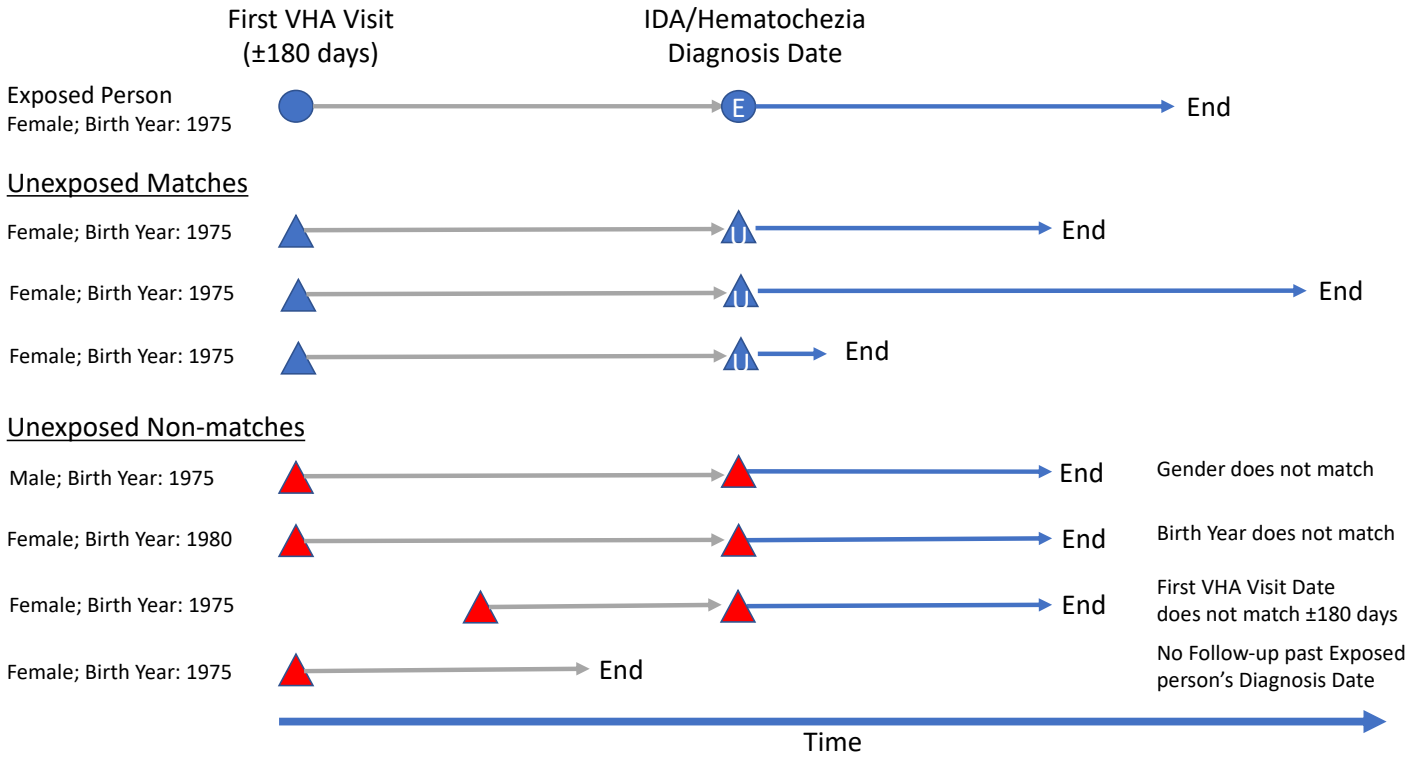
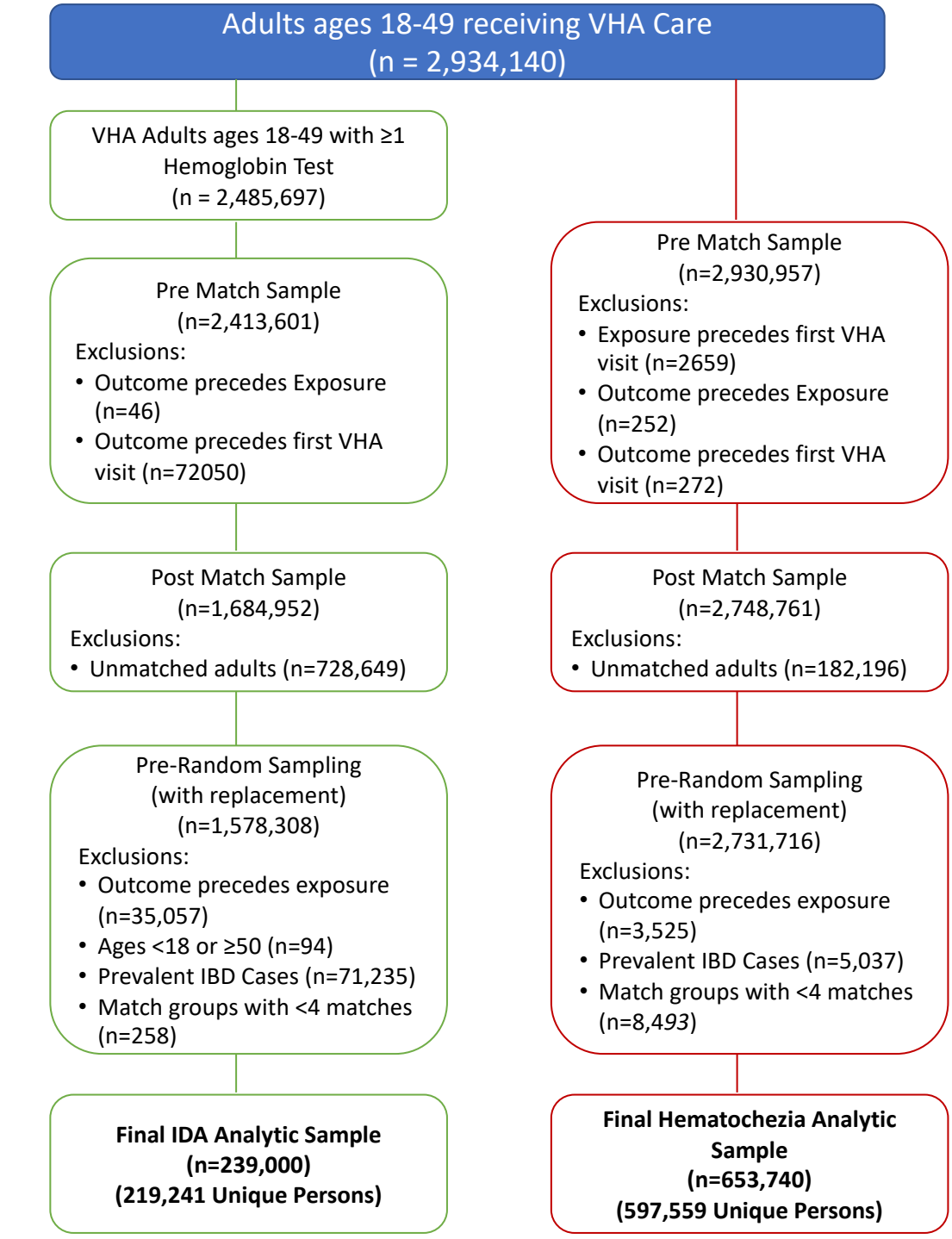
APPENDIX

Appendix Figure 1. Graphical description of matching methodology.



Matching of an exposed individual (based on gender, year of birth, first VHA visit date, and follow-up after exposed person IDA/Hematochezia diagnosis date) to unexposed individuals. Shown are valid matches (blue symbols) and examples of unexposed individuals considered "non-matches" (shown in red). "End" signifies colorectal cancer, non-CRC death, turning 50, loss to follow-up or censoring at 5 years.

Appendix Figure 2. Flowchart of VA participants eligible for cohort



Appendix Table 1. CPT Procedure Codes Used to Identify First Visit

Code	Definition
99201	Office or other outpatient visit for the evaluation and management of a new patient; problem-focused history/examination; straightforward medical decision making
99202	Office or other outpatient visit for the evaluation and management of a new patient; expanded problem-focused history/examination; straightforward medical decision making
99203	Office or other outpatient visit for the evaluation and management of a new patient; detailed problem-focused history/examination; low complexity medical decision making
99204	Office or other outpatient visit for the evaluation and management of a new patient; comprehensive problem-focused history/examination; moderate complexity medical decision making
99205	Office or other outpatient visit for the evaluation and management of a new patient; comprehensive problem-focused history/examination; high complexity medical decision making
99211	Office or other outpatient visit for the evaluation and management of an established patient that may not require the presence of a physician.
99212	Office or other outpatient visit for the evaluation and management of established patient; problem-focused history/examination; straightforward medical decision making
99213	Office or other outpatient visit for the evaluation and management of established patient; expanded problem-focused history/examination; straightforward medical decision making
99214	Office or other outpatient visit for the evaluation and management of established patient; detailed problem-focused history/examination; low complexity medical decision making
99215	Office or other outpatient visit for the evaluation and management of established patient; comprehensive problem-focused history/examination; high complexity medical decision making
G0463	Hospital outpatient clinic visit for assessment and management of a patient

Appendix Table 2. CPT Procedure Codes Used to Identify Colonoscopy

Code	Definition
44388	Colonoscopy through stoma; diagnostic, including collection of specimen(s) by brushing or washing, when performed (separate procedure)
44389	Colonoscopy through stoma; with biopsy, single or multiple
44390	Colonoscopy through stoma; with removal of foreign body
44391	Colonoscopy through stoma; with control of bleeding (eg, injection, bipolar cautery, unipolar cautery, laser, heater probe, stapler, plasma coagulator)
44392	Colonoscopy through stoma; with removal of tumor(s), polyp(s), or other lesion(s) by hot biopsy forceps or bipolar cautery
44393	Colonoscopy through stoma; with ablation of tumor(s), polyp(s), or other lesion(s) not amenable to removal by forceps, cautery or snare
44394	Colonoscopy through stoma; with removal of tumor(s), polyp(s), or other lesion(s) by snare technique
44397	Colonoscopy through stoma; with transendoscopic stent placement (includes predilation)
44401	Colonoscopy through stoma; with ablation of tumor(s), polyp(s), or other lesion(s) (includes pre-and post-dilation and guide wire passage)
44402	Colonoscopy through stoma; with endoscopic stent placement (including pre- and post-dilation and guide wire passage, when performed)
44403	Colonoscopy through stoma; with endoscopic mucosal resection
44404	Colonoscopy through stoma; with directed submucosal injection(s), any substance
44405	Colonoscopy through stoma; with transendoscopic balloon dilation
44406	Colonoscopy through stoma; with endoscopic ultrasound examination, limited to the sigmoid, descending, transverse, or ascending colon and cecum
44407	Colonoscopy through stoma; with transendoscopic ultrasound guided intramural/transmural fine needle aspiration/biopsy(s)
45355	Colonoscopy, rigid or flexible, transabdominal via colotomy, single or multiple
45378	Colonoscopy, flexible, proximal to splenic flexure; diagnostic, with or without collection of specimen(s) by brushing or washing
45379	Colonoscopy, flexible, proximal to splenic flexure; with removal of foreign body
45380	Colonoscopy, flexible, proximal to splenic flexure; with biopsy, single or multiple
45381	Colonoscopy, flexible; with directed submucosal injection(s), any substance submucosal injection(s), any substance
45382	Colonoscopy, flexible; with control of bleeding, any method bleeding (eg, injection, cautery, laser, heater probe, stapler, plasma coagulator)
45383	Colonoscopy, flexible, proximal to splenic flexure; with ablation of tumors, polyps, or other lesions not removable by forceps, cautery or snare
45384	Colonoscopy, flexible; with removal of tumor(s), polyp(s), or other lesion(s) by hot biopsy forceps cautery
45385	Colonoscopy, flexible; with removal of tumor(s), polyp(s), or other lesion(s) by snare technique
45386	Colonoscopy, flexible; with transendoscopic balloon dilation balloon, 1 or more strictures
45387	Colonoscopy, flexible, proximal to splenic flexure; with transendoscopic stent placement (includes predilation)
45388	Colonoscopy, flexible; with ablation of tumor(s), polyp(s), or other lesion(s) (includes pre- and post-dilation and guide wire passage, when performed)
45389	Colonoscopy, flexible; with endoscopic stent placement (includes pre- and post-dilation and guide wire passage, when performed)
45390	Colonoscopy, flexible; with endoscopic mucosal resection
45391	Colonoscopy, flexible, proximal to splenic flexure; with endoscopic ultrasound examination
45392	Colonoscopy, flexible, proximal to splenic flexure; with transendoscopic ultrasound guided intramural or transmural fine needle aspiration/biopsy(s)
45393	Colonoscopy, flexible; with decompression (for pathologic distention) (eg, volvulus, megacolon), including placement of decompression tube
45398	Colonoscopy, flexible; with band ligation(s) (eg, hemorrhoids)
G0105	Colorectal cancer screening; colonoscopy on individual at high risk
G0121	Colorectal cancer screening; colonoscopy on individual not meeting criteria for high risk
G6019	Colonoscopy through stoma; with ablation of tumor(s), polyp(s), or other lesion(s) not amenable to removal by forceps, cautery or snare
G6020	Colonoscopy through stoma; with transendoscopic stent placement (includes predilation)
G6024	Colonoscopy, flexible; proximal to splenic flexure; with ablation of tumors, polyps, or other lesions not removable by forceps, cautery or snare
G6025	Colonoscopy, flexible, proximal to splenic flexure; with transendoscopic stent placement (includes predilation)

Appendix Table 3. Algorithms and codes used to derive study variables.

Condition	Primary or Sensitivity Analysis	Algorithm/ Codes	Comments
Iron Deficiency Anemia	Primary	Algorithm	Hemoglobin <130 g/L in men, <120 g/L in women AND Iron test within 3 months of Hemoglobin test showing Ferritin levels ≤15 ng/mL OR Transferrin saturation levels ≤16%
	Sensitivity	Algorithm	Hemoglobin <130 g/L in men, <120 g/L in women AND Iron test within 3 months of Hemoglobin test showing Ferritin levels ≤45 ng/mL OR Transferrin saturation levels ≤16%
Hematochezia	Primary	Codes	ICD-9: 569.3, 578.1 ICD-10: K62.5, K92.1
	Sensitivity	Codes	ICD-9: 569.3, 578.1, 578.9 ICD-10: K62.5, K92.1, K92.2
Body Mass Index	Primary	Algorithm	Measurement of Height (in inches) within 5 years of Follow-up start date AND Measurement of Weight (in pounds) within 1 year of Follow-up start date
Diabetes Prevalence	Primary	Algorithm	Prescription for diabetes medication in year of Follow-up start date (insulin, sulfonylureas, biguanides, thiazolidinediones, other hypoglycemic medications) OR 2+ diabetes codes for inpatient and/or outpatient visits (VA and Medicare) over a 24-month period (ICD-9: 250, 357.2, 362.0, 366.41)
Aspirin Use	Primary	Algorithm	At least 2 prescriptions or mentions of aspirin in free-text notes up to 1 year prior to Follow-up start date
Change in Bowel Habit	Sensitivity	Codes	Within ± 60 days of Follow-up start date: ICD-9: 787.99 ICD-10: R19.4
Unexplained Weight Loss	Sensitivity	Codes	Within ± 60 days of Follow-up start date: ICD-9: 783.21 ICD-10: R63.4
Prior Menorrhagia	Sensitivity	Codes	Before Follow-up start date: ICD-9: 626.2, 627.0 ICD-10: N92.0
Prior Hysterectomy	Sensitivity	Codes	Before Follow-up start date: ICD-9: V88.01 ICD-10: Z90.71, Z90.710

Appendix Table 4. Cancer characteristics of iron deficiency anemia (IDA) and hematochezia analytic cohorts.

	Iron Deficiency Anemia Cohort			Hematochezia Cohort		
	Overall N=239,000	**No IDA N=191,200**	**IDA N=47,800**	**Overall N=653,740**	**No Hematochezia N=522,992**	**Hematochezia N=130,748**
YCRC Cases:						
Anatomic Site						
Proximal	98 (38%)	17 (23%)	81 (44%)	85 (15%)	27 (18%)	58 (14%)
Distal	104 (40%)	36 (49%)	68 (37%)	251 (45%)	85 (57%)	166 (41%)
Rectal	51 (20%)	20 (27%)	31 (17%)	214 (38%)	36 (24%)	178 (44%)
Unknown	4 (2%)	0 (0%)	4 (2%)	6 (1%)	2 (1%)	4 (1%)
Stage at Detection						
Stage I	34 (13%)	8 (11%)	26 (14%)	111 (20%)	14 (9%)	97 (24%)
Stage II	45 (18%)	10 (14%)	35 (19%)	90 (16%)	15 (10%)	75 (18%)
Stage III	53 (21%)	17 (23%)	36 (20%)	125 (22%)	28 (19%)	97 (24%)
Stage IV	47 (18%)	13 (18%)	34 (18%)	100 (18%)	37 (25%)	63 (16%)
Unknown	78 (30%)	25 (34%)	53 (29%)	130 (23%)	56 (37%)	74 (18%)

Stage at detection defined using American Joint Committee on Cancer staging.

Appendix Table 5. Absolute risk and Cox proportional hazards model findings for overall joint exposure analyses in IDA and Hematochezia analytic cohorts.

		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)
IDA Cohort	IDA\ Hematochezia	1,320	31	2.50% (1.65%, 3.36%)	2.39% (1.52%, 3.23%)	42.6 (31.2, 62.8)	65.14 (42.64, 99.50)	65.29 (42.49, 100.32)
	No IDA/ Hematochezia*	237,680	226	0.12% (0.10%, 0.14%)				
Hematochezia Cohort	Hematochezia/ IDA	1,320	31	2.45% (1.69%, 3.44%)	2.36% (1.61%, 3.35%)	42.6 (31.2, 64.4)	86.07 (58.46, 126.70)	79.04 (53.38, 117.03)
	No Hematochezia/ IDA*	652,420	525	0.09% (0.08%, 0.10%)				

*Unexposed group includes individuals exposed to only IDA (IDA analytic cohort) or only hematochezia (Hematochezia analytic cohort)
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% confidence intervals
Unadjusted Model includes matching strata variable as random intercept; Adjusted Model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use

Appendix Table 6. Sensitivity Analysis excluding National Death Index-identified cases: Absolute risk and Cox proportional hazards model findings 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	141	0.33% (0.27%, 0.39%)	0.30% (0.24%, 0.35%)	339 (329.5, 404.7)	63.51 (40.60, 99.36)	12.12 (8.65, 16.98)
	No IDA	191,200	49	0.04% (0.03%, 0.05%)				
Men	IDA	22,445	120	0.62% (0.51%, 0.74%)	0.55% (0.44%, 0.68%)	140.3 (121.3, 167.7)	14.00 (9.64, 20.32)	13.83 (9.42, 20.30)
	No IDA	89,780	36	0.06% (0.04%, 0.09%)				
Women	IDA	25,355	21	0.10% (0.06%, 0.14%)	0.08% (0.04%, 0.13%)	1056.5 (765, 1683)	6.47 (3.24, 12.93)	7.73 (3.71, 16.14)
	No IDA	101,420	13	0.02% (0.01%, 0.03%)				
Age <30	IDA	4,925	5	0.10% (0.02%, 0.21%)	0.10% (0.01%, 0.20%)	985 (540.5, 4890.4)	20.85 (2.44, 178.49)	31.26 (3.36, 290.74)
	No IDA	20,505	1	0.01% (0.00%, 0.02%)				
Age 30-39	IDA	12,255	17	0.16% (0.08%, 0.24%)	0.15% (0.07%, 0.23%)	720.9 (488.2, 1384.7)	17.63 (5.93, 52.39)	18.04 (5.88, 55.40)
	No IDA	50,152	4	0.01% (0.00%, 0.03%)				
Age 40-49	IDA	30,620	119	0.45% (0.37%, 0.54%)	0.39% (0.31%, 0.48%)	257.3 (216.9, 316.1)	10.99 (7.78, 15.53)	11.31 (7.91, 16.17)
	No IDA	120,543	44	0.06% (0.04%, 0.08%)				

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% confidence intervals

Unadjusted Model includes matching strata variable as random intercept; Adjusted Model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use

Appendix Table 7. Sensitivity Analysis excluding National Death Index-identified cases: Absolute risk and Cox proportional hazards model findings for overall, sex-stratified and age-stratified analyses in hematochezia 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	Hematochezia	130,748	360	0.29% (0.26%, 0.32%)	0.26% (0.24%, 0.30%)	363.2 (329.5, 404.7)	14.18 (11.38, 17.67)	13.19 (10.50, 16.58)
	No Hematochezia	522,992	102	0.02% (0.02%, 0.03%)				
Men	Hematochezia	114,259	329	0.30% (0.27%, 0.34%)	0.28% (0.24%, 0.31%)	347.3 (314.6, 389.2)	13.49 (10.77, 16.91)	12.39 (9.80, 15.66)
	No Hematochezia	457,036	98	0.03% (0.02%, 0.03%)				
Women	Hematochezia	16,489	31	0.20% (0.13%, 0.27%)	0.19% (0.12%, 0.26%)	531.9 (390.8, 824.7)	31.07 (10.97, 88.01)	33.00 (11.27, 96.65)
	No Hematochezia	65,956	4	0.01% (0.00%, 0.01%)				
Age <30	Hematochezia	19,700	8	0.04% (0.02%, 0.07%)	0.04% (0.01%, 0.07%)	2462.5 (1409, 6559)	16.03 (3.40, 75.48)	20.58 (4.03, 105.18)
	No Hematochezia	78,875	2	0.002% (0.00%, 0.01%)				
Age 30-39	Hematochezia	32,506	47	0.14% (0.10%, 0.19%)	0.14% (0.10%, 0.18%)	691.6 (537, 967)	31.48 (13.46, 73.64)	29.27 (12.15, 70.52)
	No Hematochezia	130,294	6	0.004% (0.00%, 0.01%)				
Age 40-49	Hematochezia	78,542	305	0.42% (0.37%, 0.47%)	0.38% (0.33%, 0.43%)	257.5 (232.1, 289.1)	13.05 (10.36, 16.45)	12.15 (9.56, 15.43)
	No Hematochezia	313,823	94	0.04% (0.04%, 0.05%)				

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% confidence intervals
Unadjusted Model includes matching strata variable as random intercept; Adjusted Model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use

Appendix Table 8. Absolute risk and Cox proportional hazards model findings for overall analyses in IDA and Hematochezia analytic Cohorts with modifications to inclusion criteria.

		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)
IDA Cohort	IDA	59,127	212	0.42% (0.37%, 0.48%)	0.37% (0.32%, 0.43%)	238.1 (207.8, 272.9)	9.78 (7.63, 12.52)	10.45 (8.07, 13.53)
	No IDA	236,508	89	0.05% (0.04%, 0.06%)				
Hematochezia Cohort	Hematochezia	209,698	592	0.30% (0.28%, 0.33%)	0.28% (0.25%, 0.30%)	333 (307.5, 361.8)	13.72 (11.59, 16.25)	14.35 (12.01, 17.15)
	No Hematochezia	838,792	174	0.02% (0.02%, 0.03%)				

IDA definition is anemia based on hemoglobin test (<13.0 mg/dL in men and <12.0 mg/dL in women) and iron deficiency (ferritin levels <45 ng/mL or transferrin saturation ≤16%)
Hematochezia definition based on ICD-9 (569.3, 578.1, 578.9) and ICD-10 (K62.5, K92.1, K92.2) diagnostic codes.
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% confidence intervals
Unadjusted Model includes matching strata variable as random intercept; Adjusted Model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use

Appendix Table 9. Absolute risk and Cox proportional hazards model findings for overall analyses in IDA and hematochezia analytic cohorts excluding cases with both IDA and hematochezia.

		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)
IDA Cohort	IDA	46,480	153	0.39% (0.32%, 0.45%)	0.33% (0.26%, 0.40%)	303.8 (261.7, 357.6)	8.85 (6.68, 11.73)	9.25 (6.90, 12.40)
	No IDA	185,920	71	0.05% (0.04%, 0.07%)				
Hematochezia Cohort	Hematochezia	129,428	375	0.31% (0.28%, 0.34%)	0.28% (0.24%, 0.31%)	345.1 (312.7, 383.6)	10.12 (8.37, 12.23)	9.91 (8.12, 12.09)
	No Hematochezia	517,712	149	0.03% (0.03%, 0.04%)				

IDA and hematochezia-exposed groups exclude all individuals who had both IDA and hematochezia exposure (n=1,320).
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% confidence intervals
Unadjusted Model includes matching strata variable as random intercept; Adjusted Model additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use

Appendix Table 10. Cox proportional hazards model findings for overall analyses in IDA and Hematochezia analytic Cohorts comparing adjusted models.

	Cox Proportional Hazards Models		
	Unadjusted HR (95% CI)	**Adjusted HR Model 1 (95% CI)**	**Adjusted HR Model 2 (95% CI)**
IDA Cohort	10.35 (7.89, 13.57)	10.81 (8.15, 14.33)	10.58 (7.98, 14.04)
Hematochezia Cohort	10.88 (9.02, 13.12)	10.66 (8.76, 12.97)	9.94 (8.15, 12.12)

Unadjusted Model includes matching strata variable as random intercept
Adjusted Model 1 additionally adjusts for race/ethnicity, BMI, prevalent diabetes, smoking status and aspirin use
Adjusted Model 2 additionally adjusts for change in bowel habits and unexplained weight loss.

Appendix Table 11. RECORD Statement

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	p. 1-2	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract (p.2): Design Abstract (p.2): Design Abstract (p.2): Design
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	p. 4-6		
Objectives	3	State specific objectives, including any prespecified hypotheses	p. 4-6		
Methods					
Study Design	4	Present key elements of study design early in the paper	p. 6-8		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	p. 6-8		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	p. 6-8	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and	p. 6-8 N/A

		Cross-sectional study - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	p.4-6	not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	p. 6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	p. 7-8	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	p. 6-8 Appendix Table 3
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	p. 6-8		
Bias	9	Describe any efforts to address potential sources of bias	p. 6-9		
Study size	10	Explain how the study size was arrived at	p. 6-8		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	p. 6-8		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how	p. 8-10 p. 8-9 p. 9 N/A		

		loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	p. 6-9		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	p. 10 p. 9
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	p. 6
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	p. 10	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Appendix Figure 2
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest	p. 10, 15 Table 2 Table 2		

		(c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	p. 10, 15		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	p. 10, 15		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	p. 8-10, Tables 3-4 N/A p. 10-18, Tables 3-4		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	p. 10-19, Appendix Tables 4-9		
Discussion					
Key results	18	Summarise key results with reference to study objectives	p. 19-21		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	p. 21	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	p. 21

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	p. 19-21		
Generalisability	21	Discuss the generalisability (external validity) of the study results	p. 19-21		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	p. 23		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	p. 23