

# UC San Diego

## UC San Diego Previously Published Works

### Title

Risk factors for colorectal cancer significantly vary by anatomic site.

### Permalink

<https://escholarship.org/uc/item/9fh8n52t>

### Journal

BMJ open gastroenterology, 6(1)

### ISSN

2054-4774

### Authors

Demb, Joshua  
Earles, Ashley  
Martínez, María Elena  
[et al.](#)

### Publication Date

2019

### DOI

10.1136/bmjgast-2019-000313

Peer reviewed

# Risk factors for colorectal cancer significantly vary by anatomic site

Joshua Demb,<sup>1</sup> Ashley Earles,<sup>2</sup> María Elena Martínez,<sup>1,3</sup> Ranier Bustamante,<sup>2</sup> Alex K Bryant,<sup>4</sup> James D Murphy,<sup>4</sup> Lin Liu,<sup>1,2,3</sup> Samir Gupta<sup>1,5,6</sup>

**To cite:** Demb J, Earles A, Martínez ME, *et al*. Risk factors for colorectal cancer significantly vary by anatomic site. *BMJ Open Gastro* 2019;**6**:e000313. doi:10.1136/bmjgast-2019-000313

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjgast-2019-000313>)

LL and SG are co-senior authors.

JD and AE are co-first authors.

Received 11 May 2019

Revised 3 July 2019

Accepted 20 July 2019

## ABSTRACT

**Objective** To conduct an anatomic site-specific case-control study of candidate colorectal cancer (CRC) risk factors.

**Design** Case-control study of US veterans with >1 colonoscopy during 1999–2011. Cases had cancer registry-identified CRC at colonoscopy, while controls were CRC free at colonoscopy and within 3 years of colonoscopy. Primary outcome was CRC, stratified by anatomic site: proximal, distal, or rectal. Candidate risk factors included age, sex, race/ethnicity, body mass index, height, diabetes, smoking status, and aspirin exposure summarised by adjusted ORs and 95% CIs.

**Results** 21 744 CRC cases (n=7017 rectal; n=7039 distal; n=7688 proximal) and 612 646 controls were included. Males had significantly higher odds relative to females for rectal cancer (OR=2.84, 95% CI 2.25 to 3.58) than distal cancer (OR=1.84, 95% CI 1.50 to 2.24). Relative to whites, blacks had significantly lower rectal cancer odds (OR=0.88, 95% CI 0.82 to 0.95), but increased distal (OR=1.27, 95% CI 1.19 to 1.37) and proximal odds (OR=1.62, 95% CI 1.52 to 1.72). Diabetes prevalence was more strongly associated with proximal (OR=1.29, 95% CI 1.22 to 1.36) than distal (OR=1.15, 95% CI 1.08 to 1.22) or rectal cancer (OR=1.12, 95% CI 1.06 to 1.19). Current smoking was more strongly associated with rectal cancer (OR=1.81, 95% CI 1.68 to 1.95) than proximal cancer (OR=1.53, 95% CI 1.43 to 1.65) or distal cancer (OR=1.46, 95% CI 1.35 to 1.57) compared with never smoking. Aspirin use was significantly more strongly associated with reduced rectal cancer odds (OR=0.71, 95% CI 0.67 to 0.76) than distal (OR=0.85, 95% CI 0.81 to 0.90) or proximal (OR=0.91, 95% CI 0.86 to 0.95).

**Conclusion** Candidate CRC risk factor associations vary significantly by anatomic site. Accounting for site may enable better insights into CRC pathogenesis and cancer control strategies.

## INTRODUCTION

Colorectal cancer (CRC) is the third leading cause of cancer and cancer-related deaths worldwide.<sup>1</sup> CRCs may be divided into three anatomic sites: proximal cancers, generally including cancers of the caecum, ascending colon, hepatic flexure, and transverse colon; distal cancers, including the descending and sigmoid colon; and rectal cancer, where the splenic flexure is variably grouped with

## Summary box

### What is already known about this subject?

► Colorectal cancer (CRC) is the third leading cause of cancer and cancer-related deaths in the world. While many studies have reported on key risk factors associated with CRC risk, such as age, sex, race/ethnicity, body mass index, height, diabetes, smoking status and aspirin exposure, anatomic site-specific differences have not been widely investigated.

### What are the new findings?

► Candidate CRC risk factor associations varied markedly by anatomic subsite. Increased CRC risk for males was most closely associated with rectal cancer. Compared with whites, blacks had reduced risk of rectal cancer, but increased risk of distal and proximal colon cancer. Diabetes prevalence most significantly associated with increased proximal cancer risk compared with distal or rectal cancer. Compared with never smoking, smoking was most significantly associated with rectal cancer compared with proximal or distal cancer. Aspirin was associated with reduced risk for all three subsites, though most strongly associated with reduced risk for rectal cancer.

### How might it impact on clinical practice in the foreseeable future?

► Our findings indicate that risk factor associations in CRC vary by anatomic subsite, which could further contribute to differences in tumour presentation. Learning more about how risk factors are associated with specific CRC sites may enable better insights into CRC pathogenesis, and guide future cancer control strategies.

either proximal or distal location.<sup>2,3</sup> Current research shows that embryologic origins, associated microbial milieu, and tumour characteristics such as mutational signatures and histological features differ by anatomic site.<sup>4–6</sup> For example, microsatellite instability is more commonly observed in proximal compared with distal or rectal cancer, and proximal cancers have also been linked more commonly with grade 3/4 cancers and mucinous histology.<sup>5,7,8</sup> Though site-specific differences in tumour characteristics are likely in



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

### Correspondence to

Dr Samir Gupta;  
s1gupta@ucsd.edu



part driven by differences in aetiological factors, most observational studies of CRC risk factors lack the ability to examine risk factors for CRC by specific anatomic sites.

The lack of site-specific analyses to date may have hindered our ability to understand whether some risk factors may be particularly important (or unimportant) for cancer development, and even possibly precluded our ability to identify risk factor associations specific to an anatomic site. More specificity could guide studies of pathogenesis and have implications for cancer control strategies.

In addition to issues with anatomic site-specific case definition, prior case–control studies of CRC risk factors often lack controls with normal colonoscopy documenting absence of polyps or CRC. Because colorectal neoplasia is common, inclusion of controls with undiagnosed neoplasia (such as polyps) in case–control studies of CRC can reduce the ability to accurately estimate risk factor associations, and potentially lead to failures to identify true associations.

To address these gaps in the current CRC literature, our aim was to conduct an anatomic site-specific case–control study of candidate risk factors for CRC with normal colonoscopy controls using large-scale national data from the US Veterans Health Administration (VHA).

## METHODS

### Study design, setting and data sources

We conducted a retrospective case–control study to explore the anatomic site-specific risk factors for CRC among US veterans receiving colonoscopy at the VHA. The Department of Veteran Affairs (VA) is one of the largest integrated healthcare providers in the USA, caring for over 6 million veterans annually.<sup>9</sup> Since 1999, all VA sites have used an integrated electronic health record (EHR) for documentation of clinical encounters, which, along with additional healthcare resources, can be accessed for research.

The VA Corporate Data Warehouse provides access to discrete data, including demographic characteristics, administrative claims-based diagnosis and procedure codes, prescriptions, and anthropometric measures (eg, weight and height), as well as free-text data, including procedure notes and pathology reports. CRC was ascertained by the VA Central Cancer Registry (VACCR), which has been shown to accurately identify 90% of CRC cases.<sup>10</sup> Follow-up was ascertained by the VHA Vital Status File including the date of last visit, represented as the date and time the last vital record was taken by the healthcare provider.<sup>11</sup>

### Study sample and selection criteria

Our study sample consisted of veterans with at least one Current Procedural Terminology (CPT) code for colonoscopy from 1999 to 2011 (see online supplementary appendix table A for codes used). We excluded veterans with a history of CRC based on VACCR entry or an

International Classification of Diseases, Ninth Revision (ICD-9) diagnosis code issued  $\geq 6$  months prior to baseline colonoscopy, as well as those with an ICD-9 code consistent with inflammatory bowel disease prior to and up to 6 months after baseline colonoscopy. We also excluded individuals with less than 3 years of follow-up. Full inclusion and exclusion criteria are noted in online supplementary appendix table B.

### Case selection

Cases were identified by the VACCR and defined using International Classification of Diseases, Oncology, Third Revision (ICD-O-3) site codes for CRC (C18.0, C18.2–C18.7, C19.9, C20.9). For cases identified within (before or after) 6 months from date of baseline (first) colonoscopy, Surveillance, Epidemiology, and End Results (SEER) programme summary stage and histology were extracted. Online supplementary appendix table C includes details of our selection criteria. We excluded cases with unknown SEER stage, carcinoma in situ, or ICD-O-3 histology codes not consistent with adenocarcinoma. If tumour histology was not specified/available, we allowed for case inclusion as long as site, stage, and diagnosis date information was available, given that the majority of CRCs are adenocarcinomas. Cases were stratified into sites based on site codes as proximal (C18.0, C18.2–C18.4), distal (C18.5–C18.7) and rectal (C19.9, C20.9).

### Control selection

Controls were veterans with no prior CRC diagnosis, normal baseline colonoscopy defined by presence of a CPT code for diagnostic colonoscopy only (45378 or G0121), and absence of a colon biopsy (as evidenced by absence of a pathology report within 30 days of baseline colonoscopy) (see online supplementary appendix table C for full outline of study selection criteria). Our prior work has shown that this approach is 96.3% sensitive and 97.5% specific for normal colonoscopy and had a positive predictive value of 97%.<sup>12 13</sup> Additionally, to avoid inclusion of controls with missed CRC at baseline colonoscopy, controls with CRC diagnosed by VACCR or an ICD-9 code within up to 3 years of baseline colonoscopy were excluded. If a candidate control had less than 3 years of follow-up (due to death or lost to follow-up at VA), they were excluded to ensure that controls were CRC free.

### Candidate risk factors

Candidate risk factors were ascertained based on presence at time of baseline colonoscopy, and included age, sex, race/ethnicity, body mass index (BMI), height, diabetes, smoking status, and aspirin exposure. BMI was characterised using previously developed criteria—using a median weight derived from 3 years of weight measurements, and a single height measure—that included removal of biologically implausible values.<sup>14</sup> Diabetes was defined using a previously validated algorithm that included inpatient

visits, outpatient visits and medications.<sup>15</sup> Smoking status was classified into current, former, never and unknown.<sup>10</sup> Aspirin exposure was defined as at least two prescriptions or two mentions of aspirin in free-text notes up to 1 year prior to colonoscopy. We have shown this approach has a positive predictive value and a negative predictive value of 99.2% and 97.5%, respectively, for capturing EHR-documented aspirin use.<sup>16</sup>

### Statistical analyses

The primary outcome was CRC, stratified by anatomic site as proximal, distal, and rectal cancers. Risk factors were summarised by descriptive statistics and compared between sites using univariate tests (Kruskal-Wallis test and  $\chi^2$  test). We used multinomial logistic regression to examine the risk factors for CRC at three anatomic subsites. The multinomial logistic regression can be expressed as  $\log(p_i/p_0) = b_0 + b_1x_1 + \dots + b_px_p$ , where  $p_i$  is the probability that the subject is in the  $i$ th group,  $p_0$  is the probability that the subject is in the reference group and  $x_i$  is the risk factor. In the initial analysis, the group 'normal control' was specified as the reference group. To control for time trends in the performance of colonoscopy, and distribution of risk factors, we considered calendar year of procedure a priori as a potential confounder in our analyses. All risk factors were included simultaneously in one model thus effect estimates are interpreted as associations independent of other risk factors. Anatomic site-specific ORs with 95% CIs for each candidate risk factor were estimated, using adjusted models. For simplicity, we interpret our outcomes of ORs in the context of 'risk'. Furthermore, we define increased risk as factors associated with OR >1, decreased risk as factors associated with OR <1, and 95% CIs not crossing unity as indicating statistical significance. If unadjusted and adjusted analyses show similar results, we presented the adjusted findings only in our results.

In order to assess whether CRC risks across anatomic subsites were statistically different, we also ran the multinomial logistic regression with 'proximal' and 'distal' specified as the reference category to allow for case-case comparisons of proximal versus distal, proximal versus rectal, and distal versus rectal risk. For case-case comparisons,  $p < 0.05$  was interpreted as statistically significant.

To consider the impact of potential immortal time bias resulting from our requirement that controls be cancer free at baseline colonoscopy through 3 years after colonoscopy follow-up, we performed a sensitivity analysis where we compared cases to controls who were cancer free within 6 months rather than 3 years. Because of observation of lower odds of CRC with increasing BMI, post hoc analyses assessed trends in BMI at 5 years and 10 years prior to index colonoscopy.

Analyses were performed using R V.3.5.1 and Stata V.15 (StataCorp, College Station, TX).<sup>17</sup>

## RESULTS

From a study base of 1 878 429 veterans with colonoscopy during 1999–2011, we identified 21 744 CRC cases ( $n=7017$  rectal;  $n=7039$  distal;  $n=7688$  proximal) and 612 646 controls which were CRC free at colonoscopy.

Table 1 shows the demographic characteristics of all cases and controls. For all CRC cases combined versus controls, median age was 68 years vs 61 years, 98% vs 95% were male, median BMI was 27.9 kg/m<sup>2</sup> vs 28.9 kg/m<sup>2</sup>, 28% vs 24% had diabetes, and 25% vs 30% were never smokers; race/ethnicity groups were similar. Proximal cases were older than distal or rectal cases, had higher rates of diabetes and aspirin exposure, and were more likely to be non-Hispanic blacks. Distal cases had the highest BMI (28.4 kg/m<sup>2</sup>) compared with proximal (27.9 kg/m<sup>2</sup>) and rectal (27.3 kg/m<sup>2</sup>) cases.

Figure 1 shows the anatomic site-specific risk factor associations for CRC cases compared with normal colonoscopy controls by anatomic site, adjusted for other candidate factors. Online supplementary appendix figure 1 depicts univariate associations between candidate factors and site-specific CRC odds.

### Demographic factors

Age was associated with increased odds of CRC in all three subsites in both unadjusted and adjusted analyses. Five-year increase in age was associated with a significantly higher odds for proximal cancer (OR=1.58, 95% CI 1.56 to 1.59) compared with distal (OR=1.34, 95% CI 1.33 to 1.36) and rectal cancer odds (OR=1.29, 95% CI 1.28 to 1.30;  $p < 0.05$  for all case-case comparisons).

CRC odds were increased for males compared with females across all sites in unadjusted analyses. However, in adjusted analyses, CRC odds were increased for males compared with females for only distal cancer (OR=1.84, 95% CI 1.50 to 2.24) and rectal cancer (OR=2.84, 95% CI 2.25 to 3.58); odds were significantly higher for the comparison between males versus females for rectal compared with distal cancer based on case-case analyses ( $p < 0.05$  for case-case comparisons).

In unadjusted analyses, compared with non-Hispanic whites, blacks had no increased odds of distal cancer, increased odds of proximal cancer (OR=1.24, 95% CI 1.17 to 1.31) and decreased odds of rectal cancer (OR=0.77, 95% CI 0.72 to 0.83). In adjusted analyses, blacks had reduced odds of rectal cancer (OR=0.88, 95% CI 0.82 to 0.95), but increased odds of proximal cancer (OR=1.62, 95% CI 1.52 to 1.72) and distal cancer (OR=1.27, 95% CI 1.19 to 1.37); the ORs for the three subsite cancers are significantly different ( $p < 0.05$  for all case-case comparisons).

In unadjusted analyses, Hispanics had increased odds of distal cancer (OR=1.28, 95% CI 1.15 to 1.43) compared with non-Hispanic whites. In adjusted analyses, Hispanics had increased odds of cancer for all three sites (OR=1.57, 95% CI 1.40 to 1.77 for distal; OR=1.36, 95% CI 1.19 to 1.54 for proximal; OR=1.32, 95% CI 1.17 to 1.50 for rectal). Among Hispanics, distal cancer odds



**Table 1** Demographic characteristics of CRC cases versus normal colonoscopy controls

Demographics	Total CRC n=21 744	Proximal n=7688	Distal n=7039	Rectal n=7017	Controls* n=612 646
Age, median (Q1–Q3)	68 (60–76)	71 (62–78)	66 (60–75)	65 (59–74)	61 (55–68)
Sex, n (%)					
Male	21 364 (98.3)	7514 (97.7)	6923 (98.4)	6927 (98.7)	580 369 (94.7)
Female	380 (1.7)	174 (2.3)	116 (1.6)	90 (1.3)	32 277 (5.3)
Race/ethnicity, n (%)					
Non-Hispanic white	14 838 (68.2)	5076 (66.0)	4752 (67.5)	5010 (71.4)	420 131 (68.6)
Non-Hispanic black	3735 (17.2)	1559 (20.3)	1216 (17.3)	960 (13.7)	104 292 (17.0)
Hispanic	959 (4.4)	303 (3.9)	347 (4.9)	309 (4.4)	23 994 (3.9)
Asian/Pacific Islander	177 (0.8)	53 (0.7)	60 (0.9)	64 (0.9)	7306 (1.2)
American Indian/Alaska Native	95 (0.4)	32 (0.4)	25 (0.4)	38 (0.5)	3177 (0.5)
Other/multiracial	383 (1.8)	124 (1.6)	142 (2.0)	117 (1.7)	11 991 (2.0)
Unknown	1557 (7.2)	541 (7.0)	497 (7.1)	519 (7.4)	41 755 (6.8)
BMI, median (Q1–Q3)	27.9 (24.7–31.7)	27.9 (24.7–31.7)	28.4 (25.2–32.3)	27.3 (24.1–31.0)	28.9 (25.8–32.5)
BMI (categorical), n (%)					
Underweight	299 (1.4)	100 (1.3)	74 (1.1)	125 (1.8)	3624 (0.6)
Normal	4788 (22.0)	1667 (21.7)	1350 (19.2)	1771 (25.2)	101 309 (16.5)
Overweight	7187 (33.1)	2558 (33.3)	2340 (33.2)	2289 (32.6)	215 161 (35.1)
Obese	6437 (29.6)	2266 (29.5)	2337 (33.2)	1834 (26.1)	221 178 (36.1)
Unknown	3033 (13.9)	1097 (14.3)	938 (13.3)	998 (14.2)	71 374 (11.7)
Height, median (Q1–Q3)	69.8 (68–71.5)	69.5 (67.8–71.5)	70 (68–71.8)	69.8 (68–71.5)	69.6 (67.5–71.5)
Diabetes, n (%)	6085 (28.0)	2371 (30.8)	2004 (28.5)	1710 (24.4)	148 293 (24.2)
Smoking status, n (%)					
Current	4486 (20.6)	1482 (19.3)	1362 (19.3)	1642 (23.4)	128 211 (20.9)
Former	3928 (18.1)	1451 (18.9)	1301 (18.5)	1176 (16.8)	117 281 (19.1)
Never	5390 (24.8)	2100 (27.3)	1750 (24.9)	1540 (21.9)	181 118 (29.6)
Unknown	7940 (36.5)	2655 (34.5)	2626 (37.3)	2659 (37.9)	186 036 (30.4)
Aspirin exposure, n (%)	7639 (35.1)	3008 (39.1)	2511 (35.7)	2120 (30.2)	210 518 (34.4)

\*The demographic variables were compared among groups using Kruskal-Wallis test and  $\chi^2$  test, all p values were <0.001. BMI, body mass index; CRC, colorectal cancer; Q1, first quartile; Q3, third quartile.

were significantly higher than rectal or proximal cancer odds ( $p < 0.05$  for case-case comparisons).

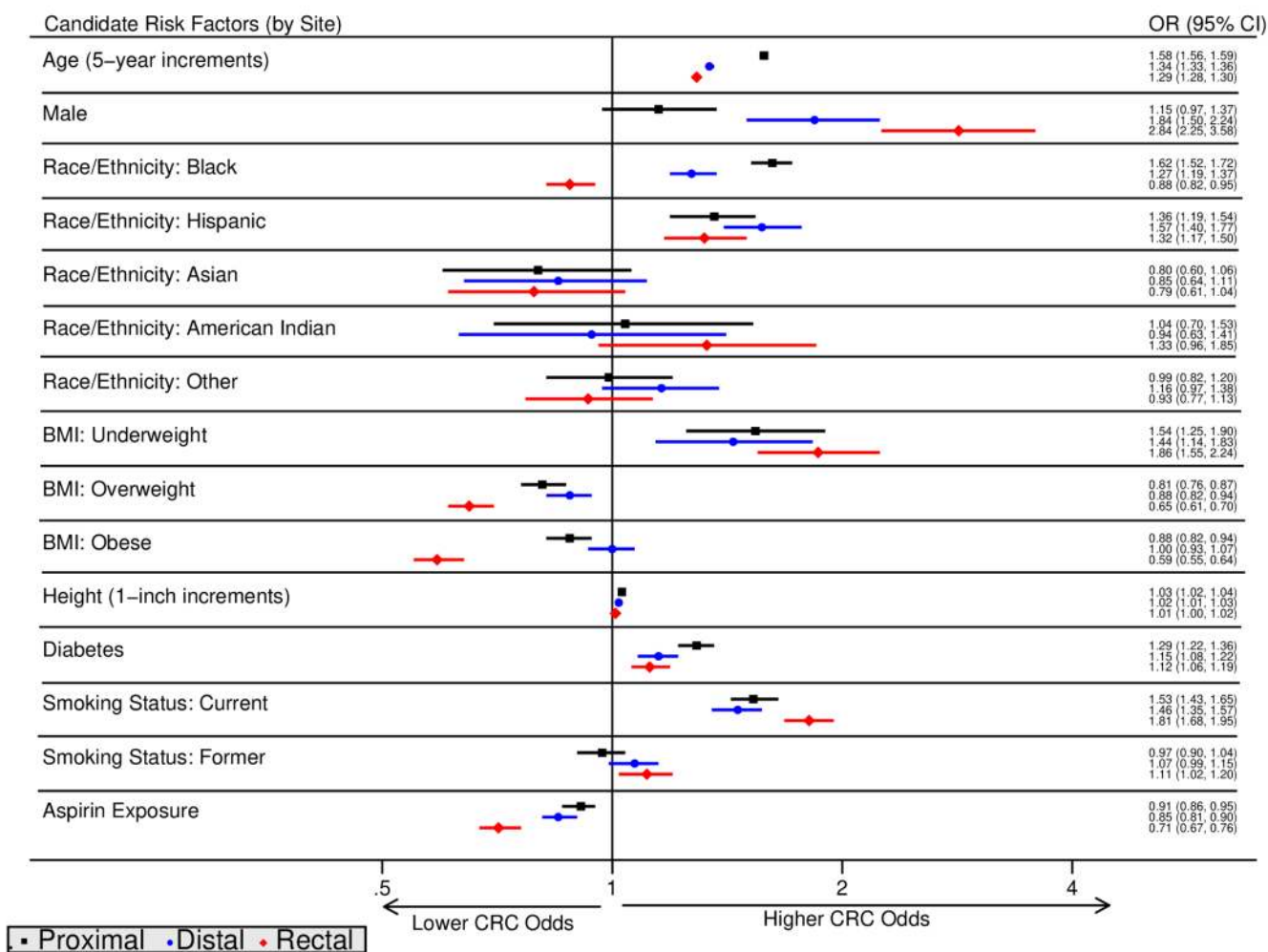
In unadjusted analyses, individuals classified as obese (defined by BMI  $\geq 30.0$  kg/m<sup>2</sup>) had lower odds of cancer at all three sites: (OR=0.62, 95% CI 0.58 to 0.66 for proximal; OR=0.79, 95% CI 0.74 to 0.85 for distal; OR=0.47, 95% CI 0.44 to 0.51 for rectal). In adjusted analyses, obese individuals had significantly reduced odds of rectal cancer (OR=0.59, 95% CI 0.55 to 0.64) significantly lower than that of proximal cancer (OR=0.88, 95% CI 0.82 to 0.94), compared with normal BMI individuals (<25 kg/m<sup>2</sup>;  $p < 0.05$ ).

In both unadjusted and adjusted analyses, increased height was weakly associated with increased odds of CRC. In adjusted analysis, increased height was associated with increased odds of proximal cancer (OR=1.03, 95% CI 1.02 to 1.04), distal cancer (OR=1.02, 95% CI 1.01 to 1.03) and rectal cancer (OR=1.01, 95% CI 1.00 to 1.02).

### Clinical factors

Having diabetes increased odds for proximal cancer (OR=1.40, 95% CI 1.33 to 1.47) and distal cancer (OR=1.25, 95% CI 1.18 to 1.31) in unadjusted analyses. In adjusted analyses, having diabetes increased odds for proximal cancer (OR=1.29, 95% CI 1.22 to 1.36), significantly increased more than odds for distal cancer (OR=1.15, 95% CI 1.08 to 1.22) and rectal cancer (OR=1.12, 95% CI 1.06 to 1.19) based on case-case analyses ( $p < 0.05$  for case-case comparisons).

Current smoking was associated with increased odds of distal cancer (OR=1.10, 95% CI 1.02 to 1.18) and rectal cancer (OR=1.51, 95% CI 1.41 to 1.62) compared with never smokers in unadjusted analyses. The odds of CRC were significantly increased in all three sites for current smokers compared with non-smokers in adjusted analyses. Significantly higher odds of rectal cancer were shown for current smoking (OR=1.81, 95% CI 1.68 to



**Figure 1** Risk factors for colorectal cancer by anatomic site. OR findings from adjusted multinomial logistic regression with corresponding 95% CIs, stratified by anatomic site (proximal, distal, rectal). BMI, body mass index; CRC, colorectal cancer.

1.95) than proximal cancer (OR=1.53, 95% CI 1.43 to 1.65) and distal cancer (OR=1.46, 95% CI 1.35 to 1.57);  $p < 0.05$  for case-case comparisons.

Aspirin exposure was associated with increased odds of proximal cancer (OR=1.23, 95% CI 1.17 to 1.29) and distal cancer (OR=1.06, 95% CI 1.01 to 1.11) but decreased odds of rectal cancer (OR=0.83, 95% CI 0.79 to 0.87) in unadjusted analyses. In adjusted analyses, however, aspirin exposure was associated with reduced odds for cancer at all three subsites but most strongly associated with odds for rectal cancer (OR=0.71, 95% CI 0.67 to 0.76) compared with distal cancer (OR=0.85, 95% CI 0.81 to 0.90) and proximal site cancer (OR=0.91, 95% CI 0.86 to 0.95); two-sided  $p$  values for all case-case comparisons were less than 0.05.

### Sensitivity and post hoc analyses

We performed post hoc analyses to further investigate our findings for BMI, looking at change in BMI from 10 and 5 years prior to baseline measurement (online appendix figure 2). Our findings indicated that from 5 years prior to baseline, mean BMI declined among proximal, distal

and rectal cases, but increased among controls. As such, the lower BMI seen in cases compared with controls may in part be explained by weight loss developing as a result of occult cancer in the period prior to colonoscopic diagnosis.

Our sensitivity analyses to explore the impact of potential immortal time bias introduced by requiring controls to be CRC free for 3 years showed that relaxing criteria to being CRC free for 6 months resulted in qualitatively similar results (data not shown).

### DISCUSSION

Among the 21 744 CRC cases and 612 646 normal colonoscopy controls, we found that candidate risk factor associations for CRC vary markedly by anatomic site. Significant differences in presence and magnitude of site-specific associations were found for a number of traditionally cited risk factors for CRC, including male sex, age, race/ethnicity, BMI, height, diabetes, and smoking. As such, accounting for anatomic site in epidemiological studies of CRC risk may allow for more accurate insights into



CRC pathogenesis and strategies for cancer control. Our findings extend and clarify prior research on traditional risk factors for CRC.

### Age

Consistent with prior work, we found that a 5-year age increase was associated with 1.54-fold increased odds for proximal cancer, highlighting a significant distal to proximal colon cancer shift. This distal to proximal shift has been shown in prior studies, particularly among adults over age 70.<sup>8 18–22</sup> Multiple studies have shown these age-related findings in proximal cancer were particularly strong among women.<sup>18 20 21</sup> Iida *et al* reported that increased risk due to age could be related to the change in production or composition of bile acid, which is found to be associated with colorectal carcinogenesis in the proximal colon. In postmenopausal women specifically, Iida *et al* postulated that decreased oestrogen secretion may lead to increased secondary bile acid production and subsequent increased CRC risk.<sup>18 23</sup>

### Sex

Compared with females, males had increased risk of CRC across all sites, which is in agreement with findings from prior studies.<sup>19 24 25</sup> Notably, CRC risk in males increased with more than two times the odds of distal cancer, and nearly three times the odds of rectal cancer compared with females, a finding also supported by prior research.<sup>26</sup> One potential explanation for this finding, independent of other candidate risk factors in our model, could be due to unmeasured sex-specific lifestyle factors. Poynter *et al* suggested that alcohol consumption could increase rectal cancer risk among males, but not among females.<sup>27</sup> Among females, previous studies suggested the protective nature of hormone replacement therapy or oral contraceptive use, with Gao *et al* noting that the protective effects of hormones could explain the increased difference in rectal cancer risk between males and females.<sup>24 28 29</sup>

### Race/ethnicity

Our study found that risk of proximal colon cancer was increased among non-Hispanic blacks, which aligned with findings from previous studies.<sup>2 30 31</sup> Irby *et al* hypothesised that the higher black-to-white risk ratios for proximal colon cancer could be attributed to a higher baseline diabetes prevalence among blacks as compared with non-Hispanic whites.<sup>30</sup> However, when adjusting for diabetes within our analyses, the higher risk for proximal cancer observed among blacks persisted. Some have hypothesised that blacks may be more likely to have a higher proportion of cancer with microsatellite instability, which is more likely to arise in the proximal than distal colon, but consistency of this observation has not been borne out by other studies.<sup>32 33</sup> As such, reasons for higher rates of proximal CRC among blacks in this study and other studies require further study.

Our study also found that Hispanics had increased risk for distal colon cancer compared with non-Hispanic

whites, which aligned with two prior studies.<sup>34 35</sup> Chat-tar-Cora *et al* found higher rates of distal colon among Hispanics in their study, noting that their findings might be due to more than half of their sample population being under age 65, when distal colon cancer is more likely.<sup>35</sup> While Jafri *et al* had similar findings, they cautioned that there are likely variations in effect within Hispanic subgroups. These potential differences highlight a need for more research about CRC risk, particularly distal colon cancer risk, within these subpopulations.<sup>34</sup>

### Body mass index

Our findings of a significantly protective effect of higher BMI on CRC risk at all subsites except distal were surprising, given that CRC is identified as an obesity-related cancer, and prior studies consistently show obesity to be associated with increased CRC risk, regardless of site.<sup>36–38</sup> However, in our data we observed higher BMI was associated with reduced risk for proximal and distal cancers. We speculate several potential reasons for our discordant observations. The median BMI of all included individuals in our study was consistent with overweight (28.9 for controls and 27.9 for cases), which is higher than most studies of obesity and CRC risk, and could have impacted risk associations. BMI was recorded based on measurements just prior to time of colonoscopy. CRC can cause weight loss, which could affect the temporality of our estimation and make a higher BMI seem protective. In post hoc analyses, we found that BMI decreased markedly among cases within the 5 years prior to baseline, which would align with this hypothesis.

Another potential explanation for this discrepancy is the use of BMI as a surrogate measure for overall body fat. In our study, we calculated BMI using a median weight derived from 3 years of weight measurements, and a single height measure. A previous meta-analysis found that abdominal obesity is a more sensitive indicator of CRC risk than BMI, and that visceral obesity might be the main driving factor of the association between obesity and CRC risk.<sup>36 39</sup> In this study, we were unable to differentiate between abdominal and visceral obesity, though future studies should consider this distinction to better understand the association between obesity and CRC risk. Additionally, we ascertained presence/absence of obesity over a short time frame (3 years prior to index colonoscopy); duration of obesity may be a better measure of obesity-related risk, and lack of association may have been due to inability to study persistent obesity.

### Height

Increased height was associated with a slight increased risk of all CRC types, though the effect was small. These findings align with those from prior systematic reviews showing that height is a risk factor for CRC, notably for proximal and distal cancers.<sup>40 41</sup> A recent multinational cohort study additionally found that increased height was associated with increased proximal and distal cancer risk, but not rectal cancer risk.<sup>42</sup> Potential mechanisms

explaining this relationship include increased exposure to growth factors, such as growth hormone or insulin-like growth factors in childhood or early adulthood and excess calorie consumption in early life.<sup>40 43 44</sup>

### Diabetes

Diabetes prevalence was associated with increased risk of all CRC types, but significantly higher risk of proximal colon cancer, which was consistent with prior research.<sup>45 46</sup> An underlying mechanism that could explain higher risk of proximal colon cancer is the effect of hyperinsulinaemia on the colon.<sup>45</sup> Insulin has mitogenic effects on CRC tissue, and upregulates leptin expression, which has been shown to increase cell proliferation within only the proximal colon.<sup>45 47 48</sup> However, there have been few studies examining the effect of serum or plasma insulin levels on CRC risk to support this theory.<sup>47 49 50</sup> More research is needed to understand the potential mechanisms by which diabetes may increase CRC (particularly proximal cancer) risk.

### Smoking

Current smoking was associated with increased CRC odds across all three sites, particularly proximal and rectal cancer risk, while former smoking was associated with increased risk of rectal cancer, all findings which align with the current literature.<sup>51 52</sup> Botteri *et al* indicated that smoking is associated with CRC cases displaying high microsatellite instability, which tend to arise from the serrated pathway of CRC.<sup>51</sup> The authors believed this might explain why higher smoking-related cancer risk exists in the proximal colon and rectum.<sup>51</sup> Another explanation postulated by Leufkens *et al* was that smoking might be a risk factor for flat CRC adenomas, which are more commonly found in the proximal colon.<sup>53 54</sup>

### Aspirin

Within our study, aspirin exposure was found to be significantly protective against all sites with the strongest protection against rectal cancer compared with other sites. Previous studies showed aspirin use to be protective against CRC risk, regardless of site.<sup>55 56</sup> While it is likely that the benefits of aspirin use outweigh the potential risks, more research needs to be conducted to better understand all potential mechanisms that cause aspirin to have a protective effect, particularly for rectal cancer.

### Implications

We comprehensively examined the association of seven major CRC risk factors with site-specific CRC odds. To date, most previous studies lacked adequate sample size to stratify findings by site. Given the molecular, clinical and pathological differences in CRCs arising from each site, stratifying our findings by site can enable researchers to dig deeper into site-specific mechanisms that could contribute to CRC tumorigenesis or policies that could promote more adequate prevention of certain types of CRC.

Our findings suggest that differences in CRC subsite risk exist by race and ethnicity, even within an equal access public healthcare system. Given that the VHA works to minimise financial barriers and provide quality care to all veterans, findings of racial and ethnic differences indicate that further studies are needed to learn more about factors that could predispose different racial or ethnic groups to site-specific CRC risk. Differences in risk by race/ethnicity despite an equal access system may point to differences in unmeasured biological factors or environmental exposures that may modify risk.

Increased risk of CRC regardless of site among current smokers indicates that more targeted screening efforts could help prevent CRC in these higher risk individuals. Our observation that smoking is associated with CRC risk at all subsites, but appears most closely associated with rectal and proximal cancer risk, suggests that the mechanisms driving risk may differ by anatomic subsites, and suggests a need for further study of the site-specific drivers of CRC risk.

### Strengths and limitations

Several limitations may be considered when interpreting this work. The study population is composed of veterans receiving care within the VHA. As such, the findings may not be representative of the general US population. The sample was predominantly male, reflective of older US veterans. However, we had 380 female cases and 32 277 female controls, so while women were disproportionately represented compared with males, there was still a large absolute number of females included in the study. Data for additional candidate risk factors, such as physical activity, diet, and alcohol use, were not available within the EHR used for this study, precluding exploration of the association of these factors with site-specific CRC risk. Duration/dose of risk factor exposures, particularly for aspirin and diabetes, was not extensively measured, limiting the ability to explore potential causality in detail. We also did not consider combined effects of risk factors in this analysis by testing interaction, which would serve as another important future direction as we think about the joint effect of risk factors, such as smoking status and obesity or smoking status and aspirin exposure. Risk factor data were ascertained within 1 year prior to baseline (index) colonoscopy, which can lead to concerns about bias due to left truncation. Our decision to restrict risk factor collection to within 1 year of baseline was intended to address how EHR data are not measured at the same time, while also ensuring a small enough time window that would not lead to potential concerns about misrepresenting risk factor status at baseline. Thus, we anticipate the potential bias from left truncation to be minimal.

Our study also has several strengths. This study is one of the largest case-control studies to date to measure the association of key risk factors to anatomic site-specific risk for CRC. Cases were ascertained from the VACCR, which uses a rigorous process to collect information on cancer





cases locally, and then validate them centrally. Furthermore, the use of normal colonoscopy controls without CRC or adenomas at baseline ensures greater comparability between cases and controls than previous studies could provide.

## CONCLUSION

Our study findings show that the presence and strength of association of CRC risk factors may differ by anatomic site. Based on our observations, we suggest future studies should focus on better understanding mechanisms for some of these associations, such as that of diabetes and proximal cancer risk, former smoking and rectal cancer risk, and aspirin exposure on site-specific CRC risk. Ultimately, accounting for anatomic site in epidemiological studies of CRC may enable better insights into CRC pathogenesis and potential cancer control strategies. Accordingly, anatomic site of CRC should be a key consideration in future studies of CRC risk.

### Author affiliations

<sup>1</sup>Moore's Cancer Center, University of California San Diego, La Jolla, California, USA

<sup>2</sup>Department of Research, VA San Diego Healthcare System, San Diego, California, USA

<sup>3</sup>Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, United States

<sup>4</sup>Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California, USA

<sup>5</sup>Veterans Affairs San Diego Healthcare System, San Diego, CA, United States

<sup>6</sup>Department of Medicine, Division of Gastroenterology, University of California San Diego, La Jolla, CA, United States

**Contributors** Concept and design: JD, AE, RB, LL, AKB, JM, MEM, SG. Analysis and interpretation of data: JD, AE, RB, LL, SG. Drafting of manuscript: JD, AE, RB, LL, AKB, JM, MEM, SG. Critical revision of the manuscript for important intellectual content: JD, AE, RB, LL, AKB, JM, MEM, SG. Statistical analysis: JD, RB, LL, SG. Obtained funding: SG.

**Funding** This research was supported by the VA Health Services Research and Development (Grant No 5101HX001574-04, PI: SG) and the National Cancer Institute/National Institutes of Health (Grant No 1R37CA222866-01, PI: SG; Grant No 1F32CA239360-01, PI: JD).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87–108.
- Gonzalez EC, Roetzheim RG, Ferrante JM, et al. Predictors of proximal vs. distal colorectal cancers. *Dis Colon Rectum* 2001;44:251–8.
- Testa U, Pelosi E, Castelli G. Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells. *Medical Sciences* 2018;6.
- Carethers JM. One colon lumen but two organs. *Gastroenterology* 2011;141:411–2.
- Buflin JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med* 1990;113:779–88.
- Irrazábal T, Belcheva A, Girardin SE, et al. The multifaceted role of the intestinal microbiota in colon cancer. *Mol Cell* 2014;54:309–20.
- Benedix F, Schmidt U, Mroczkowski P, et al. Colon carcinoma—classification into right and left sided cancer or according to colonic subsite?—Analysis of 29,568 patients. *Eur J Surg Oncol* 2011;37:134–9.
- Papagiorgis P, Oikonomakis I, Karapanagiotou I, et al. The impact of tumor location on the histopathologic expression of colorectal cancer. *J Buon* 2006;11:317–21.
- United States Department of Veterans Affairs. Veteran population, 2018. Available: [https://www.va.gov/vetdata/Veteran\\_Population.asp](https://www.va.gov/vetdata/Veteran_Population.asp) [Accessed 6 Jul 2018].
- McGinnis KA, Brandt CA, Skanderson M, et al. Validating smoking data from the veteran's Affairs health factors dataset, an electronic data source. *Nicotine Tob Res* 2011;13:1233–9.
- Ramanathan D. *VIREC Factbook: corporate data Warehouse (CDW) vital sign 1.1 domain*. Hines, IL, 2018.
- Gupta S, Liu L, Patterson OV, et al. A Framework for Leveraging "Big Data" to Advance Epidemiology and Improve Quality: Design of the VA Colonoscopy Collaborative. *EGEMS* 2018;6.
- Earles A, Liu L, Bustamante R, et al. Structured approach for evaluating strategies for cancer ascertainment using large-scale electronic health record data. *JCO Clinical Cancer Informatics* 2018;(2):1–12.
- Nol PH, Copeland LA, Perrin RA, et al. Vha corporate data Warehouse height and weight data: opportunities and challenges for health services research. *JRRD* 2010;47:739–50.
- Miller DR, Safford MM, Pogach LM. Who has diabetes? best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care* 2004;27 Suppl 2:B10–21.
- Bustamante R, Earles A, Murphy JD, et al. Ascertainment of aspirin exposure using structured and unstructured large-scale electronic health record data. *Med Care* 2019;1.
- Team R Development Core. R: a language and environment for statistical computing, 2018. Available: <http://www.r-project.org/>
- Iida Y, Kawai K, Tsuno NH, et al. Proximal shift of colorectal cancer along with aging. *Clin Colorectal Cancer* 2014;13:213–8.
- Murphy G, Devesa SS, Cross AJ, et al. Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *Int. J. Cancer* 2011;128:1668–75.
- Jess P, Hansen IO, Gamborg M, et al. A nationwide Danish cohort study challenging the categorisation into right-sided and left-sided colon cancer. *BMJ Open* 2013;3:e002608.
- Saltzstein SL, Behling CA. Age and time as factors in the left-to-right shift of the subsite of colorectal adenocarcinoma: a study of 213,383 cases from the California cancer registry. *J Clin Gastroenterol* 2007;41:173–7.
- Mik M, Berut M, Dziki L, et al. Right- and left-sided colon cancer – clinical and pathological differences of the disease entity in one organ. *Aoms* 2017;1:157–62.
- McMichael AJ, Potter JD, Reproduction PJD. Reproduction, endogenous and exogenous sex hormones, and colon cancer: a review and hypothesis. *J Natl Cancer Inst* 1980;65:1201–7.
- Gao R-N, Neutel CI, Wai E. Gender differences in colorectal cancer incidence, mortality, hospitalizations and surgical procedures in Canada. *J Public Health* 2008;30:194–201.
- Nguyen SP, Bent S, Chen Y-H, et al. Gender as a risk factor for advanced neoplasia and colorectal cancer: a systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology* 2009;7:676–81.
- Dubrow R, Johansen C, Skov T, et al. Age-Period-Cohort modelling of large-bowel-cancer incidence by anatomic sub-site and sex in Denmark. *Int J Cancer* 1994;58:324–9.
- Poynter JN, Haile RW, Siegmund KD, et al. Associations between smoking, alcohol consumption, and colorectal cancer, overall and by tumor microsatellite instability status. *Cancer Epidemiology Biomarkers & Prevention* 2009;18:2745–50.
- Newcomb PA, Pocobelli G, Chia V. *Why hormones protect against large bowel cancer: old ideas, new evidence*. New York, NY: Springer, 2008: 259–69.
- Charlton BM, Wu K, Zhang X, et al. Oral contraceptive use and colorectal cancer in the nurses' health study I and II. *Cancer Epidemiology Biomarkers & Prevention* 2015;24:1214–21.

30. Irby Ket *et al.* Emerging and widening colorectal carcinoma disparities between blacks and whites in the United States (1975–2002). *Cancer Epidemiology Biomarkers & Prevention* 2006;15:792–7.
31. Thornton JG, Morris AM, Thornton JD, *et al.* Racial variation in colorectal polyp and tumor location. *J Natl Med Assoc* 2007;99:723–8.
32. Ashktorab H, Smoot DT, Farzanmehr H, *et al.* Clinicopathological features and microsatellite instability (MSI) in colorectal cancers from African Americans. *Int. J. Cancer* 2005;116:914–9.
33. Berera S, Koru-Sengul T, Miao F, *et al.* Colorectal tumors from different racial and ethnic minorities have similar rates of mismatch repair deficiency. *Clinical Gastroenterology and Hepatology* 2016;14:1163–71.
34. Jafri NS, Gould M, El-Serag HB, *et al.* Incidence and survival of colorectal cancer among Hispanics in the United States: a population-based study. *Dig Dis Sci* 2013;58:2052–60.
35. Chatter-Cora D, Onime GD, Coppa GF, *et al.* Anatomic, age, and sex distribution of colorectal cancer in a new York City Hispanic population. *J Natl Med Assoc* 1998;90:19–24.
36. Dai Z, Xu Y-C, Niu L. Obesity and colorectal cancer risk: a meta-analysis of cohort studies. *World J Gastroenterol* 2007;13:4199–206.
37. Schlesinger S, Lieb W, Koch M, *et al.* Body weight gain and risk of colorectal cancer: a systematic review and meta-analysis of observational studies. *Obes Rev* 2015;16:607–19.
38. Lauby-Secretan B, Scoccianti C, Loomis D, *et al.* Body fatness and cancer — viewpoint of the IARC Working group. *N Engl J Med* 2016;375:794–8.
39. Keum N, Lee DH, Kim R, *et al.* Visceral adiposity and colorectal adenomas: dose-response meta-analysis of observational studies. *Ann Oncol* 2015;26:1101–9.
40. Aicr, WCRF. Diet, nutrition, physical activity and colorectal cancer. Available: <https://www.aicr.org/continuous-update-project/reports/colorectal-cancer-2017-report.pdf> [Accessed 26 Jun 2019].
41. Green J, Cairns BJ, Casabonne D, *et al.* Height and cancer incidence in the Million women study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol* 2011;12:785–94.
42. Murphy N, Ward HA, Jenab M, *et al.* Heterogeneity of Colorectal Cancer Risk Factors by Anatomical Subsite in 10 European Countries: A Multinational Cohort Study. *Clinical Gastroenterology and Hepatology* 2019;17:1323–31.
43. Bray I, Gunnell D, Holly JMP, *et al.* Associations of childhood and adulthood height and the components of height with insulin-like growth factor levels in adulthood: a 65-year follow-up of the Boyd Orr cohort. *J Clin Endocrinol Metab* 2006;91:1382–9.
44. Gunnell D, Okasha M, Davey Smith G, *et al.* Height, leg length, and cancer risk: a systematic review. *Epidemiol Rev* 2001;23:313–42.
45. Sun L, Yu S. Diabetes mellitus is an independent risk factor for colorectal cancer. *Dig Dis Sci* 2012;57:1586–97.
46. Peeters PJHL, Bazelier MT, Leufkens HGM, *et al.* The risk of colorectal cancer in patients with type 2 diabetes: associations with treatment stage and obesity. *Diabetes Care* 2015;38:495–502.
47. Limburg PJ, Anderson KE, Johnson TW, *et al.* Diabetes mellitus and subsite-specific colorectal cancer risks in the Iowa women's health study. *Cancer Epidemiol Biomarkers Prev* 2005;14:133–7.
48. Aparicio T, Guilmeau S, Goiot H, *et al.* Leptin reduces the development of the initial precancerous lesions induced by azoxymethane in the rat colonic mucosa. *Gastroenterology* 2004;126:499–510.
49. Schoen RE, Tangen CM, Kuller LH, *et al.* Increased blood glucose and insulin, body size, and incident colorectal cancer. *J Natl Cancer Inst* 1999;91:1147–54.
50. Saydah SH, Platz EA, Rifai N, *et al.* Association of markers of insulin and glucose control with subsequent colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 2003;12:412–8.
51. Botteri E, Iodice S, Bagnardi V, *et al.* Smoking and colorectal cancer. *JAMA* 2008;300.
52. Cheng J, Chen Y, Wang X, *et al.* Meta-Analysis of prospective cohort studies of cigarette smoking and the incidence of colon and rectal cancers. *Eur J Cancer Prev* 2015;24:6–15.
53. Leufkens AM, Van Duijnhoven FJB, Siersema PD, *et al.* Cigarette smoking and colorectal cancer risk in the European prospective investigation into cancer and nutrition study. *Clin Gastroenterol Hepatol* 2011;9:137–44.
54. Anderson JC, Stein B, Kahi CJ, *et al.* Association of smoking and flat adenomas: results from an asymptomatic population screened with a high-definition colonoscope. *Gastrointest Endosc* 2010;71:1234–40.
55. Ruder EH, Laiyemo AO, Graubard BI, *et al.* Non-Steroidal anti-inflammatory drugs and colorectal cancer risk in a large, prospective cohort. *Am J Gastroenterol* 2011;106:1340–50.
56. Rothwell PM, Wilson M, Elwin C-E, *et al.* Long-Term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *The Lancet* 2010;376:1741–50.