Type 2 Diabetes and Risk of Early-Onset Colorectal Cancer

Zitong Li, MD^{1*}, Hanyu Chen, MD^{1*}, Cassandra D.L Fritz, MD^{2*}, Xiaobin Zheng, MD, PhD¹, Xiaoyu Zong, MPH¹, Katelin B. Nickel, MPH³, Andrew Tipping, MA³, Long H. Nguyen, MD^{4,5}, Andrew T. Chan, MD, MPH ^{4,5,6,7}, Edward L. Giovannucci, MD, ScD^{8,9}, raham A. Colditz, MD, DrPH^{1,10}, Margaret A. Olsen, PhD, MPH^{1,3}, Peter T. Campbell, PhD, MSc¹⁷

Graham A. Colditz, MD, DrPH^{1,10}, Margaret A. Olsen, PhD, MPH^{1,3}, Peter T. Campbell, PhD, MSc¹¹, Nicholas O. Davidson, MD, DSc², Ryan C. Fields, MD^{10,12}, Yin Cao, MPH, ScD^{1,4,10}

¹ Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

² Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

³ Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St Louis, MO, USA

⁴ Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

⁵ Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

⁶ Broad Institute of MIT and Harvard, Cambridge, Massachusetts

⁷ Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁸Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA

⁹Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

¹⁰ Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA

¹¹ Department of Population Science, American Cancer Society, Atlanta, GA, USA

¹² Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

*contributed equally as first authors

Short title: Diabetes and Early-Onset Colorectal Cancer

Correspondence:

Yin Cao, MPH, ScD, Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, 660 S. Euclid Ave., Campus Box 8100, St. Louis, MO 63110.

Tel: 314-747-3925; Email: vin.cao@wustl.edu

Keywords: diabetes; colorectal cancer; early-onset; epidemiology

Abbreviations: CI, confidence interval; CRC, colorectal cancer; FOBT, fecal occult blood test; GI, gastrointestinal; HCUP CCS, Healthcare Cost and Utilization Project's Clinical Classification Software; HMO, Health Maintenance Organization; IBD, inflammatory bowel diseases; IQR, interquartile range; OR, odds ratio; PPO, Preferred Provider Organization; SD, standard deviation.

Word count: 2962 (Limit: 4000); Tables: 4; References: 53

ABSTRACT

Objective: Early-onset colorectal cancer (CRC) is increasing in many developed countries. Type 2 diabetes mellitus has increased substantially in younger adults; however, its role in early-onset CRC remains unidentified.

Design: We conducted a claims-based nested case-control study using IBM® MarketScan® Commercial Database (2006-2015). Incident early-onset CRC diagnosed at ages 18-49 were identified by ICD-9-CM diagnosis code, and the first coded diagnostic pathology date was assigned as the index date. Controls were frequency matched with cases. Type 2 diabetes, stratified by severity, was identified through ICD-9-CM using the Klabunde algorithm. Multivariate logistic regressions were used to estimate odds ratios (ORs) and 95% confidence intervals (Cls). **Results:** A total of 6001 early-onset CRC and 52104 controls were included. Type 2 diabetes was associated with an increased risk of early-onset CRC (5.0% in cases vs. 3.7% in controls; OR 1.24; 95% Cl 1.09 to 1.41). The positive association was more pronounced for uncontrolled (OR 1.37; 95% Cl 1.12 to 1.67) or complicated (OR 1.59; 95% Cl 1.08-2.35) type 2 diabetes compared to controlled diabetes (OR 1.13; 95% Cl 0.94 to 1.36). The positive association was driven by proximal (OR 1.35; 95% Cl 1.03 to 1.77) and distal (OR 1.67; 95% Cl 1.30 to 2.15) colon cancer but not rectal cancer.

Conclusions: Individuals with type 2 diabetes have a higher risk of early-onset CRC, with stronger associations for uncontrolled/complicated diabetes. The rising prevalence of type 2 diabetes among younger adults in the US may partially contribute to the increasing incidence of early-onset CRC.

Word count: 249 (Limit: 250 words)

INTRODUCTION

Colorectal cancer (CRC) is the second most common cause of cancer-related death in the United States (U.S.).¹ Due to increased CRC screening and uptake of colonoscopy, the incidence and mortality rates of CRC have declined for several decades among adults aged 50 years and older.² In contrast, the incidence and mortality of early-onset CRC (individuals younger than 50 years) have been increasing since the mid-1990s.³⁻⁵ During 2012-2016, the incidence of proximal colon, distal colon, and rectal cancer all rose at 1.8% annually among adults younger than 50.⁴ Such an alarming increase in early-onset CRC contributed to a 6 year drop in the median age of CRC diagnosis, from 72 years during 1988-1989 to 66 years during 2015-2016.⁵ Further elucidation of risk factors that contributed to this alarming increase is pivotal.

Thus far, obesity⁶ and sedentary lifestyle⁷ are among the potential contributors to the rise in early-onset CRC, pointing to a possible role of insulin dysregulation. However, the role of type 2 diabetes in early-onset CRC has not been elucidated from large U.S. population-based studies.^{8 9} Although the association between type 2 diabetes and average to late-onset CRC is established,¹⁰⁻¹² the emerging molecular characteristics of early-onset CRC¹³ support the necessity to revisit such association in a younger population. The link between early-onset CRC with obesity and prolonged sitting, both of which are risk factors for type 2 diabetes,^{14 15} further lend support to such unmet need. Additionally, type 2 diabetes, if etiologically relevant to early-onset CRC, likely contributes to the rising incidence of early-onset CRC due to the paralleled increase of type 2 diabetes¹⁶⁻²³ and the rise in early-onset CRC in the U.S. and globally.²⁴ Specifically in the U.S., between 1988 and 2012, type 2 diabetes has increased from 2.7% to 4.5% for ages 20-44 years, and from 13.3% to 16.2% among ages 45-64.¹⁷ It is also worth noting that compared with adults with type 2 diabetes at older ages, type 2 diabetes before age 45 appears to be a more aggressive disease with increased risk of requiring insulin.²⁵ Therefore, investigating the role of type 2 diabetes in early-onset CRC in a population-based

study will likely generate significant insights into the etiology, prevention, and early detection of earlyonset CRC.

To address these critical knowledge gaps, we used the IBM® MarketScan® Commercial Database (2006-2015), a longitudinal database that contains individual-level commercial health insurance claims data from over 113 million individuals from all geographic areas of the U.S., to comprehensively examine the association between type 2 diabetes and risk of early-onset CRC.

METHODS

Study population

We conducted a nested case-control study of early-onset CRC using the MarketScan Database (2006-2015), a longitudinal, de-identified, individual-level healthcare claims database comprised of more than 113 million commercially insured U.S. adults under age 65.²⁶ The database captures information on outpatient and inpatient insurance-reimbursable services, prescription data, type of health plan, and demographic information. Compared with all non-elderly people with employer-sponsored insurance in the U.S., MarketScan enrollees have a similar age and sex distribution.²⁷ The Institutional Review Board (IRB) at Washington University in St. Louis decided that this project was not considered to meet federal definitions under the jurisdiction of an IRB and therefore falls outside the purview of the Human Research Protection Office.

Ascertainment of cases and controls

All adults with an incident diagnosis of CRC between ages 18 and 49 were considered as early-onset CRC and were identified by an International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis code (153.0-153.9, 154.0, 154.1, and 154.8). To reduce false positives, we included only pathology-coded CRC cases and assigned the first diagnostic pathology date as the index date. Confirmed pathology diagnoses are automated into pathology ICD-9-CM

diagnosis codes, which were considered accurate in reporting pathology findings.²⁸ We restricted our analyses to adults with at least two years of enrollment before the index dates, as well as 90 days of enrollment after the index dates to derive metastatic status. CRCs were further classified into proximal colon (153.0-153.1, 153.4-153.6), distal colon (153.2-153.3, 153.7), unspecified colon (153.8-153.9), and rectal (154.0-154.1) tumor. Metastatic status was imputed using coded diagnosis and/or treatment records for liver/lung metastasis within 90 days of diagnosis.²⁹ We excluded CRC patients with any prior/concurrent cancer history (V10.x) or genetic susceptibility to malignant neoplasm (V84.0x), as well as CRC patients with any cancer except non-melanoma skin cancer identified through Cost and Utilization Project's Clinical Classification Software (HCUP CCS)³⁰ within two years before the index dates. The identification of CRC as cases or the other cancers for exclusion for both cases and controls were based on Klabunde *et al*,³¹ which requires at least one inpatient facility claim and/or two outpatient or provider claims 31-365 days apart.

Controls without CRC were frequency matched with cases by up to an 8:1 ratio based on age (18-24 and every five years thereafter), sex (female, male), geographical region (Northeast, North Central, South, West, unknown) due to geographic variations in the incidence of early-onset CRC,³² duration of insurance enrollment before index diagnosis (years), and prescription drug coverage (yes, no). Controls were selected to ensure that the distribution of the control index dates matched the distribution of index dates among the cases to account for changes over time. Controls were assigned random index dates and were selected to match the year of the corresponding case's index date. Controls with genetic susceptibility and personal cancer history were also excluded.

Ascertainment of type 2 diabetes mellitus

To reduce bias due to increased detection of diabetes and other comorbidities for CRC cases during the workup period but not among controls,³³ we restricted our exposures and covariates from 91 days

to two years before the index dates. Two years before the index dates were chosen to maximize statistical power. More importantly, this time period would allow us to capture almost all patients with diabetes in the target population as 96% of patients with diabetes had at least one diabetes-related appointment with a healthcare provider within a 2-year period.³⁴ Type 2 diabetes was defined using ICD-9-CM diagnosis codes (250.00, 250.02, 250.10, 250.12, 250.20, 250.22, 250.30, 250.32, 250.40, 250.42, 250.50, 250.52, 250.60, 250.62, 250.70, 250.72, 250.80, 250.82, 250.90, 250.92) and the Klabunde algorithm.³¹ Type 2 diabetes was further classified as controlled/not stated as uncontrolled without complications (with 250.02, 250.12, 250.20, 250.30 for all the encounters), uncontrolled without complications (with 250.02, 250.12, 250.22, 250.32 in any of the encounters),³⁵ or complicated (250.40, 250.42, 250.50, 250.52, 250.60, 250.52, 250.60, 250.62, 250.70, 250.62, 250.70, 250.72, 250.80, 250.72, 250.80, 250.82, 250.80, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.91, 250.22, 250.50, 250.52, 250.60, 250.52, 250.70, 250.72, 250.80, 250.82, 250.80, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.82, 250.90, 250.92, 250.80, 250.82, 250.90, 250.82, 250.90, 250.92, 250.92, 250.90, 250.92, 250.82, 250.90, 250.92, 250.92, 250.92, 250.90, 250.92, 250.82, 250.90, 250.92, 250.92, 250.92, 250.92, 250.82, 250.90, 250.92, 250.92, 250.92, 250.92, 250.92, 250.82, 250.90, 250.92, 250.92, 250.92, 250.92, 250.92, 250.82, 250.90, 250.92, 250.92, 250.92, 250.92, 250.92, 250.92, 250.92, 250.82, 250.90, 250.92,

Assessment of covariates and other clinical information

We extracted demographic information, including employment status, urban/rural residence, geographical region, health plan, and derived the Charlson Comorbidity Index³⁶ without diabetes. We also extracted information on potential confounders, including inflammatory bowel diseases (IBD), obesity, and family history of gastrointestinal cancer between 91 days and 2 years prior to the index dates. Information on fecal occult blood testing (FOBT) and screening/other colonoscopies during the same time period was retrieved. We also obtained information on a list of pre-specified early signs/symptoms for CRC, including gastrointestinal bleeding, abdominal pain, anemia, change of bowel habits, diarrhea, constipation, and weight loss between 91 days and 2 years prior to the index dates.³⁷

Statistical analysis

To evaluate the association between type 2 diabetes and risk of early-onset CRC, multivariable logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (Cls). We first adjusted for matching factors including age (years), sex, duration of insurance enrollment (number of completed years of enrollment), geographical region (Northeast, North Central, South, West, unknown), and prescription drug coverage before the index dates. We then additionally adjusted for full time employment status, residence (urban, rural, unknown), health plan (Preferred Provider Organization, Health Maintenance Organization, others), Charlson Comorbidity Index (continuous), IBD, obesity, family history of gastrointestinal cancer, screening colonoscopy, other colonoscopy, and FOBT. Individuals without geographic region or residence (<2%) were included in the adjustment using missing indicators. To examine the robustness and generalizability of these findings to an asymptomatic population, we conducted sensitivity analyses by restricting to participants without the following: IBD, prior colonoscopy/FOBT, or pre-specified early signs/symptoms of CRC. We also examined whether the association differed according to the severity of type 2 diabetes (controlled, uncontrolled, complicated).

We further examined if the association between type 2 diabetes and early-onset CRC differed according to tumor anatomical site (colon [proximal colon, distal colon, unspecified colon], rectum). We also conducted stratified analyses to evaluate the association among subgroups, including sex, age at the index date (18-45 vs. 46-49 years), birth year (\leq 1965 vs.>1965), and geographical region (South vs. others). *P* value for interaction was estimated using a Wald test on the cross-product term of type 2 diabetes and each stratification factor. All the analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina, USA). All the statistical tests were two-sided and *P* values <0.05 were considered statistically significant.

RESULTS

A total of 6001 early-onset CRC cases and 52104 controls were included in the analyses (**table 1**). The mean age of early-onset CRC patients was 43.0 years. Compared with controls, early-onset CRC cases were more likely to have IBD and be coded for obesity. They also had higher rates of colonoscopies other than for screening and FOBT tests before the index dates.

Type 2 diabetes was present in 5.0% of early-onset CRC patients, compared with 3.7% among the controls (**table 2**). In comparison to those without type 2 diabetes, individuals with type 2 diabetes had a 24% increased risk of early-onset CRC (OR, 1.24; 95% CI, 1.09 to 1.41), after adjusting for the matching factors and a range of potential confounders, including full time employment status, residence, type of commercial health insurance, Charlson Comorbidity Index, IBD, obesity, family history of gastrointestinal cancer, personal history of screening colonoscopy, other colonoscopy, and FOBT. The positive association remained similar when we restricted the analysis to individuals without IBD, without family history of gastrointestinal cancer, without previous colonoscopy/FOBT, or without a list of early signs/symptoms of CRC. Secondary analyses of early-onset CRC according to the severity of type 2 diabetes revealed a significant positive association for uncontrolled (OR, 1.37; 95% CI, 1.12 to 1.67) and complicated type 2 diabetes (OR, 1.59; 95% CI, 1.08 to 2.35). However, there was no association between controlled type 2 diabetes and risk of early-onset CRC (OR, 1.13; 95% CI, 0.94 to 1.36).

We further evaluated the association between type 2 diabetes and early-onset CRC according to anatomical location of the CRC (**table 3**). The positive association was largely driven by proximal colon (OR, 1.35; 95% CI, 1.03 to 1.77) and distal colon cancer (OR, 1.67; 95% CI, 1.30 to 2.15) rather than by rectal cancer (OR, 1.13; 95% CI, 0.92 to 1.40).

In stratified analyses, the associations between type 2 diabetes and early-onset CRC according to sex (female vs. male), age (18-45 vs. 46-49 years), and geographic region (South vs. others) were similar and no significant interactions were identified (all P >0.05 for interaction). The

association appeared stronger for persons born after 1965 (OR, 1.31; 95% CI, 1.09 to 1.42) compared to those before 1965 (OR, 1.19; 95% CI, 0.99 to 1.42), although no interaction was observed (P = 0.44 for interaction) (**table 4**).

DISCUSSION

In this nested case-control study leveraging real-world claims data with 6001 early-onset CRC cases, we found that type 2 diabetes was associated with a 24% increased risk of developing early-onset CRC compared to individuals without type 2 diabetes. The positive association, largely observed for proximal and distal colon cancer, remained similar when restricted to individuals without IBD, family history of gastrointestinal cancer, previous colonoscopy/FOBT, or early signs/symptoms of CRC. We also found that this association was more pronounced for uncontrolled or complicated type 2 diabetes compared to controlled type 2 diabetes. Our findings suggest that type 2 diabetes contributes, in part, to the rising incidence of early-onset CRC.

A recent systematic review³⁸ estimated that type 2 diabetes was associated with a 27% increased risk of CRC among older individuals. Our study supports a similar association between type 2 diabetes and risk of early-onset CRC. Findings from recent studies attempting to examine the link between type 2 diabetes and early-onset CRC were mixed. While Low *et al* did not observe an association between type 2 diabetes and early-onset CRC, ⁹ Khan *et al* reported that in a Swedish nationwide cohort, type 2 diabetes was associated with increased risk of early-onset CRC.⁸ However, only 31 patients with type 2 diabetes were documented prior to early-onset CRC. Moreover, the prevalence of type 2 diabetes (1.4%) in this analytic cohort was significantly lower than that in the US population. With a total of 302 patients with type 2 diabetes preceding 6,001 early-onset CRC cases, our study is among the first with adequate power to provide a reliable effect estimate for the association between type 2 diabetes and early-onset CRC. Notably, our results are likely

generalizable to the U.S. population as the prevalence of type 2 diabetes among our controls corresponds to the U.S. national data.¹⁷ The similar associations observed across strata of sex, birth year, and geographical regions (South, historically described as the "diabetes belt," ³⁹ vs. others) further ensure the generalizability of our findings.

The underlying mechanisms of the association between type 2 diabetes and CRC are not fully understood. Insulin resistance, hyperglycemia, and hyperinsulinemia may play important roles.⁴⁰⁴¹ Our observed positive association for uncontrolled but not for controlled type 2 diabetes, further supports that hyperglycemia/hyperinsulinemia are likely critical to colorectal carcinogenesis. Impaired insulin receptor activation and subsequent defective PI3K signaling pathway could lead to insulin resistance and hyperinsulinemia,⁴² as well as high levels of insulin-like growth factor (IGF) 1.⁴³ The insulin-PI3K pathway has been shown to have profound effects on cancer initiation⁴² by stimulating colonic mucosal cell growth and sustaining tumor growth. IGFs are key regulators in signal transduction networks that have important roles in neoplasia.⁴⁴ Uncontrolled hyperglycemia in type 2 diabetes might contribute to DNA damage and aberrant RNA expression that could promote carcinogenesis and cancer progression.⁴⁵ Moreover, a dysregulated immune system in type 2 diabetes patients and chronic inflammation accompanied with elevated cytokine level such as interleukin-6, tumor necrosis factor-α, and C-reactive protein could promote CRC tumorigenesis.⁴⁶ Finally, altered host-microbiota crosstalk⁴⁷ and increased colonic transit time in type 2 diabetes have been linked to dysregulation of bile acid metabolism,⁴⁸ which could also contribute to colorectal carcinogenesis.

It is worth noting that the observed associations between type 2 diabetes and early-onset CRC were evident only for colon but not rectal cancer. Emerging evidence indicates that the molecular features of CRC vary by anatomic subsites. Microsatellite instability (MSI), CpG island methylator phenotype (CIMP), and *BRAF* mutation gradually increase from the rectum to the ascending colon.⁴⁹

Rectal cancers also exhibit more *TP53* mutations and fewer *PIK3CA* mutations or *CTNNB1* mutations.⁵⁰ Given the heterogeneous nature of CRC, it is expected that risk factors also differ by anatomic location. In line with our findings in younger adults, a case-control analysis involving 21744 CRCs from Veterans (median age 68 years) reported that diabetes was associated with 29%, 15%, and 12% increased risk of proximal, distal, and rectal cancer, respectively.⁵¹ It is hypothesized that IGF1 activates the PI3K/AKT pathway through *PIK3CA* mutation.⁴² The lower frequency of *PIK3CA* mutations in rectal cancer may in part explain the lack of association between type 2 diabetes and rectal cancer. Further elucidations of mechanisms underlying these differential associations are critical for the development of precision prevention strategies.

Our study has several strengths. This large, nested case-control study leveraged longitudinal claims data from close to half of the U.S. adult population. Such an unprecedented sample size provided a unique opportunity to examine the association of interest. This type of examination between type 2 diabetes and early-onset CRC is not otherwise feasible in existing prospective cohort studies or other real-world EHR-based databases, due to the relatively low prevalence of type 2 diabetes and low incidence of early-onset CRC in younger adults. Our rigorous study design restricted CRC cases to those only with confirmed pathology claims, and type 2 diabetes patients were identified through ICD-9-CM coding followed by an established algorithm to maximize reliability. To minimize potential detection bias from patients presenting with signs/symptoms that directly led to a diagnosis of CRC, we leveraged claims data from 91 days to two years before CRC diagnosis and adjusted for a list of variables associated with detection. We also conducted sensitivity analyses to minimize the influence of comorbidities/symptoms that may have led to differential detection of type 2 diabetes among cases and controls.

The study also has a few limitations. First, we were not able to reliably identify the first date of type 2 diabetes diagnosis and thus could not assess the impact of duration of diabetes. However,

among the prospective studies that allowed such assessment, there were a limited number of CRC cases with diabetes,⁵² and such investigation is largely infeasible for early-onset CRC. Second, residual confounding could not be ruled out due to limited information on confounders such as smoking, alcohol intake, diet, and physical activity. Studies on the associations between these factors and early-onset CRC are limited. Furthermore, these lifestyle factors are likely modest risk factors for type 2 diabetes and early-onset CRC. Indeed, in analyses among an older population, these factors minimally confound the association between type 2 diabetes and CRC of later-onset.⁵³ Third, the MarketScan database does not provide information on race/ethnicity and is restricted to individuals with commercial insurance. Thus, further validation in diverse groups is warranted.

Conclusions

In this large U.S. claims-based nested case-control study, type 2 diabetes was associated with increased risk of early-onset CRC, suggesting that the rising incidence of early-onset CRC may be partially attributed to the surging prevalence of type 2 diabetes. The more pronounced association for uncontrolled or complicated diabetes further highlights the importance of early detection and intervention of type 2 diabetes at younger ages. Our findings lend support to the promise of type 2 diabetes control as an emerging CRC prevention strategy among younger adults.

Table 1.	Characteristic	s of participants	s according to	case and	control s	status,	MarketScan
Commer	cial Database ((2006-2015)	-				

	Cases	Controls*
	(n = 6001)	(n = 52104)
Age at the index dates (y), mean \pm SD	43.0 ± 5.8	42.8 ± 5.8
Female, n (%)	2922 (48.7)	25704 (49.3)
Geographical region, n (%)		
South	2631 (43.8)	22624 (43.4)
Northeast	936 (15.6)	8088 (15.5)
North Central	1380 (23.0)	11744 (22.5)
West	956 (15.9)	8824 (16.9)
Unknown	98 (1.6)	824 (1.6)
Duration of insurance enrollment (y), median (IQR)	3.5 (2.6-5.0)	3.5 (2.6-4.9)
Prescription drug coverage, n (%)	4951 (82.5)	43139 (82.8)
Full time employed, n (%)	3075 (51.2)	27722 (53.2)
Residence, n (%)		
Urban	5005 (83.4)	43937 (84.3)
Rural	902 (15.0)	7370 (14.1)
Unknown	94 (1.6)	797 (1.5)
Health plan, n (%)		
PPO	3881 (64.7)	32 672 (62.7)
НМО	776 (12.9)	7359 (14.1)
Other	1344 (22.4)	12073 (23.2)
Inflammatory bowel disease†, n (%)	287 (4.8)	1372 (2.6)
Obesity†, n (%)	428 (7.1)	2972 (5.7)
Family history of gastrointestinal cancer†, n (%)	75 (1.2)	510 (1.0)
Fecal occult blood test†, n (%)	452 (7.5)	3078 (5.9)
Screening colonoscopy†, n (%)	68 (1.1)	1036 (2.0)
Other colonoscopy†, n (%)	207 (3.4)	914 (1.8)
Charlson Comorbidity Index [‡] , mean ± SD	0.1 ± 0.5	0.06 ± 0.4
Type 2 diabetes†, n (%)	302 (5.0)	1946 (3.7)
Anatomical site§, n (%)		
Colon	3846 (64.1)	-
Proximal colon	1156 (19.3)	-
Distal colon	1115 (18.6)	-
Unspecified colon	1575 (26.2)	-
Rectum	2114 (35.2)	-
Metastatic status , n (%)		
Non-metastatic	4560 (76.0)	-
Metastatic	1441 (24.0)	-

Abbreviations: HMO = Health Maintenance Organization; IQR = interquartile range; PPO = Preferred Provider Organization; SD = standard deviation.

*Controls were matched based on age (18-24 and every 5 years thereafter), sex (female, male), geographical region (Northeast, North Central, South, West, unknown), duration of insurance enrollment before index diagnosis (years), and prescription drug coverage (yes, no). †Between 91 days and 2 years before the index date.

‡Charlson Comorbidity Index was calculated without accounting for diabetes.

§A total of 41 cases with more than one anatomical site were excluded.

||Metastatic colon cancer was determined using treatment encounters of liver or lung metastasis within 3 months after the index dates.

Table 2. Type 2 diabetes and risk of early-onset colorectal cancer

	Participants with type 2 diabetes, No. (%)		Multivariable adjusted	Multivariable adjusted	
	Cases	Controls	OR (95% CI)*	OR (95% CI)†	
Any type 2 diabetes					
All participants	302 (5.0)	1946 (3.7)	1.34 (1.19 to 1.52)	1.24 (1.09 to 1.41)	
Without IBD	282 (4.9)	1862 (3.7)	1.33 (1.17 to 1.52)	1.25 (1.09 to 1.42)	
Without family history of GI cancer	300 (5.1)	1928 (3.7)	1.35 (1.19 to 1.53)	1.25 (1.10 to 1.42)	
Without colonoscopy/FOBT	267 (5.0)	1731 (3.7)	1.36 (1.19 to 1.55)	1.25 (1.09 to 1.43)	
Without early signs/symptoms‡	176 (4.5)	1278 (3.1)	1.39 (1.18 to 1.64)	1.35 (1.14 to 1.59)	
Type 2 diabetes severity§					
Controlled	138 (2.3)	993 (1.9)	1.20 (1.00 to 1.44)	1.13 (0.94 to 1.36)	
Uncontrolled	114 (1.9)	670 (1.3)	1.47 (1.20 to 1.80)	1.37 (1.12 to 1.67)	
Complicated	32 (0.5)	146 (0.3)	1.89 (1.29 to 2.77)	1.59 (1.08 to 2.35)	

Abbreviations: CI = confidence interval; FOBT = fecal occult blood test; GI = gastrointestinal; IBD = inflammatory bowel disease; OR = odds ratio.

*Adjusted for matching factors including age (year), sex (female, male), duration of insurance enrollment (year), region (Northeast, North Central, South, West, unknown), and prescription drug coverage (yes/no).

†Additionally adjusted for employment status (full time/others), residence (urban, rural, unknown), health plan (PPO, HMO, others), Charlson Comorbidity Index without diabetes (continuous), and any of the following conditions between 91 days to 2 years before the index dates: IBD (yes/no), obesity (yes/no), family history of gastrointestinal cancer (yes/no), screening colonoscopy (yes/no), other colonoscopy (yes/no), and fecal occult blood test (yes/no).

‡Early signs/symptoms included any of the following conditions between 91 days to 2 years before the index dates: gastrointestinal bleeding, abdominal pain, anemia, change of bowel habits, diarrhea, constipation, and weight loss. §A total of 18 type 2 diabetes (out of 302) patients among cases and 137 among controls (out of 1946) could not be classified as controlled, uncontrolled, or complicated.

able 3. Type 2 diabetes and risk of ea	ly-onset colorectal cancer accordin	g to anatomical site
--	-------------------------------------	----------------------

	Participants with type 2 diabetes, No. (%)		Multivariable adjusted	Multivariable adjusted	
	Cases	Controls	OR (95% CI)	OR (95% CI)	
Colon cancer‡	204 (5.3)	1946 (3.7)	1.43 (1.23 to 1.66)	1.32 (1.13 to 1.54)	
Proximal colon	60 (5.2)	1946 (3.7)	1.48 (1.14 to 1.93)	1.35 (1.03 to 1.77)	
Distal colon	72 (6.5)	1946 (3.7)	1.72 (1.35 to 2.19)	1.67 (1.30 to 2.15)	
Unspecified colon	72 (4.6)	1946 (3.7)	1.19 (0.94 to 1.52)	1.06 (0.83 to 1.36)	
Rectal cancer‡	94 (4.6)	1946 (3.7)	1.22 (0.99 to 1.50)	1.13 (0.92 to 1.40)	

Abbreviations: CI = confidence interval; OR = odds ratio.

*Adjusted for the same set of covariates as model * in Table 2.

†Adjusted for the same set of covariates as model † in Table 2.

‡A total of 41 cases with more than one anatomical site were excluded.

	Participants with type 2 diabetes, No. (%)		Multivariable adjusted	P for	
	Cases	Controls	OR (95% CI) *	interaction†	
Sex					
Female	135 (4.6)	873 (3.4)	1.26 (1.04 to 1.52)	0.92	
Male	167 (5.4)	1073 (4.1)	1.23 (1.03 to 1.46)		
Age at the index date					
≤ 45	96 (3.3)	590 (2.3)	1.28 (1.02 to 1.60)	0.66	
46-50	206 (6.7)	1356 (5.0)	1.25 (1.07 to 1.45)		
Birth year					
≤ 1965	152 (6.3)	1005 (5.0)	1.19 (0.99 to 1.42)	0.44	
> 1965	150 (4.2)	941 (2.9)	1.31 (1.09 to 1.57)		
Geographical region					
South	155 (5.9)	994 (4.4)	1.26 (1.06 to 1.51)	0.99	
Others	147 (4.4)	952 (3.2)	1.23 (1.03 to 1.48)		

Table 4. Stratified analyses of type 2 diabetes and risk of early-onset colorectal cancer

Abbreviations: CI = confidence interval; OR = odds ratio.

*Adjusted for the same covariates as the model † in Table 2 without the stratification factor.

†P for interaction was calculated by Wald test using the cross-product term of type 2 diabetes and each stratification factor.

REFERENCES

- 1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. *CA: a Cancer Journal for Clinicians* 2021;71(1):7-33.
- Welch HG, Robertson DJ. Colorectal Cancer on the Decline--Why Screening Can't Explain It All. The New England journal of medicine 2016;374(17):1605-7. doi: 10.1056/NEJMp1600448 [published Online First: 2016/04/28]
- Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *Journal of the National Cancer Institute* 2017;109(8) doi: 10.1093/jnci/djw322 [published Online First: 2017/04/05]
- 4. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin* 2020 doi: 10.3322/caac.21601 [published Online First: 2020/03/07]
- 5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70(1):7-30. doi: 10.3322/caac.21590 [published Online First: 2020/01/09]
- 6. Liu P-H, Wu K, Ng K, et al. Association of obesity with risk of early-onset colorectal cancer among women. *JAMA oncology* 2019;5(1):37-44.
- 7. Nguyen LH, Liu P-H, Zheng X, et al. Sedentary behaviors, TV viewing time, and risk of youngonset colorectal cancer. *JNCI cancer spectrum* 2018;2(4):pky073.
- Khan UA, Fallah M, Tian Y, et al. Personal history of diabetes as important as family history of colorectal cancer for risk of colorectal cancer: a nationwide cohort study. *American Journal of Gastroenterology* 2020;115(7):1103-09.
- 9. Low EE, Demb J, Liu L, et al. Risk factors for early-onset colorectal cancer. *Gastroenterology* 2020;159(2):492-501. e7.
- Gonzalez N, Prieto I, del Puerto-Nevado L, et al. 2017 update on the relationship between diabetes and colorectal cancer: epidemiology, potential molecular mechanisms and therapeutic implications. *Oncotarget* 2017;8(11):18456-85. doi: 10.18632/oncotarget.14472
- Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. Journal of the National Cancer Institute 2005;97(22):1679-87. doi: 10.1093/jnci/dji375 [published Online First: 2005/11/17]
- 12. Ling S, Brown K, Miksza JK, et al. Association of type 2 diabetes with cancer: A meta-analysis with bias analysis for unmeasured confounding in 151 cohorts comprising 32 million people. *Diabetes Care* 2020;43(9):2313-22.
- 13. Willauer AN, Liu Y, Pereira AA, et al. Clinical and molecular characterization of early , onset colorectal cancer. *Cancer* 2019;125(12):2002-10.
- 14. Hu FB, Li TY, Colditz GA, et al. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. *Jama* 2003;289(14):1785-91.
- Eckel RH, Kahn SE, Ferrannini E, et al. Obesity and type 2 diabetes: what can be unified and what needs to be individualized? *The Journal of Clinical Endocrinology & Metabolism* 2011;96(6):1654-63.
- Mayer-Davis EJ, Lawrence JM, Dabelea D, et al. Incidence Trends of Type 1 and Type 2 Diabetes among Youths, 2002-2012. N Engl J Med 2017;376(15):1419-29. doi: 10.1056/NEJMoa1610187 [published Online First: 2017/04/14]
- Menke A, Casagrande S, Geiss L, et al. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *Jama* 2015;314(10):1021-9. doi: 10.1001/jama.2015.10029 [published Online First: 2015/09/09]
- Dabelea D, Mayer-Davis EJ, Saydah Š, et al. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *Jama* 2014;311(17):1778-86. doi: 10.1001/jama.2014.3201 [published Online First: 2014/05/06]

- Xu G, Liu B, Sun Y, et al. Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study. *BMJ* 2018;362:k1497. doi: 10.1136/bmj.k1497 [published Online First: 2018/09/06]
- Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14(2):88-98. doi: 10.1038/nrendo.2017.151 [published Online First: 2017/12/09]
- 21. Lascar N, Brown J, Pattison H, et al. Type 2 diabetes in adolescents and young adults. *The lancet Diabetes & endocrinology* 2018;6(1):69-80. doi: 10.1016/S2213-8587(17)30186-9 [published Online First: 2017/08/30]
- Cheng YJ, Kanaya AM, Araneta MRG, et al. Prevalence of Diabetes by Race and Ethnicity in the United States, 2011-2016. JAMA 2019;322(24):2389-98. doi: 10.1001/jama.2019.19365 [published Online First: 2019/12/21]
- 23. Campbell PT. The role of diabetes and diabetes treatments in colorectal cancer mortality, incidence, and survival. *Current Nutrition Reports* 2013;2(1):37-47.
- 24. Siegel RL, Torre LA, Soerjomataram I, et al. Global patterns and trends in colorectal cancer incidence in young adults. *Gut* 2019;68(12):2179-85.
- 25. Hillier TA, Pedula KL. Complications in young adults with early-onset type 2 diabetes: losing the relative protection of youth. *Diabetes Care* 2003;26(11):2999-3005. doi: 10.2337/diacare.26.11.2999 [published Online First: 2003/10/28]
- 26. IBM Watson Health. White paper: IBM MarketScan Research Databases for health services researchers. 2018
- Aizcorbe A, Liebman E, Pack S, et al. Measuring health care costs of individuals with employersponsored health insurance in the U.S.: A comparison of survey and claims data. *Statistical journal of the IAOS* 2012;28(1-2):43-51. doi: 10.3233/sji-2012-0743 [published Online First: 2012/01/01]
- 28. Institute TC. Pathology/lab coding alert: receive fair payment for physcian interpretation of pap smear. *Durham: The Coding Institute* 2001
- Anaya DA, Becker NS, Richardson P, et al. Use of administrative data to identify colorectal liver metastasis. J Surg Res 2012;176(1):141-6. doi: 10.1016/j.jss.2011.07.022 [published Online First: 2011/10/04]
- 30. Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS) 2015. US Agency for Healthcare Research and Quality; 2015, 2014.
- 31. Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. J Clin Epidemiol 2000;53(12):1258-67. doi: 10.1016/s0895-4356(00)00256-0 [published Online First: 2001/01/09]
- 32. Siegel RL, Medhanie GA, Fedewa SA, et al. State variation in early-onset colorectal cancer in the United States, 1995–2015. *JNCI: Journal of the National Cancer Institute* 2019;111(10):1104-06.
- 33. Johnson J, Bowker S, Richardson K, et al. Time-varying incidence of cancer after the onset of type 2 diabetes: evidence of potential detection bias. *Diabetologia* 2011;54(9):2263.
- 34. Gibson DM. Frequency and predictors of missed visits to primary care and eye care providers for annually recommended diabetes preventive care services over a two-year period among US adults with diabetes. *Preventive Medicine* 2017;105:257-64.
- 35. Association AH. AHA coding clinic for ICD-9-CM. AHA, Chicago 2008
- Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases* 1987;40(5):373-83.
- 37. Hamilton W, Round A, Sharp D, et al. Clinical features of colorectal cancer before diagnosis: a population-based case–control study. *British journal of cancer* 2005;93(4):399-405.

- Tsilidis KK, Kasimis JC, Lopez DS, et al. Type 2 diabetes and cancer: umbrella review of metaanalyses of observational studies. *BMJ* 2015;350:g7607. doi: 10.1136/bmj.g7607 [published Online First: 2015/01/06]
- 39. Geiss LS, Kirtland K, Lin J, et al. Changes in diagnosed diabetes, obesity, and physical inactivity prevalence in US counties, 2004-2012. *PLoS One* 2017;12(3)
- 40. Giovannucci E. Insulin and colon cancer. *Cancer Causes & Control* 1995;6(2):164-79.
- 41. McKeown-Eyssen G. Epidemiology of colorectal cancer revisited: are serum triglycerides and/or plasma glucose associated with risk? *Cancer Epidemiology and Prevention Biomarkers* 1994;3(8):687-95.
- 42. Hopkins BD, Goncalves MD, Cantley LC. Insulin-PI3K signalling: an evolutionarily insulated metabolic driver of cancer. *Nat Rev Endocrinol* 2020 doi: 10.1038/s41574-020-0329-9 [published Online First: 2020/03/05]
- 43. Ma J, Pollak MN, Giovannucci E, et al. Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. *J Natl Cancer Inst* 1999;91(7):620-5. doi: 10.1093/jnci/91.7.620 [published Online First: 1999/04/15]
- 44. Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. *Nature Reviews Cancer* 2008;8(12):915-28.
- 45. Ryu TY, Park J, Scherer PE. Hyperglycemia as a risk factor for cancer progression. *Diabetes Metab J* 2014;38(5):330-6. doi: 10.4093/dmj.2014.38.5.330 [published Online First: 2014/10/29]
- 46. Donath MY, Dinarello CA, Mandrup-Poulsen T. Targeting innate immune mediators in type 1 and type 2 diabetes. *Nat Rev Immunol* 2019;19(12):734-46. doi: 10.1038/s41577-019-0213-9 [published Online First: 2019/09/11]
- 47. Al-Obaide MAI, Singh R, Datta P, et al. Gut Microbiota-Dependent Trimethylamine-N-oxide and Serum Biomarkers in Patients with T2DM and Advanced CKD. *Journal of clinical medicine* 2017;6(9) doi: 10.3390/jcm6090086 [published Online First: 2017/09/20]
- Ahmad TR, Haeusler RA. Bile acids in glucose metabolism and insulin signalling mechanisms and research needs. *Nat Rev Endocrinol* 2019;15(12):701-12. doi: 10.1038/s41574-019-0266-7 [published Online First: 2019/10/17]
- 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(6):847-54. doi: 10.1136/gutjnl-2011-300865 [published Online First: 2012/03/20]
- 50. Loree JM, Pereira AA, Lam M, et al. Classifying colorectal cancer by tumor location rather than sidedness highlights a continuum in mutation profiles and consensus molecular subtypes. *Clinical Cancer Research* 2018;24(5):1062-72.
- 51. Demb J, Earles A, Martínez ME, et al. Risk factors for colorectal cancer significantly vary by anatomic site. *BMJ open gastroenterology* 2019;6(1):e000313. doi: 10.1136/bmjgast-2019-000313 [published Online First: 2019/09/17]
- 52. Ma Y, Yang W, Song M, et al. Type 2 diabetes and risk of colorectal cancer in two large U.S. prospective cohorts. *British journal of cancer* 2018;119(11):1436-42. doi: 10.1038/s41416-018-0314-4 [published Online First: 2018/11/08]
- 53. Hu Y, Zhang X, Ma Y, et al. Incident type 2 diabetes duration and cancer risk: a prospective study in two US cohorts. *JNCI: Journal of the National Cancer Institute* 2020

Conflict of Interest: ATC previously served as a consultant for Janssen Pharmaceuticals, Pfizer, Inc, and Bayer Pharma AG for work unrelated to the topic. YC previously served as a consultant for Geneoscopy for work unrelated to the topic. The remaining authors disclose no conflicts.

Funding: This work was supported by the National Institutes of Health (NIH) grant P30CA091842. The Center for Administrative Data Research is supported in part by the Washington University Institute of Clinical and Translational Sciences grant UL1 TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the NIH and R24 HS19455 through the Agency for Healthcare Research and Quality (AHRQ). CDF is supported by T32 DK007130. XBZ was supported by the International Program for PhD candidates, Sun Yat-sen University. ATC is Stuart and Suzanne Steele MGH Research Scholar.

The role of the funder: The funding sources had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the article for publication.

Author contributions:

Drs. Li, Chen, Fritz, and Cao 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.

Study concept and design: ZL, HC, CF, XBZ, MAO, YC.

Acquisition of data: KBN, AT, MAO, YC.

Analysis and interpretation of data: All authors.

Drafting of the manuscript: ZL, HC, CF, XBZ, YC.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: ZL, HC, XYZ, YC.

Obtained funding: YC

Administrative, technical, or material support: YC.

Study supervision: YC.

All authors approved the final submitted draft of the manuscript.