

Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies

Xiaoliang Wang ,^{1,2} Kelli O'Connell,³ Jihyouon Jeon,⁴ Mingyang Song,^{5,6} David Hunter,^{7,8} Michael Hoffmeister,⁹ Yi Lin,¹ Sonja Berndt,¹⁰ Hermann Brenner,⁹ Andrew T Chan,^{11,12} Jenny Chang-Claude,¹³ Jian Gong,¹ Marc J Gunter,¹⁴ Tabitha A Harrison,¹ Richard B Hayes,¹⁵ Amit Joshi,^{7,16} Polly Newcomb,^{1,2} Robert Schoen,¹⁷ Martha L Slattery,¹⁸ Ashley Vargas,¹⁹ John D Potter,^{1,2} Loic Le Marchand,²⁰ Edward Giovannucci,^{5,6} Emily White,^{1,2} Li Hsu,¹ Ulrike Peters,^{1,2} Mengmeng Du³

To cite: Wang X, O'Connell K, Jeon J, *et al.* Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. *BMJ Open Gastro* 2019;**6**:e000339. doi:10.1136/bmjgast-2019-000339

Received 15 August 2019
Revised 17 October 2019
Accepted 25 October 2019



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

For numbered affiliations see end of article.

Correspondence to

Dr Xiaoliang Wang;
xwang23@fredhutch.org

ABSTRACT

Objective 'Environmental' factors associated with colorectal cancer (CRC) risk include modifiable and non-modifiable variables. Whether those with different non-modifiable baseline risks will benefit similarly from reducing their modifiable CRC risks remains unclear.

Design Using 7945 cases and 8893 controls from 11 population-based studies, we combined 17 risk factors to characterise the overall environmental predisposition to CRC (environmental risk score (E-score)). We estimated the absolute risks (ARs) of CRC of 10 and 30 years across E-score using incidence-rate data from the Surveillance, Epidemiology, and End Results programme. We then combined the modifiable risk factors and estimated ARs across the modifiable risk score, stratified by non-modifiable risk profile based on genetic predisposition, family history and height.

Results Higher E-score was associated with increased CRC risk (OR_{quartile} 1.33; 95% CI 1.30 to 1.37). Across E-scores, 30-year ARs of CRC increased from 2.5% in the lowest quartile (Q1) to 5.9% in the highest (Q4) quartile for men, and from 2.1% to 4.5% for women. The modifiable risk score had a stronger association in those with high non-modifiable risk (relative excess risk due to interaction=1.2, 95% CI 0.5 to 1.9). For those in Q4 of non-modifiable risk, a decrease in modifiable risk reduced 30-year ARs from 8.9% to 3.4% for men and from 6.0% to 3.2% for women, a level lower or comparable to the average population risk.

Conclusions Changes in modifiable risk factors may result in a substantial decline in CRC risk in both sexes. Those with high inherited risk may reap greater benefit from lifestyle modifications. Our results suggested comprehensive evaluation of environmental factors may facilitate CRC risk stratification.

INTRODUCTION

Colorectal cancer (CRC) is one of the most common and fatal cancers in the world.¹ In

Significance of this study

What is already known on this subject?

- ▶ Epidemiological studies have successfully identified many anthropometric, dietary, lifestyle, and pharmacological factors associated with colorectal cancer (CRC) risk.
- ▶ The American Cancer Society has provided a guideline on nutrition and physical activity for overall cancer prevention.

What are the new findings?

- ▶ By comprehensively evaluating the overall environmental predisposition from known risk factors, we found that a higher environmental risk score was associated with higher CRC risk.
- ▶ The 30-year absolute risk of CRC among participants with the highest non-modifiable baseline risk can be dropped to a level comparable to or below population average risks by changing modifiable risk scores from the highest to the lowest quartile.

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

- ▶ Comprehensive evaluation of environmental factors can facilitate targeted risk management and screening strategies for CRC prevention.
- ▶ Individuals with higher baseline CRC risk due to genetic predisposition or family history can reduce the long-term CRC risk by modifying various environmental and lifestyle factors.
- ▶ Individuals with higher CRC risk may have options to change other lifestyle factors to achieve the same risk management effect if modifying a particular risk factor, such as aspirin use, is not clinically advisable or feasible.

the USA, there were an estimated 140 250 new cases and 50 630 deaths in 2018.² Epidemiological studies have successfully



identified many anthropometric, dietary, lifestyle, and pharmacological factors associated with CRC risk (collectively referred to here as 'environmental' factors). Risk-increasing factors include greater height,³ obesity,⁴ smoking,⁵⁻⁷ alcohol intake,^{8,9} red and processed meat intake,^{10,11} and diabetes.¹² In contrast, factors associated with reduced risk include physical activity,^{13,14} use of aspirin or other non-steroidal anti-inflammatory drugs (NSAIDs),¹⁵ use of postmenopausal hormone (PMH) in women,¹⁶ and intakes of fruits, vegetables, calcium, folate, and fibre.^{17,18} These data have informed current guidance for primary CRC prevention. The American Cancer Society (ACS), for instance, provided a guideline on nutrition and physical activity for overall cancer prevention.¹⁹ The US Preventive Services Task Force also recently recommended aspirin as a chemopreventive agent for those at moderate cardiovascular disease risk.²⁰

Previous studies have focused on a single or a restricted group of risk factors. A case-control study found that having more healthy lifestyle factors was associated with increasingly lower risks of CRC, regardless of genetic risk.²¹ A higher at-risk lifestyle score from four factors was found to be associated with higher risk of colon and rectal cancer in two population-based case-control studies.²² However, other CRC risk factors have not been evaluated in a comprehensive summary score. Moreover, an individual's CRC risk also depends on factors that are unlikely to be modified, such as age, sex, height, CRC family history, and common genetic predisposition.²³⁻³³ It is unknown whether changing modifiable risk factors has similar benefits among those with high versus low non-modifiable CRC risk profiles. In a breast cancer consortium of prospective studies, Maas *et al* demonstrated that improvement in estimating absolute risk of breast cancer can identify subsets of the population at an elevated risk who

would benefit most from risk-reduction strategies such as altering modifiable factors, suggesting the utility of comprehensive risk modelling.³⁴

Here, we assessed whether comprehensively aggregating information across environmental factors can improve CRC risk stratification in the general population. We first developed a framework to build an overall environmental risk score (E-score) based on risk factors identified in published studies. We evaluated the relative risk and long-term absolute risks of CRC by E-score. Finally, we estimated absolute risks across modifiable risk factors stratified by non-modifiable baseline risks.

METHODS

Study population

We included 16 838 participants (7945 cases and 8893 controls) from seven nested case-control studies in prospective US cohorts and four case-control studies from the USA and Europe, from the Genetics and Epidemiology of Colorectal Cancer Consortium. Details have been published^{25,35} and are summarised in [table 1](#).

Cases were identified as those with incident, invasive CRC (n=7155) or advanced adenoma (n=790), confirmed by medical record, pathology report or death certificate. Population-based controls were selected based on study-specific eligibility and matching criteria (mostly age and sex). For the small subset of advanced adenoma cases, matched controls also had a polyp-free sigmoidoscopy or colonoscopy at the time of adenoma selection.³⁶

Only participants with European ancestry were included, and race/ethnicity was confirmed using principal component analysis.³⁷

Patient and public involvement

Patients were not involved in the conduct of this study.

Table 1 Demographic characteristics of participating studies

Study	Country	Study design	Cases	Controls	Female, n (%)	Median age (range)
Colo2&3	USA	Case-control	87	122	94 (45)	65 (38-86)
DACHS	Germany	Case-control	2363	2200	1820 (40)	69 (33-99)
DALS	USA	Case-control	1116	1174	1028 (45)	64 (28-79)
HPFS	USA	Cohort	483	685	0 (0)	63 (48-82)
MEC	USA	Cohort	314	334	305 (47)	63 (45-76)
NHS	USA	Cohort	837	1422	2259 (100)	59 (44-69)
PHS	USA	Cohort	367	384	0 (0)	59 (40-85)
PLCO	USA	Cohort	434	683	437 (39)	65 (55-74)
PMH	USA	Case-control	228	99	327 (100)	63 (48-73)
VITAL	USA	Cohort	254	263	239 (46)	66 (50-76)
WHI	USA	Cohort	1462	1527	2989 (100)	66 (50-79)
Total			7945	8893	9498 (56)	65 (28-99)

Colo 2&3, a case-control study from the University of Hawai'i; DACHS, Darmkrebs: Chancen der Verhütung durch Screening Study; DALS, diet, activity and lifestyle study; HPFS, Health Professionals Follow-up Study; MEC, multiethnic cohort; NHS, Nurses' Health Study; PHS, Physicians' Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PMH, the hormones and colon cancer study; VITAL, Vitamins and Lifestyle Study; WHI, Women's Health Initiative.



Assessment of environmental and genetic variables

Demographics and environmental exposures were self-reported either at in-person interviews or via structured self-administered questionnaires, given each study's protocols. We applied a multistep, iterative data harmonisation procedure (online supplementary eMethod 1). In brief, variables were combined into a single dataset with common definition, standardised coding and permissible values. Quality-control checks were performed for variable ranges and coding logics. Outlying values were truncated to an established range for each variable.

Demographic and medical information included age, sex, height, education, first-degree CRC family history and history of endoscopy (colonoscopy/sigmoidoscopy). Age was defined as age at diagnosis for cases and age at selection for controls. Height was either measured or self-reported at baseline. Lifestyle variables included body mass index (BMI), smoking, physical activity, regular use of aspirin and non-aspirin NSAIDs, PMH use in women, diabetes, and dietary intakes. BMI was calculated based on weight and height at baseline (kg/m^2). Smoking was defined using two variables: (1) ever/never smokers and (2) pack-years of smoking among ever smokers. Sex-study-specific quartiles of pack-years were derived. Physical activity was defined as binary (yes if vigorous/moderate physical activity of <1 hour/week). Study-specific definitions of regular use of aspirin and non-aspirin that captured both duration and frequency were used.³⁸ PMH use in women was defined as current use at study baseline. History of diabetes was defined as diagnosis of diabetes at baseline. Dietary covariates were ascertained using food frequency questionnaires or diet histories, including intakes of alcohol, fruits, vegetables, dietary fibre, red meat, processed meat, total (dietary plus supplemental) calcium, total folate, and total energy. Sex study-specific quartiles were created for all dietary variables except alcohol. Alcohol was categorised by gram of alcohol intake per day: non-drinkers, 1–28 and >28 g/day.

DNA for genotyping was mostly obtained from blood samples, with some from buccal swabs. Details on genotyping, imputation, and quality controls have been described previously.^{39 40}

Building risk scores

Overall E-score

To capture the overall environmental risk profile for each participant, we calculated the E-score based on 17 environmental factors: BMI (kg/m^2), height (cm), smoking (ever/never, pack-years), alcohol consumption (0, 1–28 or >28 g/day), physical activity (yes/no), aspirin use (yes/no), other NSAID use (yes/no), PMH use (yes/no) in women, sex study-specific quartiles of dietary factors (red meat, processed meat, fruits, vegetables, fibre, total calcium and total folate), and diabetes (yes/no). We selected these factors based on known and plausible CRC risk factors previously reported in the literature.

We excluded participants with missing data on 5 or more of the 17 environmental risk factors ($n=1312$).

We then replaced missing values (missing proportion, 0%–16%) with the sex study-specific mean for all factors. We also performed multiple imputation in a sensitivity analysis. As we aimed to create a score that summarised an individual's overall environmental risk profile, the reference category for each factor was that associated with the lowest CRC risk in published studies (eg, NSAID use is coded as 0 vs NSAID non-user is coded as 1), so that the estimated weights for each risk factor represented an increase in CRC risk. To give variable weights optimal control for confounding by the intentions of endoscopy, weights were calculated in the subset of studies in which endoscopy was used for screening rather than diagnosis. Separately in men and women using this subset, we created a weighted risk score by (1) estimating the sex-specific coefficients using multivariable logistic regression that included all risk factors in one model, adjusting for age, study, screening, education, and total energy consumption (online supplementary eFigure 1); and (2) for each subject, multiplying his/her value by the corresponding regression coefficient for each variable, then summing across all variables.

Because all environmental variables were included simultaneously in the same logistic regression model, the resulting estimated weights for each variable accounted for the influence of other variables included in the E-score, as well as additional potential confounders (eg, age, study, screening, education, and total energy consumption). Weights estimated in our dataset for each variable were consistent with previously reported associations between known and potential risk factors and CRC risk.^{3–18} The resulting sum was then recoded as sex-study-specific quartiles based on cut points in controls and was modelled as an ordinal variable. The study-specific overall E-score was created using a similar approach, additionally adjusting for sex.

Modifiable and non-modifiable risk scores

We further developed a modifiable risk score, including BMI, physical activity, smoking, intakes of alcohol, processed meat, red meat, fruit, vegetables, fibre, total calcium and total folate, aspirin and NSAID use, PMH use in women, and diabetes. The non-modifiable risk score included height, CRC family history and common genetic predisposition. To capture common genetic predisposition, we calculated a genetic risk score combining estimated effects of 63 genome-wide association study (GWAS)-identified single-nucleotide polymorphisms (SNPs),^{33 35} from a multivariable logistic regression of all 63 SNPs, adjusting for age, sex, study, genotyping platform, and principal components of genetic ancestry. The full list of variables included as modifiable or non-modifiable risk factors is listed in online supplementary eTable 1.

Similar to the approach in developing the E-score, multivariable logistic regression was used to estimate sex-specific coefficients for all the variables in both modifiable and non-modifiable risk scores, adjusting for age,

study, education, screening, and total energy consumption. Sex-specific coefficients were then summed into two risk scores. The sex study-specific quartiles were calculated using cut points in controls for both risk scores.

Statistical analyses

All statistical analyses were conducted centrally on individual-level data. To estimate the association between E-score and CRC risk, we used multivariable logistic regression, adjusting for age, sex, study, education, screening and total energy consumption. In sensitivity analyses, we compared estimated associations from mean imputation⁴¹ and multiple imputation⁴² to address missing data.

We calculated absolute cumulative risks of 10 and 30 years for a 50-year-old individual within each quartile of E-score, as well as by low (<10%), medium (45%–55%), and high (>90%) E-scores. The details of estimating absolute risks of CRC have been described previously (online supplementary eMethod 2).^{35 43 44} Briefly, we used external age-specific population incidence rates among the white population from the Surveillance, Epidemiology, and End Results (SEER) between 1982 and 2014.⁴⁵ We then multiplied the external incidence rate with one minus sex-specific population attributable risk, which was estimated by taking the average of the inverse exponential of risk scores among cases.⁴⁴ We also accounted for competing risks from death in the absolute risk estimation, where the mortality rates were obtained from the National Center for Health Statistics. We obtained the 95% CIs of the 10-year absolute risk

estimates of CRC with 100 bootstrap samples. Sensitivity analyses were performed stratified by study design and history of screening.

In addition, we estimated the absolute risks of CRC of 10 and 30 years for individuals with different modifiable risk score quartiles, stratified by non-modifiable risk score quartiles. Relative excess risk due to interaction (RERI) was assessed to test for additive interaction between modifiable and non-modifiable risk scores.⁴⁶

In the secondary analysis, we evaluated the impact of individual risk factors that were statistically significant when building the E-score, stratified by quartiles of the non-modifiable risk score. Variables analysed included BMI (obese vs normal), pack-years of smoking (highest (Q4) vs lowest (Q1) quartiles), aspirin use (yes/no), other NSAID use (yes/no), total calcium intake (Q4 vs Q1) and PMH use (yes/no) in women.

All analyses were conducted using R V.3.4.1 (<http://www-r-project.org>) and SAS V.9.4.

RESULTS

Overall E-score and CRC risk

Higher E-scores were associated with increased CRC risk (OR_{quartile}, 1.33; 95% CI 1.29 to 1.37; [figure 1](#)). In the sensitivity analysis, the estimated association between E-score and CRC risk was identical when we used multiple imputation for missing data (OR_{quartile}, 1.33; 95% CI 1.18 to 1.50).

Based on the SEER data, the average risks of CRC of 10 and 30 years for a 50-year-old man were 0.68% and 4.1%,

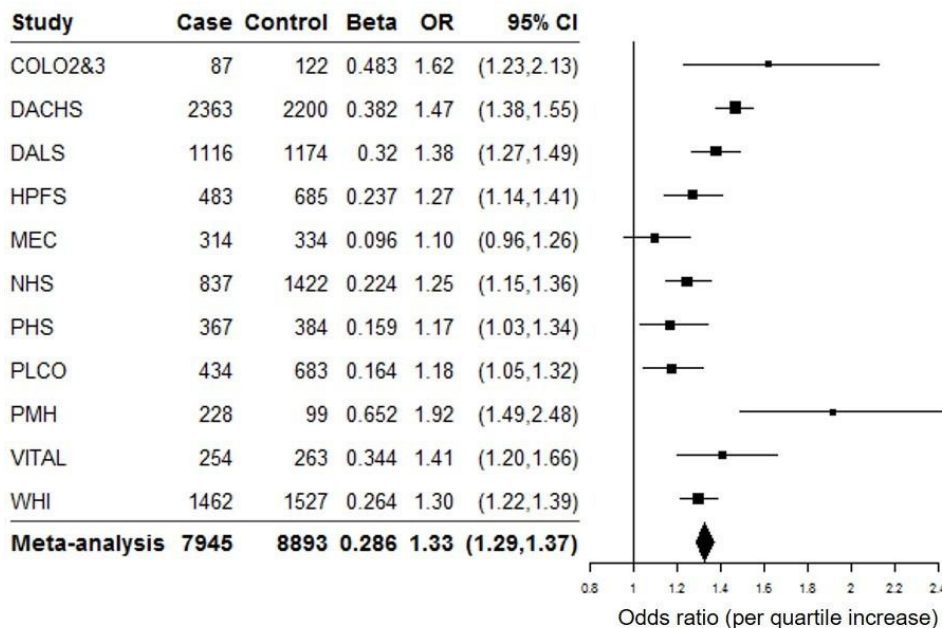


Figure 1 Associations of weighted study sex-specific quartiles of environmental risk scores and colorectal cancer risk among studies. P heterogeneity=0.0002; Adjusted for age, total energy consumption, history of screening, and education. Colo2&3, a case-control study from the University of Hawai'i; DACHS, Darmkrebs: Chancen der Verhütung durch Screening Study; DALS, diet, activity and lifestyle study; HPFS, Health Professionals Follow-up Study; MEC, multiethnic cohort; NHS, Nurses' Health Study; PHS, Physicians' Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PMH, the hormones and colon cancer study; VITAL, Vitamins and Lifestyle Study; WHI, Women's Health Initiative.

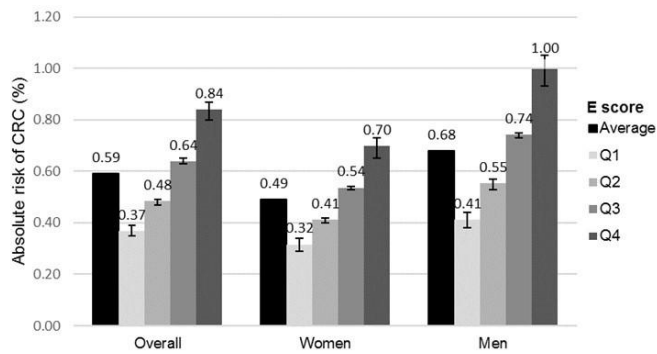
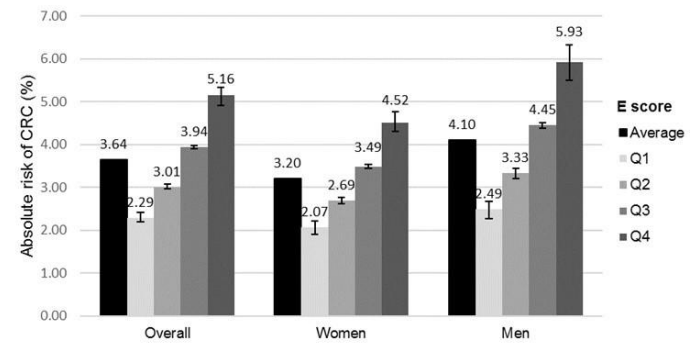
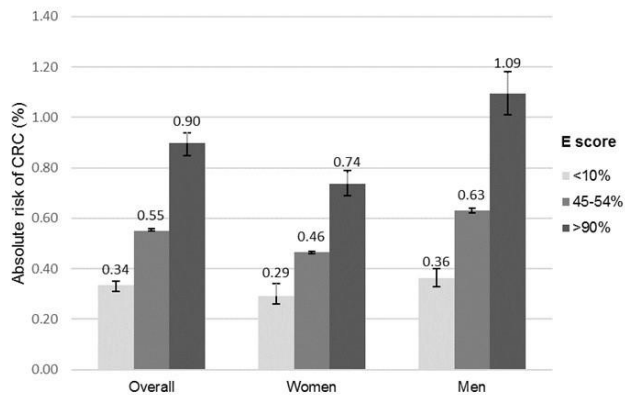
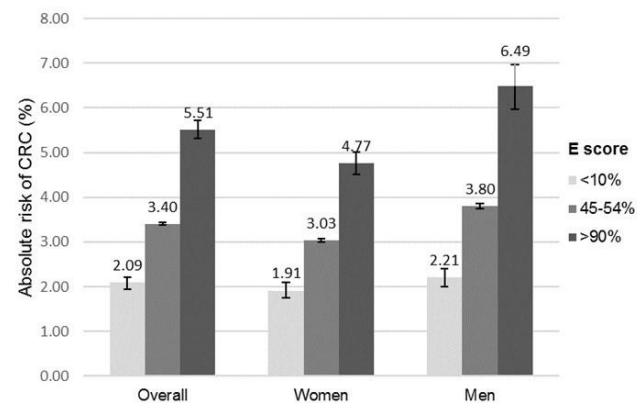
A 10-year absolute risk of colorectal cancer by environmental risk quartiles

B 30-year absolute risk of colorectal cancer by environmental risk quartiles

C 10-year absolute risk of colorectal cancer among participants with low, medium and high environmental risk scores

D 30-year absolute risk of colorectal cancer among participants with low, medium and high environmental risk scores


Figure 2 (A) 10-year and (B) 30-year absolute risk of CRC for a 50-year old individual by E-score. (A) E-score included body mass index (kg/m^2), height (cm), smoking (ever/never, pack-years), alcohol consumption (non-drinkers, 1–28 g/day, >28 g/day), physical activity (sedentary, yes/no), regular use of aspirin (yes/no), regular use of other non-steroidal anti-inflammatory drugs (yes/no), regular use of postmenopausal hormone in women (yes/no), sex-specific and study-specific quartiles of dietary factors (red meat, processed meat, fruits, vegetables, fibre, folate and calcium) and history of diabetes (yes/no). (B) Adjusted for age, study, total energy consumption, history of screening and education. CRC, colorectal cancer; E-score, environmental risk score.

respectively, and were 0.49% and 3.2% for a woman, respectively. Compared with these average-risk estimates, the absolute risks of CRC of both 10 and 30 years increased with higher E-score (figure 2 and online supplementary eTable 2). The 30-year absolute risk of CRC in Q4 of E-scores was 5.9% (95% CI 5.5% to 6.3%) among men and 4.5% (95% CI 4.3% to 4.8%) among women, compared with 2.5% (95% CI 2.3% to 2.7%) in Q1 of E-score among men and 2.1% (95% CI 1.9% to 2.2%) among women (figure 2B). The 30-year absolute risk of CRC among individuals within the highest E-score decile (>90%) was 2.5 times, and 2.9 times higher than among those within the lowest E-score decile (<10%) among men and women, respectively (figure 2D).

Modifiable risk scores and CRC risk by non-modifiable risk scores

We found a statistically significant additive interaction between modifiable and non-modifiable risks (RERI, 1.23; 95% CI 0.51 to 1.95; $p < 0.001$). As expected, the absolute risks of CRC of both 10 and 30 years increased with higher non-modifiable risk scores, and absolute risks increased

with higher modifiable risk scores within each quartile of non-modifiable risk score (figure 3 and online supplementary eFigure 2). The difference in absolute risk between Q4 and Q1 of the modifiable risk score was largest among those with the highest non-modifiable risk score. The trend was the similar for both sexes, although the absolute risks were higher in men. The 30-year absolute risk of CRC for men in Q1 of non-modifiable risks varied by 2.7% (from 4.1% to 1.4% comparing Q4 with Q1 of modifiable risk); however, the difference was more than double this among individuals in Q4 of non-modifiable risks (5.5%, from 8.9% to 3.4%; figure 3C). Similarly, the 30-year absolute risk of CRC decreased from 3.2% to 1.4% among women in Q1 of non-modifiable risk, compared with the change of 6.0% to 3.2% among women in Q4 of non-modifiable risk score (figure 3D).

Individual modifiable risk factors and CRC risk by non-modifiable risk scores

The estimated absolute risks of CRC showed similar trends across non-modifiable risk quartiles, for individual risk factors (table 2). The lowest absolute risks of CRC were

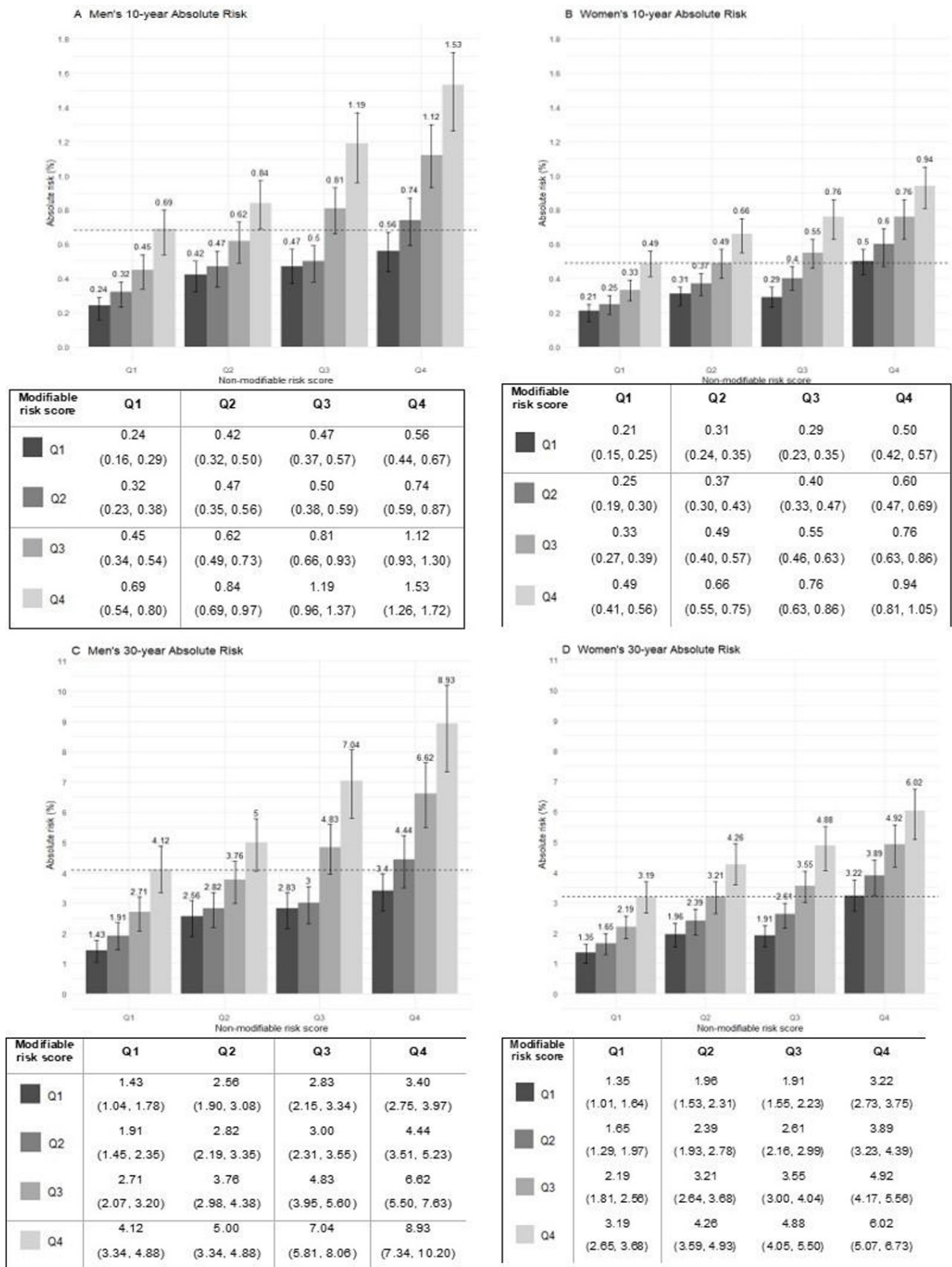


Figure 3 Distribution of absolute risk associated with modifiable risk score (A) stratified by non-modifiable risk score quartiles (B) in the USA. Dashed lines indicate the average absolute risks of CRC for a 50-year-old person: 0.68% and 0.49% for 10-year absolute risk in men (A) and women (B), and 4.1% and 3.2% for 30-year absolute risk in men (C) and women (D), respectively. (A) Modifiable risk score included body mass index, sedentary, smoking, pack-years of smoking, intakes of alcohol, fibre, calcium, folate, processed meat, red meat, fruit and vegetables, use of aspirin and non-aspirin non-steroidal anti-inflammatory drugs and postmenopausal hormone use among women and diabetes. (B) Non-modifiable risk score included age, sex, height, family history of CRC, and common genetic predisposition based on 63 genome-wide association study (GWAS)-identified single-nucleotide polymorphisms. CRC, colorectal cancer.

Table 2 Estimated absolute risks of colorectal cancer by individual modifiable risk factors stratified by non-modifiable risk quartiles

Non-modifiable risk*	BMI†		Pack-years of smoking		NSAID		Aspirin		PMH		Calcium	
	<25	≥30	Q1	Q4	Y	N	Y	N	Y	N	Q4	Q1
10-year absolute risk												
Q1 (lowest)	0.33	0.50	0.34	0.46	0.26	0.40	0.30	0.40	0.33	0.50	0.31	0.44
Q2	0.45	0.68	0.39	0.67	0.43	0.54	0.47	0.56	0.48	0.67	0.42	0.63
Q3	0.52	0.81	0.47	0.75	0.49	0.64	0.49	0.66	0.55	0.76	0.56	0.67
Q4 (highest)	0.72	1.07	0.69	0.96	0.74	0.88	0.69	0.91	0.75	1.07	0.71	0.99
30-year absolute risk												
Q1 (lowest)	2.06	3.08	2.13	2.85	1.62	2.50	1.90	2.52	2.04	3.09	1.95	2.73
Q2	2.82	4.22	2.42	4.15	2.69	3.37	2.90	3.45	2.95	4.11	2.64	3.88
Q3	3.25	4.97	2.91	4.60	3.06	3.93	3.02	4.09	3.43	4.70	3.46	4.12
Q4 (highest)	4.43	6.55	4.25	5.86	4.56	5.37	4.28	5.56	4.65	6.51	4.36	6.03

*Non-modifiable risk score included age, sex, height, family history of CRC, and common genetic predisposition based on 63 genome-wide association study (GWAS)-identified single-nucleotide polymorphisms.

†BMI was calculated as weight in kilogram divided by height in metre squared.

BMI, body mass index; N, non-user; PMH, postmenopausal hormone; Y, regular user.

among regular users of NSAIDs and aspirin. However, changing single risk factors did not reduce the long-term absolute risks of CRC to the same extent as modifying multiple risk factors.

DISCUSSION

Incorporating information on most known risk factors of CRC, we showed that higher E-score was associated with higher risk of CRC and that the absolute risks of CRC varied largely by E-score among a population of European ancestry. We also showed that the absolute risk of developing CRC was substantially lower with lower modifiable risk scores. This difference may be especially pronounced for men and for those who were at higher risk due to non-modifiable risks.

Our results on the overall E-score and CRC risks reinforce published data and guidelines regarding the effect of lifestyle patterns.^{19 47 48} The World Cancer Research Fund estimated that approximately 47% of CRC in the USA can be attributed to lifestyle factors, including low dietary fibre intake, high red/processed meat intake, obesity, lack of physical activity, and alcohol consumption.⁴⁷ Similarly, in our study, we observed that a large proportion of CRC in both men and women may be preventable by changing lifestyle factors. The ACS guidelines for cancer prevention also suggest maintaining a healthy weight, adopting a physically active lifestyle, and consuming a healthy diet with an emphasis on plant foods.¹⁹ The Women's Health Initiative Observational Study reported that CRC risk was statistically significantly lower among women with higher adherence scores to the ACS guidelines,⁴⁸ and this association was similarly seen in a multicentre prospective cohort.⁴⁹ Other cohort studies that evaluated dietary patterns found that healthier dietary patterns were associated with lower CRC risk.^{50–52} In the Nurses' Health Study and Health Professionals Follow-up Study, those with a healthy lifestyle pattern

had a substantially lower population attributable risk of CRC than the US population.⁵³ Furthermore, we found that the estimated absolute risks were similar among those with and without history of screening endoscopy (online supplementary eTable 3). This is consistent with a previous study that also found no differences in association between healthy lifestyles and CRC risks among subgroups by history of colonoscopy.²¹

These data, taken together, highlight that lifestyle modification may markedly reduce CRC risk. However, previous studies have evaluated only a limited number of risk factors, whereas others have not been considered, including use of aspirin or NSAIDs, intakes of fibre, folate and calcium, diabetes and PMH use in women. In this study, we created a comprehensive E-score by estimating the weights of 17 risk factors of CRC simultaneously, which helped to not only account for relationships between risk factors but also provide a more comprehensive estimate of the overall environmental impact on CRC risk.

Furthermore, our data are consistent with the interpretation that a shift in modifiable risk profiles can appreciably reduce the long-term absolute risks of developing CRC. Notably, our results suggest the elevated risks of CRC for those in the highest quartile of non-modifiable risk scores can potentially be reduced to the average population risk through lifestyle modification. This may have important implications for targeted risk communication, risk management, and behavioural interventions for CRC prevention. For example, an individual with positive family history or higher genetic risk of CRC could potentially achieve population average risk by incorporating a combination of lifestyles, including maintaining normal BMI, quitting smoking, and reducing red and processed meat intake. Such lifestyle-based risk management strategy could provide more cost-effective ways for individuals with higher 'baseline' risk of CRC in both sexes, but especially in men.



Previous studies have shown promise with using pharmacological agents in targeted populations. The Breast Cancer Prevention Trial showed that tamoxifen use reduced the incidence of invasive breast cancer by 49% in high-risk women.⁵⁴ Long-term use of aspirin was also shown to reduce CRC incidence by 24% among individuals with higher risk of cardiovascular diseases.¹⁵ Individual agents, such as aspirin and NSAIDs, may have a marked effect in specific populations; however, our results further suggest that changes in individual exposures or behaviours might not translate to a reduced long-term CRC risk to the same extent as modifying multiple risk factors. This is useful information because risk factors may not be equally modifiable for each individual, and prevention decisions need to be made based on more individualised risk–benefit evaluation. For instance, an individual who experiences gastrointestinal bleeding from using aspirin can still reduce the risk of developing CRC by quitting smoking, increasing physical activity, and adopting a healthier diet. Taking into consideration the whole modifiable risk profile not only makes greater sense for the prevention of multiple chronic diseases but also helps to provide more opportunities for primary prevention of CRC, particularly among those at higher risk of CRC due to their non-modifiable risk profile.

Similar to the improved model for risk stratification for breast cancer prevention,³⁴ our model provided a practical framework to use current existing genetic and environmental data from large cancer consortium for guiding public health strategies. Our study is among the first to estimate the combined association of a wide range of established and plausible environmental CRC factors while accounting for non-modifiable risks. We estimated the E-score from multiple population-based studies, and estimated absolute risks based on information from a nationally representative database. Risk factors were carefully harmonised across participating studies, enabling us to simultaneously incorporate a comprehensive set of environmental risk factors. In addition to environmental data, genetic data were also available from all study participants, allowing us to incorporate a common genetic predisposition in estimating non-modifiable risk profiles. Our study is the largest to date and includes numerous studies derived from different study populations. This helped yield precise estimates of risk factor associations and ensured the robustness of observed findings.

Despite these strengths, there are some limitations. First, environmental factors were self-reported, where measurement errors might have attenuated associations. However, self-reported lifestyle and diet have been shown to have modest to high accuracy in prior studies.^{55 56} Second, the sex study-specific mean imputation approach for missing data reduced the variance of distributions and could result in biased estimates. However, sensitivity analysis using multiple imputation did not show appreciable differences. We also included both case–control and cohort studies. Environmental factors are assessed after cancer diagnosis in case–control studies and are

susceptible to differential recall bias. Stratifying by study design, we found the differences in absolute risks across E-score quartiles were larger in case–control studies (online supplementary eFigure 3). However, the absolute risks increased with higher E-scores in a similar manner between case–control and cohort studies. We chose to maximise our sample size (and thus precision) for developing the risk scores, instead of splitting our data into a discovery and validation set. We do not anticipate overfitting (or bias due to estimation) appreciably altered our findings given (1) our large sample size, which has been shown to greatly reduce the overestimation of regression coefficients,^{57 58} and (2) we did not perform variable selection but rather included only published CRC risk factors in our models. Moreover, our estimates of effect sizes for each known risk factor were comparable to those previously reported in the literature. We also used external incidence rates to estimate absolute risks with bootstrapping methods to assess uncertainty. Furthermore, our study population is generally older, and most environmental factors were assessed within 2–10 years prior to baseline. Our results may need further evaluation on younger populations. Lastly, we included only individuals with European ancestry. The observed association may differ in other ethnicity groups.

In conclusion, we demonstrated the value of comprehensive assessment of environmental risk factors to estimate the relative and absolute risks of CRC. Our results suggest that risk factor modification may reduce long-term CRC risk, particularly among individuals at higher baseline risk due to non-modifiable factors. These results provide key insights to help inform targeted CRC prevention guidelines and perhaps allow better practice of screening (using risk scores rather than just age) in the general population. Additional studies for further validation would strengthen our findings.

Author affiliations

¹Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

²Epidemiology, University of Washington, Seattle, Washington, USA

³Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA

⁴Epidemiology, University of Michigan, Ann Arbor, Michigan, USA

⁵Channing Division of Network Medicine, Harvard Medical School, Boston, Massachusetts, USA

⁶Nutrition, Harvard University T H Chan School of Public Health, Boston, Massachusetts, USA

⁷Epidemiology, Harvard University T H Chan School of Public Health, Boston, Massachusetts, USA

⁸Nuffield Department of Population Health, University of Oxford, Oxford, UK

⁹Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany

¹⁰Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

¹¹Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

¹²Harvard Medical School, Boston, Massachusetts, USA

¹³Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

¹⁴Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France

¹⁵Epidemiology, New York University School of Medicine, New York, New York, USA

¹⁶Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

¹⁷Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

¹⁸Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah, USA

¹⁹Office of Disease Prevention, National Institutes of Health, Bethesda, Maryland, USA

²⁰Epidemiology Program, University of Hawai'i Cancer Center, Honolulu, Hawaii, USA

Twitter Xiaoliang Wang @_zhanly

Acknowledgements Darmkrebs: chancen der Verhütung durch Screening Study: we thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benschaid, Muhabbet Celik and Ursula Eilber for excellent technical assistance. Harvard cohorts (Health Professionals Follow-up Study (HPFS), Nurses' Health Study (NHS) and Physicians' Health Study (PHS)): The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. Harvard: We thank the participants and staff of the HPFS, NHS and PHS for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA and WY. The authors assume full responsibility for analyses and interpretation of these data. Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO): The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible. The Hormones and Colon Cancer Study (PMH): The authors thank the study participants and staff of the Hormones and Colon Cancer study. Women's Health Initiative (WHI): The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found online (<http://www.whi.org/researchers/DocumentsWritePaper/WHIInvestigatorShortList.pdf>).

Contributors MD and JG planned the study proposal; KOC, JJ and YL conducted the analyses; XW and MD led the study, manuscript preparation and manuscript revision; EW, LH and UP oversaw the study plan and manuscript preparation and revision; MS, DH, MH, YL, SB, HB, ATC, JC-C, JG, MJG, TAH, RBH, AJ, PN, RS, MLS, AV, JDP, LLM, EG and EW supported the proposal review, data curation, manuscript submission and revision.

Funding Genetics and Epidemiology of Colorectal Cancer Consortium: National Cancer Institute, National Institutes of Health, US Department of Health and Human Services (R01 CA059045, U01 CA164930 and R01 201407). This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. COLO2&3: National Institutes of Health (R01 CA60987). DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B). DALs: National Institutes of Health (R01 CA48998 to M. L. Slattery). Harvard cohorts (HPFS, NHS and PHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003) and PHS by the National Institutes of Health (R01 CA042182). MEC: National Institutes of Health (R37 CA54281, P01 CA033619 and R01 CA063464). PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, PMH: National Institutes of Health (R01 CA076366 to PN). Vitamins and Lifestyle Study: National Institutes of Health (K05 CA154337). Women's Health Initiative (WHI): The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C.

Competing interests MJG: where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do

not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Patient consent for publication Not required.

Ethics approval Informed consent was given by all participants. Studies were approved by their respective institutional review boards.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

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

ORCID iD

Xiaoliang Wang <http://orcid.org/0000-0001-5184-2935>

REFERENCES

- 1 Arnold M, Sierra MS, Laversanne M, *et al*. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017;66:683–91.
- 2 Siegel RL, Miller KD, Fedewa SA, *et al*. Colorectal cancer statistics, 2017. *CA Cancer J Clin* 2017;67:177–93.
- 3 Thrift AP, Gong J, Peters U, *et al*. Mendelian randomization study of height and risk of colorectal cancer. *Int J Epidemiol* 2015;44:662–72.
- 4 Renehan AG, Tyson M, Egger M, *et al*. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371:569–78.
- 5 Botteri E, Iodice S, Bagnardi V, *et al*. Smoking and colorectal cancer: a meta-analysis. *JAMA* 2008;300:2765–78.
- 6 Liang PS, Chen T-Y, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *Int J Cancer* 2009;124:2406–15.
- 7 Walter V, Jansen L, Hoffmeister M, *et al*. Smoking and survival of colorectal cancer patients: systematic review and meta-analysis. *Ann Oncol* 2014;25:1517–25.
- 8 Fedirko V, Tramacere I, Bagnardi V, *et al*. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol* 2011;22:1958–72.
- 9 Moskal A, Norat T, Ferrari P, *et al*. Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. *Int J Cancer* 2007;120:664–71.
- 10 Bouvard V, Loomis D, Guyton KZ, *et al*. Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* 2015;16:1599–600.
- 11 Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* 2006;119:2657–64.
- 12 Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. *J Natl Cancer Inst* 2005;97:1679–87.
- 13 Harriss DJ, Atkinson G, Batterham A, *et al*. Lifestyle factors and colorectal cancer risk (2): a systematic review and meta-analysis of associations with leisure-time physical activity. *Colorectal Dis* 2009;11:689–701.
- 14 Samad AKA, Taylor RS, Marshall T, *et al*. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* 2005;7:204–13.
- 15 Rothwell PM, Wilson M, Elwin C-E, *et al*. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *Lancet* 2010;376:1741–50.
- 16 Lin KJ, Cheung WY, Lai JY-C, *et al*. The effect of estrogen vs. combined estrogen-progestogen therapy on the risk of colorectal cancer. *Int J Cancer* 2012;130:419–30.
- 17 Huncharek M, Muscat J, Kupelnick B. Colorectal cancer risk and dietary intake of calcium, vitamin D, and dairy products: a meta-analysis of 26,335 cases from 60 observational studies. *Nutr Cancer* 2009;61:47–69.
- 18 Park Y, Hunter DJ, Spiegelman D, *et al*. Dietary fiber intake and risk of colorectal cancer: a pooled analysis of prospective cohort studies. *JAMA* 2005;294:2849–57.
- 19 Kushi LH, Doyle C, McCullough M, *et al*. American cancer Society guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J Clin* 2012;62:30–67.

- 20 Chubak J, Whitlock EP, Williams SB, *et al.* Aspirin for the prevention of cancer incidence and mortality: systematic evidence reviews for the U.S. preventive services Task force. *Ann Intern Med* 2016;164:814–25.
- 21 Carr PR, Weigl K, Jansen L, *et al.* Healthy lifestyle factors associated with lower risk of colorectal cancer irrespective of genetic risk. *Gastroenterology* 2018;155:1805–15.
- 22 Slattery ML, John EM, Torres-Mejia G, *et al.* Genetic variation in genes involved in hormones, inflammation and energetic factors and breast cancer risk in an admixed population. *Carcinogenesis* 2012;33:1512–21.
- 23 Al-Tassan N, Chmiel NH, Maynard J, *et al.* Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 2002;30:227–32.
- 24 Houlston RS, Cheadle J, Dobbins SE, *et al.* Meta-Analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 2010;42:973–7.
- 25 Hutter CM, Chang-Claude J, Slattery ML, *et al.* Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res* 2012;72:2036–44.
- 26 Peters U, Bien S, Zubair N. Genetic architecture of colorectal cancer. *Gut* 2015;64:1623–36.
- 27 Schumacher FR, Schmit SL, Jiao S, *et al.* Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun* 2015;6:7138.
- 28 Tenesa A, Farrington SM, Prendergast JGD, *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 2008;40:631–7.
- 29 Tomlinson IPM, Carvajal-Carmona LG, Dobbins SE, *et al.* Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet* 2011;7:e1002105.
- 30 Zanke BW, Greenwood CMT, Rangrej J, *et al.* Genome-Wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;39:989–94.
- 31 Zeng C, Matsuda K, Jia W-H, *et al.* Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology* 2016;150:1633–45.
- 32 Zhang B, Jia W-H, Matsuda K, *et al.* Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* 2014;46:533–42.
- 33 Huyghe JR, Bien SA, Harrison TA, *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;51:76–87.
- 34 Maas P, Barrdahl M, Joshi AD, *et al.* Breast cancer risk from modifiable and Nonmodifiable risk factors among white women in the United States. *JAMA Oncol* 2016;2:1295–302.
- 35 Jeon J, Du M, Schoen RE, *et al.* Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology* 2018;154:e19:2152–64.
- 36 Gong J, Hutter CM, Newcomb PA, *et al.* Genome-wide interaction analyses between genetic variants and alcohol consumption and smoking for risk of colorectal cancer. *PLoS Genet* 2016;12:e1006296.
- 37 Price AL, Patterson NJ, Plenge RM, *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
- 38 Wang X, Chan AT, Slattery ML, *et al.* Influence of smoking, body mass index, and other factors on the preventive effect of nonsteroidal anti-inflammatory drugs on colorectal cancer risk. *Cancer Res* 2018;78:4790–9.
- 39 Peters U, Hutter CM, Hsu L, *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet* 2012;131:217–34.
- 40 Peters U, Jiao S, Schumacher FR, *et al.* Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology* 2013;144:799–807.
- 41 Burton A, Altman DG. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer* 2004;91:4–8.
- 42 Honaker J, King G, Blackwell M. Amelia II: a program for missing data. *J Stat Softw* 2011;45:1–47.
- 43 Hsu L, Jeon J, Brenner H, *et al.* A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology* 2015;148:e14:1330–9.
- 44 Gail MH, Brinton LA, Byar DP, *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
- 45 SEER*Explorer: an interactive website for SEER cancer statistics surveillance research program, National cancer Institute.
- 46 Lundberg M, Fredlund P, Hallqvist J, *et al.* A SAS program calculating three measures of interaction with confidence intervals. *Epidemiology* 1996;7:655–6.
- 47 World Cancer Research Fund International/American Institute for Cancer Research. *Cancer preventability estimates*, 2019.
- 48 Thomson CA, McCullough ML, Wertheim BC, *et al.* Nutrition and physical activity cancer prevention guidelines, cancer risk, and mortality in the women's health Initiative. *Cancer Prev Res* 2014;7:42–53.
- 49 Romaguera D, Vergnaud A-C, Peeters PH, *et al.* Is concordance with world cancer research Fund/American Institute for cancer research guidelines for cancer prevention related to subsequent risk of cancer? results from the EPIC study. *Am J Clin Nutr* 2012;96:150–63.
- 50 Jones-McLean E, Hu J, Greene-Finestone LS, *et al.* A DASH dietary pattern and the risk of colorectal cancer in Canadian adults. *Health Promot Chronic Dis Prev Can* 2015;35:12–20.
- 51 Reedy J, Wirfalt E, Flood A, *et al.* Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study. *Am J Epidemiol* 2010;171:479–87.
- 52 Slattery ML, Boucher KM, Caan BJ, *et al.* Eating patterns and risk of colon cancer. *Am J Epidemiol* 1998;148:4–16.
- 53 Song M, Giovannucci E. Preventable incidence and mortality of carcinoma associated with lifestyle factors among white adults in the United States. *JAMA Oncol* 2016;2:1154–61.
- 54 Fisher B, Costantino JP, Wickerham DL, *et al.* Tamoxifen for the prevention of breast cancer: current status of the National surgical adjuvant breast and bowel project P-1 study. *J Natl Cancer Inst* 2005;97:1652–62.
- 55 Hu FB, Satija A, Rimm EB, *et al.* Diet assessment methods in the nurses' health studies and contribution to evidence-based nutritional policies and guidelines. *Am J Public Health* 2016;106:1567–72.
- 56 Wolf AM, Hunter DJ, Colditz GA, *et al.* Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol* 1994;23:991–9.
- 57 Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in genetic association studies. *Genet Epidemiol* 2009;33:453–62.
- 58 Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in quantitative-trait association studies. *Genet Epidemiol* 2011;35:133–8.