

Colon Cancer Risk Following Intestinal *Clostridioides difficile* Infection: A Longitudinal Cohort Study

David A. Geier^a, Mark R. Geier^{a, b}

Abstract

Background: The gut microbiome may play an important role in the etiology and progression of colon cancer. The present hypothesis-testing study compared the colon cancer incidence rate among adults diagnosed with intestinal *Clostridioides* (formerly *Clostridium*) *difficile* (Cdiff) (the Cdiff cohort) to adults not diagnosed with intestinal Cdiff infection (the non-Cdiff cohort).

Methods: De-identified eligibility and claim healthcare records within the Independent Healthcare Research Database (IHRD) from a longitudinal cohort of adults (the overall cohort) enrolled in the Florida Medicaid system between 1990 through 2012 were examined. Adults with ≥ 8 outpatient office visits over 8 years of continuous eligibility were examined. There were 964 adults in the Cdiff cohort and 292,136 adults in the non-Cdiff cohort. Frequency and Cox proportional hazards models were utilized.

Results: Colon cancer incidence rate in the non-Cdiff cohort remained relatively uniform over the entire study period, whereas a marked increase was observed in the Cdiff cohort within the first 4 years of a Cdiff diagnosis. Colon cancer incidence was significantly increased (about 2.7-fold) in the Cdiff cohort (3.11 per 1,000 person-years) compared to the non-Cdiff cohort (1.16 per 1,000 person-years). Adjustments for gender, age, residency, birthdate, colonoscopy screening, family history of cancer, and personal history of tobacco abuse, alcohol abuse/dependence, drug abuse/dependence, and overweight/obesity, as well as consideration of diagnostic status for ulcerative and infection colitis, immunodeficiency, and personal history of cancer did not significantly change the observed results.

Conclusions: This is the first epidemiological study associating Cdiff with an increased risk for colon cancer. Future studies should further evaluate this relationship.

Keywords: *Clostridioides difficile*; Colon cancer; Intestinal infection; Longitudinal cohort

Introduction

Colon cancer is dependent upon multiple factors, including chronic inflammatory processes, carcinogens, genetic susceptibility, and microbiota [1-3]. The gut microbiome may play an important role in the etiology and progression of colon cancer [4]. Colon cancer is associated with fecal and mucosal dysbiosis [5-7], and gavage with fecal/colonic mucosal slurries from colon cancer patients induced colonic tumors in susceptible mouse model systems [8-10]. A recent study unexpectedly demonstrated the ability of *Clostridioides* (formerly known as *Clostridium*) *difficile* (Cdiff) and its cytotoxin, toxin B (TcdB) to promote colonic carcinogenesis [11].

The present study evaluated the potential relationship between intestinal Cdiff infection and the subsequent colon cancer risk. A longitudinal cohort of adults enrolled in the Florida Medicaid system between 1990 through 2012 was examined in the Independent Healthcare Research Database (IHRD). The incidence rate of colon cancer was compared among adults following an intestinal Cdiff infection (the Cdiff cohort) to adults not diagnosed with intestinal Cdiff infection (the non-Cdiff cohort).

Materials and Methods

IHRD

As described previously in great detail, the IHRD contains de-identified healthcare records from the Florida Medicaid system [12-18]. Unique recipient identifier codes for each person allow for linkage of their eligibility and claim records. The Liberty Institutional Review Board (IRB) (Deland, FL) approved data assembly and accessing of the IHRD. This study was conducted in compliance with the ethical standards of the responsible institution on human subjects as well as with the Helsinki Declaration. The IHRD was examined using the SAS system for Windows, version 9.4 (Cary, NC, USA).

Study participants

A schematic flowchart of the data examined within the IHRD is shown in Figure 1. A cohort of 9,358,645 persons of all ages with eligibility at specific times for Florida Medicaid from

Manuscript submitted April 10, 2023, accepted June 2, 2023
Published online xx xx, 2023

^aResearch Department, Institute of Chronic Illnesses, Inc., Silver Spring, MD, USA
^bCorresponding Author: Mark R. Geier, Research Department, Institute of Chronic Illnesses, Inc., Silver Spring, MD 20905, USA.
Email: mgeier@comcast.net

doi: <https://doi.org/10.14740/jocmr4919>

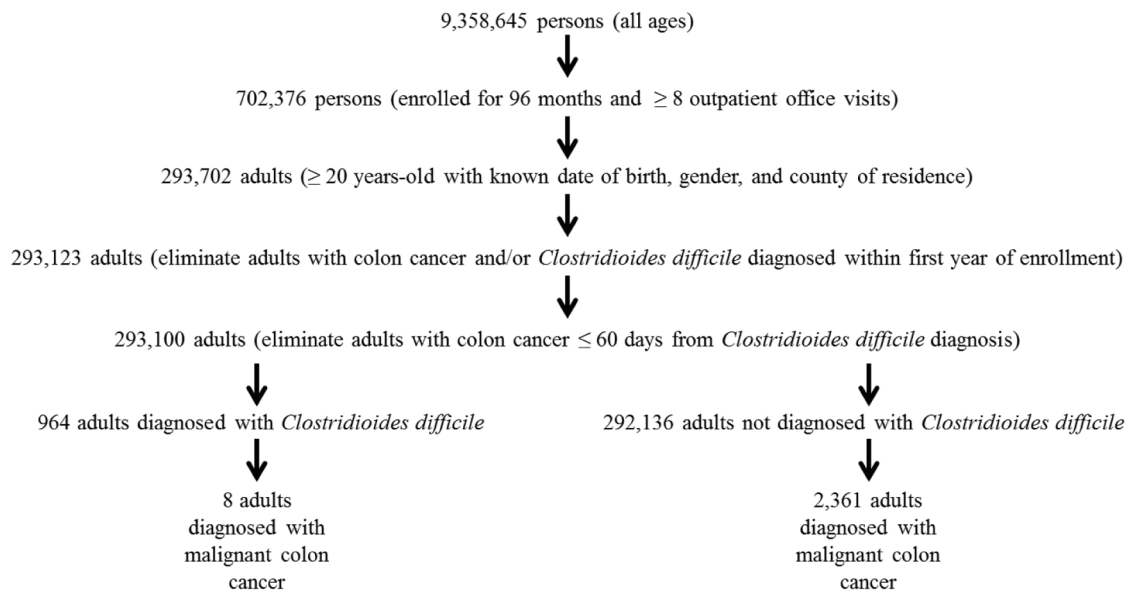


Figure 1. A schematic flowchart of the data examined in the present study.

July 1990 through June 2012 was initially evaluated. Inclusion criteria required that persons examined be ≥ 20 years old with a known date of birth, gender, county of residence at initial enrollment and county of residence at the end of enrollment, and be continuously eligible for Florida Medicaid for 96 months with ≥ 8 outpatient office visits during their enrollment period. This resulted in the cohort being reduced to 293,702 adults.

Cohorts examined

For this study, in order to capture newly diagnosed Cdiff (International Code for Disease, 9th revision (ICD-9) code: 008.45), only adults diagnosed with Cdiff on at least one filed claim after the first year of enrollment were included. Those diagnosed with Cdiff in the first year of enrollment were excluded from the study. Adults without a Cdiff diagnosis were assumed to not have Cdiff and were included in the non-Cdiff cohort.

Outcomes

The outcome variable examined was colon cancer (ICD-9 code: 153.xx). The diagnoses examined were of the hepatitis flexure, transverse colon, descending colon, sigmoid colon, cecum, appendix vermiformis, ascending colon, splenic flexure, other specified sites of large intestine, and unspecified site. The date of onset for colon cancer was the date of service for the first claim in chronological order specifying a colon cancer diagnosis. In order to capture newly diagnosed colon cancer, the study inclusion criteria required adults to be diagnosed with colon cancer on at least one filed claim after the first year of enrollment. Those adults diagnosed with colon cancer in the first year of enrollment were excluded. Adults not meeting the study inclusion criteria for a colon cancer diagnosis were as-

sumed to not have colon cancer. Finally, for the Cdiff cohort, adults diagnosed with colon cancer before their Cdiff (these would not be informative as to the relationship between Cdiff and subsequent colon cancer risk) or colon cancer within 60 days (2 months) after their Cdiff (these were deemed to be too close in temporal proximity to allow for a biological plausible relationship between Cdiff and colon cancer) were excluded. A breakdown of the initial colon cancer diagnoses in the cohorts examined were as follows: hepatitis flexure (Cdiff cohort = 1, non-Cdiff cohort = 81), transverse colon (Cdiff cohort = 0, non-Cdiff cohort = 84), descending colon (Cdiff cohort = 0, non-Cdiff cohort = 64), sigmoid colon (Cdiff cohort = 0, non-Cdiff cohort = 341), cecum (Cdiff cohort = 0, non-Cdiff cohort = 164), appendix vermiformis (Cdiff cohort = 0, non-Cdiff cohort = 14), ascending colon (Cdiff cohort = 0, non-Cdiff cohort = 156), splenic flexure (Cdiff cohort = 0, non-Cdiff cohort = 21), other specified sites of large intestine (Cdiff cohort = 1, non-Cdiff cohort = 124), and unspecified site (Cdiff cohort = 6, non-Cdiff cohort = 1,312).

Colon cancer screening procedures

In light of the potential importance of colon cancer screening procedures to detect colon cancer, the total number of colonoscopy screening procedures was determined for each adult examined. Colonoscopy screening procedures were identified from claims that specified the current procedural terminology (CPT) code of 45378. Colonoscopy screening procedures in the non-Cdiff cohort were counted from the end of the first year of enrollment until the end of the enrollment period for those without colon cancer and from the end of the first year of enrollment until the date of a colon cancer diagnosis for those with colon cancer. Colonoscopy screening procedures were counted in the Cdiff cohort from the initial Cdiff diagnosis un-

til the end of the enrollment period for those without colon cancer and from the initial Cdiff diagnosis until the data of a colon cancer diagnosis for those with colon cancer.

Medical diagnoses potentially influencing colon cancer risk

In light of the potential for previous medical diagnoses to influence the risk for colon cancer [19-22], the diagnostic status for each adult examined was determined for ulcerative colitis (ICD-9 code: 556.xx), infectious colitis (ICD-9 code: 009.0), immunodeficiency (ICD-9 code: 279.0x, 279.1x, 279.2, 279.3, or 279.8), personal history of any types of cancer, except colon cancer (ICD-9 code: 140.xx through 194.xx, 200.xx through 209, and 233.xx), tobacco abuse (ICD-9 code: 305.1x), drug abuse/dependence (ICD-9 code: 304.xx, 305.2x, 305.3x, 305.34, 305.5x, 305.6x, 305.7x, 305.8x or 305.9x), alcohol abuse/dependence (ICD-9 code: 303.xx or 305.0x), overweight/obesity (ICD-9 code: 278.00, 278.01, or 278.02), and family history of any type of cancer (ICD-9 code: V16.xx). In the non-Cdiff cohort, within the period of enrollment examined, the diagnostic status (diagnosed or undiagnosed) for each person was determined for ulcerative colitis, infectious colitis, immunodeficiency, and cancer, except for those diagnoses occurring after a colon cancer diagnosis. In the Cdiff cohort, within the period of enrollment examined, the diagnostic status (diagnosed or undiagnosed) for each person was determined for ulcerative colitis, infectious colitis, immunodeficiency, and cancer, except for those diagnoses occurring after a Cdiff diagnosis. In both the non-Cdiff and Cdiff cohorts, the diagnostic status (diagnosed or undiagnosed) for each person was determined (diagnosed at any time during enrollment) for family history of cancer and personal history of tobacco abuse, drug abuse/dependence, alcohol abuse/dependence, and overweight/obesity.

Statistical analyses

In statistical analyses, the statistical package in SAS and StatsDirect version 3.3.5 (Birkenhead, UK) were utilized, and a two-sided P-value < 0.05 was considered statistically significant. The null hypothesis was that there would be no difference in the incidence rate of colon cancer in the Cdiff cohort as compared to the non-Cdiff cohort. The study size was sufficient to ensure adequate statistical power for the statistical analyses undertaken.

In order to consider differences in follow-up time in the Cdiff cohort as compared to the non-Cdiff cohort, person-years of follow-up for each adult examined were calculated. Follow-up time began in the non-Cdiff cohort at the end of the first year of enrollment. For adults in non-Cdiff cohort without colon cancer, the end of follow-up time was the end of the enrollment period. For adults in the non-Cdiff cohort with colon cancer, the end of follow-up time was the initial date of a colon cancer diagnosis. Follow-up time in the Cdiff cohort began after the initial diagnosis of Cdiff. For adults in the Cdiff cohort

without colon cancer, the end of follow-up time was the end of the enrollment period. For adults in the Cdiff cohort with colon cancer, the end of follow-up time was the initial date of a colon cancer diagnosis.

Frequency modeling using an exact Fisher statistical test in StatsDirect was utilized to evaluate the incidence rate of colon cancer per person-year in the Cdiff cohort as compared to the non-Cdiff cohort. The risk ratio and the attributable rate per 1,000 person-years were also calculated. An assessment of the data examined showed that the power of the present study to detect an association based upon a normal approximation was 82.7%.

Cox proportional hazards models in SAS were also constructed to evaluate the incidence rate of colon cancer per person-year in the Cdiff cohort as compared to the non-Cdiff cohort. The models constructed were unadjusted and adjusted. The overall hazard ratio (HR) for each model constructed was determined.

In adjusted model I, the covariates of gender (categorical variable: female versus male), date of birth in years (continuous variable), age in years at initial enrollment (continuous variable), and county of residency at initial enrollment and at the end of enrollment (continuous variable) were included. There are a total of 67 counties in the state of Florida and each was assigned a numeric value between 1 and 67. The counties are grouped by the state of Florida based upon geographical areas into 11 districts, with each district containing several counties located in a similar geographical area in the state of Florida. In adjusted model II, all of the covariates from model I were included, but an additional covariate was created for colonoscopy screening status. The colonoscopy screening variable was defined as those adults receiving ≥ 1 colonoscopy screening procedures as compared to those adults receiving no colonoscopy screening procedures (categorical variable: yes versus no). In adjusted model III, all of the covariates from model I were included, but an additional covariate was created to reflect the number of colonoscopy screening procedures (continuous variable). In adjusted model IV, all of the covariates from adjusted model III were included, but family history of any type of cancer diagnoses (categorical variable: diagnosed vs. not diagnosed), tobacco abuse (categorical variable: diagnosed vs. not diagnosed), alcohol abuse/dependence (categorical variable: diagnosed vs. not diagnosed), and drug abuse/dependence (categorical variable: diagnosed vs. not diagnosed) were included as covariates. In adjusted model V, all of the covariates and exclusion criteria from adjusted model IV were applied, but anyone counted with diagnoses of ulcerative colitis, infectious colitis, immunodeficiency, or any types of cancer (except for colon cancer) were excluded.

In light of significant reductions in colon cancer survivability within the first 8 years post-colon cancer diagnosis [23], the mean follow-up period from a colon cancer diagnosis until the end of the enrollment period for the Cdiff cohort (1.73 ± 1.03 years, $n = 8$) as compared to the non-Cdiff cohort (3.46 ± 2.16 years, $n = 2,361$) revealed significant differences (unpaired *t*-test statistic, assuming unequal variances, $P < 0.05$). As a result, the initial period (from 1 to 4 years) of excluding adults diagnosed with Cdiff and/or colon cancer for both cohorts was increased. The consequence of this adjustment was

that the mean follow-up period from a colon cancer diagnosis until the end of the enrollment period for the Cdiff cohort (1.46 ± 0.77 years, $n = 5$) as compared to the non-Cdiff cohort (1.85 ± 1.12 years, $n = 1,374$) were similar (unpaired *t*-test statistic, assuming equal variances, $P = 0.44$). The methodology employed previously was utilized to examine the remaining data in adjusted statistical models II -V.

Results

A demographic overview of the overall cohort ($n = 293,100$) examined is presented in Table 1. A total of 964 adults were diagnosed with Cdiff and 292,136 adults were not diagnosed with Cdiff. An examination of the overall cohort revealed the Cdiff incidence rate was 0.47 per 1,000 person-years (95% confidence interval (CI) = 0.44 to 0.50 per 1,000 person-years) and the colon cancer incidence rate was 1.16 per 1,000 person-years (95% CI = 1.11 to 1.21 per 1,000 person-years). In both the Cdiff cohort and the non-Cdiff cohort, there was a female preponderance (female/male ratio of about 2:1) and the mean age in years at initial enrollment was similar (about 50 years old). In addition, the distribution of residency locations for both cohorts was similar at initial and final enrollment. The mean follow-up period in years was significantly greater in the non-Cdiff cohort (6.97 ± 0.37) as compared to the Cdiff cohort (2.67 ± 1.97).

An overview of adults diagnosed with colon cancer examined is shown in Table 2. The incidence rate of diagnosed colon cancer was 1.16 per 1,000 person-years in the non-Cdiff cohort as compared to 3.11 per 1,000 person-years in the Cdiff cohort. The mean onset for diagnosed colon cancer after Cdiff was 564 ± 434 days. The mean age of initial colon cancer diagnosis (about 61 years old) and the female preponderance (female/male ratio of about 2:1) was similar in both cohorts examined. The percentage of adults administered ≥ 1 colonoscopy screening procedure (about 7%) and the mean total number of colonoscopy screening procedures per person (about 0.11) were similar in both cohorts examined. As for the location of the initial colon cancer diagnoses most adults received a diagnosis of colon, unspecified in the Cdiff cohort (75%) and the non-Cdiff cohort (55.57%).

Frequency modeling revealed that the risk ratio for the incidence rate of colon cancer in the Cdiff cohort was 2.68-fold (95% CI = 1.16 to 5.29) significantly increased as compared to the non-Cdiff cohort (Table 3). The attributable rate per 1,000 person-years for colon cancer post-Cdiff was 1.95 per 1,000 person-years (95% CI = 0.63 to 3.26 per 1,000 person-years).

Cox proportional hazards modeling revealed the incidence rate of colon cancer in the Cdiff cohort as compared to the non-Cdiff cohort was about 2.6-fold significantly increased in the unadjusted and adjusted models (Table 4). The result of a life-test model examining the same data to compare the incidence rate of colon cancer in the Cdiff cohort as compared to the non-Cdiff cohort overtime is shown in Figure 2. The results of the life-test model showed that most colon cancer was diagnosed from about 3 months through 4 years post-Cdiff.

As shown in Table 5, adjusted model II and model III revealed that colonoscopy screening procedure status did not im-

part the significant association between Cdiff and colon cancer risk. In addition, adjusted model IV revealed that consideration of the diagnostic status for family history of any type of cancer and personal history of tobacco abuse, drug abuse/dependence, or alcohol abuse/dependence did not impact the significant association between Cdiff and colon cancer risk. Similarly, adjusted model V revealed that the covariates considered in adjusted model IV, with the additional exclusion of persons previously diagnosed with ulcerative colitis, infectious colitis, immunodeficiency, or any type of cancer did not impact the significant association between Cdiff and colon cancer risk.

Table 6 shows that by increasing the initial enrollment period without a Cdiff and/or colon cancer diagnosis from 1 to 4 years in adjusted models II-V, the risk for a colon cancer diagnosis following Cdiff was significantly increased, and the magnitude of the association was several-fold higher as compared to the HRs observed in Tables 3-5. The greatest risks for colon cancer after Cdiff were observed in adjusted models IV-V (HR = about 4.0). None of the covariates, colonoscopy procedure variables, or other diagnosis variables impacted the association between Cdiff and colon cancer.

Discussion

The observed study results provide important, new insights into the potential association between Cdiff and the risk of colon cancer. The longitudinal cohort design employed allowed for an evaluation of the relationship between Cdiff and its association with colon cancer over a multi-year period. As shown in Figure 2, the incidence rate of colon cancer over time in the non-Cdiff cohort remained relatively uniform over the entire study period. By contrast, also shown in Figure 2, the Cdiff cohort experienced a marked initial increase in colon cancer diagnoses from 60 days through 4 years and a leveling off from 4 to 7 years. The observed clustering of colon cancer diagnoses within the first 4 years may indicate a temporal window, in which Cdiff infection exerts its physiological effects on the host. In the alternative, the clustering observed may reflect that there were relatively few persons in the Cdiff cohort with more than 4 years of follow-up post-Cdiff diagnosis, and, hence, the study sample was underpowered to detect additional colon cancers from 4 to 7 years after a Cdiff diagnosis.

The statistical models utilized considered numerous potential confounders such as gender, age in years, date of birth in years, residency, and colonoscopy screening status. In addition, exclusion of persons with medical diagnoses such as ulcerative colitis, infectious colitis, immunodeficiency, previous history of any type of cancer, as well as considering the impact of a family history of any type of cancer and a personal history of tobacco abuse, drug abuse/dependence, alcohol abuse/dependence, and overweight/obesity, all which may significantly impact colon cancer risk, did not alter the significant association observed between Cdiff and colon cancer. Finally, consideration of colon cancer survivability revealed that by ensuring that the period of follow-up between a colon cancer diagnosis and the end of the enrollment period in the Cdiff and non-Cdiff cohorts was similar, the risk of a colon cancer diagnosed after Cdiff actually increased relative to the other

Table 1. Demographic Characteristics of the Adults Examined^a

Parameter examined	Non-Cdiff cohort (n = 292,136)	Cdiff cohort (n = 964)
Gender (%)		
Male	83,473 (28.57%)	337 (34.96%)
Female	208,663 (71.43%)	627 (64.97%)
Date of birth in years, mean ± SD (range)	1942 ± 18.27 (1,898 - 1,984)	1947 ± 14.47 (1,901 - 1,982)
Age in years at initial study enrollment, mean ± SD (range)	51.24 ± 17.96 (20 - 96)	50.11 ± 14.18 (20 - 91)
Residency at initial enrollment (%) ^b		
District 1	13,356 (4.57%)	39 (4.05%)
District 2	18,296 (6.26%)	60 (6.22%)
District 3	27,652 (9.47%)	111 (11.52%)
District 4	23,551 (8.06%)	105 (10.89%)
District 5	17,663 (6.05%)	89 (9.23%)
District 6	31,024 (10.62%)	82 (8.51%)
District 7	23,647 (8.09%)	91 (9.44%)
District 8	11,817 (4.04%)	51 (5.29%)
District 9	15,682 (5.37%)	49 (5.08%)
District 10	14,522 (4.97%)	52 (5.39%)
District 11	94,926 (32.50%)	235 (24.38%)
Residency at the end enrollment (%) ^b		
District 1	13,352 (4.57%)	41 (4.25%)
District 2	18,359 (6.28%)	58 (6.02%)
District 3	28,369 (9.71%)	118 (12.24%)
District 4	23,740 (8.13%)	103 (10.69%)
District 5	17,919 (6.13%)	91 (9.44%)
District 6	30,980 (10.60%)	86 (8.92%)
District 7	23,843 (8.16%)	89 (9.23%)
District 8	12,030 (4.12%)	51 (5.29%)
District 9	15,714 (5.38%)	46 (4.77%)
District 10	14,853 (5.08%)	47 (4.88%)
District 11	92,977 (31.83%)	234 (24.27%)
Personal history		
Tobacco abuse (%) ^c	1,938 (0.66%)	11 (1.14%)
Alcohol abuse/dependence (%) ^d	17,475 (5.98%)	153 (15.87%)
Drug abuse/dependence (%) ^e	17,088 (5.85%)	159 (16.49%)
Overweight/obesity (%) ^f	15,287 (5.23%)	162 (16.80%)
Family history		
Cancer (%) ^g	2,684 (0.92%)	24 (2.49%)
Person-years		
Total	2,036,767	2,575
Mean ± SD (range)	6.97 ± 0.37 (0.008 - 7.00)	2.67 ± 1.98 (0.003 - 7.00)

^aAll adults were enrolled for 96 consecutive months. All adults had ≥ 8 outpatient office visits during their enrollment period. All adults diagnosed with Cdiff and/or colon cancer within the first year of enrollment were excluded. ^bAll adults were specified to have a county of residence at the time of initial enrollment and at the end of the enrollment period. There are a total of 67 counties in the state of Florida, which are grouped by the state of Florida based upon geographical areas into 11 districts. ^cPersonal history of tobacco abuse (ICD-9 code: 305.1x). ^dPersonal history of alcohol abuse/dependence (ICD-9 code: 303.xx or 305.0x). ^ePersonal history of drug abuse/dependence (ICD-9 code: 304.xx, 305.2x, 305.3x, 305.34, 305.5x, 305.6x, 305.7x, 305.8x or 305.9x). ^fPersonal history of overweight/obesity (ICD9 code: 278.00, 278.01, or 278.02). ^gFamily history of any type of cancer (ICD-9 code: V16.xx). Cdiff: *Clostridioides* (formerly known as *Clostridium*) *difficile*; CI: confidence interval; SD: standard deviation.

Table 2. Demographic Characteristics of the Adults Diagnosed With Colon Cancer^a Examined^b

Cohort examined (n)	Gender (%)	Date of birth, mean ± SD (range)	Age of initial diagnosis, mean ± SD (range)	Incidence rate per 1,000 person-years (n)
All adults (n = 293,100)	Male: 753 (31.79%)	1933 ± 15.3 (1,899 - 1,980)	61.76 ± 13.77 (21 - 92)	1.16 (2,369)
	Female: 1,616 (68.21%)			
Non-Cdiff cohort (n = 292,136)	Male: 751 (31.81%)	1933 ± 15.1 (1,899 - 1,979)	61.78 ± 13.77 (21 - 92)	1.16 (2,361)
	Female: 1,610 (68.19%)			
Cdiff cohort (n = 964)	Male: 2 (25.00%)	1943 ± 10.5 (1,924 - 1,955)	61.9 ± 12.9 (47 - 85)	3.11 (8)
	Female: 6 (75.00%)			

^aAdults diagnosed with colon cancer were diagnosed according to the International Code for Disease 9th revision with the code 153.xx. ^bAll adults were enrolled for 96 consecutive months. All adults had ≥ 8 outpatient office visits during their enrollment period. All adults diagnosed with Cdiff and/or colon cancer within the first year of enrollment were excluded. Cdiff: *Clostridioides* (formerly known as *Clostridium*) *difficile*; SD: standard deviation.

Table 3. Frequency Modeling Results to Evaluate the Incidence Rate of Colon Cancer^a in the Cdiff Cohort as Compared to the Non-Cdiff Cohort^b

Parameter examined	Cdiff cohort (n = 964)	Non-Cdiff cohort (n = 292,136)	Risk ratio (95% CI)
			Attributable rate per 1,000 person-years (95% CI)
Colon cancer (n)	8	2,361	
Incidence rate per 1,000 person-years	3.11	1.16	2.68 (1.16 to 5.29)* 1.95 (0.63 to 3.26)

*P < 0.05. ^aAdults diagnosed with colon cancer were diagnosed according to the International Code for Disease 9th revision with the code 153.xx. ^bAll adults were enrolled for 96 consecutive months. All adults had ≥ 8 outpatient office visits during their enrollment period. All adults diagnosed with Cdiff and/or colon cancer within the first year of enrollment were excluded. Cdiff: *Clostridioides* (formerly known as *Clostridium*) *difficile*; CI: confidence interval.

models constructed.

In evaluating the results observed, at present, there are no dedicated epidemiological studies examining the potential association between Cdiff and the risk of colon cancer [24, 25]. The results of the present study provide the first longitudinal cohort epidemiological data associating Cdiff with a long-term increased risk for colon cancer. Despite the lack of previous epidemiological studies, the results observed in the present study are biologically plausible. For example, investigators

analyzed data from a population-based case-control study of colon cancer and measured antibody levels against flagellin of salmonella (FliC). These investigators observed FliC antibody levels were significantly higher in patients with colorectal polyps and cancer than in controls [26]. Another study examined the potential link between *Fusobacterium nucleatum* intestinal infection and the risk of colon cancer [27]. The investigators observed that the amount of *F. nucleatum* DNA in colorectal cancer tissue was associated with shorter colon cancer sur-

Table 4. Cox Proportional Hazards Model Results Examining Colon Cancer^a Diagnoses in the Cdiff Cohort as Compared to Non-Cdiff Cohort^b

Model type	Variable	Hazards ratio (95% CI)
Unadjusted model	Cdiff (diagnosed vs. not diagnosed)	2.646 (1.321 to 5.301)*
Adjusted model I	Cdiff (diagnosed vs. not diagnosed)	2.657 (1.325 to 5.330)*
	Gender (female vs. male)	0.840 (0.770 to 0.917)**
	Date of birth in years	1.034 (1.024 to 1.043)**
	Residency at initial enrollment	1.000 (0.995 to 1.006)
	Residency at end of enrollment	0.997 (0.991 to 1.003)
	Age in years at initial enrollment	1.072 (1.062 to 1.082)**

*P < 0.01. **P < 0.001. ^aAdults diagnosed with colon cancer. ^bAll adults were enrolled for 96 consecutive months. All adults had ≥ 8 outpatient office visits during their enrollment period. All adults diagnosed with Cdiff and/or colon cancer within the first year of enrollment were excluded. Cdiff: *Clostridioides* (formerly known as *Clostridium*) *difficile*; CI: confidence interval.

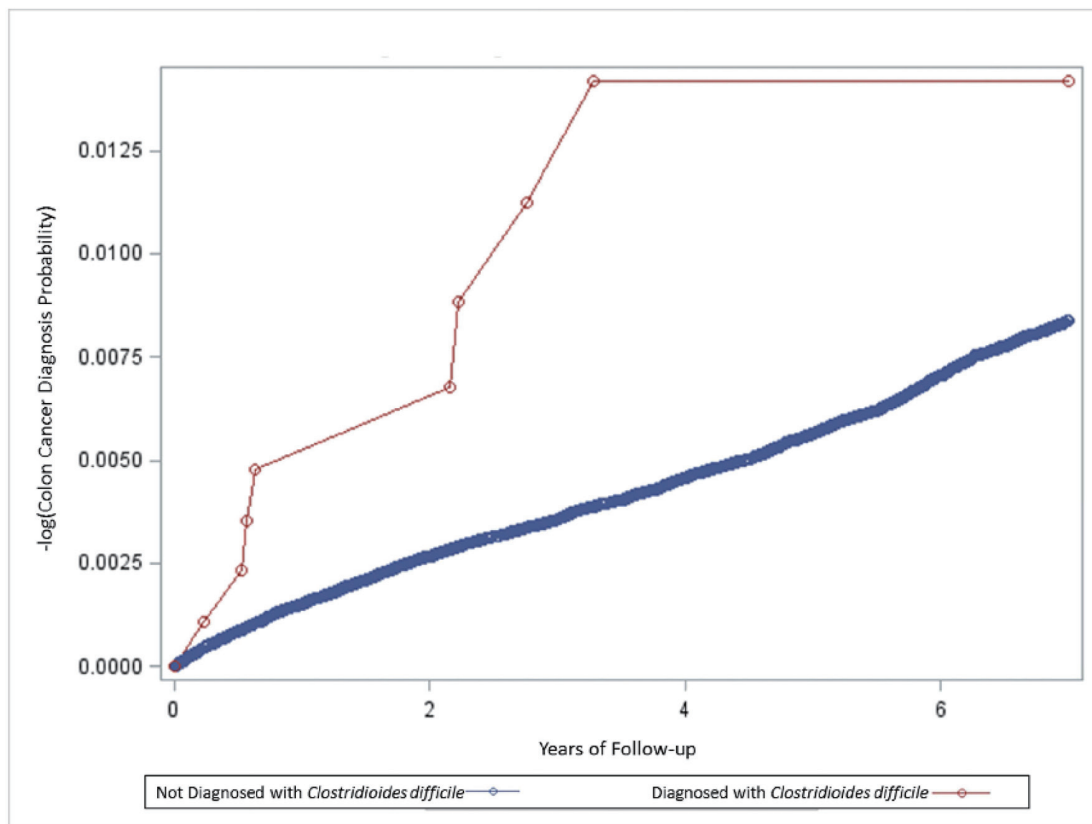


Figure 2. A life-test model summary of the relationship between the incidence rate of colon cancer in the *Clostridioides difficile* cohort as compared to the non-*C. difficile* cohort.

vival. Other investigators undertook a retrospective study on hospitalized adults with bacteremia but without a previous colorectal cancer diagnosis and compared them to a group of patients not diagnosed with bacteremia [28]. These investigators observed that several strains of bacteria identified were associated with marked increases in the risk for the patients subsequently being diagnosed with colorectal cancer.

It is also important to consider that the present study design addressed many of the common potential limitations of epidemiological studies. The diagnoses/procedures examined were performed prospectively on the adults examined as part of their routine medical care, and were undertaken independently of the present study design. As a result, phenomena such as recall bias or *post hoc* assumptions by researchers as to presumed associations or lack of associations from retrospective interpretations of previously collected data were minimized.

All diagnoses examined were made by licensed medical staff using ICD-9 coding. This helps to minimize potential differences induced by changing diagnostic criteria and allows the present study results to be comparable with many other sources of medical data. Previous studies also revealed high accuracy between ICD-9 coded colon cancers as compared to chart review [29] and between ICD-9 coded *Cdiff* as compared to microbiological test results [30]. Further, for countries presently utilizing ICD-10 diagnostic codes, the types of colon cancer examined in this study using ICD-9 coding are virtually identical

to the types of colon cancer present in ICD-10 coding.

The cohorts examined were demographically similar, and covariates such as age, date of birth, gender, residency, colonoscopy screening status, and prior personal and family medical histories were considered in the statistical models constructed. It is important to note that the phenomena observed in this study with covariates, such as more recent dates of birth, older ages, and male as compared to female gender are all significant risk factors for a colon cancer diagnosis, consistent with previous studies [31, 32]. It was also observed that personal histories of overweight/obesity, tobacco abuse, alcohol abuse/dependence, and drug abuse/dependence were more common in the *Cdiff* cohort as compared to the undiagnosed cohort, but consideration of these variables in the statistical models constructed did not alter the association between *Cdiff* infection and colon cancer risk. Finally, the cohorts examined were continuously enrolled in the Florida Medicaid system for an 8-year period, actively utilizing the Florida Medicaid system, and extensive measures were employed to ensure that only new cases of *Cdiff* and colon cancer were examined.

Conclusion

This retrospective cohort study of healthcare data from the IHRD provides important, new evidence supporting the hy-

Table 5. Cox Proportional Hazards Model Results Examining Colon Cancer^a Diagnoses in the Cdiff Cohort as Compared to Non-Cdiff Cohort^b When Excluding Adults Diagnosed With Cdiff and/or Colon Cancer Within the First Year of Enrollment

Model type	Variable	Hazards ratio (95% CI)
Adjusted model II		
	Cdiff (diagnosed vs. not diagnosed)	2.673 (1.333 to 5.360)**
	Gender (female vs. male)	0.836 (0.767 to 0.913)***
	Date of birth in years	1.031 (1.022 to 1.041)***
	Residency at initial enrollment	1.000 (0.995 to 1.006)
	Residency at end of enrollment	0.997 (0.991 to 1.003)
	Age in Years at Initial Enrollment	1.069 (1.059 to 1.080)***
	≥ 1 colonoscopy screening procedure ^c (yes vs. no)	1.210 (1.059 to 1.381)**
Adjusted model III		
	Cdiff (diagnosed vs. not diagnosed)	2.667 (1.3330 to 5.348)**
	Gender (female vs. male)	0.836 (0.767 to 0.913)***
	Date of birth in years	1.031 (1.022 to 1.041)***
	Residency at initial enrollment	1.000 (0.995 to 1.006)
	Residency at end of enrollment	0.997 (0.991 to 1.003)
	Age in years at initial enrollment	1.069 (1.059 to 1.080)***
	Total colonoscopy screening procedures ^d	1.161 (1.048 to 1.285)**
Adjusted model IV		
	Cdiff (diagnosed vs. not diagnosed)	2.471 (1.232 to 4.958)*
	Gender (female vs. male)	0.821 (0.752 to 0.897)***
	Date of birth in years	1.024 (1.015 to 1.034)***
	Residency at initial enrollment	1.000 (0.994 to 1.006)
	Residency at end of enrollment	0.997 (0.991 to 1.003)
	Age in years at initial enrollment	1.064 (1.054 to 1.075)***
	Total colonoscopy screening procedures ^d	1.105 (0.996 to 1.226)
	Family history of any type of cancer	3.547 (3.589 to 5.534)***
	Personal history of tobacco abuse	0.996 (0.516 to 1.923)
	Personal history of alcohol abuse/dependence	1.195 (0.975 to 1.466)
	Personal history of drug abuse/dependence	1.121 (0.892 to 1.409)
	Personal history of overweight/obesity	1.251 (1.043 to 1.500)3*
Adjusted model V ^e		
	Cdiff (diagnosed vs. not diagnosed)	2.970 (1.478 to 5.970)**
	Gender (female vs. male)	0.758 (0.681 to 0.844)***
	Date of birth in years	1.027 (1.015 to 1.039)***
	Residency at initial enrollment	1.000 (0.993 to 1.007)
	Residency at end of enrollment	0.997 (0.990 to 1.004)
	Age in years at initial enrollment	1.071 (1.058 to 1.084)***
	Total colonoscopy screening procedures ^d	1.082 (0.946 to 1.237)
	Family history of any type of cancer	5.242 (3.975 to 6.913)***
	Personal history of tobacco abuse	1.075 (0.480 to 2.405)
	Personal history of alcohol abuse/dependence	1.217 (0.944 to 1.570)
	Personal history of drug abuse/dependence	1.153 (0.866 to 1.535)
	Personal history of overweight/obesity	1.242 (0.990 to 1.560)

*P < 0.05. **P < 0.01. ***P < 0.001. ^aAdults diagnosed with colon cancer. ^bAll adults were enrolled for 96 consecutive months. All adults had ≥ 8 outpatient office visits during their enrollment period. ^cAdults were assessed to determine whether at least one colonoscopy screening procedure was undertaken within the appropriate study period. ^dAdults were assessed to determine the total number of colonoscopy screening procedures undertaken within the appropriate study period. ^eAdults diagnosed with ulcerative colitis, infectious colitis, immunodeficiency, or any type of cancer before a Cdiff diagnosis were excluded. Adults diagnosed with ulcerative colitis, infectious colitis, immunodeficiency, or any types of cancer before a colon cancer diagnosis were excluded. Cdiff: *Clostridioides* (formerly known as *Clostridium*) *difficile*; CI: confidence interval.

Table 6. Cox Proportional Hazards Model Results Examining Colon Cancer^a Diagnoses in the Cdiff Cohort as Compared to Non-Cdiff Cohort^b When Excluding Adults Diagnosed With Cdiff and/or Colon Cancer Diagnoses Within the First 4 Years of Enrollment

Model type	Variable	Hazards ratio (95% CI)
Adjusted model II	Cdiff (diagnosed vs. not diagnosed)	4.064 (1.686 to 9.800)**
	Gender (female vs. male)	0.790 (0.705 to 0.885)***
	Date of birth in years	1.023 (1.011 to 1.035)***
	Residency at initial enrollment	1.001 (0.994 to 1.009)
	Residency at end of enrollment	0.995 (0.988 to 1.003)
	Age in years at initial enrollment	1.062 (1.048 to 1.076)***
	≥ 1 colonoscopy screening procedure ^c (yes vs. no)	1.468 (1.189 to 1.811)***
Adjusted model III	Cdiff (diagnosed vs. not diagnosed)	4.039 (1.675 to 9.739)**
	Gender (female vs. male)	0.790 (0.705 to 0.885)***
	Date of birth in years	1.023 (1.010 to 1.035)***
	Residency at initial enrollment	1.001 (0.994 to 1.009)
	Residency at end of enrollment	0.995 (0.988 to 1.003)
	Age in years at initial enrollment	1.062 (1.048 to 1.075)***
	Total colonoscopy screening procedures ^d	1.353 (1.140 to 1.605)***
Adjusted model IV	Cdiff (diagnosed vs. not diagnosed)	3.761 (1.559 to 9.078)**
	Gender (female vs. male)	0.786 (0.700 to 0.881)***
	Date of birth in years	1.016 (1.004 to 1.029)*
	Residency at initial enrollment	1.001 (0.994 to 1.009)
	Residency at end of enrollment	0.995 (0.988 to 1.003)
	Age in years at initial enrollment	1.057 (1.043 to 1.071)***
	Total colonoscopy screening procedures ^d	1.302 (1.095 to 1.547)**
	Family history of any type of cancer	3.387 (2.427 to 4.725)***
	Personal history of tobacco abuse	0.929 (0.384 to 2.243)
	Personal history of alcohol abuse/dependence	1.369 (1.060 to 1.770)*
	Personal history of drug abuse/dependence	1.082 (0.799 to 1.465)
Personal history of overweight/obesity	1.329 (1.047 to 1.686)*	
Adjusted model V ^e	Cdiff (diagnosed vs. not diagnosed)	4.562 (1.885 to 11.041)***
	Gender (female vs. male)	0.735 (0.638 to 0.848)***
	Date of birth in years	1.019 (1.004 to 1.035)*
	Residency at initial enrollment	0.999 (0.990 to 1.008)
	Residency at end of enrollment	0.997 (0.988 to 1.007)
	Age in years at initial enrollment	1.064 (1.047 to 1.081)***
	Total colonoscopy screening procedures ^d	1.320 (1.061 to 1.644)*
	Family history of any type of cancer	4.402 (2.954 to 6.559)***
	Personal history of tobacco abuse	1.231 (0.458 to 3.306)
	Personal history of alcohol abuse/dependence	1.377 (0.995 to 1.905)
	Personal history of drug abuse/dependence	1.086 (0.739 to 1.596)
	Personal history of overweight/obesity	1.357 (1.010 to 1.825)*

*P < 0.05. **P < 0.01. ***P < 0.001. ^aAdults diagnosed with colon cancer. ^bAll adults were enrolled for 96 consecutive months. All adults had ≥ 8 outpatient office visits during their enrollment period. ^cAdults were assessed to determine whether at least one colonoscopy screening procedure was undertaken within the appropriate study period. ^dAdults were assessed to determine the total number of colonoscopy screening procedures undertaken within the appropriate study period. ^eAdults diagnosed with ulcerative colitis, infectious colitis, immunodeficiency, or any type of cancer before a Cdiff diagnosis were excluded. Adults diagnosed with ulcerative colitis, infectious colitis, immunodeficiency, or any types of cancer before a colon cancer diagnosis were excluded. Cdiff: *Clostridioides* (formerly known as *Clostridium*) *difficile*; CI: confidence interval.

hypothesis that *Cdiff* is associated with an increased risk of colon cancer. In light of the importance of *Cdiff* infection in the etiology of colon cancer, future studies should elucidate potential risk factors for *Cdiff* infections. It is also recommended that future studies should focus on various types of infections (including recurrent infections) on colon cancer risk, and consideration given to the importance of changing screening guidelines for these patient populations.

Acknowledgments

None to declare.

Financial Disclosure

This research was funded by the non-profit Institute of Chronic Illnesses, Inc. Mark R. Geier and David A. Geier are co-directors of the Institute of Chronic Illnesses, Inc.

Conflict of Interest

Mark R. Geier and David A. Geier have a patent pending for a vaccine to prevent *Clostridioides difficile* associated colon cancer.

Informed Consent

This Liberty IRB waived the informed consent requirement for the IHRD data analyzed in this study.

Author Contributions

David A. Geier and Mark R. Geier: conceptualization, methodology, writing - original draft, writing - review and editing.

Data Availability

Florida's Agency for Health Care Administration was the source for the Florida Medicaid system eligibility and claim data within the IHRD, and they are the agency that should be contacted to gain access to Florida Medicaid data.

Abbreviations

Cdiff: *Clostridioides* (formerly *Clostridium*) *difficile*; IHRD: Independent Healthcare Research Database; TcdB: cytotoxin toxin B; IRB: Institutional Review Board; CPT: current procedural terminology; ICD-9: International Code for Disease, 9th revision; HR: hazard ratio; CI: confidence interval; FliC: flagellin of *Salmonella*

References

1. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. *Cell Host Microbe*. 2014;15(3):317-328. doi [pubmed pmc](#)
2. Garrett WS. The gut microbiota and colon cancer. *Science*. 2019;364(6446):1133-1135. doi [pubmed](#)
3. Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, Flavell RA. Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer*. 2013;13(11):759-771. doi [pubmed](#)
4. Parsonnet J. Bacterial infection as a cause of cancer. *Environ Health Perspect*. 1995;103(Suppl 8):263-268. doi [pubmed pmc](#)
5. Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, Roslani AC, et al. High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes*. 2017;3:34. doi [pubmed pmc](#)
6. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, Beghini F, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med*. 2019;25(4):667-678. doi [pubmed pmc](#)
7. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 2019;25(4):679-689. doi [pubmed pmc](#)
8. Wong SH, Zhao L, Zhang X, Nakatsu G, Han J, Xu W, Xiao X, et al. Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. *Gastroenterology*. 2017;153(6):1621-1633.e1626. doi [pubmed](#)
9. Tomkovich S, Dejea CM, Winglee K, Drewes JL, Chung L, Housseau F, Pope JL, et al. Human colon mucosal biofilms from healthy or colon cancer hosts are carcinogenic. *J Clin Invest*. 2019;129(4):1699-1712. doi [pubmed pmc](#)
10. Baxter NT, Zackular JP, Chen GY, Schloss PD. Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. *Microbiome*. 2014;2:20. doi [pubmed pmc](#)
11. Drewes JL, Chen J, Markham NO, Knippel RJ, Domingue JC, Tam AJ, Chan JL, et al. Human colon cancer-derived *clostridioides difficile* strains drive colonic tumorigenesis in mice. *Cancer Discov*. 2022;12(8):1873-1885. doi [pubmed pmc](#)
12. Geier DA, Geier MR. A longitudinal cohort study of childhood MMR vaccination and seizure disorder among American children. *Brain Dev*. 2021;43(2):251-267. doi [pubmed](#)
13. Geier DA, Kern JK, Geier MR. Childhood MMR vaccination and the incidence rate of measles infection: a ten year longitudinal cohort study of American children born in the 1990s. *BMC Pediatr*. 2019;19(1):325. doi [pubmed pmc](#)
14. Geier DA, Kern JK, Geier MR. Down syndrome as a genetic model to evaluate the role of oxidative stress and transsulfuration abnormalities in autism spectrum disorder.

- der: A 10-year longitudinal cohort study. *Dev Neurobiol*. 2019;79(9-10):857-867. [doi pubmed](#)
15. Kern JK, Geier DA, Homme KG, Geier MR. A ten year longitudinal examination of the incidence rate and age of childhood encephalopathy diagnoses in an autism spectrum disorder diagnosed cohort. *Acta Neurobiol Exp (Wars)*. 2020;80(1):66-75. [pubmed](#)
 16. Geier DA, Geier MR. A longitudinal cohort study of precocious puberty and autism spectrum disorder. *Horm Res Paediatr*. 2021;94(5-6):219-228. [doi pubmed](#)
 17. Geier DA, Geier MR. Fetal alcohol syndrome and the risk of neurodevelopmental disorders: A longitudinal cohort study. *Brain Dev*. 2022;44(10):706-714. [doi pubmed](#)
 18. Geier DA, Geier MR. Childhood MMR Vaccination Effectiveness Against Rubella: A Longitudinal Cohort Study. *Glob Pediatr Health*. 2022;9(2333794X221094266). [doi pubmed pmc](#)
 19. Ekobom A, Helmick C, Zack M, Adami HO. Ulcerative colitis and colorectal cancer. A population-based study. *N Engl J Med*. 1990;323(18):1228-1233. [doi pubmed](#)
 20. Mughini-Gras L, Schaapveld M, Kramers J, Mooij S, Neeffjes-Borst EA, Pelt WV, Neeffjes J. Increased colon cancer risk after severe Salmonella infection. *PLoS One*. 2018;13(1):e0189721. [doi pubmed pmc](#)
 21. Lin OS. Acquired risk factors for colorectal cancer. *Methods Mol Biol*. 2009;472:361-372. [doi pubmed](#)
 22. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol*. 2001;96(10):2992-3003. [doi pubmed](#)
 23. Gloeckler Ries LA, Reichman ME, Lewis DR, Hankey BF, Edwards BK. Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. *Oncologist*. 2003;8(6):541-552. [doi pubmed](#)
 24. Chopra T, Alangaden GJ, Chandrasekar P. Clostridium difficile infection in cancer patients and hematopoietic stem cell transplant recipients. *Expert Rev Anti Infect Ther*. 2010;8(10):1113-1119. [doi pubmed](#)
 25. Burgner D, Siarakas S, Eagles G, McCarthy A, Bradbury R, Stevens M. A prospective study of Clostridium difficile infection and colonization in pediatric oncology patients. *Pediatr Infect Dis J*. 1997;16(12):1131-1134. [doi pubmed](#)
 26. Kato I, Boleij A, Kortman GA, Roelofs R, Djuric Z, Severson RK, Tjalsma H. Partial associations of dietary iron, smoking and intestinal bacteria with colorectal cancer risk. *Nutr Cancer*. 2013;65(2):169-177. [doi pubmed pmc](#)
 27. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, Yang J, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut*. 2016;65(12):1973-1980. [doi pubmed pmc](#)
 28. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, Tsoi KKK, et al. Association between bacteremia from specific microbes and subsequent diagnosis of colorectal cancer. *Gastroenterology*. 2018;155(2):383-390.e388. [doi pubmed](#)
 29. Cozzolino F, Bidoli E, Abraha I, Fusco M, Giovannini G, Casucci P, Orso M, et al. Accuracy of colorectal cancer ICD-9-CM codes in Italian administrative healthcare databases: a cross-sectional diagnostic study. *BMJ Open*. 2018;8(7):e020630. [doi pubmed pmc](#)
 30. Scheurer DB, Hicks LS, Cook EF, Schnipper JL. Accuracy of ICD-9 coding for Clostridium difficile infections: a retrospective cohort. *Epidemiol Infect*. 2007;135(6):1010-1013. [doi pubmed pmc](#)
 31. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, Jemal A. Colorectal cancer incidence patterns in the United States, 1974-2013. *J Natl Cancer Inst*. 2017;109(8):djw322. [doi pubmed pmc](#)
 32. Siegel RL, Jemal A, Ward EM. Increase in incidence of colorectal cancer among young men and women in the United States. *Cancer Epidemiol Biomarkers Prev*. 2009;18(6):1695-1698. [doi pubmed](#)