Recalibrating the Genetics and Epidemiology of Colorectal Cancer Consortium Environmental Risk Score for Use in US Veterans

April R. Williams[1], Thomas S. Redding IV[2], Brian A. Sullivan[2,3], Robin N. Baidya[4], Belinda Ear[2], Kelly Cho[5,6,7], Kerry L. Ivey[5,6,7], Christina D. Williams[2,3], Jason A. Dominitz[8,9], David Lieberman[10,11], and Elizabeth R. Hauser[2,3]; on behalf of the VA Million Veteran Program

ABSTRACT

Background: Risk for colorectal cancer may accumulate through multiple environmental factors. Understanding their effects, along with genetics, age, and family history, could allow improvements in clinical decisions for screening protocols. We aimed to extend the previous work by recalibrating an environmental risk score (e-Score) for colorectal cancer among a sample of US veteran participants of the Million Veteran Program.

Methods: Demographic, lifestyle, and colorectal cancer data from 2011 to 2022 were abstracted from survey responses and health records of 227,504 male Million Veteran Program participants. Weighting for each environmental factor's effect size was recalculated using Veterans Affairs training data to create a recalibrated e-Score. This recalibrated score was compared with the original weighted e-Score in a validation sample of 113,752 (*n* cases = 590). Nested multiple logistic regression models tested associations between quintiles for recalibrated and original e-Scores. Likelihood ratio tests were used to compare model performance.

Results: Age ($P < 0.0001$), education ($P < 0.0001$), diabetes ($P < 0.0001$), physical activity ($P < 0.0001$), smoking ($P < 0.0001$), NSAID use ($P < 0.0001$), calcium ($P = 0.015$), folate ($P = 0.020$), and fruit consumption ($P = 0.019$) were significantly different between colorectal cancer case and control groups. In the validation sample, the recalibrated e-Score model significantly improved the base model performance ($P < 0.001$), but the original e-Score model did not ($P = 0.07$). The recalibrated e-Score model quintile 5 was associated with significantly higher odds for colorectal cancer compared with quintile 1 (Q5 vs. Q1: 1.79; 95% CI, 1.38–2.33).

Conclusions: Multiple environmental factors and the recalibrated e-Score quintiles were significantly associated with colorectal cancer cases.

Impact: A recalibrated, veteran-specific e-Score could be used to help personalize colorectal cancer screening and prevention strategies.

Introduction

Colorectal cancer is the second leading cause for cancer-related mortality in the United States (1). Colorectal cancer risk may be influenced, in part, by the cumulative effects of sociodemographic factors and lifestyle collectively referred to as environmental factors. Understanding the effects of environmental factors on colorectal cancer risk could be useful to inform clinical decision-making to tailor screening and surveillance protocols, which traditionally only account for age, family history (2, 3), and colonoscopy findings (4). An understanding of these effects could support clinicians having meaningful discussions during clinical visits that may motivate their patients to modify relevant lifestyle behaviors and thereby lower their colorectal cancer risk.

In the work conducted by the Colorectal Transdisciplinary Study and Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Jeon and colleagues (5) found that a sex-specific environmental risk score (e-Score) comprising a weighted sum of lifestyle and sociodemographic factors combined with a polygenic risk score was more predictive of colorectal cancer risk than family history alone. Their findings also suggested that a comprehensive assessment could provide a risk-adjusted age for screening initiation. Data were harmonized across 14 cohort studies comprising an overall homogeneous sample of participants. The socioeconomic and lifestyle factors of participants in the GECCO studies are not representative of US veterans that receive care at the Veterans Health Administration—a unique population of individuals at substantially elevated risk of chronic diseases, including colorectal cancer (6). It is accepted that risk scores perform inconsistently across different populations (7). Thus, we sought to "recalibrate" the GECCO e-Score to help us understand its use in cancer prevention studies for veterans.

[1] U.S. Department of Veterans Affairs Million Veteran Program Coordinating Center, Boston, Massachusetts. [2] Cooperative Studies Program Epidemiology Center-Durham, Durham VA Health Care System, Durham, North Carolina. [3] Duke University, Durham, North Carolina. [4] National Oncology Program, U.S. Department of Veterans Affairs, Washington, District of Columbia. [5] Massachusetts Veterans Epidemiology and Research Information Center (MAVERIC) and the VA Million Veteran Program, Boston VA Healthcare System, Boston, Massachusetts. [6] Division of Aging, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. [7] Department of Medicine, Harvard Medical School, Boston, Massachusetts. [8] National Gastroenterology and Hepatology Program, Veterans Health Administration Washington, Washington, District of Columbia. [9] University of Washington School of Medicine, Seattle, Washington. [10] VA Portland Health Care System, Portland, Oregon. [11] Oregon Health & Science University, Portland, Oregon.

Membership of the VA Million Veteran Program is provided in Supplement S1.

Corresponding Author: April R. Williams, U.S. Department of Veterans Affairs Million Veteran Program Coordinating Center, 2 Avenue de Lafayette, Boston, MA 02111. E-mail: April.Williams9@VA.gov

Cancer Epidemiol Biomarkers Prev 2024;33:1456–64

doi: 10.1158/1055-9965.EPI-24-0791

Although genetic testing for colorectal cancer risk has also become a more commonplace for those with family history, its utility is limited by our imperfect understanding of the interactions between environmental factors and genetic expression. Screening colonoscopy is an ideal time to assess risk because of both genetics and lifestyle/environment. With the GECCO work as a foundation, lifestyle data from a pilot sample (n = 2,846) of male US veteran participants of the Cooperative Studies Program #380 (8) were used to construct an e-Score weighted according to the GECCO study findings. This e-Score was tested for associations with colonoscopy findings of advanced neoplasia (AN; n cases = 405), colorectal cancer precursor lesions with variable neoplastic potential. The results of this pilot study did not detect a significant association between AN and the e-Score, possibly because of low statistical power, potential differences in population demographics, environmental exposures, risk factor profiles, and other clinical factors of AN compared with colorectal cancer. Furthermore, there are known disparities for colorectal cancer across race with unclear mechanisms (9), but both GECCO and the Cooperative Studies Program #380 study participants were nearly all of European ancestry (10).

Using environmental and lifestyle survey and electronic health record data collected from Veterans Affairs (VA) Million Veteran Program (MVP) participants, the objectives of the current study were to test for associations between colorectal cancer diagnoses and composite e-Scores among US veterans controlling for age, family history, and race using (i) an e-Score weighted according to the GECCO findings and (ii) a recalibrated (11) veteran-specific e-Score weighted using data from the US veteran MVP sample. The hypothesis tested in this study is that higher risk e-Scores compared with lower risk e-Scores are significantly associated with higher odds for colorectal cancer cases.

Materials and Methods

Study design and population

In 2011, the VA's Office of Research and Development launched the MVP (12). MVP is an observational cohort study and biobank of US veterans used by researchers to study how genes, lifestyle, military experiences, and exposures affect health and wellness among a diverse cohort of US veterans. By December 2022, there were 913,319 participants enrolled. The program includes participant surveys designed to obtain individual self-reported social and lifestyle information to use alongside participants' genomic data for clinical research (13). As this research project is a data-only study using previously collected data from the MVP Central Research Database, it uses a Waiver of Health Insurance Portability and Accountability Act Authorization.

Data used in this analysis were collected between January 2011 and December 2022. Every participant completed the MVP baseline and lifestyle surveys, in which the lifestyle survey completion date is considered the index date for the participant in the study. Survey items included questions about family history of colorectal cancer, education level, smoking status, alcohol use, dietary habits, physical activity, and medication use (13) described in detail below. Additional data used to ascertain colonoscopy history and colorectal cancer diagnosis were obtained from participants' health records relative to the index date through December 2022. Excluded MVP participants had a colorectal cancer diagnosis identified in their VA health records prior to index date, a diagnosis of inflammatory bowel disease or a hereditary colorectal cancer syndrome, or incomplete baseline and lifestyle survey data necessary to construct the e-Score. As participants were not required to fully respond to either survey, we have missingness for the study variables that accumulated to approximately 50%. Given the large sample size and rare outcome, complete cases were used in the study. The final sample comprised participants with complete data to calculate the e-Score (n = 227,504). **Figure 1** depicts the sample selection.

Measures

e-Score

Factors used to construct the e-Score were the same as those used in the GECCO study described by Jeon and colleagues (5); this was chosen so that a direct comparative analysis could be conducted between their published weights and our recalculated weighting for each lifestyle and environmental variable. These include body mass index (BMI; kg/m^2), height (cm), prior type 2 diabetes diagnosis (yes = 1; no = 0), regular use of aspirin (yes = 0; no = 1), regular use of NSAID (yes = 0; no = 1), educational attainment (<high school = "category 1"; high school graduate = "category 2"; some college or technical school = "category 3"; college or graduate degree = "category 4"), physical activity level that meets the recommendations of the American Heart Association's Life's Simple Seven (yes = 0; no = 1; ref. 14), ever-smoked (yes = 1; no = 0), smoking pack-years (if ever-smoked = "yes", assigned a quartile from sample distribution: 1,2, 3, or 4), alcohol use (<1 g/day = 1; 1–28 g/day = 0; >28 g/day = 2), and dietary intake: fiber (g/day), calcium (mg/day), folate (μg/day), processed meat (servings/day), red meat (servings/day), fruit (servings/day), vegetable (servings/day; all assigned a quartile from sample distribution: 0, 1, 2, or 3 in the order of increasing risk); and total energy (scaled by dividing by the standard error 1.06).

Covariates

In all main analyses, covariates included self-reported family history of colorectal cancer among siblings, parents, or grandparents, age at index date, and race (categorized as White, Black, or unknown/underrepresented/multiple races). Race is a social construct that was either self-reported in the MVP baseline survey or the most common recorded race found in the veterans' health records, if conflicting with the self-report. For additional sensitivity analyses, colonoscopy history prior to index date was derived from procedure codes in health records, and history of other cancers was identified using data from the VA Central Cancer Registry.

Colorectal cancer diagnosis

The primary outcome is a diagnosis of colorectal cancer (yes/no) identified using either of two ascertainment approaches previously described (15). The first is a published list of International Classification of Diseases (ICD) 9 and ICD10 codes used by VA Cooperative Studies Program Epidemiology Analytics Resource to identify colorectal cancer diagnoses for population summaries of colorectal cancer incidence and prevalence rates (https://www.vacsp.research.va.gov/CSPEC/Studies/CSPEAR/Main.asp). Colorectal cancer was ascertained if at least one inpatient or at least two outpatient health records contained relevant colorectal cancer diagnosis ICD9/10 codes. The second uses data stored in the Veterans Health Administration Corporate Data Warehouse from the Veterans Affairs Central Cancer Registry, a population-based cancer registry of VA cancer cases (https://www.data.va.gov/dataset/Veterans-Affairs-Central-Cancer-Registry-VACCR-/jvmd-8fgj). Colorectal cancer cases were ascertained by colon-specific tumor sites. Cases ascertained as colorectal cancer that had characteristics not associated with colorectal cancer such as certain tumor sites (e.g., appendiceal or anal canal) or histology (e.g., neuroendocrine) were not considered colorectal cancer cases.

Statistical analyses

Descriptive statistics

MVP participant demographics and relevant health factors were assessed for distribution, missingness, and outliers. Subsequent analyses were restricted to participants with complete data needed to construct the e-Score. Descriptive statistics included frequencies and proportions for categorical variables and means, SDs, medians, and ranges for the continuous variables, stratified by controls and colorectal cancer cases. Bivariate analyses for each demographic, clinical indicator, and colorectal cancer risk factor considered in this study were conducted to assess differences between the control and colorectal cancer case groups using either χ^2 or Kruskal–Wallis as appropriate.

e-Score derivation and calibration

The independent 3variable of interest in this analysis is the sex-specific e-Score among the male sample, which is a weighted sum of risk factors, in which lower e-Scores suggest lower cumulative risk for colorectal cancer. e-Scores were not calculated for the female participants in the study because of insufficient colorectal cancer cases among females in the MVP cohort. However, a description of the female MVP participants with complete data needed to calculate an e-Score is provided in Supplementary Table S1.

Two sets of e-Scores were calculated using the male sample ($N = 227{,}504$) by applying (i) GECCO study–derived weights (GECCO e-Score) and (ii) MVP recalibrated weights (recalibrated e-Score). To calculate the GECCO and recalibrated e-Scores, the MVP male sample was first split randomly into two halves to produce training and validation samples (each $n = 113{,}752$; colorectal cancer cases $n = 590$). To accomplish this, the SAS procedure SURVEYSELECT used simple random sampling with strata based on the colorectal cancer variable such that an equal distribution of cases and controls exist in each sample. The published GECCO weights were applied to respective risk factors and summed to calculate the GECCO e-Score among the male validation sample. For the recalibrated e-Score, a multiple logistic regression model used the training sample with colorectal cancer as the outcome and all the risk factors as independent variables controlling for age and family history of colorectal cancer to produce parameter estimates. To calculate the recalibrated e-Score, parameter estimates from the training model were applied as weights to respective risk factors among the validation sample and summed. See Supplementary Table S2 for a table of the parameter estimates from the GECCO study and the recalibrated training model. Both the GECCO and recalibrated e-Scores were standardized to percentages to be used as continuous variables, and participants were categorized into quintiles for use in statistical models. Higher e-Scores are expected to indicate higher risk for colorectal cancer.

Both the GECCO and recalibrated e-Score percentages and quintiles were assessed for distribution by control and colorectal cancer case groups.

Statistical models

Nested logistic regression models among the all-male validation sample were created using the SAS (SAS Institute, RRID: SCR_008567) procedure LOGISTIC with the outcome of colorectal cancer case yes/no. The base model used age, race, and family history of colorectal cancer as the independent variables. Two separate expanded models included the GECCO e-Score quintiles and the recalibrated e-Score quintiles. Adjusted ORs and 95% confidence intervals (CI) are presented for all models. Statistical significance of the test that the OR = 1 was defined as a Wald test P value < 0.05. The likelihood ratio comparison test (LRT; ref. 16) was used to compare fit of the GECCO and recalibrated e-Score expanded models with the base model. Model fit statistics Akaike information criterion and AUC are also presented for all nested models to show model fit comparisons.

Sensitivity analyses were performed comparing additional models with the recalibrated e-Score–expanded models to evaluate robustness to the assumption of population homogeneity. We fit a model that incorporated an interaction term between race and the recalibrated e-Score and the interaction term was not significant. Additional sensitivity analyses considered a full validation sample model and the following categorical variables for cohort definition for additional sensitivity analyses: (i) age 50+; (ii) Black race; (iii) at least one colonoscopy prior to index date; and (iv) absence of any type of cancer prior to index date. The LRT (16) was used to compare fit of the GECCO and recalibrated e-Score expanded models with the base model, and the sensitivity analyses models with the recalibrated e-Score expanded models. Model fit statistics Akaike information criterion and AUC are also presented for all nested models to show model fit comparisons. All data manipulations and analyses were performed using SAS Enterprise Guide 8.3 (SAS Institute, RRID: SCR_008567) software.

Data availability

Statistical code that supports this study is available upon reasonable request to the corresponding author. Due to the sensitivity of data collected for this study, analytic datasets underlying this study cannot be shared outside the VA, except as required under the Freedom of Information Act, per VA policy. However, upon formal request and pending approval from the Veterans Health Administration Office of Research Oversight (ORO), a deidentified, anonymized dataset for this study can be created and shared (https://www.mvp.va.gov/pwa/researcher-FAQ). Summary-level information, including phenotype metadata, can be shared on the publicly accessible VA Centralized Interactive Phenomics Resource (https://phenomics.va.ornl.gov/).

Results

Selection of the final sample used in the study ($n = 227{,}504$) is depicted in **Fig. 1**. Descriptive statistics of demographics and colorectal cancer risk factors used to construct the e-Score for the full male sample are presented in **Table 1**. Differences in demographics and risk factors between control and colorectal cancer case groups varied; age ($P < 0.0001$), education ($P < 0.0001$), history of diabetes ($P < 0.0001$), physical activity level ($P < 0.0001$), smoking status ($P < 0.0001$), NSAID use ($P < 0.0001$), calcium ($P = 0.015$), folate ($P = 0.020$), and fruit consumption ($P = 0.019$) were significantly different between the groups.

The multiple logistic regression training model yielded a set of recalibrated β-estimates that were used as weights in the recalibrated e-Score calculation (see Supplementary Table S2). Variables with significant parameter estimates in the original GECCO study were some college or college graduate educational attainment, BMI, heavy alcohol intake, aspirin use, and vegetable intake. Among the

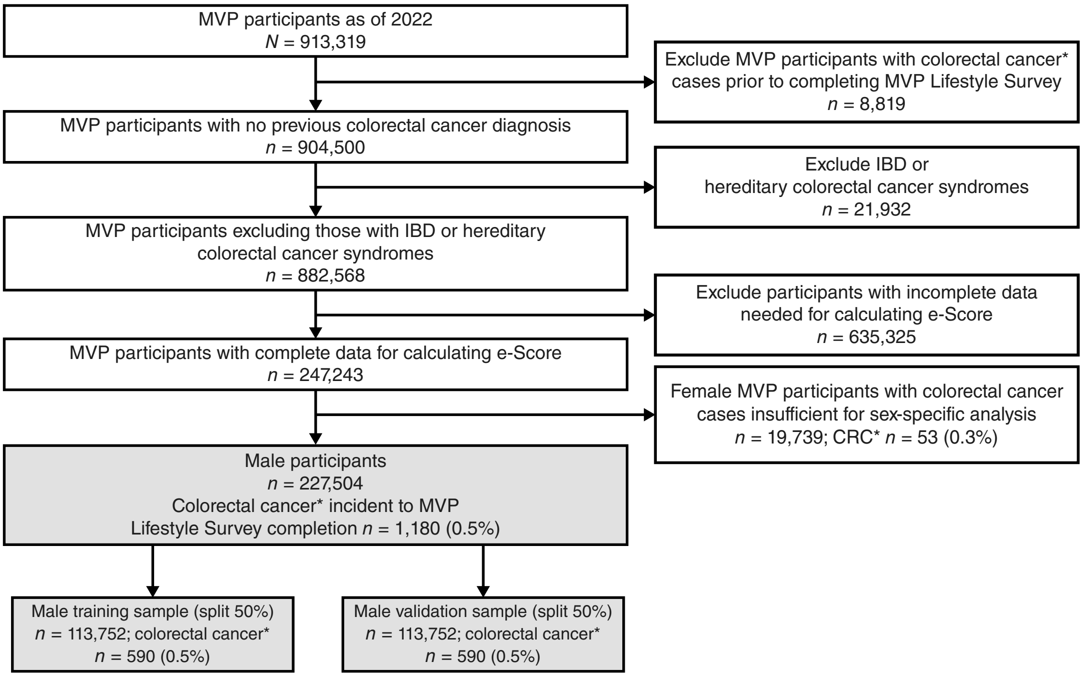
MVP participants as of 2022
N = 913,319
Exclude MVP participants with colorectal cancer* cases prior to completing MVP Lifestyle Survey
n = 8,819
MVP participants with no previous colorectal cancer diagnosis
n = 904,500
Exclude IBD or hereditary colorectal cancer syndromes
n = 21,932
MVP participants excluding those with IBD or hereditary colorectal cancer syndromes
n = 882,568
Exclude participants with incomplete data needed for calculating e-Score
n = 635,325
MVP participants with complete data for calculating e-Score
n = 247,243
Female MVP participants with colorectal cancer cases insufficient for sex-specific analysis
n = 19,739; CRC* n = 53 (0.3%)
Male participants
n = 227,504
Colorectal cancer* incident to MVP Lifestyle Survey completion n = 1,180 (0.5%)
Male training sample (split 50%)
n = 113,752; colorectal cancer*
n = 590 (0.5%)
Male validation sample (split 50%)
n = 113,752; colorectal cancer*
n = 590 (0.5%)

Figure 1.

Sample selection for the final analytic sample. *, Colorectal cancer diagnosis identified by Oncology Domain of the VA Corporate Data Warehouse or by ICD9/10 in inpatient/outpatient records, excluding non–colorectal cancer histologic findings. IBD, inflammatory bowel disease.

MVP male training sample, variables with significant parameter estimates were college graduate educational attainment, height, diabetes diagnosis, lack of recommended physical activity, NSAID use, calcium, fruit, and total energy intake.

Unadjusted and adjusted odds of colorectal cancer comparing colorectal cancer cases ($n = 590$) with controls ($n = 113,162$) for GECCO e-Score and recalibrated e-Score quintiles among the male validation sample are shown in **Table 2**. The proportion of colorectal cancer cases ranged 0.39% to 0.69% and trended upward across participant quintiles of increasing recalibrated e-Scores, except for a slight decrease between quintiles 3 and 4. However, the proportion of colorectal cancer cases across quintiles increasing GECCO e-Scores varied with no discernible trend. This distinction was validated by the adjusted odds for colorectal cancer across GECCO and recalibrated e-Score quintiles (P for trend 0.58 and <0.0001, respectively). Participants with recalibrated e-Scores in quintiles 3, 4, and 5 had significantly higher odds for colorectal cancer compared with those in quintile 1 (ORs Q3 vs. Q1: 1.44, 95% CI, 1.10–1.89; Q4 vs. Q1: 1.34, 95% CI, 1.02–1.77; and Q5 vs. Q1: 1.79, 95% CI, 1.38–2.33, respectively). Across all three models, Black race had higher odds for colorectal cancer compared with the White reference group (base model OR: 1.46, 95% CI, 1.13–1.92; GECCO e-Score model OR: 1.44, 95% CI, 1.09–1.88; and recalibrated e-Score model OR: 1.38, 95% CI, 1.05–1.82). Age, but not family history, was also significantly associated with colorectal cancer.

Results from the nested model analyses to test for significance of the GECCO e-Score and recalibrated e-Score associations with the outcome of colorectal cancer diagnosis are presented in **Table 3**. Using the LRT to compare performance with that of the base model that controlled for age, race, and family history of colorectal cancer, the GECCO e-Score model was not statistically significantly better [$\chi^2(4) = 8.61$, $P = 0.07$]. However, adding the recalibrated e-Score to the base model significantly improved model performance [$\chi^2(4) = 23.95$, $P < 0.001$]. In sensitivity analyses of the recalibrated e-Score among subset cohorts of the full validation sample, which included all the following: (i) those aged 50+, (ii) those with at least one colonoscopy prior to index date; and (iii) those with no history of any cancer prior to index date, the expanded model including the recalibrated e-Score showed statistically significant improvement in performance compared with a base model (see Supplementary Table S3). However, in the nested models among the Black race category only cohort, the expanded model with the recalibrated e-Score was not statistically significantly better.

Discussion

Given the challenges associated with effectively screening all eligible individuals using existing colorectal cancer prevention paradigms, it is important to explore strategies that could personalize screening recommendations for various populations across a variety of clinical settings. In this study, we sought to test the predictive performance for colorectal cancer of an evidence-based environmental risk score based on prior work and recalibrate (17) the data point in an independent population of male US veterans. Despite the robust methods used in the work by the GECCO study upon which our study was based, an e-Score using their study's weighting was not significantly associated with colorectal cancer among the male veteran sample. However, data from nearly a quarter million ($n = 227,504$) male US veteran participants of the MVP were used to recalibrate the e-Score to be veteran specific, and the recalibrated e-Score validation showed that it was significantly associated with colorectal cancer. Race was used as a covariate and in all models, was significantly associated with the risk for

Table 1. Participant demographics and environmental risk characteristics for the male sample ($N = 227{,}504$) controls and colorectal cancer cases.

Characteristic	Controls (*n* = 226,324)	Colorectal cancer cases (*n* = 1,180)	Total sample (*N* = 227,504)	*P* value
Age, years, mean (SD)	66.6 (10.8)	69.4 (8.8)	66.6 (10.8)	<0.0001[a]
Race, *n* (%)				0.2186[a]
Black or African American	19,522 (8.6%)	117 (9.9%)	19,639 (8.6%)	
White	188,509 (83.3%)	976 (82.7%)	189,485 (83.3%)	
Unknown/underrepresented/multiple	18,293 (8.1%)	87 (7.4%)	18,380 (8.1%)	
Ethnicity, *n* (%)				0.3431[a]
Hispanic/Latinx	14,176 (6.3%)	66 (5.6%)	14,242 (6.3%)	
Not Hispanic/Latinx	212,148 (93.7%)	1,114 (94.4%)	213,262 (93.7%)	
Education[b], *n* (%)				<0.0001[c]
Less than high school	7,287 (3.2%)	54 (4.6%)	7,341 (3.2%)	
High school or GED	50,464 (22.3%)	314 (26.6%)	50,778 (22.3%)	
Some college or technical school	98,308 (43.4%)	528 (44.7%)	98,836 (43.4%)	
College graduate or more	70,265 (31.0%)	284 (24.1%)	70,549 (31.0%)	
Height, cm, mean (SD)	177.7 (7.0)	178.1 (6.9)	177.7 (7.0)	0.0745[c]
BMI[d], kg/cm^2, mean (SD)	29.5 (5.4)	29.5 (5.5)	29.5 (5.4)	0.6072[c]
Diabetes[e], *n* (%)	75,675 (33.4%)	472 (40.0%)	76,147 (33.5%)	<0.0001[c]
Physical activity (vigorous)[f], *n* (%)	95,464 (42.2%)	405 (34.3%)	95,869 (42.1%)	<0.0001[c]
Smoking status, *n* (%)				
Ever-smoker[g]	158,565 (70.1%)	907 (76.9%)	159,472 (70.1%)	<0.0001[c]
Smoking pack-years[h], mean (SD)	19.4 (25.6)	24.1 (27.9)	19.5 (25.6)	<0.0001[a]
Alcohol consumption, *n* (%)				<0.0518[a]
<1 g/day	108,372 (47.9%)	605 (51.3%)	108,977 (47.9%)	
1–28 g/day	97,401 (43.0%)	468 (39.7%)	97,869 (43.0%)	
>28 g/day	20,551 (9.1%)	107 (9.1%)	20,658 (9.1%)	
Pharmaceutical use[i], *n* (%)				
Aspirin	32,377 (14.3%)	167 (14.2%)	32,544 (14.3%)	0.8809[a]
NSAID	62,843 (27.8%)	245 (20.8%)	63,088 (27.7%)	<0.0001[a]
Nutrients[j], mean (SD)				
Fiber (g/day)	14.0 (5.6)	13.8 (5.4)	14.0 (5.6)	0.2584[c]
Calcium (mg/day)	797.0 (409.3)	768.6 (395.0)	796.9 (409.2)	0.0150[c]
Folate (μg/day)	637.2 (410.8)	611.2 (401.9)	637.0 (410.7)	0.0204[c]
Diet[j], mean (SD) of daily servings, mean (SD)				
Processed meat	0.4 (0.5)	0.5 (0.5)	0.4 (0.5)	0.3998[c]
Red meat	0.7 (0.6)	0.8 (0.6)	0.7 (0.6)	0.2442[c]
Fruit	1.2 (1.2)	1.2 (1.2)	1.2 (1.2)	0.0188[c]
Vegetable	1.3 (1.3)	1.3 (1.2)	1.3 (1.3)	0.2483[c]
Total energy[k] mean (SD)	1,429.6 (613.3)	1,424.5 (629.6)	1,429.6 (613.4)	0.4445[c]
Family history of colon cancer[l], *n* (%)	28,857 (12.8%)	159 (13.5%)	29,016 (12.8%)	<0.4569[a]
≥1 Colonoscopy[m], *n* (%)	139,196 (61.5%)	1,003 (85.0%)	140,199 (61.6%)	<0.0001[c]

[a] Result from the Kruskal–Wallis test for differences between groups; *P* value < 0.05 is significant.

[b] Education—educational attainment response on the MVP baseline survey.

[c] Result from the χ^2 test for differences between groups; *P* value < 0.05 is significant.

[d] BMI—kg/m^2 derived from MVP baseline survey and VA health record data sources.

[e] Diabetes—Previous diagnosis of type 2 diabetes derived from (a) either ≥1 use of ICD code 250.xx at a primary care visit, or ≥2 uses of the code in any setting and (b) an outpatient prescription of a diabetes drug.

[f] Vigorous physical activity—physical activity level that meets the recommendations of the American Heart Association's Life's Simple Seven[43] reported on the MVP Lifestyle Survey.

[g] Ever-smoked—responded "Yes" to the question "In your lifetime, have you smoked a total of at least 100 cigarettes?" on the MVP Lifestyle Survey.

[h] Pack-years—number of years smoked times number of packs/day (current or former) derived from responses to questions on the MVP Lifestyle Survey.

[i] NSAID and aspirin use—used 2 or more days/week as reported on the MVP Lifestyle Survey.

[j] Nutrient and diet categories—derived from the MVP Lifestyle Survey food frequency questionnaire responses.

[k] Total calories—kCal/day derived from MVP nutrient tables.

[l] Family history of colon cancer—responded "Yes" to questions on the MVP baseline survey that ask whether a grandparent/parent/sibling had colon cancer.

[m] ≥1 colonoscopy—at least 1 procedure code (CPT) for colonoscopy found in participant's VA health records prior to completing the MVP Lifestyle Survey.

colorectal cancer. Synthesizing these findings shows that the recalibrated, veteran-specific e-Score may have clinical utility for improving colorectal cancer preventive care by predicting risk-based colorectal cancer screening needs for individuals within the Veterans Health Administration.

We found that in this sample of US veterans, the e-Score recalibrated using sample-derived weights performed better than the e-Score that used GECCO's published weights. Investigations that produce and test predictive risk scores may require calibration when developing such a tool for use in diverse populations (11). Indeed, US

Table 2. Logistic regression tests for e-Score quintile associations with colorectal cancer among all-male MVP validation sample ($n = 113{,}752$).

	Odds for colorectal cancer					
Metrics	**e-Score Quintile 1 (referent)**	**e-Score Quintile 2**	**e-Score Quintile 3**	**e-Score Quintile 4**	**e-Score Quintile 5**	***P* for trend**
GECCO[a] e-Score (median; range)	8.1; 0–12.2	15.6; 12.2–18.7	21.9; 18.7–25.4	29.5; 25.4–34.7	43.3; 34.7–100.0	
Number of participants	22,751	22,750	22,751	22,750	22,750	
Number (%) of colorectal cancer cases	122 (0.54)	112 (0.49)	94 (0.41)	130 (0.57)	132 (0.58)	
Unadjusted model[b]OR (95% CI)	1.00	0.99 (0.77, 1.23)	0.71 (0.55, 0.93)	0.85 (0.66, 1.09)	0.92 (0.72, 1.18)	0.436
Multivariable-adjusted model[c]OR (95% CI)	1.00	0.99 (0.77, 1.26)	0.71 (0.55, 0.93)	0.86 (0.67, 1.11)	0.95 (0.74, 1.22)	0.587
Recalibrated[d] e-Score (median; range)	32.5; 0–37.6	41.4; 37.6–44.7	47.8; 44.7–50.8	54.1; 50.8–57.9	63.1; 57.9–100.0	
Number of participants	22,751	22,749	22,752	22,749	22,751	
Number (%) of colorectal cancer cases	88 (0.39)	99 (0.44)	128 (0.56)	119 (0.52)	156 (0.69)	
Unadjusted model[e]OR (95% CI)	1.00	1.13 (0.84, 1.50)	**1.46 (1.11, 1.91)**	**1.35 (1.03, 1.79)**	**1.78 (1.37, 2.31)**	**<0.0001**
Multivariable-adjusted model[f]OR (95% CI)	1.00	1.12 (0.84, 1.49)	**1.44 (1.10, 1.89)**	**1.34 (1.02, 1.77)**	**1.79 (1.38, 2.33)**	**<0.0001**

[a] e-Score calculated using β estimates derived from a GECCO training model.

[b] Unadjusted colorectal cancer modeled for GECCO e-Score.

[c] GECCO e-Score multivariable adjusted: colorectal cancer modeled for GECCO e-Score, adjusted for age, race (Black, White, or unknown/underrepresented/multiple), and family history of colorectal cancer.

[d] e-Score calculated using β estimates derived from a recalibrated e-Score training model.

[e] Unadjusted colorectal cancer modeled for recalibrated e-Score.

[f] Recalibrated e-Score multivariable adjusted: colorectal cancer modeled for recalibrated e-Score, adjusted for age, race (Black, White, or unknown/underrepresented/multiple), and family history of colorectal cancer.

Bold *P* value < 0.05 is significant.

veterans that receive VA healthcare are known to have unique demographic characteristics (18) and lifestyle behaviors (19) compared with the US population as a whole, as well as an increased risk for AN during colorectal cancer screening (20). We found risk factors that comprised the e-Score carried different weights of association with colorectal cancer among the MVP study sample compared with the GECCO study sample as shown in Supplementary Table S2. This may be due to distinct differences in characteristics between the all-veteran MVP and the GECCO study samples. In the GECCO study, researchers harmonized data across 14 cohort and case/control studies to produce a matched sample of cases and controls in sex-specific cohorts of White/European ancestry participants. Others have also experienced improvements with recalibrated predictive scoring in other settings, such as surgical outcomes, in a Veteran population (21, 22). We found that our results using the full cohort were consistent in sensitivity analyses among various cohorts of the MVP sample, with overall similar findings of a dose–response relationship between the e-Score and odds for colorectal cancer. See Supplementary Table S3 for nested model fit statistics results for all models. Thus, the results of our study indicate that the recalibrated e-Score may be useful in personalizing screening protocols by age and other risk factors. Potential future uses of the e-Score in all-veteran populations include informing clinical guidelines for screening, focus areas for training clinicians and staff on preventive lifestyle and behavior changes in clinical settings, and development of messaging and prompts in health records to alert both patients and providers to individuals who are potentially at high risk for colorectal cancer.

Findings from this study are additive to the current literature about colorectal cancer risk. The future of colorectal cancer

Table 3. e-Score model comparisons and LRT.

Model	**χ^2 (df) *P***	**AIC**	**AUC**	**LRT[a]**
Primary analytic models[b]				GECCO e-Score: 8.608 > 9.49; *P* = 0.07 **Recalibrated e-Score: 23.969 > 9.49; *P* < 0.001**
Base model	56.610 (4) <0.0001	7339.076	0.579	7329.076
GECCO e-Score expanded model	65.218 (8) <0.0001	7338.468	0.589	7320.468
Recalibrated e-Score expanded model	80.579 (8) <0.0001	7323.107	0.600	7305.107

Abbreviation: AIC, Akaike information criterion.

[a] LRT: If difference between the −2LogL for the base model and expanded model is greater than the χ^2 statistic for difference in degrees of freedom based on an $\alpha = 0.05$, then the expanded model is significantly better.

[b] Primary analytic models use the male validation sample in the following form for the base model: colorectal cancer ~ age + race + family history; with the addition of the GECCO e-Score and recalibrated e-Score quintiles in separate expanded models.

prevention seems to be most informed by a combination of both genetic and environmental risks. The GECCO study showed that genetic test findings combined with lifestyle risks may be more useful than genetic test findings alone when personalizing age for colorectal cancer screenings. Wells and colleagues (23) developed a risk calculator for colorectal cancer using data from >180,000 patients who had colorectal cancer, which showed good accuracy. Erben and colleagues (24) also used a lifestyle risk score in combination with genetic risk for colorectal cancer and found the strongest associations for ANs among the highest tertile lifestyle risk scores compared with the lowest risk tertile (1.96, 95% CI, 1.53–2.51). Zheng and colleagues (25) calibrated a family history colorectal cancer risk prediction model using lifestyle factors as well, of which calcium and fruit were significantly associated with reduced risk for colorectal cancer, mirroring our findings of dietary components in the recalibrated e-Score. Furthermore, a risk score can supplement the heuristics of clinical decision-making. Kostopoulou and colleagues (26) tested the effects of sharing a cancer risk score on referral-making to specialists after presenting vignettes of patient scenarios with general practitioners in the United Kingdom. They found that the clinicians were willing to take into consideration the risk score results alongside their own assessment of the patient and their decision to refer to an oncologist. Other risk scores and risk prediction models have been posited to be used for personalizing screening recommendations and addressing colorectal cancer prevention overall. For example, in one study, less invasive and lower cost fecal immunochemical tests were offered before colonoscopy to those with a lower risk for colorectal cancer based on age, sex, family history, and smoking status (27). Given overlap in risk factors, an e-Score may have utility in the prediction and screening of other chronic diseases as well.

Risk factors comprising the MVP e-Score that were significantly associated with odds for colorectal cancer in our study are known contributors to colorectal cancer risk. These include educational attainment (28), height (29), type 2 diabetes (30, 31), physical activity (32, 8), regular use of NSAIDs (33, 34), and specific dietary factors [calcium (35) and fruits (36)]. Interestingly, intake of red meat (37), alcohol (38), and fiber (36) was not related to odds for a colorectal cancer diagnosis as have been seen in population studies of colorectal cancer. There are also known disparities in colorectal cancer outcomes across race in the general population and among US veterans (39). Interestingly, results from all the validation models did not show any associations between family history and colorectal cancer. This may be due to the older age of the MVP study participants as family history has a smaller effect beyond the age of risk for early-onset colorectal cancer. Furthermore, our use of family history did not include the age of family member's diagnosis, which is important for risk considerations (40).

Unlike the work of GECCO, which used cohorts of only European ancestry participants, the MVP cohort included participants with one or more of several racial identities. This allowed us to build on the work of the GECCO study. In our study, the use of race was limited as a moderator for the e-Score and was defined as a crude, three-category variable that included Black, White, and unknown/underrepresented/multiple, in which the last category included American Indian/Alaskan Native, Pacific Islander, and Asian categories. We found in all models of the primary analyses that Black men had significantly higher odds for colorectal cancer compared with White men, but the other race categories in this study were not significantly associated with colorectal cancer. We found no interaction effects between race and the recalibrated e-Score. In sensitivity analysis among the sample of Black MVP participants, there was no significant association between the recalibrated e-Score and a colorectal cancer diagnosis, which is possibly due to a sparsity of colorectal cancer cases in that sample ($n = 9{,}693$; cases = 61). Our study's results that Black male veterans had higher odds for colorectal cancer were notable and consistent with other population studies' findings. One study that explored the benefits of colorectal cancer screening as it has become more widely available showed that although rates of colorectal cancer have been decreasing among screen-eligible individuals, the rates of decrease seem largest among Whites in the United States (39). Another simulation study that used population data concluded that screening, not risk, was the most important factor related to disparities in odds for colorectal cancer among Black individuals in the United States (9). Further work, particularly in understanding screening uptake and accessibility among all veterans, is needed to better understand these complex multifactorial disparities surrounding equitable care across populations.

There are limitations to this work and the interpretation of its findings. First, the available data used in this study derived from complete case, self-report data collected for the MVP study are subjected to recall and self-selection bias (41, 42). Thus, representativeness is unknown. Second, for many research studies that utilize veteran health data, the findings are limited in generalizability to other populations. Next, the sample of Black veterans used in sensitivity analysis may have been insufficient to draw meaningful conclusions using the methods in this study. It should be noted that the GECCO work and most other explorations of colorectal cancer risk are sex-specific, and we sought to examine colorectal cancer risk factors among the female sample of MVP participants in our study. However, there were insufficient data to calculate veteran-specific e-Score weights in a training sample of females. Alternative methods may be explored in a follow-up study given that females comprise the largest growing population among US veterans. Both Black and female US veteran populations are sorely underrepresented in research and are important to consider in follow-up studies of colorectal cancer risk including endoscopy, family history, and the risk factors discussed in this study.

The development of a recalibrated veteran-specific e-Score in our study provides a foundation for future studies that address the challenges of utilizing risk prediction tools in understudied populations, as well as explore the utility of the e-Score in providing clinical guidance for personalized screening protocols and colorectal cancer prevention strategies. For example, more work is needed to determine if using an e-Score in a veteran population to determine whether lifestyle choices in aggregate should affect starting age of screening. Likewise, if someone reduces a high-risk e-Score, assessing whether risk truly goes down with lifestyle or behavioral changes, or if there is some permanent or other inherent risk that persists is important. Finally, ongoing studies are testing the ability to enhance these risk prediction models with genetic factors for improved colorectal cancer prevention and more effective use of colorectal cancer screening resources.

Authors' Disclosures

B.A. Sullivan reports grants from the American Gastroenterological Association (Research Scholar Award—AGA2021-13-03) during the conduct of the study, as well as grants and personal fees from Exact Sciences outside the submitted work. D. Lieberman reports other support from Geneoscopy, UDX, and ColoWrap outside the submitted work. No disclosures were reported by the other authors.

Disclaimer

This publication does not represent the views of the Department of Veterans Affairs or the US Government.

Authors' Contributions

A.R. Williams: Conceptualization, formal analysis, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **T.S. Redding:** Conceptualization, formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **B.A. Sullivan:** Conceptualization, formal analysis, investigation, writing–original draft, writing–review and editing. **R.N. Baidya:** Conceptualization, methodology, writing–original draft, writing–review and editing. **B. Ear:** Conceptualization, project administration, writing–review and editing. **K. Cho:** Resources, supervision, writing–review and editing. **K.L. Ivey:** Conceptualization, formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **C.D. Williams:** Conceptualization, supervision, writing–review and editing. **J.A. Dominitz:** Conceptualization, supervision, writing–original draft, writing–review and editing. **D. Lieberman:** Conceptualization, resources, supervision, writing–original draft, writing–review and editing. **E.R. Hauser:** Conceptualization, resources, formal analysis, supervision, investigation, methodology, writing–original draft, writing–review and editing.

Acknowledgments

This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by the Department of Veterans Affairs (VA) Office of Academic Affiliations Advanced Fellowship: The Big Data Scientist Training Enhancement Program, Biomedical Laboratory Research and Development (BLRD) Grant #1 I01 BX005718-01, and the American Gastroenterological Association Research Foundation's American Gastroenterological Association Research Scholar Award #AGA2021-13-03. Veteran health records and analyses were conducted using computational resources made available by the VA Informatics and Computing Infrastructure. VA Informatics and Computing Infrastructure is a research and development environment jointly funded by the VA's Office of Research and Development and the Office of Information and Technology. Additional acknowledgments for the Million Veteran Program are found in Supplement S1. US Department of Veterans Affairs. System of Records Notice 97VA10P1: Consolidated Data Information System-VA. 76 FR 25409. May 4, 2011. Amended March 3, 2015.

Note

Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Received May 27, 2024; revised August 8, 2024; accepted September 6, 2024; published first September 10, 2024.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74:12–49.
2. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin 2018;68:250–81.
3. Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, Davis EM, et al; US Preventive Services Task Force. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA 2021;325: 1965–77.
4. Sullivan BA, Redding TS IV, Hauser ER, Gellad ZF, Qin X, Gupta S, et al. High-risk adenomas at screening colonoscopy remain predictive of future high-risk adenomas despite an intervening negative colonoscopy. Am J Gastroenterol 2020;115:1275–82.
5. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. Gastroenterology 2018;154: 2152–64.e19.
6. Betancourt JA, Stigler Granados P, Pacheco GJ, Shanmugam R, Kruse CS, Fulton LV. Obesity and morbidity risk in the U.S. Veteran. Healthcare (Basel) 2020;8:191.
7. Sussman JB, Wiitala WL, Zawistowski M, Hofer TP, Bentley D, Hayward RA. The Veterans Affairs cardiac risk score: recalibrating the atherosclerotic cardiovascular disease score for applied use. Med Care 2017;55:864–70.
8. Lieberman DA, Prindiville S, Weiss DG, Willett W; VA Cooperative Study Group 380. Risk factors for advanced colonic neoplasia and hyperplastic polyps in asymptomatic individuals. JAMA 2003;290:2959–67.
9. Rutter CM, Nascimento de Lima P, Maerzluft CE, May FP, Murphy CC. Black-White disparities in colorectal cancer outcomes: a simulation study of screening benefit. J Natl Cancer Inst Monogr 2023;2023:196–203.
10. Mbemi A, Khanna S, Njiki S, Yedjou CG, Tchounwou PB. Impact of gene–environment interactions on cancer development. Int J Environ Res Public Health 2020;17:8089.
11. Wei J, Shi Z, Na R, Resurreccion WK, Wang CH, Duggan D, et al. Calibration of polygenic risk scores is required prior to clinical implementation: results of three common cancers in UKB. J Med Genet 2022;59:243–7.
12. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 2016;70:214–23.
13. Nguyen X-MT, Whitbourne SB, Li Y, Quaden RM, Song RJ, Nguyen H-NA, et al. Data resource profile: self-reported data in the Million Veteran Program: survey development and insights from the first 850 736 participants. Int J Epidemiol 2023;52:e1–17.
14. Sanchez E. Life's simple 7: vital but not easy. J Am Heart Assoc 2018;7: e009324.
15. Earles A, Liu L, Bustamante R, Coke P, Lynch J, Messer K, et al. Structured approach for evaluating strategies for cancer ascertainment using large-scale electronic health record data. JCO Clin Cancer Inform 2018;2:1–12.
16. Buse A. The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. Am Stat 1982;36:153–7.
17. Nosek BA, Errington TM. What is replication? PLoS Biol 2020;18: e3000691.
18. Eibner C, Krull H, Brown KM, Cefalu M, Mulcahy AW, Pollard M, et al. Current and projected characteristics and unique health care needs of the patient population served by the Department of Veterans Affairs. Rand Health Q 2016;5:13.
19. Dong D, Stewart H, Carlson AC. An Examination of Veterans' Diet Quality. Washington (DC): U.S. Department of Agriculture, Economic Research Service; 2019. p. 32.
20. El-Halabi MM, Rex DK, Saito A, Eckert GJ, Kahi CJ. Defining adenoma detection rate benchmarks in average-risk male veterans. Gastrointest Endosc 2019;89:137–43.
21. Mahmud N, Fricker Z, Hubbard RA, Ioannou GN, Lewis JD, Taddei TH, et al. Risk prediction models for post-operative mortality in patients with cirrhosis. Hepatololy 2021;73:204–18.
22. Kaplan DE, Dai F, Skanderson M, Aytaman A, Baytarian M, D'Addeo K, et al. Recalibrating the child–turcotte–pugh score to improve prediction of transplant-free survival in patients with cirrhosis. Dig Dis Sci 2016;61: 3309–20.
23. Wells BJ, Kattan MW, Cooper GS, Jackson L, Koroukian S. Colorectal cancer predicted risk online (CRC-PRO) calculator using data from the multi-ethnic cohort study. J Am Board Fam Med 2014;27:42–55.
24. Erben V, Carr PR, Guo F, Weigl K, Hoffmeister M, Brenner H. Individual and joint associations of genetic risk and healthy lifestyle score with colorectal neoplasms among participants of screening colonoscopy. Cancer Prev Res (Phila) 2021;14:649–58.
25. Zheng Y, Hua X, Win AK, MacInnis RJ, Gallinger S, Marchand LL, et al. A new comprehensive colorectal cancer risk prediction model incorporating family history, personal characteristics, and environmental factors. Cancer Epidemiol Biomarkers Prev 2020;29:549–57.
26. Kostopoulou O, Arora K, Pálfi B. Using cancer risk algorithms to improve risk estimates and referral decisions. Commun Med 2022;2:1–9.
27. Chiu H-M, Ching JYL, Wu KC, Rerknimitr R, Li J, Wu D-C, et al. A risk-scoring system combined with a fecal immunochemical test is effective in screening high-risk subjects for early colonoscopy to detect advanced colorectal neoplasms. Gastroenterology 2016;150:617–25.e3.
28. Schudde L, Bernell K. Educational attainment and nonwage labor market returns in the United States. AERA Open 2019;5:10.1177/2332858419874056.

29. Thrift AP, Gong J, Peters U, Chang-Claude J, Rudolph A, Slattery ML, et al. Mendelian randomization study of height and risk of colorectal cancer. Int J Epidemiol 2015;44:662–72.
30. Peeters PJHL, Bazelier MT, Leufkens HGM, de Vries F, De Bruin ML. The risk of colorectal cancer in patients with type 2 diabetes: associations with treatment stage and obesity. Diabetes Care 2015;38:495–502.
31. Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. J Natl Cancer Inst 2005;97:1679–87.
32. Slattery ML. Physical activity and colorectal cancer. Sports Med 2004;34: 239–52.
33. Chubak J, Kamineni A, Buist DS, Anderson ML, Whitlock EP. Aspirin Use for the Prevention of Colorectal Cancer: An Updated Systematic Evidence Review for the U.S. Preventive Services Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2015. p. 45.
34. Nan H, Hutter CM, Lin Y, Jacobs EJ, Ulrich CM, White E, et al. Association of aspirin and NSAID use with risk of colorectal cancer according to genetic variants. JAMA 2015;313:1133–42.
35. Carroll C, Cooper K, Papaioannou D, Hind D, Pilgrim H, Tappenden P. Supplemental calcium in the chemoprevention of colorectal cancer: a systematic review and meta-analysis. Clin Ther 2010;32:789–803.
36. Dahm CC, Keogh RH, Spencer EA, Greenwood DC, Key TJ, Fentiman IS, et al. Dietary fiber and colorectal cancer risk: a nested case-control study using food diaries. J Natl Cancer Inst 2010;102:614–26.
37. Alexander DD, Weed DL, Miller PE, Mohamed MA. Red meat and colorectal cancer: a quantitative update on the state of the epidemiologic science. J Am Coll Nutr 2015;34:521–43.
38. Park SY, Wilkens LR, Setiawan VW, Monroe KR, Haiman CA, Le Marchand L. Alcohol intake and colorectal cancer risk in the Multiethnic Cohort Study. Am J Epidemiol 2019;188:67–76.
39. Murphy CC, Sandler RS, Sanoff HK, Yang YC, Lund JL, Baron JA. Decrease in incidence of colorectal cancer among individuals 50 years or older after recommendations for Population-based screening. Clin Gastroenterol Hepatol 2017;15:903–9.e6.
40. Roos VH, Mangas-Sanjuan C, Rodriguez-Girondo M, Medina-Prado L, Steyerberg EW, Bossuyt PMM, et al. Effects of family history on relative and absolute risks for colorectal cancer: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2019;17:2657–67.e9.
41. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211–17.
42. Elston DM. Participation bias, self-selection bias, and response bias. J Am Acad Dermatol 2021;S0190-9622–4.