



# The role of socioeconomic disparity in colorectal cancer stage at presentation

Aesha Patel<sup>1</sup> · Owen Gantz<sup>1</sup> · Pavel Zagadailov<sup>2</sup> · Aziz M. Merchant<sup>1</sup>

Received: 23 November 2018 / Accepted: 12 February 2019 / Published online: 20 February 2019  
© Italian Society of Surgery (SIC) 2019

## Abstract

Colorectal cancer, despite multiple screening measures being available, is the second leading cause of death due to cancer. Cancer stage at diagnosis is an important determinant of survival, where earlier stages have significantly increased rates of survival. By looking at various social health disparities (at a patient and geographic level) and their effect on stage at presentation, we will gain a better understanding of the effect they have on cancer outcomes. Data were collected from the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) database for the years 2007–2014. Covariates extracted were patient-level variables such as age, race, primary site, state/county, insurance status as well as county-level data which included percent urban population, median family income, rural–urban continuum code classification, percent of population that has not completed high school, percent of population below the poverty line, percent of population foreign-born, percent of language-isolated persons, and unemployment rate. The primary outcome analyzed was cancer staging at diagnosis. A  $\chi^2$  analysis and multivariate binary logistic regression was modeled to elucidate the associations between study covariates and late stage of cancer presentation. Chi-squared analysis demonstrated significant associations (at  $p < 0.05$ ) between stage of diagnosis with race, age, insurance status, location of primary site, percent of population below poverty line, percent of language-isolated persons, and percent of unemployed. To help reduce these disparities, community resources and increased screening and prevention techniques must be implemented to target the unique populations at greatest risk for developing the disease.

**Keywords** Cancer outcomes · Cancer screening · Colorectal cancer · Cancer risk factors · Social health disparities · Stage at diagnosis

## Introduction

According to the National Cancer Institute, colorectal cancer is the fourth leading cause of cancer in the US among both sexes, but the second leading cause of death due to cancer when both men and women are combined [1, 2]. Of those who are diagnosed, about 58% have a 10-year survival rate [3]. The incidence of colorectal cancer has decreased modestly in the past few decades due to new treatments, improvements in recognition and control of risk factors such

as smoking, and early screening, diagnosis, and interventions [4]. Despite these overall improvements, colon cancer incidence and mortality remain substantially higher among blacks than whites and higher among men than among women [4, 5].

While prevention is always the best treatment, cancer stage at diagnosis is one of the most important determinants of survival [6]. If diagnosed at earlier stages, colorectal cancer has an improved survival rate than during late stage diagnosis [7, 8]. Furthermore, from a health disparity perspective, when analyzed by the stage of cancer (localized, regional, distant), black race was associated with poorer survival rates than white race at each stage [9]. Moreover, insurance coverage accounts for a significant amount of the disparity in survival rate between younger black versus white patients with colon cancer [10]. In developed countries, later stage diagnoses of colon, as well as other cancers, are

✉ Aziz M. Merchant  
Aziz.merchant@rutgers.edu

<sup>1</sup> Department of Surgery, Rutgers-New Jersey Medical School, 185 South Orange Avenue, Suite MSB G530, Newark, NJ, USA

<sup>2</sup> Clinical Outcomes Research Group, CORG LLC, 178 Meadow Brook Road, Grantham, NH 03753, USA

associated with socioeconomic disparity, potentially exacerbating the already known difference in overall survival [11].

By continuing to understand the effect of various social health disparities, such as age, gender, race, type of insurance coverage, socioeconomic status, education level, unemployment rates, and poverty rates, on colorectal cancer stage at presentation, we may gain a better understanding of the effect they have on cancer outcomes. Studies have focused on the role of race and insurance status in outcomes of colon cancer stratified by stage at presentation, but few studies have looked at the role of income status, education level, and geographic urban versus rural disparity, all of which are known disparities in cancer care. To this end, we endeavored to understand the role that race, age, insurance status, education level, income level, and geographic disparity may play in the presentation of late stage colon cancer, potentially shedding light on how health disparities lead to increased cancer mortality. We performed a retrospective database analysis to understand the association that colorectal cancer stage (late versus early) at presentation may have with patient-specific and geographic disparities.

## Methods

### Data Source

Data were sourced from the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) database for the years 2007–2014, which is supplemented by county-level demographic data from the US Census and American Community Survey (ACS). SEER is a nationally representative survey of cancer registries that collects incidence and survival data from population-based registries that cover approximately 28% of the US population [12]. SEER diagnosis records prior to the year 2011 were matched to data from the 2007–2011 ACS, while diagnosis records after the year 2011 were matched to ACS data for 2011–2015 to ensure temporal accuracy for the county-level attributes used in this analysis.

### Study Population and Variables

Patients were selected on the basis of a malignant primary site in the colon or rectum and with a year of diagnosis between 2007 and 2014 and an age of diagnosis greater than or equal to 40 years old. Records with missing gender or age data were excluded from selection as well as those records missing valid staging data or a state/county assignment. Covariates abstracted from SEER included patient-level variables such as age, race, year of diagnosis, primary site, and state/county, insurance status, and derived American Joint Committee on Cancer (AJCC) 6th edition staging

at diagnosis. Additionally, county-level data available from the ACS were sourced for each record matched for the state/county data. This included percent urban population, median family income, rural–urban continuum code classification, percent of population that has not completed high school, percent of population below the poverty line, percent of population foreign-born, percent of language-isolated persons, and unemployment rate. The primary outcome analyzed in this study was cancer staging at diagnosis, which we defined as early for cancer with AJCC staging I or II and late for neoplasm staged III or IV at diagnosis [13].

### Statistical methods

A contingency table was first generated to identify counts of early and late stage diagnosis for each covariate under study and a  $\chi^2$  analysis was performed as a first approximation of the associations between these covariates and stage at presentation (Table 1). Continuous county-level covariates were grouped into quintiles for the purposes of the contingency table. This was performed as a univariate analysis to inform the independent associations of each covariate with stage at presentation.

However, since the  $\chi^2$  test does not demonstrate the directionality or independence of these associations, we generated a model using a multivariate binary logistic regression on the same data to further elucidate the associations between study covariates and stage of cancer presentation. Study covariates that were determined to be significant in the Chi-squared analysis were included in this model, which classified cases into early or late stage at presentation and accepted as input the significant study covariates, with continuous county-level covariates modeled as continuous variables. The risk of a late stage at diagnosis for colorectal cancer was calculated as an adjusted odds ratio (OR) and 95% confidence interval (CI) (Table 2). The regression and  $\chi^2$  analyses were performed in the R statistical computing software using the speedglm package [14, 15]. Statistical tests were two-tailed with  $\alpha = 0.05$ . This methodology is validated by previous literature analyzing disparities in healthcare outcomes with the SEER data set [16].

## Results

Presented in Table 1 is a contingency table showing the stage at presentation of the analyzed cases by patient demographics and county-level ACS attributes. Our study consisted of a total of 259,828 patients. Chi-squared analysis demonstrated significant associations (at  $p < 0.05$ ) between stage of diagnosis with sex, race, age, insurance status, location of primary site, percent of population below poverty line, percent of language-isolated persons, and percent of unemployed.

**Table 1** Contingency table and  $\chi^2$  analysis

	Stage at presentation		Total	$\chi^2$ analysis <i>p</i> value
	Early (I–II)	Late (III–IV)		
<b>Sex</b>				
Female	62 112 (52.1%)	57 171 (47.9%)	119 283	0.01
Male	66 749 (51.4%)	63 068 (48.6%)	129 817	
<b>Race</b>				
Black	13 667 (46.9%)	15 453 (53.1%)	29 120	< 0.01
Other <sup>a</sup>	11 088 (50.2%)	10 991 (49.8%)	22 079	
White	104 106 (52.6%)	93 795 (47.4%)	197 901	
<b>Age (years)</b>				
40–44	2 703 (38.8%)	4 260 (61.2%)	6 963	< 0.01
45–49	5 253 (40.5%)	7 708 (59.5%)	12 961	
50–54	10 864 (47.9%)	11 818 (52.1%)	22 682	
55–59	12 353 (47.6%)	13 599 (52.4%)	25 952	
60–64	14 600 (49.0%)	15 195 (51.0%)	29 795	
65–69	17 048 (53.2%)	14 987 (46.8%)	32 035	
70–74	17 059 (54.4%)	14 274 (45.6%)	31 333	
75–79	17 075 (55.9%)	13 448 (44.1%)	30 523	
80–84	15 987 (56.4%)	12 370 (43.6%)	28 357	
85+	15 919 (55.9%)	12 580 (44.1%)	28 499	
<b>Insurance status</b>				
Any medicaid	13 030 (45.9%)	15 361 (54.1%)	28 391	< 0.01
Insured	109 157 (57.6%)	97 680 (42.4%)	206 837	
Uninsured	3 090 (40.4%)	4 565 (59.6%)	7 655	
Unknown	3 584 (57.6%)	2 633 (42.4%)	6 217	
<b>Location</b>				
Ascending colon	20 669 (56.2%)	16 104 (43.8%)	36 773	< 0.01
Cecum	20 792 (49.6%)	21 153 (50.4%)	41 945	
Descending colon	5 938 (53.5%)	5 152 (46.5%)	11 090	
Hepatic Flexure	4 955 (54.6%)	4 121 (45.4%)	9 076	
Large Intestine, NOS	1 634 (21.5%)	5 954 (78.5%)	7 588	
Rectosigmoid Junction	9 089 (46.9%)	10 303 (53.1%)	19 392	
Rectum	26 271 (52.8%)	23 462 (47.2%)	49 733	
Sigmoid colon	26 224 (52.9%)	23 329 (47.1%)	49 553	
Splenic flexure	2 983 (50.1%)	2 971 (49.9%)	5 954	
Transverse colon	10 306 (57.3%)	7 690 (42.7%)	17,996	
<b>Percent urban</b>				
0–20%	4 333 (52.5%)	3 927 (47.5%)	8 260	0.23
21–40%	5 093 (52.3%)	4 640 (47.7%)	9 733	
41–60%	8 655 (51.3%)	8 201 (48.7%)	16 856	
61–80%	13 057 (52.1%)	12 028 (47.9%)	25 085	
81–100%	97 707 (51.7%)	91 437 (48.3%)	189 144	
<b>Median family income</b>				
\$24 800–\$50 200	13 324 (51.2%)	12 705 (48.8%)	26 029	0.17
\$50 201–\$75 400	70 463 (51.8%)	65 691 (48.2%)	136 154	
\$75 401–\$101 000	35 239 (51.7%)	32 860 (48.3%)	68 099	
\$101 001–\$126 000	9 819 (52.2%)	8 977 (47.8%)	18 796	
<b>By RU cont code</b>				
Rural	2 156 (52.3%)	1 968 (47.7%)	4 124	0.58
Suburban	111 951 (51.7%)	104 684 (48.3%)	216 635	
Unknown	262 (52.1%)	241 (47.9%)	503	

**Table 1** (continued)

	Stage at presentation		Total	$\chi^2$ analysis <i>p</i> value
	Early (I–II)	Late (III–IV)		
Urban	14 492 (52.1%)	13 346 (47.9%)	27 838	
Percent <HS education				0.54
0–9.89%	27 069 (51.8%)	25 233 (48.2%)	52 302	
9.90–18.4%	64 316 (51.6%)	60 390 (48.4%)	124 706	
18.5–26.9%	31 993 (52.0%)	29 585 (48.0%)	61 578	
27.0–35.4%	5 248 (52.1%)	4 819 (47.9%)	10 067	
35.5–43.9%	219 (51.5%)	206 (48.5%)	425	
Percent below poverty line				< 0.01
2.95–12%	45 933 (52.2%)	42 103 (47.8%)	88 036	
13–21%	66 282 (51.6%)	62 101 (48.4%)	128 383	
22–30%	15 254 (51.0%)	14 630 (49.0%)	29 884	
31–39%	1 270 (49.5%)	1 294 (50.5%)	2 564	
40–48%	106 (50.2%)	105 (49.8%)	211	
Percent foreign born				0.26
0–8.38%	42 621 (51.7%)	39 881 (48.3%)	82 502	
8.39–16.8%	25 524 (51.6%)	23 894 (48.4%)	49 418	
16.9–25.1%	30 642 (52.1%)	28 151 (47.9%)	58 793	
25.2–33.5%	12 009 (51.6%)	11 261 (48.4%)	23 270	
33.6–42%	18 049 (51.4%)	17 046 (48.6%)	35 095	
Percent language isolation				< 0.01
0–4.6%	60 351 (51.6%)	56 681 (48.4%)	117 032	
4.7–9.2%	36 496 (52.2%)	33 471 (47.8%)	69 967	
9.3–13.8%	23 225 (51.2%)	22 097 (48.8%)	45 322	
13.9–18.4%	8 538 (52.4%)	7 752 (47.6%)	16 290	
18.5–23%	235 (50.3%)	232 (49.7%)	467	
Percent unemployment				< 0.01
1.19–7.44%	26 634 (52.0%)	24 555 (48.0%)	51 189	
7.45–13.7%	94 971 (51.8%)	88 503 (48.2%)	183 474	
13.8–19.9%	7 183 (50.2%)	7 119 (49.8%)	14 302	
20.0–26.2%	57 (50.4%)	56 (49.6%)	113	

<sup>a</sup>American Indian/AK native, Asian/Pacific Islander

### Chi-squared analysis: demographics

Race, age, and insurance status were found to be significant in this analysis. Black patients presented with 53.1% late-stage cancer while white patients presented with 47.4% late-stage cancer. Increasing age showed a negative trend for late stage presentation with the older populations being more likely to present with early-stage cancers and the younger population presented with a higher percentage of late-stage cancers than early. Insurance status also demonstrated a significant association with both Medicaid patients and uninsured patients presenting with a higher percentage of late-stage cancer (54.1% and 59.6%, respectively) than early (45.9% and 40.4%, respectively), while insured patients presented with a lower percentage of late-stage cancer (42.4%) than early (57.6%). Location of primary site showed that the large intestine, NOS (78.5%) showed statistically

significant greater rates of late-stage cancer than the rectum (52.8%). When breaking the colon down into parts, the cecum (50.4%) and the rectosigmoid junction (53.1%) had statistically significant greater rates of late-stage cancer. The ascending colon (56.2%), descending colon (53.5%), hepatic flexure (54.6%), sigmoid colon (52.9%), splenic flexure (50.1%), and transverse colon (56.9%) showed greater rates of early-stage cancer.

### Chi-squared analysis: county-level

County-level ACS data were also examined to identify geographic associations with stage at diagnosis. Percent below poverty line, percent of language isolation, and percent unemployed all showed to be significant in this analysis. The effect of poverty was analyzed using the percentage of the population below poverty line in a patient's county and

**Table 2** Logistic regression model for late stage at presentation

	Odds ratio (95% CI)	<i>p</i> value
<b>Race</b>		
White	Ref	
Black	1.19 (1.1,1.28)	< 0.01
Other <sup>a</sup>	1.12 (1.03,1.22)	< 0.01
Unknown	0.50 (0.37,0.66)	< 0.01
<b>Age at Dx (years)</b>		
40–44	Ref	
45–49	0.91 (0.77,1.07)	0.26
50–54	0.59 (0.51,0.69)	< 0.01
55–59	0.66 (0.56,0.76)	<0.01
60–64	0.65 (0.56,0.75)	<0.01
65–69	0.55 (0.47,0.63)	< 0.01
70–74	0.52 (0.45,0.60)	< 0.01
75–79	0.49 (0.43,0.57)	< 0.01
80–84	0.49 (0.42,0.57)	< 0.01
85+	0.52 (0.45,0.61)	< 0.01
<b>Sex</b>		
Female	Ref	
Male	1.02 (0.97,1.06)	0.50
<b>Insurance status</b>		
Insured	Ref	
Any medicaid	1.22 (1.14,1.31)	< 0.01
Uninsured	1.36 (1.17,1.58)	< 0.01
Unknown	0.85 (0.73,1.00)	0.04
<b>Continuous county-level variables</b>		
Percent urban 2010	1.00 (1.00,1.00)	0.27
Median family income ACS	1.00 (1.00,1.00)	0.39
Percent < high school education ACS	1.01 (1.00,1.02)	0.11
Percent of persons below poverty ACS	0.99 (0.98,1.00)	0.20
Percent of language isolation ACS (households)	0.99 (0.97,1.00)	0.01
Percent of those moved previous year from outside US ACS (ages 5+)	1.11 (1.02,1.21)	0.02
Percent of unemployed ACS	1.00 (0.99,1.01)	0.89

ACS American community survey

showed higher rates of late-stage cancer for patients living in the most impoverished counties compared to the least (49.8% and 47.8%, respectively). Percent of language isolation in a geographic area showed higher levels of early-stage cancer for all percentages of language isolation. Unemployment rates showed a positive association with late-stage presentation. Counties with the lowest unemployment had 48.0% late-stage presentation versus counties with the greatest unemployment rates, which had 49.6% late-stage presentation (Table 1).

A logistic regression model was also used to find significant disparities in stage at presentation and demonstrate the factors that contribute independently to stage at presentation. This analysis showed significant results for race,

age, insurance, percent of language isolation, and percent of those who moved to US in the past year.

Compared to white patients, black patients had the highest likelihood of a late-stage presentation (OR 1.19,  $p < 0.01$ ) followed by patients of “other” race (American Indian, Alaska Native, Asian/Pacific Islander) who had the next highest likelihood of a late-stage presentation (OR 1.12,  $p < 0.01$ ). Additionally, increasing age was found to have an inverse relationship with late-stage colon cancer and instead found to be significantly associated with an earlier stage at presentation. This trend becomes significant in the 50–54 age group (OR 0.59,  $p < 0.01$ ) and continues to the oldest age group in the data set, 85+ (OR 0.52,  $p < 0.01$ ). Insurance status was also found to be predictive of stage at

presentation, with Medicaid (OR 1.22,  $p < 0.01$ ) and uninsured (OR 1.36,  $p < 0.01$ ) patients significantly more likely than insured patients to have a late stage at presentation.

County-level attributes also demonstrated significant associations with stage at presentation, specifically percent of population that is language-isolated (OR 0.99,  $p = 0.01$ ) and percent of population that moved to the US from another country in the past year (OR 1.11,  $p = 0.02$ ).

Sex, median family income, percent urban, percent below poverty line, percent below high school education, and unemployment percentage were not found to be significant predictors of a patient's cancer staging at diagnosis.

## Discussion

Our study is a Type II prognostic study, which is used to identify factors associated with subsequent clinical outcomes in a patient with a given disease. Our study uniquely examines the association between important disparities such as income factors, education level, and geographic disparities (as well as the traditional disparities of race, gender, and insurance status) for colorectal cancer over a 7-year period for a large nationally representative cohort. In this study, we assessed the role of patient and geographic disparities on stage at presentation of colorectal cancer.

At the patient level, the black race and being uninsured or on Medicaid all demonstrated significance in presenting with late-stage colon cancer. However, the results showed increasing age to be significant for early-stage colon cancer compared to late-stage colon cancer. This could likely be due to the United States Preventative Services Task Force Grade “A” recommendation to get colon cancer screening starting at the age of 50 (40 if there is a family history), and thus helping older patients detect the cancer earlier [17]. Another explanation and likely contributor is, a different more biologically aggressive tumor occurs in younger patients. This has been hypothesized due to the higher percentage of poorly differentiated and mucin-producing cancers found in the younger population. Patients of this age are also more likely to have colon cancer syndromes, such as familial adenomatous polyposis or Lynch syndrome [18]. In addition, we found in our univariate analysis that gender did have a significant relationship with stage of presentation; however, this association was later found to be explained by other risk factors in our multivariate analysis.

At the geographic level, percent of population that is language isolated and percent of population that immigrated to the US within the past year were significant predictors of late-stage presentation of colorectal cancer. Percent urban population, percent below high school education, percent below poverty line, and percent of unemployment were all

not significant predictors of late-stage cancer. Economic factors, such as median family income, were not associated with late-stage presentation of colorectal cancer; however, median family income is not a fully representative variable in the SEER database. The lowest data point in SEER for median family income begins close to the US poverty line of \$25,100 for a four-person household, therefore it does not accurately represent the population of patients below the poverty line [18].

Previous studies that used data from single states or single disparities have found similar associations between colorectal cancer stage at presentation and patient or geographic disparities [7, 19–27]. A similar study in 2009 by Halpern et al. [27] used the National Cancer Database to look at multiple patient characteristics associated with stage at diagnosis. In the study, they found insurance status, race, gender, and age to be significantly associated with colorectal cancer stage at diagnosis. Our study did not find a significant association for gender in the multivariate analysis, but had consistent results for insurance status, race, and age. This study also found increasing age to be significant for higher presentation of early-stage colon cancer further validating our results. Their study found that women had increased odds of presenting with late-stage cancer than men. The differences in significance of gender can be due to different national databases used or it can indicate that this disparity has subsided over time. While both are large national databases, SEER registries cover about 28% of the entire US population while National Cancer Databases have about 70% of all newly diagnosed cancers accounted for [28, 29]. Another reason might be due to the Halpern et al. study on patients diagnosed with cancer between 1998 and 2004, being done almost 10 years ago. The disparity among gender may have gone away since then and could be why it was not found in our study. Our study went further and assessed geographic disparities as well as trends in the past 10 years of the significant associations. Overall, most of the disparities mentioned in the previous studies have not changed from the past years and continue to remain a disparity.

A previous study by Valeri et al. used data from SEER to apply a counterfactual framework to colorectal cancer survival rates. This framework allows one to see what actually happened versus what would have happened if a variable was eliminated. These frameworks help show the importance of the variable and the strong effect it has on the results or in this case, the disparity. They used this method to estimate the extent to which race disparities among colorectal cancer survival would be reduced if the differences in stage at diagnosis were eliminated between races [30]. The study looked specifically at black versus white patients, and consistent with our results, found that black patients are more likely to be diagnosed with Stage IV cancer than white patients. When removing these disparities in stage at diagnosis, they

found that would reduce the overall difference in cancer outcomes between black and white patients by 35%. This is a significant reduction and demonstrates the importance of stage at diagnosis to survival outcomes.

A study by Winawer et al. found that those who got a colonoscopic polypectomy, which removes adenomatous polyps, as a preventative measure showed lower-than-expected incidence of colorectal cancer. Out of 1418 patients that had a colonoscopic polypectomy, only five patients had asymptomatic early-stage colorectal cancer detected later on, and no symptomatic cancers were detected [31]. Many patients are still not up-to-date with screenings despite various studies like this proving colorectal cancer screening to be an effective preventative measure. Another study by Siegel et al. looked at various colorectal cancer statistics. It identified the percentage of United States adults over 50 who got screening tests done, broken down by several disparities. It found that 68.3% of adults over 65 get screened, while only 57.8% of adults ages 50–64 get screened. Asian and Hispanic adults had the least screening percentages with 49.4% and 49.9% getting screened, respectively, compared to 65.4% of white adults get screened. Education level showed 71.3% of adults who graduated college get screened while only 47.4% of patients with a high school degree or less get screened. The largest difference exists in insurance status with 59.6% of insured patients getting screened, while only 25.1% of uninsured patients get screened [32]. This study shows the same disparities exist in lower screening rates as they do for later stage at diagnosis, further indicating the importance of reducing/eliminating disparities to improve colorectal cancer outcomes.

An article by Zonderman et al. [33], stated that one of the most potent nonbiological factors influencing the development of health disparities is poverty. The article attributed lower education levels and insurance status to lower socioeconomic status. This in turn can lead to infrequent doctors' visits, lower health literacy, and more unhealthy practices like smoking, which can all increase cancer risks dramatically. To help reduce these disparities, community resources and culturally appropriate techniques must be implemented that target the unique populations at greatest risk for developing the disease.

This study has several limitations. First, the SEER data are broadly representative of the United States cancer population but there are minor differences in foreign-born patients and urban inhabitants being overrepresented [33]. However, while the data are broadly representative, and SEER is considered the gold standard for data quality amongst cancer registries in the US and globally, the database is still incomplete and has inaccuracies since its data are collected from selected registries that may not be representative of the entire US. Inaccuracies can be due to miscoding of the data transmitted to SEER or the data

made available to the registrar for coding were not accurate. Data about socioeconomic status in particular are lacking [34]. In addition, patient migration is an important limitation of SEER. Patients moving into and out of SEER and non-SEER regions would be lost from the data leading to bias in the conclusions. This is an intrinsic limitation of all cancer databases that use registries. Using SEER, we used their Rural-Urban Continuum Code method to classify rural versus urban patients, which classifies patients based on the county they reside in. It is possible that patients might not seek care in the same areas of their residence causing inaccurate data. Also, SEER, as well as other observational studies, provide detailed data on diagnosis, stage, and treatment at the time of diagnosis, but long-term outcomes and follow-ups are not available. Although assessing the stage at presentation of colorectal cancer is a common method to determine outcomes, it is not completely indicative of outcomes and variations can exist.

SEER also does not collect some important variables that can help understand causes of poor outcomes; for example, comorbidities (such as diabetes, polyp history, genetic disorders), previous treatments, use of screening tests, or lifestyle factors (such as smoking, alcohol, tobacco, obesity, lack of physical activity, diet low in fruits/vegetables) and therefore, those variables cannot be controlled for [36–38]. A 2013 study by Johnson et al. performed a meta-analysis using 12 established non-screening colorectal cancer risk factors to quantify each of these risk factors' impact on colorectal cancer risk. They found inflammatory bowel disease and history of colorectal cancer in a first degree relative to be associated with the highest risk of colorectal cancer. They found increased Body Mass Index, red meat intake, cigarette smoking, low physical activity, low vegetable consumption, and low fruit consumption to be associated with moderate risk of colorectal cancer. Contrary to common associations, they did not find any significant associations between alcohol use, post-menopausal hormone therapy, processed meat, and colorectal cancer. They also did not find any significant difference in the risk of colorectal cancer with 5 years of using Aspirin/Non-steroidal anti-inflammatory drugs compared to no use [37].

In addition, there are several limitations of the statistical methods. SEER database has only a select number of confounding variables that can be controlled for. The results can be biased due to certain confounding variables that are not accounted for in the database. Further, data may be multicollinear, leading to unclear associations between the independent variables. This will lead to an unclear effect on the true association between the independent variable with the dependent variable. In addition, SEER classifies anyone with private insurance or Medicare as insured and it is not possible to extract solely Medicare data. Since majority of patients with colorectal cancer are elderly and have



Medicare, this could be an interesting result to see; however, it is not possible with SEER.

Further, while certain associations are noted to be significant due to a  $p$  value  $<0.01$ , the odds ratio may not deviate significantly from 1.0. This is due to the large sample size in the study. While these associations are statistically significant, they may not be clinically significant. We have presented the odds ratios in Table 2 to help assess which variables might have significant clinical findings as well.

Lastly, although the data assessed were from the most recent SEER data, and inherent time lag exists in the database, and therefore, the data might not reflect the most recent associations with colorectal cancer disparities and stage at presentation.

## Conclusion

In conclusion, this study revealed that there are still significant associations with certain disparities like black race, lower socioeconomic class, and uninsured patients. Previous studies over 10 years ago have also demonstrated similar disparities, indicating enough progress has not been made to reduce these disparities and thus reduce colorectal cancer incidence and result in improved survival rates.

**Author contributions** All authors of this manuscript have made substantial contributions to the conception and design of the study, and data analysis and interpretation. Drafting and revising the manuscript. Final approval of the manuscript submission.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest, financial or otherwise.

**Research involving human participants and/or animals** This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed consent** For this type of study, informed consent is not required.

## References

- [1] Key Statistics for Colorectal Cancer. American Cancer Society. <https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html>. Accessed on 12-15-2017.
- Cancer Stat Facts: Colon and Rectum Cancer. Surveillance Research Program, National Cancer Institute. <https://seer.cancer.gov/statfacts/html/colorect.html>. Accessed on 12-15-2017.
- Colorectal Cancer: Statistics. American Society of Clinical Oncology. <https://www.cancer.net/cancer-types/colorectal-cancer/statistics>. Updated on 8-2017. Accessed on 12-15-2017.
- US Cancer Statistics Working Group. United States Cancer Statistics: 1999-2013 Incidence and Mortality Web-based Report. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2016. <https://www.cdc.gov/uscs>. Accessed Jan 6, 2018.
- Abdel-Rahman O (2018) Challenging a dogma: 5-year survival does not equal cure in all colorectal cancer patients. *Expert Rev Anticancer Ther* 18(2):187–192
- Massarweh NN, Chiang YJ, Xing Y et al (2014) Association between travel distance and metastatic disease at diagnosis among patients with colon cancer. *J Clin Oncol* 32(9):942–948
- Oh HS, Chung HJ, Kim HK, Choi JS (2007) Differences in overall survival when colorectal cancer patients are stratified into new TNM staging strategy. *Cancer Res Treat* 39(2):61–64
- What are the Survival Rates for Colorectal Cancer, by Stage? American Cancer Society. <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html>. Accessed on 12-15-2017.
- White A, Joseph D, Rim SH, Johnson CJ, Coleman MP, Allemani C (2017) Colon cancer survival in the United States by race and stage (2001–2009): findings from the CONCORD-2 study. *Cancer* 24:5014–5036
- Sineshaw HM, Ng K, Flanders WD, Brawley OW, Jemal A (2017) Factors that contribute to differences in survival of black vs white patients with colorectal cancer. *Gastroenterology* 17:36351–36355
- Kweon SS, Kim MG, Kang MR, Shin MH, Choi JS (2017) Difference of stage at cancer diagnosis by socioeconomic status for four target cancers of the national cancer screening program in Korea: results from the Gwangju and Jeonnam cancer registries. *J Epidemiol* 27(7):299–304
- Overview of the Seer Program. Surveillance, Epidemiology, and End Results Program. <https://seer.cancer.gov/about/overview.html>. Access January 18, 2018.
- Chen AY, Schrag NM, Halpern MT, Ward EM (2007) The impact of health insurance status on stage at diagnosis of oropharyngeal cancer. *Cancer* 110(2):395–402
- Marco Enea (2017). speedglm: Fitting Linear and Generalized Linear Models to Large Data Sets. R package version 0.3-2. <https://CRAN.R-project.org/package=speedglm>
- Core Team R (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna <https://www.R-project.org/>
- Ratnapradipa KL, Lian M, Jeffe DB, Davidson NO, Eberth JM, Pruitt SL, Schootman M (2017) Patient, hospital, and geographic disparities in laparoscopic surgery use among surveillance, epidemiology, and end results-medicare patients with colon cancer. *Dis Colon Rectum* 60(9):905–913
- Screening for Colorectal Cancer (2016) US Preventive Services Task Force Recommendation Statement. *JAMA* 315(23):2564–2575
- Hargan ED (2018) annual update of the HHS poverty guidelines. *Fed Reg* 83(12):2642–2644
- Askari A, Nachiappan S, Currie A et al (2017) The relationship between ethnicity, social deprivation and late presentation of colorectal cancer. *Cancer Epidemiol* 47:88–93
- Farkas DT, Greenbaum A, Singhal V, Cosgrove JM (2012) Effect of insurance status on the stage of breast and colorectal cancers in a safety-net hospital. *J Oncol Pract* 8(3 Suppl):16s–21s
- Tawk R, Abner A, Ashford A, Brown CP. Differences in Colorectal Cancer Outcomes by Race and Insurance. Edberg M, Hayes BE, Rice VM, Tchounwou PB, eds. *International Journal of Environmental Research and Public Health*. 2016;13(1):48.



22. Parikh AA, Robinson J, Zaydfudim VM, Penson D, Whiteside MA (2014) The effect of health insurance status on the treatment and outcomes of patients with colorectal cancer. *J Surg Oncol* 110(3):227–232
23. Ward EM, Halpern M, Schrag N et al (2008) Association of insurance with cancer care utilization and outcomes. *A Cancer J Clin* 58:9–31
24. Ward EM, Fedewa SA, Cokkinides V, Virgo K (2010) The association of insurance and stage at diagnosis among patients aged 55 to 74 years in the national cancer database. *Cancer J* 16(6):614–621
25. Halpern MT, Ward EM, Pavluck AL, Schrag NM, Bian J, Chen AY (2008) Association of insurance status and ethnicity with cancer stage at diagnosis for 12 cancer sites: a retrospective analysis. *Lancet Oncol* 9(3):222–231
26. Paquette I, Finlayson SR (2007) Rural versus urban colorectal and lung cancer patients: differences in stage at presentation. *J Am Coll Surg* 205(5):636–641
27. Halpern MT, Pavluck AL, Ko CY, Ward EM (2009) Factors associated with colon cancer stage at diagnosis. *Dig Dis Sci* 54:2680
28. Surveillance, Epidemiology, and End Results Program. <https://seer.cancer.gov/about/overview.html>. Accessed 15 Mar 2017
29. National Cancer Database. American College of Surgeons. <https://www.facs.org/quality-programs/cancer/ncdb>. Accessed 21 Feb 2017
30. Valeri L, Chen JT, Garcia-Albeniz X, Krieger N, VanderWee TJ, Coull BA (2016) The role of stage at diagnosis in colorectal cancer black-white survival disparities: a counterfactual causal inference approach. *Cancer Epidemiol Biomark Prev* 25(1):83–89
31. Winawer SJ, Zauber AG, Ho MN et al (1993) Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med* 329:1977–1981
32. Siegel RL, Miller KD, Fedewa SA et al (2017) (2017) Colorectal cancer statistics. *Cancer J Clin* 67(3):177–193
34. Yu JB, Smith BD (2009) NCI seer public-use data: applications and limitations in oncology research. *Oncology* 23:288–295
35. Duggan MA, Anderson WF, Altekruse S, Penberthy L, Sherman ME (2016) The surveillance, epidemiology and end results (SEER) program and pathology: towards strengthening the critical relationship. *Am J Surg Pathol* 40(12):e94–e102
33. Zonderman AB, Norbeck Ejiogu N, J, Evans MK, (2014) The influence of health disparities on targeting cancer prevention efforts. *Am J Prev Med* 46(3):87–97
36. Colorectal Cancer. Centers for Disease Control and Prevention. [https://www.cdc.gov/cancer/colorectal/basic\\_info/risk\\_factors.htm](https://www.cdc.gov/cancer/colorectal/basic_info/risk_factors.htm). Published June 20, 2018. Accessed July 19, 2018.
37. Haggard FA, Boushey RP (2009) colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* 22(4):191–197
38. Johnson CM, Wei C, Ensor JE et al (2013) Meta-analyses of colorectal cancer risk factors. *Cancer Cau Control* 24(6):1207–1222

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