European Journal of Nutrition (2025) 64:33
https://doi.org/10.1007/s00394-024-03513-9

ORIGINAL CONTRIBUTION

Association of dietary patterns derived by reduced-rank regression with colorectal cancer risk and mortality

Zegeye Abebe[1,2] · Molla Mesele Wassie[1] · Phuc D. Nguyen[1] · Amy C. Reynolds[1] · Yohannes Adama Melaku[1,3]

Received: 18 February 2024 / Accepted: 10 October 2024

Abstract

Purpose Unhealthy dietary patterns contribute to an increased risk of colorectal cancer (CRC). Limited prior research has used reduced rank regression (RRR) to assess dietary patterns relative to CRC risk. This study aimed to identify dietary patterns derived by RRR and assess their associations with CRC risk and mortality.

Methods We used data from the multicentre Prostate, Lung, Colorectal, and Ovarian Cancer Screening (PLCO) trial. Dietary intake was assessed using a Dietary History Questionnaire. In the RRR intake of fibre, folate, and the percentage of energy from carbohydrates, saturated and unsaturated fatty acids were used as response variables. Cox models and competing risk survival regression, with age as the time scale, were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for CRC risk and mortality, respectively.

Results The median follow-up time for CRC risk ($n = 1044$) and mortality ($n = 499$) was 9.4 years (Interquartile Range: 8.0, 10.1) and 16.9 years (11.9, 18.6), respectively. Two dietary patterns were identified: the first was characterised by high carbohydrate, folate and low fatty acid intake, and the second by high fibre and unsaturated fatty acid. Compared to participants in the first tertile of the high fibre and unsaturated fatty acid pattern, those in the third tertile had a lower risk of CRC (HR = 0.88; 95% CI: 0.76, 1.03), and colon cancer (HR = 0.85; 95% CI: 0.72, 1.01). Conversely, the high carbohydrate, high folate and low fatty acid pattern had no association with CRC outcomes. None of the dietary patterns showed associations with rectal cancer or CRC mortality.

Conclusion A diet enriched with high fibre and unsaturated fatty acids may reduce the risk of CRC. These results highlight the potential protective effect of adequate fibre intake in conjunction with high consumption of unsaturated fatty acids against CRC.

Keywords Colorectal cancer · Healthy diet · Hybrid dietary data analysis · Longitudinal data · Reduced rank regression · Survival analysis

Background

Colorectal cancer (CRC) is one of the leading causes of cancer-related morbidity and mortality [1], and is the third most common cancer worldwide [2]. Incidence of CRC is steadily increasing, with the number of CRC cases and associated deaths reaching 2.17 million in 2022. It is estimated that 3.2 million new CRC cases will be diagnosed, and 1.6 million people will die from CRC-related causes in 2040 [3].

The development and progression of CRC are influenced by various factors, including genetic predisposition [4, 5] and unhealthy lifestyle behaviours such as physical inactivity [6] and poor diet [7]. Diet is a modifiable risk factor that plays an important role in the development of CRC [8].

✉ Zegeye Abebe
abeb0011@flinders.edu.au

1 Flinders Health and Medical Research Institute, Flinders University, Adelaide, South Australia 5042, Australia

2 Department of Human Nutrition, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, P. O. Box 196, Gondar, Ethiopia

3 Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia

Published online: 28 November 2024

Among the nutrients that have been investigated are carbohydrates, fibre, folate, saturated and unsaturated fatty acids [9–12]. However, the synergic or interaction effect of these nutrients on CRC risk and mortality has not received much attention, as most existing studies focus on specific nutrients or food items, and do not use dietary data analysis methods which consider all consumed food items and nutrients collectively.

This is problematic, as previous analyses of associations between single nutrient or food items and cancer risk have failed to capture the complexity of dietary patterns [13]. Nutrients and food items are not consumed in isolation, and we need approaches that consider food intake more holistically. To address this complexity, more recent studies have adopted various methodological approaches to capture the multidimensionality of dietary intake and interactions between nutrients and food items [14]. These include a priori methods (based on adherence to a pre-specified dietary pattern) [15], *a posteriori* methods (based on observed dietary intake) [16], or hybrid methods [17].

Several studies previously assessed the association between dietary patterns and the risk of CRC using both *priori* and *posteriori* methods. Systematic reviews and meta-analyses indicate that high scores of a healthy eating index [18, 19], the Mediterranean diet [18, 20], dietary approach to stop hypertension [18, 21], and anti-inflammatory diet [18, 19] were associated with a reduced risk of CRC. Similarly, *posteriori* methods have shown that prudent dietary patterns protective against CRC [22, 23]. On the contrary, the western dietary pattern has been associated with an increased risk of CRC [22, 24, 25].

These dietary indices do not consider nutrient-nutrient interaction and correlations, do not describe the overall diet quality (quality scores focus on selected aspects of the diet), and are limited by the available scientific knowledge of food and health outcomes [15, 16]. While p*osteriori* methods consider the interaction of different foods or nutrients, do not take into account the previous knowledge of diet and its relationship with CRC [26].

Methods that used both combination of a priori and post priori methods may provide a better understanding of between dietary intake and CRC. Reduced rank regression (RRR) is one of the hybrid methods that combines dietary data with a priori knowledge of disease-related factors, such as biomarkers or nutrients, in order to derive dietary patterns that are directly associated with health outcomes [27]. Previous research has indicated that dietary patterns derived through RRR exhibit a robust association with conditions like cardiovascular disease [17], breast cancer [28] and bone mass density [29] when compared to dietary patterns derived through a priori and *a posteriori* methods. While promising applications are apparent in other health outcomes, the utilization of RRR in the context of prospective studies examining dietary patterns and CRC risk is limited to date. Additionally, a limited number of studies have considered fibre, folate and macronutrients as response variables to assess the association between dietary intake and risk of CRC. These studies have found differential effects of carbohydrates, fat, fibre, and folate on CRC risk, with some increasing and others reducing risk [12, 30–34].

Using dietary patterns that account for the greatest variation in these nutrients can aid in evaluating the relationship between dietary intake and the risk of CRC. Consequently, RRR can be employed to identify dietary patterns associated with CRC, particularly through the pathways involving fibre, folate, and the intake of carbohydrates, saturated and unsaturated fatty acids. Thus, we aimed to identify dietary patterns explaining the variability in total fibre and folate intake, as well as the percentage of energy from carbohydrates, saturated and unsaturated fatty acids, to determine their association with CRC risk and mortality.

Methods

Subjects and study design

The current analyses utilise data from a large multicentre randomized controlled trial study of the Prostate, Lung, Colorectal, and Ovarian Cancer Screening (PLCO). The PLCO was designed by the National Cancer Institute (NCI) in the United States of America (USA) and was established in 1993 to determine the effects of screening on cancer-related mortality in men and women aged 55 to 74 years. The detailed trial profile is described elsewhere [35].

Briefly, from November 1993 to July 2001, around 155,000 participants were enlisted at 10 screening centres across the USA, and randomly assigned to either the intervention or control group. The intervention group underwent screening examinations. Digital rectal examination and prostate-specific antigen testing were used to detect prostate cancer in males, cancer antigen 125 and transvaginal ultrasound were used to detect ovarian cancer in females, and chest radiography and flexible sigmoidoscopy were used to detect lung and CRC in both males and females. The control group received standard care. Both groups completed a baseline questionnaire, and the intervention group additionally completed a dietary questionnaire. In December 1998, a second dietary measure, the Dietary History Questionnaire (DHQ), was introduced for both groups to assess dietary intake. Control arm participants randomized before December 1998 were provided the DHQ in 1999 or 2000, while those randomized in or after December 1998 were offered the DHQ at baseline.

Physical activity status was assessed between 2006 and 2007. Data on cancer diagnoses were collected through 2009, and mortality data were tracked through 2018. For this analysis, participants who completed DHQ and had a valid dietary DHQ (with caloric intake falling between the 1st and 99th percentiles, and had less than 8 missing frequency responses on the DHQ) were included [36]. Participants with a personal history of cancer before DHQ collection were excluded. Further exclusions were applied to individuals with missing values on key confounding variables (< 1%). As a result, the final sample consisted of 97,561 participants (Fig. 1).

Dietary assessment

The DHQ questionnaire is widely used in prior studies [36–40] and collects information about the frequency of 77 food items typically consumed, as well as the usual dietary intake of the study participants in the previous 12 months [41]. The questionnaire also assesses the typical portion size. The questionnaire can be publicly accessed here: http://riskfactor.cancer.gov/DHQ/. The DHQ has been evaluated against a widely used food frequency questionnaire, the Willett and Block food frequency questionnaire (FFQ), in Eating at America's Table Study [42]. The results showed that DHQ is best compared to Willett and Block FFQs. Notably, strong correlations of 0.48, 0.45, and 0.18 were observed for women, and 0.49, 0.45, and 0.21 for men, respectively, between true energy and DHQ, Block FFQ, and Willett FFQ [42]. Additionally, DHQ exhibited higher correlations for 26 nutrients in contrast to Block and Willett FFQs [42]. Furthermore, validation by Thompson et al. [43] involved a comparison of DHQ concerning the design of dietary data collection approaches, such as grouping (single vs. multiple separate questions), different forms of food (consumption frequency of each food vs. consumption frequency of main food), additions (e.g., adding sugar to coffee), and units (portion size vs. frequency of units). The findings from this validation suggested that the FFQ in DHQ enhanced the accuracy of reporting dietary data.

Fig. 1 Flow diagram of subjects included in the prospective analysis of the PLCO Trial. Note: DHQ, Dietary History Questionnaire; CRC, Colorectal cancer; PLCO, Prostate, Lung, Colorectal, and Ovarian

To determine nutrient and food intake amounts, US dietary data and the pyramid food group servings database from the US Department of Agriculture (USDA) were utilized [44]. The food and nutrient values obtained were then used to create food groups based on the USDA's My Pyramid Equivalents Database (MPED) [45]. The database translates the intake components of the National Health and Nutrition Examination Survey (NHANES) 2003–2004 into the 27 major groups and subgroups as grain group (whole grain and refined grain), vegetable group (dark green vegetables, orange vegetables, white potatoes, other starch vegetables, tomatoes and other vegetables), fruit group (citrus fruits and other fruits), milk group (milk, yogurt, and cheese), meat and beans group (meat, organ meats, frankfurters, poultry, fish and shellfish high in n-3 fatty acids, fish and shellfish low in n-3 fatty acids, eggs, cooked dry beans and peas, soybean products, nuts and seeds), oils (discretionary oil), and extras (discretionary solid fat, added sugars and alcoholic beverages [45]. In this study, a total of 29 food groups based on MyPyramid major group and subgroup classification, with 2 two additional groups created for alcohol, as detailed in Supplementary Table 1, were used to construct dietary patterns.

CRC outcomes

The primary outcome of interest was incident CRC cases. The identification of CRC cases primarily involved self-reporting through annually administered follow-up questionnaires. Each participant received an annual follow-up email prompting them to provide details about cancer diagnoses, including type, site, date of diagnosis, and contact information for their healthcare providers. To ensure the accuracy of reported cancer cases, a thorough review of medical records was conducted using a standard form [46]. Within the screening arm, additional cases were identified through clinical follow-up of a positive screening test [47]. CRC cases were coded according to the criteria defined by the International Classification of Diseases for Oncology (ICD-O-2) codes. Specifically, codes C180-C189 were associated with colon cancer, while codes C199, C209, C212, and C218 were linked to rectum cancer [37].

The secondary outcome was CRC mortality. CRC mortality was also identified through annual study update questionnaires and determined from death certificate information gathered through annual follow-up procedures and annual searches of the National Death Index (NDI). CRC specific mortality was determined according to the International Classification of Diseases version 9 (ICD-9): codes 153XX-154XX (except 1535X) as CRC. Finally, an independent Death Review Committee further reviewed deaths if participants (1) were diagnosed with a confirmed PLCO cancer; (2) had a death certificate stating they died of a PLCO cancer and/or (3) had death certificate which was ambiguous as to whether the cause of death was a PLCO cancer [46].

Confounding variables

Socio-demographic characteristics were collected at baseline via questionnaire. Smoking habits were categorized as *current*, *former smoker*, *and never smoked*. Education status was operationalized as *completed less than 8 years; attending 8–11 years; attending up to 12 years or complete high school; post high school training other than college; some college; college graduate*, *and postgraduate*. Occupational status was defined as *working*, *homemaker*, *unemployed*, *retired*, *extended sick leave*, *disabled*, *and others*. Body mass index (BMI) was classified as *underweight (< 18.5)*, *normal (18.5–24.9)*, *overweight (25-29.9)*, *or obese (> 30 kg/m^2)*. In addition, family history of any cancer other than CRC (*yes/no*), family history of CRC (*yes/no*), Aspirin use (*yes/no*), diabetes (*yes/no*), colorectal polyps (*yes/no*), and diverticulitis (*yes/no*) were collected at baseline.

Physical activity level of the study participants over 12 months was assessed using a questionnaire (Supplementary Table 2). In the PLCO study, participants were queried about moderate physical activity with the question, "On average, how many days per week did you engage in any moderate physical activity where you worked up a light sweat or increased your breathing and heart rate to moderately high levels?" Response options included none or less than 1 day per week, 2 to 3 days per week, 4 to 5 days per week, and 6 to 7 days per week. For each frequency category, participants were also asked about the duration of each session, with options ranging from none or less than 15 min to 40 min or more. A similar question-and-response format was used for vigorous physical activity. To analyse the data, we adopted the procedures outlined in a previous PLCO publication [48] to calculate the weekly time spent in moderate and vigorous physical activity.

Initially, we summed the lower and higher bounds of each response category and divided by two to obtain the average days of moderate physical activity per week. This resulted in the following categories: *none or less than 1 day per week as 0*, *2 to 3 days per week as 2.5*, *4 to 5 days per week as 4.5*, *and 6 to 7 days per week as 6.5*. Similarly, the average duration of each session was categorized as follows: *none or less than 15 min as 0*, *16 to 19 min as 17.5*, *20 to 29 min as 24.5*, *30 to 39 min as 34.5*, *and 40 min or more as 40*. The same procedures were applied to strenuous physical activity.

Then, the total amount of time spent in moderate or vigorous physical activity per week was calculated as a product of weekly frequency and duration for each physical activity.

Finally, participants who reported greater than 75 min of vigorous activity per week or 150 min of moderate activity per week were classified as meeting the recommended physical activity guidelines for health [49]. Those not meeting these criteria were classified as not fulfilling the recommended physical activity per week.

We used a direct acyclic graph (DAG) to identify potential confounding variables. The dietary habits of individuals can vary based on demographic and behavioural factors, including age, sex, family history of cancer, educational background, occupation, physical activity level, and smoking habits. Additionally, the presence of colon-related health issues, liver problems, and colorectal polyps may prompt modifications in dietary choices. These factors, along with demographic [50, 51] and behavioural characteristics [52], as well as colon comorbidity, liver problems [53], and colorectal polyps [54], showed associations with CRC and are considered confounders in our analysis.

Three models were built from the DAG (Supplementary Figs. 1–3). The first model (Model 1) was adjusted for age, sex, marital status, family history of any other cancer, family history of CRC, educational status, occupation, physical activity, colon comorbidity, liver problems, diverticulitis, colorectal polys, physical activity and smoking (Supplementary Fig. 1). The second model additionally adjusted for diabetes, hypertension, and BMI (Model 2) (Supplementary Fig. 2). The third model included adjustment for Aspirin use (Model 3) (Supplementary Fig. 3) and, the last model was additionally adjusted for total energy intake (Model 4). Since diabetes, hypertension, and overweight/obesity were assessed at the time of dietary data collection, we used model-based approaches where we assumed these conditions as mediators (Model 1) and confounders (Model 2, 3 and 4) in separate models. Dietary intake is recognized as a risk factor for conditions such as diabetes [55], hypertension [56], and overweight/obesity [57]. Consequently, these health conditions also pose a risk for CRC [58–60]. The presence of diabetes, hypertension, and overweight/obesity may lead to adjustments in dietary habits and can be considered as confounding variables. In addition, it has been documented that the use of aspirin is associated with an increased risk of CRC. However, the relationship between dietary patterns and aspirin uses, or vice versa, is not established. Consequently, aspirin use is considered a covariate in our analysis.

Dietary pattern analysis

To determine the dietary patterns (DP) in relation to CRC, RRR was used. RRR is specifically designed to derive dietary patterns that maximize the variation explained by selected response variables, based on an a priori hypothesis that these variables are associated with the outcome of interest [16].

In this analysis, the response variables used to identify the RRR-based dietary patterns were selected intermediate response variables following evidence from the World Health Organization (WHO) report on chronic disease prevention, the World Cancer Research Fund/America Institute for Cancer Research guidelines on diet and nutrition and physical activity for cancer prevention and relevant literature [33, 61–64]. The chosen response variables included fibre density (g/d) [62], folate density (g/d) [12], proportion of energy derived from carbohydrates [63], unsaturated fatty acids [33], and saturated fatty acids [61]. Fibre and folate densities were determined by dividing the total daily intake of fibre and folate from the diet (mg/day) by the total daily energy intake (kJ/day) and then multiplying by 100. The proportion of energy derived from total carbohydrates, unsaturated fatty acids, and saturated fatty acids was calculated by dividing the energy intake from carbohydrates (kJ), unsaturated fatty acids (kJ), and saturated fatty acids (kJ) by the total energy intake and multiplying each result by 100.

Diets rich in simple carbohydrates and sugars have been associated with an increased risk of CRC, possibly due to their impact on insulin levels and inflammation [65]. In contrast, dietary fibre reduces the risk of CRC due to its potential to regulate blood sugar levels, promote regular bowel movements, dilute carcinogens in the colon, and provide a substrate for beneficial gut bacteria [66, 67].

Diets high in saturated fatty acids, which are commonly found in red meat, full-fat dairy products, and certain oils, have been associated with an increased risk of CRC in several studies [31, 32]. The mechanism behind this association is thought to involve the promotion of inflammation and the production of carcinogenic compounds during the digestion of saturated fats [68]. On the contrary, unsaturated fatty acids, especially the omega-3 fatty acids found in fatty fish, flaxseeds, and walnuts, have been investigated for their potential protective effect against CRC [33, 34]. Omega-3 fatty acids have anti-inflammatory properties and may help suppress the growth of cancer cells [69, 70]. Folate is found in foods like leafy greens, citrus fruits, and legumes, plays a crucial role in DNA synthesis and repair. Some research has suggested that adequate folate intake may reduce the risk of CRC, especially in individuals with certain genetic variants that affect folate metabolism [12].

In the RRR analyses, 29 food groups were used as predictors. To calculate the factor scores, the intake of these food groups was standardized. Subsequently, five-factor scores were generated, but only the first two-factor scores were retained for further analysis. This decision was made because each of the first two-factor scores accounted for more than 10% of the variation in dietary intake observed

among the study population and interpretability of the dietary patterns. An absolute value of factor loading ≥ 0.2 was used to declare a significant contribution of a food group to the dietary patterns and calculate factor scores [71]. The factor scores were categorized into tertiles, the first tertile (lowest intake and the third tertile (highest intake). The correlation between factor scores and intermediate response variables was calculated and used to label the names of dietary patterns.

Statistical analysis

Descriptive statistics were used to summarize the baseline patient characteristics across the tertiles of each dietary pattern. Mean (± SD), median (interquartile range, IQR), and proportion were used to summarize the results. Survival analysis was used to assess the association between RRR-derived dietary patterns and CRC risk. Kaplan-Meier survival curve was used to compare the risk and mortality of CRC. A Cox model was used to estimate the hazard ratios of CRC risk associated with dietary patterns, in which age was used as the time scale. In the analysis, tests for the linear trend of association across tertiles of each factor score were performed by incorporating the tertiles of dietary patterns as continuous parameters in the models.

The follow-up time commenced for each participant based on age at DHQ assessment. Follow-up ended according to age at CRC diagnosis (event), and censored at loss to follow-up, at death, or at end of follow-up.

Similarly, to assess the association of dietary patterns with CRC mortality, competing risk survival regression considering death due to other causes as competing risk was used [72]. Individuals who had no event of interest during the follow-up period were considered censored (assigned 0) and the death of individuals from CRC was considered as the event (assigned 1). Death due to causes other than CRC was considered as a competing event (assigned 2).

To estimate the association between dietary patterns and the anatomical site of cancer, the analysis was repeated for colon cancer and rectal cancer separately. Finally, since the number of CRC cases and mortality is relatively small compared to the sample size, we combined the incidence and mortality of cases together and labelled them as composite outcomes, and re-ran the analyses [73].

Missing data

Overall, 25.4% ($n = 24{,}776$) of the participants had missing data on physical activity. As missing information on physical activity was determined to be missing at random, multiple imputation method based on chained equations was used to impute missing values in covariates [74, 75]. We identified activity history as an auxillary variable to impute physical activity level (correlation = 0.4) [76]. We specified imputation models to impute values based on a logistic regression model for physical activity. All other baseline variables including CRC status were included in these models. We imputed 10 complete imputations [77]. The associations of dietary patterns and CRC incidence and mortality were estimated in each imputed data set, and the 10 estimates combined using Rubin's rule [78]. All other covariates were missing less than 1% data and were excluded from the multiple imputation process [79].

Sensitivity analysis

Three sensitivity analyses were conducted to check the robustness of the findings. First, the analysis was conducted by removing those who developed CRC within one year after dietary assessment to rule out potential reverse causation. Second, we excluded participants reporting excessive alcohol consumption (30 g/day and 20 g/day for men and women, respectively) as the amount of ethanol more than these thresholds is a convincing cause of CRC. Finally, as 25.4% of the study participants had missing physical activity data, a complete case analysis was conducted to compare the results with the findings reported using multiple imputations for physical activity data.

Dietary patterns, descriptive and complete case Cox proportional hazard analyses were carried out using Stata version 17 [80]. R with RStudio [81] was used for multiple imputation using multiple imputation with chained equation (*mice*) package [82] and competing risk survival regression analysis using *cmprsk* package [83].

Results

Sociodemographic characteristics

A total of 97,561 individuals (47,313 males and 50,248 females) were included in the analyses, among whom 1, 044 CRC cases and 499 CRC mortalities were identified in a median follow-up time of 9.4 (2.9) years, with a maximum follow-up of 11.9 years during a total of 7,248,447.5 person-years. The incidence and mortality rate of CRC during the follow-up period was 14.4 (95% CI: 13.6, 15.3) and 33.9 (95% CI: 31.1, 37.0) per 100,000-person-years. The median age of the study participants was 62 years (58.0, 66.0). Approximately 10.2% of the participants had a family history of CRC, while around 55.9% had a family history of other cancers. A third (32.5%) and 6.6% of the study participants had hypertension and diabetes, respectively (Table 1).

Dietary patterns

Five dietary patterns were identified. Figure 2 provides the factor loadings for the first two dietary patterns. Dietary pattern I (DP1) was characterized by a higher intake of citrus fruit, other fruit, non-wholegrain, wholegrain, sugar, and dark green vegetables. Conversely, dietary pattern II (DP2) was characterized by a higher intake of dietary fat oil and solid, dark green vegetables, and other vegetables, and a lower intake of sugar, beer, and liquor.

DP1 reflected positive correlations with fibre density ($r=0.7$), folate density ($r=0.6$), and percentage of carbohydrate intake ($r=0.8$) and negative correlations with the percentage of saturated fatty acid intake (r=-0.7) and unsaturated fatty acid intake (r=-0.8) (Fig. 3). Consequently, this pattern was labelled as the "high carbohydrate/fibre, folate, and low fatty acid pattern" and accounted for 50.7% and 67.5% of the variance in the intermediate response variables and total food intake, respectively (Fig. 3).

In contrast, DP2 displayed positive correlations with fibre density ($r=0.5$), folate density ($r=0.4$), and unsaturated fatty acid intake ($r=0.5$) (Fig. 3). It was labelled as the "high fibre, and unsaturated fatty acid pattern." DP2 explained 12.9% and 17.2% of the variance in the intermediate response variables and total food intake, respectively.

The first two dietary patterns combined explained 64% and 84% of the total variance in the intermediate response variables and dietary intake, respectively. On the other hand, dietary pattern III (DP3), dietary pattern IV (DP4), and dietary pattern V (DP5) contributed less than 10% of the variance in the total food intake and were therefore excluded from further analysis (Supplementary Tables 3–5).

Baseline dietary pattern characteristics

There were significant differences for sex, age group, marital status, occupation, smoking status, BMI and presence of diabetes across tertiles of dietary patterns. However, there was no difference in colon or liver comorbidity across tertiles of both DP1 and DP2 (Supplementary Table 6).

In addition, there was a baseline difference in the daily nutrient and energy intake among the study participants across dietary patterns. Study participants in the highest tertile of DP1 reported consuming higher fibre, folate and carbohydrates, but lower protein, saturated fat, mono- and polyunsaturated fatty acid, and alcohol at baseline. Study participants in the highest tertile of DP2 consumed higher fibre, discretionary fat, protein, mono and polyunsaturated fatty acids, and folate. However, participants in the highest tertile of DP2 consumed lower carbohydrates and alcohol (Supplementary Table 7). Furthermore, Supplementary Table 8 showed the daily serving size consumption of the 29 food groups by the study participants across DP1 and DP2.

Dietary patterns and CRC incidence

In comparison to the first tertile of DP2, those in the second (HR = 0.87; 95%CI: 0.74, 1.01) and third (HR = 0.88; 95%CI: 0.76, 1.03) tertiles had lower CRC risk(P-trend = 0.101). DP1 had no association with the risk of CRC (highest tertile: HR = 1.00; 95% CI: 0.85, 1.18, P-trend = 0.997).

When the analysis was stratified by anatomical site of the tumour, comparing the highest tertiles to the first tertiles, DP2 was associated with a reduced risk of colon cancer (HR = 0.85; 95% CI: 0.72, 1.01, P-trend = 0.059) but not with rectal cancer. DP1 had no association with both colon and rectal cancers (Table 2). The Kaplan-Meier failure curve comparing the risk of developing CRC across tertiles of DP1 and DP2 is reported in in Supplementary Figs. 4 and 5.

Dietary patterns and CRC mortality

The median follow-up period was 16.9(11.9–19.9) years and there were 499 mortalities attributed to CRC. Results from the fully adjusted model showed that higher adherence to neither DP2 (SHR = 0.91; 95%CI: 0.72, 1.14, P-trend = 0.7356) nor DP1 (SHR = 0.95; 95%CI: 0.75, 1.21, P-trend = 0.2189) was associated with CRC mortality (Table 3). The Kaplan-Meier survival curve comparing the probability of surviving from CRC across tertiles of DP1 and DP2 is shown in Supplementary Figs. 6 and 7.

Sensitivity analysis

For CRC risk, similar results were found when the analysis was conducted without excessive alcohol consumers, and those who developed CRC within one year of dietary data collection (Supplementary Tables 9 and 10). Using complete case analysis, the association between dietary patterns and CRC risk was also similar with the analysis from the imputed data (Supplementary Table 11).

Discussion

Two dietary patterns were identified through the utilization of RRR, focusing on the density of fibre and folate, as well as the proportion of energy derived from carbohydrates, saturated fatty acids, and unsaturated fatty acids. The high fibre and unsaturated fatty acid pattern exhibited an inverse association with the risk of CRC and colon cancer suggesting a protective effect against CRC. Dietary patterns (D1)

Table 1 Baseline characteristics of the study participants in the PLCO study ($N = 97,561$)

Variables	*n*	Percentage [a]
Median (IQR) age at baseline	62 (58, 66)	
Female	50,248	51.5
Race		
White, Non-Hispanic	88,851	91.1
Black, Non-Hispanic	3,156	3.2
Hispanic	1,426	1.5
Asian	3,479	3.6
Pacific Islander	453	0.5
American Indian	196	0.2
Educational status		
< 8 Years	593	0.6
8–11 Years	5,181	5.3
12 years or completed high school	22,724	23.3
Post high school training other than college	12,621	12.9
Some college	20,984	21.5
College graduate	17,182	17.6
Postgraduate	18,276	18.7
Marital status		
Married or living as married	76,593	78.5
Widowed	7,859	8.1
Divorced	9,287	9.5
Separated	750	0.8
Never married	3,072	3.2
Occupational status		
Homemaker	11,482	11.8
Working	39,244	40.2
Unemployed	837	0.9
Retired	42,027	43.1
Extended sick leave	141	0.1
Disabled	1,619	1.7
Other	2,211	2.3
Smoking status		
Never smoked cigarettes	46,674	47.8
Current cigarette smoker	8,975	9.2
Former cigarette smoker	41,912	43.0
Physical activity meets the requirements for health	25,685	26.3
Missing	24,776	25.4
Family history of CRC		
No	85,198	87.3
Yes, immediate family member	9,987	10.2
Possibly- relative or cancer type not clear	2,376	2.4
Family history of any other cancer	54,486	55.9
BMI Baseline		
0-18.5	652	0.7
18.5–25	32,873	33.7
25–30	41,614	42.7
30+	22,422	23.0
Aspirin use	45,920	47.1
Colon Comorbidities[#]	1,299	1.3
Diabetes	6,509	6.7
Diverticulitis	6,525	6.7
Hypertension	31,759	32.5
Liver comorbidities[@]	3,487	3.6

Table 1 (continued)

Variables	*n*	Percentage [a]
Osteoporosis	4,932	5.1
Colorectal Polyps	6,439	6.6

[a] All values are percentages unless specified

[#] Ulcerative Colitis, Crohn's Disease, Gardner's Syndrome, or Familial Polyposis

[@]hepatitis or cirrhosis

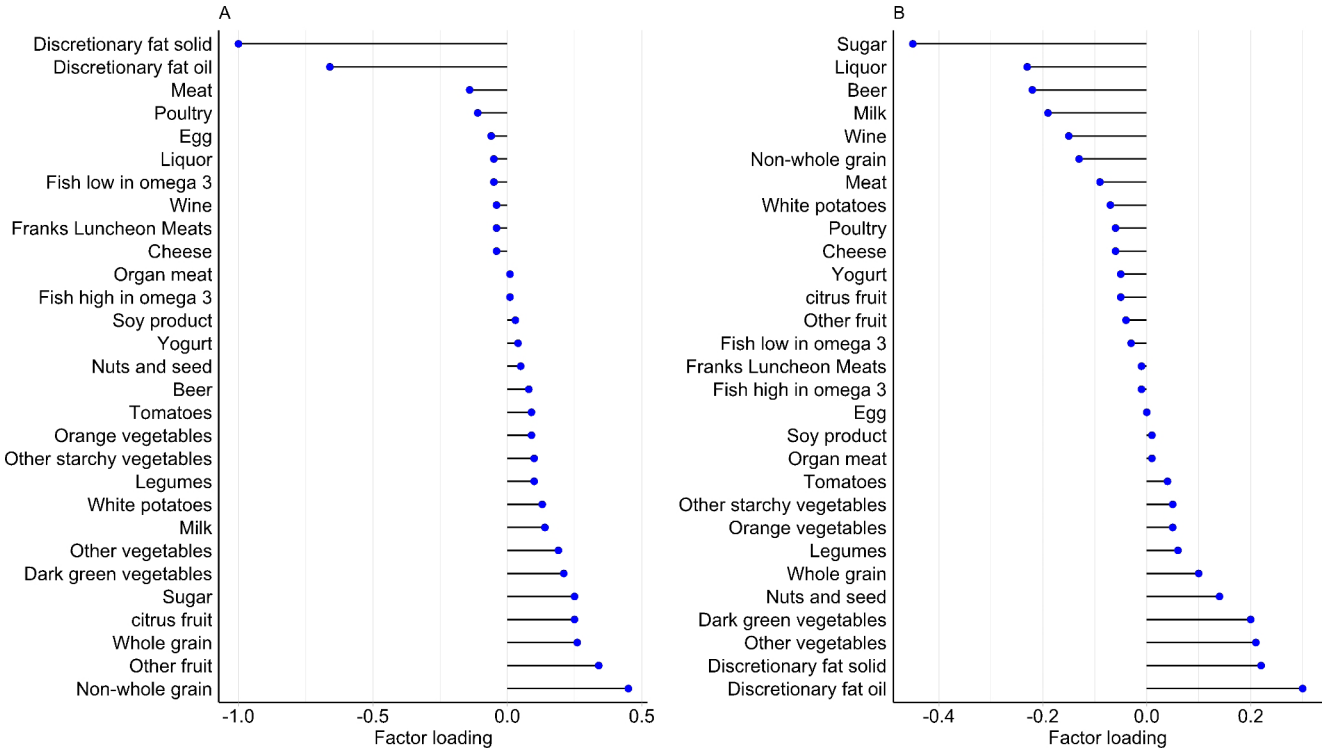

Fig. 2 Factor loading of dietary patterns generated by RRR among study participants in PLCO study. **A**) dietary pattern I(DP1); **B**) dietary pattern II (DP2)

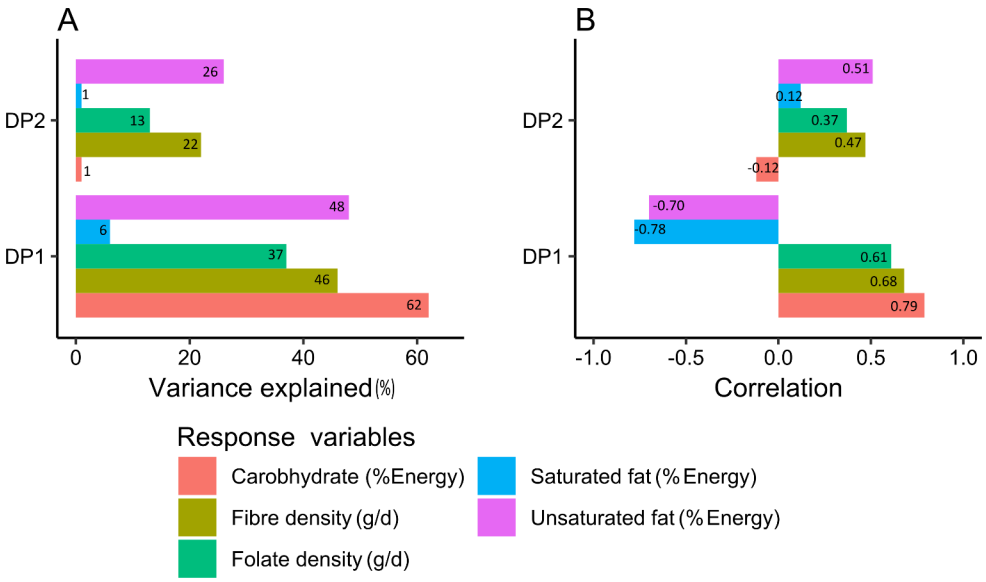

Fig. 3 Explained variation (%) in response variables and correlation between dietary patterns and response variables using RRR. **A**) correlation between dietary patterns and response variables; **B**) Explained variance in the response variables by each dietary patterns

Table 2 Hazard ratio (95% CI) for risk of CRC by tertiles of DP1 and DP2 among study participants in PLCO study

Model	DP1				DP2			
	T1	T2	T3	*P*-trend	T1	T2	T3	*P*-trend
			HR a (95% CI)					
CRC								
Model 1	Ref	1.01(0.87, 1.17)	0.98(0.84, 1.15)	0.809	Ref	0.88(0.76, 1.02)	0.91(0.78, 1.05)	0.200
Model 2	Ref	1.02(0.87, 1.18)	1.00(0.85, 1.17)	0.970	Ref	0.86(0.75, 1.00)	0.88(0.76, 1.03)	0.101
Model 3	Ref	1.02(0.87, 1.18)	1.00(0.85, 1.17)	0.994	Ref	0.86(0.75, 1.00)	0.88(0.76, 1.03)	0.101
Model 4	Ref	1.02(0.87, 1.19)	1.00(0.85, 1.18)	0.997	Ref	0.87(0.74, 1.01)	0.88(0.76, 1.03)	0.101
Colon								
Model 1	Ref	1.04(0.88, 1.23)	1.01(0.85, 1.20)	0.885	Ref	0.86(0.73, 1.01)	0.89(0.75, 1.05)	0.150
Model 2	Ref	1.05(0.89, 1.24)	1.03(0.87, 1.23)	0.742	Ref	0.84(0.71, 0.99)	0.85(0.72, 1.01)	0.060
Model 3	Ref	1.05(0.89, 1.24)	1.03(0.87, 1.23)	0/714	Ref	0.84(0.71, 0.99)	0.85(0.72, 1.01)	0.060
Model 4	Ref	1.04(0.87, 1.24)	1.03(0.86, 1.22)	0.793	Ref	0.83(0.71, 0.98)	0.85(0.72, 1.01)	0.059
Rectum								
Model 1	Ref	0.88(0.60, 1.27)	0.83(0.56, 1.23)	0.363	Ref	0.99(0.68, 1.43)	1.03(0.70, 1.51)	0.871
Model 2	Ref	0.89(0.61, 1.29)	0.85(0.57, 1.26)	0.415	Ref	1.00(0.69, 1.45)	1.06(0.72, 1.55)	0.770
Model 3	Ref	0.88(0.61, 1.29)	0.84(0.57, 1.25)	0.264	Ref	1.00(0.69, 1.45)	1.06(0.72, 1.55)	0.976
Model 4	Ref	0.93(0.63, 1.37)	0.88(0.58, 1.31)	0.402	Ref	1.03(0.70, 1.51)	1.06(0.73, 1.56)	0.686
Combined results ($n = 1239$)								
Model 1	Ref	0.95(0.83, 1.09)	0.90(0.78, 1.03)	0.135	Ref	0.89(0.78, 1.02)	0.90(0.78, 1.03)	0.133
Model 2	Ref	0.97(0.84, 1.11)	0.92(0.80, 1.07)	0.281	Ref	0.87(0.76, 0.99)	0.86(0.75, 0.99)	0.033
Model 3	Ref	0.97(0.84, 1.11)	0.92(0.80, 1.07)	0.289	Ref	0.87(0.76, 0.99)	0.86(0.75, 0.99)	0.033
Model 4	Ref	0.98(0.84, 1.13)	0.93(0.80, 1.08)	0.354	Ref	0.87(0.76, 1.00)	0.86(0.75, 0.99)	0.034

Acronyms: T1 = First tertile (lowest adherence); T2 = Second tertile; T3 = third tertile (highest adherence); DP1 = Dietary pattern 1; DP2 = dietary pattern 2; CRC = colorectal cancer; PLCO = Prostate, Lung, Colorectal, and Ovarian; Ref = reference

Model 1: Adjusted for age, sex, race, marital status, educational status, occupational status, smoking status, family history of cancer other than CRC, family history of CRC, colon comorbidities, diverticulitis, liver comorbidities, colorectal polyps, and physical activity

Model 2: Adjusted for variables in model 1 plus hypertension, diabetes, body mass index

Model 3: Adjusted for variables in model 2 plus aspirin use

Model 4: Adjusted for variables in model 3 plus total energy intake

Table 3 Sub-distribution hazard ratio (95%CI) for mortality of CRC by tertiles of DP1 and DP2 among study participants in PLCO study

Models	DP1				DP2			*P*-trend
	T1	T2	T3	*P*-trend	T1	T2	T3	
			HR a (95% CI)					
Model 1	Ref	0.90(0.72, 1.13)	0.92(0.73, 1.15)	0.377	Ref	0.98(0.79, 1.21)	0.93(0.74, 1.16)	0.5326
Model 2	Ref	0.92(0.73, 1.14)	0.94(0.75, 1.19)	0.1507	Ref	0.97(0.78, 1.20)	0.91((0.72, 1.14)	0.7285
Model 3	Ref	0.91(0.73, 1.14)	0.94(0.75, 1.19)	0.1626	Ref	0.97(0.78, 1.20)	0.91(0.72, 1.13)	0.7263
Model 4	Ref	0.93(0.73, 1.18)	0.95(0.75, 1.21)	0.2189	Ref	0.97(0.78, 1.21)	0.91(0.72, 1.14)	0.7356

Acronyms: T1 = First tertile (lowest adherence); T2 = Second tertile; T3 = third tertile (highest adherence); DP1 = Dietary pattern 1; DP2 = dietary pattern 2; CRC = colorectal cancer; Ref = reference

Model 1: Adjusted for age, sex, race, marital status, educational status, occupational status, smoking status, family history of CRC, family history of cancer other than CRC, colon comorbidities, diverticulitis, liver comorbidities, colorectal polyps, and physical activity

Model 2: Adjusted for variables model 1 plus hypertension, diabetes and BMI

Model 3: Adjusted for variables in model 2 plus Aspirin use

Model 4: Adjusted for variables in model 3 plus total energy intake

which were high in fibre, folate, and carbohydrate, and low saturated and unsaturated fatty acids showed no significant risk association with CRC incidence. In addition, neither dietary pattern displayed any risk associations with rectal cancer as well as CRC mortality. A sensitivity analysis provided further validation for the robustness of the association between the two dietary patterns and the risk of CRC. These findings provide evidence that dietary intake characterized by higher intake of dietary fat oil and solid, dark green vegetables, and other vegetables, and a lower intake of sugar, beer, and liquor could potentially contribute to a lowered incidence rate of CRC.

Strengths and limitations

These analyses featured a larger sample size, longer follow-up period, comprehensive dietary information, and the inclusion of adjustments for multiple confounding variables. However, it has also the following limitations. First, while the DHQ is used to assess an individual's typical dietary intake, there is potential for social desirability or recall bias in the types and amounts of food consumed. Second, dietary intake was evaluated only once, with no consideration given to potential changes in dietary patterns over time. Third, the average age of the participants in our study corresponds to late middle-age, and it is unclear whether the findings are generalizable to younger adults. Considering the potential impact of particular dietary patterns on the rising incidence of young onset CRC, further research specifically targeting this demographic is warranted to better understand the relationship between dietary pattern and CRC risk in younger adults. Finally, RRR employs intermediate response variables to establish dietary patterns. Our analysis specifically incorporates fibre, folate, and the percentage of energy from fat and carbohydrates to investigate the correlation between dietary intake and CRC. However, the potential impact of food groups with alternative pathways influencing CRC risk remains unexplored in our study and is an area for future consideration in research.

Comparison with other studies

The association between RRR-derived dietary patterns and CRC risk and mortality has been infrequently considered. At the time of this analysis, we identified three studies on RRR-derived dietary patterns and CRC [84–86]. None of these studies were directly comparable to our findings due to the difference of used intermediate response variables to derive dietary patterns, reflecting broader challenges with comparing findings. Willemsen et al. [84] examined dietary fibre, vitamin D, fructose, and discretionary fats, and found adherence to a high fibre and discretionary fats dietary pattern reduced risk of colon cancer, while fructose pattern showed no association with risk of colon cancer. This may compliment our findings that the high fibre, and unsaturated fatty acid dietary pattern reduced the risk of CRC and colon cancer. Additionally, Cho et al. [85] and Fung et al. [86] conducted two studies, using plasma C-reactive protein (CRP) and serum C-peptide concentrations as response variables, respectively. Their results revealed a robust association between dietary patterns derived from both CRP and C-peptide and CRC. The use of biomarkers of such as CRP and C-peptide to derive dietary patterns as an intermediate response variable may better explain the association between diet and CRC compared to utilizing macronutrients as intermediate response variables [87].

In the current study, the high carbohydrate/fibre/folate, and low fatty acid dietary pattern showed no association with CRC risk. There is strong evidence about the protective effect of dietary fibre on CRC risk [11, 88] which is why dietary guidelines including dietary guidelines for Americans [30] and World Cancer Research Fund [89] recommend adequate intake of dietary fibre. In contrast, other studies have shown higher consumption of non-fibre carbohydrates (refined sugar, foods with high glycaemic load and index and simple carbohydrates) are associated with an increased risk of CRC [90–93]. This suggests that increased fibre consumption might offset the impact of simple carbohydrates on CRC risk, or vice versa.

We also found a high-fibre and unsaturated fatty acid dietary pattern was associated with lower risk of CRC and colon cancer. This finding is supported by a systematic review of observational studies, which concluded that high dietary intake of docosahexaenoic acid, docosapentaenoic acid, and eicosapentaenoic acid serve as protective factors against CRC [94]. In contrast to our results, another systematic review investigating the relationship between long-chain omega-3 (LCn3), alpha-linolenic acid (ALA), omega-6, and total polyunsaturated fatty acid (PUFA) consumption and cancer risk found that increasing LCn3 and ALA had minimal to no impact on cancer risk, while total PUFA intake might slightly elevate cancer risk [95]. This review specifically included only randomized control trials (RCTs) that compared higher versus lower levels of LCn3, ALA, omega-6, and total PUFA through interventions involving food, supplements, or advice to modify the intake of these fatty acids for at least during the study duration. The included RCTs often focussed on isolated nutrients or specific dietary components, which may not fully capture the complexity of real-world dietary patterns [96, 97]. Nutrients interact in a synergistic manner, and RCTs focusing on single components with relatively short follow-up periods might not represent the holistic nature of dietary influences on health [98]. Thus, further studies will be needed using the same intermediate response variables to confirm our findings.

Potential mechanisms of dietary intake and CRC

The relationship between dietary intake and CRC is complex and explained by the effect of dietary intake on various biological processes, including inflammation, microbiota composition, insulin resistance and oxidative stress [69]. Certain foods, such as processed foods, refined sugars, trans fats, and red or processed meats, are known to induce inflammation [99]. These foods can trigger an inflammatory

response in the body, contributing to chronic inflammation [100]. Conversely, a diet rich in anti-inflammatory foods like fruits, vegetables, whole grains, fatty fish (abundant in omega-3 fatty acids), nuts, seeds, and olive oil can mitigate inflammation, reducing the risk of chronic diseases, including CRC [70].

Complex carbohydrates, especially those from whole grains, legumes, fruits, and vegetables, have been linked to lower levels of inflammation and insulin resistance [101–103]. Fibre, a type of complex carbohydrate, supports a healthy gut microbiota, which is crucial for regulating the immune system and inflammation [67, 104]. Diets with high fruit and vegetable content, rich in fibre, stimulate the production of butyrate, a short-chain fatty acid with anti-inflammatory and anti-cancer properties [105]. Similarly, antioxidant-rich foods, such as berries, leafy greens, fruits, and vegetables, neutralize free radicals in the body, reducing oxidative stress and the associated risk of CRC [106]. Free radicals contribute to CRC risk through oxidative stress, which is closely linked to inflammation [107].

The types and amounts of food a person consumes influence the growth and activities of various microbes in the gut. Diets rich in fibre (found in fruits, vegetables, whole grains, and legumes), promote the growth of beneficial bacteria in the gut [108]. The gut microbiome plays a crucial role in reducing the risk of CRC by engaging with the normal gut barrier system, fostering the survival of epithelial cells, and, notably, safeguarding the body against external or opportunistic and harmful pathogens [109].

Diets rich in highly processed and refined carbohydrates, such as sugary snacks, white bread, and sugary beverages, can contribute to insulin resistance. Insulin resistance increase the risk of cancer through activation of insulin like growth factor 1 (IGF-1) [110]. On the other hand, consumption of diets rich in salmon fish, almonds, walnuts, and flaxseeds, turmeric, ginger, garlic, berries, citrus fruits, leafy greens, and colourful vegetables may contribute less risk of CRC [111]. These food items are rich in antioxidants such as vitamins C and E, carotenoids and polyphenols [112], which neutralize free radicals, thereby reducing oxidative stress and the risk of CRC [113].

Conclusion

The findings of this study underline that diets rich in fibre, folate, and unsaturated fatty acids intake may reduce the risk of CRC. However, there is no observed association between dietary patterns and CRC mortality. Our findings align with existing recommendations regarding the benefits of consuming food items rich in fibre and unsaturated fatty acids to reduce the risk of CRC. These findings suggest that adherence to dietary pattern characterized by higher intake of dietary fat oil and solid, dark green vegetables, and other vegetables, and a lower intake of sugar, beer, and liquor could potentially reduce the risk of CRC occurrence. The findings hold significance for public health, highlighting the need to promote nutrition education and counselling, particularly for individuals at risk of developing CRC due to unhealthy dietary habits. Policymakers and public health initiatives can strengthen dietary recommendations by promoting the consumption of foods rich in fibre and unsaturated fats through awareness creation and increasing the availability of such food options. Future studies incorporating repeated measurement of dietary intake, encompassing diverse age groups, and using the same intermediate response variables are needed to confirm these findings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00394-024-03513-9.

Acknowledgements Zegeye Abebe is thankful for the scholarship provided by the Australian Government Research Training Program. The authors express their gratitude to the National Cancer Institute for providing access to the data collected during the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial (PLCO-1166). The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

Author contributions ZA, MMW, PDN, ACR, YAM designed the analysis; ZA conducted the analysis and wrote the draft manuscript; MMW, ACR and YAM revised and edited the manuscript. All authors read and approved the final manuscript.

Funding No specific fund was secured for this study. YAM and MMW are supported by a National Health and Medical Research Council of Australia (NHMRC) Investigator Grants (2009776 and 2009050, respectively).

Data availability The data presented in this manuscript can be accessed from the NCI upon request.

Declarations

Ethical approval The usage of the CRC data in the PLCO study was authorised by the National Cancer Institute (PLCO-1166) and the Human Research Ethics Committee of Flinders University (project number 6435).

Competing interests All authors declare that they have no competing interests.

References

1. Rawla P, Sunkara T, Barsouk A (2019) Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 14(2):89–103
2. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A (2023) Colorectal cancer statistics, 2023. Cancer J Clin 73(3):233–254

3. Xi Y, Xu P (2021) Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol 14(10):101174
4. Valle L (2014) Genetic predisposition to colorectal cancer: where we stand and future perspectives. World J Gastroenterology: WJG 20(29):9828
5. Munteanu I, Mastalier B (2014) Genetics of colorectal cancer. J Med Life 7(4):507
6. Shaw E, Farris MS, Stone CR, Derksen JW, Johnson R, Hilsden RJ et al (2018) Effects of physical activity on colorectal cancer risk among family history and body mass index subgroups: a systematic review and meta-analysis. BMC Cancer 18:1–15
7. Parsa N (2012) Environmental factors inducing human cancers. Iran J Public Health 41(11):1
8. Elizabeth L, Machado P, Zinöcker M, Baker P, Lawrence M (2020) Ultra-processed foods and health outcomes: a narrative review. Nutrients 12(7):1955
9. Lotfi K, Salari-Moghaddam A, Yousefinia M, Larijani B, Esmaillzadeh A (2021) Dietary intakes of monounsaturated fatty acids and risk of mortality from all causes, cardiovascular disease and cancer: a systematic review and dose-response meta-analysis of prospective cohort studies. Ageing Res Rev 72:101467
10. Hu J, Wang J, Li Y, Xue K, Kan J (2023) Use of Dietary fibers in reducing the risk of several Cancer types: an Umbrella Review. Nutrients 15(11):2545
11. Arayici ME, Mert-Ozupek N, Yalcin F, Basbinar Y, Ellidokuz H (2022) Soluble and insoluble Dietary Fiber Consumption and Colorectal Cancer risk: a systematic review and Meta-analysis. Nutr Cancer 74(7):2412–2425
12. Fu H, He J, Li C, Deng Z, Chang H (2023) Folate intake and risk of colorectal cancer: a systematic review and up-to-date meta-analysis of prospective studies. Eur J Cancer Prev 32(2):103–112
13. Agnoli C, Pounis G, Krogh V (2019) Dietary pattern analysis. Analysis in Nutrition Research: Elsevier; pp. 75–101
14. Part D (2020) Chapter 8: Dietary Patterns. Guidelines Advisory Committee. 2020-07
15. Zhao J, Li Z, Gao Q, Zhao H, Chen S, Huang L et al (2021) A review of statistical methods for dietary pattern analysis. Nutr J 20(1):1–18
16. Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 13(1):3–9
17. Hoffmann K, Zyriax B-C, Boeing H, Windler E (2004) A dietary pattern derived to explain biomarker variation is strongly associated with the risk of coronary artery disease. Am J Clin Nutr 80(3):633–640
18. Moazzen S, van der Sloot KWJ, Bock GHd, Alizadeh BZ (2021) Systematic review and meta-analysis of diet quality and colorectal cancer risk: is the evidence of sufficient quality to develop recommendations? Crit Rev Food Sci Nutr 61(16):2773–2782
19. Steck SE, Guinter M, Zheng J, Thomson CA (2015) Index-based dietary patterns and colorectal cancer risk: a systematic review. Adv Nutr 6(6):763–773
20. Morze J, Danielewicz A, Przybyłowicz K, Zeng H, Hoffmann G, Schwingshackl L (2021) An updated systematic review and meta-analysis on adherence to mediterranean diet and risk of cancer. Eur J Nutr 60(3):1561–1586
21. Mohseni R, Mohseni F, Alizadeh S, Abbasi S (2020) The association of dietary approaches to stop hypertension (DASH) diet with the risk of colorectal cancer: a meta-analysis of observational studies. Nutr Cancer 72(5):778–790
22. Yusof AS, Isa ZM, Shah SA (2012) Dietary patterns and risk of colorectal cancer: a systematic review of cohort studies (2000–2011). Asian Pac J Cancer Prev 13(9):4713–4717
23. Feng Y-L, Shu L, Zheng P-F, Zhang X-Y, Si C-J, Yu X-L et al (2017) Dietary patterns and colorectal cancer risk: a meta-analysis. Eur J Cancer Prev 26(3):201–211
24. Kesse E, Clavel-Chapelon F, Boutron-Ruault M-C (2006) Dietary patterns and risk of colorectal tumors: a cohort of French women of the National Education System (E3N). Am J Epidemiol 164(11):1085–1093
25. Jafari Nasab S, Ghanavati M, Bahrami A, Rafiee P, Sadeghi A, Clark CC et al (2021) Dietary nutrient patterns and the risk of colorectal cancer and colorectal adenomas: a case-control study. Eur J Cancer Prev 30(1):46–52
26. Tucker KL (2010) Dietary patterns, approaches, and multicultural perspective. Appl Physiol Nutr Metab 35(2):211–218
27. Schulz CA, Oluwagbemigun K, Nöthlings U (2021) Advances in dietary pattern analysis in nutritional epidemiology. Eur J Nutr 60(8):4115–4130. https://doi.org/10.1007/s00394-021-02545-9
28. Pot GK, Stephen AM, Dahm CC, Key TJ, Cairns BJ, Burley VJ et al (2014) Dietary patterns derived with multiple methods from food diaries and breast cancer risk in the UK Dietary Cohort Consortium. Eur J Clin Nutr 68(12):1353–1358
29. Melaku YA, Gill TK, Taylor AW, Adams R, Shi Z (2018) A comparison of principal component analysis, partial least-squares and reduced-rank regressions in the identification of dietary patterns associated with bone mass in ageing australians. Eur J Nutr 57:1969–1983
30. U.S. Department of Agriculture and U.S. Department of Health and Human Services (2023) Dietary Guidelines for Americans, 2020–2025. 9th Edition. https://www.dietaryguidelines.gov/ available on September 21
31. Fan L, Cai Y, Wang H, Zhang H, Chen C, Zhang M, Lu Z, Li Y, Zhang F, Ning C, Wang W (2023) Saturated fatty acid intake, genetic risk and colorectal cancer incidence: a large-scale prospective cohort study. Int J Cancer 153(3):499–511
32. Fan L, Cai Y, Wang H, Zhang H, Chen C, Zhang M et al (2023) Saturated fatty acid intake, genetic risk and colorectal cancer incidence: a large-scale prospective cohort study. Int J Cancer 153(3):499–511. https://doi.org/10.1002/ijc.34544
33. Shekari S, Fathi S, Roumi Z, Akbari ME, Tajadod S, Afsharfar M et al (2022) Association between dietary intake of fatty acids and colorectal cancer, a case-control study. Front Nutr 9:856408
34. Xu AA, Kennedy LK, Hoffman K, White DL, Kanwal F, El-Serag HB et al (2022) Dietary fatty acid intake and the colonic gut microbiota in humans. Nutrients 14(13):2722
35. Black A, Huang W-Y, Wright P, Riley T, Mabie J, Mathew S et al (2015) PLCO: evolution of an epidemiologic resource and opportunities for future studies. Rev Recen Clin Trial 10(3):238–245
36. Guinter MA, McLain AC, Merchant AT, Sandler DP, Steck SE (2018) A dietary pattern based on estrogen metabolism is associated with breast cancer risk in a prospective cohort of postmenopausal women. Int J Cancer 143(3):580–590
37. Xiao Y, Wang Y, Gu H, Xu Z, Tang Y, He H et al (2023) Adherence to the Paleolithic diet and paleolithic-like lifestyle reduce the risk of colorectal cancer in the United States: a prospective cohort study. J Translational Med 21(1):482
38. Zhong G-C, Li Z, You A-J, Zhu Q, Wang C-R, Yang P-F (2023) Plant-based diets and the risk of pancreatic cancer: a large prospective multicenter study. Am J Clin Nutr 117(2):235–242
39. Li Z, Wang K, Shivappa N, Hébert JR, Chen H, Liu H et al (2022) Inflammatory potential of diet and colorectal carcinogenesis: a prospective longitudinal cohort. Br J Cancer 126(12):1735–1743
40. Agricultural Health Study Questionnaires & Study Data. https://aghealth.nih.gov/collaboration/questionnaires.html. Access on 11 December 2023 [
41. National Cancer Institute. PLCO - the cancer data access system, Washington DC National Cancer Institute https://cdas.cancer.gov/learn/plco/early-qx/ [
42. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S et al (2001) Comparative validation of the Block, Willett, and National Cancer Institute food frequency

questionnaires: the eating at America's table study. Am J Epidemiol 154(12):1089–1099
43. Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB et al (2002) Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 102(2):212–225
44. Tippett Kand Cypel Y (1997) Design, Operation The Continuing Survey of Food Intakes by individuals and the Diet and Health Knowledge Survey, 1994–96. US Department of Agriculture, Springfiled. Agricultural Research Service, Nationwide Food Surveys Report. 96–1
45. Bowman SA, Friday JE, Moshfegh AJ (2008) MyPyramid Equivalents Database, 2.0 for USDA survey foods, 2003–2004: documentation and user guide. US Department of Agriculture
46. ClinicalTrials.gov (2022) Screening for Colorectal Cancer in Older Patients (PLCO Screening Trial). https://classic.clinicaltrials.gov/ct2/show/study/NCT01696981#wrapper
47. Kitahara CM, Trabert B, Katki HA, Chaturvedi AK, Kemp TJ, Pinto LA et al (2014) Body mass index, physical activity, and serum markers of inflammation, immunity, and insulin resistance. Cancer Epidemiol Biomarkers Prev 23(12):2840–2849
48. Wolin KY, GRUBB R III, Ragard L, Mabie J, Andriole GL et al (2015) Physical activity and benign prostatic hyperplasia-related outcomes and nocturia. Med Sci Sports Exerc 47(3):581
49. WHO G. Global physical activity questionnaire (GPAQ) analysis guide. Geneva: World Health Organization (2012) 1–22
50. Virostko J, Capasso A, Yankeelov TE, Goodgame B (2019) Recent trends in the age at diagnosis of colorectal cancer in the US National Cancer Data Base, 2004-2015. Cancer 125(21):3828–3835
51. Baraibar I, Ros J, Saoudi N, Salvà F, García A, Castells M et al (2023) Sex and gender perspectives in colorectal cancer. ESMO open 8(2):101204
52. Liu L, Shi Y, Li T, Qin Q, Yin J, Pang S et al (2016) Leisure time physical activity and cancer risk: evaluation of the WHO's recommendation based on 126 high-quality epidemiological studies. Br J Sports Med 50(6):372–378
53. Komaki Y, Komaki F, Micic D, Ido A, Sakuraba A (2017) Risk of colorectal cancer in chronic liver diseases: a systematic review and meta-analysis. Gastrointest Endosc 86(1):93–104 e5
54. Johnston L, Carey F (2020) Pathology of colorectal polyps and cancer. Surg (Oxford) 38(1):12–17
55. Shu L, Shen X-M, Li C, Zhang X-Y, Zheng P-F (2017) Dietary patterns are associated with type 2 diabetes mellitus among middle-aged adults in Zhejiang Province, China. Nutr J 16(1):1–9
56. Ndanuko RN, Tapsell LC, Charlton KE, Neale EP, Batterham MJ (2016) Dietary patterns and blood pressure in adults: a systematic review and meta-analysis of randomized controlled trials. Adv Nutr 7(1):76–89
57. Min M, Li-Fa X, Dong H, Jing W, Ming-Jie B (2017) Dietary patterns and overweight/obesity: a review article. Iran J Public Health 46(7):869
58. Mandic M, Li H, Safizadeh F, Niedermaier T, Hoffmeister M, Brenner H (2023) Is the association of overweight and obesity with colorectal cancer underestimated? An umbrella review of systematic reviews and meta-analyses. Eur J Epidemiol 38(2):135–144
59. Ta HDK, Nguyen NN, Ho DKN, Nguyen HD, Ni Y-C, Yee KX et al (2023) Association of diabetes mellitus with early-onset colorectal cancer: A systematic review and meta-analysis of 19 studies including 10 million individuals and 30,000 events. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 17(8):102828
60. Xuan K, Zhao T, Sun C, Patel AS, Liu H, Chen X et al (2021) The association between hypertension and colorectal cancer: a meta-analysis of observational studies. Eur J Cancer Prev 30(1):84–96
61. World Health Organization (2023) Saturated fatty acid and trans-fatty acid intake for adults and children: WHO guideline. Saturated fatty acid and trans-fatty acid intake for adults and children. WHO guideline
62. Humphreys KJ, Conlon MA, Young GP, Topping DL, Hu Y, Winter JM et al (2014) Dietary manipulation of oncogenic microRNA expression in human rectal mucosa: a randomized trial. Cancer Prev Res 7(8):786–795
63. Seidelmann SB, Claggett B, Cheng S, Henglin M, Shah A, Steffen LM et al (2018) Dietary carbohydrate intake and mortality: a prospective cohort study and meta-analysis. Lancet Public Health 3(9):e419–e28
64. World Cancer Research Fund International (2018) Diet, nutrition, physical activity and cancer: a global perspective: a summary of the Third Expert Report. World Cancer Research Fund International
65. Włodarczyk J, Włodarczyk M, Zielińska M, Jędrzejczak B, Dziki Ł, Fichna J (2021) Blockade of fructose transporter protein GLUT5 inhibits proliferation of colon cancer cells: proof of concept for a new class of anti-tumor therapeutics. Pharmacol Rep 73(3):939–945
66. Jiao J, Xu J-Y, Zhang W, Han S, Qin L-Q (2015) Effect of dietary fiber on circulating C-reactive protein in overweight and obese adults: a meta-analysis of randomized controlled trials. Int J Food Sci Nutr 66(1):114–119
67. Krishnamurthy VMR, Wei G, Baird BC, Murtaugh M, Chonchol MB, Raphael KL et al (2012) High dietary fiber intake is associated with decreased inflammation and all-cause mortality in patients with chronic kidney disease. Kidney Int 81(3):300–306
68. Song M, Garrett WS, Chan AT (2015) Nutrients, foods, and colorectal cancer prevention. Gastroenterology 148(6):1244–1260 e16
69. Bojková B, Winklewski PJ, Wszedybyl-Winklewska M (2020) Dietary fat and cancer—which is good, which is bad, and the body of evidence. Int J Mol Sci 21(11):4114
70. Gleissman H, Johnsen JI, Kogner P (2010) Omega-3 fatty acids in cancer, the protectors of good and the killers of evil? Exp Cell Res 316(8):1365–1373
71. Denova-Gutierrez E, Tucker KL, Flores M, Barquera S, Salmeron J (2016) Dietary patterns are associated with predicted cardiovascular disease risk in an urban Mexican adult population. J Nutr 146(1):90–97
72. Abdel-Qadir H, Fang J, Lee DS, Tu JV, Amir E, Austin PC et al (2018) Importance of considering competing risks in time-to-event analyses: application to stroke risk in a retrospective cohort study of elderly patients with atrial fibrillation. Circulation: Cardiovasc Qual Outcomes 11(7):e004580
73. Nobbs HM, Yaxley A, Thomas J, Delaney C, Koczwara B, Luszcz M et al (2016) Do dietary patterns in older age influence the development of cancer and cardiovascular disease: a longitudinal study of ageing. Clin Nutr 35(2):528–535
74. White IR, Royston P (2009) Imputing missing covariate values for the Cox model. Stat Med 28(15):1982–1998
75. Andersen K, Mariosa D, Adami H-O, Held C, Ingelsson E, Lagerros YT et al (2014) Dose–response relationship of total and leisure time physical activity to risk of heart failure: a prospective cohort study. Circulation: Heart Fail 7(5):701–708
76. Cornish R, Macleod J, Carpenter J, Tilling K (2017) Multiple imputation using linked proxy outcome data resulted in important bias reduction and efficiency gains: a simulation study. Emerg Themes Epidemiol 14(1):1–13
77. Graham JW, Olchowski AE, Gilreath TD (2007) How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 8:206–213
78. Little R, Rubin D (1987) Multiple imputation for nonresponse in surveys. John Wiley Sons Inc Doi 10:9780470316696

79. Jakobsen JC, Gluud C, Wetterslev J, Winkel P (2017) When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. BMC Med Res Methodol 17(1):1–10
80. StataCorp (2021) Stata: Release 17. Statistical Software. StataCorp LLC, College Station, TX
81. Posit team. RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston MA (2023) URL http://www.posit.co/ [
82. Van Buuren S, Groothuis-Oudshoorn K (2011) Mice: Multivariate imputation by chained equations in R. J Stat Softw 45:1–67
83. Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94(446):496–509
84. Willemsen RF, McNeil J, Heer E, Johnson ST, Friedenreich CM, Brenner DR (2022) Dietary patterns with combined and site-specific cancer incidence in Alberta's tomorrow project cohort. Eur J Clin Nutr 76(3):360–372
85. Cho YA, Lee J, Oh JH, Chang HJ, Sohn DK, Shin A et al (2018) Inflammatory dietary pattern, IL-17F genetic variant, and the risk of Colorectal Cancer. Nutrients. 10(6)
86. Fung TT, Hu FB, Schulze M, Pollak M, Wu T, Fuchs CS et al (2012) A dietary pattern that is associated with C-peptide and risk of colorectal cancer in women. Cancer Causes Control 23:959–965
87. Weikert C, Schulze MB (2016) Evaluating dietary patterns: the role of reduced rank regression. Curr Opin Clin Nutr Metab Care 19(5):341–346
88. Xie Y, Ren YW, Li XL, Li ZN (2020) A review of the associations between Dietary Fiber Intake and Cancer Prevention or Prognosis. J Nutritional Oncol 5(3):123–131
89. World Cancer Research Fund/American Institute for Cancer Research. Diet, nutrition, physical activity and cancer: aglobal perspective. Continuous Update Project Expert Report (2018) https://www.wcrf.org/wp-content/uploads/2021/02/Summary-of-Third-Expert-Report-2018.pdf accessed on April 14, 2022
90. Grasgruber P, Hrazdira E, Sebera M, Kalina T (2018) Cancer incidence in Europe: an ecological analysis of nutritional and other environmental factors. Front Oncol 8:151
91. Zelenskiy S, Thompson CL, Tucker TC, Li L (2014) High dietary glycemic load is associated with increased risk of colon cancer. Nutr Cancer 66(3):362–368
92. Higginbotham S, Zhang Z-F, Lee I-M, Cook NR, Giovannucci E, Buring JE et al (2004) Dietary glycemic load and risk of colorectal cancer in the women's Health Study. J Natl Cancer Inst 96(3):229–233
93. Sieri S, Krogh V, Agnoli C, Ricceri F, Palli D, Masala G et al (2015) Dietary glycemic index and glycemic load and risk of colorectal cancer: results from the EPIC-Italy study. Int J Cancer 136(12):2923–2931
94. Lu Y, Li D, Wang L, Zhang H, Jiang F, Zhang R et al (2023) Comprehensive Investigation on associations between Dietary Intake and blood levels of fatty acids and colorectal Cancer risk. Nutrients 15(3):730
95. Hanson S, Thorpe G, Winstanley L, Abdelhamid AS, Hooper L (2020) Omega-3, omega-6 and total dietary polyunsaturated fat on cancer incidence: systematic review and meta-analysis of randomised trials. Br J Cancer 122(8):1260–1270
96. Hébert JR, Frongillo EA, Adams SA, Turner-McGrievy GM, Hurley TG, Miller DR et al (2016) Perspective: randomized controlled trials are not a panacea for diet-related research. Adv Nutr 7(3):423–432
97. Zeilstra D, Younes JA, Brummer RJ, Kleerebezem M (2018) Perspective: fundamental limitations of the randomized controlled trial method in nutritional research: the example of probiotics. Adv Nutr 9(5):561–571
98. Ioannidis JP (2019) Unreformed nutritional epidemiology: a lamp post in the dark forest. Eur J Epidemiol 34:327–331
99. Tristan Asensi M, Napoletano A, Sofi F, Dinu M (2023) Low-Grade inflammation and Ultra-processed Foods Consumption: a review. Nutrients 15:1546
100. Clemente-Suárez VJ, Beltrán-Velasco AI, Redondo-Flórez L, Martín-Rodríguez A, Tornero-Aguilera JF (2023) Global impacts of western Diet and its effects on Metabolism and Health: a narrative review. Nutrients 15(12):2749
101. Chassard C, Lacroix C (2013) Carbohydrates and the human gut microbiota. Curr Opin Clin Nutr Metabolic Care 16(4):453–460
102. Oliphant K, Allen-Vercoe E (2019) Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. Microbiome. 7: 91. PUBMED; 2019
103. Florea N, Perde D, Shami A (2022) ROLE OF THE GUT MICROBIOTA IN NUTRITION AND HEALTH. Sci Collect «InterConf» 107:363–366
104. Mora-Flores LP, Moreno-Terrazas Casildo R, Fuentes-Cabrera J, Pérez-Vicente HA, de Anda-Jáuregui G, Neri-Torres EE (2023) The role of Carbohydrate Intake on the gut microbiome: a weight of evidence systematic review. Microorganisms 11(7):1728
105. Cheng X, Zheng J, Lin A, Xia H, Zhang Z, Gao Q et al (2020) A review: roles of carbohydrates in human diseases through regulation of imbalanced intestinal microbiota. J Funct Foods 74:104197
106. Zehiroglu C, Ozturk Sarikaya SB (2019) The importance of antioxidants and place in today's scientific and technological studies. J Food Sci Technol 56:4757–4774
107. Bardelčíková A, Šoltys J, Mojžiš J (2023) Oxidative stress, inflammation and colorectal Cancer: an overview. Antioxidants 12(4):901
108. Zhang P (2022) Influence of foods and nutrition on the gut microbiome and implications for intestinal health. Int J Mol Sci 23(17):9588
109. Tremaroli V, Bäckhed F (2012) Functional interactions between the gut microbiota and host metabolism. Nature 489(7415):242–249
110. Komninou D, Ayonote A, Richie JP Jr, Rigas B (2003) Insulin resistance and its contribution to colon carcinogenesis. Experimental Biology Med 228(4):396–405
111. Halvorsen BL, Carlsen MH, Phillips KM, Bøhn SK, Holte K, Jacobs DR Jr et al (2006) Content of redox-active compounds (ie, antioxidants) in foods consumed in the United States. Am J Clin Nutr 84(1):95–135
112. Rahaman MM, Hossain R, Herrera-Bravo J, Islam MT, Atolani O, Adeyemi OS et al (2023) Natural antioxidants from some fruits, seeds, foods, natural products, and associated health benefits: an update. Food Sci Nutr 11(4):1657–1670
113. Chaudhary P, Janmeda P, Docea AO, Yeskaliyeva B, Abdull Razis AF, Modu B et al (2023) Oxidative stress, free radicals and antioxidants: potential crosstalk in the pathophysiology of human diseases. Front Chem 11:1158198