**Supplementary Methods**

**Assessment of dietary and lifestyle factors**

We used information collected from biennial questionnaires on major lifestyle factors for CRC, such as body weight, physical activity, smoking, family history of CRC, endoscopic screening, multivitamin use, and aspirin and nonsteroidal anti-inflammatory drug (NSAID) use. Body mass index (BMI, kg/m2), defined as weight in kilograms divided by the square of height in meters was calculated to assess overall adiposity. Using a previously validated assessment ([1](#_ENREF_1), [2](#_ENREF_2)), physical activity was assessed by summing the products of time spent at each recreational or leisure-time activity with the average metabolic equivalent (MET) for that activity. Dietary information was obtained from the validated food frequency questionnaires (FFQs) administered in 1980, 1984, 1986 and 1990 ([3](#_ENREF_3)). To assess the overall dietary pattern, we calculated a summary score ranging from 2.5 (worst) to 87.5 (best) based on individual food intake for each participant according to the Alternate Healthy Eating Index (AHEI), which is designed to target food choices and macronutrient sources associated with reduced chronic disease risk. Adherence to AHEI has been associated with a lower risk of major chronic diseases in this cohort.([4](#_ENREF_4)) To better capture long-term lifestyle and nutritional status for finer control of confounding, we calculated cumulative averages for lifestyle factors collected from questionnaires obtained at baseline enrollment (1976 questionnaire) through the time of blood collection (1990 questionnaire). Missing information was carried forward from available information from prior questionnaires. For participants with missing data across all questionnaires (<1%), we imputed using the sample median.

**Statistical analyses**

Because biomarkers were measured in at least two batches, we corrected biomarker concentrations for measurement batch using the average batch correction method,([5](#_ENREF_5)) with adjustment for age, body mass index (BMI), physical activity and case control status. We compared means (standard deviation) and medians (interquartile ranges) of continuous variables between case and control participants using paired t test and Wilcoxon signed rank test, respectively. Mantel-Haenszel chi-square test was used to compare the distributions of categorical variables between cases and controls.

We assessed the potential nonlinear relationship between biomarkers and CRC risk using stepwise restricted cubic spline analysis,([6](#_ENREF_6)) with a *P*<0.05 as the criteria for both inclusion and retention in the model. We used a likelihood ratio test to determine the significance of the non-linearity.

**Supplementary Table 1. Subsite-specific risk of colorectal cancer according to vitamin D-related biomarkers in the Nurses' Health Study (1990-2011)a**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Biomarker** | **Colon cancer** | |  | **Proximal colon cancer** | |  | **Distal colon cancer** | |  | **Rectal cancer** | |
| No. of cases/controls | OR (95% CI)b |  | No. of cases/controls | OR (95% CI)b |  | No. of cases/controls | OR (95% CI)b |  | No. of cases/controls | OR (95% CI)b |
| **Total 25(OH)D** |  |  |  |  |  |  |  |  |  |  |  |
| Quartile 1 | 83/126 | 1 (referent) |  | 53/77 | 1 (referent) |  | 30/49 | 1 (referent) |  | 27/39 | 1 (referent) |
| Quartile 2 | 79/126 | 0.90 (0.58-1.40) |  | 50/78 | 0.83 (0.47-1.47) |  | 29/48 | 0.98 (0.44-2.18) |  | 21/36 | 1.00 (0.42-2.38) |
| Quartile 3 | 70/138 | 0.74 (0.47-1.17) |  | 45/92 | 0.63 (0.36-1.11) |  | 25/46 | 0.87 (0.37-2.04) |  | 15/29 | 0.80 (0.31-2.05) |
| Quartile 4 | 56/139 | 0.64 (0.39-1.06) |  | 38/96 | 0.51 (0.27-0.96) |  | 18/43 | 0.84 (0.33-2.15) |  | 13/31 | 0.58 (0.22-1.55) |
| *P* for trend |  | 0.07 |  |  | 0.03 |  |  | 0.71 |  |  | 0.22 |
| **VDBP** |  |  |  |  |  |  |  |  |  |  |  |
| Quartile 1 | 77/132 | 1 (referent) |  | 48/85 | 1 (referent) |  | 29/47 | 1 (referent) |  | 22/35 | 1 (referent) |
| Quartile 2 | 68/133 | 0.96 (0.62-1.49) |  | 43/79 | 1.02 (0.59-1.77) |  | 25/54 | 0.75 (0.34-1.66) |  | 19/30 | 1.33 (0.49-3.62) |
| Quartile 3 | 66/135 | 0.90 (0.58-1.38) |  | 44/91 | 0.88 (0.52-1.51) |  | 22/44 | 0.97 (0.42-2.23) |  | 19/32 | 0.99 (0.39-2.49) |
| Quartile 4 | 77/129 | 1.08 (0.71-1.63) |  | 51/88 | 1.11 (0.66-1.88) |  | 26/41 | 1.08 (0.51-2.28) |  | 16/38 | 0.69 (0.25-1.92) |
| *P* for trend |  | 0.87 |  |  | 0.94 |  |  | 0.70 |  |  | 0.25 |
| **Free 25(OH)D** |  |  |  |  |  |  |  |  |  |  |  |
| Quartile 1 | 83/120 | 1 (referent) |  | 53/76 | 1 (referent) |  | 30/44 | 1 (referent) |  | 29/47 | 1 (referent) |
| Quartile 2 | 68/135 | 0.77 (0.50-1.20) |  | 40/90 | 0.58 (0.33-1.04) |  | 28/45 | 1.01 (0.46-2.18) |  | 21/30 | 0.95 (0.41-2.20) |
| Quartile 3 | 68/140 | 0.71 (0.46-1.10) |  | 48/92 | 0.69 (0.40-1.20) |  | 20/48 | 0.62 (0.28-1.36) |  | 10/25 | 0.58 (0.22-1.53) |
| Quartile 4 | 69/134 | 0.78 (0.50-1.20) |  | 45/85 | 0.71 (0.41-1.25) |  | 24/49 | 0.78 (0.36-1.71) |  | 16/33 | 0.88 (0.36-2.17) |
| *P* for trend |  | 0.28 |  |  | 0.39 |  |  | 0.47 |  |  | 0.65 |
| **Bioavailable 25(OH)D** |  |  |  |  |  |  |  |  |  |  |  |
| Quartile 1 | 75/122 | 1 (referent) |  | 46/78 | 1 (referent) |  | 29/44 | 1 (referent) |  | 24/42 | 1 (referent) |
| Quartile 2 | 68/133 | 0.90 (0.56-1.45) |  | 40/84 | 0.77 (0.41-1.41) |  | 28/49 | 1.01 (0.44-2.33) |  | 20/33 | 1.17 (0.46-2.98) |
| Quartile 3 | 66/139 | 0.85 (0.54-1.35) |  | 49/94 | 0.86 (0.48-1.54) |  | 17/45 | 0.58 (0.25-1.39) |  | 14/30 | 0.72 (0.28-1.86) |
| Quartile 4 | 79/135 | 1.03 (0.66-1.61) |  | 51/87 | 0.99 (0.56-1.73) |  | 28/48 | 1.03 (0.46-2.33) |  | 18/30 | 1.39 (0.52-3.72) |
| *P* for trend |  | 0.79 |  |  | 0.77 |  |  | 0.92 |  |  | 0.66 |

Abbreviations: 25(OH)D, 25-hydroxyvitamin D; CI, confidence interval; OR, odds ratio; VDBP, vitamin D binding protein.

a *P* for heterogeneity between colon and rectal cancers: 0.79 for total 25(OH)D, 0.26 for VDBP, 0.86 for free 25(OH)D, 0.50 for bioavailable 25(OH)D. *P* for heterogeneity between proximal colon, distal colon and rectal cancers: 0.70 for total 25(OH)D, 0.49 for VDBP, 0.99 for free 25(OH)D, 0.78 for bioavailable 25(OH)D. Contrast test method was used to calculate the *P* for hetergoneity in the associations between subsites.

b Conditional logistic regression adjusted for matching factors (age and time of blood draw), family history of colorectal cancer, multivitamin use, pack-years of smoking (0, 0-9.9, 10-24.9, 25-49.9, ≥50), body mass index (continuous), physical activity (≤5, 5.1-10, 10.1-20, >20 METs), regular aspirin/NSAID use, postmenopausal status and hormone use (premenopausal, never use of postmenopausal hormone, ever use of postmenopausal hormone), calcium intake (in quartiles), and Alternative Healthy Eating Index (in quartiles).

**Supplementary Table 2. Stratified analysis of the association of plasma total 25(OH)D and VDBP with risk of colorectal cancer in the Nurses' Health Study (1990-2011)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Analysis** | **Quartile 1** | **Quartile 2** | **Quartile 3** | **Quartile 4** | ***P* for trend** | ***P* for interaction a** |
| **Total 25(OH)D** |  |  |  |  |  |  |
| Age |  |  |  |  |  | 0.07 |
| < 60 years |  |  |  |  |  |  |
| No. of cases/controls | 59/106 | 49/84 | 47/90 | 43/85 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.06 (0.63-1.79) | 0.84 (0.49-1.42) | 0.87 (0.50-1.51) | 0.38 |  |
| ≥ 60 years |  |  |  |  |  |  |
| No. of cases/controls | 56/66 | 56/88 | 39/81 | 29/89 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.71 (0.42-1.22) | 0.55 (0.31-0.98) | 0.35 (0.19-0.67) | 0.004 |  |
| Body mass index |  |  |  |  |  | 0.03 |
| < 25.0 kg/m2 |  |  |  |  |  |  |
| No. of cases/controls | 56/90 | 60/105 | 55/113 | 50/123 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.89 (0.54-1.46) | 0.70 (0.43-1.15) | 0.61 (0.36-1.04) | 0.05 |  |
| 25.0-29.9 kg/m2 |  |  |  |  |  |  |
| No. of cases/controls | 30/57 | 29/45 | 28/43 | 20/39 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.27 (0.62-2.62) | 1.13 (0.56-2.28) | 0.96 (0.45-2.08) | 0.66 |  |
| ≥ 30.0 kg/m2 |  |  |  |  |  |  |
| No. of cases/controls | 29/25 | 16/22 | 3/15 | 2/12 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.63 (0.26-1.53) | 0.16 (0.04-0.65) | 0.14 (0.03-0.76) | 0.002 |  |
| Physical activity |  |  |  |  |  | 0.78 |
| < 10 MET-hours/week |  |  |  |  |  |  |
| No. of cases/controls | 64/104 | 50/80 | 42/90 | 29/62 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.00 (0.61-1.65) | 0.72 (0.42-1.21) | 0.75 (0.40-1.41) | 0.08 |  |
| ≥ 10 MET-hours/week |  |  |  |  |  |  |
| No. of cases/controls | 51/68 | 55/92 | 44/81 | 43/112 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.78 (0.46-1.33) | 0.68 (0.39-1.18) | 0.49 (0.28-0.86) | 0.01 |  |
| Smoking status |  |  |  |  |  | 0.25 |
| Never smoking |  |  |  |  |  |  |
| No. of cases/controls | 50/74 | 47/98 | 43/72 | 30/70 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.69 (0.40-1.17) | 0.87 (0.50-1.50) | 0.60 (0.33-1.09) | 0.23 |  |
| Ever smoking |  |  |  |  |  |  |
| No. of cases/controls | 65/98 | 58/74 | 43/99 | 42/104 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.14 (0.70-1.87) | 0.57 (0.34-0.95) | 0.58 (0.34-0.99) | 0.02 |  |
| Alcohol consumption |  |  |  |  |  | 0.74 |
| < 5 g/day | 74/110 | 70/121 | 59/108 | 38/99 |  |  |
| No. of cases/controls |  |  |  |  |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.87 (0.56-1.35) | 0.77 (0.47-1.24) | 0.54 (0.31-0.92) | 0.01 |  |
| ≥ 5 g/day |  |  |  |  |  |  |
| No. of cases/controls | 41/62 | 35/51 | 27/63 | 34/75 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.96 (0.52-1.78) | 0.60 (0.31-1.15) | 0.67 (0.36-1.25) | 0.03 |  |
| Regular use of aspirin/NSAIDs |  |  |  |  |  | 0.03 |
| Non-users | 59/93 | 60/96 | 46/82 | 43/77 |  |  |
| No. of cases/controls |  |  |  |  |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.99 (0.60-1.62) | 0.84 (0.49-1.44) | 0.87 (0.49-1.54) | 0.59 |  |
| Users |  |  |  |  |  |  |
| No. of cases/controls | 56/79 | 45/76 | 40/89 | 29/97 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.85 (0.50-1.43) | 0.60 (0.35-1.03) | 0.39 (0.21-0.70) | 0.006 |  |
| **VDBP** |  |  |  |  |  |  |
| Age |  |  |  |  |  | 0.65 |
| < 60 years |  |  |  |  |  |  |
| No. of cases/controls | 49/94 | 47/91 | 51/84 | 51/96 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.02 (0.31-3.31) | 2.32 (0.72-7.42) | 1.10 (0.36-3.35) | 0.62 |  |
| ≥ 60 years |  |  |  |  |  |  |
| No. of cases/controls | 54/79 | 44/81 | 37/89 | 45/75 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.94 (0.56-1.59) | 1.22 (0.73-2.04) | 1.01 (0.62-1.66) | 0.99 |  |
| Body mass index |  |  |  |  |  | 0.41 |
| < 25.0 kg/m2 |  |  |  |  |  |  |
| No. of cases/controls | 49/104 | 61/107 | 51/105 | 60/115 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.26 (0.77-2.06) | 1.09 (0.65-1.80) | 1.22 (0.75-1.97) | 0.93 |  |
| 25.0-29.9 kg/m2 |  |  |  |  |  |  |
| No. of cases/controls | 30/57 | 29/45 | 28/43 | 20/39 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.85 (0.42-1.73) | 0.91 (0.46-1.80) | 1.03 (0.51-2.06) | 0.90 |  |
| ≥ 30.0 kg/m2 |  |  |  |  |  |  |
| No. of cases/controls | 54/69 | 30/65 | 37/68 | 36/56 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.25 (0.08-0.74) | 0.41 (0.15-1.13) | 0.36 (0.12-1.08) | 0.16 |  |
| Physical activity |  |  |  |  |  | 0.71 |
| < 10 MET-hours/week |  |  |  |  |  |  |
| No. of cases/controls | 48/80 | 46/89 | 43/83 | 48/84 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.90 (0.53-1.52) | 0.89 (0.53-1.49) | 1.00 (0.59-1.68) | 0.91 |  |
| ≥ 10 MET-hours/week |  |  |  |  |  |  |
| No. of cases/controls | 55/93 | 45/83 | 45/90 | 48/87 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.93 (0.54-1.58) | 0.89 (0.53-1.49) | 0.93 (0.56-1.55) | 0.55 |  |
| Smoking status |  |  |  |  |  | 0.85 |
| Never smoking |  |  |  |  |  |  |
| No. of cases/controls | 52/81 | 33/74 | 43/92 | 42/67 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.66 (0.38-1.16) | 0.70 (0.42-1.18) | 0.93 (0.55-1.59) | 0.62 |  |
| Ever smoking |  |  |  |  |  |  |
| No. of cases/controls | 51/92 | 58/98 | 45/81 | 54/104 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 1.15 (0.70-1.88) | 1.10 (0.65-1.87) | 1.00 (0.61-1.65) | 0.61 |  |
| Alcohol consumption |  |  |  |  |  | 0.94 |
| < 5 g/day |  |  |  |  |  |  |
| No. of cases/controls | 68/114 | 61/110 | 52/113 | 60/101 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.96 (0.61-1.52) | 0.80 (0.50-1.26) | 0.99 (0.63-1.55) | 0.71 |  |
| ≥ 5 g/day |  |  |  |  |  |  |
| No. of cases/controls | 35/59 | 30/62 | 36/60 | 36/70 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.80 (0.42-1.50) | 1.05 (0.57-1.95) | 0.93 (0.50-1.74) | 0.80 |  |
| Regular use of aspirin/NSAIDs |  |  |  |  |  | 0.50 |
| Non-users |  |  |  |  |  |  |
| No. of cases/controls | 51/83 | 49/86 | 49/93 | 59/86 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.94 (0.55-1.60) | 0.87 (0.51-1.47) | 1.10 (0.66-1.83) | 0.95 |  |
| Users |  |  |  |  |  |  |
| No. of cases/controls | 52/90 | 42/86 | 39/80 | 37/85 |  |  |
| Multivariable OR (95% CI) b | 1 (referent) | 0.86 (0.50-1.49) | 0.93 (0.55-1.57) | 0.82 (0.48-1.40) | 0.22 |  |

Abbreviations: 25(OH)D, 25-hydroxyvitamin D; CI, confidence interval; OR, odds ratio; VDBP, vitamin D binding protein.

a Wald test was used for the product term between stratified variable (binary) and main exposure (continuous) except BMI, for which a likelihood ratio test was used to compare the model with the product terms of BMI (three categories) and main exposure (continuous) to that without these terms.

b Adjusted for matching factors (age and time of blood draw), family history of colorectal cancer, multivitamin use, pack-years of smoking (0, 0-9.9, 10-24.9, 25-49.9, ≥50), alcohol consumption (0, 0-5.0, 5.1-15.0, >15 g/d), body mass index (continuous), physical activity (≤5, 5.1-10, 10.1-20, >20 METs), regular aspirin/NSAID use, postmenopausal status and hormone use (premenopausal, never use of postmenopausal hormone, ever use of postmenopausal hormone), calcium intake (in quartiles), and Alternative Healthy Eating Index (in quartiles).

**References**

1. Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. Int J Epidemiol. 1994;23:991-9.

2. Chasan-Taber S, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. Epidemiology. 1996;7:81-6.

3. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. American journal of epidemiology. 1985;122:51-65.

4. McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, et al. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. The American journal of clinical nutrition. 2002;76:1261-71.

5. Rosner B, Cook N, Portman R, Daniels S, Falkner B. Determination of blood pressure percentiles in normal-weight children: some methodological issues. Am J Epidemiol. 2008;167:653-66.

6. Durrleman S, Simon R. Flexible regression models with cubic splines. Statistics in medicine. 1989;8:551-61.