OPEN ACCESS

Original research

Young-onset colorectal cancer risk among individuals with iron-deficiency anaemia and haematochezia

Joshua Demb [1], Lin Liu [2], Caitlin C. Murphy [3], Chyke A. Doubeni [4], María Elena Martínez [2], Samir Gupta [1,5]

[1]Department of Medicine, Division of Gastroenterology, University of California San Diego, La Jolla, California, USA
[2]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA
[3]Population and Data Sciences, UT Southwestern Medical Center, Dallas, Texas, USA
[4]Center for Health Equity and Community Engagement; Family Medicine, Mayo Clinic Cancer Center; Center for Clinical and Translational Science, Mayo Clinic, Rochester, Minnesota, USA
[5]Department of Medicine, Section of Gastroenterology, VA San Diego Healthcare System, San Diego, California, USA

Correspondence to
Dr Samir Gupta, Section of Gastroenterology, Department of Medicine, VA San Diego Healthcare System, San Diego, CA 92161, USA; s1gupta@health.ucsd.edu

Received 15 May 2020
Revised 15 September 2020
Accepted 17 September 2020
Published Online First 18 December 2020

Check for updates

To cite: Demb J, Liu L, Murphy CC, *et al. Gut* 2021;**70**:1529–1537.

ABSTRACT

Objective Young-onset colorectal cancer (YCRC) incidence is rising. Scant data exist on YCRC risk after presentation with concerning symptoms such as iron-deficiency anaemia (IDA) or haematochezia. We examined the association between IDA and YCRC, and haematochezia and YCRC.

Design Cohort study of US Veterans aged 18–49 years receiving Veterans Health Administration (VHA) care 1999–2016. IDA analytic cohort was created matching individuals without incident IDA to those with IDA 4:1 based on sex, birth year and first VHA visit date (n=239 000). We used this approach to also create a distinct haematochezia analytic cohort (n=653 740). Incident YCRC was ascertained via linkage to cancer registry and/or cause-specific mortality data. We computed cumulative incidence, risk difference (RD) and HRs using Cox models in each cohort.

Results Five-year YCRC cumulative incidence was 0.45% among individuals with IDA versus 0.05% without IDA (RD: 0.39%, 95% CI: 0.33%–0.46%), corresponding to an HR of 10.81 (95% CI: 8.15–14.33). Comparing IDA versus no IDA, RD was 0.78% for men (95% CI: 0.64%–0.92%) and 0.08% for women (95% CI: 0.03%–0.13%), and RD increased by age from 0.14% for <30 years to 0.53% for 40–49 years. YCRC cumulative incidence was 0.33% among individuals with haematochezia versus 0.03% without haematochezia (RD: 0.30%, 95% CI: 0.26%–0.33%), corresponding to an HR of 10.66 (95% CI: 8.76–12.97). Comparing haematochezia versus no haematochezia, RD increased by age from 0.04% for <30 years to 0.43% for 40–49 years.

Conclusion Colonoscopy should be strongly considered in adults aged <50 years with IDA or haematochezia without a clinically confirmed alternate source.

Significance of this study

What is already known on this subject?

- Young-onset colorectal cancer (YCRC) incidence is rising, but scant data exist on YCRC risk after presentation with concerning symptoms such as iron-deficiency anaemia (IDA) or haematochezia.

What are the new findings?

- Substantially increased YCRC risk was observed after either IDA or haematochezia diagnosis. Notably, risk was higher among men with IDA and among adults aged ≥30 years with IDA or haematochezia. Despite increased risk, there was low uptake of diagnostic colonoscopy among individuals with IDA or haematochezia.

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

- Colonoscopy should be strongly considered in adults aged <50 years with IDA or haematochezia without a clinically confirmed alternate source.

BACKGROUND

Colorectal cancer (CRC) is the second leading cause of cancer death in the USA.[1] Proportion of CRC diagnosed in adults aged <50 years—hereafter called young-onset CRC (YCRC)—has increased over time, with cases often diagnosed at later stages requiring more intense treatment.[2–10] Some have advocated for more aggressive work-up of individuals aged <50 years presenting with purported 'red flag' signs or symptoms for CRC, iron-deficiency anaemia (IDA) and haematochezia, under the postulate this may enhance timely diagnosis and treatment.[4 11 12] However, US and European clinical practice guidelines conflict regarding whether IDA or haematochezia in adults aged <50 years should trigger diagnostic colonoscopy work-up. For IDA, some guidelines recommend colonoscopy for all individuals, regardless of age or sex, while others recommend colonoscopy for those with IDA aged <50 years based on sex, and presence/absence of key clinical characteristics such as lower abdominal symptoms (table 1).[13–15] Similar variation exists across guidelines for work-up of haematochezia, with some recommending colonoscopy as first work-up only for those aged ≥40 years, and others stratifying recommendations based on age, CRC risk factors and bleeding characteristics, such as presence of bright red blood per rectum (table 1).[15 16]

Inconsistency across guidelines reflects insufficient data on YCRC risk among individuals aged <50 years with IDA or haematochezia. Clarifying risk may inform future guidelines, optimise selection of individuals for diagnostic colonoscopy, and ultimately facilitate earlier stage detection and timely YCRC treatment. To address knowledge gaps regarding YCRC risk related to IDA and haematochezia, we conducted a study examining

Table 1 Guidelines for colonoscopy work-up with presentation of iron-deficiency anaemia (IDA) or haematochezia

Symptomatic presentation	Organisation	Recommendations
Iron Deficiency Anaemia	American Gastroenterological Association[25]	Bidirectional endoscopy (esophagogastroduodenoscopy and colonoscopy) in asymptomatic postmenopausal women and all men with IDA. Bidirectional endoscopy in asymptomatic premenopausal women with IDA
	American Society of Gastrointestinal Endoscopy[13]	Colonoscopy regardless of age or sex
	British Society of Gastroenterology[14]	Upper and lower gastrointestinal investigations in postmenopausal women and all men with IDA, unless there is a history of significant non-GI blood loss.
	European Panel on the Appropriateness of Gastrointestinal Endoscopy[15]	Colonoscopy indicated for: Patients ages≥50 Men ages<50 with lower abdominal symptoms (eg, abdominal pain, change in bowel habits). Women without gynaecological symptoms and presenting with lower abdominal symptoms. Men and women ages<50 without lower abdominal symptoms but no known source of bleeding identified.
Hematochezia	American Society of Gastrointestinal Endoscopy[16]	Digital rectal exam and flexible sigmoidoscopy with or without anoscopy prior to colonoscopy among healthy individuals ages≤40 years
	European Panel on the Appropriateness of Gastrointestinal Endoscopy[15]	Colonoscopy indicated for: Adults ages≥50 Adults ages<50 without bright red blood, without source of bleeding identified at sigmoidoscopy or anoscopy, or in the presence of any CRC risk factors such as personal or family history of CRC or inflammatory bowel disease.

CRC, colorectal cancer; IDA, iron deficiency anaemia.

the association between these potential 'red flag' diagnoses and YCRC risk in a large cohort of US Veterans (aged <50 years).

METHODS

Study design, setting and data sources

We conducted a retrospective cohort study among US Veterans aged 18–49 years receiving care within Veterans Health Administration (VHA), one of the largest US healthcare providers.[17] We used a matched cohort design, matching individuals with IDA or haematochezia diagnosis to individuals without a diagnosis (online supplemental appendix figure 1). Matched cohort designs ensure balance of covariate distributions across exposure groups and comparable follow-up between exposed and unexposed individuals using a matched follow-up start date.[18–20] Matching characteristics included birth year, sex and first VHA visit date (±180 days). The 'Matching' package in R, version 3.5.1 was used to conduct matching.[21]

To identify study population data, we used several Department of Veterans Affairs (VA) data resources, including the VA Corporate Data Warehouse (CDW), VHA Vital Status file and National Death Index (NDI). The VA CDW provided discrete data, including demographic characteristics, administrative claims-based diagnosis and procedure codes, prescriptions, anthropometric measures, and free-text data including procedure notes and pathology reports. The VHA Vital Status file was used to ascertain follow-up time through date of last visit, represented as the date and time the last vital record was taken by the healthcare provider.[22] NDI cause-specific mortality data were used to assess vital status and cause of death, and offer the advantage of capturing cause of death within and outside VHA. Person-level linkage between VHA data and the NDI cause-specific mortality data was derived through collaboration between VA and Department of Defense partners, with matching based on social security number (SSN) or VA-scrambled SSN.[23]

IDA analytic cohort

Participants: IDA analytic cohort

The IDA analytic cohort included individuals aged 18–49 years receiving VHA care between 1999 and 2016. All Veterans in the IDA analytic cohort had at least one blood test measuring haemoglobin conducted within the VHA; this blood test date was defined as date of cohort entry. For each individual with IDA diagnosis, we sampled (with replacement) four matched undiagnosed individuals among those alive on the index date (date of IDA diagnosis of exposed individual). Follow-up of each 4:1 unexposed to exposed matched group—hereafter referred to as matched clusters—started on index date and continued until YCRC diagnosis, death from non-CRC causes, turning age 50, 5 years of follow-up or end of study (31 December 2016). We excluded individuals with YCRC or IBD diagnoses prior to start of follow-up. Additionally, we excluded Veterans based on any International Classification of Diseases, Ninth and Tenth Revision (ICD-9, ICD-10) diagnosis codes for IDA prior to the date of cohort entry (date of haemoglobin blood test). Only full 4:1 clusters were included.

IDA exposure variable

IDA was identified by lab diagnosis using the WHO criteria: a haemoglobin test identifying anaemia (haemoglobin <130 g/L in men, < 120 g/L in women) with a follow-up iron test within 3 months indicating iron deficiency (ferritin levels ≤15 ng/mL or transferrin saturation levels ≤16%).[24] Exposed individuals were required to have an iron test to confirm presence of iron deficiency. To account for potential variations in sensitivity and specificity of IDA diagnostic criteria, we conducted a sensitivity analysis where iron deficiency was defined by ferritin levels ≤45 ng/mL (per 2020 American Gastroenterological Association guidelines)[25] and transferrin saturation levels ≤16%.

Haematochezia analytic cohort

Participants: haematochezia analytic cohort

The haematochezia analytic cohort included individuals aged 18–49 years receiving VHA care between 1999 and 2016. For the haematochezia analytic cohort, date of cohort entry was defined by the first Current Procedural Terminology (CPT) code for an office visit initiating care within the VHA (online supplemental appendix table 1). Similar to the IDA analytic cohort, for each individual with haematochezia diagnosis, we sampled (with replacement) four matched undiagnosed individuals among those alive on the index date (date of haematochezia diagnosis of exposed individual). Follow-up of each 4:1 matched cluster started on index date and continued until YCRC diagnosis, death from non-CRC causes, turning age 50, 5 years of follow-up or end of study (31 December 2016). We excluded individuals with YCRC or IBD diagnoses prior to start of follow-up. Additionally, we excluded Veterans based on ICD-9/ICD-10 diagnosis codes for haematochezia prior to the date of cohort entry. Only full 4:1 clusters were included.

Haematochezia exposure variable

Haematochezia was identified by ICD-9 (569.3, 578.1) or ICD-10 (K62.5, K92.1) codes determined by the research team. To account for potential variations in administrative claims codes used to indicate haematochezia, we conducted a sensitivity analysis where the haematochezia exposure included ICD-9 (578.9) and ICD-10 (K92.2) codes corresponding to unspecified GI haemorrhage.

YCRC outcomes

Primary outcome was YCRC within 5 years of start of follow-up, defined by primary and secondary diagnoses identified in VA Central Cancer Registry and Oncology Raw, which can accurately identify 90% of CRC cases,[26] or NDI-identified YCRC. YCRC cases were divided into three anatomical sites based on methodology from prior studies.[27–29] American Joint Committee on Cancer stage was also derived from Oncology Raw. The 5-year time window is based on an a priori assumption that IDA or haematochezia diagnoses would be either resolved or otherwise unrelated to YCRC outcome outside this time period.

Covariates

Covariates were identified through a priori examination of the literature for potential common causes of IDA or haematochezia and YCRC. Covariates included race/ethnicity, body mass index (BMI), diabetes, aspirin use and smoking status (current, former, never). We defined race/ethnicity in six mutually exclusive categories: non-Hispanic White (White); non-Hispanic Black (Black); Hispanic; Asian or Native Hawaiian/Pacific Islander; American Indian or Alaska Native; and other (multiracial and those designating 'other' race) using race and ethnicity data within CDW. BMI and diabetes were characterised based on previously derived algorithms.[30 31] Aspirin exposure was defined as at least two prescriptions or mentions of aspirin in free-text notes up to 1 year prior to start of follow-up, an approach found to have a positive predictive value and negative predictive value of 99.2% and 97.5%, respectively.[32] Smoking status was determined from the VHA Health Factors structured data domain, classifying individuals based on terminology including 'current smoker', 'former smoker' or 'never smoker'.[33]

Statistical analysis

The IDA and haematochezia analytic cohorts were analysed separately. We used univariable analyses to compare Veterans with IDA or haematochezia versus without IDA or haematochezia diagnosis using Wilcoxon rank-sum tests or X^2 tests for continuous and categorical variables, respectively. Five-year cumulative YCRC incidence was derived using Kaplan-Meier estimation to account for censoring.[34] Cumulative incidence was used to calculate risk differences. Number needed to scope (NNS) to identify one YCRC case was estimated by postulating that cumulative incidence over 5 years represented baseline prevalence of CRC, and computing the inverse of YCRC prevalence among exposed individuals.[35] Corresponding 95% CIs for cumulative incidence, risk difference and NNS estimates were derived through bootstrapping with 1000 replications.[36] We used Cox proportional hazard models to estimate YCRC hazard ratios (HRs). Follow-up of each matched cluster started on index date and continued until YCRC diagnosis or first censoring date. We estimated HRs and corresponding 95% CIs using mixed Cox regression models adjusted for race/ethnicity, BMI (categorical), prevalent diabetes, smoking status and aspirin use, and accounting for similar covariate distributions of matched clusters using cluster-specific random intercepts.[37] Missingness in covariates was treated as an additional category to avoid data loss. An additional sensitivity analysis was performed adjusting for diagnosis of change in bowel habit (ICD-9: 787.99; ICD-10: R19.4) and unexplained weight loss (ICD-9: 783.21; ICD-10: R63.4) within ±60 days of follow-up start date in adjusted Cox regression models. Additional sex-stratified and age-stratified analyses were also performed, including analyses stratified by both age and sex. We also considered joint exposure of IDA and haematochezia on YCRC risk and sensitivity analyses excluding persons with joint exposures of IDA and haematochezia.

Additionally, we descriptively examined proportion of individuals receiving colonoscopy after IDA or haematochezia diagnosis. Colonoscopy was ascertained using CPT codes (online supplemental appendix table 2) summarised as proportion receiving colonoscopy (1) within 5 years and (2) within 60 days. Additional sensitivity analyses were performed to examine whether shorter first VHA visit date matching window (±90 days), excluding YCRC cases only ascertained from NDI records, excluding women in the IDA cohort with prior diagnoses of menorrhagia (ICD-9: 627.0, 626.2; ICD-10: N92.0) or prior hysterectomy (ICD-9: V88.01; ICD-10: Z90.71, Z90.710) impacted results qualitatively. Online supplemental appendix table 3 includes a description of algorithm and codes used to derive study variables. Analyses were performed using R, version 3.5.1.[38]

Investigators JD, LL and SG had full access to databases used for this study and used to develop the study population. This research was done without patient involvement. Patients were not invited to comment on the study design, develop patient relevant outcomes, interpret the results, or contribute to the writing or editing of this document for readability or accuracy.

RESULTS

Of 2 934 140 Veterans aged 18–49 years, 2 493 861 were eligible for the IDA analytic cohort, and 2 930 957 Veterans were eligible for the haematochezia analytic cohort. After applying predefined exclusion criteria and matching, there were 239 000 Veterans in the IDA analytic cohort and 653 740 Veterans in the haematochezia analytic cohort (online supplemental appendix figure 2).

Table 2 Sample characteristics overall and by iron-deficiency anaemia (IDA) and haematochezia analytic cohorts

	IDA cohort			Haematochezia cohort		
	Overall N=**239 000**	No IDA N=**191 200**	IDA N=**47 800**	Overall N=**653 740**	No haematochezia N=**522 992**	Haematochezia N=**130 748**
Follow-up time in years, median (Q1-Q3)	3.8 (1.7–5.0)	3.8 (1.8–5.0)	3.6 (1.6–5.0)	5.0 (3.1–5.0)	5.0 (3.1–5.0)	5.0 (3.0–5.0)
Age in years, median (Q1-Q3)	42 (36–46)	42 (36–46)	43 (36–46)	42 (34–46)	42 (34–46)	42 (34–46)
Ages <30	25 430 (10.6%)	20 505 (10.7%)	4925 (10.3%)	98 575 (15.0%)	78 875 (15.1%)	19 700 (15.0%)
Ages 30–39	62 407 (26.1%)	50 152 (26.2%)	12 255 (25.6%)	162 800 (24.9%)	130 294 (24.9%)	32 506 (24.9%)
Ages 40–49	151 163 (63.2%)	120 543 (63.0%)	30 620 (64.1%)	392 365 (60.0%)	313 823 (60.0%)	78 542 (60.1%)
Sex						
Male	112 225 (47.0%)	89 780 (47.0%)	22 445 (47.0%)	571 295 (87.4%)	457 036 (87.4%)	114 259 (87.4%)
Female	126 775 (53.0%)	101 420 (53.0%)	25 355 (53.0%)	82 445 (12.6%)	65 956 (12.6%)	16 489 (12.6%)
Race/ethnicity						
White	116 577 (48.8%)	98 077 (51.3%)	18 500 (38.7%)	349 040 (53.4%)	277 533 (53.1%)	71 507 (54.7%)
Black	72 569 (30.4%)	51 852 (27.1%)	20 717 (43.3%)	153 210 (23.4%)	119 185 (22.8%)	34 025 (26.0%)
Hispanic	16 849 (7.1%)	13 640 (7.1%)	3209 (6.7%)	49 902 (7.6%)	38 880 (7.4%)	11 022 (8.4%)
Asian/Pacific Islander	4435 (1.9%)	3646 (1.9%)	789 (1.7%)	4542 (0.7%)	3565 (0.7%)	977 (0.8%)
American Indian	1880 (0.8%)	1456 (0.8%)	424 (0.9%)	10 348 (1.6%)	8170 (1.6%)	2178 (1.7%)
Multiracial/other	4124 (1.7%)	3365 (1.8%)	759 (1.6%)	12 830 (2.0%)	10 235 (2.0%)	2595 (2.0%)
Missing	22 566 (9.4%)	19 164 (10.0%)	3402 (7.1%)	73 868 (11.3%)	65 424 (12.5%)	8444 (6.5%)
Smoking status						
Never	79 949 (33.5%)	61 011 (31.9%)	18 938 (39.6%)	169 458 (25.9%)	131 643 (25.2%)	37 815 (28.9%)
Former	22 569 (9.4%)	17 700 (9.3%)	4869 (10.2%)	58 781 (9.0%)	44 809 (8.6%)	13 972 (10.7%)
Current	60 234 (25.2%)	49 021 (25.6%)	11 213 (23.5%)	180 728 (27.6%)	138 255 (26.4%)	42 473 (32.5%)
Missing	76 248 (31.9%)	63 468 (33.2%)	12 780 (26.7%)	244 773 (37.4%)	208 285 (39.8%)	36 488 (27.9%)
Prevalent diabetes	18 847 (7.9%)	12 755 (6.7%)	6092 (12.7%)	44 390 (6.8%)	34 071 (6.5%)	10 319 (7.9%)
BMI, median (Q1-Q3)	28.8 (25.0–33.0)	28.7 (25.1–32.9)	28.9 (24.8–33.6)	29.0 (25.7–32.9)	28.9 (25.6–32.7)	29.5 (26.1–33.5)
Underweight	1695 (0.7%)	1052 (0.6%)	643 (1.4%)	2503 (0.4%)	1890 (0.4%)	613 (0.5%)
Normal	41 900 (17.5%)	31 768 (16.6%)	10 132 (21.2%)	88 345 (13.5%)	67 484 (12.9%)	20 861 (16.0%)
Overweight	58 852 (24.6%)	46 021 (24.1%)	12 831 (26.8%)	161 697 (24.7%)	121 257 (23.2%)	40 440 (30.9%)
Obese	73 541 (30.8%)	55 679 (29.1%)	17 862 (37.4%)	189 632 (29.0%)	136 360 (26.1%)	53 272 (40.7%)
Missing	63 012 (26.4%)	56 680 (29.6%)	6332 (13.2%)	211 563 (32.4%)	196 001 (37.5%)	15 562 (11.9%)
Aspirin use	17 605 (7.4%)	11 358 (5.9%)	6247 (13.1%)	42 446 (6.5%)	29 457 (5.6%)	12 989 (9.9%)

BMI, body mass index; Q, quartile.

IDA and YCRC risk

In the IDA analytic cohort, there were 0.8 million person-years of follow-up time and 257 YCRC diagnoses. Median age at index date was 42 (quartile 1–quartile 3 (Q1–Q3): 36–46) with a median 3.8 years of follow-up time (Q1–Q3: 1.7–5.0; table 2). Most were aged 40–49 years (63%), 49% were White and 55% were overweight or obese. There were more Veterans with IDA who were Black (43% vs 27%), obese (37% vs 29%) and aspirin users (13% vs 6%), compared with Veterans without IDA. YCRC anatomic site distribution was 38% proximal, 40% distal, 20% rectal and 2% unknown, with 39% of cancers diagnosed at stage III or stage IV (online supplemental appendix table 4).

Among 47 800 Veterans with IDA, there were 184 YCRCs (cumulative incidence: 0.45%) versus 73 YCRCs in 191 200 Veterans without IDA (cumulative incidence: 0.05%; table 3), corresponding to a risk difference (RD) of 0.39% (95% CI: 0.33%–0.46%). YCRC risk was higher among those with IDA versus without IDA (HR: 10.81, 95% CI: 8.15–14.33). There were 8482 Veterans (17%; 8482/47 800) who received a colonoscopy within 5 years of IDA diagnosis, with 2409 (28%; 2409/8482) receiving colonoscopies within 60 days.

Five-year cumulative incidence among men with IDA versus without IDA was 0.85% (95% CI: 0.72%–1.00%) compared with 0.08% (95% CI: 0.06%–0.11%; RD: 0.78%, 95% CI: 0.64%–0.92%). Five-year cumulative incidence among women with IDA versus without IDA was 0.11% (95% CI: 0.07%–0.17%) compared with 0.03% (95% CI: 0.02%–0.05%; RD: 0.08%; 95% CI: 0.03%–0.13%; table 3). Sensitivity analyses excluding women with prior menorrhagia or hysterectomy yielded similar results. In age-stratified analyses, 5-year cumulative incidence increased with increasing age for those with IDA: 0.14% (95% CI: 0.04%–0.27%) for ages <30 years, 0.20% (95% CI: 0.12%–0.28%) for ages 30–39 years and 0.61% (95% CI: 0.51%−0.72%) for ages 40–49 years (table 3). Age-stratified RDs similarly increased among those with IDA: 0.14% for ages <30 years (95% CI: 0.04%−0.26%), 0.18% for ages 30–39 years (95% CI: 0.10%–0.26%) and 0.53% for ages 40–49 years (95% CI: 0.41%–0.63%). In analyses stratified by age and sex, men aged 40–49 years with IDA had a 5-year cumulative incidence of 1.02% (95% CI: 0.84%–1.19%) compared with 0.10% (95% CI: 0.07%–0.14%) without IDA, yielding an RD of 0.91% (95% CI: 0.74%–1.09%). Age-stratified results among women were qualitatively similar.

Haematochezia diagnosis and YCRC risk

In the haematochezia analytic cohort, there were 2.62 million person-years of follow-up time and 556 YCRC cases. Median age at index date was 42 (Q1–Q3: 34–46) with median follow-up 5 years (Q1–Q3: 3.1–5.0), with 60% ages 40–49 years, 87% men, and 53% White. More Veterans with haematochezia were overweight or obese (72% vs 49%) and aspirin users (10% vs 6%)

Table 3 Absolute risk and Cox proportional hazards models for overall, sex-stratified and age-stratified analyses in IDA analytic cohort

		Cohort characteristics		Absolute estimates			Cox proportional hazards models	
		Baseline at risk (N)	Number of YCRC cases	5-year cumulative incidence % (95% CI)	Risk difference % (95% CI)	NNS (95% CI)	Unadjusted HR (95% CI)	Adjusted HR (95% CI)
Overall	IDA	47 800	184	0.45 (0.38–0.51)	0.39 (0.33–0.46)	259.8 (226.8–301.6)	10.35 (7.89–13.57)	10.81 (8.15–14.33)
	No IDA	191 200	73	0.05 (0.04–0.07)				
Ages <30 years	IDA	4925	7	0.14 (0.04–0.27)	0.14 (0.04–0.26)	703.6 (377.2–2422)	29.22 (3.59–237.47)	147.67 (3.43–6350.88)
	No IDA	20 505	1	0.01 (0.00–0.02)				
Ages 30–39 years	IDA	12 255	21	0.20 (0.12–0.28)	0.18 (0.10–0.26)	583.6 (399.2–943.8)	14.53 (5.86–36.00)	14.00 (5.48–35.76)
	No IDA	50 152	6	0.02 (0.00–0.03)				
Ages 40–49 years	IDA	30 620	156	0.61 (0.51–0.72)	0.53 (0.41–0.63)	196.3 (171.1 –235.9)	9.62 (7.21–12.83)	10.06 (7.47–13.56)
	No IDA	120 543	66	0.09 (0.07–0.11)				
Men	IDA	22 445	160	0.85 (0.72–1.00)	0.78 (0.64–0.92)	140.3 (121.5–165.6)	14.04 (10.17–19.39)	14.00 (10.04–19.54)
	No IDA	89 780	48	0.08 (0.06–0.11)				
Men, aged <30 years	IDA	1235	7	0.58 (0.17–1.04)	0.56 (0.15–1.00)	176.4 (95.2–625)	29.44 (3.62–239.31)	38.76 (4.50–334.13)
	No IDA	5141	1	0.02 (0.00–0.06)				
Men, aged 30–39 years	IDA	3812	13	0.37 (0.28–0.58)	0.34 (0.14–0.55)	293.2 (185.2–625)	14.35 (4.68–44.02)	11.96 (3.76–38.08)
	No IDA	16 201	4	0.03 (0.01–0.07)				
Men, aged 40–49 years	IDA	17 398	140	1.02 (0.84–1.19)	0.91 (0.74–1.09)	124.3 (106.4–149.3)	13.56 (9.64–19.09)	13.95 (9.81–19.85)
	No IDA	68 438	43	0.10 (0.07–0.14)				
Women	IDA	25 355	24	0.11 (0.07–0.17)	0.08 (0.03–0.13)	1056.5 (741–1692)	3.85 (2.20–6.74)	4.24 (2.34–7.69)
	No IDA	101 420	25	0.03 (0.02–0.05)				
Women, aged <30 years	IDA	3690	0					
	No IDA	15 364	0					
Women, aged 30–39 years	IDA	8443	8	0.12 (0.04–0.22)	0.11 (0.03–0.21)	1055.4 (588.2–2500)	16.09 (3.42–75.76)	13.99 (2.76–70.85)
	No IDA	33 951	2	0.01 (0.00–0.02)				
Women, aged 40–49 years	IDA	13 222	16	0.14 (0.07–0.22)	0.07 (0.00–0.16)	826.4 (555.6–1428.6)	2.74 (1.45–5.19)	3.23 (1.65–6.31)
	No IDA	52 105	23	0.07 (0.04–0.10)				

Risk difference corresponds to difference between exposed and unexposed cumulative incidence results. Number needed to scope (NNS) is the inverse of the CRC prevalence among exposed individuals.
Absolute estimates derived from 5-year cumulative incidence curve models accounting for censoring with bootstrapped 95% CIs.
Unadjusted model includes matching strata variable as random intercept; adjusted model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use.
Empty cells reflect having zero cases, limiting the ability to conduct stratified analyses.
BMI, body mass index; CRC, colorectal cancer; IDA, iron-deficiency anaemia; YCRC, young-onset colorectal cancer.

compared with Veterans without haematochezia (table 2). YCRC anatomic site distribution was 15% proximal, 45% distal, 38% rectal and 1% unknown, with 40% of cancers diagnosed at stage III or stage IV (online supplemental appendix table 4).

Among 130 748 Veterans with haematochezia, there were 406 YCRCs (5-year cumulative incidence: 0.33%, 95% CI: 0.30%–0.36%) compared with 150 YCRCs in 522 992 Veterans without haematochezia (5-year cumulative incidence: 0.03%, 95% CI: 0.03%–0.04%), corresponding to an RD of 0.30% (95% CI: 0.26%–0.33%; table 4). YCRC risk among Veterans with haematochezia was 10.66-fold higher (adjusted HR: 10.66, 95% CI: 8.76–12.97). Among those with haematochezia, 59

Table 4 Absolute risk and Cox proportional hazards models for overall, sex-stratified and age-stratified analyses in haematochezia analytic cohort

		Cohort characteristics		Absolute estimates			Cox proportional hazards models	
		Baseline at risk (N)	Number of YCRC cases	5-year cumulative incidence % (95% CI)	Risk difference % (95% CI)	NNS (95% CI)	Unadjusted HR (95% CI)	Adjusted HR (95% CI)
Overall	Haematochezia	130 748	406	0.33 (0.30–0.36)	0.30 (0.26–0.33)	322 (293.4–359.1)	10.88 (9.02–13.12)	10.66 (8.76–12.97)
	No haematochezia	522 992	150	0.03 (0.03–0.04)				
Age <30 years	Haematochezia	19 700	9	0.05 (0.02–0.08)	0.04 (0.01–0.08)	2188.9 (1312–4985)	12.02 (3.25–44.41)	16.47 (4.09–66.29)
	No haematochezia	78 875	3	0.003 (0.00–0.01)				
Ages 30–39 years	Haematochezia	32 506	51	0.16 (0.12–0.20)	0.15 (0.11–0.19)	637.4 (497.4–865.6)	17.09 (9.11–32.05)	17.21 (8.88–33.35)
	No haematochezia	130 294	12	0.01 (0.00–0.01)				
Ages 40–49 years	Haematochezia	78 542	346	0.50 (0.44–0.55)	0.43 (0.38–0.49)	227 (206.3–253.3)	10.32 (8.45–12.59)	10.12 (8.22–12.45)
	No haematochezia	313 823	135	0.06 (0.05–0.07)				
Men	Haematochezia	114 259	371	0.35 (0.31–0.38)	0.31 (0.27–0.35)	308 (280.3–342.9)	10.43 (8.60–12.65)	10.16 (8.30–12.44)
	No haematochezia	457 036	143	0.04 (0.03–0.04)				
Men, aged <30 years	Haematochezia	16 367	6	0.04 (0.01–0.07)	0.03 (0.01–0.06)	2727.8 (1428.6–10 000)	8.02 (2.01–32.07)	10.24 (2.31–45.30)
	No haematochezia	65 571	3	0.004 (0.00–0.01)				
Men, aged 30–39 years	Haematochezia	27 761	40	0.14 (0.10–0.19)	0.14 (0.10–0.18)	694 (526.3–1000)	16.07 (8.04–32.13)	15.68 (7.56–32.52)
	No haematochezia	111 167	10	0.01 (0.00–0.02)				
Men, aged 40–49 years	Haematochezia	70 131	325	0.52 (0.47–0.58)	0.46 (0.39–0.52)	215.8 (196.1–243.9)	10.07 (8.22–12.34)	9.85 (7.97–12.18)
	No haematochezia	280 298	130	0.07 (0.06–0.08)				
Women	Haematochezia	16 489	35	0.22 (0.15–0.30)	0.21 (0.14–0.29)	471.1 (350.2–686.1)	20.05 (8.91–45.13)	20.60 (8.84–48.00)
	No haematochezia	65 956	7	0.01 (0.00–0.02)				
Women, aged <30 years	Haematochezia	3333	3	0.09 (0.03–100)	0.09 (0.00–0.21)	1111 (476.2, Und)		
	No haematochezia	13 304	0	0.00 (0–100)				
Women, aged 30–39 years	Haematochezia	4745	11	0.23 (0.11–0.39)	0.22 (0.09–0.38)	431.4 (263.2–909.1)	22.22 (4.93–100.26)	25.35 (5.30–121.30)
	No haematochezia	19 127	2	0.01 (0.00–0.03)				
Women, aged 40–49 years	Haematochezia	8411	21	0.27 (0.16–0.39)	0.25 (0.14–0.37)	400.5 (263.2–666.7)	16.79 (6.33–44.53)	15.99 (5.81–44.05)
	No haematochezia	33 525	5	0.02 (0.00–0.04)				

Risk difference corresponds to difference between exposed and unexposed cumulative incidence results. Number needed to scope (NNS) is the inverse of the CRC prevalence among exposed individuals.
Absolute estimates derived from 5-year cumulative incidence curve models accounting for censoring with bootstrapped 95% CIs.
Unadjusted model includes matching strata variable as random intercept; adjusted model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use.
Empty cells reflect having zero cases, limiting the ability to conduct stratified analyses.
.BMI, body mass index; CRC, colorectal cancer; Und, undefined; YCRC, young-onset colorectal cancer.

936 (46%) received a colonoscopy within 5 years follow-up, with 59% (35 298/59 936) of colonoscopies within 60 days of haematochezia diagnosis.

Five-year cumulative incidence among men with haematochezia versus without haematochezia was 0.35% (95% CI: 0.31%–0.38%) compared with 0.04% (95% CI: 0.03%–0.04%; RD: 0.31%; 95% CI: 0.27%–0.35%; table 4). Five-year cumulative incidence among women with haematochezia versus without haematochezia was 0.22% (95% CI: 0.15%–0.30%) compared with 0.01% (95% CI: 0.00%–0.02; RD:

Demb J, *et al*. *Gut* 2021;**70**:1529–1537. doi:10.1136/gutjnl-2020-321849

0.21%; 95% CI: 0.14%–0.29%). Five-year cumulative incidence among Veterans with haematochezia increased with increasing age: 0.05% (95% CI: 0.02%–0.08%) for age <30 years, 0.16% (95% CI: 0.12%–0.20%) for ages 30–39 years and 0.50% (95% CI: 0.44%−0.55%) for ages 40–49 years. Age-stratified RDs similarly increased with increasing age: 0.04% for ages <30 years (95% CI: 0.01%−0.08%), 0.15% for ages 30–39 years (95% CI: 0.11%–0.19%) and 0.43% for ages 40–49 years (95% CI: 0.38%–0.49%).

In analyses stratified by age and sex, men aged 40–49 years with haematochezia had a 5-year cumulative incidence of 0.52% (95% CI: 0.47%–0.58%) compared with 0.07% (95% CI: 0.06%–0.08%) without haematochezia, yielding an RD of 0.46% (95% CI: 0.39%–0.52%). Among women, age-stratified RDs increased with increasing age: 0.22% for ages 30–39 years (95% CI: 0.09%–0.38%) and 0.25% for ages 40–49 years (95% CI: 0.14%–0.37%).

NNS to detect one YCRC

NNS was 259.8 (95% CI: 226.8–301.6) in the IDA analytic cohort and 322 (95% CI: 293.4–359.1) in the haematochezia analytic cohort. NNS was 140.3 (95% CI: 121.5–165.6) for men and 1056.5 (95% CI: 741–1692) for women in the IDA analytic cohort and 308 (95% CI: 280.3–342.9) for men and 471.1 (95% CI: 350.2–686.1) for women in the haematochezia analytic cohort. NNS decreased with increasing age. In the IDA analytic cohort, NNS was 703.6 (95% CI: 377.2–2422) for ages <30 years, 583.6 (95% CI: 399.2–943.8) for ages 30–39 years and 196.3 (95% CI: 171.1–235.9) for ages 40–49 years. The NNS was lowest among men aged 40–49 years at 124.3 (95% CI: 106.4 –149.3).

In the haematochezia analytic cohort, NNS was 2188.9 (95% CI: 1312–4985) for ages <30 years, 637.4 (95% CI: 497.4–865.6) for ages 30–39 years and 227 (95% CI: 206.3–253.3) for ages 40–49 years. Among men, NNS decreased with increasing age: 2727.8 for ages <30 years (95% CI: 1428.6–10 000), 694 for ages 30–39 years (95% CI: 526.3–1000) and 215.8 for ages 40–49 years (95% CI: 196.1–243.9). Among women, NNS was lowest among those aged 40–49 years (NNS: 400.5, 95% CI: 263.2–666.7).

Additional analyses

Among 1320 Veterans with concurrent IDA and haematochezia, there were 31 YCRCs (5-year cumulative incidence: 2.50%, 95% CI: 1.65%–3.36%), compared with 226 YCRCs among 237 680 Veterans with only either or neither IDA or haematochezia diagnosis (5-year cumulative incidence: 0.12%, 95% CI: 0.10%–0.14%), corresponding to an RD of 2.39% (online supplemental appendix table 5).

In a sensitivity analysis excluding cases ascertained only from NDI records, 5-year cumulative incidence estimates slightly decreased, but were similar to the main analyses (online supplemental appendix tables 6 and 7). In sensitivity analysis modifying exposure definitions of IDA to include individuals with ferritin levels ≤45 ng/mL and haematochezia to include those with unspecified GI haemorrhage, the findings were similar to those of the primary analyses (online supplemental appendix table 8). In sensitivity analyses excluding individuals with joint exposures to IDA and haematochezia from each analytic cohort to examine independent effects of each exposure, the findings remained robust (online supplemental appendix table 9). In sensitivity analyses additionally adjusting for change in bowel habit and abnormal weight loss in Cox models, there was no meaningful difference in the effect of IDA or haematochezia on YCRC risk (online supplemental appendix table 10).

DISCUSSION

In two distinct analytic cohorts derived from a sample of 2.9 million Veterans aged 18–49 years, we found both IDA and haematochezia diagnosis were associated with 10-fold increased YCRC risk. YCRC risk was particularly elevated for men with IDA or haematochezia, and risk increased with increasing age in both cohorts. Given current evidence gaps in YCRC risk and burden, our findings could inform clinical guidelines and improve timely YCRC detection and treatment.

NNS was 140.3 for men and 1056.5 for women in the IDA analytic cohort, and 308 for men and 471.1 for women in the haematochezia analytic cohort. Prior work estimated the NNS to detect CRC among asymptomatic individuals undergoing screening colonoscopy is 333 among adults aged 50–75 years.[39] The estimated NNS results for men in both cohorts were lower than these thresholds, suggesting these groups should be strongly considered for colonoscopy to rule out CRC. Notably, when adjusting for both age and sex, men aged 40–49 years had the lowest NNS at 124.3 in the IDA analytic cohort and 215.8 in the haematochezia analytic cohort. While NNS for women in the IDA and haematochezia analytic cohorts were markedly higher (1056.5 and 471.1, respectively), women aged 40–49 years in the IDA analytic cohort (NNS: 826.4) and the haematochezia analytic cohort (NNS: 400.5) had lower NNS values. In age-stratified analyses, risk difference increased by age, such that Veterans aged 40–49 years in both IDA (NNS: 196.3) and haematochezia (NNS: 227) cohorts had NNS below the 333 threshold.

YCRC risk among individuals exposed versus unexposed to IDA has not been widely studied. Hung *et al* found a positive association between IDA diagnosis and CRC, aligning with our findings.[40] Perhaps because of the paucity of prior evidence on YCRC risk associated with IDA versus without IDA, practice guidelines regarding work-up including colonoscopy vary widely (table 1). While our study findings strengthen evidence to support guidelines recommending colonoscopy for men aged <50 years with IDA, they also raise questions concerning recommendations by some to restrict colonoscopy to postmenopausal women aged <50 years rather than all younger women. Women in our IDA analytic cohort had markedly lower YCRC risk compared with men but still had increased risk among those with IDA, even after excluding those with prior menorrhagia or hysterectomy. As such, YCRC risk findings among women aged <50 years with IDA diagnosis suggest harms and benefits of colonoscopy uptake should be carefully considered and studied further. The findings also clarify YCRC risk by age of IDA diagnosis, as the risk difference increases sharply with age, justifying discussion about age-specific colonoscopy referral to rule out YCRC.

YCRC risk among individuals with haematochezia versus without haematochezia has also not been widely studied. Prior studies retrospectively assessed symptoms in patients with YCRC at time of diagnosis, finding between 37% and 59% of symptomatic individuals had rectal bleeding at diagnosis; these symptoms sometimes led to delays in diagnosis because they were attributed to haemorrhoids.[41–45] Our findings suggest clinicians should recommend patients aged <50 years with haematochezia for complete diagnostic colonoscopy work-up, particularly those between ages 40 and 49 years, or if symptoms persist. If these results are replicated by other population-based studies, practice guidelines may need to be updated to recommend colonoscopy as the primary test to evaluate haematochezia.

Our findings highlight some of the decision-making challenges clinicians face in determining the threshold for colonoscopy follow-up for individuals with IDA or haematochezia but provide a range of data for decision-making. While there may not yet be consensus on whether to base decision-making on relative risk increases, absolute risk increases or cumulative 5-year YCRC incidence, our data provide all of these parameters. As such, clinicians and guideline makers may use these data based on the parameters they feel are most important for clinical decisions, including for shared decision-making with patients.

Despite observed increased YCRC risk after IDA or haematochezia diagnosis, few patients received follow-up colonoscopy. Only 17% with IDA diagnosis and 46% with haematochezia diagnosis received colonoscopy, despite being at a higher YCRC risk than individuals without these clinical findings. No published studies, to our knowledge, have measured colonoscopy uptake among adults aged <50 years with IDA or haematochezia. The disconnect between increased risk and low uptake may herald an opportunity for health services interventions to promote colonoscopy uptake, particularly for individuals at highest risk for YCRC, such as men with IDA, and individuals with IDA or haematochezia aged 40–49 years.

Several limitations may be considered in interpreting our results. First, findings of association between IDA or haematochezia diagnosis and YCRC risk are not causal, but instead identify potential YCRC warning signs, and whether colonoscopy is indicated in these scenarios. As Veterans may receive healthcare services outside of VHA, the results potentially underestimate burden of IDA and haematochezia in the Veteran population and resultant colonoscopy uptake. Despite the sizeable proportion of YCRC cases that arise as a result of a family history of CRC,[46] we were unable to account for family history of CRC due to inadequate documentation within the data source. We relied on commonly used diagnostic codes and prior laboratory criteria to inform ascertainment of haematochezia and IDA exposures, respectively. While we did not validate our IDA or haematochezia definitions, which could impact measurement precision, we conducted multiple sensitivity analyses using different exposure definitions, none of which meaningfully impacted our primary results. We also could not distinguish severity of haematochezia among patients; cancer risks might be markedly different among individuals with minor bleeding (blood on toilet paper) versus more obvious, persistent blood in stool. VHA disproportionately cares for a higher number of men versus women, which can over-represent an effect more prevalent among men. However, our IDA and haematochezia study cohorts included significant absolute numbers of women compared with other VHA-based studies.

Our study also has several strengths. To our knowledge, it is one of the largest studies to examine association between clinically suspicious CRC signs and YCRC risk. We identified IDA diagnosis using lab reports, which is a more robust methodology than relying on diagnosis codes. Cohort matching enabled adjustment for potential factors that might induce bias, specifically birth year, sex, first VHA visit and time of follow-up initiation. Further, sensitivity analyses to account for possible variations in clinical care and sampling methodology helped ensure robustness of our findings.

CONCLUSION

We found YCRC risk is elevated among Veterans aged 18–49 years after IDA or haematochezia diagnosis. Among Veterans with IDA, risk is highest among men and individuals aged ≥30 years. Among Veterans with haematochezia, YCRC risk is similar among men and women and highest among individuals aged ≥30 years. Despite increased observed risk, colonoscopy uptake after either IDA or haematochezia diagnosis was low. Our results offer the opportunity to inform clinical decision-making and practice guidelines that may facilitate earlier detection and treatment of the rising number of adults aged <50 years at risk for incident and fatal YCRC.

Twitter Joshua Demb @dembj13 and Samir Gupta @samirguptaGI

Acknowledgements We acknowledge the Joint Veterans Administration/Department of Defense Mortality Data Repository and the Center for Disease Prevention Control for providing access to National Data Index data.

Contributors JD, LL, CM, CAD, MEM and SG—concept and design. JD, LL, CM, CAD, MEM and SG—analysis and Interpretation of data. JD and SG—drafting of the manuscript. JD, LL, CM, CAD, MEM and SG—critical revision of the manuscript for Important intellectual content. JD, LL, CM, CAD, MEM and SG—statistical analysis.

Funding This research was supported by Grant #: 5F32CA239360-02 (PI: JD) from National Cancer Institute/National Institutes of Health; Grant # 5I01HX001574-05 (PI: SG) from VA Health Services Research and Development; and Grant #: 5R37CA222866-02 (PI: SG) from National Cancer Institute/National Institutes of Health.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the VA San Diego Healthcare System: H130441 and the University of California San Diego Institutional Review Boards: 191787.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Joshua Demb http://orcid.org/0000-0002-3338-1730
Caitlin C. Murphy http://orcid.org/0000-0001-9365-0691
Samir Gupta http://orcid.org/0000-0003-4192-5002

REFERENCES

1 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;68:7–30.
2 Siegel RL, Miller KD, Fedewa SA, *et al*. Colorectal Cancer Facts and Figures 2017-2019. *Am Cancer Soc*, 2017. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf [Accessed 13 Jun 2018].
3 SEER*Explorer application: colon and rectum long-term trends in seer incidence rates, 1975-2015. National cancer Institute: surveillance, epidemiology, and end results program, 2018. Available: https://seer.cancer.gov/explorer/application.php?site=20&data_type=1&graph_type=1&compareBy=sex&chk_sex_1=1&chk_race_1=1&chk_race_3=3&chk_race_2=2&chk_age_range_9=9&advopt_precision=1&advopt_display=1&showDataFor=race_1_and_age_range_9. Published [Accessed 3 Jul 2018].
4 Doubeni CA. Early-onset colorectal cancer: What reported statistics can and cannot tell us and their implications. *Cancer* 2019;125:3706–8.
5 Chen FW, Sundaram V, Chew TA, *et al*. Advanced-Stage colorectal cancer in persons younger than 50 years not associated with longer duration of symptoms or time to diagnosis. *Clin Gastroenterol Hepatol* 2017;15:728–37.

6 Siegel RL, Miller KD, Fedewa SA, *et al*. Colorectal cancer statistics, 2017. *CA Cancer J Clin* 2017;67:177–93.
7 Smith RA, Andrews KS, Brooks D, *et al*. Cancer screening in the United States, 2018: a review of current American cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2018;68:297–316.
8 Siegel RL, Fedewa SA, Anderson WF, *et al*. Colorectal cancer incidence patterns in the United States, 1974–2013. *JNCI J Natl Cancer Inst* 2017;109:djw322.
9 Patel SG, Ahnen DJ. Colorectal cancer in the young. *Curr Gastroenterol Rep* 2018;20:15.
10 You YN, Xing Y, Feig BW, *et al*. Young-Onset colorectal cancer: is it time to pay attention? *Arch Intern Med* 2012;172:287–9.
11 Peterse EFP, Meester RGS, Siegel RL, *et al*. The impact of the rising colorectal cancer incidence in young adults on the optimal age to start screening: Microsimulation analysis I to inform the American cancer Society colorectal cancer screening guideline. *Cancer* 2018;124:2964–73.
12 Ahnen DJ, Wade SW, Jones WF, *et al*. The increasing incidence of young-onset colorectal cancer: a call to action. *Mayo Clin Proc* 2014;89:216–24.
13 ASGE Standards of Practice Committee, Early DS, Ben-Menachem T, *et al*. Appropriate use of Gi endoscopy. *Gastrointest Endosc* 2012;75:1127–31.
14 Goddard AF, James MW, McIntyre AS, *et al*. Guidelines for the management of iron deficiency anaemia. *Gut* 2011;60:1309–16.
15 Peytremann-Bridevaux I, Arditi C, Froehlich F, *et al*. Appropriateness of colonoscopy in Europe (EPAGE II). iron-deficiency anemia and hematochezia. *Endoscopy* 2009;41:227–33.
16 ASGE Standards of Practice Committee, Pasha SF, Shergill A, *et al*. The role of endoscopy in the patient with lower Gi bleeding. *Gastrointest Endosc* 2014;79:875–85.
17 United States department of Veterans Affairs. veteran population. National center for veterans analysis and statistics. Available: https://www.va.gov/vetdata/Veteran_Population.asp [Accessed 6 Jul 2018].
18 Greenland S, Morgenstern H. Matching and efficiency in cohort studies. *Am J Epidemiol* 1990;131:151–9.
19 Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *Int J Epidemiol* 2013;42:860–9.
20 Sjölander A, Greenland S. Ignoring the matching variables in cohort studies - when is it valid and why? *Stat Med* 2013;32:4696–708.
21 Jasjeet SS. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J Stat Softw* 2011;42:1–52.
22 Ramanathan D. *VIReC Factbook: corporate data Warehouse (CDW) vital sign 1.1 domain*. Hines, IL, 2018.
23 Prevention C of E for S. Joint department of Veterans Affairs (Va) and department of defense (DOD) suicide data Repository – national death index (NDI), 2020. Available: http://www.dspo.mil/Portals/113/Documents/SDR Fact Sheet.pdf
24 Iron Deficiency Anaemia - Assessment, Prevention, and Control. *A guide for programme managers*. Geneva, Switzerland, 2001.
25 CW K, Siddique S, Patel A, *et al*. American gastroenterological association Institute guideline on the gastrointestinal evaluation of iron deficiency anemia. *Am Gastroenterol Assoc* 2020.
26 Zullig LL, Jackson GL, Dorn RA, *et al*. Cancer incidence among patients of the U.S. Veterans Affairs health care system. *Mil Med* 2012;177:693–701.
27 Gonzalez EC, Roetzheim RG, Ferrante JM, *et al*. Predictors of proximal vs. distal colorectal cancers. *Dis Colon Rectum* 2001;44:251–8.
28 Testa U, Pelosi E, Castelli G. Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells. *Med Sci* 2018;6. doi:10.3390/medsci6020031. [Epub ahead of print: 13 Apr 2018].
29 Demb J, Earles A, Martínez ME, *et al*. Risk factors for colorectal cancer significantly vary by anatomic site. *BMJ Open Gastroenterol* 2019;6:e000313.
30 Noël PH, Copeland LA, Perrin RA, *et al*. Vha corporate data Warehouse height and weight data: opportunities and challenges for health services research. *J Rehabil Res Dev* 2010;47:739–50 http://www.ncbi.nlm.nih.gov/pubmed/21141302
31 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.
32 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;57:e60–4.
33 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.
34 Aalen O. Undefined. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. JSTOR, 1978. Available: https://www.jstor.org/stable/4615704 [Accessed 28 Oct 2019].
35 Ferlitsch M, Reinhart K, Pramhas S, *et al*. Sex-Specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. *JAMA* 2011;306:1352–8.
36 Zhang Z, Ambrogi F, Bokov AF, *et al*. Estimate risk difference and number needed to treat in survival analysis. *Ann Transl Med* 2018;6:120.
37 Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. J. Wiley, 2002. https://books.google.com/books?hl=en&lr=&id=BR4Kq-a1MIMC&oi=fnd&pg=PR7&dq=kalbfleish+prentice+2002&ots=xDlgbGRV9W&sig=CUdXCdrKj97ieJ73fbOMttVdwZc#v=onepage&q=kalbfleish prentice 2002&f=false
38 Team R Development Core. R: a language and environment for statistical computing, 2018. Available: http://www.r-project.org/
39 Heitman SJ, Ronksley PE, Hilsden RJ, *et al*. Prevalence of adenomas and colorectal cancer in average risk individuals: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2009;7:1272–8.
40 Hung N, Shen C-C, Hu Y-W, *et al*. Risk of cancer in patients with iron deficiency anemia: a nationwide population-based study. *PLoS One* 2015;10:e0119647.
41 Riaz R, Masood N, Benish A. Red flag symptoms: detailed account of clinicopathological features in young-onset colorectal cancer. *Intest Res* 2017;15:203.
42 Dozois EJ, Boardman LA, Suwanthanma W, *et al*. Young-Onset colorectal cancer in patients with no known genetic predisposition. *Medicine* 2008;87:259–63.
43 Myers EA, Feingold DL, Forde KA, *et al*. Colorectal cancer in patients under 50 years of age: a retrospective analysis of two institutions' experience. *World J Gastroenterol* 2013;19:5651–7.
44 Olivo R, Ratnayake S. Colorectal cancer in young patients: a retrospective cohort study in a single institution. *ANZ J Surg* 2019;89:905–7.
45 Cha J-M, Kozarek RA, La Selva D, *et al*. Findings of diagnostic colonoscopy in young adults versus findings of screening colonoscopy in patients aged 50 to 54 years: a comparative study stratified by symptom category. *Gastrointest Endosc* 2015;82:138–45.
46 Syed AR, Thakkar P, Horne ZD, *et al*. Old *vs* new: Risk factors predicting early onset colorectal cancer. *World J Gastrointest Oncol* 2019;11:1011–20.