Original Paper

Pharmaceutical Agents as Potential Drivers in the Development of Early-Onset Colorectal Cancer: Case-Control Study

Irit Ben-Aharon[1*], DPhil, MD; Ran Rotem[2*], DPhil; Cheli Melzer-Cohen[3], MSci; Gilad Twig[4,5,6], MD; Andrea Cercek[7], MD; Elizabeth Half[1], MD; Tal Goshen-Lago[1], BA; Gabriel Chodik[5*], DPhil; David Kelsen[7,8*], MD

[1]Department of Gastroenterology, Rambam Healthcare Campus, Haifa, Israel

[2]Harvard T Chan School of Public Health, Boston, MA, United States

[3]KSM Research and Innovation Center, Maccabi Healthcare Services, Tel-Aviv, Israel

[4]The Institute of Endocrinology Diabetes and Metabolism, Sheba Medical Center, Ramat Gan, Israel

[5]Department of Preventive Medicine and Epidemiology, School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[6]Gertner Institute for Epidemiology & Health Policy Research, Sheba Medical Center, Ramat Gan, Israel

[7]Gastrointestinal Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[8]Weill Cornell Medical College, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
David Kelsen, MD
Gastrointestinal Oncology Service
Department of Medicine
Memorial Sloan Kettering Cancer Center
300 East 66th Street
Office 1027
New York, NY, 10065
United States
Phone: 1 646 888 4179
Email: kelsend@mskcc.org

Abstract

Background: The incidence of early-onset colorectal cancer (EOCRC) rose abruptly in the mid 1990s, is continuing to increase, and has now been noted in many countries. By 2030, 25% of American patients diagnosed with rectal cancer will be 49 years or younger. The large majority of EOCRC cases are not found in patients with germline cancer susceptibility mutations (eg, Lynch syndrome) or inflammatory bowel disease. Thus, environmental or lifestyle factors are suspected drivers. Obesity, sedentary lifestyle, diabetes mellitus, smoking, alcohol, or antibiotics affecting the gut microbiome have been proposed. However, these factors, which have been present since the 1950s, have not yet been conclusively linked to the abrupt increase in EOCRC. The sharp increase suggests the introduction of a new risk factor for young people. We hypothesized that the driver may be an off-target effect of a pharmaceutical agent (ie, one requiring regulatory approval before its use in the general population or an off-label use of a previously approved agent) in a genetically susceptible subgroup of young adults. If a pharmaceutical agent is an EOCRC driving factor, regulatory risk mitigation strategies could be used.

Objective: We aimed to evaluate the possibility that pharmaceutical agents serve as risk factors for EOCRC.

Methods: We conducted a case-control study. Data including demographics, comorbidities, and complete medication dispensing history were obtained from the electronic medical records database of Maccabi Healthcare Services, a state-mandated health provider covering 26% of the Israeli population. The participants included 941 patients with EOCRC (≤50 years of age) diagnosed during 2001-2019 who were density matched at a ratio of 1:10 with 9410 control patients. Patients with inflammatory bowel disease and those with a known inherited cancer susceptibility syndrome were excluded. An advanced machine learning algorithm based on gradient boosted decision trees coupled with Bayesian model optimization and repeated data sampling was used to sort through the very high-dimensional drug dispensing data to identify specific medication groups that were consistently linked with EOCRC while allowing for synergistic or antagonistic interactions between medications. Odds ratios for the identified medication classes were obtained from a conditional logistic regression model.

Results: Out of more than 800 medication classes, we identified several classes that were consistently associated with EOCRC risk across independently trained models. Interactions between medication groups did not seem to substantially affect the risk. In our analysis, drug groups that were consistently positively associated with EOCRC included beta blockers and valerian *(Valeriana officinalis)*. Antibiotics were not consistently associated with EOCRC risk.

Conclusions: Our analysis suggests that the development of EOCRC may be correlated with prior use of specific medications. Additional analyses should be used to validate the results. The mechanism of action inducing EOCRC by candidate pharmaceutical agents will then need to be determined.

Introduction

An abrupt increase in the incidence of early-onset colorectal cancer (EOCRC), generally defined as the occurrence of colon or rectal cancer in people less than 50 years of age, began in the mid 1990s; this was first noted in the early 2000s [1-3]. The majority of EOCRC cases are not due to inherited cancer susceptibility genes or inflammatory bowel disease (IBD), which have long been known as risk factors for EOCRC [3-5]. This implies that the driving factors for the increase in EOCRC are environmental, although there is presumably an interplay between genes and the environment. The rise in the incidence of EOCRC has now been documented in many countries [3,6,7]. While an increasing prevalence of obesity, high processed foods diets, diabetes, smoking, alcohol consumption, and sedentary lifestyles have been implicated in EOCRC, these lifestyle and comorbidity factors have been present since the 1950s, and the marked increase in EOCRC was noted beginning in the 1990s [8,9]. Further, many patients with EOCRC are neither obese nor sedentary; data on obesity as an increased risk factor for EOCRC report both increased and decreased risk compared to controls [10-13]. The extensive use of antibiotics has also been suggested as a causative factor through the induction of changes to the spectrum of the gut microbiome, but data regarding the gut microbiome and metabolites in patients with EOCRC compared to unaffected patients are conflicting [14,15].

The observation that the increase in EOCRC incidence was abrupt rather than occurring gradually over decades suggested to us that a factor leading to colonic neoplasia that was not previously widely used was made available to young people over a relatively short period of time. We thus hypothesized that a driver for EOCRC could be an off-target effect of a medication that acts directly or indirectly on the colonic mucosa to result in the activation of oncogenic pathways or silencing of protective pathways in individuals who otherwise have a benign genetic makeup. These medications were either not previously available or not previously widely used and were prescribed by physicians or obtained over-the-counter (OTC) by young people. These medications could be new drug approvals by regulatory agencies through new drug applications (NDAs) or supplementary NDAs or drugs that began to be used for uses other than the approved indication (ie, off-label use). The pharmaceutical agent could act either directly on the colonic mucosa or indirectly (eg, via the gut microbiome or its metabolites). Even though EOCRC is increasing in incidence, it still affects only a small fraction of all young people. Young patients developing EOCRC may have a genetic predisposition, such as a genetic polymorphism (possibly resulting in altered drug metabolism), which translates into an off target carcinogenic effect.

To test our hypothesis, we sought to identify pharmaceutical agents used more frequently in patients with EOCRC than in their peer controls and determine which may be risk factors for EOCRC by using a novel machine learning method to analyze a large Israeli electronic medical records (EMRs) database that includes digitized pharmacy records for patients with EOCRC and matched controls.

Methods

Study Population

To identify pharmaceutical agents related to EOCRC, a nested case-control study was performed using data from Maccabi Healthcare Services (MHS), an Israeli state-mandated health provider that serves 2.6 million members (26% of the Israeli population). EMRs have been used in MHS since the mid 1990s. Each patient's EMR contains individual-level demographic and clinical information, including clinical diagnoses, hospitalizations, medical procedures, laboratory test results, and filled prescriptions.

Case Ascertainment and Matching

MHS maintains a cancer registry though linkage with the Israel National Cancer Registry (INCR). The INCR was established in 1965 with the aim of continuously collecting data on newly diagnosed cancer cases from all medical institutions in Israel. Since 1981, all pathology results from diagnosed cancer cases must be submitted to the INCR. To further ensure complete case ascertainment, the MHS cancer registry supplements the national database with 2 additional sources of data, namely, histological findings from the MHS central laboratory and purchase authorization records for cancer-specific medications.

We used the MHS cancer registry to retrieve information on all patients with colorectal cases first diagnosed between January 1, 2001, to December 31, 2019, who received their first cancer diagnosis at age 50 years or younger (n=1461). To minimize

misclassification of medication exposure, we restricted the analysis to patients who were continuously enrolled in MHS for at least 3 years prior to the index date (n=1174). We also excluded 237 cases who had EMR indications of IBD (Crohn disease or ulcerative colitis) or a personal or family history of suspicion for an inherited cancer susceptibility gene. Thus, the final case population included 941 patients with colorectal cancer.

For each case we matched 10 controls through density (risk-set) matching. Controls had to be continuously enrolled in MHS for at least 3 years before the index date, have no personal history of colorectal cancer, and have no evidence of colorectal cancer at the index date. Matching was done using sex, residential district, sociodemographic status (based on enumeration area, which is the smallest geostatistical unit of the Israeli central bureau of statistics), minority subpopulation at the residential area (Israeli Arab or Jewish Orthodox and Ultra Orthodox or Jewish secular), birth year (±2 years), and periphery index (proximity to a large urban center).

Exposure Assessment

MHS members can fill prescriptions at any of more than 700 affiliated pharmacies across the country with a minimum copayment. OTC medications can also be purchased through these pharmacies at a discount. For both cases and controls, we obtained information on all dispensing of medications (OTC and prescription) prior to the index date, implementing a 2-year lag prior to the index date to reduce the likelihood of reverse-causation. To reduce the dimensionality of the data and ease interpretation, individual medications were grouped according to the fifth level (chemical substance) of the Anatomical Therapeutic Chemical classification system. Drugs that were used by less than 0.1% (103/10,351) of the cohort were excluded.

Comorbidity Information

Our goal in this analysis was to identify medication classes that were linked with EOCRC to inform future studies. Thus, we did not aim to make casual interpretations. We collected information on several comorbidities recorded before the index state that have been linked with EOCRC [16-18]. These included diabetes (International Classification of Diseases, Ninth Revision [ICD-9] code 250.x); hypertension (ICD-9 code 401.x); cardiovascular conditions including ischemic heart disease (ICD-9 codes 410.x, 412, 429.7, and 429.79), non–myocardial infarction (ICD-9 codes 36.x, 411.x, 414.x, and 429.2), congestive heart failure (ICD-9 codes 404.x and 428.x), peripheral vascular disease (ICD-9 codes 440.x, 441.x, 442.x, and 443.x), stroke (ICD-9 codes 433.x1 and 438.x), transient ischemic attack (ICD-9 code 435.x), cerebrovascular atherosclerosis (ICD-9 code 434.x0), and atrial fibrillation (ICD-9 code 427.3x); and obesity (ICD-9 code 278.02). Data on comorbidities were obtained from the MHS automated patient registries, which apply case ascertainment algorithms to define patients with chronic diseases according to multiple data sources and disease-specific international guidelines. These include separate registries for cardiovascular disease, diabetes, and hypertension. MHS physicians are required to measure patients' body weight annually and document this in the EHR

[17-20]. To assess potential surveillance bias, we also calculated the average frequency of monthly physician visits for each patient by summing all physician visits recorded for a given patient by each patient's follow-up time in the system up to 2 years before the index date, as this can affect the likelihood and timing of receiving an EOCRC diagnosis as well as the nature of medication prescription patterns.

Statistical Analysis

The very high dimensionality of the medication dispensing data limits the use of conventional statistical methods for the analysis since such methods cannot account for the frequent coexposure to many different medication classes. Thus, to conduct a high-throughput screening of medications and identify specific medication groups, individual or in combination, that were consistently linked with EOCRC, we used eXtreme Gradient Boosting (XGBoost), a highly efficient implementation of gradient tree boosting [21]. Briefly, gradient boosting is an ensemble technique that combines the outputs of many individual decision trees to capture complex associations with an outcome of interest. The combining of trees is done iteratively so that each new tree is fit on the residuals of the previous one. XGBoost uses parallel processing to reduce computing time and advanced regularization to enhance the stability of the predictions. The model requires defining several data-specific regularization parameters to optimize performance, which were selected though a 5-fold cross-validation coupled with a Bayesian model optimization [22].

To enhance the reproducibility of our findings, we trained 50 independent XGBoost models, each time setting aside a randomly selected 20% of the data. A list of medication groups associated with EOCRC was derived through global Shapley additive explanations (SHAP) values [23], focusing on medication groups that were associated with the outcome in at least 50% of the independent runs or those with global SHA*P* value in the top 2.5% of the distribution in at least 1 of the runs. A final XGBoost model was then trained with these selected medication groups using the full data to evaluate the presence of informative interactions, assessed through the tree depth hyperparameter following final model optimization. Since machine learning models, including XGBoost, do not provide specific effect estimates, the final list of medications was then included in a conditional logistic regression model to obtain odds ratios (ORs) with 95% CIs for each medication category while accounting for the matched risk sets and adjusting for the frequency of physician visits as well as the patients' history of diabetes, cardiovascular disease, and excess body weight at baseline that were found to be significant in the multivariable model. All statistical analyses were performed using R (version 4.0.2, R Foundation for Statistical Computing).

Ethical Considerations

The study protocol was approved by the MHS institutional review board (MHS IRB 0034-24). Informed consent was waived by the institutional review board because patients' identifying details were removed. There was no compensation provided to individual patients for participation in the study.

Results

Population characteristics are shown in Table 1. The average age at the time of EOCRC diagnosis was 43.6 (SD 6.7) years. The average number of follow-up months and the frequency of physician visits was similar for cases and controls. Cases and controls were well-matched.

Of the more than 800 medication classes used by the study participants, we identified 5 classes that were consistently associated (>50%) with EOCRC risk across the independently trained models. An additional 10 medication groups not consistently associated with EOCRC also had SHAP values in

the top 2.5% in at least 1 run (Table 2). Interactions between medication groups did not seem to be informative for risk prediction.

An examination of the effect estimates for the identified medication groups from the conditional regression model adjusted for the aforementioned comorbidities suggested several medication classes whose use was associated with increased odds of EOCRC. These included a beta-blocker, an angiotensin converting enzyme (ACE) inhibitor, and valerian, an herbal medication. Some protective associations were also observed, including for paroxetine, a selective serotonin uptake inhibitor. Antibiotics were not consistently associated with increased risks (Table 3).

Table 1. Demographics of patients with colorectal cancer and controls from data covering the period 2001-2019.

Demographics	Colorectal cases (n=941)	Controls (n=9410)
Sex, n (%)		
Male	459 (48.8)	4590 (48.8)
Female	482 (51.2)	4820 (51.2)
Population subgroup, n (%)		
Secular Jewish	829 (88.1)	8290 (88.1)
Jewish Ultra Orthodox	64 (6.8)	640 (6.8)
Israeli Arab	48 (5.1)	480 (5.1)
Residential area, n (%)		
North	176 (18.7)	1760 (18.7)
Center	183 (19.4)	1830 (19.4)
Jerusalem	211 (22.4)	2110 (22.4)
Sharon	212 (22.5)	2120 (22.5)
South	159 (16.9)	1590 (16.9)
Sociodemographic status[a], mean (SD)	6.5 (1.9)	6.5 (1.9)
Weeks of follow-up before index date, mean (SD)[a]	641 (273)	646 (278)
Age at index date[b] (years), mean (SD)	43.6 (6.7)	43.6 (6.6)
Physician visits over follow-up period[c], yearly mean (SD)	7.2 (6.0)	7.2 (6.0)

[a]Sociodemographic status was measured on a 10-point scale (1: lowest; 10: highest).

[b]The index date was defined as the first date of colorectal cancer diagnosis.

[c]Used as a proxy measure of the frequency of contact with the medical system, which could affect the likelihood and timing of receiving a colorectal cancer diagnosis and medication prescription patterns.

XSL•FO

RenderX

Table 2. Prevalence of the use of pharmaceutical agents in patients with EOCRC and controls using eXtreme Gradient Boosting from data covering the period 2001-2019.

ATC[a] class	ATC description	Prevalence of use in cases, n (%)	Prevalence of use in controls, n (%)	Rate of runs in which an association with CRC[b] was observed[c] (%)	Rate of runs with a global SHAP[d] value ≥97.5% of the distribution[e] (%)
c07aa05	Propranolol	42 (4.5)	228 (2.4)	70	5
c09aa02	Enalapril	44 (4.7)	161 (2.8)	60	3
n05cm09	Valerian	47 (5)	291 (3.1)	58	2
n06ab05	Paroxetine	15 (1.6)	376 (4)	54	10
c10aa01	Simvastatin	99 (10.5)	724 (7.7)	50	4
a10ba02	Metformin	40 (4.3)	234 (2.5)	46	2
d04aa13	Dimetindene	67 (7.1)	906 (9.6)	40	1
s01ca01	Dexamethansone	175 (18.6)	1986 (21.1)	36	3
d07ac13	Mometasone	103 (10.9)	1249 (13.3)	32	1
d11ax18	Diclofenac	86 (9.1)	672 (7.1)	30	1
a02ba03	Famotidine	161 (17.1)	1381 (14.7)	28	1
s01aa01	Chloramphenicol	301 (32)	2782 (29.6)	24	2
j01ce02	Phenoxymethylpenicillin	335 (35.6)	3443 (36.6)	20	1
r01ba52	Pseuroephedrine, combination	243 (25.8)	2631 (28)	20	1
b03ad01	Ferrous amino acid complex	174 (18.5)	1590 (16.9)	16	1
g01aa10	Clindamycin	79 (8.4)	685 (7.3)	10	1
r03ac03	Terbutaline	51 (5.4)	451 (4.8)	2	1

[a]ATC: Anatomical Therapeutic Chemical classification system.

[b]CRC: colorectal cancer.

[c]Based on global SHAP values calculated for each of 50 runs. A global SHAP value >0 for a given feature was considered evidence for an association with the outcome.

[d]SHAP: Shapley additive explanation.

[e]Considering each of the 50 different global SHAP values distributions individually.

Table 3. Odds ratios considering each of the 50 different global Shapley additive explanation value distributions from data covering the period 2001-2019.

ATC[a] class	ATC description	Odds ratio[b] (95% CI)	P value
c07aa05	Propranolol	1.94 (1.37-2.77)	<.001
n05cm09	Valerian	1.61 (1.15-2.25)	.01
c09aa02	Enalapril	1.34 (0.9-1.98)	.15
d11ax18	Diclofenac	1.33 (1.04-1.71)	.03
c10aa01	Simvastatin	1.3 (1-1.69)	.05
g01aa10	Clindamycin	1.22 (0.94-1.59)	.14
s01aa01	Chloramphenicol	1.19 (1.02-1.4)	.03
a02ba03	Famotidine	1.17 (0.97-1.42)	.10
r03ac03	Terbutaline	1.17 (0.86-1.59)	.32
b03ad01	Ferrous amino acid complex	1.15 (0.95-1.4)	.16
a10ba02	Metformin	1.14 (0.66-1.96)	.65
j01ce02	Phenoxymethylpenicillin	0.96 (0.82-1.11)	.58
r01ba52	Pseuroephedrine, combination	0.86 (0.73-1.02)	.07
s01ca01	Dexamethansone	0.8 (0.67-0.97)	.02
d07ac13	Mometasone	0.79 (0.63-0.98)	.04
d04aa13	Dimetindene	0.71 (0.54-0.92)	.01
n06ab05	Paroxetine	0.33 (0.2-0.57)	.001

[a]ATC: Anatomical Therapeutic Chemical classification system.

[b]Effect estimates from conditional logistic regression accounting for the matched risk sets, with additional adjustment for the average frequency of physician visits and any history of diabetes, cardiovascular disease, or excess weight prior to the index date.

Discussion

Principal Results

Our analysis of the use of pharmaceutical agents based on data from a quarter of the Israeli population during the period from 2001 to 2019, shortly after the marked increase in EOCRC incidence began, identified several medications that were used significantly more frequently by patients affected by early-onset colorectal cancer than their age-matched controls. Antibiotics, proposed in other studies as the responsible agent for the increase in EOCRC cases by changing the gut microbiome, were not used more frequently in patients with EOCRC than in controls. Agents used more frequently in patients with EOCRC included beta-blockers, ACE inhibitors, and an herbal supplement, valerian.

Comparison With Other EOCRC Research Studies

The incidence of EOCRC has steadily increased since the mid 1990s; it is estimated that of the approximately 153,000 new cases of colorectal cancer diagnosed in 2023 in the United States, 19,550 will be in people younger than 50 years of age [24,25]. By 2030, it is estimated that 1 in 10 new cases of colon cancer and 1 in 4 cases of rectal cancer will occur in young people. EOCRC is on the rise in many countries [7,26].

The abrupt increase in the incidence of EOCRC suggests that the causative agent is an environmental factor, as neither IBD colitis-associated cancers nor inherited cancer susceptibility genes have increased in incidence, and neither are found in the large majority of patients with EOCRC [3,4,27,28]. The continued increasing incidence of EOCRC further suggests that the environmental factor is still present; while the absolute number of cases is small, age cohort analyses suggest the increasing incidence is higher in younger age cohorts (eg, 20-29 years of age) and that succeeding age cohorts have a higher incidence of EOCRC [29-31]. The large majority of EOCRC cases are located on the left side of the colon, especially in the rectum [32,33]. While some studies have shown little difference in the spectrum of tumor genomic alterations between EOCRC and sporadic cancer in older patients, others have not [4,34]. A recent analysis noted differences based on gender and ethnicity in nonhypermutated tumors in patients with EOCRC [4,34,35]. However, these somatic tumor genomic analyses did not focus on the initiating driving factors in "normal" colonic mucosae.

The suspicion that an environmental factor is responsible for the increasing incidence of EOCRC has led to analyses of comorbidities including obesity, diabetes mellitus, and lifestyle factors, such as a sedentary rather than active lifestyle, smoking and alcohol use, and the widespread use of antibiotics leading to a change in the gut microbiome [8,13,15]. For alcohol use, a recent population-based study from South Korea found an association between moderate and heavy alcohol use and an increased incidence of left-sided colon and rectal cancers, but not proximal cancers [36]. The effect was greater in men than in women. However, alcohol use was self-reported and assessed at a single point, not over time, and comorbidities including IBD and a family history of colorectal cancer were not excluded. For other environmental factors, such as obesity, some studies

have found an increased EOCRC risk, but others have not [10-12]. Factors such as obesity, diabetes mellitus, more sedentary lifestyles, smoking, and alcohol use have increased gradually over decades and do not address the abrupt increase in the incidence of EOCRC.

The sharp increase in incidence suggests that a new inciting factor was introduced to a target population. This led us to hypothesize that a medication not previously widely used in young people had been introduced into general use and either directly or indirectly initiates neoplasia in colonic mucosa. A medication may have been introduced either through regulatory approval (NDAs), approval for an additional indication for an already approved drug (supplementary NDAs), a change in practice in which an already approved agent is more widely used by physicians to treat a condition other than the approved indication, or a new use of an OTC agent. While the incidence of EOCRC is increasing, it still affects only a small percentage of people under the age of 50 years. We therefore further hypothesize that in younger people who develop EOCRC, an otherwise benign genetic variant (not resulting in colonic neoplasia in the absence of exposure to the suspected medication) is potentiated by exposure to the medication to initiate or propagate neoplasia.

In order to obtain data for an initial evaluation of our hypothesis, it is necessary to have access to both detailed demographic and clinical information linked to detailed medication use data. In Israel, all citizens have access to health care through 4 health maintenance organizations. Detailed and extensive demographic, clinical, and pharmaceutical data are available to allow an analysis of pharmaceutical agents and clinical outcomes. The databases of Israeli health maintenance organizations have enabled the assessment of the risk of vaccine toxicities during the COVID-19 pandemic, evaluating, for example, the incidence and patients affected by myocarditis [37]. For our study, we used digitized records including pharmaceutical use both of prescribed and OTC agents from a large segment of the Israeli population as the source for our data. Our analysis indicated that, as we hypothesized, the use of several pharmaceutical agents was more common in young people diagnosed with EOCRC when compared to their control peers.

In regard to the agents that we have identified in our initial analysis, for propranolol, an agent one would not ordinarily expect to be prescribed to young people, we suspect its use was to treat anxiety or for side effects of antianxiety agents. Antihypertensives, such as ACE inhibitors, may be prescribed more commonly in young people as both stress and other conditions, such as obesity, increase the risk for hypertension. Both propranolol and valerian have been reported to have suppressive effects on established cancer cells; to our knowledge, there are no data as to their effects in inducing dysplasia and cancer in normal colon mucosae cells [38-40]. Data concerning ACE inhibitors and the risk for cancer are conflicting, with a recent report indicating null associations between use of ACE inhibitors and colorectal cancer risk [41]. Previous studies have suggested an association between pharmaceutical agents and cancer, including insulin use and breast cancer, antihypertensives and skin cancer, and sitagliptin and pancreas cancer, but these associations have not yet been

confirmed. Yang et al [42] recently reported a web-based algorithm to assess cancer risk for 6 classes of pharmaceuticals (including antihypertensives, antihyperuricemics, antihyperlipidemics, nonsteroidal anti-inflammatory drugs, and antianxiety agents) in Taiwanese patients. However, the study only included agents used within 2 years of cancer diagnosis, which is a period during which the cancer may have already developed. For EOCRC, only 4 patients were between ages of 20-39 years, and the number of patients younger than 50 years was not provided.

The agents we have identified in our initial analysis do not include antibiotics. While it has been suggested that the widespread use of antibiotics during childhood and adolescence may underlie a carcinogenic effect on the gut microbiome, which may be translated into an increased incidence of EOCRC, the data are conflicting. In a recent British study, antibiotic consumption was associated with colorectal cancer in both older and EOCRC cohorts, while a population-based case-control study failed to find conclusive evidence that antibiotics were associated with EOCRC risk [43,44]. In a population-based analysis of the effect of cesarian section (removing exposure to the maternal vaginal flora), while a trend for a higher rate of EOCRC in women born by vaginal delivery was found, the overall analysis did not show a difference in the incidence of EOCRC in individuals born by cesarian section versus those undergoing vaginal delivery [45].

If our hypothesis is correct and validation studies confirm that a pharmaceutical agent or class of agents is a driver in the development of EOCRC, this would lead to regulatory bodies requiring strategies to mitigate risk [46]. Guidance to industry from the US Food and Drug Administration, for example, provides instructions for determining the level of risk, the patient populations at risk, and when available, biomarkers to determine if an individual patient is at risk. If a companion diagnostic test (eg, a genetic polymorphism found in patients with EOCRC but not in their unaffected peers) is developed, this would be included in the labeling indication. Examples of risk mitigation strategies for other illnesses include labeling indication instructions to identify patients with CYP2C19 alleles that poorly or almost totally do not metabolize clopidogel (Plavix) to minimize the risk of using an ineffective drug or drug dose [47]. For thalidomide, no safe dose has been identified for use of this agent in women who are pregnant, and the labeling indication includes a black box warning.

Strengths and Limitations

Strengths of our study include the use of an EMR including digitized pharmaceutical data covering a large percentage of an entire country's population, long-term follow-up, and the use of an advanced machine learning algorithm. We recognize that while we used extensive matching and adjustment for possible confounders, as well as repeated sampling of the data to enhance reproducibility, our results are still susceptible to confounding by indication. That is, rather than reflecting an association with the medications, the observed finding could reflect associations with the underlying conditions for which the medications were prescribed [48]. We also recognize the sample size limitations and have not drawn causality conclusions regarding the agents

identified as being used more frequently in patients with EOCRC compared to controls. To assess the risk of publication bias, we systematically reviewed the literature with regard to our hypothesis that a pharmaceutical agent is a driving factor in the abrupt increase in the incidence of EOCRC and found no publications to date that have studied this hypothesis specifically in EOCRC. We regard our current data as supporting our hypothesis but note that validation studies from similar health care system databases are needed. Moreover, prospective studies to identify the mechanism of action of pharmaceutical agents used more frequently by patients with EOCRC than controls on normal colonic mucosae in initiating a neoplastic process in a susceptible population of younger adults are required. We have initiated such studies.

In conclusion, our analysis suggests that EOCRC may be associated with the prior use of specific medications. Additional studies of pharmaceutical agents as possible drivers of EOCRC and the potential mechanisms of action for this effect are underway.

Data Availability

According to Israel Ministry of Health regulations and Maccabi Healthcare Services policy, individual-level data cannot be shared openly (even if deidentified). Specific requests for remote access to deidentified community-level data will be reviewed by Kahn-Sagol-Maccabi, Maccabi Healthcare Services Research and Innovation Center.

Conflicts of Interest

AC declares research support from Seagen and GSK and an advisory role with Merck, Seagen, Pfizer, Bayer, Janssen, and GSK. The other authors have no conflicts of interest to disclose.

References

1. Akimoto N, Ugai T, Zhong R, Hamada T, Fujiyoshi K, Giannakis M, et al. Rising incidence of early-onset colorectal cancer - a call to action. Nat Rev Clin Oncol. 2021 Apr;18(4):230-243 [FREE Full text] [doi: 10.1038/s41571-020-00445-1] [Medline: 33219329]
2. Cavestro GM, Mannucci A, Zuppardo RA, Di Leo M, Stoffel E, Tonon G. Early onset sporadic colorectal cancer: worrisome trends and oncogenic features. Dig Liver Dis. 2018 Jun;50(6):521-532 [doi: 10.1016/j.dld.2018.02.009] [Medline: 29615301]
3. Spaander MCW, Zauber AG, Syngal S, Blaser MJ, Sung JJ, You YN, et al. Young-onset colorectal cancer. Nat Rev Dis Primers. 2023 Apr 27;9(1):21 [FREE Full text] [doi: 10.1038/s41572-023-00432-7] [Medline: 37105987]
4. Cercek A, Chatila WK, Yaeger R, Walch H, Fernandes GDS, Krishnan A, et al. A comprehensive comparison of early-onset and average-onset colorectal cancers. J Natl Cancer Inst. 2021 Nov 29;113(12):1683-1692 [FREE Full text] [doi: 10.1093/jnci/djab124] [Medline: 34405229]
5. Ben-Aharon I, van Laarhoven HWM, Fontana E, Obermannova R, Nilsson M, Lordick F. Early-onset cancer in the gastrointestinal tract is on the rise-evidence and implications. Cancer Discov. 2023 Mar 01;13(3):538-551 [doi: 10.1158/2159-8290.CD-22-1038] [Medline: 36757194]
6. Siegel RL, Torre LA, Soerjomataram I, Hayes RB, Bray F, Weber TK, et al. Global patterns and trends in colorectal cancer incidence in young adults. Gut. 2019 Dec;68(12):2179-2185 [doi: 10.1136/gutjnl-2019-319511] [Medline: 31488504]
7. Anugwom C, Braimoh G, Sultan A, Johnson WM, Debes JD, Mohammed A. Epidemiology and genetics of early onset colorectal cancer-African overview with a focus on Ethiopia. Semin Oncol. 2023;50(1-2):28-33 [doi: 10.1053/j.seminoncol.2023.03.007] [Medline: 37032270]
8. Li H, Chen X, Hoffmeister M, Brenner H. Associations of smoking with early- and late-onset colorectal cancer. JNCI Cancer Spectr. 2023 Jan 03;7(1):pkad004 [FREE Full text] [doi: 10.1093/jncics/pkad004] [Medline: 36759940]
9. Chen J, Zhang IL, Terry MB, Yang W. Dietary factors and early-onset colorectal cancer in the United States-an ecologic analysis. Cancer Epidemiol Biomarkers Prev. 2023 Feb 06;32(2):217-225 [FREE Full text] [doi: 10.1158/1055-9965.EPI-22-0442] [Medline: 36129804]
10. Stoffel EM, Murphy CC. Epidemiology and mechanisms of the increasing incidence of colon and rectal cancers in young adults. Gastroenterology. 2020 Jan;158(2):341-353 [FREE Full text] [doi: 10.1053/j.gastro.2019.07.055] [Medline: 31394082]
11. Liu P, Wu K, Ng K, Zauber AG, Nguyen LH, Song M, et al. Association of obesity with risk of early-onset colorectal cancer among women. JAMA Oncol. 2019 Jan 01;5(1):37-44 [FREE Full text] [doi: 10.1001/jamaoncol.2018.4280] [Medline: 30326010]
12. Low EE, Demb J, Liu L, Earles A, Bustamante R, Williams CD, et al. Risk factors for early-onset colorectal cancer. Gastroenterology. 2020 Aug;159(2):492-501.e7 [FREE Full text] [doi: 10.1053/j.gastro.2020.01.004] [Medline: 31926997]
13. Hua H, Jiang Q, Sun P, Xu X. Risk factors for early-onset colorectal cancer: systematic review and meta-analysis. Front Oncol. 2023;13:1132306 [FREE Full text] [doi: 10.3389/fonc.2023.1132306] [Medline: 37213277]

XSL•FO

RenderX

14. Zhang J, Haines C, Watson AJM, Hart AR, Platt MJ, Pardoll DM, et al. Oral antibiotic use and risk of colorectal cancer in the United Kingdom, 1989-2012: a matched case-control study. Gut. 2019 Nov;68(11):1971-1978 [doi: 10.1136/gutjnl-2019-318593] [Medline: 31427405]

15. Kong C, Liang L, Liu G, Du L, Yang Y, Liu J, et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. Gut. 2023 Jun;72(6):1129-1142 [doi: 10.1136/gutjnl-2022-327156] [Medline: 35953094]

16. Zhuo X, Melzer Cohen C, Chen J, Chodick G, Alsumali A, Cook J. Validating the UK prospective diabetes study outcome model 2 using data of 94,946 Israeli patients with type 2 diabetes. J Diabetes Complications. 2022 Jan;36(1):108086 [doi: 10.1016/j.jdiacomp.2021.108086] [Medline: 34799250]

17. Weitzman D, Chodick G, Shalev V, Grossman C, Grossman E. Prevalence and factors associated with resistant hypertension in a large health maintenance organization in Israel. Hypertension. 2014 Sep;64(3):501-507 [doi: 10.1161/HYPERTENSIONAHA.114.03718] [Medline: 24958503]

18. Chodick G, Omer-Gilon M, Derazne E, Puris G, Rotem R, Tzur D, et al. Adolescent body mass index and changes in pre-pregnancy body mass index in relation to risk of gestational diabetes. EClinicalMedicine. 2021 Dec;42:101211 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101211] [Medline: 34849479]

19. Shalev V, Chodick G, Goren I, Silber H, Kokia E, Heymann AD. The use of an automated patient registry to manage and monitor cardiovascular conditions and related outcomes in a large health organization. Int J Cardiol. 2011 Nov 03;152(3):345-349 [doi: 10.1016/j.ijcard.2010.08.002] [Medline: 20826019]

20. Chodick G, Heymann AD, Shalev V, Kookia E. The epidemiology of diabetes in a large Israeli HMO. Eur J Epidemiol. 2003;18(12):1143-1146 [doi: 10.1023/b:ejep.0000006635.36802.c8] [Medline: 14758871]

21. Chen T. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16; August 13-17, 2016; San Francisco [doi: 10.1145/2939672.2939785]

22. Bernd B, Bossek J, Horn D, Thomas J, Lang M. mlrMBO: a modular framework for model-based optimization of expensive black-box functions. arXiv. Preprint posted online on March 9, 2017. [FREE Full text]

23. Lundberg SM, Lee S. A unified approach to interpreting model predictions. 2017 Presented at: 31st Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach

24. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020 May;70(3):145-164 [FREE Full text] [doi: 10.3322/caac.21601] [Medline: 32133645]

25. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233-254 [FREE Full text] [doi: 10.3322/caac.21772] [Medline: 36856579]

26. Huang J, Lucero-Prisno DE, Zhang L, Xu W, Wong SH, Ng SC, et al. Updated epidemiology of gastrointestinal cancers in East Asia. Nat Rev Gastroenterol Hepatol. 2023 May;20(5):271-287 [doi: 10.1038/s41575-022-00726-3] [Medline: 36631716]

27. Mosher CA, Brown GR, Weideman RA, Crook TW, Cipher DJ, Spechler SJ, et al. Incidence of colorectal cancer and extracolonic cancers in veteran patients with inflammatory bowel disease. Inflamm Bowel Dis. 2018 Feb 15;24(3):617-623 [doi: 10.1093/ibd/izx046] [Medline: 29390104]

28. Malham M, Jakobsen C, Paerregaard A, Virta LJ, Kolho K, Wewer V. The incidence of cancer and mortality in paediatric onset inflammatory bowel disease in Denmark and Finland during a 23-year period: a population-based study. Aliment Pharmacol Ther. 2019 Jul;50(1):33-39 [doi: 10.1111/apt.15258] [Medline: 31069829]

29. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, et al. Colorectal cancer incidence patterns in the United States, 1974-2013. J Natl Cancer Inst. 2017 Aug 01;109(8):djw322 [FREE Full text] [doi: 10.1093/jnci/djw322] [Medline: 28376186]

30. Dwyer AJ, Murphy CC, Boland CR, Garcia R, Hampel H, Limburg P, et al. A summary of the fight colorectal cancer working meeting: exploring risk factors and etiology of sporadic early-age onset colorectal cancer. Gastroenterology. 2019 Aug;157(2):280-288 [FREE Full text] [doi: 10.1053/j.gastro.2019.04.049] [Medline: 31095950]

31. Tanaka LF, Figueroa SH, Popova V, Klug SJ, Buttmann-Schweiger N. The rising incidence of early-onset colorectal cancer. Dtsch Arztebl Int. 2023 Feb 03;120(Forthcoming):59-64 [FREE Full text] [doi: 10.3238/arztebl.m2022.0368] [Medline: 36471648]

32. Tom CM, Mankarious MM, Jeganathan NA, Deutsch M, Koltun WA, Berg AS, et al. Characteristics and outcomes of right-versus left-sided early-onset colorectal cancer. Dis Colon Rectum. 2023 Apr 01;66(4):498-510 [doi: 10.1097/DCR.0000000000002273] [Medline: 35001052]

33. Sinicrope FA. Increasing incidence of early-onset colorectal cancer. N Engl J Med. 2022 Apr 21;386(16):1547-1558 [doi: 10.1056/NEJMra2200869] [Medline: 35443109]

34. Marx O, Mankarious M, Yochum G. Molecular genetics of early-onset colorectal cancer. World J Biol Chem. 2023 Mar 27;14(2):13-27 [FREE Full text] [doi: 10.4331/wjbc.v14.i2.13] [Medline: 37034132]

35. Holowatyj AN, Wen W, Gibbs T, Seagle HM, Keller SR, Edwards DRV, et al. Racial/ethnic and sex differences in somatic cancer gene mutations among patients with early-onset colorectal cancer. Cancer Discov. 2023 Mar 01;13(3):570-579 [FREE Full text] [doi: 10.1158/2159-8290.CD-22-0764] [Medline: 36520636]

36. Jin EH, Han K, Shin CM, Lee DH, Kang SJ, Lim JH, et al. Sex and tumor-site differences in the association of alcohol intake with the risk of early-onset colorectal cancer. J Clin Oncol. 2023 Aug 01;41(22):3816-3825 [FREE Full text] [doi: 10.1200/JCO.22.01895] [Medline: 37315287]

37. Witberg G, Barda N, Hoss S, Richter I, Wiessman M, Aviv Y, et al. Myocarditis after Covid-19 vaccination in a large health care organization. N Engl J Med. 2021 Dec 02;385(23):2132-2139 [FREE Full text] [doi: 10.1056/NEJMoa2110737] [Medline: 34614329]

38. Anselmino LE, Baglioni MV, Malizia F, Laluce NC, Etichetti CB, Marignac VLM, et al. Repositioning metformin and propranolol for colorectal and triple negative breast cancers treatment. Sci Rep. 2021 Apr 14;11(1):8091 [FREE Full text] [doi: 10.1038/s41598-021-87525-z] [Medline: 33854147]

39. Liao P, Song K, Zhu Z, Liu Z, Zhang W, Li W, et al. Propranolol suppresses the growth of colorectal cancer through simultaneously activating autologous CD8 T cells and inhibiting tumor AKT/MAPK pathway. Clin Pharmacol Ther. 2020 Sep;108(3):606-615 [doi: 10.1002/cpt.1894] [Medline: 32418204]

40. Shi F, Li Y, Han R, Fu A, Wang R, Nusbaum O, et al. Valerian and valeric acid inhibit growth of breast cancer cells possibly by mediating epigenetic modifications. Sci Rep. 2021 Jan 28;11(1):2519 [FREE Full text] [doi: 10.1038/s41598-021-81620-x] [Medline: 33510252]

41. Brasky TM, Flores KF, Larson JC, Newton AM, Shadyab AH, Watanabe JH, et al. Associations of angiotensin-converting enzyme inhibitor or angiotensin receptor blocker use with colorectal cancer risk in the women's health initiative. Cancer Epidemiol Biomarkers Prev. 2021 May;30(5):1029-1032 [FREE Full text] [doi: 10.1158/1055-9965.EPI-20-1401] [Medline: 33627382]

42. Yang H, Islam MM, Nguyen PAA, Wang C, Poly TN, Huang C, et al. Development of a web-based system for exploring cancer risk with long-term use of drugs: logistic regression approach. JMIR Public Health Surveill. 2021 Feb 15;7(2):e21401 [FREE Full text] [doi: 10.2196/21401] [Medline: 33587043]

43. McDowell R, Perrott S, Murchie P, Cardwell C, Hughes C, Samuel L. Oral antibiotic use and early-onset colorectal cancer: findings from a case-control study using a national clinical database. Br J Cancer. 2022 Apr;126(6):957-967 [FREE Full text] [doi: 10.1038/s41416-021-01665-7] [Medline: 34921228]

44. Nguyen LH, Cao Y, Batyrbekova N, Roelstraete B, Ma W, Khalili H, et al. Antibiotic therapy and risk of early-onset colorectal cancer: a national case-control study. Clin Transl Gastroenterol. 2022 Jan 13;13(1):e00437 [FREE Full text] [doi: 10.14309/ctg.0000000000000437] [Medline: 35029165]

45. Cao Y, Nguyen LH, Tica S, Otegbeye E, Zong X, Roelstraete B, et al. Evaluation of birth by cesarean delivery and development of early-onset colorectal cancer. JAMA Netw Open. 2023 Apr 03;6(4):e2310316 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.10316] [Medline: 37103933]

46. Guidance for industry: warnings and precautions, contraindications, boxed warning sections of labeling for human prescription drug and biological products – content and format. Food and Drug Administration. URL: https://www.fda.gov/downloads/drugs/guidances/ucm075096.pdf [accessed 2023-11-24]

47. FDA drug safety communication: reduced effectiveness of Plavix (clopidogrel) in patients who are poor metabolizers of the drug. Food and Drug Administration. URL: http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm203888.htm [accessed 2023-11-24]

48. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. JAMA. 2016 Nov 01;316(17):1818-1819 [doi: 10.1001/jama.2016.16435] [Medline: 27802529]

Abbreviations

ACE: angiotensin converting enzyme
EMR: electronic medical record
EOCRC: early-onset colorectal cancer
IBD: inflammatory bowel disease
ICD-9: International Classification of Diseases, Ninth Revision
INCR: Israel National Cancer Registry
MHS: Maccabi Healthcare Services
NDA: new drug application
OR: odds ratio
OTC: over-the-counter
SHAP: Shapley additive explanations
XGBoost: eXtreme Gradient Boosting