

Supplementary Table 1: Comparison of the covariates and factors in the published models with their reconstruction in the external validation

Variable	Authors	Covariate or baseline factor for population incidence calculation	Covariate or baseline factor for population incidence calculation in the UK Biobank	Covariate or baseline factor for population incidence calculation in EPIC
Age	All	Age on entering the risk period	Variable available	Variable available
Ethnicity	Freedman et al <sup>1</sup>	Ethnicity of the participant. The original model was for White participants. However the published code <sup>2</sup> also provided tailoring of the model to African-Americans, Hispanics and Asian-Americans.	Variable available. Due to the predominance of White participants and the tailoring of the model to the USA population, the model was validated only in the White population.	Variable not available. No suitable surrogates available. Set all to White.
	Wells et al <sup>3</sup>	Ethnicity of the participant. Categorical variable: (i) Latino, (ii) Hawaiian, (iii) White, (iv) Black and (v) Japanese	Variable available. Due to the predominance of White participants and the tailoring of the model to the USA population, the model was validated only in the White population.	Variable not available. No suitable surrogates available. Set all to White.
Education	Wells et al <sup>3</sup>	Years of education. Continuous variable. The online version of this model had a maximum value of 20 years and so this cap was used in the validation. <sup>4</sup>	Variable not available. Constructed by deducting school starting age from age at completing full time education. If this resulted in a negative value it was reset to zero.	Variable not available. Constructed by deducting country specific school starting ages from the age at leaving school. If this resulted in a negative value it was reset to zero.
Height	Colditz et al <sup>5</sup>	Height Dichotomous variable (threshold 5 feet 7 inches or taller for women and 5 feet 10 inches or taller for men)	Variable available.	Variable available. The adjusted value of height was used which incorporated measured heights in EPIC along with the addition of calibrated self-reported heights from the Oxford Health Conscious Group. <sup>6</sup>
	Shin et al <sup>7</sup>	Height Categorical variable: (i) ≤151cm, (ii) > 151 to ≤ 155cm, (iii) >155 to ≤ 158cm, (iv) >158cm.	Variable available.	Variable available. The adjusted value of height was used which incorporated measured heights in EPIC along with the addition of calibrated self-reported heights from the Oxford Health Conscious Group. <sup>6</sup>
Body mass index	Colditz et al <sup>5</sup> Driver et al <sup>8</sup> Freedman et al <sup>1</sup> Ma et al <sup>9</sup> Shin et al <sup>7</sup> Steffen et al <sup>10</sup> Wells et al <sup>3</sup>	Body Mass Index (kilograms per meter squared) Continuous or categorical variable.	Variable available	Variable available The adjusted value of body mass index was used. This was calculated from measured heights and weights (adjusted for clothing) with the addition of calibrated self-reported heights and weights from the Oxford Health Conscious Group. <sup>6</sup>

Family History	Colditz et al <sup>5</sup>	Family History (parent or sibling) of colon cancer. Dichotomous variable (presence or absence)	Variable not available Constructed by incorporating history of “bowel cancer” in (i) father, (ii) mother and (iii) siblings. It was not possible to differentiate between colorectal and colon cancer, therefore all cases were assumed to be colon cancer.	Variable not available. As the variables for the construction of a surrogate were not widely available and participant numbers were further constrained by the requirement for complete observations across the model’s variables it was set to 0 for all participants.
	Freedman et al <sup>1</sup>	Family History (first degree relative) of colorectal cancer. Categorical variable: (i) 0 relatives, (ii) 1 relative and (iii) 2 or more relatives.	Variable not available. Constructed by incorporating history of “bowel cancer” in (i) father, (ii) mother and (iii) siblings. No available data for history of “bowel cancer” in a participant’s offspring.	Variable not available. As the variables for its construction were not widely available and participant numbers were further constrained by the requirement for complete observations across the model’s variables it was set to 0 for all participants.
	Shin et al <sup>7</sup>	Family History of cancer Dichotomous variable (Presence or absence)	Variable not available. Constructed based on family history of “bowel cancer”, “breast cancer”, “lung cancer” and “prostate cancer” in (i) father, (ii) mother and (iii) siblings. No available data for history of cancer in a participant’s offspring.	Variable not available. Set to zero for all participants.
	Taylor et al <sup>11</sup>	Family History (first through third degree relatives) of colorectal cancer. Categorical variable with multiple categories.	Variable not available. Constructed by incorporating history of “bowel cancer” in (i) father, (ii) mother and (iii) siblings. No available data for history of “bowel cancer” in a participant’s offspring. As the cohort only contained first degree relatives the familial relative risk for this degree of resolution was taken from a previously published paper. <sup>12</sup>	Variable not available. However as, aside from age, this was the only variable included in the model it was reconstructed since the numbers eligible were not further constrained by the availability of data for other variables. It was constructed by incorporating history of colorectal cancer in (i) father, (ii) mother and (iii) siblings. There was no available data for history of colorectal cancer in a participant’s offspring. As the model only contained first degree relatives the familial relative risk for this degree of resolution was taken from a previously published paper. <sup>12</sup>
	Wells et al <sup>3</sup>	Family History (first degree relative) of colon cancer Dichotomous variable (presence or absence)	Variable not available. Constructed by incorporating history of “bowel cancer” in (i) father, (ii) mother and (iii) siblings. It was not possible to differentiate between colorectal and colon cancer, therefore all cases were assumed to be colon cancer. No available data for history of “bowel cancer” in a participant’s offspring.	Variable not available. As the variables for the construction of a surrogate were not widely available and participant numbers were further constrained by the requirement for complete observations across the model’s variables it was set to 0 for all participants.
Diabetes	Steffen et al <sup>10</sup>	History of Diabetes Dichotomous variable (presence or absence)	Variable available	Variable available
	Wells et al <sup>3</sup>	History of Diabetes Dichotomous variable (presence or absence)	Variable available	Variable available

Endoscopy / Screening	Colditz et al <sup>5</sup>	History of faecal occult blood testing or sigmoidoscopy in the preceding 10 years beginning at age 50. Dichotomous variable (presence or absence)	Variable not available Utilised similar variables which detailed prior bowel cancer screening (faecal occult blood tests or endoscopy) as well as the time since the last screen. This differed from the original variable construction by the inclusion of colonoscopy investigations in addition to sigmoidoscopy. The 50 year age threshold was not applied as this was not a limitation in our data.	Variable not available. No suitable surrogates available so set to absent for all participants.
	Freedman et al <sup>1</sup>	Endoscopy history in the preceding 10 years and the presence of polyps. Categorical variable: (i) Endoscopy and no history of polyps in the last 10 years. (ii) No endoscopy in the last 10 years. (iii) Endoscopy and history of polyps in the last 10 years. (iv) Endoscopy and polyps unknown.	Variable not available. Constructed from variables detailing prior bowel cancer screening (faecal occult blood tests or endoscopy), time since the last screen and the presence of colon or rectal adenoma/polyps. This differed from the original variable construction as (i) it includes faecal occult blood tests, therefore widening the remit and (ii) there was no indication of when the polyp (if present) was diagnosed. As such the 10 year threshold was changed to “ever”, again widening the remit of the model.	Variable not available. Constructed from variables that detailed the presence of large bowel polyps and the age at which they were identified. It was inferred that those with polyps diagnosed within the preceding 10 years had undergone endoscopy and so were placed in category (iii), while those with polyps diagnosed more than 10 years prior would have undergone further surveillance and so were placed in category (i). All others were placed in category (iv).
	Steffen et al <sup>10</sup>	Ever had colorectal cancer screening Dichotomous variable (presence or absence)	Variable available.	Variable not available. No suitable surrogates were available so set to absent for all participants.
Inflammatory Bowel Disease	Colditz et al <sup>5</sup>	Inflammatory Bowel Disease for 10 or more years	Variable not available It was constructed by utilising self-reported data for diagnosis but without quantification of the duration.	Variable not available. No suitable surrogates available so set to absent for all participants.
Menopause/ oestrogen use	Colditz et al <sup>5</sup>	Use of birth control pills for 5 years or more Dichotomous variable (presence or absence)	Variable not available Constructed from variables detailing (i) ever use of the oral contraceptive pill (OCP), (ii) age at starting the OCP and (iii) age at which the OCP was last used.	Variable not available. Constructed from variables detailing length of OCP use and ever use of the OCP.
	Colditz et al <sup>5</sup>	Use of postmenopausal hormones for 5 years or more Dichotomous variable (presence or absence)	Variable not available Constructed from variables detailing (i) ever use of hormone replacement therapy (HRT), (ii) age at starting HRT and (iii) age at which HRT was last used.	Variable not available. Constructed from variables detailing length of HRT use and ever use of HRT.
	Freedman et al <sup>1</sup>	Oestrogen status in the preceding two years. Assessed by menopause status and the use of HRT. Dichotomous variable (positive or negative).	Variable not available. Constructed using (i) age at menopause, (ii) history of oophorectomy, (iii) age at oophorectomy, (iv) HRT ever use, (v) age at starting HRT and (vi) age at last use of HRT.	Variable not available. Constructed using (i) age at menopause, (ii) history of oophorectomy, (iii) age at oophorectomy, (iv) HRT ever use, (v) age at starting HRT and (vi) duration of HRT use
	Wells et al <sup>3</sup>	Oestrogen use. Defined as “female hormones administered by pill, injection or patch for menopause or other reasons” <sup>3</sup> . Categorical variable: (i) current use, (ii) prior use, (iii) never use.	Variable not available. Constructed from (i) ever use of the OCP, (ii) age last used OCP, (iii) ever use of HRT and (iv) age last used HRT.	Variable not available Constructed from (i) ever use of the OCP, (ii) current use of the OCP, (iii) ever use of HRT and (iv) current use of HRT.

Non-steroidal anti-inflammatory drugs (NSAIDs)	Colditz et al <sup>5</sup>	Daily aspirin use for 15 years or more. Dichotomous variable (presence or absence)	Variable not available. It was constructed by utilising a variable that detailed the regular use of aspirin but without further quantification of the frequency or duration of use.	Variable not available. A variable was present which identified regular users of NSAIDs but without further quantification of the type, frequency or duration of use. Furthermore it was only available in a limited number of individuals. As such it was set to nil for all participants.
	Freedman et al <sup>1</sup>	In the original model this was classified as use of aspirin and other NSAIDs for at least three times a week, for a minimum of 1 month, in the preceding year. However the published script for men separated this into aspirin and non-aspirin containing NSAIDs. <sup>2</sup> Dichotomous variable (presence or absence)	Variable not available. It was constructed by utilising variables that detailed the regular use of aspirin and ibuprofen but without further quantification of the frequency or duration of use. In the female model the two variables subtypes were combined while in the male model they were entered as separate variables.	Variable not available. A variable was present which identified regular users of NSAIDs but without further quantification of the type, frequency or duration of use. Furthermore it was only available in a limited number of individuals. As such it was set to nil for all participants (including in the male model where NSAIDs were subdivided).
	Wells et al <sup>3</sup>	Use of Aspirin (men) or NSAIDs (women) twice per week. Categorical variable: (i) yes, currently, (ii) yes, not currently and (iii) no.	Variable not available. It was constructed by utilising variables that detailed the regular use of aspirin and ibuprofen but without further quantification of the frequency or duration of use.	Variable not available. A variable was present which identified regular users of NSAIDs but without further quantification of the type, frequency or duration of use. Furthermore it was only available in a limited number of individuals. As such it was set to nil for all participants.
Physical activity	Colditz et al <sup>5</sup>	In its validation <sup>13</sup> physical activity was defined in metabolic equivalent tasks (MET). Dichotomous variable (threshold 25 MET hours per week or greater).	Variable not available. Constructed from variables detailing the number of days per week the participant undertook moderate or vigorous physical activity and the usual number of minutes spent doing so. This was converted into a total MET score (hours per week) following the International Physical Activity Questionnaire Guidelines. <sup>14</sup>	Variable available (recreational and household activities)
	Freedman et al <sup>1</sup>	Current leisure-time vigorous activity. Categorical variable: (i) 0 hours per week, (ii) 0-2 hours per week, (iii) 2-4 hours per week, (iv) 4+ hours per week.	Variable not available. Constructed from variables detailing the number of days per weeks the participant undertook vigorous physical activity as well as the usual number of minutes spent undertaking it on a typical day. This was converted into the number of hours per week following the International Physical Activity Questionnaire Guidelines. <sup>14</sup>	Variable available (MET hours per week from vigorous recreational physical activity were converted into hours per week)
	Ma et al <sup>9</sup>	MET hours per day were calculated based on time spent at: (i) heavy physical work or strenuous activity, (ii) walking or standing, (iii) sedentary and (iv) asleep or others. Continuous variable.	Variable not available. Constructed from variables detailing the number of days per week and duration the participant walked, or undertook moderate or vigorous physical activity (following the International Physical Activity Questionnaire Guidelines <sup>14</sup> ) as well as the duration of sleep per day. Walking was the sole contributor to the walking or standing variable, as was sleep to the asleep or others variable. Moderate and vigorous physical activity were summed to create the heavy physical work / strenuous activity variable. Sedentary was then derived from the remaining time in the day. These were then all	Variable not available. Constructed from variables which provided MET hours per week from walking, cycling, gardening, sport, housework, DIY and stair climbing after converting them into hours of activity per day. All but walking were summed to create the heavy physical work/strenuous activity variable, while walking was the sole contributor to the walking or standing variable. The remaining time was then classified as asleep or others unless it exceeded 8 hours in which case the excess time contributed to the sedentary category. These were then all converted into METs using the authors' conversion factors and summed.

			converted into METs using the authors' conversion factors and summed.	
	Wells et al <sup>3</sup>	Hours of moderate physical activity per day. However the description of the type of physical activity used within their derivation cohort was "moderate or strenuous activity per day" and so this definition was utilised. The online version of this model had a maximum value of 4 hours and so this cap was used in the validation. <sup>4</sup> Continuous variable.	Variable not available. Constructed from variables detailing the number of days per weeks the participant undertook moderate or vigorous physical activity and the usual duration. This was summed to calculate total number of hours per day following the International Physical Activity Questionnaire Guidelines. <sup>14</sup>	Variable not available. Constructed by converting variables detailing MET hours per week from walking, cycling, gardening, sport, housework, DIY and stair climbing into hours per day and summing.
Smoking	Driver et al <sup>8</sup>	Smoking Status Dichotomous variable (never smoker or current/previous smoker)	Variable available.	Variable available.
	Freedman et al <sup>1</sup>	Years of cigarette smoking Categorical variable: (i) 0, (ii) >0 to <15, (iii) ≥15 to <35 and (iv) ≥35	Variable not available. Constructed from variables detailing (i) age started cigarette smoking, (ii) age stopped cigarette smoking and (iii) periods of stopping smoking.	Variables available.
	Freedman et al <sup>1</sup>	Usual number of cigarettes smoked per day for current and former smokers. Categorical variable: (i) never, (ii) >0 to <11, (iii) ≥11 to ≤20 and (iv) >20.	Variable not available. Constructed from variables available detailing number of cigarettes currently or previously smoked a day.	Variables available.
	Ma et al <sup>9</sup>	Smoking Status Categorical variable: (i) Never smokers, (ii) current smokers (iii) former smokers.	Variable available.	Variable available.
	Steffen et al <sup>10</sup>	Smoking Status Categorical variable: (i) Never smokers, (ii) current smokers (iii) former smokers.	Variable available.	Variable available.
	Wells et al <sup>3</sup>	Pack years of smoking. The online version of this model had a maximum value of 50 pack years and so this cap was used in the validation <sup>4</sup> Continuous variable.	Variable not available. Constructed using a purpose built algorithm that incorporated variables detailing number of cigarettes smoked, duration of smoking and periods of smoking cessation. This was based on the pack year calculation developed for the UK Biobank data by Professor Martin Tobin, University of Leicester. <sup>15</sup>	Variable not available. Constructed from variables detailing lifetime average number of cigarettes smoked and duration of smoking.
Alcohol	Colditz et al <sup>5</sup>	Frequency of alcohol consumption (more than seven servings per week) Dichotomous variable (presence or absence)	Variable available (the category of "daily or almost daily" was regarded to equate to more than seven servings per week).	Variable not available Constructed from a surrogate, detailing grams of alcohol per day, after conversion to drinks per day (12 grams of alcohol per drink).
	Driver et al <sup>8</sup>	Frequency of alcohol consumption (once a week or more) Dichotomous variable (presence or absence)	Variable available (the category of "once or twice a week" was regarded to equate to once a week or more).	Variable not available Constructed from a surrogate, detailing grams of alcohol per day, after conversion to drinks per week (12 grams of alcohol per drink).

	Ma et al <sup>9</sup>	Alcohol consumption Categorical variable: (i) never, (ii) occasional, (iii) < 300 grams per week, (iv) ≥300 grams per week.	Variable not available. Constructed from variables detailing (i) the frequency of alcohol consumption as a whole as well as (ii) the frequency and volume of different types of alcoholic drinks. Typical values for units of alcohol in different drinks as well as the number of grams of alcohol per unit (8g) were obtained from a UK Government publication <sup>16</sup> .	Variable not available Constructed from variables detailing frequency of alcohol consumption and grams of alcohol per day.
	Shin et al <sup>7</sup>	Alcohol consumption Categorical variable: (i) 0 g/day, (ii) 1-14.9 g/day, (iii) 15-24.9 g/day, (iv) 25 or more g/day.	Variable not available. Constructed from variables detailing the frequency and volume of different types of alcoholic drinks. Typical values for units of alcohol in different drinks as well as the number of grams of alcohol per unit (8g) were obtained from a UK Government publication. <sup>16</sup>	Variable available.
	Steffen et al <sup>10</sup>	Drinks per day Continuous variable	Variable not available. Constructed from variables documenting the frequency of drinking different types of alcoholic drinks.	Variable not available Constructed from a surrogate, detailing grams of alcohol per day, after conversion to drinks per day (12 grams of alcohol per drink).
	Wells et al <sup>3</sup>	Drinks per day Continuous variable The online version of this model had a maximum value of 12 drinks and so this cap was used in the validation. <sup>4</sup>	Variable not available. Constructed from variables documenting the frequency of drinking different types of alcoholic drinks.	Variable not available Constructed from a surrogate, detailing grams of alcohol per day, after conversion to drinks per day (12 grams of alcohol per drink).
Red Meat/Meat	Colditz et al <sup>5</sup>	Eating 1.2 servings or more of red meat a day Updated definition and relative risk were taken from a subsequent validation paper. <sup>13</sup> Dichotomous variable (presence or absence)	Variable not available Constructed by merging beef, lamb/mutton, pork and processed meat intake. Incorporated processed meat due to the fact that red meat would be a substantial constituent of it.	Variable not available Constructed by merging red and processed meat (due to the fact that red meat would be a substantial constituent of processed meat). A serving was taken to be 113g.
	Shin et al <sup>7</sup>	Meat consumption per week Categorical variable: (i) ≤1 time, (ii) 2-3 times, (iii) ≥ 4 times.	Variable not available. Constructed by merging beef, lamb/mutton, pork, poultry and processed meat intake.	Variable not available. Variable detailing the total weight of meat intake per day was extrapolated to per week and converted to servings (113g was taken to one serving).
	Wells et al <sup>3</sup>	Ounces of red meat intake per day The online version of this model had a maximum value of 5 ounces and so this cap was used in the validation. <sup>4</sup> Continuous variable.	Variable not available Constructed by merging beef, lamb/mutton, pork and processed meat intake. Incorporated processed meat due to the fact that red meat would be a substantial constituent of it.	Variable not available Constructed by merging red and processed meat (due to the fact that red meat would be a substantial constituent of processed meat). Converted to ounces.
Vegetables	Colditz et al <sup>5</sup>	Eating 3 or more servings of vegetables per day Dichotomous variable (presence or absence)	Variable not available Constructed by summing two surrogates detailing tablespoons of raw and cooked vegetable eaten per day.	Variable not available. Constructed from a surrogate, detailing grams of vegetables eaten per day.
	Freedman et al <sup>1</sup>	Eating 5 or more servings of vegetables per day Dichotomous variable (presence or absence)	Variable not available Constructed by summing two surrogates detailing tablespoons of raw and cooked vegetable eaten per day.	Variable not available. Constructed from a surrogate, detailing grams of vegetables eaten per day.

---

Multivitamins	Colditz et al <sup>5</sup>	Regular multivitamin use Dichotomous variable (presence or absence)	Variable available	Variable not available "Have you taken vitamins/minerals" was used as a surrogate
	Wells et al <sup>3</sup>	Regular multivitamin use Dichotomous variable (presence or absence)	Variable available	Variable not available "Have you taken vitamins/minerals" was used as a surrogate

---

Supplementary Table 2: General characteristics of the populations used for the original model derivation, The European Prospective Investigation into Cancer and Nutrition (EPIC) validation and UK Biobank validation

Model	Population Origin	Country	Characteristics / Eligibility Criteria	Mean Age (or as described)	Mean BMI (kg/m <sup>2</sup> ) (or as described)	Participants (cases)	Outcome Cancer type
<b>Derivation Cohort</b>							
Colditz <sup>5</sup>	Model was constructed through expert opinion, and did not utilise a specific cohort						
Driver <sup>8</sup>	Physician's Health Study. Followed from 1982 to 2004.	USA	Male aged 40-84. Apparently healthy. No history of cardiovascular disease or other serious illnesses, no history of cancer (bar non-melanoma skin cancer) and no indication or contraindication to analgesia usage.	40-49 (41.1%) 50-59 (33.9%) 60-69 (18.5%) ≥70 (6.5%)	<25 (57.6%) 25-29.9 (38.2%) ≥30 (4.2%)	21,581 (485)	Colorectal
Freedman <sup>1</sup>	Case-control 1991 to 1994	USA	Male and Female. Cases excluded if prior colorectal cancer diagnosis, familial adenomatous polyposis, ulcerative colitis or Crohn's documented in the pathology report. Controls matched by sex and age (5 year age group)	50-59: Case = 319, Control = 347 60-69: Case = 606, Control = 751 ≥70: Case = 671, Control = 876	Male: ≤24.9: Case = 220, Control = 337 25-≤30: Case = 426, Control = 514 >30: Case = 245, Control = 207  Female (Distal Colon Cases only) ≤29.9: Case= 246, Control = 726 ≥30: Case= 88, Control = 190	1,599 cases 1,974 controls	Colon
	Case-control 1997 to 2002	USA	Male and Female. Cases excluded if prior colorectal cancer diagnosis, familial adenomatous polyposis, ulcerative colitis or Crohn's documented in the pathology report. Controls matched by sex and age (5 year age group)	50-59: Case = 175, Control = 225 60-69: Case = 261, Control = 316 ≥70: Case = 228, Control = 318	Female: ≤29.9: Case= 192, Control =298 ≥30: Case= 75, Control =83	664 cases 859 controls	Rectal
Ma <sup>9</sup>	Japan Public Health Center-based Prospective Study. Cohort II. Followed from 1993 to 2005.	Japan	Male aged 40 – 69. Excluded if a history of cardiovascular disease or cancer.	52.9 (S.D. 8.8)	23.4 (S.D. 2.9)	28,115 (543)	Colorectal
Shin <sup>7</sup>	Korean National Health System. Followed from 1996-1997 to 2004-2005	Korea	Male and Female aged 30 - 80. Participated in a National Health Insurance Corporation medical examination.	Not available	Not available	1,326,058 (9,147)	Colorectal



Excluded if prior cancer diagnosis or missing variables

Steffen <sup>10</sup>	45 and Up Study. Followed from 2006-2008 to 2011	Australia	Male and Female aged 45 or older Excluded if prevalent cancer (bar non-melanoma skin cancer), missing study entry date, body mass index <15 or >50 kg/m <sup>2</sup> or extreme values for activity or food intake.	61.2 (S.D. 10.8)	27.0 (S.D. 4.9)	197,874 (1,103)	Colorectal
Taylor <sup>11</sup>	Utah Population Database (UPDB).	USA	Male and Female Familial relative risk was calculated from the descendants of original Utah pioneers with ≥3 generations of genealogy data. <sup>12</sup> The CRC morbidity rates and all-cause mortality rates were calculated from UPDB data (1981 to 1985). <sup>11</sup>			2,327,327 (94,931)	Colorectal
Wells <sup>3</sup>	Multiethnic Cohort Study. Followed from 1993-1996 to 2004	USA	Male and Female aged over 45. Excluded if a prior diagnosis of colorectal cancer or adenomatous polyps	Male Case = 64.2 (S.D. 7.8) Non-case = 59.8 (S.D. 8.9)  Female Case = 64 (S.D. 7.9) Non-case = 59.5 (S.D. 8.8)	Male Case = 26.6 (S.D. 4.2) Non-case = 26.6 (S.D. 4.1)  Female Case = 26.6 (S.D. 5.7) Non-case = 26.4 (S.D. 5.5)	180,630 (2,762)	Colorectal
<b>Validation Cohorts</b>							
	UK Biobank †. Followed 2006-2010 to 2015.	Great Britain	Male and Female aged 37 – 73 Excluded if prior cancer diagnosis (bar non-melanoma skin cancer), zero or negative follow-up. Additionally new diagnosis of appendiceal cancers were excluded with the participants being censored at diagnosis.	56.3 (S.D. 8.1)	27.4 (S.D. 4.8)	475,629 (2,617)	Colorectal
	European Prospective Investigation into Cancer and Nutrition †. Followed 1991-2001 to 2013	Europe	Male and Female aged 17 – 98 Excluded if prior cancer diagnosis, zero length follow-up, missing the date of cancer diagnosis, missing the data of vital status check-up. Additionally new diagnosis of appendiceal cancers were excluded with the participants being censored at diagnosis.	51.3 (S.D. 10.0)	25.8 (S.D. 4.3)‡	491,992 (6,467)	Colorectal

† The description of the validation cohorts is as they were available for use with all models after eligibility criteria had been applied. Participants available for validation of specific models varied depending on the variables required. ‡ Variable available for 400,746 participants.

Supplementary Table 3: Comparison of eligible and ineligible male participants in the UK Biobank

	UKB Male	Colditz <sup>5</sup>		Driver <sup>8</sup>		Freedman <sup>1</sup>		Ma <sup>9</sup>		Shin <sup>7</sup>		Steffen <sup>10</sup> <sup>β</sup>		Taylor <sup>11</sup> <sup>β</sup>		Wells <sup>3</sup>	
		Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible
Number of Participants (% of cohort)	219,385	157,638 (71.9)	61,747 (28.1)	216,440 (98.7)	2,945 (1.3)	118,439 (54.0)	100,946 (46.0)	196,524 (89.6)	22,861 (10.4)	145,723 (66.4)	73,662 (33.6)	185,914 (84.7)	33,471 (15.3)	177,031 (80.7)	42,354 (19.3)	93,608 (42.7)	125,777 (57.3)
Colorectal Cases ■ (C.I.R. /100,000 years)	1,258 (118.9)			1,236 (118.4)	22 (157.3)	823 (144.8)	435 (88.9)	1,102 (116.2)	156 (142.6)			1,085 (121.2)	173 (106.7)	990 (115.8)	268 (132.0)	559 (123.9)	699 (115.3)
Colon (C.I.R. /100,000 years)	768 (72.6)	518 (68.0)	250 (84.4)					670 (70.7)	98 (89.6)	229 Ω (32.5)	131 Ω (37.0)	651 (72.7)	117 (72.2)				
Rectal (C.I.R. /100,000 years)	482 (45.6)							424 (44.7)	58 (53.0)			428 (47.8)	54 (33.3)				
Follow-up (years)	1,057,724	761,536	296,188	1,043,740	13,983	568,480	489,244	948,335	109,388	703,758	353,966	895,575	162,148	854,696	203,028	451,300	606,424
<u>Age<sup>Δ</sup></u>																	
% with no data	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Of those with data (S.D.)	56.5 (8.2)	56.2 (8.2)	57.5 (8.2)	56.5 (8.2)	56.8 (8.5)	60.0 (5.4)	52.5 (9.0)	56.4 (8.2)	57.4 (8.3)	56.5 (8.1)	56.7 (8.4)	56.7 (8.2)	55.8 (8.4)	56.3 (8.2)	57.6 (8.2)	56.9 (8.1)	56.3 (8.3)
<u>BMI kg/m<sup>2</sup>Δ</u>																	
% with no data	0.7	0.0	2.6	0.0	54.0	0.0	1.6	0.0	7.0	0.0	2.2	0.0	4.8	0.6	1.4	0.0	1.3
Of those with data (S.D.)	27.8 (4.2)	27.7 (4.1)	28.3 (4.5)	27.8 (4.2)	28.4 (4.7)	27.8 (4.1)	27.9 (4.4)	27.8 (4.2)	28.5 (4.7)	27.6 (4.0)	28.3 (4.6)	27.7 (4.1)	28.4 (4.8)	27.7 (4.2)	28.3 (4.5)	28.1 (4.2)	27.6 (4.3)
<u>Smoking status</u>																	
% with no data	0.6	0.2	1.7	0.0	46.3	0.0	1.4	0.0	6.0	0.2	1.4	0.0	4.1	0.3	2.0	0.0	1.1
Of those with data																	
% Never	49.2	51.1	44.3	49.2	46.1	47.9	50.7	49.8	43.4	49.4	48.7	48.9	51.1	50.5	43.5	46.4	51.3
% Previous	38.2	37.7	39.4	38.2	31.7	42.1	33.5	38.2	37.5	39.2	36.0	39.1	32.7	37.7	40.1	40.8	36.2
% Current	12.7	11.3	16.4	12.6	22.1	10.1	15.8	12.0	19.1	11.3	15.4	12.1	16.2	11.8	16.4	12.8	12.6
<u>Alcohol status</u>																	
% with no data	0.3	0.0	1.2	0.0	25.1	0.0	0.7	0.0	3.2	0.0	1.0	0.0	2.2	0.1	1.5	0.0	0.6
Of those with data																	
% Never	2.8	2.2	4.5	2.8	8.9	1.6	4.3	2.5	6.0	2.9	2.7	3.1	1.5	2.5	4.2	1.6	3.8
% Previous	3.5	3.0	4.9	3.5	7.0	2.9	4.2	3.3	5.4	3.7	3.1	3.9	1.4	3.2	4.9	3.1	3.8
% Current	93.7	94.8	90.6	93.8	84.1	95.5	91.5	94.2	88.6	93.4	94.2	93.1	97.1	94.3	90.9	95.3	92.5
<u>Days/week of vigorous physical activity<sup>Δ</sup></u>																	
% with no data	4.8	0.0	17.0	4.4	31.6	0.0	10.4	0.0	45.9	2.7	8.9	3.7	10.6	3.4	10.7	0.0	8.3
Of those with data (S.D.)	2.1 (2.1)	2.1 (2.0)	2.0 (2.1)	2.1 (2.0)	2.0 (2.3)	2.0 (2.0)	2.2 (2.1)	2.1 (2.0)	1.9 (2.3)	2.1 (2.0)	2.0 (2.1)	2.1 (2.0)	2.0 (2.1)	2.1 (2.0)	2.1 (2.2)	2.1 (2.1)	2.0 (2.0)
<u>Processed meat consumption</u>																	
% with no data	0.5	0.0	1.8	0.2	19.8	0.0	1.0	0.1	3.9	0.0	1.5	0.2	2.4	0.1	2.0	0.0	0.9
Of those with data																	
% Never	5.4	5.4	5.6	5.4	9.4	4.6	6.5	5.4	6.1	5.2	6.0	5.3	6.2	5.4	5.5	3.8	6.7
% < 1 / week	21.4	21.6	20.7	21.4	20.0	22.0	20.6	21.5	20.1	21.0	22.1	21.1	22.7	21.5	20.9	20.0	22.4
% 1 / week	29.8	30.1	29.0	29.8	28.2	30.8	28.6	29.9	29.3	30.3	28.8	30.0	28.9	30.0	28.8	30.5	29.3
% 2-4 / week	36.9	36.7	37.4	36.9	34.7	36.6	37.2	36.9	37.1	37.3	36.0	37.2	35.2	36.8	37.4	39.1	35.2
% 5-6 / week	5.2	5.0	5.6	5.2	5.7	4.8	5.6	5.1	5.6	5.0	5.4	5.1	5.4	5.0	5.7	5.3	5.1
% ≥ 1 / day	1.4	1.2	1.7	1.4	2.0	1.1	1.6	1.3	2.0	1.2	1.7	1.3	1.7	1.3	1.7	1.3	1.4

The UK Biobank (UKB) was assessed over a 5 year time horizon from recruitment. <sup>β</sup> These models were not sex-specific and so both male and female participants were eligible. ■ In the UKB some colorectal cancer diagnosis included more than one anatomical site. If the cancer encompassed both colon and rectum at the date of first diagnosis it was classified as colorectal and was not included in the subsite analysis (e.g. colon or rectal). <sup>Δ</sup> Mean values with standard deviation (S.D.). Ω Right colon cancer only. Cancer site was defined by the International Statistical Classification of Disease and Related Health Problems, 10<sup>th</sup> Revision,<sup>17</sup> colorectal encompassed C18-C20, colon C18 and rectal C19 and C20.

Supplementary Table 4: Comparison of eligible and ineligible male participants in The European Prospective Investigation into Cancer and Nutrition

	EPIC Male	Colditz <sup>5</sup>		Driver <sup>8</sup>		Freedman <sup>1</sup>		Ma <sup>9</sup>		Shin <sup>7</sup>		Steffen <sup>10,β</sup>		Taylor <sup>11,β</sup>		Wells <sup>3</sup>	
		Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible
Number of Participants (% of cohort)	148,007	93,863 (63.4)	54,144 (36.6)	122,650 (82.9)	25,357 (17.1)	56,290 (38.0)	91,717 (62.0)	110,784 (74.9)	37,223 (25.1)	124,293 (84.0)	23,714 (16.0)	111,667 (75.4)	36,340 (24.6)	25,273 (17.1)	122,734 (82.9)	41,587 (28.1)	106,420 (71.9)
Colorectal Cases (C.I.R. /100,000 years)	1,472 (107.0)			1,278 (111.9)	194 (83.5)	771 (149.2)	701 (81.8)	1,191 (115.7)	281 (81.6)			463 (85.2)	179 (101.8)	261 (108.3)	1,211 (106.9)	395 (101.5)	1,077 (109.3)
Colon (C.I.R. /100,000 years)	855 (62.2)	607 (69.1)	248 (50.0)					676 (65.7)	179 (52.0)	115 Ω (19.0)	30 Ω (26.2)	258 (47.5)	117 (66.6)				
Rectal (C.I.R. /100,000 years)	617 (44.9)							515 (50.0)	102 (29.6)			205 (37.7)	62 (35.3)				
Follow-up time (years)	1,374,039	878,437	495,602	1,141,586	232,453	516,656	857,383	1,029,610	344,430	605,178	114,315	543,701	175,792	241,050	1,132,990	389,014	985,025
<u>Age<sup>α</sup></u>																	
% with no data	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Of those with data (S.D.)	52.3 (10.2)	54.2 (7.0)	49.0 (13.4)	53.7 (7.3)	45.3 (17.0)	57.5 (4.9)	49.1 (11.2)	54.0 (7.2)	47.2 (14.9)	53.7 (7.3)	44.6 (17.3)	53.4 (7.2)	48.7 (15.6)	53.3 (7.8)	52.0 (10.6)	51.8 (7.5)	52.5 (11.1)
<u>BMI kg/m<sup>2α</sup></u>																	
% with no data	2.3	0.0	6.2	0.0	13.3	0.0	3.7	0.0	9.0	0.0	14.2	0.0	9.3	6.8	1.3	0.0	3.2
Of those with data (S.D.)	26.5 (3.7)	26.7 (3.6)	26.2 (3.8)	26.7 (3.6)	25.4 (3.7)	27.0 (3.6)	26.2 (3.7)	26.8 (3.6)	25.5 (3.7)	26.7 (3.6)	25.3 (3.7)	26.8 (3.7)	25.6 (3.6)	27.6 (3.5)	26.3 (3.7)	27.0 (3.6)	26.3 (3.7)
<u>Smoking status</u>																	
% with no data	2.0	0.4	4.6	0.0	11.5	0.0	3.2	0.0	7.8	1.3	5.3	0.0	8.0	3.3	1.7	0.0	2.7
Of those with data:																	
% Never	33.3	29.7	39.8	31.2	44.6	33.8	33.0	29.7	44.8	31.2	44.6	31.1	40.8	31.2	33.7	34.2	32.9
% Previous	36.9	39.1	32.8	38.2	29.8	39.6	35.2	39.1	29.8	38.2	29.8	37.6	34.5	36.9	36.2	37.8	36.5
% Current	29.8	31.2	27.4	30.6	25.7	26.6	31.9	31.2	25.4	30.6	25.7	31.4	24.7	31.9	29.4	28.0	30.6
<u>Alcohol status</u>																	
% with no data	24.9	15.6	41.0	20.3	47.6	4.5	37.5	11.8	64.1	20.7	47.3	21.0	37.1	10.8	27.8	0.5	34.5
Of those with data:																	
% Never	1.6	1.3	2.4	1.5	2.5	1.5	1.8	1.5	2.5	1.5	2.5	1.5	2.2	3.2	1.3	1.8	1.5
% Previous	4.4	4.3	4.8	4.4	4.7	4.5	4.3	4.4	4.8	4.4	4.7	3.9	6.4	10.7	2.8	5.0	4.1
% Current	94.0	94.4	92.8	94.1	92.8	94.0	93.9	94.1	92.8	94.1	92.8	94.6	91.4	86.1	95.9	93.2	94.4
<u>Vigorous recreation activity (MET hours / week)<sup>α</sup></u>																	
% with no data	24.5	19.2	33.6	23.2	30.5	0.0	39.5	15.4	51.4	23.6	29.1	23.3	28.0	7.5	27.9	0.8	33.7
Of those with data (S.D.)	18.1 (28.7)	17.6 (27.6)	19.0 (30.8)	17.5 (27.9)	21.2 (32.4)	18.4 (28.5)	17.8 (28.8)	17.5 (27.8)	21.0 (32.5)	17.5 (27.9)	21.4 (32.5)	17.4 (27.9)	20.3 (31.0)	8.7 (22.0)	20.6 (29.7)	13.9 (26.1)	20.5 (29.8)
<u>Processed meat consumption (g/day)<sup>α</sup></u>																	
% with no data	1.9	0.0	5.3	0.0	11.4	0.0	3.1	0.0	7.7	0.0	12.1	0.0	7.9	6.8	1.0	0.0	2.7
Of those with data (S.D.)	42.3 (41.2)	45.4 (40.8)	36.5 (41.3)	44.4 (41.7)	30.5 (36.0)	40.9 (40.5)	43.1 (41.6)	44.5 (42.4)	35.1 (36.2)	44.2 (41.6)	30.8 (36.5)	45.4 (42.5)	31.8 (34.4)	41.7 (37.7)	42.4 (41.8)	48.7 (47.8)	39.7 (37.9)

The European Prospective Investigation into Cancer and Nutrition (EPIC) was assessed over a 10 year time horizon from recruitment, with the exception of Shin et al<sup>7</sup> and Steffen et al<sup>10</sup> which were assessed over 5 years. <sup>β</sup> These models were not sex-specific and so both male and female participants were eligible. <sup>α</sup> Mean values with standard deviation (S.D.). Ω Right colon cancer only. Cancer site was defined by the International Classification of Diseases for Oncology, 2<sup>nd</sup> revision (ICD-O-2),<sup>18</sup> colorectal encompassed C18-C20, colon C18 and rectal C19 and C20.

Supplementary Table 5: Comparison of eligible and ineligible female participants in the UK Biobank

	UKB Female	Colditz <sup>5</sup>		Freedman <sup>1</sup>		Shin <sup>7</sup>		Steffen <sup>10 β</sup>		Taylor <sup>11 β</sup>		Wells <sup>3</sup>	
		Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible
Number of Participants (% of cohort)	256,244	172,807 (67.4)	83,437 (32.6)	123,991 (48.4)	132,253 (51.6)	178,665 (69.7)	77,579 (30.3)	201,704 (78.7)	54,540 (21.3)	219,484 (85.7)	36,760 (14.3)	117,367 (45.8)	138,877 (54.2)
Colorectal Cases ■ (C.I.R. /100,000 years)	892 (71.7)			521 (86.7)	371 (57.7)			696 (71.2)	196 (73.5)	768 (72.1)	124 (69.6)	426 (74.7)	466 (69.1)
Colon (C.I.R. /100,000 years)	639 (51.4)	402 (47.9)	237 (58.6)					490 (50.1)	149 (55.9)				
Rectal (C.I.R. /100,000 years)	240 (19.3)					176 (20.3)	64 (17.0)	194 (19.8)	46 (17.3)				
Follow-up time (years)†	1,244,045	839,681	404,364	600,585	643,460	867,746	376,299	977,418	266,627	1,065,777	178,268	570,107	673,938
<u>Age</u> <sup>^</sup>													
% with no data	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Of those with data (S.D.)	56.1 (8.0)	55.4 (8.0)	57.7 (7.9)	59.5 (5.3)	53.0 (8.8)	56.1 (7.9)	56.1 (8.2)	56.1 (8.0)	56.2 (8.2)	56.0 (8.0)	57.2 (8.1)	56.7 (7.9)	55.7 (8.1)
<u>BMI</u> kg/m <sup>2</sup> <sup>^</sup>													
% with no data	0.5	0.0	1.6	0.0	1.0	0.1	1.5	0.0	2.5	0.4	1.2	0.0	1.0
Of those with data (S.D.)	27.1 (5.2)	26.7 (5.1)	27.8 (5.4)	26.9 (4.9)	27.3 (5.4)	26.6 (4.9)	28.2 (5.8)	26.8 (5.0)	28.2 (5.7)	26.9 (5.1)	28.0 (5.5)	27.3 (5.2)	26.9 (5.2)
<u>Smoking status</u>													
% with no data	0.6	0.2	1.3	0.2	0.9	0.3	1.1	0.0	2.6	0.3	2.1	0.0	1.0
Of those with data													
% Never	59.9	61.0	57.7	58.7	61.1	58.4	63.3	59.2	62.6	60.5	56.0	58.7	60.9
% Previous	31.1	30.9	31.6	34.0	28.4	33.2	26.4	32.3	26.7	30.9	32.3	31.7	30.6
% Current	9.0	8.2	10.8	7.3	10.6	8.5	10.2	8.5	10.7	8.6	11.7	10.0	8.5
<u>Alcohol status</u>													
% with no data	0.3	0.0	0.8	0.0	0.5	0.0	0.9	0.0	1.3	0.1	1.6	0.0	0.5
Of those with data													
% Never	5.9	4.8	8.2	4.4	7.3	8.0	0.9	7.1	1.4	5.5	8.6	4.9	6.7
% Previous	3.6	3.2	4.6	3.3	3.9	5.1	0.3	4.5	0.5	3.4	4.9	3.7	3.6
% Current	90.5	92.0	87.3	92.3	88.8	86.9	98.8	88.4	98.1	91.1	86.5	91.4	89.7
<u>Days/week of vigorous physical activity</u> <sup>^</sup>													
% with no data	6.0	0.0	18.5	0.0	11.7	4.7	9.0	5.0	9.9	4.9	12.6	4.9	7.0
Of those with data (S.D.)	1.7 (1.9)	1.7 (1.8)	1.6 (1.9)	1.6 (1.8)	1.7 (1.9)	1.7 (1.8)	1.6 (1.9)	1.7 (1.9)	1.6 (1.9)	1.7 (1.8)	1.7 (2.0)	1.6 (1.9)	1.7 (1.9)
<u>Processed meat consumption</u>													
% with no data <sup>^</sup>	0.4	0.0	1.3	0.0	0.8	0.0	1.4	0.2	1.4	0.2	2.0	0.1	0.7
Of those with data													
% Never	12.6	12.9	12.0	12.0	13.3	12.6	12.8	12.6	12.7	12.6	12.7	10.3	14.7
% < 1 / week	38.0	38.6	36.9	39.8	36.4	37.8	38.5	37.9	38.4	38.3	36.5	38.1	38.0
% 1 / week	28.7	28.6	29.0	28.8	28.6	29.1	27.7	28.9	28.0	28.7	28.8	30.1	27.5
% 2-4 / week	18.8	18.3	20.0	17.9	19.7	18.7	19.0	18.8	19.0	18.7	19.7	19.8	18.0
% 5-6 / week	1.5	1.4	1.7	1.3	1.6	1.4	1.6	1.4	1.5	1.4	1.8	1.4	1.5
% ≥ 1 / day	0.4	0.3	0.5	0.3	0.5	0.3	0.5	0.4	0.5	0.4	0.5	0.3	0.4

The UK Biobank (UKB) was assessed over a 5 year time horizon from recruitment. <sup>β</sup> These models were not sex-specific and so both male and female participants were eligible. ■ In the UKB some colorectal cancer diagnosis included more than one anatomical site. If the cancer encompassed both colon and rectum at the date of first diagnosis it was classified as colorectal and was not included in the subsite analysis (e.g. colon or rectal). <sup>^</sup> Mean values with standard deviation (S.D.). Cancer site was defined by the International Statistical Classification of Disease and Related Health Problems, 10<sup>th</sup> Revision,<sup>17</sup> colorectal encompassed C18-C20, colon C18 and rectal C19 and C20.

Supplementary Table 6: Comparison of eligible and ineligible female participants in The European Prospective Investigation into Cancer and Nutrition

	EPIC Female	Colditz <sup>5</sup>		Freedman <sup>1</sup>		Shin <sup>7</sup>		Steffen <sup>10</sup> <sup>β</sup>		Taylor <sup>11</sup> <sup>β</sup>		Wells <sup>3</sup>	
		Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible	Eligible	Ineligible
Number of Participants (% of cohort)	343,985	149,164 (43.4)	194,821 (56.6)	80,560 (23.4)	263,425 (76.6)	207,887 (60.4)	136,098 (39.6)	189,023 (55.0)	154,962 (45.0)	84,752 (24.6)	259,233 (75.4)	69,154 (20.1)	274,831 (79.9)
Colorectal Cases (C.I.R. /100,000 years)	2,010 (61.8)			714 (94.5)	1,296 (51.9)			533 (57.6)	316 (41.7)	411 (51.2)	1,599 (65.4)	422 (64.4)	1,588 (61.2)
Colon (C.I.R. /100,000 years)	1,312 (40.4)	744 (52.6)	568 (30.9)					341 (36.9)	205 (27.1)				
Rectal (C.I.R. /100,000 years)	698 (21.5)					217 (21.3)	86 (12.9)	192 (20.8)	111 (14.7)				
Follow-up time (years)	3,250,144	1,414,842	1,835,302	755,233	2,494,912	1,017,063	665,275	924,823	757,515	803,450	2,446,694	655,623	2,594,521
<u>Age<sup>α</sup></u>													
% with no data	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Of those with data (S.D.)	50.9 (9.9)	53.7 (7.0)	48.7 (11.1)	57.9 (4.9)	48.8 (10.0)	53.6 (7.2)	46.7 (11.7)	53.5 (7.2)	47.7 (11.6)	52.6 (7.0)	50.3 (10.6)	51.9 (7.3)	50.6 (10.4)
<u>BMI kg/m<sup>2</sup> <sup>α</sup></u>													
% with no data	25.6	0.0	45.1	0.0	33.4	0.2	64.3	0.0	56.7	46.0	18.9	0.0	32.0
Of those with data (S.D.)	25.5 (4.6)	25.6 (4.5)	25.2 (4.8)	26.4 (4.7)	25.0 (4.5)	25.8 (4.6)	24.0 (4.3)	25.8 (4.6)	24.4 (4.4)	26.2 (4.9)	25.3 (4.6)	26.3 (4.7)	25.2 (4.6)
<u>Smoking status</u>													
% with no data	2.6	0.7	4.0	0.8	3.1	1.3	4.5	0.0	5.7	4.6	1.9	0.0	3.2
Of those with data:													
% Never	57.0	57.1	56.9	64.1	54.7	57.4	56.3	57.5	56.3	70.5	52.7	66.1	54.6
% Previous	23.1	23.9	22.4	20.6	23.8	23.0	23.2	22.5	23.8	18.7	24.5	18.8	24.2
% Current	20.0	19.0	20.7	15.3	21.4	19.6	20.5	20.0	20.0	10.8	22.9	15.1	21.2
<u>Alcohol status</u>													
% with no data	24.6	6.9	38.2	2.7	31.3	17.0	36.1	17.2	33.7	4.6	31.1	0.3	30.7
Of those with data:													
% Never	10.5	9.9	11.2	11.5	10.1	10.5	10.4	11.0	9.7	16.3	7.9	13.6	9.4
% Previous	4.9	5.7	3.9	6.3	4.3	5.6	3.5	5.1	4.6	8.0	3.5	6.9	4.2
% Current	84.6	84.4	84.9	82.2	85.7	83.9	86.2	83.9	85.8	75.7	88.7	79.5	86.5
<u>Vigorous recreation activity (MET hours / week) <sup>α</sup></u>													
% with no data	30.1	19.4	38.4	0.0	39.4	26.9	35.1	27.1	33.8	8.3	37.3	1.1	37.4
Of those with data (S.D.)	12.6 (23.8)	11.2 (22.2)	14.1 (25.2)	11.7 (23.1)	13.1 (24.1)	11.7 (22.9)	14.2 (25.1)	11.6 (22.9)	14.0 (24.8)	10.1 (22.8)	13.8 (24.1)	8.8 (19.4)	14.1 (25.1)
<u>Processed meat consumption (g/day) <sup>α</sup></u>													
% with no data	1.0	0.0	1.7	0.0	1.3	0.0	2.4	0.0	2.1	2.4	0.5	0.0	1.2
Of those with data (S.D.)	27.4 (25.8)	26.0 (24.6)	28.5 (26.6)	24.0 (24.8)	28.5 (26.0)	25.9 (25.3)	29.8 (26.3)	26.1 (25.6)	29.1 (25.8)	29.5 (24.9)	26.7 (26.0)	27.7 (28.6)	27.3 (25.0)

The European Prospective Investigation into Cancer and Nutrition (EPIC) was assessed over a 10 year time horizon from recruitment, with the exception of Shin et al<sup>7</sup> and Steffen et al<sup>10</sup> which were assessed over 5 years. <sup>β</sup> These models were not sex-specific and so both male and female participants were eligible. <sup>α</sup> Mean values with standard deviation (S.D.). Cancer site was defined by the International Classification of Diseases for Oncology, 2<sup>nd</sup> revision (ICD-O-2),<sup>18</sup> colorectal encompassed C18-C20, colon C18 and rectal C19 and C20.

Supplementary Table 7: Published and derived coefficients for the validated risk prediction models

Author	Colditz <sup>5</sup>	Driver <sup>8</sup>	Freedman <sup>1</sup>	Ma <sup>9</sup>	Shin <sup>7</sup>	Steffen <sup>10</sup>	Taylor <sup>11</sup>
Sex						-0.254 CRC (women vs men)	
						-0.053 CC (women vs men)	
						-0.702 RC (women vs men)	
Age		0.81 Male CRC (50-59)	-0.60 Female DCC (>65)	0.08 Male CRC (year)	0.10 Male right CC (age-mean age, years)	0.065 CRC (year)	
		1.48 Male CRC (60-69)		0.085 Male CC (year)	0 Male right CC (age-mean age <sup>2</sup> , years <sup>2</sup> )	0.073 CC (year)	
		1.83 Male CRC (≥70)		0.067 Male RC (year)	0.08 Female RC (age-mean age, years)	0.045 RC (year)	
					0 Female RC (age – mean age <sup>2</sup> , years <sup>2</sup> )		
Height	0.26 Male CC (≥ 5 feet 10 inches)				0.06 Female RC (>151 - ≤155 cm)		
	Female CC (≥ 5 feet 7 inches)				0.14 Female RC (>155 - ≤158 cm)		
					0.21 Female RC (>158 cm)		

Body mass index	0.41 CC ( $\geq 27$ kg/m <sup>2</sup> )	0.23 Male CRC (25.0-29.9 kg/m <sup>2</sup> ) 0.48 Male CRC ( $\geq 30$ kg/m <sup>2</sup> )	0.23 Male PCC (25.0- $\leq 30$ kg/m <sup>2</sup> ) 0.46 Male PCC ( $>30$ kg/m <sup>2</sup> ) 0.32 Male DCC (25.0- $\leq 30$ kg/m <sup>2</sup> ) 0.64 Male DCC ( $>30$ kg/m <sup>2</sup> ) 0.08 Female DCC ( $>30$ kg/m <sup>2</sup> ) 0.34 Female RC ( $>30$ kg/m <sup>2</sup> )	0.047 Male CRC (kg/m <sup>2</sup> ) 0.049 Male CC (kg/m <sup>2</sup> )	0.10 Male right CC ( $\geq 25.0$ kg/m <sup>2</sup> )	0.020 CRC (kg/m <sup>2</sup> ) 0.027 CC (kg/m <sup>2</sup> ) 0.006 RC (kg/m <sup>2</sup> )
Family history of cancer	0.59 CC (sibling or parent with colon cancer)	0.59 Male PCC (1 first degree relative with CRC) 1.19 Male PCC ( $\geq 2$ first degree relatives with CRC) 0.52 Male DCC (1 first degree relative with CRC) 1.03 Male DCC ( $\geq 2$ first degree relatives with CRC) 0.40 Male RC (1 first degree relative with CRC)	0.25 Male right CC (family history of cancer)	-0.12 ■ (no affected first-degree relative) 0.65 ■ (1 affected first-degree relative) 1.10 ■ (2 affected first-degree relatives) 1.49 ■ (3 affected first-degree relative)		

Family history of cancer  
(continued)

0.41  
Female PCC (1 first degree  
relative with CRC)  
0.82  
Female PCC ( $\geq 2$  first  
degree relatives with CRC)  
0.37  
  
Female DCC (1 first degree  
relative with CRC)  
0.74  
Female DCC ( $\geq 2$  first  
degree relatives with CRC)  
  
0.43  
Female RC (1 first degree  
relative with CRC)

Diabetes

0.221  
CRC (prevalent diabetes)  
0.086  
CC (prevalent diabetes)  
0.480  
RC (prevalent diabetes)

Screening / Endoscopy

-0.69  
CC (flexible sigmoidoscopy  
or FOBT within 10 years)

0.35  
Male PCC (No Endoscopy  
in last 10 years)  
0.57  
Male PCC (Endoscopy in  
last 10 years and history of  
polyps)  
0.46  
Male PCC (Endoscopy and  
polyps unknown)

-0.531  
CRC (ever had CRC  
screening)  
-0.345  
CC (ever had CRC  
screening)  
-0.953  
RC (ever had CRC  
screening)



Screening / Endoscopy  
(continued)

1.04

Male DCC (No Endoscopy  
in last 10 years)

0.29

Male DCC (Endoscopy in  
last 10 years and history of  
polyps)

0.96

Male DCC (Endoscopy and  
polyps unknown)

1.35

Male RC (No Endoscopy in  
last 10 years)

0.65

Male RC (Endoscopy in last  
10 years and history of  
polyps)

-0.67

Male RC (Endoscopy and  
polyps unknown)

0.60

Female PCC (No  
Endoscopy in last 10 years)

0.96

Female PCC (Endoscopy in  
last 10 years and history of  
polyps)

-0.49

Female PCC (Endoscopy  
and polyps unknown)

Screening / Endoscopy  
(continued)

1.24  
Female DCC (No  
Endoscopy in last 10 years)  
1.47  
Female DCC (Endoscopy in  
last 10 years and history of  
polyps)  
1.15  
Female DCC (Endoscopy  
and polyps unknown)  
  
1.10  
Female RC (No Endoscopy  
in last 10 years)  
1.16  
Female RC (Endoscopy in  
last 10 years and history of  
polyps)  
-0.99  
Female RC (Endoscopy and  
polyps unknown)

---

Inflammatory bowel  
disease                      0.41  
                                    CC ( $\geq 10$  years)

---

Oestrogen status                      -0.39  
    Female PCC (Negative in  
    last 2 years)  
    -0.73  
    Female DCC (Negative in  
    last 2 years)

---

Oestrogen status (continued)		-0.40	
		Female RC (Negative in last 2 years)	
BMI- Oestrogen interaction		0.99	
		Female DCC	
Hormone replacement therapy	-0.22		
	Female CC ( $\geq 5$ years)		
Oral Contraceptive Pill	-0.36		
	Female CC ( $\geq 5$ years)		
Non-steroidal anti- inflammatory drugs	-0.36	-0.43	
	CC (Aspirin daily use $\geq 15$ years)	Male PCC (regular use)	
		-0.34	
		Male DCC (regular use)	
		-0.42	
		Male RC (regular use)	
		-0.46	
		Female PCC (regular use)	
		-0.36	
		Female DCC (regular use)	
		-0.36	
		Female RC (regular use)	
Physical Activity	-0.51 *	-0.19	-0.019
	CC ( $\geq 25$ MET-hour/week)	Male RC ( $>0 - \leq 2$ h/week)	Male CRC (MET-h/day)
		-0.37	
		Male RC ( $>2 - \leq 4$ h/week)	-0.020
		-0.56	Male CC (MET-h/day)
		Male RC ( $>4$ h/week)	

Physical Activity  
(continued)

-0.15	-0.020
Female PCC (>0 - ≤ 2 h/week)	Male RC (MET-h/day)
-0.29	
Female PCC (>2 - ≤ 4 h/week)	
-0.43	
Female RC (>4 h/week)	
-0.37	
Female RC (>0 - ≤ 2 h/week)	
-0.24	
Female RC (>2 - ≤ 4 h/week)	
-0.46	
Female RC (>4 h/week)	

Smoking

0.35	0.26	0.071	0.240
Male CRC (current/previous smoker)	Male PCC (>0 - <11 cigarettes/day)	Male CRC (former)	CRC (former)
	0.53	0.239	0.269
	Male PCC (≥11 - ≤20 cigarettes/day)	Male CRC (current)	CRC (current)
	0.80	0.186	0.109
	Male PCC (>20 cigarettes/day)	Male CC (former)	CC (former)
		0.347	0.165
		Male CC (current)	CC (current)
			0.536
	-0.51		RC (former)
	Male PCC (>0 - <15 years smoking)		0.476
			RC (current)
	-0.13		
	Male PCC (≥15 - <35 years smoking)		

Smoking  
(continued)

-0.40  
Male PCC (≥35 years  
smoking)

Alcohol

0.34 \*  
CC (> 7 servings/week)

0.31  
Male CRC (≥once/week)

-0.163  
Male CRC (never)  
0.358  
Male CRC (regular, <300  
g/w)  
0.659  
Male CRC (regular ≥  
300g/week)

0.13  
Male right CC (1-14.9  
g/day)  
0.18  
Male right CC (15-24.9  
g/day)  
0.22  
Male right CC (≥25 g/day)

0.080  
CRC (drinks/day)  
0.095  
CC (drinks/day)  
0.051  
RC (drinks/day)

-0.140  
Male CC (never)  
0.419  
Male CC (regular, <300  
g/week)  
0.655  
Male CC (regular ≥  
300g/week)

0  
Female RC (1-14.9 g/day)  
0.39  
Female RC (≥ 15 g/day)

-0.094  
Male RC (never)  
0.365  
Male RC (regular, <300  
g/week)  
0.745  
Male RC (regular ≥  
300g/week)

Red Meat/Meat	0.41*			0.09
	CC ( $\geq 1.2$ servings/day)			Male right CC (2-3 times / week)
				0.21
				Male right CC ( $\geq 4$ times / week)
				0.03
				Female RC (2-3 times / week)
				0.33
				Female RC ( $\geq 4$ times / week)
Vegetables	-0.36	-0.54		
	CC ( $\geq 3$ servings/day)	Male PCC ( $\geq 5$ servings /day)		
		-0.33		
		Female PCC ( $\geq 5$ servings /day)		
Vitamins	-0.69			
	CC (Regular multivitamin use)			

CRC = colorectal cancer, CC = colon cancer, PCC = proximal colon cancer, DCC = distal colon cancer, RC = rectal cancer, g = grams, h = hours, kg = kilogram, m<sup>2</sup> = meters squared, MET = Metabolic equivalent task. Where odds ratios, hazard ratios or relative risks were published these were transformed into coefficients using the natural log. \* Coefficient values were calculated from relative risks published in a subsequent validation paper or the variable definition from the validation was utilised <sup>13</sup>. ■ Coefficient values were calculated from relative risks in a preceding paper published by several of the authors <sup>12</sup>. The reference categories for categorical variables are not included. Please see the original publications for further description of the variables definitions and model construction. Permission was not obtained to present the previously unpublished coefficients of the Wells et al <sup>3</sup> model and so the values are omitted from this table.



## References:

1. Freedman AN, Slattery ML, Ballard-Barbash R, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* 2009;27(5):686-93. doi: 10.1200/jco.2008.17.4797 [published Online First: 2008/12/31]
2. National Cancer Institute Division of Cancer Epidemiology and Genetics. Colon Cancer Risk Assessment - SAS Macro [Available from: <https://dceg.cancer.gov/tools/risk-assessment/ccratsasmacro> accessed 23rd June 2017.
3. Wells BJ, Kattan MW, Cooper GS, et al. Colorectal cancer predicted risk online (CRC-PRO) calculator using data from the multi-ethnic cohort study. *J Am Board Fam Med* 2014;27(1):42-55. doi: 10.3122/jabfm.2014.01.130040 [published Online First: 2014/01/07]
4. Wells BJ, Kattan MW, Cooper GS, et al. Colorectal Cancer - 10 year Predicted Risk Online (CRC-PRO) 2014 [Available from: <http://riskcalc.org/ColorectalCancer/> accessed 20th June 2017.
5. Colditz GA, Atwood KA, Emmons K, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control* 2000;11(6):477-88. [published Online First: 2000/07/06]
6. Haftenberger M, Friedenreich C, Lahmann PH, et al. Anthropometric Data in the EPIC Study. In: International Agency for Research on Cancer, ed. Unpublished Internal Document, 2015.
7. Shin A, Joo J, Yang HR, et al. Risk prediction model for colorectal cancer: National Health Insurance Corporation study, Korea. *PLoS One* 2014;9(2):e88079. doi: 10.1371/journal.pone.0088079 [published Online First: 2014/02/18]
8. Driver JA, Gaziano JM, Gelber RP, et al. Development of a risk score for colorectal cancer in men. *Am J Med* 2007;120(3):257-63. doi: 10.1016/j.amjmed.2006.05.055 [published Online First: 2007/03/14]
9. Ma E, Sasazuki S, Iwasaki M, et al. 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men. *Cancer Epidemiol* 2010;34(5):534-41. doi: 10.1016/j.canep.2010.04.021 [published Online First: 2010/06/18]
10. Steffen A, MacInnis RJ, Joshy G, et al. Development and validation of a risk score predicting risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2014;23(11):2543-52. doi: 10.1158/1055-9965.EPI-14-0206
11. Taylor DP, Stoddard GJ, Burt RW, et al. How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. *Genet Med* 2011;13(5):385-91. doi: 10.1097/GIM.0b013e3182064384 [published Online First: 2011/01/29]
12. Taylor DP, Burt RW, Williams MS, et al. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* 2010;138(3):877-85.
13. Kim DJ, Rockhill B, Colditz GA. Validation of the Harvard Cancer Risk Index: a prediction tool for individual cancer risk. *Journal of Clinical Epidemiology* 2004;57(4):332-40.



14. International Physical Activity Questionnaire. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) - Short and Long Forms, 2005.
15. Uk biobank. Data\_Field 20161 [Available from: <http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20161> accessed 10th July 2017.
16. Department of Health. Alcohol Units, a brief guide. In: Department of Health, ed., 2008.
17. UK Biobank. Data-Field 40006, Notes [Available from: <http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=40006> accessed 3rd July 2017.
18. European Prospective Investigation into Cancer and Nutrition. Guidelines\_Cancer\_Followup\_2015\_Nov. Unpublished internal document.