</sup> and Justin X. Moore, PhD, MPH<sup>11</sup>

**INTRODUCTION:** Despite overall reductions in colorectal cancer (CRC) morbidity and mortality, survival disparities by sex persist among young patients (age <50 years). Our study sought to quantify variance in early-onset CRC survival accounted for by individual/community-level characteristics among a population-based cohort of US women.

**METHODS:** Geographic hot spots—counties with high early-onset CRC mortality rates among women—were derived using 3 geospatial autocorrelation approaches with Centers for Disease Control and Prevention national mortality data. We identified women (age: 15–49 years) diagnosed with CRC from 1999 to 2016 in the National Institutes of Health/National Cancer Institute's Surveillance, Epidemiology, and End Results program. Patterns of community health behaviors by hot spot classification were assessed by Spearman correlation ( $\rho$ ). Generalized  $R^2$  values were used to evaluate variance in survival attributed to individual/community-level features.

**RESULTS:** Approximately 1 in every 16 contiguous US counties identified as hot spots (191 of 3,108), and 52.9% of hot spot counties ( $n = 101$ ) were located in the South. Among 28,790 women with early-onset CRC, 13.7% of cases ( $n = 3,954$ ) resided in hot spot counties. Physical inactivity and fertility were community health behaviors that modestly correlated with hot spot residence among women with early-onset CRC ( $\rho = 0.21$  and  $\rho = -0.23$ , respectively;  $P < 0.01$ ). Together, individual/community-level features accounted for distinct variance patterns in early-onset CRC survival among women (hot spot counties: 33.8%; non-hot spot counties: 34.1%).

**DISCUSSION:** Individual/community-level features accounted for approximately one-third of variation in early-onset CRC survival among women and differed between hot spot vs non-hot spot counties. Understanding the impact of community health behaviors—particularly in regions with high early-onset CRC mortality rates—is critical for tailoring strategies to reduce early-onset CRC disparities.

**SUPPLEMENTARY MATERIAL** accompanies this paper at <http://links.lww.com/CTG/A446>; <http://links.lww.com/CTG/A447>; and <http://links.lww.com/CTG/A448>.

*Clinical and Translational Gastroenterology* 2020;11:e00266. <https://doi.org/10.14309/ctg.000000000000266>

## INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the third leading cause of cancer deaths among women and men in the United States, with an estimated 147,950 new cases and 53,200 deaths in 2020 (1). Despite reductions in the overall CRC burden, incidence rates are rising at an alarming rate with unknown etiologies for men and women younger than 50 years

(2–4). Initial studies support the role of health behaviors in early-onset colorectal carcinogenesis—as obesity, physical inactivity, and sedentary behaviors have been linked to increased early-onset CRC risk among women (5–7). However, the role of health behaviors on CRC outcomes among young women remains unknown. Given that sex disparities in survival among patients with early-onset CRC have been previously reported (3), a

<sup>1</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA; <sup>2</sup>Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, USA; <sup>3</sup>Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, Utah, USA; <sup>4</sup>Division of Research, Kaiser Permanente Northern California, Oakland, California, USA; <sup>5</sup>Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri, USA; <sup>6</sup>Huntsman Cancer Institute, Salt Lake City, Utah, USA; <sup>7</sup>Fundacion Jimenez Diaz University Hospital, Madrid, Spain; <sup>8</sup>Alvin J. Siteman Cancer Center, St. Louis, Missouri, USA; <sup>9</sup>Department of Family and Preventive Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA; <sup>10</sup>Division of Medical Oncology, University of Colorado Denver, Aurora, Colorado, USA; <sup>11</sup>Division of Epidemiology, Augusta University at the Medical College of Georgia, Augusta, Georgia, USA. **Correspondence:** Andreana N. Holowatyj, PhD, MS. E-mail: [andreana.holowatyj@vumc.org](mailto:andreana.holowatyj@vumc.org).

**Received April 7, 2020; accepted October 12, 2020; published online November 23, 2020**

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of The American College of Gastroenterology

better understanding of sex-specific differences in colorectal carcinogenesis—particularly among young patients—is needed to optimize detection and treatment strategies.

Sex differences in disease prevalence and outcomes persist among all-comers diagnosed with CRC. Women have a higher incidence of right-sided CRCs (8,9), and women with right-sided tumors tend to exhibit a higher rate of comorbidities (9). Lieberman et al. observed an association between proximal large polyp risk and female sex (10), although the prevalence of polyps among young patients with CRC remains unknown. Importantly, sex differences in the performance of fecal occult blood testing and colonoscopies have been reported by sex (11,12) because colonoscopy is considered to be a more difficult procedure among women. Previous reports have indicated that approximately one-third of colonoscopies performed in women were considered technically difficult vs only 1 in every 6 cases among men (12). As differences in colon length, body fatness, energy balance, and secretion of hormones in proportion to fat also persist by sex (12,13), these findings suggest that differences in tumor biology and individual health behaviors may be associated with sex-specific disparities in CRC outcomes—particularly among young patients.

Given the impact of sex-specific differences on CRC presentation and prognosis (1,3), we hypothesize that biological-, individual-, and community-level factors may be contributing to sex-specific disparities in disease-specific outcomes, as well as distinct patterns in geographic variation of early-onset CRC survival by sex. The purpose of our population-based cohort study was to define county-level early-onset CRC mortality hot spots among women using geospatial methodology and to determine the contribution of individual- and community-level characteristics on geographic variation in CRC-specific survival among women diagnosed with a first primary invasive CRC aged 15–49 years—using nationally representative data from the Centers for Disease Control and Prevention (CDC) and National Institutes of Health/National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program.

## METHODS

### Early-onset CRC hot spots

Data were obtained from the CDC underlying causes of death file which collects national mortality data for every US resident (14), and from the National Institutes of Health/National Cancer Institute’s SEER Program, which collects cancer incidence/mortality data from 18 population-based (Figure 1) cancer registries covering approximately 28% of the US population (15) (Figure 1). Early-onset CRCs were defined as cancers of the colon and rectum diagnosed among women aged 15–49 years. National mortality data from the CDC between years 1980 and 2018 within the contiguous United States were used to examine early-onset CRC mortality (on the basis of available death certificate information) among women aged 15–54 years (to identify women with early-onset CRC [age <50 years] with standardized 5-year follow-up) by county-level estimates using geospatial methodology as we have previously described (16).

Geographic hot spots—counties of high mortality rates for early-onset CRC—were derived using 3 geospatial methods. First, we estimated the smoothed Empirical Bayes (EB) early-onset CRC mortality rates using the proportion of women with deaths attributed to CRC, with smoothing performed using the EB tools in GeoDa1.6.7.9 (<http://geodacenter.asu.edu>) (17).

Briefly, the smoothed EB rate method allows for more stable estimation of CRC mortality rates by accounting for the overall county population. EB smoothing includes the variance of each county by using the corresponding county population, and thus, counties with higher populations have smaller variance while those with smaller populations have larger variance. Furthermore, the EB rates are smoothed using an inverse function of the variance for shrinkage toward the overall mean for early-onset CRC mortality rates (e.g., counties with larger populations and smaller variance gave higher weights toward the observed weights). We categorized the EB smoothed early-onset CRC mortality rates into quintiles, and we defined counties as high-risk if the smoothed EB early-onset CRC mortality rates for women were in the fifth quintile. Second, we used Local Indicators of Spatial Association to measure similarity of early-onset CRC mortality between counties and calculate values both within and across geographic boundaries, additionally identifying spatial outliers (17–19). For each US county, we estimated Local Moran’s I Statistic values, using associated z-scores and *P* values to assess the magnitude of spatial autocorrelation and statistical significance, respectively (17). Statistically significant positive z-scores indicated counties surrounded by areas with similar early-onset CRC mortality rates for women—either similarly high or similarly low (positive spatial autocorrelation) (17). Finally, we used the Getis-Ord  $G_i^*$  statistic to identify areas where early-onset CRC mortality rates with either high or low values clustered within the context of the neighboring county (17,20–22). To estimate the  $G_i^*$  statistic, we used age-adjusted early-onset CRC mortality rates.

We considered a county to be a “hot spot” for early-onset CRC among women if it was identified as high risk using the following 3 geospatial analyses: (i) fifth quintile of smoothed EB early-onset CRC mortality rates, (ii) high-high cluster using Local Indicators of Spatial Association, and (iii) as defined by Getis-Ord  $G_i^*$  statistic. All other US counties were categorized as non-hot spots. Therefore, “hot spots” were defined conservatively to represent areas of high early-onset CRC mortality among women using all 3 approaches (see Table S1, Supplementary Digital Content 1, <http://links.lww.com/CTG/A446>).

### Study population

A case listing was run on the SEER 18 Registries Custom Data with additional treatment fields data set (SEER\*Stat 8.3.6) (15) to obtain clinicodemographic and survival data on first primary invasive early-onset CRC cases. Analysis was restricted to include 28,790 women diagnosed between years 1999 and 2016 in the contiguous United States with known age and race/ethnicity classified as non-Hispanic white, non-Hispanic black (NHB), Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, or unknown (Figure 1). Demographic and clinical characteristics examined included diagnosis age, race/ethnicity, county of CRC diagnosis, marital status, American Joint Committee on Cancer (AJCC) clinical stage, tumor site (colon or rectosigmoid junction/rectum), histopathologic subtype, cancer sequence, receipt of surgery, chemotherapy, and radiation therapy. Receipt of surgery, chemotherapy, and radiation therapy as part of first course of treatment were each evaluated as dichotomous variables (yes or no/unknown). Survival time was calculated from diagnosis date to last follow-up or death date. Follow-up for each patient is current within 22 months of the annual submission date (November 2018).

### SEER data linkage to CDC, American Community Survey, and County Health Rankings databases

To obtain characteristics on community-level health behaviors, SEER cases were linked with CDC data, as well as data from the 2014 County Health Rankings (CHR) and the 2014 American Community Survey (ACS)—by Federal Information Processing Standard county codes (Figure 1). The CHR and ACS comprise nationally representative data collected from a sample of the total noninstitutionalized population aged 18+ years residing in households. The CHR database uses several survey samples (e.g., Food Environment Atlas) to provide estimates of county-level factors, and the ACS provides aggregated estimates for demographic statistics over 5 years (2014 ACS: years 2010–2014).

From the 2014 CHR, we considered the county-level proportions of adult obesity (individuals aged 20+ years who report a body mass index 30+ kg/m<sup>2</sup>), adult smoking (adult population that currently smokes daily or most days and has smoked 100+ cigarettes in their lifetime), fertility (women aged 15–50 years with a live birth in the past year), foreign born individuals, limited access to healthy foods (population that is low income [having an annual family income ≤200% of the federal poverty threshold for the family size] and does not live close to a grocery store [rural areas: 10+ miles; nonrural: 1+ miles from a grocery store]), physical inactivity (adults [age: 20+ years] reporting no leisure-time physical activity), unemployment (civilian labor force aged 16+ years unemployed but seeking work), uninsured (population aged 18–65 years with no health insurance coverage), and primary care physicians (ratio of primary care physicians per 100,000 persons).

From the 2014 ACS, we collected county-level proportions for the following demographic characteristics: race/ethnicity (non-Hispanic white, NHB, or Hispanic), household income, and population density. Population density was evaluated as a dichotomous variable (rural/urban) as previously described (16) using 2010 Rural-Urban Commuting Area classifications: Urban areas were defined as population centers with 50,000+ residents and rural/nonurban areas were defined as towns or small cities with population centers with fewer than 50,000 residents.

### Statistical analysis

Differences in clinicodemographic characteristics by CRC mortality rate regions (hot spot vs non-hot spot counties) were compared using  $\chi^2$  tests for categorical variables and ANOVA or Wilcoxon rank-sum tests for continuous variables as appropriate. Multilevel regression models were used to investigate hot spot associations with CRC-specific survival. Cox proportional hazards models were used to assess hazard ratios and 95% confidence intervals for hot spot classification, age at diagnosis (5-year groups), race/ethnicity, tumor site, AJCC clinical stage, histopathologic subtype, and surgery in adjusted models. Multivariable-adjusted models were adjusted for individual-level clinicodemographic features that reached statistical significance in bivariate analysis ( $P < 0.05$ ) and county-level sociodemographic factors with Spearman correlations  $\rho \geq |0.20|$ , as well as the proportion of the population with household income <\$20,000 annually (as a proxy for socioeconomic status) with early-onset CRC hot spot residence. We tested the proportional hazards assumption in the multivariate model for CRC-specific survival by adding an interaction term with hot spot classification and follow-up time to the final models. Generalized R<sup>2</sup> values were calculated to quantify the variance in early-onset CRC

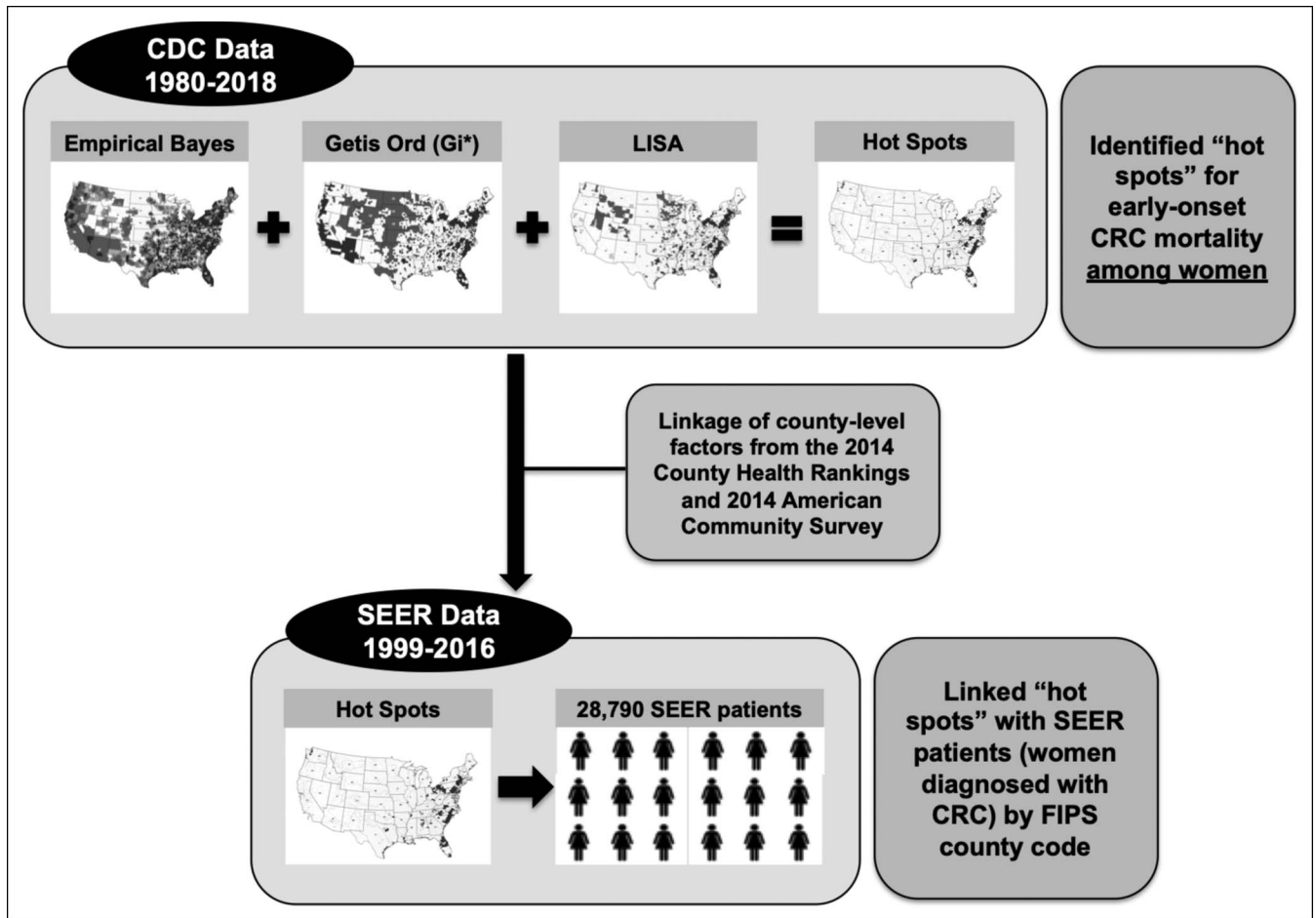
survival accounted for by individual-level clinicodemographic and county-level sociodemographic characteristics, and total variance accounted for by these features, using Cox proportional hazards regression. All data were analyzed using SAS v9.4 statistical software (SAS Institute, Cary, NC) and GeoDa v1.6.7.9 software. ArcGIS v10.5 geospatial processing software was used to visualize the geographic distribution of primary invasive early-onset CRC cases among women by hot spot classification. Statistical tests were 2-sided, with  $P < 0.05$  considered to be statistically significant.

### RESULTS

To define areas of high early-onset CRC mortality (hot spots) among women, we created a composite hot spot map with 3 geospatial analyses using CDC national mortality data (Figure 2). Approximately 1 in every 16 contiguous US counties identified as hot spots (191 of 3,108), and 52.9% of hot spot counties among women ( $n = 101$ ) were located in the Southern region of the United States (Figure 2 and see Table S1, Supplementary Digital Content 1, <http://links.lww.com/CTG/A446>). Using SEER, we identified a total of 28,790 female early-onset CRC cases between 1999 and 2016. Linkage of cases with geographic data by Federal Information Processing Standard county codes revealed that approximately 1 in every 7 women (13.7%) with early-onset CRC resided in hot spot counties (Table 1). Race/ethnicity, stage, histopathology, and receipt of first-course therapies significantly differed by hot spot residence ( $P \leq 0.001$ ; Table 1). NHBs accounted for nearly one-quarter of early-onset CRC cases among women in hot spot counties (23.7%) compared with 14.3% of all early-onset CRC cases among women in non-hot spot counties ( $P < 0.0001$ ). More than one-third of all women diagnosed with early-onset CRC had tumors of the rectosigmoid junction and rectum (36.8%). Despite the potential for chemotherapy and radiation therapy misclassification due to unknown therapy information, we noted differences in uptake of radiation therapy, chemotherapy, and surgery by early-onset CRC hot spot classification in our study population ( $P \leq 0.001$ ). Survival among women with early-onset CRC did not significantly differ by hot spot classification in adjusted models (see Table S2, Supplementary Digital Content 2, <http://links.lww.com/CTG/A447>).

Patterns of community health behaviors by early-onset CRC hot spot classification are presented in Table 2. Physical inactivity and fertility as well as the county-level proportion of NHB individuals were county-level features modestly correlated with hot spot residence ( $\rho = 0.21, -0.23, \text{ and } 0.26$ , respectively). On average, nearly one-quarter of adults residing in early-onset CRC hot spots (24.1%) reported no leisure-time physical activity, 4.9% of women (age 15–50 years) in early-onset CRC hot spot counties had a live birth in the past year, and NHB individuals comprised 19.3% of the population in hot spot counties (Table 2).

Next, to quantify the total and individual variance explained by these patient-level clinicodemographic and county-level sociodemographic characteristics on early-onset CRC survival among women, we used Cox proportional hazards regression models (Figure 3 and see Table S3, Supplementary Digital Content 3, <http://links.lww.com/CTG/A448>). Although race and ethnicity accounted for <0.5% of variation in early-onset CRC survival among women in non-hot spot counties, in hot spot counties, this factor explained 1.4% of the variation in CRC-specific survival among women with early-onset disease. Similar patterns were observed for individual-level features including



**Figure 1.** Composition of study population and data sources for this study. CDC, Centers for Disease Control and Prevention; CRC, colorectal cancer; FIPS, Federal Information Processing Standard; SEER, Surveillance, Epidemiology, and End Results.

tumor histopathology (2.1% vs 0.8%, respectively) and age at diagnosis (0.30% vs 0.09%), which explained greater variance in CRC-specific survival among hot spot counties vs non-hot spot counties. County-level proportions of the NHB population, women (age 15–50 years) with a live birth in the last year, and annual household income of <\$20,000 also explained greater variance in cancer-specific survival among young women residing in hot spot vs non-hot spot counties. Inversely, first course of therapy (including surgery, radiation therapy, and chemotherapy) each explained greater variance among women with early-onset CRC residing in non-hot spot counties vs hot spot counties (Figure 3 and see Table S3, Supplementary Digital Content 3, <http://links.lww.com/CTG/A448>). Irrespective of hot spot classification, AJCC clinical stage explained nearly 30% of variance in cancer-specific survival among women with early-onset CRC. Moreover, assessment of multicollinearity reported no evidence of high collinearity between tumor stage (variance inflation factor [VIF] = 1.03), chemotherapy (VIF = 1.20), and radiation (VIF = 1.18) (data not shown). Together, patient-level clinicodemographic and county-level sociodemographic features combined to account for one-third (34.0%) of the variation in early-onset CRC survival among women (see Table S3, Supplementary Digital Content 3, <http://links.lww.com/CTG/A448>). By hot spot classification, these factors accounted for different

proportions of variance in early-onset CRC survival among women residing in hot spot vs non-hot spot counties (33.8% vs 34.1%, respectively), although more than 60% of variance in early-onset CRC survival among women (66.2% in hot spot counties and 65.9% in non-hot spot counties) remained unexplained.

## DISCUSSION

Our analysis of the variation in CRC survival accounted for by individual-level and community-level characteristics among 28,790 women diagnosed with a first primary invasive cancer before age 50 years revealed distinct variance patterns in early-onset CRC survival by geographic region (hot spot vs non-hot spot counties). In addition, physical inactivity and fertility were community health behaviors that moderately correlated with hot spot residence among women diagnosed with early-onset CRC. Our study is the first to define areas of high early-onset CRC mortality (hot spots) in the contiguous United States among women, an analysis undertaken to minimize sex differences in early-onset CRC-specific outcomes, and to assess factors associated with geographic variation in early-onset CRC survival among women in the United States.

Chronic physical inactivity and excess caloric intake lead to energy imbalance, which over time results in the development of

**Table 1.** Summary of demographic and clinical characteristics by early-onset colorectal cancer hot spot counties in women diagnosed with colorectal cancer: SEER 18, 1999–2016

Characteristic	Total (N = 28,790) N (%) or mean (SE) <sup>c</sup>	Early-onset CRC county residence, N (%)		P <sup>b</sup>
		Hot spot <sup>a</sup> (n = 3,954 [13.7]) N (%) or mean (SE) <sup>c</sup>	Non-hot spot (n = 24,836 [86.3]) N (%) or mean (SE) <sup>c</sup>	
Age at diagnosis (%)				0.63
15–24 yr	717 (2.5)	112 (2.8)	605 (2.4)	
25–29 yr	1,022 (3.5)	150 (3.8)	872 (3.5)	
30–34 yr	2,094 (7.3)	279 (7.1)	1,815 (7.3)	
35–39 yr	3,734 (13.0)	505 (12.8)	3,229 (13.0)	
40–44 yr	7,248 (25.2)	999 (25.3)	6,249 (25.2)	
45–49 yr	13,975 (48.5)	1,909 (48.3)	12,066 (48.6)	
Mean, yr	42.4 (6.8)	42.4 (6.7)	42.3 (6.9)	0.26
Race/ethnicity (%)				<0.0001
Non-Hispanic white	16,800 (58.4)	2,279 (57.6)	14,521 (58.5)	
Non-Hispanic black	4,497 (15.6)	936 (23.7)	3,561 (14.3)	
Hispanic/Spanish/Latino	4,604 (16.0)	507 (12.8)	4,097 (16.5)	
Other <sup>d</sup>	2,889 (10.0)	232 (5.9)	2,657 (10.7)	
AJCC stage				0.0003
0–I	4,777 (16.6)	632 (16.0)	4,145 (16.7)	
II	4,853 (16.9)	731 (18.5)	4,122 (16.6)	
III	7,241 (25.2)	952 (24.1)	6,289 (25.3)	
IV	6,263 (21.8)	799 (20.2)	5,464 (22.0)	
Unknown	5,656 (19.6)	840 (21.2)	4,816 (19.4)	
Tumor site				0.29
Colon	18,199 (63.2)	2,529 (64.0)	15,670 (63.1)	
Rectosigmoid junction/rectum	10,591 (36.8)	1,425 (36.0)	9,166 (36.9)	
Histopathology				0.0001
Adenocarcinoma	21,690 (75.3)	2,883 (72.9)	18,807 (75.7)	
Mucinous adenocarcinoma	2,357 (8.2)	331 (8.4)	2,026 (8.2)	
Signet ring cell carcinoma	543 (1.9)	72 (1.8)	471 (1.9)	
Other	4,200 (14.6)	668 (16.9)	3,532 (14.2)	
Surgery				0.009
None	3,509 (12.2)	532 (13.5)	2,977 (12.0)	
Yes	25,281 (87.8)	3,422 (86.6)	21,859 (88.0)	
Radiation				0.001
None/unknown	23,535 (81.7)	3,310 (83.7)	20,225 (81.4)	
Yes	5,255 (18.3)	644 (16.3)	4,611 (18.6)	
Chemotherapy				0.0004
None/unknown	12,917 (44.9)	1,876 (47.5)	11,041 (44.5)	
Yes	15,873 (55.1)	2,078 (52.6)	13,795 (55.5)	

AJCC, American Joint Committee on Cancer; ANOVA, analysis of variance; CRC, colorectal cancer; SEER, Surveillance, Epidemiology, and End Results.

<sup>a</sup>Patients residing in counties with high rates of early-onset colorectal cancer mortality (fulfilling all 3 criteria for geographic clustering).

<sup>b</sup>Significance determined using  $\chi^2$ , ANOVA, or Wilcoxon rank-sum test as appropriate. P value calculations exclude unknown values.

<sup>c</sup>Presented as N (column percentage) or mean (SE).

<sup>d</sup>Other race/ethnicity includes American Indian/Alaska Native; Asian or Pacific Islander; or unknown.

**Table 2. Patterns of community health behaviors by female early-onset colorectal cancer hot spot classification among women diagnosed with colorectal cancer: SEER 18 linked with American Community Survey and County Health Rankings county-level data**

County-level characteristic	Early-onset CRC county residence		P <sup>b</sup>	Correlation (ρ <sup>c</sup> )
	Hot spots <sup>a</sup> (N = 3,954)	Non-hot spots (N = 24,836)		
	Presented as mean (SE)			
Race/ethnicity				
% Non-Hispanic white	57.8 (0.29)	54.8 (0.13)	<0.01	0.06
% Non-Hispanic black	19.3 (0.25)	11.5 (0.09)	<0.01	0.26
% Hispanic	15.1 (0.17)	22.2 (0.11)	<0.01	-0.09
% Household income <\$20,000/yr	15.1 (0.07)	17.5 (3.9)	<0.01	-0.12
Food environment index	7.9 (0.02)	7.5 (0.01)	<0.01	0.19
% Obesity	25.2 (0.07)	26.4 (0.03)	<0.01	-0.07
% Smoking	15.9 (0.04)	15.7 (0.03)	<0.01	0.09
% Completed college	35.5 (0.16)	29.3 (6.6)	<0.01	0.19
% Physical inactivity	24.1 (0.05)	21.7 (0.03)	<0.01	0.21
% Unemployment	9.4 (0.03)	9.2 (0.01)	<0.01	0.06
% Uninsured	15.9 (0.07)	18.4 (0.03)	<0.01	-0.17
PCP per 100,000 persons	82.9 (0.38)	73.6 (0.17)	<0.01	-0.13
% Nonurban (rural)	5.4 (0.14)	14.9 (0.14)	<0.01	-0.15
% Foreign born	1.02 (0.01)	0.97 (0.01)	<0.01	0.04
% Women aged 15–50 yr with a live birth in past year	4.9 (0.01)	5.4 (0.01)	<0.01	-0.23

CRC, colorectal cancer; ANOVA, analysis of variance; FIPS, Federal Information Processing Standard; PCP, primary care physician; SEER, Surveillance, Epidemiology, and End Results.

<sup>a</sup>County characteristic determined by the patient FIPS code. Patients residing in counties with high early-onset colorectal cancer mortality (fulfilling all 3 criteria for geographic clustering).

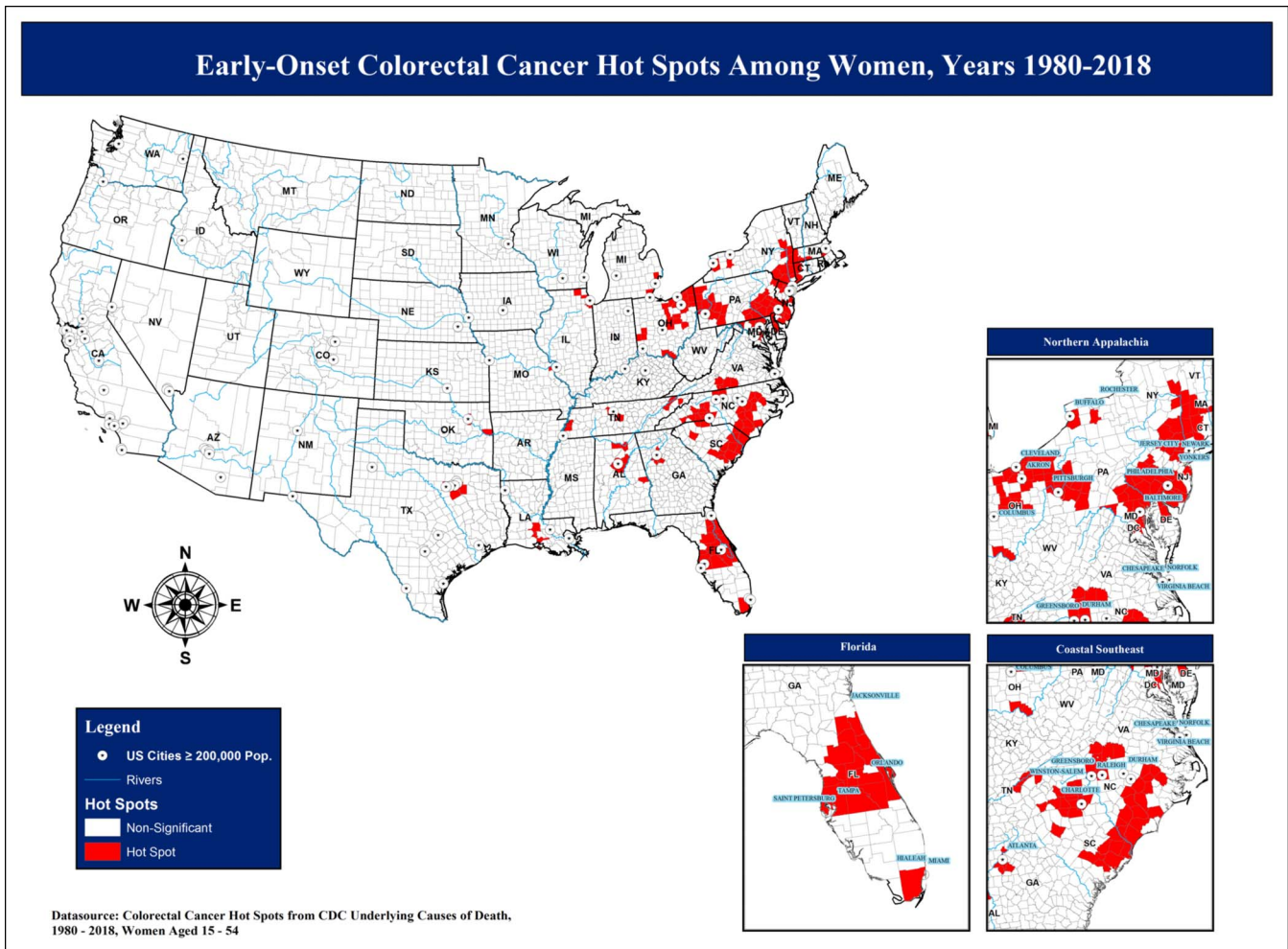
<sup>b</sup>Significance determined using ANOVA or Wilcoxon test.

<sup>c</sup>Spearman correlation with being a county-level early-onset CRC mortality hot spot.

obesity—a risk factor for early-onset CRC among women (5) and a prognostic factor of poor outcomes among patients diagnosed with gastrointestinal cancers (7). In this study, we revealed that physical inactivity was modestly and positively correlated with areas of high early-onset CRC mortality among women—approximately one-quarter of adults reported no leisure-time physical activity in hot spot vs one-fifth of adults in non-hot spot counties. In addition, female early-onset CRC hot spot counties were also composed of a higher proportion of NHB individuals compared with non-hot spot counties. Geographically, the highest prevalence of physical inactivity and poverty rates (23,24), the greatest proportion of the US black population (25), as well as a disproportionately high burden of chronic diseases (e.g., diabetes and cardiovascular disease) have been reported in the Southern United States—a region that accounts for 53% of female early-onset CRC hot spots in our study. Racial/ethnic disparities in physical inactivity persist among US adults; however, low levels of leisure-time physical activity in black Americans are almost entirely explained by poverty (26). Given the strong association between obesity and income poverty level (27,28), together with pronounced racial/ethnic disparities in obesity, physical inactivity, and early-onset CRC outcomes, counties with lower SES and reduced access to care may lend to early-onset CRC diagnosis at later stages. The differences in variance of early-onset CRC

survival among women by hot spot classification that is explained by first course of therapy regimen may in part be attributable to differences in receipt of treatment across geographic regions. Although we are limited in our ability to interpret these findings given inherent potential for misclassification of missing data in SEER (e.g., women classified as not having received chemotherapy/radiation therapy when those data were actually not recorded/captured), this difference may be explained by community-level factors, including high-quality healthcare access, urban density, distance to care provider, and residential setting (29–31). Consequently, studies designed to distinguish possible bidirectional causality between community health behaviors/sociodemographics and women diagnosed with early-onset CRC are also warranted. Together, these findings may inform cancer prevention/intervention strategies tailored to young patients and may ultimately help to inform knowledge to reverse early-onset CRC incidence trends and improve patient outcomes. Additional studies to examine early-onset CRC incidence patterns by hot spot classification among women will be helpful to discern whether disease incidence rates are higher in regions with disproportionately high early-onset CRC mortality.

For all-comers diagnosed with CRC, previous studies—including reports leveraging spatial mapping methodology—have indicated that CRC death rates are highest in Appalachia

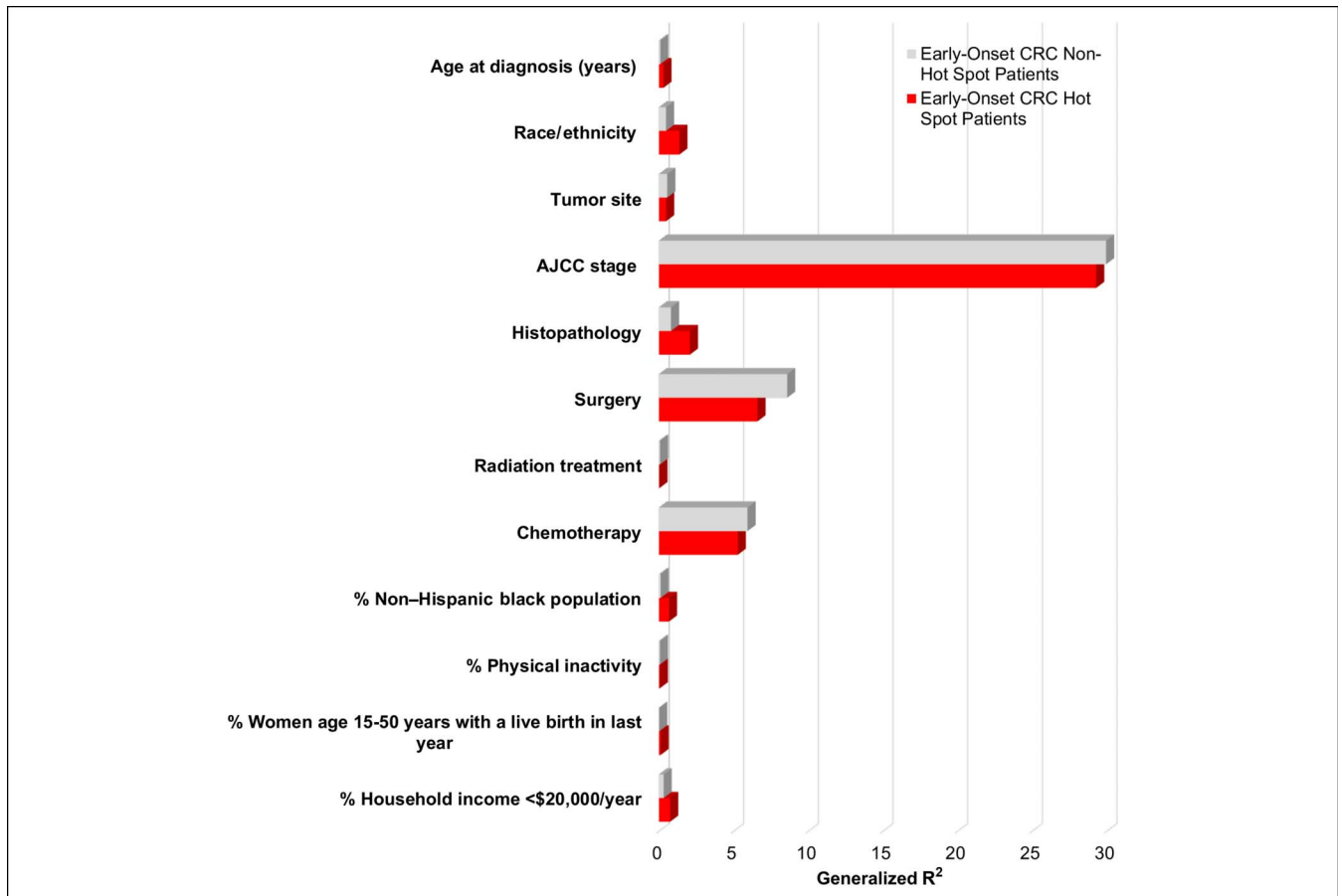


**Figure 2.** Early-onset colorectal cancer survival hot spot regions: Centers for Disease Control and Prevention (CDC), 1980–2017. Pop, population.

and areas of the South (32). Our findings among young women are consistent with these geographic disparities in CRC mortality among individuals of all ages. By contrast, our results differ from breast cancer mortality hot spots among women—because 53% of female early-onset CRC hot spot counties vs 72.5% of female breast cancer hot spot counties were in the South (16). Strikingly, a recent study of hot spots in early-onset CRC mortality aggregated for both men and women between 1999 and 2017 revealed that 92% of hot spot counties were in the Southern United States. (33). Yet from 1980 to 2018, we report a marked shift in female early-onset CRC hot spots because nearly half of all female hot spot counties were in the Midwest and Northeast regions of the United States, including northern Appalachia. Furthermore, there persists a strong differential impact of socioeconomic status on overall well-being, quality of life, income, and psychological and physical health, by sex. Poverty rates for all groups of adult women are higher than for their male counterparts (34), women with low income are more likely to develop alcohol and drug addictions influenced by the social stressors linked to poverty (35), and the effects of pregnancy on work/educational opportunities and costs associated for pregnancy are higher for women than men (36). Together, these factors may uniquely contribute to sex-specific disparities in early-onset cancer etiology and outcomes (3), including differences in community-level features by sex, and

are critical to unravel the underpinnings of the early-onset CRC epidemic.

The use of data from the population-based SEER registry program is a strength of this study because it allowed for a large number of pathologically verified cases to be identified among women across the United States with standardized 5-year follow-up. However, we acknowledge the inherent limitations in cancer registry data. One weakness of this study is that state-level colonoscopy data for this population could not be assessed. Although CRC screening among individuals younger than 49 years was not considered routine during our study period, differences in colonoscopy screening could partly contribute to variations in early-onset CRC survival among women. However, recent reports—which indicate that trends in colonoscopy screening do not fully align with early-onset CRC incidence patterns—further suggest the rising early-onset CRC burden is not fully explained by screening practices (37). We also acknowledge the inability to assess changes in female early-onset CRC mortality hot spots over time due to CDC data limitations at the county level (CDC data are suppressed at the county level when there are fewer than 10 deaths). SEER also lacks data on patient-level characteristics that can impact young patient outcomes, such as comorbid conditions (e.g., diabetes and insulin use) (38), comprehensive tumor histopathology and molecular phenotypes (e.g., microsatellite



**Figure 3.** Generalized  $R^2$  values for independently fit models among women with early-onset CRC by hot spot classification using the Cox proportional hazards regression: SEER 18 linked with American Community Survey and County Health Rankings county-level data.  $R^2$  values represent the variance explained by each independent factor. ACS, American Community Survey; AJCC, American Joint Committee on Cancer; CRC, colorectal cancer; SEER, Surveillance, Epidemiology, and End Results.

instability and somatic mutations), body mass index, history of gastrointestinal polyps, family history of cancer, and individual-level socioeconomic factors. However, use of county-level proportions of the population with an annual household income <\$20,000 as a proxy for socioeconomic status allowed us to explore potential differences in healthcare access that may contribute to geographic disparities in CRC outcomes. Although our findings raise the possibility that these individual/community-level features uniquely contribute to variation in early-onset CRC survival by geographic region, we are unable to provide evidence for causation between these features and early-onset CRC outcomes among women given the ecologic design of the study.

In summary, our findings emphasize the importance of defining patterns of variance in early-onset CRC survival to understand the impact of community health behaviors on early-onset CRC outcomes. We observed that physical inactivity and fertility were community health behaviors that modestly correlated with regions of high early-onset CRC mortality among women. We also observed that individual- and community-level factors accounted for approximately one-third of the variation in early-onset CRC survival among women and yielded distinct patterns by hot spot residence. Further study of community health behaviors and healthcare access, as well as modifiable CRC risk factors, is critical to

elucidate the unexplained variance in early-onset CRC survival among women in the United States—particularly in counties with high rates of early-onset CRC mortality—to reduce disparities in the early-onset CRC burden and improve young patient outcomes.

#### CONFLICTS OF INTEREST

**Guarantor of the article:** Andreana N. Holowatyj, PhD, MS.

**Specific author contributions:** A.N.H. and J.X.M.: had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. A.N.H., M.L., Y.H., R.V., and J.X.M.: contributed to the planning and conducting the study, collecting and interpreting data, and drafting/critical revision of the manuscript. All authors participated in the interpretation of data and drafting and critical revision of the manuscript for important intellectual content. A.N.H. and J.X.M.: obtained funding and provided support and supervision for this study.

**Financial support:** A. N. Holowatyj was supported by the Department of Medicine at the Vanderbilt University Medical Center and the National Institutes of Health (NIH) under Ruth L. Kirschstein National Research Service Award T32 HG008962 from the National Human Genome Research Institute. M. E. Langston was supported by the National Institutes of Health K12 DK111028 from the National Institute of Diabetes and Digestive and Kidney Diseases.



R. Viskochil was supported by the “Driving out Diabetes: A Larry H. Miller Family Wellness Initiative.” Y. Cao was supported by National Cancer Institute K07 CA218377. C. R. Rogers was supported by National Cancer Institute K01 CA234319. J. X. Moore was supported by a training grant from the National Cancer Institute of the National Institutes of Health under award number T32 CA190194. This work was also supported by the Huntsman Cancer Foundation and the ACSM Paffenbarger-Blair Fund for Physical Activity Epidemiology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the Driving out Diabetes: A Larry H. Miller Family Wellness Initiative. A. N. Holowatyj was also supported by the National Institutes of Health K12 HD043483 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

**Potential competing interests:** None to report.

## Study Highlights

### WHAT IS KNOWN

- ✓ CRC incidence rates are rising for men and women younger than 50 years (early-onset) with unknown etiologies.
- ✓ Disparities in survival among patients with early-onset CRC persist by sex.
- ✓ The role of health behaviors on CRC outcomes among young women remains unknown.

### WHAT IS NEW HERE

- ✓ We defined areas of high early-onset CRC mortality (hot spots) among women at the county level using geospatial methodology.
- ✓ Approximately 1 in every 16 contiguous US counties identified as early-onset CRC hot spots among women—53% of hot spot counties were located in the South.
- ✓ Proportions of physical inactivity and fertility were community health behaviors that modestly correlated with hot spot residence among women with early-onset CRC.
- ✓ Individual/community-level features together accounted for one-third of variance in early-onset CRC survival among women.
- ✓ Individual/community-level features accounted for distinct variance patterns in early-onset CRC survival among women.

### TRANSLATIONAL IMPACT

- ✓ Understanding the impact of community health behaviors is critical for tailoring strategies to reduce early-onset CRC disparities, particularly among women.

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70(1):7–30.
2. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal cancer incidence patterns in the United States, 1974–2013. *J Natl Cancer Inst* 2017;109(8):djw322.
3. Holowatyj AN, Ruterbusch JJ, Rozek LS, et al. Racial/ethnic disparities in survival among patients with young-onset colorectal cancer. *J Clin Oncol* 2016;34(18):2148–56.
4. Naishadham D, Lansdorp-Vogelaar I, Siegel R, et al. State disparities in colorectal cancer mortality patterns in the United States. *Cancer Epidemiol Biomarkers Prev* 2011;20(7):1296–302.
5. Liu PH, Wu K, Ng K, et al. Association of obesity with risk of early-onset colorectal cancer among women. *JAMA Oncol* 2019;5(1):37–44.
6. Nguyen LH, Liu PH, Zheng X, et al. Sedentary behaviors, TV viewing time, and risk of young-onset colorectal cancer. *JNCI Cancer Spectr* 2018;2(4):pky073.
7. Ulrich CM, Himbert C, Holowatyj AN, et al. Energy balance and gastrointestinal cancer: Risk, interventions, outcomes and mechanisms. *Nat Rev Gastroenterol Hepatol* 2018;15(11):683–98.
8. Hansen IO, Jess P. Possible better long-term survival in left versus right-sided colon cancer: A systematic review. *Danish Med J* 2012;59(6):A4444.
9. Benedix F, Kube R, Meyer F, et al. Comparison of 17,641 patients with right- and left-sided colon cancer: Differences in epidemiology, perioperative course, histology, and survival. *Dis Colon Rectum* 2010;53(1):57–64.
10. Lieberman DA, Williams JL, Holub JL, et al. Race, ethnicity, and sex affect risk for polyps >9 mm in average-risk individuals. *Gastroenterology* 2014;147(2):351–8; quiz e314–55.
11. Brenner H, Haug U, Hundt S. Sex differences in performance of fecal occult blood testing. *Am J Gastroenterol* 2010;105(11):2457–64.
12. Saunders BP, Fukumoto M, Halligan S, et al. Why is colonoscopy more difficult in women? *Gastrointest Endosc* 1996;43(2 Pt 1):124–6.
13. Enzi G, Gasparo M, Biondetti PR, et al. Subcutaneous and visceral fat distribution according to sex, age, and overweight, evaluated by computed tomography. *Am J Clin Nutr* 1986;44(6):739–46.
14. Centers for Disease Control and Prevention. Wide-Ranging Online Data for Epidemiologic Research (CDC-Wonder). 2019. (<http://wonder.cdc.gov>). Accessed November 1, 2019.
15. Surveillance, Epidemiology, and End Results (SEER) Program. SEER\*Stat Database: Incidence—SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2018 Sub (1973–2016 varying)—Linked To County Attributes—Total U.S., 1969–2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission. ([www.seer.cancer.gov](http://www.seer.cancer.gov)).
16. Moore JX, Royston KJ, Langston ME, et al. Mapping hot spots of breast cancer mortality in the United States: Place matters for blacks and Hispanics. *Cancer Causes Control* 2018;29(8):737–50.
17. Nassel AF, Root ED, Haukoos JS, et al. Multiple cluster analysis for the identification of high-risk census tracts for out-of-hospital cardiac arrest (OHCA) in Denver, Colorado. *Resuscitation* 2014;85:1667–73.
18. Anselin L. Local indicators of spatial association: LISA. *Geogr Anal* 1995;27:93–115.
19. Anselin L. Exploring Spatial Data with GeoDa: A Workbook. University of Illinois Urbana-Champaign: Champaign, IL, 2004.
20. Li H, Calder CA, Cressie NAC. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geogr Anal* 2007;39:357–75.
21. Li H, Calder CA, Cressie NAC. One-step estimation of spatial dependence parameters: Properties and extensions of the APLE statistic. *J Multivariate Anal* 2012;105:68–84.
22. Getis A, Ord K. The analysis of spatial association by use of distance statistics. *Geogr Anal* 1992;24:189–206.
23. Oates GR, Jackson BE, Partridge EE, et al. Sociodemographic patterns of chronic disease: How the mid-south region compares to the rest of the country. *Am J Prev Med* 2017;52(1 Suppl 1):S31–9.
24. Adult Physical Inactivity Prevalence Maps by Race/Ethnicity. 2020. (<https://www.cdc.gov/physicalactivity/data/inactivity-prevalence-maps/index.html>). Accessed August 30, 2020.
25. The Black Population: 2020. (<https://www.census.gov/prod/cen2010/briefs/c2010br-06.pdf>). Accessed August 30, 2020.
26. Merrill RM. Leisure-time physical inactivity's association with environmental, demographic, and lifestyle factors in the United States. *J Phys Act Health* 2020;17(4):412–22.
27. Akil L, Ahmad HA. Effects of socioeconomic factors on obesity rates in four southern states and Colorado. *Ethn Dis* 2011;21(1):58–62.
28. Lee H, Harris KM, Gordon-Larsen P. Life course perspectives on the links between poverty and obesity during the transition to young adulthood. *Popul Res Policy Rev* 2009;28(4):505–32.
29. Panchal JM, Lairson DR, Chan W, et al. Geographic variation in oxaliplatin chemotherapy and survival in patients with colon cancer. *Am J Ther* 2016;23(3):e720–9.
30. Panchal JM, Lairson DR, Chan W, et al. Geographic variation and sociodemographic disparity in the use of oxaliplatin-containing chemotherapy in patients with stage III colon cancer. *Clin Colorectal Cancer* 2013;12(2):113–21.
31. Soneji S, Iyer SS, Armstrong K, et al. Racial disparities in stage-specific colorectal cancer mortality: 1960–2005. *Am J Public Health* 2010;100(10):1912–6.
32. Siegel RL, Sahar L, Portier KM, et al. Cancer death rates in US congressional districts. *CA Cancer J Clin* 2015;65(5):339–44.

33. Rogers CR, Moore JX, Qeadan F, et al. Examining factors underlying geographic disparities in early-onset colorectal cancer survival among men in the United States. *Am J Cancer Res* 2020;10(5): 1592–607.
34. Eichener A, Robbins G. National Snapshot: Poverty Among Women & Families. 2014. (<https://nwlc.org/resources/national-snapshot-poverty-among-women-families-2014/>). Accessed August 30, 2020.
35. Mulia N, Schmidt L, Bond J, et al. Stress, social support and problem drinking among women in poverty. *Addiction* 2008;103(8): 1283–93.
36. Cawthorne A. The Straight Facts on Women in Poverty. 2008. ([https://www.americanprogress.org/issues/women/reports/2008/10/08/5103/the-straight-facts-on-women-in-poverty/#\\_edn1](https://www.americanprogress.org/issues/women/reports/2008/10/08/5103/the-straight-facts-on-women-in-poverty/#_edn1)). Accessed August 30, 2020.
37. Fedewa SA, Siegel RL, Jemal A. Are temporal trends in colonoscopy among young adults concordant with colorectal cancer incidence? *J Med Screen* 2019;26(4):179–85.
38. Holowatyj AN, Viskochil R, Ose D, et al. Diabetes, body fatness, and insulin prescription among adolescents and young adults with cancer. *J Adolesc Young Adult Oncol* 2020. doi: 10.1089/jayao.2020.0071.

---

**Open Access** This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.