## ARTICLE

Epidemiology

# Dietary patterns with combined and site-specific cancer incidence in Alberta's Tomorrow Project cohort

Romy F. Willemsen[1,2], Jessica McNeil[2,3], Emily Heer[2], Steven T. Johnson[4], Christine M. Friedenreich[2,5] and Darren R. Brenner 🆔 [2,5 ✉]

**BACKGROUND/OBJECTIVES:** Poor diet quality has been associated with an increased risk of cancer. Here, we examine the association between dietary patterns derived with two methods, and combined and site-specific cancer incidence in Canada.
**SUBJECTS/METHODS:** Dietary data were obtained from participants enrolled in Alberta's Tomorrow Project, a prospective cohort study, between 2000 and 2008. Principle component analysis (PCA) and reduced rank regression (RRR) were used to derive dietary patterns, and data linkage with the Alberta Cancer Registry was used for incident cancer cases. Cox proportional hazard regressions were used to estimate multivariable-adjusted models for the association between each dietary pattern score with combined and site-specific cancer incidence.
**RESULTS:** PCA revealed three dietary patterns ("western", "prudent", and "sugar, fruits, and dairy") and RRR resulted in four patterns ("dietary fiber", "vitamin D", "fructose", and "discretionary fat"). Five cancer sites were included in our site-specific analysis: lung, colon, breast, prostate, and endometrial cancers. The most protective dietary patterns for combined cancer sites were the "Prudent" pattern (HR = 0.82, CI = 0.73–0.92) and the "Dietary fiber" pattern (HR = 0.82, CI = 0.69–0.97). The "Fructose" pattern was associated with increased risk of combined cancers (HR = 1.14, CI = 1.02–1.27). Three dietary patterns were protective against colon cancer ("Prudent", "Dietary fiber", and "Discretionary fats"), and other risk reductions were seen for the "sugar, fruit, and dairy" pattern (lung cancer), and the "Dietary fiber" pattern (prostate cancer).
**CONCLUSIONS:** These results support cancer prevention strategies for a diet high in vegetables, fruits, fish, and whole grains. Further studies should explore the possible association between discretionary fats and colon cancer.

*European Journal of Clinical Nutrition* (2022) 76:360–372; https://doi.org/10.1038/s41430-021-00958-7

## INTRODUCTION

Almost 40% of all cancer cases can be attributable to modifiable risk factors [1], and a poor quality diet is linked to an increased risk of several cancer types [2]. Nutrients are often studied as active components in the diet-cancer relation, but strong associations can be difficult to detect because a food product or a nutrient constitutes only a small portion of the overall diet [3]. Hence, dietary patterns are proposed as an alternative approach to study diet-disease relations [4].

A dietary pattern is a complex collection of nutrients and food products, including the quantities and varieties of foods that comprise different diets [5, 6]. The World Cancer Research Fund (WCRF) recently concluded that there is some evidence that a healthy dietary pattern consisting of fruits, vegetables, and fish decreases cancer risk in the mouth, pharynx, and larynx [7], whereas a Western-type dietary pattern consisting of meat, potatoes, refined grains, and sugars increases the risk of obesity, which is an established risk factor for several cancers [7]. Previous research in Asia, Europe, and the United States has identified associations between dietary patterns and risk of

overall or site-specific cancers [4]; however, insufficient evidence exists regarding the etiologic associations between dietary patterns with combined and site-specific cancer incidence in Canada.

Alberta's Tomorrow Project (ATP) is a prospective cohort study designed to investigate the etiology of cancer and other chronic diseases. Information on lifestyle, behavioral, and environmental exposures was collected using surveys among participants in Alberta, Canada between 2001 and 2015 [8]. Using data from this cohort, we investigated the associations between a posteriori-derived dietary patterns using principal component analysis (PCA) and reduced rank regression (RRR) (a combination of a posteriori and a priori approach), with combined and site-specific cancer incidence in ATP participants.

## METHODS

### Study population

In total, 55,530 Albertans aged 35–69 years were enrolled in ATP, for which detailed information about participant recruitment, enrollment, and data

[1]Division of Human Nutrition and Health, Wageningen University, Wageningen, The Netherlands. [2]Department of Cancer Epidemiology and Prevention Research, CancerControl Alberta, Alberta Health Services, Calgary, AB, Canada. [3]Department of Kinesiology, Faculty of Health and Human Sciences, University of North Carolina Greensboro, Greensboro, North Carolina, USA. [4]Faculty of Health Disciplines, Athabasca University, Athabasca, AB, Canada. [5]Departments of Oncology and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ✉email: darren.brenner@ucalgary.ca

collection methods for Phase I have been previously described [8]. The current analysis included participants who completed the Canadian Diet History Questionnaire between 2000 and 2008 ($n = 26,462$). Ethics approval was obtained from the Health Research Ethics Board of Alberta – Cancer Committee (HREBA.CC-16-0495). Informed conset was obtained from all participants.

## Dietary intake assessment

The Canadian Diet History Questionnaire is a validated 124-item food frequency questionnaire of foods, beverages, and dietary supplements that assesses dietary intake over the preceding 12 months and is a modified version of the US National Cancer Institute's Diet History Questionnaire [9]. Modifications were made to the National Cancer Institute's Diet History Questionnaire to reflect food availability, brand names, nutrient composition, and food fortification in Canada [10]. Data on energy intake, 66 nutrients, and 30 food groups were available using a method described elsewhere [11, 12]. A previous validation study by Csizmadi et al. [13] compared the validity of the Canadian Diet History Questionnaire to 24 h food recalls collected by the Canadian Community Health Survey—Cycle 2.2. The study reported that the Canadian Diet History Questionnaire was able to adequately capture macro and micronutrient intakes collected from the 24 h recalls with ranges from 79–90% for carbohydrate intake (a proxy for fructose intake), total fat intake (a proxy for discretionary fat intake), and vitamin D and fiber.

## Cancer incidence rates

Data on primary cancers were obtained through a linkage with the Alberta Cancer Registry using participants' personal health numbers up to June 2018. The Alberta Cancer Registry is a population-based cancer registry that has over 95% case ascertainment. Site-specific cancers in this study were included if there was previous epidemiologic evidence and biologic plausibility for an association between dietary intake and cancer risk [7] and if at least 100 incident cases had occurred in the participants who had completed the Canadian Diet History Questionnaire at baseline. The threshold of 100 cancer cases for inclusion was used to achieve reasonable statistical power for the analyses as used in previous studies from this cohort [14]. There were sufficient cases accrued in this cohort to include separate analyses for breast, lung, colon, endometrium, and prostate cancers.

## Covariates

Information on covariates was obtained from the baseline Health and Lifestyle Questionnaire. Covariates were first added to the final multivariable models if they were known confounders in the association between dietary intake and cancer risk. These included age (years), sex (male/female for non-sex specific cancer sites), body mass index (BMI) (kg/m$^2$) and total energy intake (kCal). Furthermore, for lung cancer incidence, smoking was also forced into the model because of its known association with this cancer type. For the remaining covariates (smoking status: daily/occasional/former/never; physical activity: total MET hours/week; marital status: married or living with someone/divorced, separated or widowed/single, or never married), we assessed whether these variables changed the beta coefficient for the association between dietary patterns and cancer incidence by at least 10% when added to the model (ref). This led to the inclusion of the following additional covariates for each cancer site: smoking status for combined, breast, colon, endometrial, and prostate cancers; physical activity for combined, breast, lung, colon, endometrial, and prostate cancers; marital status for endometrial cancer.

## Follow-up time

Follow-up time was defined as the difference in years between age of cancer diagnosis (based on the exact age when primary cancer was diagnosed), or the end of follow-up for the present analyses (based on exact age at the time of data linkage with the Alberta Cancer Registry in June 2018) and the age of entry into the study (based on exact age at Canadian Diet History Questionnaire completion). The contributed person-time per participant is time point of completion of Canadian Diet History Questionnaire to diagnosis of cancer, or to the end of follow-up in June 2018, whichever came first.

## Dietary patterns

PCA is a data-driven approach for variable reduction that can be used to assess overall dietary patterns in contrast to large list of specific food variables [15]. In this application of PCA as a dimension-reduction technique, we seek to examine the impact of patterns within a regular diet on cancer risk, rather than a reductionist approach when examining single foods or nutrients. The output from PCA generates a smaller number of variables which are the principal components [15]. The principal components in this study are then examined as dietary patterns. Hence, these dietary patterns explain a considerable amount of the variation of specific food item intake within this population as possible. This approach is data driven in that it ignores any knowledge on plausible biologic associations with cancer incidence [15]. For this method, dietary patterns were derived using the 30 food groups generated with the Diet*Calc® software. Components (dietary patterns) were selected based on the eigenvalues (below or above 0.35) [16], scree plots (visual break), and their interpretability. Food groups with coefficients ≥0.35 were used to describe each component (dietary pattern) and provide the dietary pattern interpretation. The coefficients represent covariance between the foods and the overall dietary pattern. Foods with positive coefficients were positively associated with the dietary pattern, whereas the opposite was true for negative coefficients.

Orthogonal Varimax rotation was applied to reduce correlation and increase interpretability between the components (dietary patterns) [17]. Subsequently, for each participant, dietary pattern scores were estimated for every identified dietary pattern. Each specific dietary pattern has a score for each food group. For each individual, dietary pattern scores were calculated by multiplying the score for each of the 30 food groups by the intake of the food groups. These dietary pattern scores were then standardized and expressed as z-scores. The individual dietary pattern scores represent the alignment between that individual's diet and the derived dietary patterns.

RRR was the second statistical approach used to derive dietary patterns. For this method, disease-specific response variables (nutrients, combinations of nutrients, and biomarkers) are used to establish dietary patterns that explain as much variation in the population concerning that nutrient as possible [15, 18]. Response variables are nutrients selected based on scientific evidence [7] and biologic plausibility. For this study, we chose dietary fiber, vitamin D, fructose, and discretionary fats because: (a) current evidence on these nutrients in the WCRF report [7] (b) they are present in different types of foods and dietary patterns; (c) data were readily available; (d) the Food Frequency Questionnaire had been validated for these nutrients and validity was high and; (e) they are known to be associated with several specific cancers (e.g., colorectal and pancreatic cancers) [19–23]. Although these nutrients are linked to other site-specific cancers, evidence has been inconsistent and unsubstantial [19, 20, 24–28].

In RRR, dietary patterns are derived based on the variance in preselected nutrients (response variables), whereas in PCA, dietary patterns are derived based on the variance in food items and groups (predictor variables). The partial least squares procedure was used to perform the RRR and is explained in more detail elsewhere [15]. Briefly, the number of selected response variables (nutrients) dictates the number of derived dietary patterns, which was four in our study. RRR uses a covariance matrix of responses. Food groups with coefficients below or above |0.2| were used to describe the dietary pattern for each of the selected nutrients. These dietary pattern scores were standardized and expressed as z-scores.

## Statistical analysis

Baseline characteristics were estimated as means ± standard deviations (continuous variables) or as counts and percentages (categorical variables). Cox proportional hazard regression models were used to estimate multivariable-adjusted models for the association between each dietary pattern score derived by PCA and RRR with combined and site-specific cancer incidence. The multivariable model included adjustments as defined above, and an additional model accounting for reverse causality was created by excluding cancer cases diagnosed <2 years after Canadian Diet History Questionnaire completion (latency-adjusted multivariable model).

All statistical analyses were performed with SAS 9.4 (SAS Institute Inc., Cary, NC, USA).

## RESULTS

This analysis included 26,462 participants (10,109 males and 16,353 females), of which 2,721 (10.3%) developed cancer (1177 males and 1544 females) during the follow-up period. Mean follow-up time was $13.3 \pm 3.3$ years, which equaled 204,364 person-years (Table 1). The average age at baseline was $50.8 \pm 9.2$ years. The majority of participants had a BMI >25 kg/m$^2$ (66%),

**Table 1.** Baseline characteristics of Alberta's Tomorrow Project study participants ($n = 26{,}242$) by sex and case status, 2001–2008.

| Characteristic | Mean (SD) or N (Percent) | | | |
|---|---|---|---|---|
| | Males | | Females | |
| | Cancer Cases ($n = 1177$) | Non-Cancer Cases ($n = 8932$) | Cancer Cases ($n = 1544$) | Non-Cancer Cases ($n = 14{,}809$) |
| **Follow-up time (years)** | 7.88 (4.10) | 14.04 (2.51) | 7.88 (4.07) | 13.80 (2.52) |
| **Age (years)** | 57.10 (8.17) | 50.06 (8.95) | 55.47 (9.03) | 50.30 (9.11) |
| **Body mass index (kg/m$^2$)** | 31.27 (49.17) | 30.32 (47.22) | 32.61 (65.54) | 29.65 (48.87) |
| **Body mass index category** | | | | |
| *Body mass index < 25 kg/m2* | 223 (18.95%) | 2096 (23.47%) | 536 (34.72%) | 6074 (41.02%) |
| *Body mass index ≥ 25 kg/m2* | 951 (80.80%) | 6815 (76.30%) | 1001 (64.83%) | 8696 (58.72%) |
| *Missing* | 3 (0.25%) | 21 (0.24%) | 7 (0.45%) | 39 (0.26%) |
| **Marital Status** | | | | |
| *Married or living with someone* | 991 (84.20%) | 7424 (83.12%) | 1118 (72.41%) | 11,331 (76.51%) |
| *Divorced, separated, or widowed* | 124 (10.54%) | 921 (10.31%) | 332 (21.50%) | 2707 (18.28%) |
| *Single, never married* | 62 (5.27%) | 586 (6.56%) | 93 (6.02%) | 768 (5.19%) |
| *Missing* | - | 1 (0.01%) | 1 (0.06%) | 3 (0.02%) |
| **Smoking status** | | | | |
| *Never* | 389 (33.05%) | 3830 (42.88%) | 609 (39.44%) | 7013 (47.36%) |
| *Former* | 551 (46.81%) | 3501 (39.20%) | 557 (36.08%) | 5356 (36.17%) |
| *Occasional* | 31 (2.63%) | 320 (3.58%) | 41 (2.66%) | 422 (2.85%) |
| *Daily* | 206 (17.50%) | 1276 (14.29%) | 336 (21.76%) | 2003 (13.53%) |
| *Missing* | - | 5 (0.06%) | 1 (0.06%) | 15 (0.10%) |
| **Presence of at least one medical condition or co-morbidity** | | | | |
| *No* | 570 (48.43%) | 5229 (58.54%) | 861 (55.76%) | 9611 (64.90%) |
| *Yes* | 598 (50.81%) | 3656 (40.93%) | 672 (43.52%) | 5091 (34.38%) |
| *Missing* | 9 (0.76%) | 47 (0.53%) | 11 (0.71%) | 107 (0.72%) |
| **Family history of cancer** | | | | |
| *No* | 503 (42.74%) | 4469 (50.03%) | 595 (38.54%) | 6798 (45.90%) |
| *Yes* | 674 (57.26%) | 4463 (49.97%) | 949 (61.46%) | 8009 (54.08%) |
| *Missing* | - | - | - | 2 (0.01%) |
| Education | | | | |
| *High school or less* | 385 (32.71%) | 2163 (24.22%) | 548 (35.49%) | 4351 (29.38%) |
| *Some post-high school education or a post-high school degree or certificate* | 792 (67.29%) | 6768 (75.77%) | 996 (64.51%) | 10,455 (70.60%) |
| *Missing* | - | 1 (0.01%) | - | 3 (0.02%) |
| **Total household income** | | | | |
| *$0–$49,999* | 390 (33.14%) | 2037 (22.81%) | 716 (46.37%) | 5048 (34.09%) |
| *$50,000–$99,999* | 504 (42.82%) | 3989 (44.46%) | 548 (35.49%) | 5806 (49.21%) |
| *≥$100,000* | 262 (22.26%) | 2783 (31.16%) | 226 (14.64%) | 3529 (23.83%) |
| *Missing* | 21 (1.78%) | 123 (1.38%) | 54 (3.50%) | 426 (2.88%) |
| **Employment status** | | | | |
| *Working full time* | 662 (56.24%) | 6924 (77.52%) | 533 (34.52%) | 6642 (44.85%) |
| *Working part time* | 116 (9.86%) | 544 (6.09%) | 338 (21.89%) | 3431 (23.17%) |
| *Not employed, but looking for work* | 30 (2.55%) | 173 (1.94%) | 27 (1.75%) | 256 (1.73%) |
| *Homemaker* | 0 (0.00%) | 20 (0.22%) | 222 (14.38%) | 1911 (12.90%) |
| *Student* | 1 (0.08%) | 32 (0.36%) | 4 (0.26%) | 102 (0.69%) |
| *Retired* | 327 (27.78%) | 973 (10.89%) | 366 (23.70%) | 1963 (13.26%) |
| *Other* | 40 (3.40%) | 263 (2.94%) | 53 (3.43%) | 496 (3.35%) |
| *Missing* | 1 (0.08%) | 3 (0.03%) | 1 (0.06%) | 8 (0.05%) |
| **Total Physical activity (MET-hours/week)** | 151.45 (75.6) | 174.02 (74.91) | 144.34 (68.66) | 157.23 (65.01) |
| **Total energy intake (kcal)*** | 2176 (1011) | 2245 (1025) | 1595 (649) | 1645 (671) |
| Nutrients | | | | |
| *Dietary fiber (g)* | 20 (9) | 21 (10) | 18 (9) | 18 (9) |
| *Vitamin D (microg)* | 5 (4) | 5 (4) | 4 (3) | 4 (3) |
| *Fructose (g)* | 26 (19) | 28 (22) | 23 (18) | 23 (18) |
| *Saturated fatty acids (g)* | 27 (15) | 28 (15) | 19 (10) | 20 (10) |

**Table 1** continued

| Characteristic | Mean (SD) or *N* (Percent) | | | |
|---|---|---|---|---|
| | **Males** | | **Females** | |
| | **Cancer Cases** (*n* = 1177) | **Non-Cancer Cases** (*n* = 8932) | **Cancer Cases** (*n* = 1544) | **Non-Cancer Cases** (*n* = 14,809) |
| **Dietary patterns (PCA)** | | | | |
| *Western dietary pattern* | 0.27 (−0.27–1.00) | 0.30 (−0.28–1.02) | −0.41 (−0.81–0.06) | −0.40 (−0.79–0.07) |
| *Prudent dietary pattern* | −0.34 (−0.84–0.22) | −0.35 (−0.78–0.25) | −0.20 (−0.60–0.40) | −0.10 (−0.56–0.53) |
| *Sugar, fruit, and dairy dietary pattern* | −0.15 (−0.68–0.48) | −0.07 (−0.61–0.63) | −0.26 (−0.71–0.33) | −0.21 (−0.68–0.40) |
| **Dietary patterns (RRR)** | | | | |
| *Dietary fiber* | −0.01 (−0.96–1.16) | 0.19 (−0.74–1.32) | −0.63 (−1.35–0.26) | −0.50 (−1.24–0.43) |
| *Vitamin D* | −0.07 (−0.54–0.41) | −0.13 (−0.59–0.33) | 0.05 (−0.33–0.46) | 0.02 (−0.37–0.45) |
| *Fructose* | −0.03 (−0.41–0.46) | 0.04 (−0.40–0.44) | 0.02 (−0.30–0.39) | −0.38 (−0.03–0.29) |
| *Discretionary fats* | −0.17 (−0.56–0.18) | −0.15 (−0.56–0.24) | 0.02 (−0.30–0.39) | 0.06 (−0.27–0.42) |

were married (79%), highly educated (72%), employed (73%), former or never smokers (82%), and had no co-morbidities (62%).

## Principal component analysis

Three dietary patterns were identified using PCA analysis: "Western", "Prudent", and "sugar, fruits, and dairy" (Supplemental Tables 1, 2). A "Western" pattern was characterized by a high intake of grain servings, especially non-whole grains, vegetables, white potatoes, cheese, meats from lamb, pork, beef and luncheon meats (red and processed meats), discretionary fats, and teaspoons of added sugar. The "Prudent" pattern was characterized by a high intake of vegetables, fruits, and lean meat from fish and other sea food. The "sugar, fruits, and dairy" pattern was comprised of grain servings, especially whole grains, fruits, dairy, and teaspoons of added sugar. These three dietary patterns explained 42.4% of the variance in dietary intake.

Differences in standardized dietary pattern scores in males between cases and non-cases were mostly observed for a "sugar, fruit, and dairy" pattern. Female cases had lower "Prudent" pattern scores in cases compared to non-cases. Pattern scores for "Western" were similar across subgroups, although females had lower scores than males (Table 1).

## Reduced rank regression

Four dietary patterns were derived by RRR based on the four selected nutrients (Supplemental Table 3). A "dietary fiber" pattern was characterized by intake of grain servings, vegetables, and fruits and explained 52.4% of the response variation. The "vitamin D" pattern had high factor loadings for dairy and fish and other seafood and explained 19.4% of response variation. A "fructose" pattern was characterized by fruits and teaspoons of added sugar intake and explained 8.8% of response variation. Lastly, a "discretionary fat" pattern was characterized by intake of excess solid fats present within the "Milk" and "Meat and Beans" categories (e.g., whole vs. skim milk) and explained 7.9% of response variation. In total, these patterns explained 88.3 and 48.9% of the variation in response and predictor variables, respectively.

Female cases had higher standardized coefficient scores for a "fructose" pattern and lower scores for a "dietary fiber pattern" compared to non-cases (Table 1). Pattern scores were similar across groups for the "vitamin D pattern" and "discretionary fat" pattern.

## Combined cancer cases

The "Prudent" pattern was associated with reduced incidence of combined cancers in both the multivariable (HR = 0.83, 95% CI = 0.74–0.93) and latency multivariable-adjusted models (HR = 0.82, 95% CI = 0.73–0.92), when comparing the highest vs. lowest quartiles with evidence of statistically significant trends (p = 0.002

for both models). No other consistent trends were found using PCA derived patterns (Table 2).

For the RRR derived patterns, the "dietary fiber" and "discretionary fat" patterns were associated with reduced risk of combined cancer incidence, with significant trends observed for both patterns in the latency multivariable-adjusted model (p$_\text{dietary fiber}$ = 0.003; p$_\text{discretionary fatrs}$ = 0.008). The "fructose" pattern was associated with increased cancer risk in the latency multivariable-adjusted model. Those in the highest quartile had a 14% increased risk compared to the lowest quartile (HR = 1.14, 95% CI = 1.02–1.27) and a significant trend was observed (p = 0.01). (Table 3)

## Site-specific cancers

For lung cancer, the "sugar, fruit, and dairy" pattern was associated with reduced risk in all PCA models. In the latency multivariable-adjusted model, a statistically significant trend was observed (p = 0.01) and individuals in the third quartile had a reduced risk compared to those in the first quartile (HR = 0.61, 95% CI = 0.42–0.89) (Table 2). Associations with lung cancer were also found for all dietary patterns derived using RRR (Table 3). The most consistent association was with the "fructose" pattern, where individuals in higher quartiles were at increased risk of lung cancer compared to those in the first quartile. In the latency multivariable-adjusted model, the fourth quartile had a 53% greater risk of lung cancer compared to those in the first quartile (HR = 1.53, 95% CI = 1.07–2.20).

The "prudent" pattern was consistently associated with a reduced risk of colon cancer in all models, with a significant trend found in the latency multivariable-adjusted model (p = 0.037) (Table 2). Individuals in the highest quartile of this pattern had a 37% reduced risk of colorectal cancer compared to those in the lowest quartile (HR = 0.63, 95% CI = 0.42–0.96). The "dietary fiber" and "discretionary fats" patterns were associated with reduced colon cancer in the RRR analysis (Table 3). Both patterns had significant trends for the latency multivariable-adjusted model, with reductions in risk near 50% when comparing the fourth to the first quartile (HR$_\text{dietary fiber}$ = 0.54, 95% CI = 0.31–0.92; HR$_\text{discretionary fats}$ = 0.52, 95% CI = 0.33–0.82).

Reduced risk of prostate cancer was found for individuals with a "prudent" pattern, with significant trends found for both the multivariable-adjusted and latency multivariable-adjusted models (Table 2). A statistically significant trend was also observed for the "dietary fiber" pattern in the multivariable and latency multivariable-adjusted models (p = 0.03 and p = 0.01, respectively) (Table 3). In the latter model, individuals in the fourth quartile were at significantly reduced risk of prostate cancer compared to individuals in the first quartile of this pattern (HR = 0.65, 95% CI = 0.42–0.99). No statistically significant findings were observed for breast or endometrial cancer.

**Table 2.** Cox regression hazard ratios of cancer incidence for dietary patterns derived by PCA within Alberta's Tomorrow Project ($n = 26,242$), 2001–2008.

| Cancer site | Cases | Non-cases | Multivariable-adjusted HR[a] (95% CI) | Cases | Non-cases | Latency multivariable-adjusted HR[b] (95% CI) |
|---|---|---|---|---|---|---|
| **All cancers** | | | | | | |
| *Western dietary pattern* | | | | | | |
| Q1 | 676 | 5921 | 1.00 (ref) | 634 | 5921 | 1.00 (ref) |
| Q2 | 645 | 5928 | 1.00 (0.89–1.11) | 615 | 5928 | 1.02 (0.91–1.14) |
| Q3 | 706 | 5916 | 1.09 (0.97–1.22) | 662 | 5916 | 1.09 (0.92–1.23) |
| Q4 | 687 | 5907 | 1.06 (0.92–1.22) | 645 | 5907 | 1.07 (0.92–1.24) |
| Trend (p value) | | | 0.21 | | | 0.20 |
| *Prudent dietary pattern* | | | | | | |
| Q1 | 760 | 5924 | 1.00 (ref) | 718 | 5924 | 1.00 (ref) |
| Q2 | 681 | 5913 | 0.91 (0.82–1.01) | 642 | 5913 | 0.91 (0.82–1.01) |
| Q3 | 672 | 5917 | 0.90 (0.81–1.00) | 632 | 5917 | 0.90 (0.81–1.00) |
| Q4 | 601 | 5918 | 0.83 (0.74–0.93)* | 564 | 5918 | 0.82 (0.73–0.92) |
| Trend (p value) | | | 0.002** | | | 0.002** |
| **Sugar, fruit, and dairy dietary pattern** | | | | | | |
| Q1 | 753 | 5909 | 1.00 (ref) | 711 | 5909 | 1.00 (ref) |
| Q2 | 682 | 5920 | 0.95 (0.86–1.06) | 648 | 5920 | 0.96 (0.86–1.07) |
| Q3 | 678 | 5925 | 0.99 (0.89–1.10) | 641 | 5925 | 0.99 (0.89–1.11) |
| Q4 | 601 | 5918 | 0.94 (0.84–1.06) | 556 | 5918 | 0.92 (0.82–1.04) |
| Trend (p-value) | | | 0.44 | | | 0.40 |
| **Breast cancer** | | | | | | |
| *Western dietary pattern* | | | | | | |
| Q1 | 140 | 3964 | 1.00 (ref) | 129 | 3950 | 1.00 (ref) |
| Q2 | 152 | 3918 | 1.13 (0.90–1.43) | 146 | 3905 | 1.19 (0.94–1.51) |
| Q3 | 130 | 3948 | 0.99 (0.77–1.27) | 119 | 3940 | 0.99 (0.77–1.30) |
| Q4 | 120 | 3941 | 0.95 (0.71–1.28) | 114 | 3928 | 1.01 (0.74–1.37) |
| Trend (p-value) | | | 0.61 | | | 0.83 |
| *Prudent dietary pattern* | | | | | | |
| Q1 | 133 | 3992 | 1.00 (ref) | 126 | 3978 | 1.00 (ref) |
| Q2 | 163 | 3956 | 1.22 (0.97–1.54) | 153 | 3946 | 1.22 (0.96–1.54) |
| Q3 | 124 | 3924 | 0.94 (0.74–1.21) | 117 | 3907 | 0.95 (0.73–1.22) |
| Q4 | 122 | 3899 | 0.94 (0.72–1.22) | 112 | 3892 | 0.92 (0.70–1.21) |
| Trend (p-value) | | | 0.27 | | | 0.24 |
| **Sugar, fruit, and dairy dietary pattern** | | | | | | |
| Q1 | 135 | 3977 | 1.00 (ref) | 126 | 3978 | 1.00 (ref) |
| Q2 | 134 | 3942 | 1.02 (0.80–1.30) | 153 | 3946 | 1.22 (0.96–1.54) |
| Q3 | 136 | 3947 | 1.09 (0.85–1.39) | 117 | 3907 | 0.95 (0.73–1.22) |
| Q4 | 137 | 3905 | 1.16 (0.89–1.51) | 112 | 3892 | 0.92 (0.90–1.56) |
| Trend (p value) | | | 0.25 | | | 0.20 |
| Prostate cancer | | | | | | |
| *Western dietary pattern* | | | | | | |
| Q1 | 118 | 2400 | 1.00 (ref) | 111 | 2389 | 1.00 (ref) |
| Q2 | 119 | 2419 | 1.17 (0.90–1.52) | 113 | 2401 | 1.18 (0.90–1.54) |
| Q3 | 111 | 2401 | 1.18 (0.89–1.57) | 100 | 2393 | 1.12 (0.83–1.51) |
| Q4 | 92 | 2411 | 1.22 (0.84–1.76) | 88 | 2402 | 1.20 (0.82–1.75) |
| Trend (p value) | | | 0.26 | | | 0.39 |
| *Prudent dietary pattern* | | | | | | |
| Q1 | 118 | 2444 | 1.00 (ref) | 111 | 2428 | 1.00 (ref) |
| Q2 | 106 | 2370 | 0.82 (0.63–1.06) | 100 | 2363 | 0.82 (0.63–1.08) |
| Q3 | 117 | 2414 | 0.86 (0.66–1.11) | 107 | 2403 | 0.84 (0.64–1.10) |
| Q4 | 101 | 2403 | 0.76 (0.57–1.00) | 94 | 2391 | 0.76 (0.57–1.01) |
| Trend (p value) | | | 0.077 | | | 0.081 |

**Table 2** continued

| Cancer site | Cases | Non-cases | Multivariable-adjusted HR[a] (95% CI) | Cases | Non-cases | Latency multivariable-adjusted HR[b] (95% CI) |
|---|---|---|---|---|---|---|
| **Sugar, fruit, and dairy dietary pattern** | | | | | | |
| Q1 | 120 | 2452 | 1.00 (ref) | 114 | 2427 | 1.00 (ref) |
| Q2 | 123 | 2420 | 1.06 (0.82–1.36) | 119 | 2398 | 1.05 (0.82–1.36) |
| Q3 | 109 | 2397 | 1.01 (0.78–1.31) | 99 | 2383 | 0.93 (0.70–1.22) |
| Q4 | 90 | 2396 | 1.00 (0.76–1.31) | 80 | 2377 | 0.88 (0.65–1.20) |
| Trend (p value) | | | 0.94 | | | 0.33 |
| Lung cancer | | | | | | |
| *Western dietary pattern* | | | | | | |
| Q1 | 60 | 6537 | 1.00 (ref) | 58 | 6497 | 1.00 (ref) |
| Q2 | 61 | 6512 | 1.06 (0.74–1.52) | 60 | 6483 | 1.09 (0.76–1.57) |
| Q3 | 64 | 6559 | 1.10 (0.76–1.59) | 59 | 6520 | 1.08 (0.74–1.58) |
| Q4 | 67 | 6527 | 1.10 (0.70–1.73) | 64 | 6488 | 1.17 (0.73–1.86) |
| Trend (p value) | | | 0.64 | | | 0.56 |
| *Prudent dietary pattern* | | | | | | |
| Q1 | 92 | 6592 | 1.00 (ref) | 86 | 6556 | 1.00 (ref) |
| Q2 | 61 | 6537 | 0.77 (0.55–1.07) | 57 | 6499 | 0.80 (0.57–1.12) |
| Q3 | 64 | 6539 | 0.71 (0.50–1.01) | 49 | 6500 | 0.73 (0.51–1.05) |
| Q4 | 67 | 6467 | 0.72 (0.50–1.04) | 49 | 6433 | 0.72 (0.50–1.04) |
| Trend (p value) | | | 0.50* | | | 0.059 |
| **Sugar, fruit, and dairy dietary pattern** | | | | | | |
| Q1 | 95 | 6568 | 1.00 (ref) | 90 | 6531 | 1.00 (ref) |
| Q2 | 71 | 6531 | 0.96 (0.70–1.31) | 70 | 6498 | 1.00 (0.73–1.36) |
| Q3 | 43 | 6560 | 0.63 (0.44–0.91)* | 39 | 6527 | 0.61 (0.42–0.89)* |
| Q4 | 43 | 6476 | 0.67 (0.46–0.98)* | 42 | 6432 | 0.70 (0.47–1.03) |
| Trend (p value) | | | 0.007** | | | 0.011* |
| **Colon cancer** | | | | | | |
| *Western dietary pattern* | | | | | | |
| Q1 | 42 | 6555 | 1.00 (ref) | 40 | 6515 | 1.00 (ref) |
| Q2 | 49 | 6524 | 1.24 (0.81–1.88) | 47 | 6496 | 1.24 (0.81–1.91) |
| Q3 | 54 | 6568 | 1.36 (0.88–2.08) | 48 | 6530 | 1.26 (0.81–1.97) |
| Q4 | 54 | 6540 | 1.36 (0.80–2.30) | 53 | 6499 | 1.39 (0.81–2.38) |
| Trend (p-value) | | | 0.20 | | | 0.24 |
| *Prudent dietary pattern* | | | | | | |
| Q1 | 67 | 6617 | 1.00 (ref) | 65 | 6577 | 1.00 (ref) |
| Q2 | 38 | 6556 | 0.57 (0.38–0.85)* | 38 | 6517 | 0.58 (0.39–0.87)* |
| Q3 | 47 | 6542 | 0.70 (0.47–1.02) | 43 | 6506 | 0.65 (0.44–0.96)* |
| Q4 | 47 | 6472 | 0.70 (0.47–1.04) | 42 | 6440 | 0.63 (0.42–0.96)* |
| Trend (p value) | | | 0.12 | | | 0.037* |
| **Sugar, fruit, and dairy dietary pattern** | | | | | | |
| Q1 | 61 | 6601 | 1.00 (ref) | 58 | 6562 | 1.00 (ref) |
| Q2 | 53 | 6549 | 0.90 (0.63–1.31) | 48 | 6520 | 0.86 (0.59–1.26) |
| Q3 | 47 | 6556 | 0.82 (0.56–1.20) | 45 | 6521 | 0.82 (0.55–1.22) |
| Q4 | 38 | 6481 | 0.68 (0.44–1.04) | 37 | 6437 | 0.69 (0.45–1.07) |
| Trend (p value) | | | 0.069 | | | 0.10 |
| Endometrial cancer | | | | | | |
| *Western dietary pattern* | | | | | | |
| Q1 | 31 | 4073 | 1.00 (ref) | 29 | 4050 | 1.00 (ref) |
| Q2 | 34 | 4036 | 1.21 (0.74–1.98) | 30 | 4021 | 1.12 (0.67–1.88) |
| Q3 | 42 | 4036 | 1.63 (0.99–2.67) | 42 | 4017 | 1.69 (1.02–2.80) |
| Q4 | 31 | 4030 | 1.37 (0.74–2.53) | 29 | 4013 | 1.31 (0.69–2.46) |
| Trend (p value) | | | 0.15 | | | 0.16 |

**Table 2** continued

| Cancer site | Cases | Non-cases | Multivariable-adjusted HR[a] (95% CI) | Cases | Non-cases | Latency multivariable-adjusted HR[b] (95% CI) |
|---|---|---|---|---|---|---|
| *Prudent dietary pattern* | | | | | | |
| Q1 | 41 | 4084 | 1.00 (ref) | 38 | 4066 | 1.00 (ref) |
| Q2 | 35 | 4084 | 0.83 (0.53–1.30) | 33 | 4066 | 0.84 (0.53–1.24) |
| Q3 | 29 | 4019 | 0.70 (0.43–1.14) | 27 | 3997 | 0.70 (0.42–1.16) |
| Q4 | 33 | 3988 | 0.79 (0.48–1.29) | 32 | 3972 | 0.82 (0.49–1.35) |
| Trend (*p* value) | | | 0.25 | | | 0.32 |
| **Sugar, fruit, and dairy dietary pattern** | | | | | | |
| Q1 | 31 | 4081 | 1.00 (ref) | 29 | 4061 | 1.00 (ref) |
| Q2 | 42 | 4034 | 1.31 (0.82–2.09) | 40 | 4018 | 1.33 (0.82–2.14) |
| Q3 | 33 | 4050 | 1.03 (0.63–1.71) | 32 | 4031 | 1.05 (0.63–1.76) |
| Q4 | 32 | 4010 | 0.97 (0.56–1.68) | 29 | 3991 | 0.91 (0.52–1.60) |
| Trend (*p* value) | | | 0.72 | | | 0.58 |

Quartile cut-points of dietary patterns were based on non-cases. Linear trends over quartiles for each dietary pattern separately were investigated by assigning median scores to each quartile of dietary pattern intake as a continuous variable.

*$p < 0.05$; **$p < 0.01$.

[a]Adjusted for: age, sex, BMI, energy intake, smoking status, physical activity.

[b]Cancers occurring <2 years after questionnaire completion removed.

A summary of the dietary pattern components and the associated cancer risks can be found in Table 4.

## DISCUSSION

This analysis, from a prospective cohort study of ~25,000 adults from Alberta, Canada, examined the association between dietary patterns derived using both PCA and RRR, and cancer incidence. Using these methods, we found that a "dietary fiber" pattern decreased the risk of combined cancers, as well as lung, colon, and prostate cancers. A "fructose" pattern was associated with increased risk of combined cancers and lung cancer, while a "discretionary fats" pattern was associated with a decreased risk of colon and combined cancers. When dietary patterns were derived using PCA, the "Prudent" pattern was associated with a decreased risk of colon and combined cancers, and a "sugar, fruit, and dairy" pattern was associated with a reduced risk of lung cancer. The 2018 WCRF report on cancer prevention recommends eating a diet high in fruits, vegetables, whole grains, and legumes [7]. While the report specifically indicates that there is "strong evidence" that consumption of these foods is protective against colorectal cancer, it states that these foods also protect against weight gain, which is a risk factor for many cancers [7]. These recommendations are in line with our finding that a "Prudent" dietary pattern protects against combined cancers and colon cancer. These results are also consistent with previous studies [7, 21, 22, 29–34]. Contrary to previous findings, though, we did not find an increased risk of cancer among those following a "Western" pattern [21, 22, 29, 30, 35]. A possible explanation for this difference is that the "Western" pattern was common in this population, so even those in the lowest quartiles had relatively high exposure to components of this pattern, which may have decreased the ability to detect an association with cancer incidence in this population. In general, our results support cancer prevention messaging to maintain a diet high in vegetables, fruits, fish, and whole grains.

It is important to acknowledge the apparent contradiction in results obtained through PCA and RRR methods for lung cancer as highlighted in Table 4. In the PCA model, the "sugar, fruit, and dairy" pattern is associated with reduced risk, whereas the "fructose" pattern in the RRR model is associated with an increased risk. Based on the factor loading values presented in Supplementary Tables 1, 3, added tablespoons of sugar is a bigger

contributor to the "fructose pattern", whereas the natural fruit groups were stronger contributors to the "sugar, fruits, and dairy pattern". The most recent WCRF report published in 2018 suggests that fruit intake is protective for lung cancer risk, which may partly explain these discrepancies in results. However, this interpretation of the results cannot be confirmed by our present findings because these are analyses of dietary patterns, so we cannot confirm the effects of single nutrients. This may be a worthwhile area of future research.

The protective effect of fruit intake may be due to active flavonoid compounds present in fruits, which have antioxidant properties [36]. However, other studies have found this relationship to exist only among smokers because they may benefit more from the antioxidants found in fruit [37, 38]. The WCRF report also acknowledges a protective association between fruit and vegetable intake and lung cancer, which is attributed to the presence of beta-carotene in these foods [7]. We also found a protective effect for the "dietary fiber" pattern on lung, colorectal, and prostate cancers, which is consistent with previous findings [7, 25], and may also be attributable to fruit consumption in this group.

Decreased risk of combined cancers, as well as breast and colon cancers, was associated with a higher "discretionary fat" dietary pattern. The WCRF report found a limited association between discretionary fat intake and pancreatic cancer, but there was insufficient evidence for all other cancer types [7]. One study found a moderate positive association between discretionary fat intake and breast cancer risk, but this finding was attenuated after adjustment for red meat intake [39], and a meta-analysis found no association [40]. The protective association that we observed for discretionary fat dietary patterns might have been due to social desirability. Discretionary fats are defined as excess fat that may be present in foods within the "Milk" and "Meat and Beans" categories that are high fat versions instead of low fat alternatives [41]. Both male and female non cancer cases have a higher intake of saturated fat (Table 1). However, the food group saturated/discretionary fat is prone to social desirability, particularly in individuals with an unhealthy lifestyle and individuals with overweight or obesity who are prone to underreporting discretionary fats [42]. A larger proportion of individuals with overweight and obesity are present in the case group (Table 1).

Our results show a non-statistically significant decrease in all cancer incidence with a higher intake of a "vitamin D pattern".

**Table 3.** Cox regression hazard ratios of cancer incidence for dietary patterns derived by RRR within Alberta's Tomorrow Project ($n = 26{,}242$), 2001–2008.

| Cancer site | Cases | Non-cases | Multivariable-adjusted HR[a] (95% CI) | Cases | Non-cases | Latency multivariable-adjusted HR[b] (95% CI) |
|---|---|---|---|---|---|---|
| **All cancers** | | | | | | |
| *Dietary fiber* | | | | | | |
| Q1 | 790 | 5913 | 1.00 (ref) | 741 | 5913 | 1.00 (ref) |
| Q2 | 688 | 5922 | 0.89 (0.81–1.00) | 644 | 5922 | 0.88 (0.79–0.98)* |
| Q3 | 608 | 5927 | 0.81 (0.72–0.92)* | 578 | 5927 | 0.81 (0.71–0.92)* |
| Q4 | 638 | 5910 | 0.85 (0.72–0.99)* | 593 | 5910 | 0.82 (0.69–0.97)* |
| Trend (*p* value) | | | 0.006** | | | 0.003** |
| *Vitamin D* | | | | | | |
| Q1 | 646 | 5921 | 1.00 (ref) | 604 | 5921 | 1.00 (ref) |
| Q2 | 664 | 5915 | 0.95 (0.85–1.06) | 637 | 5915 | 0.97 (0.87–1.09) |
| Q3 | 684 | 5922 | 0.94 (0.84–1.05) | 632 | 5922 | 0.93 (0.83–1.04) |
| Q4 | 720 | 5914 | 0.94 (0.85–1.05) | 683 | 5914 | 0.95 (0.85–1.07) |
| Trend (*p* value) | | | 0.31 | | | 0.31 |
| *Fructose* | | | | | | |
| Q1 | 677 | 5921 | 1.00 (ref) | 635 | 5921 | 1.00 (ref) |
| Q2 | 653 | 5920 | 1.00 (0.89–1.11) | 603 | 5920 | 0.98 (0.88–1.10) |
| Q3 | 664 | 5917 | 1.04 (0.93–1.16) | 635 | 5917 | 1.06 (0.95–1.18) |
| Q4 | 720 | 5914 | 1.12 (1.01–1.25)* | 683 | 5914 | 1.14 (1.02–1.27)* |
| Trend (*p* value) | | | 0.023* | | | 0.011* |
| *Discretionary fats* | | | | | | |
| Q1 | 757 | 5911 | 1.00 (ref) | 721 | 5911 | 1.00 (ref) |
| Q2 | 694 | 5916 | 0.94 (0.85–1.05) | 657 | 5916 | 0.93 (0.83–1.04) |
| Q3 | 664 | 5922 | 0.92 (0.82–1.03) | 619 | 5922 | 0.90 (0.80–1.00) |
| Q4 | 599 | 5923 | 0.88 (0.79–0.99)* | 559 | 5923 | 0.86 (0.77–0.96)* |
| Trend (*p* value) | | | 0.025* | | | 0.008** |
| **Breast cancer** | | | | | | |
| *Dietary fiber* | | | | | | |
| Q1 | 155 | 3986 | 1.00 (ref) | 146 | 3974 | 1.00 (ref) |
| Q2 | 127 | 3960 | 0.84 (0.66–1.08) | 118 | 3950 | 0.84 (0.65–1.09) |
| Q3 | 137 | 3923 | 0.94 (0.71–1.25) | 131 | 3908 | 0.98 (0.73–1.32) |
| Q4 | 123 | 3902 | 0.89 (0.59–1.33) | 113 | 3891 | 0.91 (0.60–1.39) |
| Trend (*p* value) | | | 0.70 | | | 0.87 |
| *Vitamin D* | | | | | | |
| Q1 | 130 | 3913 | 1.00 (ref) | 124 | 3899 | 1.00 (ref) |
| Q2 | 128 | 3951 | 0.91 (0.71–1.17) | 123 | 3943 | 0.91 (0.71–1.17) |
| Q3 | 144 | 3960 | 0.98 (0.77–1.25) | 129 | 3947 | 0.91 (0.71–1.17) |
| Q4 | 140 | 3947 | 0.94 (0.73–1.19) | 132 | 3934 | 0.92 (0.72–1.18) |
| Trend (*p* value) | | | 0.74 | | | 0.54 |
| *Fructose* | | | | | | |
| Q1 | 136 | 3941 | 1.00 (ref) | 127 | 3930 | 1.00 (ref) |
| Q2 | 140 | 3915 | 1.03 (0.81–1.31) | 127 | 3899 | 1.00 (0.78–1.28) |
| Q3 | 139 | 3958 | 1.03 (0.81–1.31) | 135 | 3950 | 1.06 (0.83–1.36) |
| Q4 | 127 | 3957 | 0.95 (0.75–1.22) | 119 | 3944 | 0.95 (0.74–1.22) |
| Trend (*p* value) | | | 0.71 | | | 0.82 |
| *Discretionary fats* | | | | | | |
| Q1 | 152 | 3952 | 1.00 (ref) | 144 | 3941 | 1.00 (ref) |
| Q2 | 134 | 3961 | 0.84 (0.66–1.06) | 128 | 3954 | 0.84 (0.65–1.07) |
| Q3 | 120 | 3941 | 0.76 (0.59–0.97)* | 111 | 3925 | 0.73 (0.57–0.95)* |
| Q4 | 136 | 3917 | 0.90 (0.71–1.13) | 125 | 3903 | 0.87 (0.68–1.10) |
| Trend (*p* value) | | | 0.27 | | | 0.16 |
| **Prostate cancer** | | | | | | |
| *Dietary fiber* | | | | | | |
| Q1 | 148 | 2433 | 1.00 (ref) | 139 | 2416 | 1.00 (ref) |
| Q2 | 113 | 2406 | 0.82 (0.65–1.06) | 108 | 2396 | 0.83 (0.64–1.08) |

**Table 3** continued

| Cancer site | Cases | Non-cases | Multivariable-adjusted HRª (95% CI) | Cases | Non-cases | Latency multivariable-adjusted HRᵇ (95% CI) |
|---|---|---|---|---|---|---|
| Q3 | 93 | 2397 | 0.69 (0.51–0.94)* | 85 | 2388 | 0.66 (0.48–0.90)* |
| Q4 | 88 | 2395 | 0.70 (0.46–1.05) | 80 | 2385 | 0.65 (0.42–0.99)* |
| Trend (p value) | | | 0.028* | | | 0.012* |
| *Vitamin D* | | | | | | |
| Q1 | 105 | 2385 | 1.00 (ref) | 95 | 2372 | 1.00 (ref) |
| Q2 | 107 | 2408 | 0.81 (0.61–1.07) | 102 | 2398 | 0.86 (0.65–1.14) |
| Q3 | 104 | 2398 | 0.71 (0.54–0.95) | 98 | 2385 | 0.75 (0.56–1.01) |
| Q4 | 126 | 2440 | 0.84 (0.64–1.09) | 117 | 2430 | 0.87 (0.66–1.15) |
| Trend (p value) | | | 0.18 | | | 0.28 |
| *Fructose* | | | | | | |
| Q1 | 115 | 2410 | 1.00 (ref) | 104 | 2398 | 1.00 (ref) |
| Q2 | 123 | 2393 | 1.08 (0.84–1.40) | 114 | 2381 | 1.11 (0.84–1.45) |
| Q3 | 108 | 2392 | 1.05 (0.80–1.37) | 102 | 2381 | 1.09 (0.82–1.44) |
| Q4 | 96 | 2436 | 1.06 (0.81–1.40) | 92 | 2425 | 1.11 (0.84–1.48) |
| Trend (p value) | | | 0.72 | | | 0.49 |
| *Discretionary fats* | | | | | | |
| Q1 | 93 | 2425 | 1.00 (ref) | 88 | 2417 | 1.00 (ref) |
| Q2 | 134 | 2405 | 1.33 (1.01–1.75) | 124 | 2392 | 1.30 (0.98–1.72) |
| Q3 | 117 | 2418 | 1.14 (0.85–1.53) | 111 | 2404 | 1.14 (0.84–1.54) |
| Q4 | 98 | 2383 | 1.03 (0.77–1.38) | 89 | 2372 | 0.98 (0.73–1.33) |
| Trend (p value) | | | 0.68 | | | 0.55 |
| **Lung cancer** | | | | | | |
| *Dietary fiber* | | | | | | |
| Q1 | 92 | 6602 | 1.00 (ref) | 90 | 6565 | 1.00 (ref) |
| Q2 | 56 | 6554 | 0.69 (0.49–0.97)* | 51 | 6515 | 0.64 (0.45–0.92)* |
| Q3 | 47 | 6488 | 0.60 (0.41–0.88)* | 46 | 6459 | 0.61 (0.41–0.90)* |
| Q4 | 57 | 6491 | 0.66 (0.41–1.06) | 54 | 6449 | 0.66 (0.40–1.08) |
| Trend (p value) | | | <0.000** | | | 0.032* |
| *Vitamin D* | | | | | | |
| Q1 | 64 | 6503 | 1.00 (ref) | 60 | 6465 | 1.00 (ref) |
| Q2 | 59 | 6521 | 0.85 (0.60–1.22) | 58 | 6495 | 0.89 (0.61–1.28) |
| Q3 | 71 | 6535 | 1.08 (0.76–1.53) | 67 | 6487 | 1.08 (0.75–1.54) |
| Q4 | 58 | 6576 | 0.79 (0.55–1.13) | 56 | 6541 | 0.80 (0.55–1.16) |
| Trend (p value) | | | <0.000** | | | 0.43 |
| *Fructose* | | | | | | |
| Q1 | 52 | 6546 | 1.00 (ref) | 49 | 6507 | 1.00 (ref) |
| Q2 | 57 | 6516 | 1.14 (0.78–1.67) | 56 | 6467 | 1.19 (0.80–1.75) |
| Q3 | 55 | 6527 | 1.04 (0.71–1.53) | 54 | 6499 | 1.08 (0.73–1.60) |
| Q4 | 88 | 6546 | 1.54 (1.09–2.18)* | 82 | 6515 | 1.53 (1.07–2.20) |
| Trend (p value) | | | <0.000** | | | 0.030* |
| *Discretionary fats* | | | | | | |
| Q1 | 76 | 6592 | 1.00 (ref) | 71 | 6561 | 1.00 (ref) |
| Q2 | 75 | 6536 | 1.17 (0.84–1.63) | 71 | 6503 | 1.16 (0.82–1.64) |
| Q3 | 64 | 6522 | 1.08 (0.76–1.54) | 63 | 6478 | 1.12 (0.78–1.60) |
| Q4 | 37 | 6485 | 0.66 (0.44–0.98)* | 36 | 6446 | 0.67 (0.45–1.01) |
| Trend (p value) | | | 0.058 | | | 0.088 |
| **Colon cancer** | | | | | | |
| *Dietary fiber* | | | | | | |
| Q1 | 63 | 6630 | 1.00 (ref) | 61 | 6593 | 1.00 (ref) |
| Q2 | 47 | 6563 | 0.71 (0.48–1.04) | 45 | 6521 | 0.69 (0.46–1.02) |
| Q3 | 43 | 6492 | 0.63 (0.41–0.96)* | 38 | 6467 | 0.56 (0.36–0.87)* |
| Q4 | 46 | 6502 | 0.57 (0.33–0.98)* | 44 | 6459 | 0.54 (0.31–0.92)* |
| Trend (p value) | | | 0.025* | | | 0.010* |

**Table 3** continued

| Cancer site | Cases | Non-cases | Multivariable-adjusted HR[a] (95% CI) | Cases | Non-cases | Latency multivariable-adjusted HR[b] (95% CI) |
|---|---|---|---|---|---|---|
| *Vitamin D* | | | | | | |
| Q1 | 39 | 6528 | 1.00 (ref) | 38 | 6487 | 1.00 (ref) |
| Q2 | 53 | 6526 | 1.28 (0.84–1.96) | 50 | 6502 | 1.24 (0.81–1.91) |
| Q3 | 54 | 6552 | 1.25 (0.82–1.91) | 53 | 6501 | 1.26 (0.82–1.93) |
| Q4 | 53 | 6581 | 1.31 (0.74–1.71) | 47 | 6550 | 1.02 (0.66–1.57) |
| Trend (*p* value) | | | 0.71 | | | 0.98 |
| *Fructose* | | | | | | |
| Q1 | 59 | 6539 | 1.00 (ref) | 54 | 6502 | 1.00 (ref) |
| Q2 | 38 | 6535 | 0.68 (0.45–1.02) | 37 | 6486 | 0.73 (0.48–1.11) |
| Q3 | 44 | 6537 | 0.81 (0.54–1.21) | 43 | 6509 | 0.88 (0.58–1.32) |
| Q4 | 58 | 6576 | 1.07 (0.74–1.54) | 54 | 6543 | 1.09 (0.75–1.60) |
| Trend (*p* value) | | | 0.59 | | | 0.51 |
| *Discretionary fats* | | | | | | |
| Q1 | 65 | 6603 | 1.00 (ref) | 63 | 6569 | 1.00 (ref) |
| Q2 | 43 | 6567 | 0.70 (0.47–1.04) | 43 | 6530 | 0.72 (0.48–1.08) |
| Q3 | 59 | 6527 | 0.98 (0.67–1.42) | 53 | 6488 | 0.90 (0.61–1.34) |
| Q4 | 32 | 6490 | 0.56 (0.36–0.86)* | 29 | 6453 | 0.52 (0.33–0.82)* |
| Trend (*p* value) | | | 0.051 | | | 0.021* |
| Endometrial cancer | | | | | | |
| *Dietary fiber* | | | | | | |
| Q1 | 33 | 4108 | 1.00 (ref) | 31 | 4089 | 1.00 (ref) |
| Q2 | 43 | 4044 | 1.26 (0.78–2.04) | 40 | 4028 | 1.23 (0.75–2.02) |
| Q3 | 33 | 4027 | 0.95 (0.54–1.69) | 31 | 4008 | 0.93 (0.52–1.66) |
| Q4 | 29 | 3996 | 0.78 (0.35–1.72) | 28 | 3976 | 0.76 (0.34–1.69) |
| Trend (*p* value) | | | 0.49 | | | 0.45 |
| *Vitamin D* | | | | | | |
| Q1 | 26 | 4017 | 1.00 (ref) | 23 | 4000 | 1.00 (ref) |
| Q2 | 41 | 4038 | 1.49 (0.90–2.46) | 40 | 4026 | 1.67 (0.99–2.82) |
| Q3 | 34 | 4070 | 1.15 (0.69–1.95) | 33 | 4043 | 1.29 (0.75–2.22) |
| Q4 | 37 | 4050 | 1.23 (0.74–2.05) | 34 | 4032 | 1.29 (0.75–2.20) |
| Trend (*p* value) | | | 0.79 | | | 0.68 |
| *Fructose* | | | | | | |
| Q1 | 37 | 4040 | 1.00 (ref) | 35 | 4022 | 1.00 (ref) |
| Q2 | 31 | 4024 | 0.83 (0.51–1.35) | 29 | 3997 | 0.82 (0.50–1.36) |
| Q3 | 28 | 4069 | 0.75 (0.46–1.25) | 26 | 4059 | 0.75 (0.44–1.26) |
| Q4 | 42 | 4042 | 1.15 (0.73–1.80) | 40 | 4023 | 1.16 (0.73–1.85) |
| Trend (*p* value) | | | 0.60 | | | 0.57 |
| *Discretionary fats* | | | | | | |
| Q1 | 32 | 4072 | 1.00 (ref) | 30 | 4055 | 1.00 (ref) |
| Q2 | 38 | 4057 | 1.15 (0.71–1.88) | 38 | 4044 | 1.24 (0.76–2.04) |
| Q3 | 36 | 4025 | 1.06 (0.65–1.74) | 32 | 4004 | 1.02 (0.61–1.71) |
| Q4 | 32 | 4021 | 0.95 (0.58–1.55) | 30 | 3998 | 0.96 (0.57–1.59) |
| Trend (*p* value) | | | 0.72 | | | 0.64 |

Quartile cut-points of dietary patterns were based on non-cases. Linear trends over quartiles for each dietary pattern separately were investigated by assigning median scores to each quartile of dietary pattern intake as a continuous variable.

*$p < 0.05$; **$p < 0.01$.

[a]Adjusted for: age, sex, BMI, energy intake, smoking status, physical activity.

[b]Cancers occurring <2 years after questionnaire completion removed.

We used dietary vitamin D to derive the "Vitamin D dietary pattern", as variance in intake of supplemental vitamin D cannot be explained within a dietary pattern. An Italian case control study found an inverse association between dietary vitamin D intake and colon cancer risk among participants in the highest categories of intake [43]. However dietary vitamin D intake in this population is far below the recommended daily intake of 15 µg of vitamin D in Canada (Table 1) [44].

This study's strengths include the large sample size of the cohort which permitted a concurrent analysis of five different cancer sites. The dietary assessment method (Canadian Diet History Questionnaire) was validated for the nutrients used in this study [13]. Moreover, we used two different statistical methods in this study to derive dietary patterns to overcome some of the inherent limitations with PCA such as unexplained variation in nutrient intake. In addition, we considered the possibility of reverse

**Table 4.** Descriptors of the main components in each dietary pattern for both PCA and RRR methods, grouped by similarity of dietary pattern (i.e., the prudent and dietary fiber patterns are both partially characterized by intake of vegetables and fruits). Shaded boxes indicate the dietary patterns confers a statistically significant reduced or increased risk of cancers at the $p < 0.05$.

| Dietary Patterns | Analytical approach | Major components of dietary pattern | Significant reduced risk | Significant increased risk |
|---|---|---|---|---|
| Western | Principle component analysis | Grains (non-whole), vegetables, white potatoes, cheese, red and processed meats, discretionary fats, and added teaspoons of sugar. | None | None |
| Discretionary fats | Reduced rank regression | Excess solid fats present in foods within the "Milk" and "Meat and Beans" categories (e.g., whole vs. skim milk) | All cancers, colon cancer* | None |
| Prudent | Principle component analysis | Vegetables, fruits, and lean meat from fish and other seafood | All cancers, colon cancer, prostate cancer* | None |
| Dietary fiber | Reduced rank regression | Grain servings, vegetables, and fruits | All cancers, colon cancer, prostate cancer* | None |
| Vitamin D | Reduced rank regression | Dairy, fish, and other seafood | None | None |
| Fruits, sugar and dairy | Principle component analysis | Whole grains, fruits, dairy, and teaspoons of added sugar | Lung cancer * | None |
| Fructose | Reduced rank regression | Fruits and teaspoons of added sugar intake | None | All cancers, lung cancer* |

*Statistically significant for reduced or increased risk at the $p < 0.05$ level.

causation by excluding cancer cases diagnosed in the first 2 years of follow-up. Finally, data on many different covariates were obtained at baseline and therefore could be included in the analysis.

The main limitation of this study is the potential for exposure misclassification due to the social desirability bias that is inherent to self-reported dietary intake questionnaires. We recognize that food frequency questionnaires, including the Canadian Diet History Questionnaire used in this study, are known for misreporting and measurement errors (e.g., recall bias, modifying dietary habits or responses on a questionnaire through self-reflection, or to reduce response burden) [6, 42]. Nevertheless, this approach to collecting information on nutrient intake in studies that include a large number of participants is an acceptable limitation in the field of nutritional epidemiology for the purpose of studying associations between nutrient intake and chronic disease risk. In addition, dietary intake was only assessed once, and changes in dietary intake were not taken into consideration. This sample is a homogeneous population, which impacts the generalizability of the study findings and increases the potential for unmeasured confounding. The RRR was performed with the data of the food groups and nutrients from the same questionnaire. Therefore, food groups and nutrients are not from independent sources [15] and correlated errors in intake assessment might have been present. We have chosen to not include supplementation use as part of our dietary patterns analysis to focus solely on the contributions of dietary intakes on cancer risk. An average of 34% of Canadians have reported taken vitamin D supplements [45], suggesting that total vitamin D intake (dietary + supplements) may have been under-reported in approximately one-third of our sample. We suggest that dietary supplementation and cancer risk be an area of future research. Finally, while data-driven in their approach, there remains a certain degree of subjectivity in the analyses of dietary components based on statistical thresholds and modeling decisions. This subjectivity should be noted when interpreting results both within and across studies.

## CONCLUSION

To our knowledge, this is the largest Canadian population to be analyzed for dietary patterns and cancer incidence to date. Using population-based survey data, we were able to identify commonly-consumed dietary patterns, and explore the associations of these dietary patterns with multiple cancer sites. These dietary patterns were assessed with two complimentary approaches to provide a richer perspective of population dietary patterns. This is an emerging field of interest in cancer research and, while strong protective associations were found for those consuming a "Prudent" dietary pattern or a "dietary fiber" pattern and the dietary pattern characterized by "fructose" consumption conferred an increased risk of lung and combined cancer incidence. While these findings generally support cancer prevention strategies that advocate for a diet high in vegetables, fruits, fish, and whole grains, a reduced risk of colorectal cancer was also observed for those with a discretionary fat dietary pattern. We suspect this may be attributed to the mono- and polyunsaturated fats that are part of discretionary fat intake. These findings provide insights into dietary patterns consumed in Alberta and their associations with incidence of several cancer types that can be further compared to other populations worldwide to strengthen the evidence for the role of diet in cancer prevention.

## REFERENCES

1. Poirier AE, Ruan Y, Volesky KD, King WD, O'Sullivan DE, Gogna P, et al. The current and future burden of cancer attributable to modifiable risk factors in Canada: summary of results. Prev Med. 2019;122:140–7. https://doi.org/10.1016/j.ypmed.2019.04.007.

2. Poirier AE, Ruan Y, Hebert LA, Grevers X, Walter SD, Villeneuve PJ, et al. Estimates of the current and future burden of cancer attributable to low fruit and vegetable consumption in Canada. Prev Med. 2019;122:20–30. https://doi.org/10.1016/j.ypmed.2019.03.013.

3. Blot WJ, Tarone RE. Doll and Peto's quantitative estimates of cancer risks: holding generally true for 35 years. J Natl Cancer Inst. 2015;107:djv044. https://doi.org/10.1093/jnci/djv044.

4. Bamia C. Dietary patterns in association to cancer incidence and survival: concept, current evidence, and suggestions for future research. Eur J Clin Nutr. 2018;72:818–25. https://doi.org/10.1038/s41430-018-0128-8.

5. Solbak NM, Xu JY, Vena JE, Csizmadi I, Whelan HK, Robson PJ. Diet quality is associated with reduced incidence of cancer and self-reported chronic disease: observations from Alberta's Tomorrow Project. Prev Med. 2017;101:178–87. https://doi.org/10.1016/j.ypmed.2017.06.009.

6. Schulze MB, Martinez-Gonzalez MA, Fung TT, Lichtenstein AH, Forouhi NG. Food based dietary patterns and chronic disease prevention. BMJ. 2018;361:k2396 https://doi.org/10.1136/bmj.k2396.

7. World Cancer Research Fund/American Institute for Cancer Research. Diet, nutrition, physical activity and cancer: a global perspective. Continuous update project expert report. London: World Cancer Research Fund International; 2018.

8. Robson PJ, Solbak NM, Haig TR, Whelan HK, Vena JE, Akawung AK, et al. Design, methods and demographics from phase I of Alberta's Tomorrow Project cohort: a prospective cohort profile. CMAJ Open. 2016;4:E515.

9. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the eating at America's Table Study. Am J Epidemiol. 2001;154:1089–99. https://doi.org/10.1093/aje/154.12.1089.

10. Csizmadi I, Kahle L, Ullman R, Dawe U, Zimmerman TP, Friedenreich CM, et al. Adaptation and evaluation of the National Cancer Institute's Diet History Questionnaire and nutrient database for Canadian populations. Public Health Nutr. 2007;10:88–96.

11. Krebs-Smith SM, Cronin FJ, Haytowitz DB, Cook DA. Food sources of energy, macronutrients, cholesterol, and fiber in diets of women. J Am Diet Assoc. 1992;92:168–74.

12. Subar AF, Midthune D, Kulldorff M, Brown CC, Thompson FE, Kipnis V, et al. Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. Am J Epidemiol. 2000;152:279–86. https://doi.org/10.1093/aje/152.3.279.

13. Csizmadi I, Boucher BA, Siou GL, Massarelli I, Rondeau I, Garriguet D, et al. Using national dietary intake data to evaluate and adapt the US Diet History Questionnaire: the stepwise tailoring of an FFQ for Canadian use. Public Health Nutr. 2016;19:3247–55.

14. McNeil J, Barberio AM, Friedenreich CM, Brenner DR. Sleep and cancer incidence in Alberta's Tomorrow Project cohort. Sleep. 2019;42:zsy252. https://doi.org/10.1093/sleep/zsy252.

15. Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. Am J Epidemiol. 2004;159:935–44.

16. Hatcher L. A step-by-step approach to using SAS for factor analysis and structural equation modeling. Cary, NC. Ist edn. In: SAS Publishing; 1994.

17. Corner S. Choosing the right type of rotation in PCA and EFA. JALT Test Evaluation SIG Newsl. 2009;13:20–25.

18. Weikert C, Schulze MB. Evaluating dietary patterns: the role of reduced rank regression. Curr Opin Clin Nutr Metab Care. 2016;19:341–6. https://doi.org/10.1097/MCO.0000000000000308.

19. Mao QQ, Lin YW, Chen H, Qin J, Zheng XY, Xu X, et al. Dietary fiber intake is inversely associated with risk of pancreatic cancer: a meta-analysis. Asia Pac J Clin Nutr. 2017;26:89–96. https://doi.org/10.6133/apjcn.102015.03.

20. Wang CH, Qiao C, Wang RC, Zhou WP. Dietary fiber intake and pancreatic cancer risk: a meta-analysis of epidemiologic studies. Sci Rep. 2015;5:10834 https://doi.org/10.1038/srep10834.

21. Chen Z, Wang PP, Woodrow J, Zhu Y, Roebothan B, McLaughlin JR, et al. Dietary patterns and colorectal cancer: results from a Canadian population-based study. Nutr J. 2015;14:8 https://doi.org/10.1186/1475-2891-14-8.

22. Yusof AS, Isa ZM, Shah SA. Dietary patterns and risk of colorectal cancer: a systematic review of cohort studies (2000–2011). Asian Pac J Cancer Prev. 2012;13:4713–7. https://doi.org/10.7314/apjcp.2012.13.9.4713.

23. Song M, Chan AT, Fuchs CS, Ogino S, Hu FB, Mozaffarian D, et al. Dietary intake of fish, omega-3 and omega-6 fatty acids and risk of colorectal cancer: a prospective study in U.S. men and women. Int J Cancer. 2014;135:2413–23. https://doi.org/10.1002/ijc.28878.

24. Chen S, Chen Y, Ma S, Zheng R, Zhao P, Zhang L, et al. Dietary fibre intake and risk of breast cancer: a systematic review and meta-analysis of epidemiological studies. Oncotarget. 2016;7:80980–9. https://doi.org/10.18632/oncotarget.13140.

25. Deschasaux M, Pouchieu C, His M, Hercberg S, Latino-Martel P, Touvier M. Dietary total and insoluble fiber intakes are inversely associated with prostate cancer risk. J Nutr. 2014;144:504–10. https://doi.org/10.3945/jn.113.189670.

26. Deschasaux M, Zelek L, Pouchieu C, His M, Hercberg S, Galan P, et al. Prospective association between dietary fiber intake and breast cancer risk. PLoS ONE. 2013;8: e79718 https://doi.org/10.1371/journal.pone.0079718.

27. Kawakita D, Lee YA, Turati F, Parpinel M, Decarli A, Serraino D, et al. Dietary fiber intake and head and neck cancer risk: a pooled analysis in the International Head and Neck Cancer Epidemiology consortium. Int J Cancer. 2017;141:1811–21. https://doi.org/10.1002/ijc.30886.

28. Xu H, Ding Y, Xin X, Wang W, Zhang D. Dietary fiber intake is associated with a reduced risk of ovarian cancer: a dose-response meta-analysis. Nutr Res. 2018;57:1–11. https://doi.org/10.1016/j.nutres.2018.04.011.

29. Shu L, Wang X-Q, Wang S-F, Wang S, Mu M, Zhao Y, et al. Dietary patterns and stomach cancer: a meta-analysis. Nutr cancer. 2013;65:1105–15.

30. Magalhaes B, Peleteiro B, Lunet N. Dietary patterns and colorectal cancer: systematic review and meta-analysis. Eur J Cancer Prev. 2012;21:15–23.

31. Krusińska B, Hawrysz I, Słowińska M, Wądołowska L, Biernacki M, Czerwińska A, et al. Dietary patterns and breast or lung cancer risk: a pooled analysis of 2 case-control studies in north-eastern Poland. Adv Clin Exp Med. 2017;26:1367–75.

32. Catsburg C, Kim RS, Kirsh VA, Soskolne CL, Kreiger N, Rohan TE. Dietary patterns and breast cancer risk: a study in 2 cohorts. Am J Clin Nutr. 2015;101:817–23.

33. Link LB, Canchola AJ, Bernstein L, Clarke CA, Stram DO, Ursin G, et al. Dietary patterns and breast cancer risk in the California Teachers Study cohort. Am J Clin Nutr. 2013;98:1524–32.

34. Demetriou CA, Hadjisavvas A, Loizidou MA, Loucaides G, Neophytou I, Sieri S, et al. The mediterranean dietary pattern and breast cancer risk in Greek-Cypriot women: a case-control study. BMC Cancer. 2012;12:113.

35. Rosato V, Edefonti V, Bravi F, Bosetti C, Bertuccio P, Talamini R, et al. Nutrient-based dietary patterns and prostate cancer risk: a case–control study from Italy. Cancer Causes Control. 2014;25:525–32.

36. Vieira AR, Abar L, Vingeliene S, Chan DS, Aune D, Navarro-Rosenblatt D, et al. Fruits, vegetables and lung cancer risk: a systematic review and meta-analysis. Ann Oncol. 2016;27:81–96. https://doi.org/10.1093/annonc/mdv381.

37. Bradbury KE, Appleby PN, Key TJ. Fruit, vegetable, and fiber intake in relation to cancer risk: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC). Am J Clin Nutr. 2014;100 Suppl 1:394S–398S. https://doi.org/10.3945/ajcn.113.071357.

38. Buchner FL, Bueno-de-Mesquita HB, Linseisen J, Boshuizen HC, Kiemeney LA, Ros MM, et al. Fruits and vegetables consumption and the risk of histological subtypes of lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). Cancer Causes Control. 2010;21:357–71. https://doi.org/10.1007/s10552-009-9468-y.

39. Farvid MS, Cho E, Chen WY, Eliassen AH, Willett WC. Premenopausal dietary fat in relation to pre- and post-menopausal breast cancer. Breast Cancer Res Treat. 2014;145:255–65. https://doi.org/10.1007/s10549-014-2895-9.

40. Cao Y, Hou L, Wang W. Dietary total fat and fatty acids intake, serum fatty acids and risk of breast cancer: a meta-analysis of prospective cohort studies. Int J Cancer. 2016;138:1894–904. https://doi.org/10.1002/ijc.29938.

41. Guenther PM, Reedy J, Krebs-Smith SM, Reeve BB, Basiotis PP. Development and evaluation of the healthy eating index-2005: technical report. Center for Nutrition Policy and Promotion, U.S. Department of Agriculture, Washington DC, 2007. https://vtechworks.lib.vt.edu/bitstream/handle/10919/18682/HEI-2005TechnicalReport.pdf?sequence=3.

42. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhouser ML, et al. Addressing current criticism regarding the value of self-report dietary data. J Nutr. 2015;145:2639–45. https://doi.org/10.3945/jn.115.219634.

43. Lipworth L, Bender TJ, Rossi M, Bosetti C, Negri E, Talamini R, et al. Dietary vitamin D intake and cancers of the colon and rectum: a case-control study in Italy. Nutr Cancer. 2009;61:70–75. https://doi.org/10.1080/01635580802348633.

44. Government of Canada. Vitamin D and calcium: updated dietary reference intakes. In: Canada H, editor. Ottawa: Government of Canada; 2019.

45. Janz T, Pearson C Health at a Glance: Vitamin D blood levels of Canadians. *Statistics Canada Catalogue 82-624-X* 2015. https://www150.statcan.gc.ca/n1/pub/82-624-x/2013001/article/11727-eng.htm.

## ACKNOWLEDGEMENTS

## AUTHOR CONTRIBUTIONS

## FUNDING

## CONFLICT OF INTEREST

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41430-021-00958-7.

**Correspondence** and requests for materials should be addressed to D.R.B.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.