Gastro Hep Advances 2022;1:186–193

ORIGINAL RESEARCH—CLINICAL

Type 2 Diabetes and Risk of Early-Onset Colorectal Cancer

Zitong Li,[1,*] Hanyu Chen,[1,*] Cassandra D. L. Fritz,[2,*] Xiaobin Zheng,[1] Xiaoyu Zong,[1] Katelin B. Nickel,[3] Andrew Tipping,[3] Long H. Nguyen,[4,5] Andrew T. Chan,[4,5,6,7] Edward L. Giovannucci,[8,9] Graham A. Colditz,[1,10] Margaret A. Olsen,[1,3] Peter T. Campbell,[11] Nicholas O. Davidson,[2] Ryan C. Fields,[10,12] and Yin Cao[1,10]

[1]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri, [2]Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, [3]Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St Louis, Missouri, [4]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, [5]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, [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, Massachusetts, [8]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, [9]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, [10]Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri, [11]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, and [12]Department of Surgery, Washington University School of Medicine, St. Louis, Missouri

BACKGROUND AND AIMS: 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. **METHODS:** 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 was identified by the International Classification of Diseases, ninth Revision, Clinical Modification 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 International Classification of Diseases, ninth Revision, Clinical Modification using the Klabunde algorithm. Multivariable logistic regressions were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs). **RESULTS:** A total of 6001 early-onset CRC and 52,104 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% CI: 1.09–1.41). The positive association was more pronounced for uncontrolled (OR = 1.37; 95% CI: 1.12–1.67) or complicated (OR = 1.59, 95% CI: 1.08–2.35) type 2 diabetes compared with controlled diabetes (OR = 1.13, 95% CI: 0.94–1.36). **CONCLUSION:** Individuals with type 2 diabetes have a higher risk of early-onset CRC, with stronger associations for uncontrolled diabetes and complicated diabetes. The rising prevalence of type 2 diabetes among younger adults may partially contribute to the increasing incidence of early-onset CRC.

Keywords: Diabetes; Colorectal Cancer; Early-onset; Epidemiology

Introduction

Colorectal cancer (CRC) is the second most common cause of cancer-related death in the United States.[1] Because of increased CRC screening, the incidence and mortality rates of CRC have declined for several decades among adults aged 50 years and older. In contrast, the incidence and mortality of early-onset CRC (diagnosis before age 50 years) have been increasing since the mid-1990s.[2] During 2012–2016, the incidence of proximal colon, distal colon, and rectal cancer all rose at 1.8% annually among adults younger than 50 years.[2] 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.[2] Further elucidation of risk factors that contributed to this increase is pivotal.

Thus far, obesity[3] and sedentary lifestyle[4] 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 fully elucidated.[5–7] Although the association between type 2 diabetes and average to late-onset CRC is established,[8] the emerging molecular characteristics of early-onset CRC[9] 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,[10,11] further lends support to

**Contributed equally as first authors.*

Abbreviations used in this paper: **CI, confidence interval; CRC, colorectal cancer; FOBT, fecal occult blood testing; IBD, inflammatory bowel disease; ICD-9-CM, International Classification of Diseases, ninth Revision, Clinical Modification; IGF, insulin-like growth factor; OR, odds ratio.**

Most current article

2772-5723
https://doi.org/10.1016/j.gastha.2021.10.009

this need. In addition, type 2 diabetes, if etiologically relevant to early-onset CRC, likely contributes to the rising incidence of early-onset CRC because of the paralleled increase of type 2 diabetes[12–14] and the rise in early-onset CRC.[15] Specifically in the United States, 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.[13] In addition, when compared with older adults with type 2 diabetes, type 2 diabetes before age 45 appears to be a more aggressive disease with an increased risk of requiring insulin.[16] 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 early-onset 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 United States, 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, deidentified, individual-level health care claims database which comprised more than 113 million commercially insured U.S. adults younger than age of 65 years.[17] The database captures information on outpatient and inpatient insurance-reimbursable services, prescription data, type of health plan, and demographic information. Compared with nonelderly people with employer-sponsored insurance, MarketScan enrollees have a similar age and sex distribution.[18] This study was exempt from institutional review board approval for its deidentified limited data set analysis. All authors had access to the study data and approved the final manuscript.

Ascertainment of Cases and Controls

All adults with an incident diagnosis of CRC between ages 18 and 49 years were considered as early-onset CRC and were identified by an International Classification of Diseases, ninth 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. As confirmed pathology diagnoses are automated into pathology ICD-9-CM diagnosis codes and are considered accurate in reporting pathology findings,[19] additional codes to confirm pathology were not required. We restricted our analyses to adults with at least 2 years of enrollment before the index dates, as well as 90 days of enrollment after the index dates to derive metastatic status. By requiring 2 years of data before index CRC and up to 90 days of data after, the index dates included were between 1/1/2008 and 6/30/2015. 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.[20] We excluded patients with CRC and with any prior/concurrent cancer history (V10.x) or genetic susceptibility to malignant neoplasm (V84.0x), as well as patients with CRC and with any cancer except nonmelanoma skin cancer identified through Healthcare Cost and Utilization Project's Clinical Classification Software (HCUP CCS)[21] within 2 years before the index dates. The identification of CRC as cases or the other cancers for exclusion for both cases and controls was based on the study by Klabunde et al,[22] which requires at least one inpatient facility claim and/or 2 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 5 years thereafter), sex (female, male), geographical region (Northeast, North Central, South, West, unknown) due to geographic variations in the incidence of early-onset CRC,[23] 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 also had at least 2 years of enrollment before the 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,[24] we restricted our exposures and covariates from 91 days to 2 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 health care provider within a 2-year period.[25] 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,[22] which has been used to identify type 2 diabetes in claims data.[26] Type 2 diabetes was further classified as controlled/not stated as uncontrolled without complications (with 250.00, 250.10, 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),[27] or complicated (250.40, 250.42, 250.50, 250.52, 250.60, 250.62, 250.70, 250.72, 250.80, 250.82, 250.90, 250.92).[21]

Assessment of Covariates and Other Clinical Information

We extracted demographic information, including employment status, urban/rural residence, geographical region, and health plan, and derived the Charlson comorbidity index.[28] In addition, the Charlson comorbidity index (without diabetes and cancer except nonmelanoma skin cancer) was calculated for both cases and controls. We extracted information on potential confounders, including inflammatory bowel diseases (IBDs), obesity, and family history of gastrointestinal cancer between

91 days and 2 years before 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 prespecified 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 before the index dates.[29]

Statistical Analysis

To evaluate the association between type 2 diabetes and the risk of early-onset CRC, multivariable logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs). We first adjusted for matching factors[30] 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 prespecified early signs/symptoms of CRC. We conducted sensitivity analyses by restricting to diabetes from 1 year up to 2 years before the index dates. We also examined whether the association differed according to the severity of type 2 diabetes (controlled, uncontrolled, and complicated).

We further examined if the association between type 2 diabetes and early-onset CRC differed as per the 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 (≤1965 vs >1965), and geographical region (South vs others). The *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, NC). All the statistical tests were 2-sided, and *P* values <.05 were considered statistically significant.

Results

A total of 6001 early-onset CRC cases and 52,104 controls were included in the analyses (Table 1). The mean age of patients with early-onset CRC 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. Our analysis included 2248 individuals with type 2 diabetes; the majority (n = 1934) had claims for diabetes starting between 1 year and 2 years before early-onset CRC diagnosis.

Type 2 diabetes was present in 5.0% of patients with early-onset CRC, compared with 3.7% among the controls (Table 2). In comparison with those without type 2 diabetes, individuals with type 2 diabetes had a 24% increased risk of early-onset CRC (OR = 1.24, 95% CI: 1.09–1.41), 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. When we restricted to diabetes from 1 year to up to 2 years before the index, the association with early-onset CRC remained similar with an OR of 1.22 (95% CI 1.07–1.40). Secondary analyses of early-onset CRC as per the severity of type 2 diabetes revealed a significant positive association for uncontrolled (OR = 1.37, 95% CI: 1.12–1.67) and complicated type 2 diabetes (OR = 1.59; 95% CI: 1.08–2.35). However, there was no association between controlled type 2 diabetes and risk of early-onset CRC (OR = 1.13; 95% CI: 0.94–1.36).

We further evaluated the association between type 2 diabetes and early-onset CRC as per anatomical location of the CRC (Table 3). The positive association was largely driven by proximal colon (OR = 1.35, 95% CI: 1.03–1.77) and distal colon cancer (OR = 1.67, 95% CI: 1.30–2.15) while not apparent for rectal cancer (OR = 1.13, 95% CI: 0.92–1.40).

In stratified analyses, the associations between type 2 diabetes and early-onset CRC as per 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 $P > .05$ for interaction). The association appeared stronger for persons born after 1965 (OR = 1.31, 95% CI: 1.09–1.42) than that for those before 1965 (OR = 1.19, 95% CI: 0.99–1.42), although no interaction was observed (*P*-interaction = .44) (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 with individuals without type 2 diabetes. The positive association, largely observed for proximal and distal colon cancer, remained after restricting to individuals without IBD, family history of gastrointestinal cancer, previous colonoscopy/FOBT, or early signs/symptoms of CRC. We found that this association was more pronounced for uncontrolled or complicated type 2 diabetes than for 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 review[31] estimated that type 2 diabetes was associated with a 27% increased risk of CRC

Table 1. Characteristics of Participants as Per Case and Control Status, MarketScan Commercial Database (2006–2015)

Characteristics	Cases ($N = 6001$)	Controls[a] ($N = 52,104$)
Age at the index dates (y), mean ± SD	43.0 ± 5.8	42.8 ± 5.8
Female, *n* (%)	2922 (48.7)	25,704 (49.3)
Geographical region, *n* (%)		
South	2631 (43.8)	22,624 (43.4)
Northeast	936 (15.6)	8088 (15.5)
North Central	1380 (23.0)	11,744 (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)	43,139 (82.8)
Full time employed, *n* (%)	3075 (51.2)	27,722 (53.2)
Residence, *n* (%)		
Urban	5005 (83.4)	43,937 (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)
HMO	776 (12.9)	7359 (14.1)
Other	1344 (22.4)	12,073 (23.2)
Inflammatory bowel disease[b], *n* (%)	287 (4.8)	1372 (2.6)
Obesity[b], *n* (%)	428 (7.1)	2972 (5.7)
Family history of gastrointestinal cancer[b], *n* (%)	75 (1.2)	510 (1.0)
Fecal occult blood test[b], *n* (%)	452 (7.5)	3078 (5.9)
Screening colonoscopy[b], *n* (%)	68 (1.1)	1036 (2.0)
Other colonoscopy[b], *n* (%)	207 (3.4)	914 (1.8)
Charlson comorbidity index[c], mean ± SD	0.1 ± 0.5	0.06 ± 0.4
Anatomical site[d], *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[e], *n* (%)		
Non-metastatic	4560 (76.0)	–
Metastatic	1441 (24.0)	–

HMO, health maintenance organization; IQR, interquartile range; PPO, preferred provider organization; SD, standard deviation.
[a]Controls were matched based on age (18–24 years 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).
[b]Between 91 days and 2 years before the index date.
[c]The Charlson comorbidity index was calculated without diabetes and cancer.
[d]A total of 41 cases with more than one anatomical site were excluded.
[e]Metastatic colon cancer was determined using treatment encounters of liver or lung metastasis within 3 months after the index dates.

among older individuals. Our study supports a similar association between type 2 diabetes and the risk of early-onset CRC. Findings from recent studies attempting to examine the link between type 2 diabetes and early-onset CRC were mixed. In determining the association between metabolic risk and early-onset CRC, Schumacher et al and Low et al did not observe an association between type 2 diabetes and early-onset CRC, but these studies were limited by the number of patients with type 2 diabetes.[6,7] However, a Swedish nationwide cohort reported that type 2 diabetes was associated with increased risk of early-onset CRC.[5] Similarly, a recent larger-scale pooled analysis also reported a suggestive association between type 2 diabetes and early-onset CRC with an OR of 1.25 (95% CI 0.93–1.68), including 168 incident early-onset CRC cases among individuals who had type 2 diabetes.[32] With a total of 302 patients with type 2 diabetes preceding 6001 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

Table 2. Type 2 Diabetes and Risk of Early-onset Colorectal Cancer

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

GI, gastrointestinal.
[a]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).
[b]Additionally adjusted for employment status (full time/others), residence (urban, rural, unknown), health plan (PPO, HMO, others), Charlson comorbidity index without diabetes and cancer (continuous), and any of the following conditions between 91 days and 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).
[c]Early signs/symptoms included any of the following conditions between 91 days and 2 years before the index dates: gastrointestinal bleeding, abdominal pain, anemia, change of bowel habits, diarrhea, constipation, and weight loss.
[d]A total of 18 patients with type 2 diabetes (out of 302) among cases and 137 among controls (out of 1946) could not be classified as controlled, uncontrolled, or complicated.

corresponds to the U.S. national data.[13] The similar associations observed across strata of sex, birth year, and geographical regions (South, historically described as the "diabetes belt",[33] 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. 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,[34] as well as high levels of insulin-like growth factor (IGF) 1.[35] The insulin-PI3K pathway has been shown to have profound effects on cancer initiation[34] by stimulating colonic mucosal cell growth and sustaining tumor growth. IGFs are key regulators in signal transduction networks that have important roles in neoplasia.[35] Uncontrolled hyperglycemia in type 2 diabetes might contribute to DNA damage and aberrant RNA expression that could promote carcinogenesis and cancer progression.[36] Moreover, a dysregulated immune system in patients with type 2 diabetes and chronic inflammation accompanied with elevated cytokine levels such as interleukin-6, tumor necrosis factor-α, and C-reactive protein could promote CRC tumorigenesis.[37] Finally, altered host-microbiota cross talk[38] and increased colonic transit time in type 2 diabetes have been linked to dysregulation of bile acid metabolism,[39] which could also contribute to colorectal carcinogenesis.

The more apparent associations between type 2 diabetes and early-onset colon compared with rectal cancer require validation. Emerging evidence indicates that the molecular

Table 3. Type 2 Diabetes and Risk of Early-onset Colorectal Cancer as Per Anatomical Site

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

[a]Adjusted for the same set of covariates as model[a] in Table 2.
[b]Adjusted for the same set of covariates as model[b] in Table 2.
[c]A total of 41 cases with more than one anatomical site were excluded.

Table 4. Stratified Analyses of Type 2 Diabetes and Risk of Early-onset Colorectal Cancer

Characteristics	Participants with type 2 diabetes, *N* (%)		Multivariable adjusted OR (95% CI)[a]	*P* for interaction[b]
	Cases	Controls		
Sex				
Female	135 (4.6)	873 (3.4)	1.26 (1.04–1.52)	.92
Male	167 (5.4)	1073 (4.1)	1.23 (1.03–1.46)	
Age at the index date				
≤45	96 (3.3)	590 (2.3)	1.28 (1.02–1.60)	.66
46–50	206 (6.7)	1356 (5.0)	1.25 (1.07–1.45)	
Birth year				
≤1965	152 (6.3)	1005 (5.0)	1.19 (0.99–1.42)	.44
>1965	150 (4.2)	941 (2.9)	1.31 (1.09–1.57)	
Geographical region				
South	155 (5.9)	994 (4.4)	1.26 (1.06–1.51)	.99
Others	147 (4.4)	952 (3.2)	1.23 (1.03–1.48)	

[a]Adjusted for the same covariates as the model[b] in Table 2 without the stratification factor.
[b]*P* for interaction was calculated by the Wald test using the cross-product term of type 2 diabetes and each stratification factor.

features of CRC vary by anatomic subsites. Microsatellite instability, CpG island methylator phenotype, and *BRAF* mutation gradually increase from the rectum to the ascending colon.[40] Rectal cancers also exhibit more *TP53* mutations and fewer *PIK3CA* mutations or *CTNNB1* mutations.[41] 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 21,744 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.[42] It is hypothesized that IGF1 activates the PI3K/AKT pathway through the *PIK3CA* mutation.[34] 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 electronic health record-based databases because of 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 patients with type 2 diabetes 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 2 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,[43] and such investigation is largely infeasible for early-onset CRC. Second, residual confounding could not be ruled out because of limited information on putative confounders such as tobacco use, alcohol intake, diet, and physical activity. However, in analyses among older populations, these factors minimally confound the association between type 2 diabetes and CRC of later-onset.[44] In addition, the associations between these factors with early-onset CRC remain to be elucidated. For instance, a recent analysis assessing nongenetic risk factors for early-onset CRC did not find an association with tobacco use.[32] Third, we acknowledge the percent of obese patients in our sample was low because of under coding in claims data. However, in claims data, diagnostic codes for obesity primarily captured morbid obesity with high positive predictive values[45] and have been widely used to define obesity in studies assessing surgical outcomes.[46] In addition, type 2 diabetes is in the causal pathway between obesity and CRC, or in other words, a major link between obesity and CRC is primarily through insulin resistance and subsequent type 2 diabetes.[47] Obesity thus serves as an instrumental variable, not a confounder, for the association between type 2 diabetes and early-onset CRC. Finally, 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.

In conclusion, 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.

References

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2021. CA Cancer J Clin 2021;71:7–33.
2. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020;70:145–164.
3. Liu P-H, Wu K, Ng K, et al. Association of obesity with risk of early-onset colorectal cancer among women. JAMA Oncol 2019;5:37–44.
4. Nguyen LH, Liu P-H, Zheng X, et al. Sedentary behaviors, TV viewing time, and risk of young-onset colorectal cancer. JNCI Cancer Spectr 2018;2:pky073.
5. 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. Am J Gastroenterol 2020;115:1103–1109.
6. Low EE, Demb J, Liu L, et al. Risk factors for early-onset colorectal cancer. Gastroenterology 2020;159:492–501.e7.
7. Schumacher AJ, Chen Q, Attaluri V, et al. Metabolic risk factors associated with early-onset colorectal adenocarcinoma: a case–control study at Kaiser Permanente Southern California. Cancer Epidemiol Biomark Prev 2021;30:1792–1798.
8. Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. J Natl Cancer Inst 2005;97:1679–1687.
9. Willauer AN, Liu Y, Pereira AA, et al. Clinical and molecular characterization of early-onset colorectal cancer. Cancer 2019;125:2002–2010.
10. 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;289:1785–1791.
11. Eckel RH, Kahn SE, Ferrannini E, et al. Obesity and type 2 diabetes: what can be unified and what needs to be individualized? J Clin Endocrinol Metab 2011;96:1654–1663.
12. 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;376:1419–1429.
13. Menke A, Casagrande S, Geiss L, et al. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. JAMA 2015;314:1021–1029.
14. Cheng YJ, Kanaya AM, Araneta MRG, et al. Prevalence of diabetes by race and ethnicity in the United States, 2011-2016. JAMA 2019;322:2389–2398.
15. Siegel RL, Torre LA, Soerjomataram I, et al. Global patterns and trends in colorectal cancer incidence in young adults. Gut 2019;68:2179–2185.
16. Hillier TA, Pedula KL. Complications in young adults with early-onset type 2 diabetes: losing the relative protection of youth. Diabetes Care 2003;26:2999–3005.
17. IBM Watson Health. White paper: IBM MarketScan Research Databases for health services researchers. Cambridge, MA: IBM Watson Health, 2018.
18. Aizcorbe A, Liebman E, Pack S, et al. Measuring health care costs of individuals with employer-sponsored health insurance in the U.S.: a comparison of survey and claims data. Stat J IAOS 2012;28:43–51.
19. Institute TC. Pathology/lab coding alert: receive fair payment for physcian interpretation of pap smear. Durham: The Coding Institute, 2001.
20. Anaya DA, Becker NS, Richardson P, et al. Use of administrative data to identify colorectal liver metastasis. J Surg Res 2012;176:141–146.
21. Elixhauser A, Steiner C, Palmer L. Clinical classifications software (CCS) 2015. Washington, DC: US Agency for Healthcare Research and Quality; 2015, Volume 2014.
22. Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. J Clin Epidemiol 2000;53:1258–1267.
23. Siegel RL, Medhanie GA, Fedewa SA, et al. State variation in early-onset colorectal cancer in the United States, 1995–2015. J Natl Cancer Inst 2019;111:1104–1106.
24. 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;54:2263.
25. 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. Prev Med 2017;105:257–264.
26. Gold HT, Shao H, Oratz R, et al. Association of diabetes and other clinical and sociodemographic factors with guideline-concordant breast cancer treatment for breast cancer. Am J Clin Oncol 2020;43:101–106.
27. Association AH. AHA coding clinic for ICD-9-CM. Chicago: AHA, 2008.
28. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373–383.
29. Hamilton W, Round A, Sharp D, et al. Clinical features of colorectal cancer before diagnosis: a population-based case–control study. Br J Cancer 2005;93:399–405.
30. Pearce N. Analysis of matched case-control studies. BMJ 2016;352:i969.
31. Tsilidis KK, Kasimis JC, Lopez DS, et al. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ 2015;350:g7607.
32. Archambault AN, Lin Y, Jeon J, et al. Nongenetic determinants of risk for early-onset colorectal cancer. JNCI Cancer Spectr 2021;5:pkab029.
33. 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:e0173428.
34. Hopkins BD, Goncalves MD, Cantley LC. Insulin-PI3K signalling: an evolutionarily insulated metabolic driver of cancer. Nat Rev Endocrinol 2020;16:276–283.
35. Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. Nat Rev Cancer 2008;8:915–928.
36. Ryu TY, Park J, Scherer PE. Hyperglycemia as a risk factor for cancer progression. Diabetes Metab J 2014;38:330–336.

37. Donath MY, Dinarello CA, Mandrup-Poulsen T. Targeting innate immune mediators in type 1 and type 2 diabetes. Nat Rev Immunol 2019;19:734–746.
38. 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. J Clin Med 2017;6:86.
39. Ahmad TR, Haeusler RA. Bile acids in glucose metabolism and insulin signalling - mechanisms and research needs. Nat Rev Endocrinol 2019;15:701–712.
40. 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.
41. 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. Clin Cancer Res 2018;24:1062–1072.
42. Demb J, Earles A, Martínez ME, et al. Risk factors for colorectal cancer significantly vary by anatomic site. BMJ Open Gastroenterol 2019;6:e000313.
43. Ma Y, Yang W, Song M, et al. Type 2 diabetes and risk of colorectal cancer in two large U.S. prospective cohorts. Br J Cancer 2018;119:1436–1442.
44. Hu Y, Zhang X, Ma Y, et al. Incident type 2 diabetes duration and cancer risk: a prospective study in two US cohorts. J Natl Cancer Inst 2021;113:381–389.
45. Ammann EM, Kalsekar I, Yoo A, et al. Assessment of obesity prevalence and validity of obesity diagnoses coded in claims data for selected surgical populations: a retrospective, observational study. Medicine (Baltimore) 2019;98:e16438.
46. Morton JM, Ponce J, Malangone-Monaco E, et al. Association of bariatric surgery and national medication use. J Am Coll Surg 2019;228:171–179.
47. Frezza EE, Wachtel MS, Chiriva-Internati M. Influence of obesity on the risk of developing colon cancer. Gut 2006;55:285–291.

Received August 18, 2021. Accepted October 26, 2021.

Correspondence:
Address correspondence to: 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, Missouri 63110. e-mail: yin.cao@wustl.edu.

Authors' Contributions:
Zitong Li, Hanyu Chen, Cassandra D.L. Fritz, and Yin Cao had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Zitong Li, Hanyu Chen, Cassandra D.L. Fritz, Xiaobin Zheng, Margaret A. Olsen, and Yin Cao contributed to study concept and design. Katelin B. Nickel, Andrew Tipping, Margaret A. Olsen, and Yin Cao contributed to acquisition, analysis, or interpretation of data. Zitong Li, Hanyu Chen, Cassandra D.L. Fritz, Xiaobin Zheng, and Yin Cao contributed to drafting of the manuscript. All authors contributed to critical revision of the manuscript for important intellectual content. Zitong Li, Hanyu Chen, Xiaoyu Zong, and Yin Cao contributed to statistical analysis. Yin Cao contributed to administrative, technical, or material support. Yin Cao contributed to study supervision.

Conflicts of Interest:
Andrew T. Chan previously served as a consultant for Janssen Pharmaceuticals, Pfizer, Inc, and Bayer Pharma AG for work unrelated to the topic. Yin Cao 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). C.D.L.F. is supported by T32 DK007130. X.Z. was supported by the International Program for PhD candidates, Sun Yat-sen University. A.T.C. is Stuart and Suzanne Steele MGH Research Scholar. 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.

Ethical Statement:
The corresponding author, on behalf of all authors, jointly and severally, certifies that their institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research.

Data Transparency Statement:
Deidentified individual data are stored in the IBM MarketScan Commercial Database (2006–2015) and cannot be shared.

Preprint Server:
This manuscript has previously been uploaded to Medrxiv. https://doi.org/10.1101/2021.06.02.21257972.