Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Women Reach the Advanced Colorectal Neoplasia Risk of Men 17 Years Later: A Cross-Sectional Study of 9,795 Colonoscopies

Oğuzhan Hakan Topgül

oguzhan_topgul@hotmail.com

Başakşehir Çam ve Sakura City Hospital

Osman Sıbıç

Haseki Training and Research Hospital

Sena Çağla Özden

Haseki Training and Research Hospital

Ahmet Zahit Kaan

Derecik State Hospital

Gamze Çıtlak

Haseki Training and Research Hospital

Research Article

Keywords: Colorectal neoplasms, Adenomatous polyps, Colonoscopy, Sex characteristics, Early detection of cancer

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11694421/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The male excess of colorectal neoplasia is well established, but has not been expressed as the age at which women reach the same absolute risk, the quantity relevant to when screening should begin.

Aims

To estimate this risk-equivalent age offset, locate where in the sequence the disparity arises, and test whether it is segment-specific.

Methods

We retrospectively analysed 9,795 consecutive single-centre colonoscopies performed between January 2020 and October 2025. Outcomes were extracted from free-text endoscopy and pathology reports by a rule-based natural language processing algorithm with organ-specific attribution. The primary outcome was advanced neoplasia (advanced adenoma or colorectal cancer). Odds ratios came from logistic regression adjusted for age and year with patient-clustered standard errors, refitted in five pre-specified subsets. Sex-specific risk curves were fitted and their horizontal distance measured at matched risk.

Results

Women reached the risk of 55-year-old men 16.8 years later (95% confidence interval 12.8 to 22.7), an offset stable across screening thresholds. Advanced neoplasia occurred in 10.7% of men and 5.4% of women (adjusted odds ratio 1.98, 1.70 to 2.32), and the excess was similar at every earlier stage, from 1.73 for cancer to 2.02 for advanced adenoma ($p < 0.001$). It did not interact with age ($p = 0.680$) and persisted proximally (2.19), distally (1.98) and in every subset.

Conclusions

On the age axis, women reach the risk of men about 17 years later. The disparity is present from the earliest precursor lesion and constant thereafter, raising the hypothesis that a risk-equivalent screening threshold is crossed earlier in men.

Introduction

Colorectal cancer is the third most common cancer and the second leading cause of cancer death worldwide [1]. Most cases arise through the adenoma–carcinoma sequence, in which adenomatous and

serrated precursors progress to invasive disease [2, 3]. Because the process is slow and its precursors removable, colonoscopy with polypectomy lowers both incidence and mortality [4, 5]. Screening therefore rests on detecting lesions before cancer develops, and the starting age is set by when the risk of clinically relevant lesions justifies the procedure [6, 7].

Male sex is among the strongest non-modifiable risk factors in this disease. Meta-analysis of screening populations shows that men carry roughly twice the advanced-neoplasia risk of women of the same age [8], and large screening cohorts report a higher prevalence of adenomas, advanced adenomas and cancer in men across all age strata [9, 10]. Sex-linked mechanisms, including oestrogen signalling in colonic mucosa, have been proposed but none is established as causal [11, 12]. Registry analyses estimated that the cancer risk of men is reached by women 5 to 10 years later, prompting the proposal of sex-specific screening start ages [13].

Two questions remain difficult to answer from the existing literature. The first is how the later presentation of women should be measured. Clinical series usually report the mean age of affected patients, a statistic confounded by the difference in underlying risk, since a lower-risk group contributes fewer and later cases regardless of biology. The quantity that bears on screening policy is the age at which women reach the same absolute risk as men, read as the horizontal distance between sex-specific risk curves, rarely estimated from endoscopic data [13]. The second is where in the sequence the disparity appears. A male excess confined to invasive cancer would point to sex differences in progression or detection [11, 12]; an excess already present in the earliest precursors, and of the same magnitude throughout, would point instead to an early determinant of lesion formation [2, 3]. Most prior work reports single endpoints rather than the whole sequence in one population under one case definition [8, 9, 10]. A third concern is anatomical: women harbour proportionally more proximal and serrated lesions, which are harder to detect [14, 15, 16], so an apparent difference could reflect lesion distribution rather than risk.

We therefore analysed a consecutive single-centre colonoscopy series with pathology-based outcomes. The **primary outcome** was advanced neoplasia (advanced adenoma or colorectal cancer). The **secondary outcomes**, ordered along the sequence, were any polyp, adenoma, advanced adenoma and colorectal cancer, together with the age offset at which women reach the advanced-neoplasia risk of men and the prevalence of adenoma and advanced neoplasia in the proximal and distal colon.

Methods

Study design and setting

This was a retrospective, single-centre, cross-sectional (prevalence) study conducted in the endoscopy unit of Haseki Training and Research Hospital, Istanbul, Türkiye. All consecutive colonoscopies performed between 2 January 2020 and 30 October 2025 were evaluated. Reporting follows the STROBE statement [17]; sample assembly is shown in Fig. 1a.

Ethics

The study followed the Declaration of Helsinki and was approved by the Non-Interventional Clinical Research Ethics Committee of Haseki Training and Research Hospital (approval number 295–2026, dated 16 September 2026). Because the design was retrospective and records were anonymised, the committee waived written informed consent.

Inclusion and exclusion criteria

All colonoscopies performed in the study period were included. The unit receives referrals both from the national screening programme and for symptomatic indications, so the series reflects routine practice rather than a screening cohort. No procedure was excluded on the basis of age, indication or completeness. The unit of analysis was the procedure; patients who underwent more than one colonoscopy contributed all of them, and the resulting non-independence was handled as described below.

Variables and outcome definitions

Age, sex, procedure date, examination type, diagnostic codes and the free-text endoscopy and pathology reports were extracted from the hospital electronic record; identifiers were permanently removed before analysis. Outcomes were extracted from the reports with a rule-based natural language processing algorithm adapted to Turkish, an approach validated for colonoscopy quality reporting in other languages [18, 19]; the outcome definitions are footnoted in Table 2. Because many reports also contained specimens from a concurrent upper gastrointestinal endoscopy, every diagnosis was attributed to its organ of origin and cancer labels were checked for negation, so that only colorectal malignancy contributed to the cancer outcome.

Adenoma was defined as any tubular, tubulovillous or villous adenoma. Advanced adenoma was defined as an adenoma with a villous component, high-grade dysplasia, or a maximum diameter of 1 cm or more. Advanced neoplasia, the primary outcome, was defined as advanced adenoma or colorectal cancer. Any polyp was a composite denoting at least one adenomatous, hyperplastic or serrated lesion; it is not a histological entity, and its components are reported separately. Lesion size was taken from the report in centimetres. Lesion location was grouped as proximal (caecum to transverse colon) or distal (splenic flexure to rectum, including the anal canal); when no segment was named, the recorded distance from the anal verge was used instead, with 60 cm as the boundary. A location could be assigned for 97 to 99% of procedures with neoplasia, and for 82.3% on named segments alone; the analysis was repeated on that subset.

Algorithm performance was checked against report text in 30 randomly selected reports and by dataset-wide logical constraints, such as every advanced adenoma also carrying an adenoma label; no violation remained.

Statistical analysis

Continuous variables are given as median with interquartile range, categorical variables as number with percentage, and proportions with Wilson 95% confidence intervals (CI). Groups were compared with the chi-square and Mann-Whitney U tests. The effect of male sex on each outcome was estimated by logistic regression adjusted for age and procedure year, reported as an adjusted odds ratio (OR) with 95% CI. Because 862 of 8,724 patients (9.9%) contributed more than one procedure, standard errors were clustered by patient. To test whether the association depended on how patients were selected or examined, the primary model was refitted in five pre-specified subsets: documented indication, documented screening indication, documented caecal intubation, documented adequate preparation, and first colonoscopy per patient.

Age confounding was addressed by stratification into decades (Fisher's exact test), direct standardisation to the pooled age distribution, and a likelihood-ratio test of the age-by-sex interaction; trends were tested with the Cochran-Armitage test. To express the later presentation of women on the age axis, logistic risk curves were fitted separately in each sex and the horizontal distance between them measured at matched absolute risk, with a bootstrap 95% CI from 2,000 replicates.

The pre-specified hypotheses were that the male excess begins at the precursor stage rather than at cancer, that it is proportionally constant across age and therefore appears as a horizontal shift of the risk curve, and that it is present in both colonic segments. Because the outcomes lie on the same biological pathway and are strongly correlated, no multiplicity adjustment was applied; all associations survive Bonferroni correction. Two-sided $p < 0.05$ was considered significant. Analyses used Python 3 (pandas, SciPy, statsmodels).

Results

Study sample

Of 9,795 colonoscopies, 4,654 (47.5%) were in men and 5,141 (52.5%) in women (Fig. 1a). Median age was 57 years in men and 54 in women ($p < 0.001$), a three-year difference (Fig. 1b). Pathology material was available for 69.3% of procedures, with no difference between the sexes ($p = 0.754$); a specimen is not taken when the examination is macroscopically normal. A narrative report was retrievable for 7,719 procedures (78.8%), the remainder being polypectomies whose operative note was not exported; among these, caecal or ileocaecal visualisation was documented in 88.0% and bowel preparation was recorded as adequate in 5.9% (Table 1).

Table 1
Patient and procedure characteristics by sex

Characteristic	Men (n = 4,654)	Women (n = 5,141)	p
Age, median (IQR), years	57 (48–66)	54 (46–64)	< 0.001
Age group, n (%)			
<45 years	899 (19.3)	1,133 (22.0)	
45–49 years	464 (10.0)	607 (11.8)	
50–69 years	2,583 (55.5)	2,792 (54.3)	
≥70 years	708 (15.2)	609 (11.8)	
Procedure code recorded in the hospital system, n (%)			< 0.001
Diagnostic colonoscopy	3,065 (65.9)	4,015 (78.1)	
Colonoscopy with polypectomy	1,276 (27.4)	830 (16.1)	
Colonoscopy with biopsy	313 (6.7)	296 (5.8)	
Narrative report retrievable, n (%)	3,403 (73.1)	4,316 (84.0)	< 0.001
Pathology material obtained, n (%)	3,231 (69.4)	3,553 (69.1)	0.754
Among procedures with a retrievable report:			
Caecal or ileocaecal valve described, n (%)	2,942 (86.5)	3,850 (89.2)	< 0.001
Bowel preparation documented adequate, n (%)	204 (6.0)	248 (5.7)	0.680
Indication documented, n (%)	362 (10.6)	465 (10.8)	0.877
Recorded indication, n (%)			0.891
Screening, faecal occult blood positive or family history	50 (1.5)	62 (1.4)	
Symptomatic	309 (9.1)	401 (9.3)	
Surveillance or post-operative	3 (0.1)	2 (0.0)	

CI confidence interval, IQR interquartile range, OR odds ratio. Odds ratios are adjusted for age and procedure year with standard errors clustered by patient.

Primary outcome: advanced neoplasia

Advanced neoplasia was found in 10.7% of men (95% CI 9.8 to 11.6) and 5.4% of women (4.8 to 6.0). Adjusted for age and procedure year, men had almost twice the odds of women (OR 1.98, 95% CI 1.70 to 2.32; p < 0.001; Table 2, Fig. 2b), unchanged when the analysis was restricted to each patient's first colonoscopy (OR 2.00, 1.70 to 2.36).

Table 2
Prevalence and adjusted odds ratios for every histological outcome, by sex

Histological outcome	Men % (95% CI)	Women % (95% CI)	Adjusted OR (95% CI)	p
Any polyp	28.3 (27.1–29.7)	17.4 (16.4–18.5)	1.80 (1.64–1.99)	< 0.001
Adenoma	22.5 (21.3–23.7)	12.6 (11.8–13.6)	1.91 (1.71–2.13)	< 0.001
Tubular adenoma	17.5 (16.5–18.7)	9.6 (8.9–10.5)	1.90 (1.68–2.14)	< 0.001
Tubulovillous adenoma	3.5 (3.0–4.0)	1.6 (1.3–2.0)	2.01 (1.54–2.63)	< 0.001
Adenoma with villous component □	4.5 (3.9–5.1)	2.3 (1.9–2.7)	1.89 (1.50–2.39)	< 0.001
Serrated lesion	0.6 (0.4–0.8)	0.2 (0.1–0.4)	2.94 (1.42–6.09)	0.004
Hyperplastic polyp	9.3 (8.5–10.2)	6.5 (5.8–7.2)	1.44 (1.24–1.68)	< 0.001
High-grade dysplasia	2.1 (1.7–2.5)	1.0 (0.8–1.3)	1.91 (1.36–2.68)	< 0.001
Advanced adenoma □	8.2 (7.5–9.0)	4.0 (3.5–4.6)	2.02 (1.69–2.40)	< 0.001
Advanced neoplasia □	10.7 (9.8–11.6)	5.4 (4.8–6.0)	1.98 (1.70–2.32)	< 0.001
Colorectal cancer □	3.4 (2.9–4.0)	1.9 (1.5–2.3)	1.73 (1.33–2.23)	< 0.001

CI confidence interval, IQR interquartile range, OR odds ratio. Odds ratios are adjusted for age and procedure year with standard errors clustered by patient.

Secondary outcomes along the sequence

Every outcome was more frequent in men: any polyp 28.3% versus 17.4%, adenoma 22.5% versus 12.6%, advanced adenoma 8.2% versus 4.0% and colorectal cancer 3.4% versus 1.9% (all p < 0.001; Table 2, Fig. 2a). The same held for the morphological subtypes, including tubular, tubulovillous and villous adenoma and high-grade dysplasia (Table 2).

The adjusted odds ratios stayed within a narrow band along the sequence: any polyp 1.80 (95% CI 1.64 to 1.99), adenoma 1.91 (1.71 to 2.13), advanced adenoma 2.02 (1.69 to 2.40) and colorectal cancer 1.73 (1.33 to 2.23), all $p < 0.001$ (Table 2). The male excess was therefore established in the earliest precursor lesion and did not grow as lesions progressed towards cancer. Procedure year was not associated with any outcome, the variance inflation factor was about 1.0 for all predictors, and discrimination was modest, as expected for a model estimating an association rather than predicting individual risk (C-statistic 0.67 for advanced neoplasia).

Effect of age on the sex difference

Age did not produce the sex difference (Table 3, Fig. 3b). Median age differed by three years, and age-standardised rate ratios were consistent from precursor lesion to cancer: 1.57 for any polyp, 1.69 for adenoma, 1.94 for advanced adenoma, 1.88 for advanced neoplasia and 1.72 for cancer (all $p < 0.001$). In every age decade from 40 years onwards the odds ratio exceeded 1 and lay close to the pooled value of about 2.0; the group below 30 years had too few events to interpret. The age-by-sex interaction was not significant (likelihood-ratio $p = 0.680$), so the relative excess in men was the same at every age.

Table 3
(A) Age-standardised rates and rate ratios, and (B) the age at which each sex reaches the same advanced-neoplasia risk, with bootstrap 95% confidence intervals

Measure	Men	Women	Rate ratio or offset (years)
A. Age-standardised rates (direct, pooled age distribution)			
Any polyp	27.9	17.8	1.57
Adenoma	22.0	13.0	1.69
Advanced adenoma	8.0	4.1	1.94
Advanced neoplasia	10.4	5.5	1.88
Colorectal cancer	3.3	1.9	1.72
B. Age at which each sex reaches the same advanced-neoplasia risk			
risk 6.6%	45 years	62.5 years	17.5 (95% CI 13.0–22.2)
risk 7.9%	50 years	67.2 years	17.2 (95% CI 13.0–22.1)
risk 9.4%	55 years	71.8 years	16.8 (95% CI 12.8–22.7)
risk 11.2%	60 years	76.5 years	16.5 (95% CI 12.0–23.3)
risk 13.3%	65 years	81.2 years	16.2 (95% CI 11.2–24.5)
C. Advanced-neoplasia odds ratio for male sex in pre-specified subsets	n	Events	Adjusted OR (95% CI)
Full cohort	9,795	775	1.98 (1.70–2.32)
Documented indication	827	31	2.00 (0.95–4.19)
Documented caecal intubation	6,792	181	1.81 (1.34–2.45)
Documented adequate bowel preparation	452	24	1.88 (0.80–4.46)
First colonoscopy per patient	8,725	584	2.08 (1.74–2.48)

CI confidence interval, IQR interquartile range, OR odds ratio. Odds ratios are adjusted for age and procedure year with standard errors clustered by patient.

How much later women reach the same risk

Because the relative excess is constant while absolute risk rises with age, a constant odds ratio corresponds to a horizontal shift of the risk curve along the age axis (Fig. 3a). Men aged 50 to 59 years had an advanced-neoplasia rate of 10.6% (95% CI 9.0 to 12.4); women of the same age had 5.5% (4.5 to 6.8) and did not reach that level until 70 years or above (10.3%, 8.2 to 13.0). Fitting the curves, women reached the risk predicted for men at age 55 only 16.8 years later (95% CI 12.8 to 22.7), and the offset changed little across screening-relevant thresholds, from 17.5 years at the risk of 45-year-old men to 17.2 years at that of 50-year-old men (Table 3). A ten-year shift would not equalise risk: women aged 60 to 69 years had a rate of 6.6%, well below the 10.6% of men aged 50 to 59 years. Cancer became more common with age, rising from 1.1% below 45 years to 1.5% at 45 to 49 years and 3.2% at 50 years or above (trend $p < 0.001$).

Robustness to selection and examination quality

The association did not depend on patient selection or examination quality. Against 1.98 (95% CI 1.70 to 2.32) in the full cohort, the adjusted odds ratio for advanced neoplasia was 2.00 (0.95 to 4.19) with a documented indication, 1.81 (1.34 to 2.45) with documented caecal intubation, 1.88 (0.80 to 4.46) with documented adequate preparation, and 2.08 (1.74 to 2.48) with one colonoscopy per patient (Table 3, panel C). Intervals in the smaller subsets are wide because they contain few events; the screening subset, with three events, was not estimable. The recorded indication mix did not differ between the sexes ($p = 0.678$), and the excess was the same in patients aged 50 years and over (OR 2.04, 1.73 to 2.42) as in those under 45 years (1.91, 1.12 to 3.25).

Lesion location

A colonic location could be assigned in 97 to 99% of procedures with neoplasia, and the male excess was of similar size in both segments (Fig. 2c, d). For adenoma the adjusted odds ratio was 1.93 (95% CI 1.63 to 2.29) proximally and 1.84 (1.63 to 2.07) distally; for advanced neoplasia it was 2.19 (1.67 to 2.87) proximally and 1.98 (1.68 to 2.35) distally, all $p < 0.001$. Named segments alone, discarding the distance rule, gave the same result (proximal adenoma 1.94, distal adenoma 1.73, proximal advanced neoplasia 2.22, distal advanced neoplasia 1.94). The sex difference was therefore not explained by a shift in where lesions occur.

Discussion

In this series of 9,795 colonoscopies, women reached the advanced-neoplasia risk of 55-year-old men 16.8 years later (95% CI 12.8 to 22.7), an offset stable across the thresholds relevant to screening. Underlying this shift, neoplasia was about twice as frequent in men at every point of the sequence, with odds ratios between 1.7 and 2.1; the excess was the same in each age decade above 40 years and occurred in both halves of the colon.

The flat profile of odds ratios implies that the sex difference is set early. Had male sex accelerated the progression of established adenomas, the odds ratio would have widened from adenoma to cancer; had it affected detection of cancer, the difference would have been largest at the cancer endpoint. Neither

occurred. The simplest reading is that men form precursor lesions more readily and carry that excess forward unchanged in relative terms. This fits work proposing that oestrogen signalling acts on early mucosal events rather than late invasion [11, 12], although our data cannot test mechanism.

The ratio we report matches screening populations, in which a twofold male excess of advanced neoplasia has been described both in meta-analysis [8] and in screening cohorts across age strata [9, 10]. A mixed-indication series cannot estimate population prevalence, and ours does not: the absolute figures exceed screening-programme yields because the patients were referred. The sex ratio is a different quantity. Both sexes entered through the same pathway, their indication mix did not differ, and the ratio was unchanged in every subset examined. Selection would have to act differently in men and women to generate a twofold ratio, and nothing here indicates that it does.

The sex effect can also be read against age: an odds ratio of 1.98 for male sex alongside 1.49 per decade means that being male is worth about 1.7 decades, reproducing the 16.8-year offset from an independent direction. That offset is larger than the figure usually quoted. Brenner and colleagues, using registry incidence data, estimated that women reach the colorectal cancer risk of men roughly 5 to 10 years later and used this to argue for sex-specific screening ages [13]. Ours is greater for two reasons: our endpoint is advanced neoplasia rather than cancer, so the same relative difference maps onto a longer stretch of the age axis, and we anchored the comparison on equal absolute risk rather than on the mean age of affected patients. The estimates measure different quantities and are not in conflict.

Together, the constant twofold excess and the 17-year offset raise a question for screening research: whether an equal-risk principle, under which screening begins once a defined level of advanced-neoplasia risk is reached, would imply a lower starting age in men and a higher one in women than the sex-neutral threshold in current guidelines [6, 7]. The size of any such adjustment matters: a difference of about a decade would be insufficient here, since women a decade older than men still carried appreciably lower risk. We pose this as a question, not a recommendation. Cancer was present in 1.1% of procedures below 45 years, an age band below most screening thresholds and one in which incidence is rising [20]. Women harbour proportionally more proximal and serrated lesions, which are more easily missed [14, 15]; raising the starting age in women could widen inequity; and screening policy rests on incidence, mortality and cost-effectiveness rather than cross-sectional detection rates. These findings support formal modelling of risk-equivalent starting ages in dedicated screening cohorts, alongside the sex-stratified quality measures already recommended [21].

The location analysis addresses the most obvious alternative explanation. If the excess simply reflected the tendency of women to harbour proximal, endoscopically subtle lesions, the difference should have been concentrated distally. Instead the odds ratios were similar in the two segments and slightly higher proximally, so the disparity involves the whole colorectal mucosa rather than one territory [16, 22].

Limitations

The main limitation is incomplete documentation of indication: an indication was recorded in 827 procedures (8.4%), the hospital diagnosis codes were default billing entries, and no screening population could be isolated. The absolute prevalences are therefore detection rates in a referred, mixed-indication population, not population estimates, and any inference about when screening should begin remains a hypothesis for screening cohorts. The sex ratio is less exposed, because both sexes were referred through the same pathway and the ratio was unchanged in every subset, but residual confounding by unmeasured indication cannot be excluded, which is why the comparison is presented as an association.

Second, outcomes were classified from free-text reports by algorithm rather than by pathological re-review; residual misclassification is likely non-differential and would bias the odds ratios towards the null. Pathology material was absent for about one third of procedures, so prevalences are procedure-level detection rates. Third, the operative note was not exported for 2,061 polypectomy procedures, so documented caecal intubation and preparation describe the 7,719 procedures with a retrievable report, and endoscopist-level detection rate was unavailable. Fourth, location was inferred from report text, less precise than endoscopic mapping, although the result was unchanged on named segments alone. The cross-sectional design cannot establish causality.

Conclusions

Colorectal neoplasia was about twice as frequent in men at every stage of the sequence, and the difference originated at the precursor lesion rather than at cancer. On the age axis, women reached the risk of 55-year-old men about 17 years later. These findings generate the testable hypothesis that a risk-equivalent screening threshold is crossed earlier in men. Whether that should translate into sex-specific starting ages cannot be settled by a cross-sectional series and requires cohorts with incidence and mortality endpoints. For the endoscopy unit, a woman referred at a given age carries the lesion burden of a man about two decades younger.

Declarations

Ethics approval and consent to participate The study was approved by the Non-Interventional Clinical Research Ethics Committee of Haseki Training and Research Hospital, Istanbul, Turkiye (approval number 295-2026, dated 16 September 2026) and conducted in accordance with the Declaration of Helsinki. The requirement for individual informed consent was waived because the study was retrospective and used anonymised records.

Consent for publication All records were irreversibly anonymised before analysis. No individually identifiable information, photographs or video appear in the manuscript, and no single patient can be identified from the data presented, so patient anonymity is fully preserved and consent for publication was not required.

Competing interests The authors declare no competing interests.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Statistical analysis All statistical analyses were planned and performed by the authors using standard methods for cross-sectional data (logistic regression with patient-clustered standard errors, direct age standardisation, and bootstrap confidence intervals). No external statistical consultation was obtained.

Availability of data and materials The analysis code and the anonymised derived dataset are available from the corresponding author on reasonable request.

Authors' contributions O.H.T. conceived and designed the study, analysed and interpreted the data and drafted the manuscript. O.S., S.Ç.Ö. and A.Z.K. acquired the endoscopy and pathology data and contributed to interpretation. G.Ç. supervised the study and critically revised the manuscript. All authors read and approved the final version.

Acknowledgements None.

Prior presentations: An abstract of this work has been accepted for presentation at the 18th National Congress of Endoscopic, Laparoscopic and Robotic Surgery, Gloria Golf Resort Hotel, Antalya, Türkiye, 29 October to 1 November 2026. The work has not yet been presented and the abstract has not been published.

Funding/Support: This study received no grant support from any funding agency in the public, commercial or not-for-profit sectors.

Financial Disclosures: None reported. No author has a financial relationship with any entity with an interest in the content of this work, during the 36 months before submission or at any other time.

Conflicts of Interest: The authors declare no conflicts of interest.

Use of artificial intelligence: Claude (Anthropic), Claude Opus 5, accessed through the Claude Science research environment, was used between July and November 2026 under author supervision. It was applied to the analysis code, to the programmatic generation of Tables 1 to 3 and Figures 1 to 3 from that code, to literature searching, and to drafting and language editing. The sections concerned are Methods (natural language processing and statistical code), Results (all reported values, tables and figures) and the Introduction and Discussion (drafting and editing of text whose content and interpretation were specified by the authors). Outputs were validated by case-by-case comparison of algorithm labels with the original reports, by dataset-wide logical constraints, and by checking every reported statistic against the saved analysis output. No data were generated, altered or imputed by the tool, and no tool is listed as an author. The study had ethics committee approval with a waiver of individual consent; direct identifiers were removed and the records screened for residual identifiers before analysis, and no patient is identifiable from the manuscript.

Video abstract: If the manuscript is provisionally accepted, a video abstract will be required prior to final acceptance.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209–249. PMID: 33538338.
2. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61:759–67. PMID: 2188735.
3. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. Br J Surg. 2002;89:845–60. PMID: 12081733.
4. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med. 1993;329:1977-81. PMID: 8247072.
5. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med. 2012;366:687–96. PMID: 22356322.
6. US Preventive Services Task Force, Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2021;325:1965–1977. PMID: 34003218.
7. Shaukat A, Kahi CJ, Burke CA, Rabeneck L, Sauer BG, Rex DK. ACG Clinical Guidelines: Colorectal Cancer Screening 2021. Am J Gastroenterol. 2021;116:458–479. PMID: 33657038.
8. Nguyen SP, Bent S, Chen YH, Terdiman JP. Gender as a risk factor for advanced neoplasia and colorectal cancer: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2009;7:676 –81.e1-3. PMID: 19514116.
9. Ferlitsch M, Reinhart K, Pramhas S, Wiener C, Gal O, Bannert C, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. JAMA. 2011;306:1352–8. PMID: 21954479.
10. Regula J, Rupinski M, Kraszewska E, Polkowski M, Pachlewski J, Orlowska J, et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. N Engl J Med. 2006;355:1863–72. PMID: 17079760.
11. Kim SE, Paik HY, Yoon H, Lee JE, Kim N, Sung MK. Sex- and gender-specific disparities in colorectal cancer risk. World J Gastroenterol. 2015;21:5167-75. PMID: 25954090.
12. Tsokkou S, Konstantinidis I, Papakonstantinou M, Chatzikomnitsa P, Liampou E, Toutziari E, et al. Sex Differences in Colorectal Cancer: Epidemiology, Risk Factors, and Clinical Outcomes. J Clin Med. 2025;14(15). PMID: 40807160.

13. Brenner H, Hoffmeister M, Arndt V, Haug U. Gender differences in colorectal cancer: implications for age at initiation of screening. Br J Cancer. 2007;96:828–31. PMID: 17311019.
14. Hazewinkel Y, de Wijkerslooth TR, Stoop EM, Bossuyt PM, Biermann K, van de Vijver MJ, et al. Prevalence of serrated polyps and association with synchronous advanced neoplasia in screening colonoscopy. Endoscopy. 2014;46:219–24. PMID: 24254386.
15. Tan CK, Suzuki S, Ang TL, Koh XH, Wang LM, Aniwan S, et al. Sessile serrated lesion prevalence and factors associated with their detection: a post-hoc analysis of a multinational randomized controlled trial from Asia. Endoscopy. 2024;56:684–693. PMID: 38857619.
16. Liang B, Liu X, Zhang T, Xing D. Sex as a Contextual Modifier in Colorectal Cancer: Integrating Tumor Sidedness, Molecular Subtype, Immune Ecology, and Early-Onset Disease. Cancers (Basel). 2026;18(8). PMID: 42073632.
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Epidemiology. 2007;18:800–4. PMID: 18049194.
18. Tinmouth J, Swain D, Chorneyko K, Lee V, Bowes B, Li Y, et al. Validation of a natural language processing algorithm to identify adenomas and measure adenoma detection rates across a health system: a population-level study. Gastrointest Endosc. 2023;97:121–129.e1. PMID: 35843286.
19. Bae JH, Han HW, Yang SY, Song G, Sa S, Chung GE, et al. Natural Language Processing for Assessing Quality Indicators in Free-Text Colonoscopy and Pathology Reports: Development and Usability Study. JMIR Med Inform. 2022;10:e35257. PMID: 35436226.
20. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, et al. Colorectal Cancer Incidence Patterns in the United States, 1974–2013. J Natl Cancer Inst. 2017;109(8). PMID: 28376186.
21. Kaminski MF, Thomas-Gibson S, Bugajski M, Bretthauer M, Rees CJ, Dekker E, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy. 2017;49:378–397. PMID: 28268235.
22. Jin EH, Han K, Shin CM, Lee DH, Kang SJ, Lim JH, et al. Sex and Tumor-Site Differences in the Association of Alcohol Intake With the Risk of Early-Onset Colorectal Cancer. J Clin Oncol. 2023;41:3816–3825. PMID: 37315287.

Figures

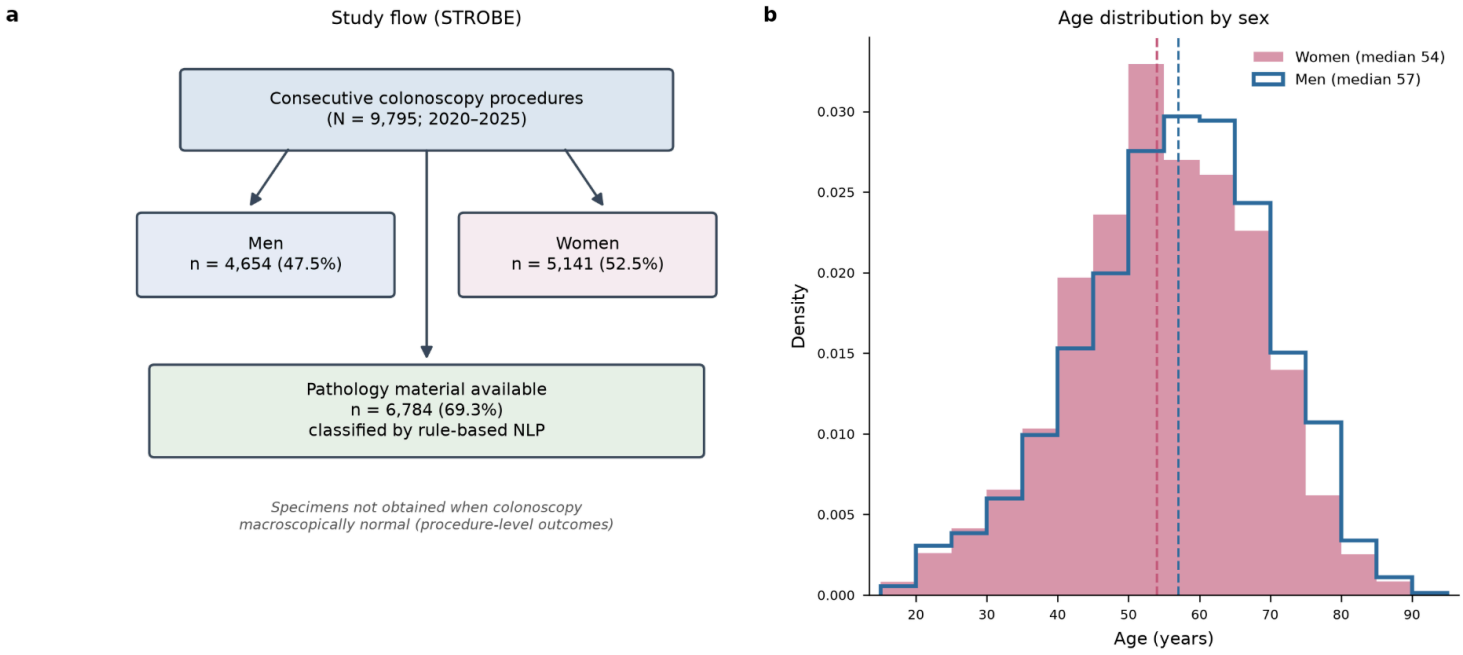

Figure 1

Study sample and age distribution. a Assembly of the study sample. b Age distribution by sex; dashed lines mark the median in each sex.

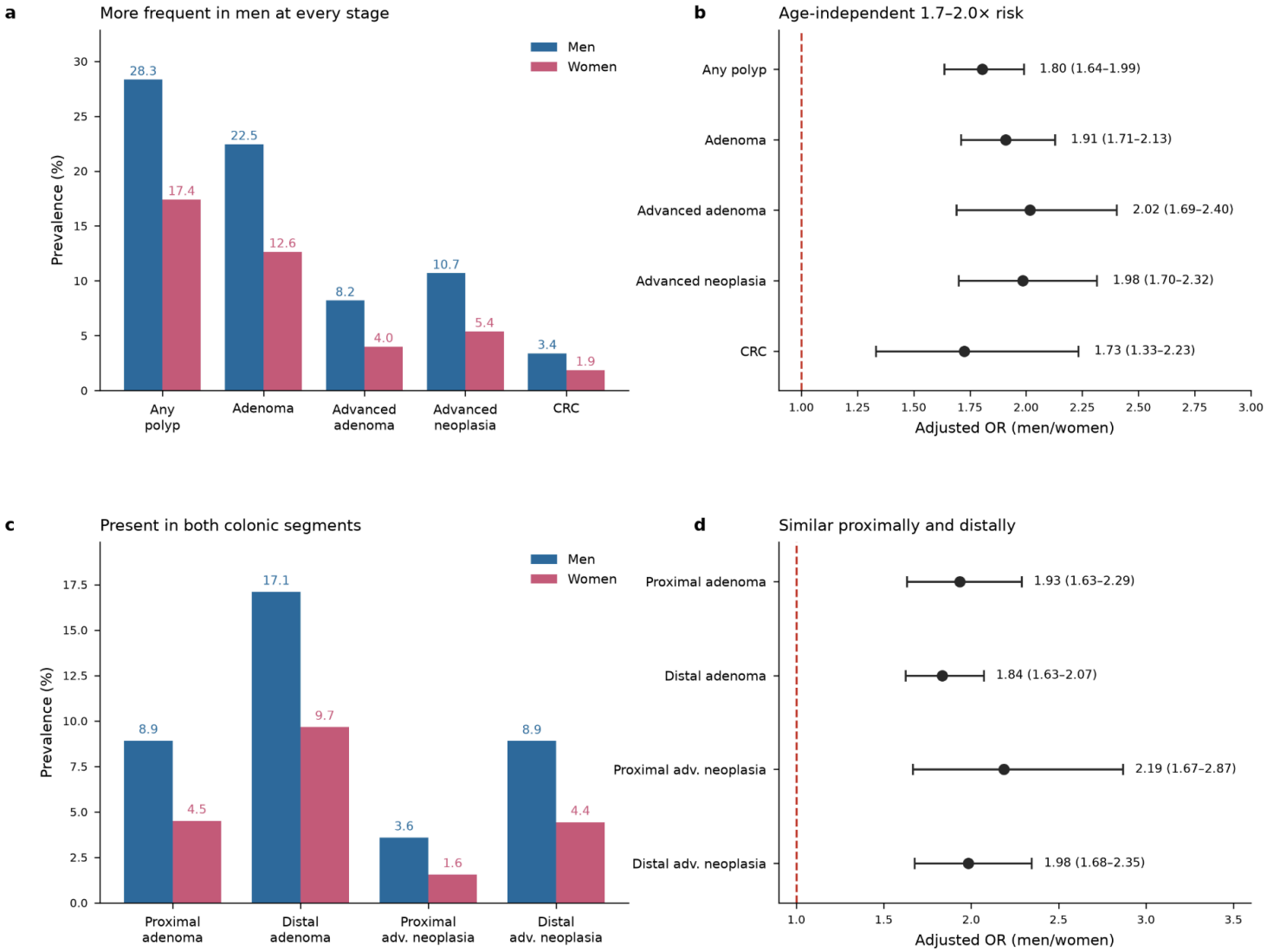

Figure 2

Sex difference across the adenoma–carcinoma sequence and by colonic segment. a Prevalence of each outcome in men and women, with values shown above the bars. b Odds ratios for male sex adjusted for age and procedure year, with 95% confidence intervals; the dashed line marks no difference. c Prevalence of proximal and distal adenoma and advanced neoplasia. d Adjusted odds ratios for male sex by segment.

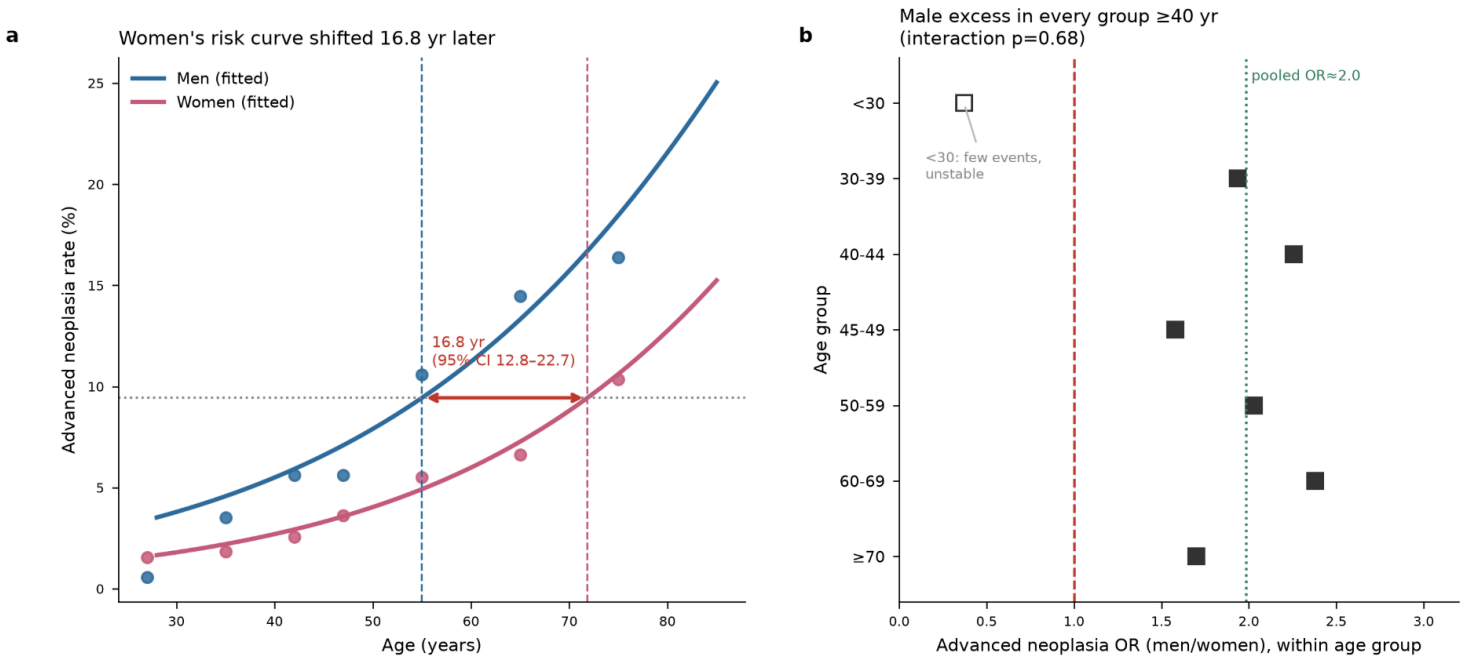

Figure 3

Effect of age on the sex difference in advanced neoplasia. a Fitted logistic risk curves by sex, with observed age-band rates shown as points; the horizontal arrows mark the age at which women reach the advanced-neoplasia risk of men aged 45 and 50 years. b Odds ratio for male sex within each age decade; the open marker denotes the group below 30 years, which had too few events to be informative, and the dotted line marks the pooled estimate.

CI confidence interval; CRC colorectal cancer; OR odds ratio.