The rise of proximal colorectal cancer: a trend analysis of subsite specific primary colorectal cancer in the SEER database

Micheal Tadros[a], Sheena Mago[b], David Miller[a], Jane A. Ungemack[c], Joseph C. Anderson[d], Helen Swede[c]

Albany Medical College, NY; University of Connecticut Health Center, Farmington; Dartmouth Geisel School of Medicine, Hanover, New Hampshire, USA

Abstract	
	Background Colorectal cancer (CRC) is the third most common cancer worldwide and the second leading cause of cancer-related deaths. Given the significant prevalence of CRC, regular preventative screening is required. CRCs in different locations of the colon have variable molecular pathogenesis, gross appearance, and general disease outcomes. While the overall incidence of CRC has been decreasing, the decrease in proximally located CRC significantly lags behind the other forms of CRC. The objective of this study was to establish independent risk factors for proximal CRC for better identification of populations at risk for closer CRC monitoring and observation.
	Methods A time-trend analysis was conducted using data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database from 1973-2007, comparing patient characteristics (age, sex, race/ethnicity, year of diagnosis, age of diagnosis, tumor grade, tumor stage, and urban-rural setting) between CRCs originating in different locations.
	Results Analysis demonstrated that black race, female sex, age over 60, and being diagnosed in the 21st century (rather than 20th) were associated with an increased risk of proximal CRC compared to CRCs originating in other locations.
	Conclusions Our study showed that black race, female sex, and age over 60 independently increased the likelihood of proximal CRC diagnosis. Furthermore, CRC trends identify an increasing proportion of all CRCs being of proximal origin. It is imperative that patients undergo regularly scheduled complete colonoscopies by trained endoscopists, especially if they belong to the high-risk groups previously identified.
	Keywords Colorectal cancer, SEER data set, colorectal risk factors, proximal colorectal cancer
	Ann Gastroenterol 2021; 34 (4): 559-567

Introduction

Studies have demonstrated that the incidence, pathogenesis, and outcome of colorectal cancer (CRC) differ depending on the location of the tumor [1-4]. Cumulative evidence through the years has indicated that tumors located in the distal (left-sided colon cancer) and proximal (right-sided colon cancer) colon have distinct biological and genetic profiles [5,6].

Proximal CRCs are more likely to be poorly differentiated, have microsatellite instability, and have an advanced stage at the time of initial diagnosis [1,6,7]. Distal CRCs, in contrast, often harbor chromosomal unstable tumors and are frequently identified earlier in the disease process secondary to overt symptoms such as alteration of bowel habits or rectal bleeding [1,8,9]. The mechanisms behind these tumor discrepancies remain uncertain to date, and whether those clinical and molecular differences translate into different outcomes and prognosis has yet to be determined. In a recent population-based study using data from the national cancer institute's Surveillance, Epidemiology, and End Results (SEER) dataset, He *et al* observed that left-sided colon cancer was associated with a slightly lower risk of overall survival and better outcomes. Left-sided CRC was more likely to metastasize to the liver and lung at the time of diagnosis, though it had a higher

[a]Department of Medicine, Division of Gastroenterology-Hepatology, Albany Medical College, Albany, New York (Micheal Tadros, David Miller); [b]Department of Medicine, University of Connecticut Health Center, Farmington, Connecticut (Sheena Mago); [c]Community Medicine and Health Care, University of Connecticut Health Center, Farmington, Connecticut (Jane A Ungemack, Helen Swede); [d]Department of Medicine, Division of Gastroenterology, Dartmouth Geisel School of Medicine, Hanover, New Hampshire (Joseph C. Anderson), USA

Correspondence to: Micheal Tadros, MD, Albany Medical College, 43 New Scotland Ave, Albany, NY 12208, USA, e-mail: tadrosm1@amc.edu

overall survival and cancer-specific survival in comparison to proximal CRC [2]. In a study by Wang *et al*, using the SEER database between 1990 and 2014, no significant mortality differences were seen between patients less than 50 years old with proximal and distal CRCs [10].

The incidence of CRC has been declining rapidly over the past several decades, but comprehensive data are scarce on the long-term trends in incidence for proximal and distal CRC. This study was conducted to compare different patient demographics and patient characteristics within the SEER registry in order to risk-stratify populations at highest risk for proximal, distal, and rectal colorectal cancer. This study was conducted in the hope that the identification of high-risk CRC groups will have implications for future CRC screening recommendations.

Materials and methods

Study design and data sources

A retrospective, population-based cohort time-trend analytic study was conducted using the National Cancer Institute's SEER-17 database. The SEER database is a population-based registry of cancer statistics comprising of over 28% of the US population, including a variety of ethnicities and states [11]. It contains data regarding patient demographics and specific cancer history-related details, including primary tumor site, tumor morphology, initial cancer staging and initial treatment.

The SEER-17 database comprises all preceding SEER registry databases (SEER 9, SEER 11, SEER 13), plus data from Greater California, Kentucky, Louisiana, and New Jersey, which all joined SEER in 2000. It contains data regarding patient demographics, primary tumor site, tumor morphology, stage at diagnosis and first course of treatment. All registries included in the SEER database were integrated into the SEER program at different times. SEER 9 consists of the Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah sites. These registries joined SEER in 1973, except for Seattle-Puget Sound and Atlanta, which joined in 1974 and 1975, respectively. SEER 11 consists of SEER 9 with the addition of Los Angeles and San Jose-Monterey; both of these areas were added to SEER in 1992. SEER 13 consists of SEER 11 as well as the Rural Georgia and the Alaska Native Tumor Registry, which both joined SEER in 1992. The last 2 registries to have joined the SEER program included race as a variable in the database. SEER 17 consists of SEER 13, and Greater California, Kentucky, Louisiana and New Jersey, which all joined SEER in 2000. Cases from Louisiana diagnosed from July to December 2005 were excluded secondary to hurricanes Katrina and Rita [11].

Study population

Patients aged ≥20 years diagnosed with CRC between 1973 and 2007 were selected for the cohort. ICD-O-3 codes were used to identify CRCs. Our study sample identified a total of 505,141 patients, aged 20 and older. A total of 24,942 cases were excluded from the analysis because of missing data for the analytic variables described below.

For descriptive analyses, based on prior epidemiological and clinical research, CRC subsites were divided into proximal, distal and rectal cancers using ICD-O-3 codes. These codes correspond to the following locations: proximal, distal, rectal, or not otherwise specified (NOS). The codes that corresponded to proximal CRCs consisted of those in the cecum (C18.0), ascending colon (C18.2), hepatic flexure of colon (C18.3), transverse colon (C18.4) and splenic flexure of colon (C18.5), whereas the distal CRCs included the descending colon (C18.6) and the sigmoid colon (C18.7). Rectal CRCs encompassed the rectosigmoid junction (C19.9) and the rectum (C20.9) [12,13]. Demographic and clinical data regarding age, sex, race/ethnicity, year of diagnosis, initial cancer diagnosis, staging, treatment, along with rural vs. urban origin were extracted from the SEER database. In the SEER registry, races were listed as White, Non-Hispanic Black, American Indian, Asian/Pacific Islander (PI) and Hispanic. Grouping based on the year of diagnosis was divided into 5-year intervals (1973-1977; 1978-1982; 1983-1987; 1988-1992; 1993-1997; 1998-2002; 2003-2007).

Patients were also classified into 3 clinically-relevant age groups (ages 20-34, 34-59, and 60 and above). The latter group was selected as the upper limit cutoff point, based on a preliminary analysis demonstrating a steep increase in proximal CRC starting at the age of 60. We modified the SEER variable "Grade" by combining Grade IV with Grade III to create 3 categories: well differentiated (I); moderately differentiated (II); and poorly differentiated (III/IV) tumors. Tumors were classified as localized, regional or distant, based on the SEER staging system for analysis. Other variables were considered as extent of disease; behavior code ICD-0-2 & 3, collaborative stage, lymph node examined or positive, derived AJCC (American Joint Committee on Cancer), SEER summary stage 1977 & 2000, and their recode and override counterpart. Unfortunately, these variables cover only certain periods in the 35-year period and are not available for the entire period. The variable "urban vs. rural" in the SEER databased was used as a surrogate for socioeconomic status and used rural-urban continuum code definitions for 2003. We recoded this variable into urban-rural with similar definitions to those used by Paquette, which defined urban as the first 3 groups and rural as 7 and 9 (not adjacent to metro area and smaller populations) (Appendix) [14]. Therefore, the total number of patients classified as urban was 439,362, whereas 18,409 were designated as rural. Because of missing data, 47,370 cases were excluded from the analysis. Geographical variation data was classified in 2 ways using the SEER registry and urban/rural variables to assess whether geographic location had an effect on the CRC subsite location. We included the 9 registries that joined SEER from 1973-1975. These registries are Atlanta, Connecticut, Detroit, Iowa, Hawaii, New Mexico, Utah, San Francisco and Seattle. The 9 registries had 338,276 cases in total. Data from other SEER registries, totaling 166,865 cases, were excluded.

Statistical analysis

The congeniality chi-square test was used to compare the categorical data regarding demographic and specific cancer characteristics among different CRC sites (proximal, distal, rectum and NOS). Various multivariable logistic regression analysis models built using forward direction and Enter were used for the final analysis to estimate the odds ratios (OR) of proximal CRC compared to distal CRC using the variables of race and ethnicity, patient age, year of diagnosis, and other clinicopathological factors. Model 1 was created to adjust for age and race, while Model 2 stratified patient age, sex, racial/ ethnic groups, and year of diagnosis. Model 3 extended Model 2 by adding tumor stage and degree of differentiation. The Hosmer and Lemeshow goodness-of-fit statistical test was used for assessment of the logistic regression models. Accordingly, ORs with corresponding 95% confidence intervals [CIs] were generated and data with a P-value of <0.05 were considered to have statistical significance. All the statistical analyses were performed using SPSS Statistics 18.0 (IBM Corporation, Armonk, NY, USA).

Results

In total, 505,141 patients diagnosed with CRC were selected for this cohort from 1973-2007. Among the total cases there were 204,168 cases of proximal CRC (40.4%), 146,011 cases of distal CRC (28.9%) and 154,962 cases of rectal CRC (30.7%) identified in the SEER 17 data set. Table 1 summarizes the characteristics of the identified patients.

The percentage of proximal CRC increased steadily over the 35-year period from 1973-2007 (P<0.001) with an increase from 35.6% during 1973-1977 to 42.5% during 2003-2007 (Table 1). Overall, the percentage of proximal CRCs was not uniform across the examined age ranges (Fig. 1). The incidence of proximal CRCs appeared to increase at the greatest rate amongst those between the ages of 60-64 and reached an incidence of up to 54% amongst those above the age of 85 (P<0.001).

When the analysis was repeated based on age subdivision into 3 major age-range groups (20-34, 35-59 and above 60 years), it was found that patients in the age group 60 and greater had a significantly higher incidence of proximal CRC (43.6%) compared to the other age groups: 34.2% in the 20-34 age group and 30.3% in 35-59 age group (P<0.001). Amongst those over 60 years old, the percentage of proximal CRCs was relatively higher in comparison to the other CRC subtypes (Table 1 and Fig. 1).

A significant variation was present amongst the incidence of proximal CRC across different racial and ethnic groups (P<0.001). Non-Hispanic Blacks had the highest percent of proximal CRCs (46.1%) followed by Non-Hispanic Whites (41.1%), American Indians (37.2%) and Hispanics (37.2%). Asians and PI exhibited the lowest incidence of proximal CRCs (30.1%). Of the racial and ethnic groups examined, Asians

and Hispanics had higher percent of rectal CRC (35.6% and 34.9%, respectively) compared to Blacks (25.8%) (Table 2). The incidence of the 3 CRC subsites showed the highest degree of variability for American Indians, probably due to the small sample size. (Table 1 and Fig. 2). Females had a significantly higher percentage of proximal CRC in comparison to males (44.9% vs. 36.0%, P<0.001; Table 1).

Tumor differentiation varied amongst subsites, with a significantly larger portion of poorly graded tumors located in the proximal colon vs. the distal colon (53.6% vs. 21.0%, P<0.001; Table 1). A majority of the higher CRC tumor grades
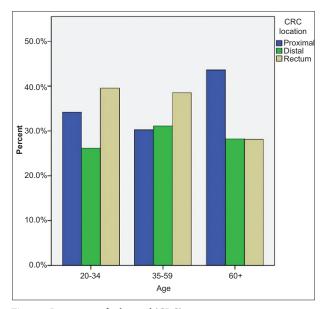

Figure 1 Percentage of colorectal (CRC) tumors across 3 age categories. All P-values <0.005
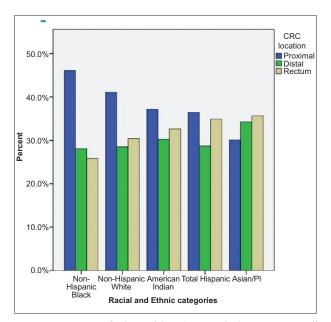

Figure 2 Percentage of colorectal (CRC) tumors by location, across all racial and ethnic groups. All P-values <0.005
PI, Pacific Islander

Table 1 Clinical characteristics of colorectal cancer at different primary sites based on the SEER database

Variable	Category	Proximal	Distal	Rectal
		n= 204,168	n= 146,011	n=154,962
		(40.5%)	(28.9%)	(34.9%)
Race	Non-Hispanic White	161,378	112,082	119,509
		(41.1%)	(28.5%)	(30.4%)
	Non-Hispanic Black	20,620	12,553	11,553
		(46.1%)	(28.1%)	(25.8%)
	American Indian	835	679	733
		(37.2%)	(30.2%)	(32.6%)
	Asian/Pacific Islander	10,085	11,489	11,950
		(30.1%)	(34.3%)	(35.6%)
	Hispanic	10756	8,470	10295
		36.4%	(28.7%)	34.9%
Age	20-34	1,714	1,313	1,983
		(34.20%)	(26.2%)	(39.6%)
	35-59	35,831	36,839	45,610
		(30.3%)	(31.1%)	(38.6%)
	60+	166,623	107,859	107,369
		(43.6%)	(28.2%)	28.10%
Sex	Male	92,134	76340	87174
		(36.0%)	29.9%	34.1%
	Female	112,034	69671	67788
		(44.9%)	27.9%	27.2%
Year of diagnosis	1973-1977	14,170	12,589	13,050
		(35.6%)	(31.6%)	(32.8%)
	1978-1982	17,117	14,797	14,604
		(36.8%)	(31.8%)	(31.4%)
	1983-1987	19,105	16,467	15,385
		(37.5%)	(32.3%)	(30.2%)
	1988-1992	21,434	16,552	16,159
		(39.6%)	(30.6%)	(29.8%)
	1993-1997	27816	19,313	19,962
		(41.5%)	(28.8%)	(29.8%)
	1998-2002	47,705	30,797	34,413
		(42.2%)	(27.3%)	(30.5%)
	2003-2007	56,821	35,496	41,389
		(42.5%)	(26.5%)	(31.0%)
Tumor grade	Well	18,999	18,174	15,588
		(36.0%)	(34.4%)	(29.5%)
	Moderate	107,186	80,977	82,605
		(39.6%)	(29.9%)	(30.5%)

(Contd...)

Table 1 (Continued)

Variable	Category	Proximal	Distal	Rectal
		n= 204,168	n= 146,011	n=154,962
		(40.5%)	(28.9%)	(34.9%)
	Poor	45,524	17,877	21,600
		(53.6%)	(21.0%)	25.4%)
Tumor stage	Local	69,014	60,742	65,905
		(35.3%)	(31.0%)	(33.7%)
	Regional	84,935	51,401	52,371
		(45.0%)	(27.2%)	(27.8%)
	Distant	42,633	28,276	24,922
		(44.5%)	(29.5%)	(26.0%)
Urban Rural	Urban	177,246	127,170	134,946
		(40.3%)	(28.9%)	(30.7%)
	Rural	7,678	5,165	5,566
		(41.7%)	(28.1%)	(30.2%)

* P-value is <0.001 in all comparisons above

Missing cases: 2154 when analyzing for race, 96,611 when analyzing for tumor grade, 24,942 when analyzing for tumor stage, and 47,370 when analyzing for urban vs. rural

Table 2 Incidence and mortality rates of colorectal cancer (CRC), 1975-2007

Year		1975	1985	2007
Age-adjusted	CRC	59.52	66.29	44.65
Incidence rates, per 100,000	Colon	41.34	47.5	31.88
	Rectum	18.18	18.42	12.70
Age-adjusted mortality rates, per 100,000	CRC	28.09	26.30	16.70

Adapted from Altekruse et al [31]

were located in the proximal colon at the time of diagnosis, in comparison to the distal colon and rectum. Of all the studied CRC patients, 44.5% of distantly spread tumors and 45% of regionally spread tumors originated from the proximal colon, as opposed to 35.3% of the locally spread tumors (P<0.001; Table 1 and Fig. 3), whereas for patients with carcinoma originating from the rectum there was more local rather than distant spread of the tumor (33.7% vs. 26.0%, P<0.001).

To investigate differences across the multiple SEER registries, the analysis was repeated using the individual SEER datasets independently of one another. Amongst the 9 locations that joined SEER around 1973, Hawaii was found to have the lowest percentage of proximal CRCs (32.1%), followed by New Mexico and Utah (37.4% and 37.5%, respectively). Iowa and Atlanta had the highest percentage of proximal tumors (43.2% and 41.8%, respectively, P<0.001).

Multiple models were built to estimate the likelihood of proximal CRC development as compared to distal CRC, using the variables of race and ethnicity, patient age, year of diagnosis,

and other clinicopathological factors. In Model 1, after adjusting for age and race, Non-Hispanic Blacks were found to have higher odds of proximal CRC than Non-Hispanic Whites (OR 1.32, 95%CI 1.3-1.35; P<0.001), while Asians and PI had a reduced likelihood of proximal CRC compared to Non-Hispanic Whites (OR 0.55, 95%CI 0.63-0.66; P<0.001; Table 3). An age greater than or equal to 60 years was associated with a higher likelihood of proximal CRC compared to patients aged 20-34 (OR 1.46, 95%CI 1.37-1.55; P<0.001).

Model 2, which also adjusted for patient's sex and year of diagnosis, in addition to the variables initially adjusted in Model 1, demonstrated that Non-Hispanic Blacks were at higher odds of having proximal CRC than Non-Hispanic Whites (OR 1.26, 95%CI 1.23-1.28; P<0.001), while Asians had the lowest odds (OR 0.63, 95%0.61-0.64; P<0.001; Table 3). Females were noted to have a higher risk of proximal CRC in comparison to males (OR 1.40, 95%CI 1.39-1.42; P<0.001), while those aged 60 and above were more likely to have proximal CRC than patients 20-34 years old (OR 1.47, 95%CI 1.39-1.56; P<0.001). With regard to incidence, the period between 2003-2007 had a significantly higher number of diagnosed proximal CRCs in comparison to those diagnosed between 1973-1997 (OR 1.44, 95%CI 1.41-1.48; P<0.001).

Model 3, further extended upon Model 2 by adjusting for initial tumor stage and degree of differentiation (Table 3). Findings did not differ appreciably from the above, though it was noted that proximal CRCs had a higher risk of having a poor grade of differentiation (OR 1.820, 95%CI 1.77-1.864; P<0.001) and distant tumor stage (OR 1.218, 95%CI 1.95-1.241; P<0.001).

Various sensitivity analyses were performed and the results proved to be consistent. For example, Model 3 was repeated

Table 3 Odds ratios (OR) and 95% confidence intervals (CI) for right-sided colon cancer compared to left-sided colorectal cancer
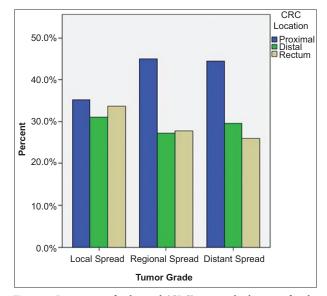
Variable		Multivariate model 1 n= 502,987		Multivariate model 2 n= 502,987		Multivariate model 3 n=395,966	
		OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value
Race	Non-Hispanic White	1.00		1.00		1.00	
	Non-Hispanic Black	1.316 (1.29-1.342)	<0.001	1.262 (1.237-1.288)	<0.001	1.331 (1.301-1.362)	<0.001
	American Indian	0.93 (0.853-1.015)	0.102	0.873 (0.801-0.953)	<0.001	0.892 (0.892-0.981)	0.019
	Asian/Pacific Islander	0.647 (0.631-0.663)	<0.001	0.627 (0.612-0.643)	<0.001	0.638 (0.621-0.655)	<0.001
	Hispanic	0.886 (0.854-0.908)	<0.001	0.834 (0.814-0.856)	<0.001	0.843 (0.820-0.867)	<0.001
Age	20-34	1.00		1.00		1.00	
	35-59	0.819 (0.771-0.870)	<0.001	0.830 (0.781-0.881)	<0.001	0.853 (0.796-0.913)	<0.001
	60+	1.455 (1.371-1.544)	<0.001	1.47 (1.386-1.561)	<0.001	1.531 (1.431-1.637)	<0.001
Sex	Male			1.00		1.00	
	Female			1.406 (1.390-1.423)	<0.001	1.403 (1.385-1.421)	<0.001
Year of diagnosis	1973-1977			1.00		1.00	
	1978-1982			1.043 (1.014-1.073)	<0.001	1.101 (1.059-1.143)	<0.001
	1983-1987			1.070 (1.041-1.100)	<0.001	1.171 (1.128-1.215)	<0.001
	1988-1992			1.183 (1.151-1.215)	<0.001	1.308 (1.262-1.356)	<0.001
	1993-1997			1.306 (1.272-1.340)	<0.001	1.448 (1399-1.499)	<0.001
	1998-2002			1.380 (1.347-1.414)	<0.001	1.560 (1.510-1.613)	<0.001
	2003-2007			1.442 (1.408-1.477)	<0.001	1.653 (1.600-1.708)	<0.001
Tumor grade	Well					1.00	
	Moderate					1.047 (1.026-1.069)	<0.001
	Poor					1.820 (1.77-1.864)	<0.001
Tumor stage	Local					1.00	
	Regional					1.28 (1.261-1.299)	<0.001
	Distant					1.218 (1.095-1.241)	<0.001

Figure 3 Percentage of colorectal (CRC) tumors by location for the different tumor grades. All P-values <0.005

among cases with confirmed adenocarcinoma only (Table 3). in addition, stratification for urban and rural origin was performed, with results comparable to those of the previously described models. No appreciable differences between urban vs. rural settings was noted amongst those with proximal CRC. Lastly, stratification was performed amongst the 9 registries that joined SEER around 1973 and again demonstrated the higher risk of proximal CRC amongst females, Non-Hispanic Blacks, a more recent year of diagnosis, and age greater than or equal to 60 years.

Discussion

A total of 204,168 (40.4%) cases of proximal CRCs, 146,011 (28.9%) cases of distal CRCs and 154,962 (30.7%) cases of rectal CRCs were identified in the SEER 17 data set. Over time, the prevalence of proximal CRC increased significantly, from 35.6% in 1973-1977 to 42.2% in 2003-2007. Across the 35 years of data, the incidence of proximal CRCs

increased over time. The year of initial diagnosis remained an independent risk factor for proximal CRC in the logistic regression analysis. A possible explanation for this finding was a change in behavioral factors that could in turn increase the risk of proximal tumors, limiting the benefit of colonoscopy. For example, a prospective study showed that a lack of exercise predisposes individuals to proximal CRCs [15]. According to this theory, the United States obesity epidemic could have led to the increased incidence of proximal tumors and may explain the proportional increase in proximal CRCs in all SEER registries [16-18].

The finding that Non-Hispanic Blacks have a higher proportional percentage of proximal tumors compared to Non-Hispanic Whites, while Asian and PI have the lowest percentage, agrees with the results of prior studies [19,20]. The higher incidence of proximal CRCs in Non-Hispanic Blacks could be postulated to be from underutilization of colonoscopy, higher rates of obesity compared to other racial groups, or insurance disparity [8,16]. As demonstrated in other studies, females were found to have significantly higher rates of proximal tumors compared to males [13,21]. These differences became more prominent over the 35-year period examined in this study. This finding may reflect higher rates of missed cancer due to higher rates of incomplete or technically challenging colonoscopies amongst females, lesser utilization of colonoscopy, a higher prevalence of flat adenomas, and/ or hormonal or biologic differences in comparison to males [22-26].

Those over the age of 60 were also found to have a significantly higher prevalence of proximal CRC compared to other types of colorectal cancer [20,21]. This finding may be explained by the anatomy of proximal CRC tumors and their general delay in symptom onset compared to distal CRC. Proximal tumors are usually noted to be bulkier and mostly asymptomatic, while distal tumors, owing to the narrower colonic lumen, are symptomatic earlier in the disease process. This hypothesis is further supported by the study's findings that proximal CRCs at the time of diagnosis had on average significantly higher tumor grades and were noted to be more advanced during staging than their distal CRC counterparts. This was further associated with worse survival rates and a lack of significant improvement in mortality and survival benefit for proximal tumors with intervention [27-29].

When geographic locations in the SEER registry were compared, Hawaii had the lowest percentage of proximal tumors. This finding may be explained either by the higher percentage of PI living there compared to other US states, or perhaps as secondary to different behavioral exposures. Utah was also found to have a low percent of proximal tumors, possibly explainable by the behavioral or dietary patterns of the state's inhabitants.

In summary, the analysis showed that, even with control for overlapping variables, an increased risk of proximal CRC is found in patients who are female, Non-Hispanic Blacks, above the age of 60, and those diagnosed in the 21st century. The Hosmer and Lemeshow goodness-of-fit test was significant (P<0.001), indicating that the variations in subsite-related CRC were not entirely explained. One unusual finding was that American Indians in Connecticut had a high OR for proximal CRC. This finding was, however, not statistically significant and most likely can be attributed to the small sample size of American Indians in the studied cohort.

Proximal CRCs lag significantly behind other located CRCs in terms of reduction in mortality rates, improved survival and reduction in incidence. This gap may be explained by the molecular and genetic differences of proximal cancers, differential risk factors, the role of flat adenoma and lastly, from missed or interval development of cancer.

Based on the findings of this study, the following are suggested areas to help narrow this gap: (i) improve endoscopy training to minimize missed or interval cancer rates; (ii) focus on measures to evaluate endoscopists (e.g., withdrawal time, and adenoma detection rate); (iii) invest in better technology to improve visualization during colonoscopy, for example high definition colonoscopy; (iv) develop better preparation regimens for colonoscopy; (v) identify those patients at risk of having difficult or incomplete colonoscopy, such as women, those of Non-Hispanic black race, and older patients; (vi) improve colonoscopy screening rates for the population, especially in women and minorities; (vii) better understand the genetic and molecular distinction between the proximal and distal colon; and (viii) develop better techniques to detect flat adenomas in order to assess subtype differences in incidence.

To the best of our knowledge, this study is one of the most comprehensive reviews of temporal and demographic trends in cancers of proximal and distal colon, and the rectum. This study included more than half a million cases over a 35-year period from 17 different registries in the United States. The SEER database is a reliable source for cancer research and has been widely used. Our findings are consistent with other studies and provide a broader and more longitudinal review of proximal, distal and rectal cancers across time, person and place [13,20,21,30,31]. This study developed a comprehensive regression analysis model to help identify demographic factors associated with an increased likelihood of proximal tumors.

One possible limitation of this study was the roughly 20,000 colorectal cancer cases within the dataset with no specified anatomic location. This was, however, a small percentage of the cases (4%). The exclusion of these cases should have a minimal effect on the results presented. Another limitation of this study is the lack of information available in the SEER dataset about CRC risk factors, such as diet, obesity, smoking, polyps and family history. Furthermore, it was difficult to identify carcinomas *in situ* in the analysis, as different classification systems were used in SEER in each represented time period. Carcinoma *in situ* has a more benign course compared to other types of cancer. The data were analyzed over time to develop a comprehensive understanding of CRC cancer by subsite. The inclusion of carcinoma *in situ* could affect mortality outcome, an aspect that was not examined in this study.

In the last 35 years, the proportion of proximal CRCs has increased, with Blacks experiencing the highest proportional percentage of proximal tumors, followed by Whites, American

Indians and Hispanics. Asians and PI, on the other hand, experience the fewest cases of proximal CRC. Females and older individuals (over 60 years) have significantly higher risks for proximal CRC. Proximal tumors also have a worse tumor grade and more advanced stage at the time of presentation and diagnosis. More research is needed to better understand why the overall decreasing incidence of proximal CRC lags behind that of the other tumors of the large intestine. This study helps identify the patient subgroups with a higher likelihood of proximal CRC, in whom screening colonoscopy would derive the greatest benefit.

Though the incidence and the mortality from colorectal cancer have decreased over the last few decades, this study sheds light on the disparity of CRC characteristics based on the originating colonic location. Our study and other emerging evidence suggest that cancer of the large bowel may reflect different processes due to biologic, anatomic, molecular, pathologic, technical and epidemiologic distinctions. We have established that patients who are female, black, over the age of 60, and with a later year of diagnosis had an independently increased likelihood of proximal CRC identification as opposed to other forms of CRC.

Summary Box

What is already known:

- Right- and left-sided colorectal cancers (CRCs) have distinct biological and genetical profiles, with right-sided CRCs having microsatellite instability, unlike left-sided, which often harbor chromosomal abnormalities
- Previous small-scale studies have demonstrated that the incidence, pathogenesis, and outcome of CRC differ depending on the geographical location
- The incidence of CRC has been declining rapidly over the past several decades, but comprehensive data are scarce on the long-term trends for CRC based on originating subsites

What the new findings are:

- Proximal CRCs lag significantly behind other located CRCs in terms of reduction in mortality rates, improved survival and reduction in incidence
- This study is one of the most comprehensive reviews of temporal and demographic trends in cancers of proximal and distal colon, and rectum, including more than half a million cases over a 35-year period from 17 different SEER registries in the United States
- There is an higher risk of proximal CRC among patients who are female, Non-Hispanic Blacks, patients aged more than 60 years, and those diagnosed in the 21st century

References

1. Lee GH, Malietzis G, Askari A, Bernardo D, Al-Hassi HO, Clark SK. Is right-sided colon cancer different to left-sided colorectal cancer? - a systematic review. *Eur J Surg Oncol* 2015;**41**:300-308.
2. He XK, Wu W, Ding YE, Ding YE, Li Y, Sun LM, Si J. Different anatomical subsites of colon cancer and mortality: a population-based study. *Gastroenterol Res Pract* 2018;**2018**:7153685.
3. Jess P, Hansen IO, Gamborg M, Jess T; Danish Colorectal Cancer Group. A nationwide Danish cohort study challenging the categorisation into right-sided and left-sided colon cancer. *BMJ Open* 2013;**3**:e002608.
4. Loupakis F, Yang D, Yau L, et al. Primary tumor location as a prognostic factor in metastatic colorectal cancer. *J Natl Cancer Inst* 2015;**107**:dju427.
5. Weiss JM, Pfau PR, O'Connor ES, et al. Mortality by stage for right- versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results—Medicare data. *J Clin Oncol* 2011;**29**:4401-4409.
6. Yamauchi M, Morikawa T, Kuchiba A, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut* 2012;**61**:847-854.
7. Ghazi S, Lindforss U, Lindberg G, Berg E, Lindblom A, Papadogiannakis N; Low-Risk Colorectal Cancer Study Group. Analysis of colorectal cancer morphology in relation to sex, age, location, and family history. *J Gastroenterol* 2012;**47**:619-634.
8. Elnatan J, Goh HS, Smith DR. C-KI-RAS activation and the biological behaviour of proximal and distal colonic adenocarcinomas. *Eur J Cancer* 1996;**32A**:491-497.
9. Hutchins G, Southward K, Handley K, et al. Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *J Clin Oncol* 2011;**29**:1261-1270.
10. Wang Y, Yang L, Zhou M, et al. Disparities in survival for right-sided vs. left-sided colon cancers in young patients: a study based on the Surveillance, Epidemiology, and End Results database (1990-2014). *Cancer Manag Res* 2018;**10**:1735-1747.
11. The Surveillance, Epidemiology, and End Results Program. 2011. Available from: http://seer.cancer.gov/ [Accessed 9 January 2021].
12. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010;**116**:544-573.
13. Murphy G, Devesa SS, Cross AJ, Inskip PD, McGlynn KA, Cook MB. Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *Int J Cancer* 2011;**128**:1668-1675.
14. Paquette I, Finlayson SR. Rural versus urban colorectal and lung cancer patients: differences in stage at presentation. *J Am Coll Surg* 2007;**205**:636-641.
15. Friedenreich C, Norat T, Steindorf K, et al. Physical activity and risk of colon and rectal cancers: the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:2398-2407.
16. Nutrition, Physical Activity, and Obesity: Data, Trends and Maps from CDC. 2020. Available from: https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html [Accessed 2 February 2021].
17. Behavioral Risk Factor Surveillance System; Prevalence and Trends Data. 2017. Available from: https://www.cdc.gov/brfss/brfssprevalence/index.html [Accessed 2 February 2021].
18. Sandler RS. Editorial: colonoscopy and colorectal cancer mortality: strong beliefs or strong facts? *Am J Gastroenterol* 2010;**105**:1633-1635.
19. Irby K, Anderson WF, Henson DE, Devesa SS. Emerging and widening colorectal carcinoma disparities between blacks and whites in the United States (1975-2002). *Cancer Epidemiol*

Biomarkers Prev 2006;**15**:792-797.

20. Wu X, Chen VW, Martin J, et al; Comparative Analysis of Incidence Rates Subcommittee, Data Evaluation and Publication Committee, North American Association of Central Cancer Registries. Subsite-specific colorectal cancer incidence rates and stage distributions among Asians and Pacific Islanders in the United States, 1995 to 1999. *Cancer Epidemiol Biomarkers Prev* 2004;**13**:1215-1222.

21. Saltzstein SL, Behling CA. Age and time as factors in the left-to-right shift of the subsite of colorectal adenocarcinoma: a study of 213,383 cases from the California Cancer Registry. *J Clin Gastroenterol* 2007;**41**:173-177.

22. Anderson JC, Gonzalez JD, Messina CR, Pollack BJ. Factors that predict incomplete colonoscopy: thinner is not always better. *Am J Gastroenterol* 2000;**95**:2784-2787.

23. Anderson JC, Messina CR, Cohn W, et al. Factors predictive of difficult colonoscopy. *Gastrointest Endosc* 2001;**54**:558-562.

24. Stock C, Knudsen AB, Lansdorp-Vogelaar I, Haug U, Brenner H. Colorectal cancer mortality prevented by use and attributable to nonuse of colonoscopy. *Gastrointest Endosc* 2011;**73**:435-443.

25. O'brien MJ, Winawer SJ, Zauber AG, et al. Flat adenomas in the National Polyp Study: is there increased risk for high-grade dysplasia initially or during surveillance? *Clin Gastroenterol Hepatol* 2004;**2**:905-911.

26. Wong R. Proximal tumors are associated with greater mortality in colon cancer. *J Gen Intern Med* 2010;**25**:1157-1163.

27. Wong RJ. Marked variations in proximal colon cancer survival by race/ethnicity within the United States. *J Clin Gastroenterol* 2010;**44**:625-630.

28. Wray CM, Ziogas A, Hinojosa MW, Le H, Stamos MJ, Zell JA. Tumor subsite location within the colon is prognostic for survival after colon cancer diagnosis. *Dis Colon Rectum* 2009;**52**:1359-1366.

29. Cheng X, Chen VW, Steele B, et al. Subsite-specific incidence rate and stage of disease in colorectal cancer by race, gender, and age group in the United States, 1992-1997. *Cancer* 2001;**92**:2547-2554.

30. Wu X, Cokkinides V, Chen VW, et al. Associations of subsite-specific colorectal cancer incidence rates and stage of disease at diagnosis with county-level poverty, by race and sex. *Cancer* 2006;**107**:1121-1127.

31. Altekruse SF, Kosary CL, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2007. 2010. Available from: https://seer.cancer.gov/archive/csr/1975_2007/ [Accessed 9 January 2021].

APPENDIX

Urban vs. Rural:

In the SEER dataset, there appears to be no meaningful difference in the distribution of CRC subsite in urban vs. rural areas.
"Counties in metro areas of 1 million population or more"
"Counties in metro areas of 250,000 to 1 million population"
"Counties in metro areas of fewer than 250,000 population"
"Urban population of 20,000 or more, adjacent to a metro area"
"Urban population of 20,000 or more, not adjacent to a metro area"
"Urban population of 2,500 to 19,999, adjacent to a metro area"
"Urban population of 2,500 to 19,999, not adjacent to a metro area"
"Completely rural or less than 2,500 urban population, adjacent to a metro area"
"Completely rural or less than 2,500 urban population, not adjacent to a metro area"

Other socioeconomic variables in the SEER county data are 1- place of birth, 2- marital status, 3-level of education (9th grade, high school, and at least Bachelor's degree), 4- median household income, 5- poverty level (person- family), 6- unemployment level and 7- white vs. blue collar. These variables require access to county data and will be analyzed as part of our future research