Diet, metabolic mediators, and cancer risk

Cody Zdan Watling

St Edmund Hall

University of Oxford

Trinity 2022

A thesis submitted for the degree of

Doctor of Philosophy in Population Health

Abstract

Diet, metabolic mediators, and cancer risk

Cody Watling, St Edmund Hall, University of Oxford, DPhil, Trinity term 2022.

Colorectal, breast, and prostate cancer account for nearly a third of all diagnosed cancers, and some dietary factors may influence risk. Moreover, circulating insulin-like growth factor-I (IGF-I) concentrations have been positively associated with risks of colorectal, breast, and prostate cancer, and IGF-I might mediate some diet-cancer associations. This thesis investigates the associations of dietary factors, IGF-I concentrations, and IGF-I related cancers among participants in the UK Biobank cohort.

Associations between intakes of food groups (438,453 participants) and macronutrients (11,815 participants) with IGF-I concentrations were explored. Intake of fish was positively associated with IGF-I concentrations, as was protein from dairy products; however, associations differed by dairy protein source in that milk and yogurt were positively associated with IGF-I concentrations whereas protein from cheese was not. Intakes of fibre and wholegrains were also positively associated with IGF-I concentrations.

The association between protein from all dairy products, milk, and cheese and colorectal, breast, and prostate cancer was explored among ~114,000 participants. Protein from all dairy products and milk were inversely associated with colorectal cancer risk, whereas a borderline positive association was observed for prostate cancer risk for men in the highest quartile of intake compared to those in the lowest quartile. IGF-I concentrations did not appear to explain these associations.

Prospective associations between carbohydrate types and sources and colorectal cancer were explored among ~114,000 participants. Non-free sugars and fibre from wholegrains were inversely associated with colorectal cancer risk. Evidence of heterogeneity by genetically predicted butyrate synthesis was observed; participants with high predicted butyrate and consuming higher wholegrains had a lower risk of colorectal cancer whereas no association was observed for those with low genetically predicted butyrate synthesis.

Finally, diet groups in relation to all cancer and common cancer types were investigated in prospective analyses of ~472,000 participants. In comparison to regular meat-eaters, low meat-eaters had lower risk of all cancer and colorectal cancer, fish-eaters had a lower risk of all cancer and prostate cancer, and vegetarians had a lower risk of all cancer, postmenopausal breast cancer, which may be due to their lower body mass index, and prostate cancer. Differences in IGF-I concentrations or free testosterone did not appear to mediate these associations.

In conclusion, research from this thesis supports dietary factors, such as dairy protein, being associated with IGF-I concentrations. However, whether these same dietary factors are positively associated with IGF-I related cancers remains unclear, suggesting other factors or mechanisms external to the IGF-I pathway. Potential novel findings suggest that wholegrain intake and colorectal cancer risk may vary by genetic factors for butyrate production and further research is needed to replicate and explore this further. Different diet groups may also be related to IGF-I related cancers; however, IGF-I concentrations did not mediate this association. Further research exploring mechanisms through and external to the IGF-I pathway is needed.

Acknowledgements

I would like to express my gratitude and appreciation to Associate Professor Aurora Pérez-Cornago and Professor Tim Key for their unwavering support, expertise, and supervision throughout the duration of this degree. I have learned and grown as a researcher under their supervision and for that, I am forever indebted.

I would also like to thank collaborators and co-authors, namely Julie Schmidt, Ruth Travis, Anika Knuppel, Carmen Piernas, Yashvee Dunneram, Ellie Watts, Sandar Tin Tin, and Tammy Tong for their contributions and input on this research.

A special thanks goes to fellow DPhil student Rebecca Kelly, whose support and brilliance motivated me to be assiduous, improved my work and learning, as well as made this process more enjoyable.

I wish to express my appreciation to the UK Biobank participants, staff, and executive team for creating this resource and their contributions to furthering health research. Without this resource, the research included in this thesis would not have been possible.

This work was funded by the Nuffield Department of Population Health doctoral scholarship, and I want to acknowledge and express my appreciation to the Nuffield Department of Population Health graduate studies team for providing me with this opportunity.

Finally, I would like to express my gratitude to my friends and family, particularly Celine, for their support and encouraging words throughout the duration of this degree.

Contents

List of Tables

List of Figures

List of Abbreviations

AHS-2: Adventist Health Study-2

BMI: Body mass index

CEU: Cancer Epidemiology Unit

CI: Confidence intervals

ER: Oestrogen receptor

EPIC: European Prospective Investigation into Cancer and Nutrition

FAP: Familial adenomatous polyposis

FFQ: Food frequency questionnaire

g/day: grams per day

GH: Growth hormone

GWAS: Genome wide association studies

HES: Hospital Episode Statistics

HR: Hazard ratio

IARC: International Agency for Research on Cancer

ICD-10: World Health Organization's International Statistical Classification of Diseases and Related Health Problems 10th revision

IGF: Insulin-like growth factor

IGF-I: Insulin-like growth factor-I

IGFBP: Insulin-like growth factor binding proteins

IORW: Inverse odds ratio weighting

kJ: Kilojoules

LRT: Likelihood ratio tests

MAP: MUTYH-associated polyposis

MAPK: Mitogen activated protein kinase

MET: Metabolic equivalent of task

MHT: Menopausal hormone therapy

mTOR: Mammalian target of rapamycin

MR: Mendelian randomisation

NDE: Natural direct effect

NHS: National Health Service

NIE: Natural indirect effect

NSAID: Non-steroidal anti-inflammatory drugs

OR: Odds ratio

PGS: Polygenic score

PI3K-Akt-mTOR: Phosphoinositide 3-kinase pathways-Akt-mammalian target of rapamycin

PSA: Prostate specific antigen

RCT: Randomised controlled trial

RR: Relative risk

SCFA: Short-chain fatty acid

SD: Standard deviation

SHBG: Sex hormone binding globulin

SMR: Scottish Morbidity Records

UK: United Kingdom

UK BiLEVE: UK Biobank Lung Exome Variant Evaluation

WCRF: World Cancer Research Fund

Chapter 1

Introduction

1.1 Rationale

Globally, cancers of the colorectum, breast, and prostate contribute to approximately 28% of all newly diagnosed cancers and nearly 20% of all cancer deaths [1]. In the United Kingdom (UK), and other countries with a high development index, cancer is the leading cause of death before the age of 70 years [1, 2], with colorectal, breast, and prostate cancer often being listed in the top 10 causes of death from any illness [2]. Each cancer site varies in risk factors; both genetic and modifiable lifestyle and environmental factors are associated with the risk of developing certain cancer types, however for prostate cancer, there have been no well-established modifiable risk factors.

One potential modifiable risk factor for colorectal, breast, and prostate cancer, which may be related to dietary intake [3-5], is circulating insulin-like growth factor-I (IGF-I), a polypeptide hormone which stimulates cell growth, proliferation, and inhibits apoptosis [6-8]. IGF-I concentrations have been positively associated with the risks of colorectal, breast, and prostate cancer in nested case-control studies [9-11], prospective cohort studies [12-14], and Mendelian randomisation (MR) studies [12-14]. Evidence has suggested that dietary intake may be related to circulating IGF-I concentrations [3-5], however, findings have generally been inconsistent and limited by small sample sizes. Moreover, the evidence linking dietary intake with risks of colorectal, breast, and prostate cancer remains uncertain, with only few dietary factors established as potential modifiable risk factors.

1.2 Research aims

The main aim of this thesis is to investigate the associations of dietary intake with colorectal, breast, and prostate cancer risk, assessing the role of potential mediators using data from the UK Biobank.

1.3 Outline of the thesis

Chapter 1: Introduces and outlines the thesis, provides the rationale, as well as the general aims of the thesis.

Chapter 2: Describes the epidemiology of colorectal, breast, and prostate cancer, including global differences in incidence and mortality, and describes known risk factors for these common cancers. This chapter also describes the associations between dietary intake and risk of colorectal, breast, and prostate cancer, and the relationship between IGF-I concentrations and risk of these cancers.

Chapter 3: This chapter describes the data source for this thesis, namely the UK Biobank prospective cohort. This chapter outlines the initial recruitment and follow-up of participants as well as describes the dietary assessments undertaken in this cohort.

Chapter 4: Investigates the cross-sectional associations of specific food group intakes with circulating IGF-I concentrations.

Chapter 5: Investigates the associations between macronutrient intakes and circulating IGF-I concentrations in a subsample of the UK Biobank that completed multiple detailed 24-hour dietary assessments.

Chapter 6: Investigates the prospective associations of protein from all dairy products, milk, and cheese with risks of colorectal, breast, and prostate cancer.

Chapter 7: Investigates the prospective associations between carbohydrate types and sources and colorectal cancer as well assesses if host genetic factors for short-chain fatty acids (SCFA) modify these associations.

Chapter 8: Investigates the prospective associations between diet groups, namely regular meat-eaters, low meat-eaters, fish-eaters (pescatarians) and vegetarians, and risks of developing all cancer, colorectal cancer, postmenopausal breast cancer, and prostate cancer.

Chapter 9: Summarises and interprets the main findings of the thesis, considers the strengths and limitations of the methodology, and discusses recommendations for future research.

1.4 Development of thesis and author's contributions

This work was conducted under the supervision of Associate Professor Aurora Pérez-Cornago (primary supervisor) and Professor Timothy Key. Prof. Key and Assoc. Prof. Pérez-Cornago conceived the ideas for Chapters 4 and 5. For the remaining chapters, I, along with Assoc. Prof. Pérez-Cornago conceived the aims.

I conducted all literature searches and have written all chapters included in this thesis with input from my supervisors. However, co-authors have read and provided suggestions on the manuscripts for publication in which the results chapters are based on.

All data were collected by the UK Biobank, and blood samples were assayed at their central laboratory, which included assays for IGF-I concentrations. Follow-up data were provided to the UK Biobank by central registrars and Hospital Episode Statistics (HES) from England, Scotland, and Wales.

I conducted all statistical analyses for the results chapters of this thesis; I adapted code developed by Ms. Georgina Fensom (statistical programmer, Cancer Epidemiology Unit (CEU)) to code outcome variables from the UK Biobank cancer, HES, and death data. I worked alongside Dr Rebecca Kelly (current DPhil student, CEU) to define exposures utilised for Chapters 5–7. I created all tables and figures included in this thesis.

1.5 Publications relating to this thesis

At the time of initial submission, three manuscripts have been published from this thesis and two other manuscripts have been submitted for publication, of which I am the first author. These are summarised below:

Chapter 4: Associations between food group intakes and circulating insulin-like growth factor-I in the UK Biobank: a cross-sectional analysis – published in the European Journal of Nutrition (Watling *et al.*, 2022. Eur. J. Nutr. July 31. DOI: https://doi.org/10.1007/s00394-022-02954-4)

Chapter 5: Macronutrient intakes and circulating IGF-I concentrations in a subsample of the UK Biobank – published in Clinical Nutrition (Watling et al., 2021. Clin Nutr. 40(7):4685-4693. Apr 19. DOI: https://doi.org/10.1016/j.clnu.2021.04.021)

Chapter 6: Associations of dairy protein intake and risks of colorectal, breast, and prostate cancer: a prospective analysis – under review at the British Journal of Cancer *(*Watling *et al.,* unpublished)

Chapter 7: Carbohydrate intakes, short-chain fatty acid genetic modifiers, and colorectal cancer risk: a prospective analysis in the UK Biobank – under review at Cancer Research (Watling et al., unpublished)

Chapter 8: Risk of cancer in regular and low meat-eaters, fish-eaters, and vegetarians: a prospective analysis in the UK Biobank – published in BMC Medicine (Watling *et al.,* 2022. BMC Med. 20, 73. Feb 24. DOI: https://doi.org/10.1186/s12916-022-02256-w)

Chapter 2

Epidemiology of colorectal, breast, and prostate cancer

Cancer is the leading, or second leading, cause of death before 70 years of age in more than 60% of countries across the world [1, 2]. Cancer is characterised by uncontrolled cell growth and division, with the potential to invade and spread to surrounding healthy tissue and other parts of the body [15]. The development of cancer begins with an error in cell regulation as a result of multiple gene mutations in a single cell over time [16]. This increases the capacity for uncontrolled proliferation and survival, leading to tumour formation with the potential to invade and spread to the surrounding healthy tissue and other parts of the body (i.e., metastasis). Cancer cells can result from various possible mutations; however, all cancer cells share common phenotypic characteristics. These phenotypic characteristics, commonly referred to as the 'Hallmarks of Cancer' include evading growth suppressors, sustained proliferative signalling, resisting cell death, enabling replicative immortality, inducing angiogenesis, deregulating cellular energetics, activating invasion and metastasis, and avoiding immune destruction [16]. Cancers are characterised by the types of cells, tissue, and location in the body from which the cancer originates.

Although cancers have common phenotypic characteristics, cancer types are heterogeneous and can have different aetiological factors. Moreover, some sites of cancer are more common, which may be due to the number of stem cell divisions within the specific tissue [17]. Collectively, cancers originating in the colorectum, breast, or prostate contribute to 28% of newly diagnosed cancers worldwide [1], and are among the most commonly diagnosed cancer types, making these sites of cancer a major public health and economic concern. Moreover, circulating IGF-I concentrations have been shown to be positively associated with all three of these cancer types. This chapter will outline the incidence, mortality, geographical trends, and risk factors for colorectal, breast, and prostate cancer, with a focus on dietary intake and circulating IGF-I concentrations.

2.1 Colorectal cancer epidemiology

The colon is the lower part of the digestive tract extending from the cecum to the rectum and functions to absorb water and salts from undigested foods as well as move waste products to the

rectum to be later excreted. The rectum connects the colon to the anus and stores faeces until defaecation. These two components of the digestive tract form the colorectum. Cancers originating in the colorectum are 95% adenocarcinomas; a type of cancer which originates in glandular epithelial cells, although other types can include mucinous carcinomas and adenosquamous carcinomas [18]. Although the colon and rectum differ in physiology, prognosis, and treatment of cancer [19], they are often grouped together as they are thought to share some common aetiological factors [20, 21] and early data may not have been available to reliably separate the colon from the rectum. However, recent evidence has suggested that some risk factors may differ depending on the anatomical location in the colorectum as well as the molecular features of the tumour [21, 22].

2.1.1 Incidence, mortality, & geographical trends

In 2020, more than 1.9 million individuals were diagnosed with colorectal cancer and over 900,000 individuals died of the disease, making it the second leading cause of cancer death in the world [1, 23]. The incidence of colorectal cancer across the globe varies dramatically, with the highest incidence found in Europe, North America, Australia, and Japan and the lowest incidence rates found in south Asia and Africa [1]. **Figure 2.1** presents the age-standardised incidence rates of colorectal cancer per 100,000 individuals. In European countries, such as Norway and Hungary, the age-standardised incidence rate is around 43 cases per 100,000 individuals, whereas in some African countries, namely Guinea and Ghana, the age-standardised incidence rate is around 3.5 cases per 100,000 individuals, reflecting over a 10-fold difference [24]. This difference in rates between countries has been attributable to environmental factors, lower socioeconomic status, as well as less colorectal cancer screening, which is readily available and implemented in countries located in Europe and North America [1]. Although the incidence of colorectal cancer is relatively low in African and some Asian countries, it is expected to increase with socioeconomic development, as colorectal cancer incidence tends to be higher among countries with greater development [25]. This change in risk with greater development of

a country is thought to reflect changes in lifestyle factors including shifts in diet, decreased physical activity, and increased prevalence of higher body weight [1].

Generally, the age-standardised incidence rates of colorectal cancer have decreased over the past 50 years in countries with high development and long standing screening programmes [26-28]; however, in countries with no large-scale screening programs and of low development, incidence rates have increased [27, 28] and colorectal cancer remains the third most commonly diagnosed cancer worldwide [1].

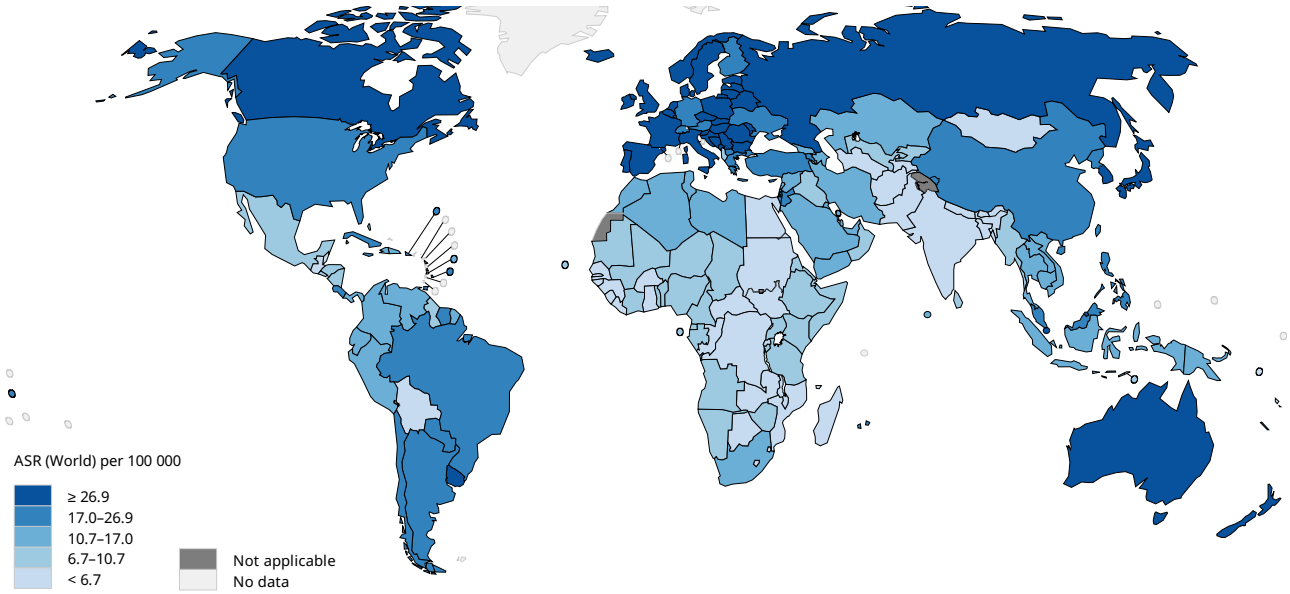


Figure 2.1 Estimated age-standardised incidence rates of colorectal cancer in 2020.
Data source [24]. Abbreviation: ASR, age-standardised rate.

The mortality of colorectal cancer generally reflects the incidence rate observed across regions of the world. For example, the age-standardised mortality rate for colorectal cancer is around 5 deaths per 100,0000 individuals in most sub-Saharan African countries whereas in European countries the mortality rate is around 10-15 deaths per 100,000 [24]. However, due to improvements in treatments, risk factor modification among populations, and most importantly implementation of screening programmes, colorectal mortality has generally decreased over time [26, 27, 29]. However, for individuals under the age of 50 years, the mortality rate has slightly

increased in the United States along with higher incidence rates [29], but the same increase in mortality has not been observed in European countries [30].

2.1.2 Risk factors

Numerous observational studies and few randomised controlled trials (RCTs) have suggested that colorectal cancer is influenced by both modifiable and non-modifiable risk factors. Colorectal cancer is more commonly diagnosed in males than females [26], which has been attributable to differences in risk factors such as higher alcohol intake, higher smoking prevalence, and higher prevalence of abdominal fat among men [23]. This section will outline established risk factors for developing colorectal cancer, emphasising risk factors with the most convincing evidence.

Age

As with almost all cancers, the incidence of colorectal cancer increases with age [23]. A diagnosis of colorectal cancer before the age of 40 is uncommon, however recent evidence has suggested that colorectal cancer incidence is increasing among individuals under the age of 50 years in some populations [23, 29-32]. This has been postulated to be a result of changing modifiable risk factors such as obesity, alcohol intake, and lower physical activity amongst younger age cohorts [23, 33], however further research is needed to better understand risk factors underlying the increased incidence among this age group [29, 32, 34, 35]. As a result of this change in incidence among younger adults, countries such as the United States have lowered age recommendations for colorectal cancer screening to start at 45 years of age instead of 50 years of age, as recommended by the American Cancer Society [36].

Genetics

Early evidence based on twins suggested that 35% of colorectal cancer cases are attributable to genetic components [37], however more recent estimates based on genetic analyses have suggested this is around 13-17% [38, 39]. There are some rare genetic variants that have a high

penetrance for developing colorectal cancer, namely familial adenomatous polyposis (FAP), MUTYH-associated polyposis (MAP), and Lynch syndrome. FAP is estimated to account for less than one percent of colorectal cancers and results in numerous colonic adenomas (a type of colonic polyp that is a major risk factor developing colorectal cancer) formed in childhood. This germline mutation on chromosome 5 on the *APC* gene is relatively uncommon among the general population, but is much more common in those from Ashkenazi Jewish ancestry [40]. Carriers of the FAP gene mutation have nearly a 100% chance of developing colorectal cancer over their lifetime [41]. MAP is an autosomal-recessive syndrome that results in the formation of colonic adenomatous polyps and overall greater risk of developing colorectal cancer. It is estimated that 43-63% of biallelic individuals will be diagnosed with colorectal cancer by the age of 60 years and 80-90% will be diagnosed at the end of their lifetime [42]. Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer, is an autosomal dominant genetic condition that accounts for approximately 3% of colorectal cancer cases. Lynch syndrome is due to germ-line mutations of any four DNA mismatch-repair genes, namely *MSH2*, *MLH1*, *MSH6*, and *PMS2,* or deletion in the *EPCAM*, which then silences the *MSH2* gene. For individuals with Lynch syndrome, colorectal cancer onset is typically diagnosed at an average age of 45 years [40]. The lifetime risk of colorectal cancer for individuals with Lynch syndrome depends on which gene is affected; this can be as high as 50% in individuals with a *MLH1* gene mutation or as low as 10% for individuals with a *PMS2* gene mutation [43, 44]. Lynch syndrome is generally rare in the greater population, affecting approximately 1 in 370 individuals [45, 46]. For individuals that carry these highly penetrant genes (FAP, MAP, and Lynch syndrome), colorectal cancer screening should be conducted more frequently and earlier compared to the general population in order to detect cancers and precancerous lesions to treat or prevent colorectal cancer [41], therefore improving prognosis [47].

Genome wide association studies (GWAS) have identified more than 100 low penetrant single nucleotide polymorphisms (SNPs) [48-50], which collectively can predict greater risk of colorectal cancer through the creation of polygenic scores, however to date, known SNPs explain a small proportion of heritability [51].

Family history

Individuals with a first-degree relative (parent, sibling, or child) diagnosed with colorectal cancer have an estimated two-fold increase in colorectal cancer risk (relative risk (RR): 2.3; 95% confidence interval (CI): 2.0-2.5), whereas those with multiple first-degree relatives who have had colorectal cancer have a four-fold increased risk (RR: 4.3; 95% CI: 3.0-6.1) compared to individuals with no family history [52]. This higher risk among family members may be due to shared genetic components such as low penetrant SNPs, which collectively may increase the risk of colorectal cancer and possibly also partly due to shared environmental factors [53, 54].

Inflammatory bowel disease

Individuals with a diagnosis of ulcerative colitis or Crohn's disease, two conditions characterised by inflammation along the gastrointestinal tract or specifically the colon, have been shown to have a higher risk of colorectal cancer [55]. In a meta-analysis of population-based cohorts, having ulcerative colitis or Crohn's disease was shown to increase the risk of colorectal 2-2.5 fold [56, 57].

Ionising radiation

Individuals who have been exposed to ionising radiation, at a large single dose or low repeated doses overtime, have an increased risk of colorectal cancer. This has been shown in observational studies for individuals exposed to high doses of radiation from atomic bombings [58], among individuals who received radiation therapy [59, 60], as well as in those with low-dose occupational exposure to ionising radiation [61].

Smoking

Like for many cancer sites, smoking is a risk factor for developing colorectal cancer. The International Agency for Research on Cancer (IARC) classifies smoking as a group 1 carcinogen for colorectal cancer [62]. Recent estimates from meta-analyses of observational studies suggest that smokers have a 15%-18% higher risk of colorectal cancer compared to non-smokers, with a

greater risk with longer duration and intensity of smoking [63, 64]. Smoking has also been shown in MR studies, with greater than 52,000 colorectal cancer cases, to be causally associated with higher colorectal cancer risk [65]. Smoking is known to have many carcinogenic compounds, such as polycyclic aromatic hydrocarbons, nitrosamines, aromatic amines, and benzene, thus greater exposure to these compounds may increasing the risk of developing colorectal cancer [66].

Non-steroidal anti-inflammatory drugs (NSAIDs)

The regular use of NSAIDs, such as aspirin and cyclo-oxygenase-2 enzyme inhibitors, has been shown to lower the risk of colorectal cancer in both observational [67] and RCTs [68-70]. RCTs assessing the risk of colorectal cancer have shown that participants who were randomised to take high dose aspirin (>500 mg) or low dose (at least 75 mg) have a lower risk of developing colorectal cancer (24% lower risk), with an even lower risk for those who have taken aspirin for a longer duration (32% lower risk) [68, 69]. For cyclo-oxygenase-2 enzyme inhibitors, similar evidence has been observed in reducing colorectal adenomas [71], however, due to the potential adverse effect of increased risk of cardiovascular events [72], they are not recommended for long-term use for colorectal cancer prevention.

Diabetes mellitus

In a meta-analysis of 30 observational studies, individuals living with diabetes have a 27% higher risk (95% CI: 21-34%) of developing colorectal cancer compared to individuals without diabetes [73]. Evidence from MR studies suggest a higher risk of colorectal cancer among individuals with a higher genetic predisposition to type II diabetes, although this may be a result of pleiotropy [74] and smaller MR studies have not observed this association [75, 76]. However, a MR analysis did show that genetic variants predicting higher fasting insulin levels were associated with a higher risk of colorectal cancer [74]. Elevated insulin levels due to insulin resistance and hyperinsulinemia, which are observed in individuals with untreated or early stage type II diabetes, have been hypothesised to partially explain the higher risk of colorectal cancer in individuals with diabetes [77]. At high levels, insulin acts as an important growth factor of

colonic cells and has been shown in *in vitro* studies to be a mitogen of colon tumour cell growth [77], and has common downstream activation for cellular proliferation and survival as IGF-I [6, 78]. Observational evidence has also suggested that the elevated risk of colorectal cancer among individuals with diabetes diminishes over time as circulating insulin levels lower [79]. Considering this and evidence from MR studies, hyperinsulinemia is likely to be an important factor in colorectal cancer carcinogenesis and may be responsible for the higher risk observed among individuals with type II diabetes [74, 79].

Adiposity

Adiposity is a major risk factor for developing colorectal cancer and an estimated 11% of colorectal cancer cases in Europe have been attributed to being overweight or obese [80]. Estimates from the World Cancer Research Fund (WCRF) meta-analysis of prospective studies have suggested that for every increase of 5 kg/m^2 in body mass index (BMI), individuals have a 5% higher risk of developing colorectal cancer, with similar relationships being observed for higher waist-to-hip ratio and waist circumference [20, 81]. This relationship between adiposity and colorectal cancer has been corroborated in MR studies [82, 83]. Adiposity is hypothesised to influence colorectal carcinogenesis through multiple mechanisms, including higher levels of insulin [74, 77, 84], greater low-grade inflammation [85], and potentially through sex hormones [85-87]; however, understanding of the mechanisms linking body size and colorectal cancer risk remains incomplete.

Physical activity

In observational and MR studies, individuals who partake in greater physical activity have a lower risk of colon cancer compared to those with lower physical activity [20, 88, 89]. Similarly, observational studies have shown that individuals with greater occupational sedentary behaviour have a greater risk of developing colon cancer [90]. However, for rectal cancer, no evidence has shown that greater physical activity modifies risk [20]. The association between physical activity and colon cancer risk is hypothesised to be in part due to maintaining lower body fat, which

may result in less insulin resistance and lower insulin levels. However, in MR studies assessing SNPs for physical activity (that were not associated with body size) in relation to the risk of colon cancer, associations remained significant [89]. Other mechanisms such as reduced inflammation [91] and changes in gut microbiota [92] may also be involved but further research is needed to understand if physical activity may influence colon cancer risk outside weight loss or decreased body size.

2.1.3 Diet and colorectal cancer risk

Dietary intake has been hypothesized to be an important modifiable risk factor for colorectal cancer for decades [93], and this remains the cancer site with the most evidence suggesting convincing and probable associations between dietary intake and risk [20]. This hypothesis of diet influencing cancer risk arose from the observed differences in cancer incidence rates between countries, as well as observations from migration studies showing that individuals moving from low incidence regions to high incidence regions would have the same incidence of cancer within one or two generations, suggesting that environmental factors, such as diet, may be important in modulating colorectal cancer risk [93-96]. **Table 2.1** presents the latest WCRF meta-analysis (updated in 2017) findings for dietary factors and risk of colorectal cancer separated by the strength of existing evidence. The next section will outline the convincing and probable associations between dietary factors and colorectal cancer risk listed in the WCRF [20] as well as the current evidence for selected dietary patterns.

Table 2.1 Dietary factors in relation to colorectal cancer risk.

		Decreases risk	Increases risk
Strong evidence	**Convincing**		Alcoholic drinks Processed meat
	Probable	Wholegrains Foods containing fibre Dairy products	Red meat
Limited evidence	**Limited - suggestive**	Foods containing vitamin C Fish	Low intakes of non-starchy vegetables Low intakes of fruit Foods containing haem iron

Table 2.1 is adapted from the World Cancer Research Fund summary of evidence from the diet, nutrition, physical activity and colorectal cancer report [20].

Alcohol

Alcohol is a major risk factor for colorectal cancer incidence accounting for approximately 8-9% of colorectal cancer cases globally [97]. In the WCRF meta-analysis, which included 19 prospective studies, a 10 grams per day (g/day) increment in alcohol intake (equivalent to about one standard drink i.e., one 5 oz glass of wine, one 12 oz beer, or 1.5 oz of spirit) was associated with a 7% higher risk of colorectal cancer. Moreover, MR studies have supported the role of alcohol in colorectal cancer carcinogenesis [98]. Intake of alcohol can lead to cancer development through multiple pathways and mechanisms, including DNA damage by acetaldehyde (a carcinogenic metabolite of ethanol) [99], oxidative stress [100], and potentially through hormone changes [101].

Processed and red meat

Early ecological studies showing that countries with higher meat intake also had higher rates of colorectal cancer suggested that colorectal cancer may be influenced by meat consumption [93].

This association was further observed in prospective cohort studies, and in the WCRF meta-analysis, which included 13 prospective studies, an increment of 50 grams of intake from processed meat (i.e., bacon, ham, or sausages) a day was associated with a 16% higher risk of colorectal cancer risk [20], which they judged to be convincing evidence. For red meat intake, such as beef, pork, and lamb, a probable positive association between red meat and colorectal cancer risk was also observed in the WCRF meta-analysis including 14 prospective studies (a 12% higher risk of colorectal cancer per 100 g/day of red meat) [20]. In 2015, the IARC listed processed meat as carcinogenic and red meat probably carcinogenic to humans [102]. Many compounds in processed and red meat have been hypothesised to underlie this association. Chemicals used to preserve processed meat, such as nitrates and nitrites, are converted in the stomach and large intestine to N-nitroso compounds, which are mutagenic [102, 103]. Moreover, cooking meat at high temperatures can create heterocyclic amines and polycyclic aromatic hydrocarbons, which are mutagenic and several of which have been shown to be carcinogenic in animal studies [103].

Foods containing dietary fibre

Dietary fibre was first hypothesized to be associated with colorectal cancer due to the noted lower incidence of colorectal cancer and high fibre intake in parts of Africa [104]. In the WCRF meta-analysis from 2017, which included 15 prospective cohort studies and 14,786 cases of colorectal cancer, a 10 g/day increment in dietary fibre was associated with a 7% reduction in risk (95% CI: 0%-13%) [20]. Higher intake of fibre as well as increasing fibre intake after a colorectal cancer diagnosis has been shown to decrease the risk of dying of the disease [105]. Fibre is hypothesized to reduce the risk of colorectal cancer through multiple mechanisms, including the reduction of faecal transit time and increase in faecal bulk, which may lessen the potential of carcinogens to interact with colonocytes [106], reduction of secondary bile acid production [107] insulin resistance [108], as well as the production of butyrate, a short-chain fatty acid (SCFA) produced from fermentation of fibre, which has been suggested to have an

anticarcinogenic effect [109]. However, mechanisms of fibre intake and colorectal cancer are not well characterised and further research is needed to better understand this.

Wholegrains

Intakes of wholegrains (e.g., wholegrain bread, oatmeal, brown rice, wholemeal pasta) have shown a probable inverse association with colorectal cancer risk [20, 110, 111]. In the most recent WCRF meta-analysis from 2017, evidence for the intake of wholegrains and colorectal cancer risk was classified as a probable inverse association. This was based on 6 prospective studies (8,320 colorectal cancer cases), all of which showed an inverse association between wholegrain intake and colorectal cancer risk, and together suggested a 17% reduction in risk (95% CI: 11%-22%) of colorectal cancer per 90 g/day of wholegrain intake [20]. Wholegrains are a good source of fibre, which may partially reduce the risk of colorectal cancer through the mechanisms outlined above. Moreover, wholegrains are a rich source of other compounds and minerals, such as lignans and polyphenolic compounds, which may act as anti-oxidants and inhibit tumour formation [112]. Despite these hypotheses, it is unclear how the intake of wholegrains modulates colorectal cancer risk and further research is needed to understand these mechanisms.

Dairy products

Probable evidence has suggested that dairy, or diets high in calcium, may be associated with a modest reduction in colorectal cancer risk [20]. In the WCRF meta-analysis from 2017 assessing the association between the intake of dairy products and the risk of colorectal cancer, 14 prospective studies suggested that a 400 g/day increase in dairy product intake was associated with a 13% lower risk of developing colorectal cancer [20]. Similar associations were observed for milk intake and diets high in calcium where a 200 g/day increase in consumption of milk and a 200 mg/day increase in calcium intake were both associated with a 6% lower risk of colorectal cancer in the WCRF meta-analysis of 13 prospective studies [20]. The association between milk intake and colorectal cancer risk has also been observed in MR analyses using the lactase gene

as a proxy for milk intake, where those with lactase persistence having a lower risk of colorectal cancer in two different cohorts [113]. Calcium may modulate colorectal carcinogenesis by binding unconjugated bile acids, thereby diminishing their potential carcinogenic effect [114, 115]. Other components of dairy products may also reduce colorectal cancer risk, such as lactoferrins and SCFAs; however, further research is needed to clarify the role of these components and mechanisms [114].

Dietary patterns

Dietary patterns, which reflect combinations of food intakes, may better capture the complexity of dietary intake as foods and nutrients are not consumed in isolation. As such, assessing whether a dietary pattern is associated with cancer risk may offer greater insight into disease prevention. To date, no dietary patterns have been established for reducing the risk of colorectal cancer [20]. Following a Mediterranean diet, that is a diet high in wholegrains, vegetables, fruits, legumes, nuts, and fish, and low in red and processed meat, has been suggested to lower the risk of colorectal cancer in some studies [116], but evidence remains inconclusive [20]. Following a vegetarian diet, that is one excluding all meat and fish, may be associated with a lower risk of developing cancer overall [117, 118], however, for colorectal cancer, the evidence remains unclear [117, 119]. In the Adventist Health Study-2 (AHS-2), a prospective cohort study with many vegetarians, individuals who did not consume meat, but consumed fish (pescatarians) had a 43% lower risk of colorectal cancer compared to meat-eaters whereas being a vegetarian was associated with a non-significant lower risk [119]. In the European Prospective Investigation into Cancer and Nutrition (EPIC) – Oxford site, another large prospective study with a large proportion of vegetarians, pescatarians had a 35% lower risk of colorectal cancer, but no association was observed for vegetarians [117]. Further research assessing dietary patterns with colorectal cancer is needed.

2.2 Female breast cancer epidemiology

Breast cancer is a malignancy originating in the breast tissue that varies in location, molecular subtypes, and grade of differentiation. Most breast cancers are carcinomas (cancers that arise in epithelial cells) and are commonly located in the ducts or lobules of the breast [120]. Breast cancers can be classified as *in situ:* pre-cancerous lesions confined to the epithelial layer; or invasive: lesions that have infiltrated the basement members of the underlying connective tissue [120]. Moreover, breast carcinomas are further characterised by molecular markers such as the presence or absence of hormone receptors, namely, oestrogen and progesterone. These hormone receptors are present in approximately 65% of all breast carcinomas and indicate that the tumour is stimulated by oestrogen and/or progesterone, thus promoting proliferation [120]. Moreover, breast carcinomas can be classified based on the presence of amplification of the *ERBB2* gene, commonly referred to as the human epidermal growth factor receptor 2 or HER2 gene. This gene codes for HER2 proteins and when over expressed can cause increased cell proliferation and is a general prognostic indicator of a more aggressive disease [120]. Considering these molecular markers, breast cancer is a heterogeneous disease with various molecular subtypes, differential grades, and metastatic potential, resulting in potential differences in aetiological factors.

2.2.1 Incidence, mortality, & geographical trends

In 2020, female breast cancer was the leading cause of cancer incidence worldwide, accounting for nearly 12% of all cancer cases [1]. An estimated 2.3 million women were diagnosed with breast cancer in 2020 and over 680,000 women died from this illness [1]. Breast cancer is much more common in North America, Europe, and Australia, where 6% of women will develop the disease before the age of 75 years [1]. This contrasts with women from sub-Saharan Africa and southern and eastern parts of Asia, where the risk is one-third lower than of women from North America, Europe, or Australia. **Figure 2.2** presents the age-standardised incidence rates of female breast cancer per 100,000 women. In some countries in Europe, such as Belgium and the Netherlands the age-standardised rate of breast cancer is 105 cases per 100,000 women, whereas

in Uganda and Myanmar the age-standardised rate is around 22 cases per 100,000 women, representing a nearly 5-fold difference in risk [24]. This difference in breast cancer rates by regions of the world is thought to reflect the differences in hormonal risk factors (i.e., earlier age of menarche, older age at first birth, fewer children, lower duration of breast feeding, and higher use of oral contraceptives and menopausal hormones), lifestyle factors such as higher alcohol intake, difference in body weight, and less physical activity, as well as the presence of breast cancer screening programmes [1, 121, 122]. However, rates of breast cancer are increasing in regions of Africa [123] and Asia [121] reflecting changes in lifestyle and environments as a result of growing economies, including women having fewer children.

Female breast cancer incidence rates have generally stayed constant in countries in North America and Europe over the past three decades, however, recent estimates have shown that breast cancer incidence is slowly increasing in the United States at a rate of $<0.5\%$ annually, as well as in some countries in Europe [122, 124]. The higher incidence in breast cancer cases has been shown to be a result of higher oestrogen receptor (ER) positive tumours, which may be due to the higher prevalence of excess body weight among these populations [124].

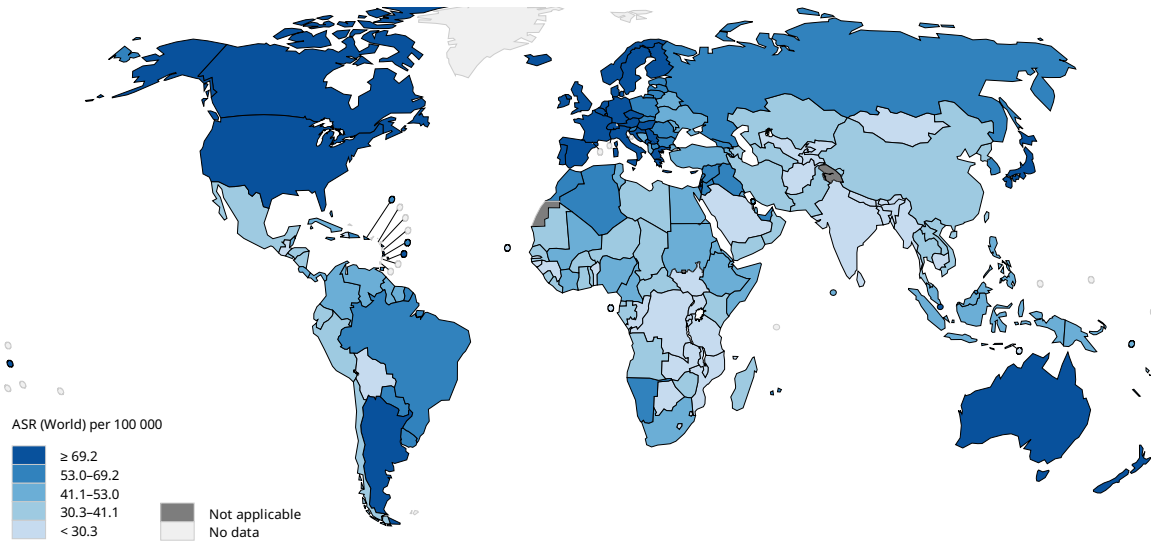


Figure 2.2 Estimated age-standardised incidence rates of breast cancer in women in 2020. Data source [24]. Abbreviation: ASR, age-standardised rate.

Despite having substantially lower cancer incidence, the mortality rate of breast cancer is substantially higher in countries located in sub-Saharan Africa [24]. For example, in Nigeria, the age-standardised rate of breast cancer is 49 cases per 100,000 women, where the mortality age-standardised rate is 26 deaths per 100,000 women, this is in contrast to the United States, where the incidence is almost two times greater (90 cases per 100,000 women) but the death rate is half that of Nigeria (12 deaths per 100,000 women) [24]. This contrast in incidence and mortality is attributed to individuals in sub-Saharan Africa presenting with advanced disease (later stage of diagnosis) than countries with higher incidence rates [125]. Population-wide breast cancer screening programmes through mammography have been proven to prevent breast cancer deaths through early detection [126], as has the development of effective treatments for breast cancer [127]. However, as most countries in sub-Saharan Africa are resource limited, introducing a mammography screening programme is difficult and potentially not cost-effective [128, 129].

2.2.2 Risk factors

Observational evidence from prospective cohort studies has established differing risk factors for developing breast cancer, which modifies a woman's risk of being diagnosed at differing points in her life. Breast cancers diagnosed before menopause (premenopausal) or after menopause (postmenopausal) have been shown to have some different risk factors and aetiologies [130] as have breast cancers diagnosed with differing hormone receptors present on the tumour cells [131]. This section will outline established risk factors for developing breast cancer, highlighting differences by menopause status.

Age

The risk of breast cancer increases rapidly with age during reproductive years until the age of 50 years, after which the incidence rate gradually slows [132]. Most breast cancers are diagnosed after the age of 50 years [133] with the median age of breast cancer diagnosis in women being 62 years [124].

Genetics

Breast cancer risk can be influenced by genetic factors, most notably by the BRCA1 and BRAC2 gene mutations, which are two highly penetrant, inheritable genes that convey a substantial increased risk of breast cancer if a woman is a carrier. These two genes, which normally produce tumour suppressor proteins, are attributable for 5% of all breast cancer cases [134]. BRCA mutations are fairly rare in most populations with about 0.2% of the population being carriers of either of these genes, however, individuals of Ashkenazi Jewish ancestry are more likely to be carriers of these genes [134]. GWAS have also identified more than 180 low penetrant SNPs related to breast cancer risk [135]. Taken together, multiple low-penetrant mutations in combination can result in a substantially higher risk of breast cancer [38].

Hormones

Endogenous hormones, specifically oestrogens and androgens, play an important role in breast cancer development [136-138]. In a meta-analysis of individual participant data from seven prospective studies, a doubling of oestradiol, free oestradiol, oestrone, testosterone and free testosterone were all positively associated with premenopausal breast cancer risk [136]. Similar results were observed for postmenopausal women in an analysis of nine prospective studies, where higher oestrogen and lower sex-hormone binding globulin (SHBG) were associated with a higher risk of breast cancer [137]. In MR studies, oestradiol and testosterone have been positively associated with breast cancer risk and SHBG has been inversely associated with breast cancer risk, however, these associations may be only for ER-positive tumours [139-141]. Oestradiol stimulates cell division and growth of breast epithelial cells and bioavailability of oestradiol is partially regulated by SHBG, thus higher levels of SHBG may reflect less available oestrogens. Sex hormones may be important mediators for risk factors associated with breast cancer risk, such as alcohol intake and obesity [142-144]; however, further research is needed to further examine this and characterise the proportion of associations due to differences in sex hormones.

Reproductive factors

As the breast tissue consists of cells that are modulated and stimulated by sex hormones, such as oestrogen and progesterone, a woman's reproductive history, which influences the exposure of endogenous hormones, is strongly associated with the risk of breast cancer. Women who have experienced early menarche, have no children (nulliparous), have few children later in their life (after the age of 30 years), and experience menopause at a later age, all have increased risk of breast cancer [145, 146]. In a meta-analysis that included individual participant data from 117 studies, each year earlier of the start of menarche was associated with a 5% higher risk of breast cancer [145]. This same meta-analysis also observed that later onset of menopause was associated with higher breast cancer risk (2.9% higher risk of breast cancer per every year later) [145]. The association between early age of menarche and breast cancer risk is supported by MR studies, suggesting a potential causal association [147]. Later age of menopause has also been shown in MR analyses to be strongly positively associated with breast cancer incidence [148]. Premenopausal women have also been shown to have a 43% higher risk of breast cancer than postmenopausal women of the same age [145]. Ovarian production of sex hormones begins at menarche and decreases at menopause and previous evidence has shown that higher sex hormones, such as oestradiol, are associated with higher breast cancer risk [136, 137]. In a prospective cohort of 1.2 million women with more than 27,000 breast cancer cases, women who had their first child after the age of 30 years had a 27% higher risk of ductal carcinoma and 79% higher risk of lobular carcinoma in comparison to women who had a child under the age of 20 years [149]. Although long-term breast cancer risk is lower in women who have given birth earlier in their life, there is a short-term increased risk of breast cancer postnatally [150]. Breastfeeding has also been shown to decrease the risk of breast cancer although the magnitude may be small; the WCRF meta-analysis from 2017, which included 13 studies, observed that each 5 month increase in breastfeeding duration was associated with a 2% lower risk of breast cancer [130].

Menopausal hormone therapy and oral contraceptives

Similar to how endogenous hormones influence breast cancer risk, the use of exogenous hormones such as oral contraceptives and menopausal hormone therapy (MHT) has been associated with increased breast cancer risk [151, 152]. Specifically, women who are current or recent users of oral contraceptives that contain oestrogen and progesterone have shown to have a 24% and 16% increased risk of breast cancer compared to non-users, respectively [151]. In a meta-analysis of individual participant data, the use of MHT had a robust positive association with breast cancer risk with even stronger associations for longer duration and for combined oestrogen-progestogen therapies [152]. For example, women who took an oestrogen-progestogen MHT for over 15 years had a 2.5-fold greater risk of being diagnosed with breast cancer compared to never users and this risk continues for multiple years after stopping [152]. From this evidence, the IARC classifies both oral contraceptives and menopausal hormone therapy as group I carcinogens for breast cancer [153]

Breast density

Breast density is classified from the amount of fibro-glandular tissue present in the breast in comparison to fat tissue. The density of breast tissue has been shown to be positively associated with breast cancer risk, with estimates suggesting a 4 to 6-fold increase in risk among women with extremely dense breasts (>75% mammographic density) compared to women who have fatty non-dense breasts [154].

Radiation

Exposure to ionising radiation through occupational exposures or through X-rays, particularly during childhood and pubertal years, has been shown to increase the risk breast cancer [155, 156]. In female flight attendants, the risk of breast cancer has been observed to be higher, which has been hypothesised due to their increased exposure to cosmic radiation while flying [157]. From the observed association between radiation and breast cancer, and on other cancer sites, the IARC classifies radiation as a group I carcinogen [158].

Diabetes mellitus

Evidence from a meta-analysis of prospective cohort studies has suggested that women with diabetes have a 23% higher risk (95% CI: 12-35%) of breast cancer than women without diabetes [159], however, this association is not observed amongst women with type I diabetes. Higher circulating insulin has been shown in a MR study to be positively associated with breast cancer risk [160]. As such, women with early stage and untreated type II diabetes may have an elevated risk of breast cancer due to hyperinsulinemia, similar to the mechanisms that influence colorectal cancer carcinogenesis described previously.

Adiposity

Adiposity, commonly measured using BMI, is a complex risk factor for breast cancer dependent on a woman's menopausal status [130]. In postmenopausal women, it is well established that having overweight or obesity puts a woman at greater risk of breast cancer; the latest WCRF meta-analysis (2017) utilising 56 prospective studies estimates a 12% greater risk of postmenopausal breast cancer per 5 kg/m^2 increase in BMI [130]. In contrast, having overweight or obesity may be protective against premenopausal breast cancer, with a 5 kg/m^2 increase in BMI reflecting a 7% lower risk of premenopausal breast cancer in a meta-analysis of 37 prospective studies [130]. Some evidence has suggested that the inverse association between BMI and premenopausal breast cancer may be stronger for hormone receptor positive cancers [161]. Mechanistically, this inverse association between higher BMI and premenopausal breast cancer has been hypothesized to be a result of overweight women having less exposure to sex hormones, dysregulation or fewer ovulatory cycles, less susceptibility to carcinogens throughout puberty and young adulthood, and greater differentiation of breast tissue [162, 163]. For postmenopausal breast cancer, higher BMI has been suggested to increase the risk of breast cancer through higher circulating hormones, including oestrogens, and potentially greater low-grade inflammation [164, 165]. For postmenopausal women, adipose tissue becomes the main source of oestrogen through aromatisation of androgens, and women who have higher BMIs may have lower levels of SHBG [166], which may lead to greater free oestradiol. In turn, this may lead to

higher circulating serum oestrogen and greater locally produced oestrogens among individuals with greater body fat [144].

The observations between BMI and premenopausal breast cancer are supported by MR studies using genetic variants for adiposity [167], however, MR studies have suggested an inverse association with postmenopausal breast cancer, which contrasts with the observations from prospective studies [167, 168]. This discrepancy may be due to genetic variants predicting early-life body size; in a longitudinal analysis assessing BMI genetic variants, the 97 genetic variants were more strongly associated with earlier life BMI than later life [169] and MR analyses have shown that genetic variants predicting larger body size in earlier life are inversely associated with breast cancer risk, whereas genetic variants for adult body size have been non-significantly positively associated after adjusting for early life BMI [170]. As such, the discrepancy between MR and prospective studies for postmenopausal breast cancer may be driven by adult weight gain [171]. In the latest WCRF meta-analysis from 2017, which included 15 prospective studies, 5 kg weight gain in adulthood was associated with a 6% higher risk of postmenopausal breast cancer [130] and genetic variants utilized in MR analyses might be less capable of adequately capturing this phenotype [171].

Height

Attained adult height has been positively associated in prospective studies with breast cancer incidence for both premenopausal (6% higher risk per 5 cm) and postmenopausal (9% higher risk per 5 cm) breast cancer [130], as well as in MR studies using genetic variants for height [172]. Height may be a good marker of the number of cells in a body and therefore risk [173], with a combination of environmental and genetic factors leading to a woman's attained height. More precisely, the combination of genetics, hormones, and early-life nutrition all influence body composition and therefore nutritional factors that influence height may also influence cancer risk, and height may be a surrogate marker of this exposure. Furthermore, hormonal influences are probably important for breast cancer risk in association with height, as particular tissues in taller people are exposed to higher levels of insulin, pituitary-derived growth hormone (GH), and

IGF hormones. IGF-I is a major regulator of growth in utero and during childhood and adolescence [6] and multiple genetic variants on the IGF signally pathway have been associated with height [174]. Further understanding of the factors associated with determinants of height and how these influences breast cancer carcinogenesis is needed.

Physical activity

Being physically active or undertaking vigorous physical activity has been shown to be inversely associated with both premenopausal and postmenopausal breast cancer risk [130]. In the latest WCRF meta-analysis of prospective studies, comparing women who did the highest amount of vigorous activity versus the lowest, those in the highest category had a 17% and 10% lower risk of premenopausal and postmenopausal breast cancer, respectively [130]. In a MR analysis assessing the association of physical activity and breast cancer risk, genetic variants that predicted physical activity were inversely associated with breast cancer risk even after excluding SNPs associated with body size or after controlling for BMI [89], suggesting an independent association of physical activity with breast cancer risk. Multiple hypothesised mechanisms exist for why greater physical activity may confer a lower risk of breast cancer outside of having a lower body weight [175], this includes lower circulating concentrations of hormones including insulin, and oestradiol [176, 177], improved immune response to cancerous cells [178], and reduction of systemic inflammatory markers [91, 179]; however, these mechanisms have not been fully explored.

Smoking

Although previously thought to be not associated with breast cancer risk [133], smoking has been shown to be modestly associated with a higher risk of breast cancer. In a recent meta-analysis, which included 14 prospective cohort studies, current smokers had a 12% higher risk of developing breast cancer in comparison to never smokers [180] and this positive association has also been observed in MR analyses including more than 120,000 breast cancer cases [65]. Tobacco smoke contains many carcinogenic compounds, such as ethylene oxide and benzene,

that have induced breast cancer in rodent models [181], and smoking metabolites have been found in the breast tissue of smokers [182]. Some evidence has suggested that smoking may be positively associated with sex hormones [166]. However, earlier studies have shown that smoking tobacco may have an anti-oestrogenic effect, which might mask the influence of carcinogenic compounds present in smoke on breast tissue [183, 184]. This may explain why earlier studies have not observed an association between smoking and breast cancer risk [142], however, this proposed anti-oestrogenic effect remains inconclusive.

2.2.3 Diet and breast cancer risk

Besides alcohol consumption, the associations of specific diets or dietary components with breast cancer risk remains unclear. One of the first hypotheses that gained considerable research attention was the association between fat consumption and breast cancer risk [185]. When looking from an ecological perspective, there appeared to be an association between fat consumption and breast cancer incidence, specifically the more fat a country consumed on average the greater breast cancer rates were observed [185]. However, this association was not observed in individual-level prospective cohort studies [186], nor in trials with women randomised to follow a low-fat diet [187, 188], suggesting that this initial observation was likely due to confounding or other factors resulting in an ecological fallacy. Although other dietary factors have been studied extensively in association with breast cancer, little evidence remains consistent, potentially due to the natural heterogeneity of breast cancer.

Dietary intake has been hypothesised to potentially influence sex hormones through multiple mechanisms; animal-derived foods may contain endogenous hormones and nutrient components, such as cholesterol, are important substrates in sex hormone production [189]. RCTs have also observed differences in sex hormone levels by dietary intake, with early evidence suggesting that sex hormones may be influenced by fat intake [190], however, low-fat intake interventions have found little or no changes in oestrogens without weight loss [191-193]. In one RCT with an intervention aimed at reducing energy intake and promoting weight-loss, oestradiol concentrations decreased by 16.2% [176]. As such, teasing out the independent

association of specific dietary components from the influence of BMI is difficult. Dietary factors have also been shown to be associated with the timing of puberty and age of menopause; some evidence has suggested that intake of protein from animal sources is associated with earlier menarche [194-196] whereas following a vegetarian diet has been associated with earlier onset of menopause [197-199]. However, how dietary factors specifically influence timing of reproductive events, and how this translates into differences in breast cancer risk remains unclear. **Table 2.2** presents a summary of the 2017 WCRF meta-analysis for dietary factors and breast cancer risk by strength of evidence as well as by menopausal status.

Table 2.2 Dietary factors in relation to breast cancer risk

		Decreases risk	**Increases risk**
Strong evidence	**Convincing**		Alcoholic drinks[2]
	Probable		Alcoholic drinks[1]
Limited evidence	**Suggestive**	Dairy products[1] Non-starchy vegetables (oestrogen receptor-negative cancers only) Foods containing carotenoids Diets high in calcium	

Table 2.2 is adapted from the World Cancer Research Fund summary of evidence from the diet, nutrition, physical activity and breast cancer report [130]
[1] represents associations for premenopausal breast cancer where [2] represents associations for postmenopausal breast cancer.

Alcohol

Alcohol remains the most convincing dietary related risk factor for breast cancer; in a meta-analysis of 22 prospective studies, alcohol intake (per 10 g/day) was associated with a 9%

greater risk of postmenopausal breast cancer, and in meta-analysis of 10 prospective cohorts, a 10 g/day increase in alcohol intake was associated with a 5% greater risk of premenopausal breast cancer [130]. From these results, the WCRF listed the evidence for the association of alcohol intake and breast cancer risk as strong and convincing for postmenopausal breast cancer and strong and probable for premenopausal breast cancer [130], and the IARC classify alcohol intake as a group I carcinogen for female breast cancer [200, 201]. Conversely, the assessment of genetic variants related to alcohol intake and breast cancer have found no association with breast cancer risk [202]. Reasons for this null association may be that the genetic instruments for alcohol intake are only weakly associated with intake, there is no causal effect, or the causal effect is small [202]. Multiple proposed mechanisms exist for why alcohol may induce breast carcinogenesis including elevating sex hormones, metabolization of alcohol to acetaldehyde, which may induce DNA damage, or lowering folate status which may change DNA methylation patterns [166, 203, 204]. Moreover, alcohol intake has been observed to have a stronger association with ER-positive tumours compared to ER-negative tumours [130], which may suggest that hormonal changes induced by alcohol consumption are important. However, specific mechanisms are not well established, and it is likely that multiple factors interplay to induce a higher risk of breast cancer.

Diets high in calcium

Limited – suggestive evidence exists that higher intake of calcium may be associated with lower breast cancer risk for both premenopausal and postmenopausal breast cancer [130]. In the latest WCRF meta-analysis from 2017, which included five prospective studies, a 300 mg/day increase in calcium intake was associated with a 13% lower risk of premenopausal breast cancer and in 6 prospective studies, a 300 mg/day increase in calcium was associated with a 4% lower risk of postmenopausal breast cancer [130]. In a recent MR study assessing the association between micronutrients and risk of breast cancer, a suggestive inverse association was observed between calcium concentrations and overall breast cancer risk [205]. *In vitro* studies have suggested that calcium modulates breast carcinogenesis by regulating cell proliferation, differentiation, and

apoptosis [206]. Calcium is tightly regulated by the body, with approximately only 1% of total calcium being found in the blood and intracellular tissue [207]. Inadequate intake of calcium or poor absorption can decrease circulating ionised calcium, which will trigger homeostatic mechanisms to maintain a sufficient level. This variation in dietary calcium intake may impact carcinogenesis although this remains unclear and specific mechanisms are yet to be elucidated [130].

Dairy – premenopausal breast cancer

Some evidence suggests that consuming dairy products is associated with a slight lower risk of premenopausal breast cancer. Specifically, in the WCRF meta-analysis of seven prospective studies, a 200 g/day increase of dairy product consumption was associated with a 5% reduced risk of premenopausal breast cancer, whereas no significant association was observed for postmenopausal breast cancer [130]. The inverse observation for dairy intake and premenopausal breast cancer may be potentially due to the high calcium content in the dairy products. However, in a MR analysis using the lactase gene as a proxy for milk intake, the most commonly consumed dairy product, there was no association between lactase persistence and breast cancer risk utilising three different studies [113]. However, the null finding in this MR study may be a result of assessing risk with total breast cancer cases rather than by menopausal status. Regardless, it remains unclear whether these associations are causal and if they differ by menopause status [130].

Non-starchy vegetables and foods containing carotenoids

The 2017 WCRF meta-analysis of prospective studies found that a 200 g/day increase in non-starchy vegetables (approximately 2 servings) was associated with a 21% lower risk of ER-negative breast cancer in three prospective studies [130]. However, the observation between non-starchy vegetables was not observed for overall breast cancer risk by menopausal status, or for any other hormone receptor subtype (i.e., ER+), and the number of studies with information on hormone status is limited.

Limited evidence suggests that the intake of foods that contain carotenoids (a phytonutrient found in certain vegetables), such as alpha- and beta-carotene, have an inverse association with overall breast cancer risk [130]. These associations have been observed from dietary questionnaires but are stronger when measured in blood samples of circulating carotenoids [208]. In a meta-analysis of nested case-control studies, circulating carotenoids were inversely associated with breast cancer risk, whereas carotenoid intake measured by dietary questionnaires has been less consistent [130, 208] potentially due to poor measurement of intake, and the association being stronger for ER-negative tumours [130]. Serum measurements of carotenoids may reflect a diet rich in many bioactive phytochemicals derived from fruits and vegetables which in combination act to reduce breast cancer risk [208]. Carotenoids may reduce proliferation of cells and have antioxidant properties in *in vitro* studies, which may prevent reactive oxygen and free radicals from damaging DNA [209]; however, further research is needed to understand these relationships.

Dietary patterns

Following specific dietary patterns such as the Mediterranean diet, a low-fat diet, or adherence to dietary guidelines have not shown a consistent association with breast cancer risk, regardless of menopausal status [130]. In cross-sectional studies, consuming a Western dietary pattern (high in meat and processed foods) has been modestly positively associated with total oestrogen and free oestradiol levels [189, 210] and meat intake may be related to lower SHBG concentrations [211]; however, how much this is due to differences in BMI and whether following this type of diet is associated with breast cancer risk remains unclear [130].

Following a vegetarian or pescatarian diet has not been shown to be associated with the risk of breast cancer in EPIC-Oxford [117] nor in the AHS-2 [212]. Although vegetarians and pescatarians may consume a more plant-rich diet, with greater variation, whether this influences breast cancer risk remains inconclusive. Some studies have suggested that red or processed meat may be carcinogenic to breast tissue [213], but in general, epidemiological studies have found no association [130]. Vegetarians also consume greater amounts of soy products, which contain

phytoestrogens that may potentially reduce the risk of breast cancer [214]. Vegetarians and pescatarians also consume more fibre, on average, than meat-eaters [215, 216], which has been suggested to potentially reduce the risk of breast cancer in a recent meta-analysis of prospective studies [217]. Despite this, evidence to date has suggested that following any dietary pattern does not modify breast cancer risk and further studies with a large number of breast cancer cases are needed to explore this.

2.3 Prostate cancer epidemiology

Prostate cancer originates in the prostate gland, which functions to provide fluid to transport and nourish sperm. Over 95% of diagnosed prostate cancer are adenocarcinomas, with the two main types being acinar adenocarcinoma and to a lesser extent ductal adenocarcinoma [218]. Prostate cancer typically develops slowly with some tumours being indolent, however other types of prostate cancer can be aggressive, spread to other parts of the body, and result in death [219, 220]. Prostate cancer aggressiveness is scored using a combination of size and spread of the tumour, if it has infiltrated local lymph nodes, if there is distant metastasis to other parts of the body, as well as how differentiated tumour cells are. The Gleason score is primarily used to score prostate tumours based on differentiation. This score originated in the 1960s from a prospective cohort study that summarised the histological growth patterns of the tumour and compared the results with prognosis and staging [221, 222]. A Gleason score of $\leq$7 indicates a low-grade tumour or well/moderately differentiated tumour and is therefore less aggressive whereas a Gleason score $\geq$8 indicates a high-grade tumour, which is undifferentiated and is potentially more aggressive. Risk factors for prostate cancer may differ by stage and grade of the tumour making prostate cancer a heterogeneous disease with differences in molecular subtypes and overall prognosis.

2.3.1 Incidence, mortality, & geographical trends

Global estimates suggest that over 1.4 million men were diagnosed with prostate cancer and nearly 380,000 men died because of the disease in 2020, making it the second most common

cancer and fifth leading cause of cancer death in men [1]. Prostate cancer incidence is much higher in countries with a high development index; in countries such as Sweden and Canada the age-standardised rates are 80 and 100 cases per 100,000 men per year, respectively, whereas in Asian countries, where incidence rates are lower, the age-standardised rate is approximately 5-10 cases per 100,000 men [24]. **Figure 2.3** presents the age-standardised incidence rates of prostate cancer by countries across the world.

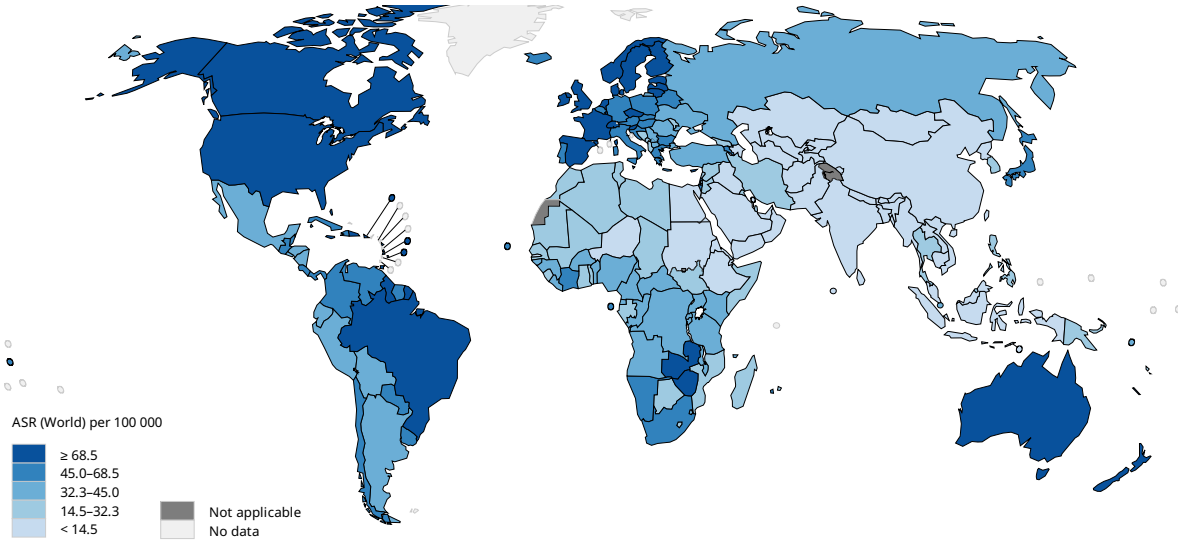


Figure 2.3 Estimated age-standardised incidence rates of prostate cancer in men in 2020. Data source [24]. Abbreviation: ASR, age-standardised rate.

Prostate cancer is the leading cause of cancer death among men in 48 countries [1, 24] and mortality rates are increasing in many countries located in Central and Eastern Europe, Asia, and Africa [223]. The mortality rate from prostate cancer by country is less variable than the incidence rate, with the highest mortality rates being observed in Caribbean and African countries [1]. In the Caribbean and countries located in the middle of Africa, the age-standardised mortality rate is estimated to be 27.9 and 24.8 deaths per 100,000 men per year, respectively [1]. Mortality rates are lowest in South-Central Asia, where the age-standardised rate is approximately 3.1 prostate cancer deaths per 100,000 men per year. In countries with high prostate cancer incidence such as in Europe and North America, the mortality rate is

relatively low with an age-standardised prostate cancer death rate of approximately 8-13 deaths per 100,000 men.

The difference in incidence rates between countries in North America and Europe and those in Asia and eastern Africa is likely partly due to the use of prostate specific antigen (PSA) testing. Unlike colorectal and breast cancer, for which there is organised screening in many countries where individuals are screened at a certain age, there is no formal screening programme for prostate cancer in any country. However, PSA testing, which became clinically available in the late 1980s, has been utilised as a monitoring, screening, and diagnostic tool for prostate cancer [224, 225]. PSA is produced by the prostate gland and elevated levels in circulation indicate greater prostate cell activity, which can be a result of a prostate cancer tumour. Upon indication of abnormal PSA levels, a biopsy will confirm if prostate cancer is present. However, it is estimated that up to 67% of prostate cancers detected by PSA testing may not have become clinically apparent during a man's life and thus would not require treatment [225, 226]. As such, with the introduction of PSA testing, large increases in prostate cancer diagnoses were observed, particularly for non-aggressive disease and among countries where such testing was common [227]. This difference in incidence may reflect both overdiagnosis of non-aggressive tumours in these countries and earlier diagnosis of aggressive disease, which may allow for enhanced treatment options [228, 229]. In recent years, the use of PSA screening has decreased with governing bodies such as the United States Preventative Services Task Force ruling against the use of PSA testing among men in the general population to screen for prostate cancer [230]. This decrease in use of PSA testing may in part explain why there is a decrease in incidence of prostate cancer among some countries where testing was common. To date, PSA testing is still conducted in countries such as the UK and the United States, however systematic implementation of the test has not been recommended for asymptomatic men and current guidelines suggest shared decision-making between healthcare providers and individuals that weighs the uncertainties, risks, as well as possible benefits [231-233]. One study that followed more than 450,000 UK men aged 45-69 estimated that around 39% had a PSA test over a ten-year period [234].

2.3.2 Risk factors

Despite prostate cancer being so common, few aetiological risk factors have been elucidated. Moreover, investigating potential modifiable risk factors for prostate cancer has been challenging due to the high proportion of prostate cancer that is 'over-diagnosed' from PSA testing, which is often low-grade and non-aggressive disease that may not be, or ever become, clinically relevant [225, 226]. For modifiable risk factors, only circulating IGF-I concentrations and free testosterone have been suggested as established risk factors. This next section will outline the few risk factors associated with higher prostate cancer risk.

Age

Older age is a major risk factor for prostate cancer with very few (~1%) prostate cancers being diagnosed before the age of 50 years [235, 236]. After the age of 50, the age-standardised incidence rate of prostate cancer increases until approximately 80 years of age in which the incidence rate declines [235]. Evidence from autopsy studies have suggested that nearly 60% of all men over the age of 79 that die from other causes will have histological evidence of prostate cancer [237]. Despite the lower incidence rate, potentially due to less diagnostic activity, men over the age of 75 years may be more likely to be diagnosed at an advanced stage and die of prostate cancer [238, 239].

Family history and genetics

Prostate cancer may have a large heritable component [240, 241]. In a meta-analysis of 26 studies, having a first-degree relative or two or more first degree relatives with a prostate cancer diagnosis was associated with a 2.5-fold higher risk and 4.4-fold higher risk of prostate cancer, respectively, compared to men with no family history [242].

To date, more than 160 low-penetrant genetic variants have been identified that are associated with prostate cancer [241, 243-245]. Together, these low-penetrant genetic variants can contribute to a substantially higher risk of developing prostate cancer; men in the top 10% of a genetic risk score derived from these genetic variants have a 2.9-fold higher risk of prostate

cancer compared to the population average [241]. To date, a third of familial risk of prostate cancer can be explained by known genetic variants [245].

The G84E mutation is a high-penetrant rare mutation located on the *HOXB13* gene and is estimated to confer a 4-fold to 8-fold increased risk of total prostate cancer [246]. Moreover, individuals with BRCA2 mutations or Lynch syndrome have also been shown to have a 2-fold to 4-fold higher risk of developing prostate cancer [247, 248].

Relatively little is known about genetic factors associated with advanced or aggressive prostate cancer as most studies examining genetic variants have used overall prostate cancer as the outcome. However, a GWAS has identified 16 SNPs related to aggressive prostate cancer, all of which were also associated with overall prostate cancer risk [245]. Further research is needed to compare genetic variants of nonaggressive and aggressive tumours to better identify individuals who may be at a greater risk of developing advanced disease.

Ethnicity

In the United States, non-Hispanic black men are 76% more likely to be diagnosed with prostate cancer and 2.3 times more likely to die of prostate cancer than white men [249]. Approximately 1 in 7 black men in the United States will be diagnosed with prostate cancer over their lifetime and the diagnosis of prostate cancer is often in younger men with more aggressive tumours [249]. This difference in risk has also been observed in the UK, where black men have 2- to 3-fold higher risk of being diagnosed with prostate cancer compared to white men [250]. The underlying reason for this difference in risk by ethnic group is not fully understood but may be due to genetic, molecular, and lifestyle differences [245, 251, 252]. GWAS have identified a few genetic variants that are more common among males from African ancestry than European ancestry and are associated with a higher risk of developing prostate cancer [245, 253]. However, the cause of the difference in ethnicity is likely multifactorial, with differences in genetics, hormones, and lifestyle contributing to this disparity making it difficult to disentangle each causal component.

Sex hormones

Androgens are involved in the development and growth of the prostate gland and are integral for normal prostate functioning [254]. Androgen deprivation therapy, which reduces the production of androgens and blocks their action on the body, influences prostate cancer risk and acts to slow progression among men with prostate cancer [255, 256]. Prostate cancer has been shown to overexpress the androgen receptor, thus implicating testosterone and dihydrotestosterone in the aetiology of the disease. In circulation, testosterone is bound to SHBG and albumin with approximately only 2% of testosterone circulating unbound and considered "free" and bioavailable [257]. In a recent analysis from a collaboration of 20 nested case-control studies, which included 6933 prostate cancer cases, men in the lowest tenth of free testosterone concentrations had a 23% lower risk of total prostate cancer compared to all the other men, but a non-significantly 56% higher risk of high-grade tumours [258]. This finding may reflect a saturation effect; after a certain threshold the androgen receptors on prostate cells become saturated and additional increases in concentrations of free testosterone have no biological effect [259]. MR analyses have implicated testosterone to be important in prostate cancer development, with recent evidence suggesting that free testosterone is associated with aggressive and early-onset prostate cancer as well as overall disease risk [14, 260]. However, observational analyses using pre-diagnostic estimated free testosterone levels have shown no association with aggressive prostate cancer risk but a positive association with early-onset prostate cancer [260]. Possible explanations for these differences could include limitations of observational evidence (i.e., confounding), or genetic variants representing exposure to higher free testosterone at key developmental periods rather than adulthood [261].

Height

In observational studies, taller adult-attained height has been associated with higher risk of prostate cancer [262]. In the latest WCRF meta-analysis from 2014, which included 34 prospective studies, a 5 cm increment in height was associated with a 4% higher risk of total, advanced, or fatal prostate cancer [262]. However, in MR analyses, genetic variants for height

have not been shown to be associated with prostate cancer risk [263, 264]. As such, genetic determinants of height may not be causally associated with prostate cancer, but environmental factors that influence adults' height, such as early life health and nutrition, may explain the differences in risk due to height [264]. Hormones such as GH and IGF-I play an important role in the regulation of bone growth [265, 266] and may be important mediators between the association of height and prostate cancer risk.

Body fat

In the most recent WCRF meta-analysis, body fatness was listed as being probably associated with risk for advanced prostate cancer [262]. In 23 studies, a 5 kg/m^2 increase in BMI was associated with an 8% higher risk of advanced prostate cancer, with similar results being observed for waist circumference and waist-hip ratio [262]. For total prostate cancer, no association with BMI has been observed [267], and in some prospective studies, an inverse association has been observed with total prostate cancer [268, 269], which may be due to differences in detection. A recent meta-analysis of 19 prospective cohort studies, which included 19,633 prostate cancer deaths, found evidence that BMI was associated with a 10% higher risk of dying of prostate cancer per 5 kg/m^2 increment in BMI [270]. Despite the consistent associations observed in observational evidence, MR analyses have not found an association between adult BMI and aggressive prostate cancer [264, 271], however one analysis did find a positive association between a genetic score for birthweight and aggressive prostate cancer risk [271]. The observed association between BMI and aggressive disease and mortality may be subject to detection bias [272]. In men who are obese, PSA concentrations are on average lower [267], prostates are typically larger, and a digital rectal exam is difficult to conduct [273], which may lead to missing an indication for a potential prostate tumour. Men who are overweight may be less health conscious and less likely to seek medical care if symptoms of prostate cancer appear. This may lead to a diagnosis of prostate cancer at a later stage and thus a worse prognosis.

Diabetes mellitus

As described, living with diabetes has been shown to increase the risk of numerous cancers including breast and colorectal cancer [73], however for prostate cancer, an inverse association has been observed between having diabetes and prostate cancer in prospective studies [250, 274, 275]. In a meta-analysis of 32 prospective cohort studies, a 14% lower risk of total prostate cancer was observed for men with diabetes [274], and in two recent large prospective studies in Europe estimated a 26% and 30% lower risk of prostate cancer among men with diabetes [250, 275]. However, MR studies do not support a causal association between type II diabetes and prostate cancer risk with genetic predictors for type II diabetes, insulin, and glucose all not being associated with prostate cancer risk [73, 75, 276]. Multiple hypothesized mechanisms exist for why men with diabetes may have a lower risk of prostate cancer including men with diabetes having lower circulating IGF-I and testosterone concentrations [277, 278] and potentially the anti-carcinogenic effect of metformin, which is a commonly prescribed drug for individuals living with diabetes [279], however further research is needed to better understand this.

Smoking

Previous meta-analyses have shown that smoking is not associated with prostate cancer incidence [280]. However, in a meta-analysis of 51 studies from 2014, smoking cigarettes was associated with a 24% higher risk of prostate cancer death [281], however, whether this is due to residual confounding, smokers being diagnosed later, or not adhering to PSA testing in follow-up remains unknown [281]. Smoking may be associated with prostate cancer risk by modifying circulating sex hormones [261], however this needs to be further explored.

2.3.3 Diet and prostate cancer risk

Overall, the evidence between diet and prostate cancer risk remains inconclusive, with no strong evidence suggesting an association for dietary factors in prostate cancer aetiology [262]. In the most recent WCRF meta-analysis of prospective studies, the only dietary factors that were

listed were dairy products and diets high in calcium as having "limited – suggestive" evidence for increasing the risk of prostate cancer. **Table 2.3** presents a summary of the 2014 WCRF meta-analysis for dietary factors and prostate cancer risk. This next section will outline the association of dairy products and diets high in calcium and discuss the evidence for other dietary factors and patterns.

Table 2.3 Dietary factors with limited evidence in relation to prostate cancer risk

		Decreases risk	**Increases risk**
Limited evidence	**Suggestive**		Dairy products Diets high in calcium

Table 2.3 is adapted from the World Cancer Research Fund summary of evidence from the diet, nutrition, physical activity and prostate cancer report[262]

Dairy products

In the latest WCRF meta-analysis from 2014, which included 15 prospective studies, a 400 g/day increase in total dairy product consumption was associated with a 7% higher overall prostate cancer risk (95% CI: 2%-12%). Few prospective studies have looked at associations by stage and grade of disease and thus results from the WCRF meta-analysis were insufficiently powered to stratify by subtypes of prostate cancer risk; however, preliminary evidence indicated a higher risk for non-advanced prostate cancer, and no association for advanced prostate cancer [262]. Milk intake was also explored with prostate cancer risk and total milk intake was positively associated with total prostate cancer risk, but this may only be for low-fat milk consumption [262]. Some evidence has also suggested that higher milk intake is associated with prostate cancer progression [282, 283], however evidence remains unclear if these associations are causal and what are the underlying mechanisms for this. In MR analyses using the lactase gene as a proxy for milk intake, no associations were observed combining three large studies,

however, a significant positive association was observed in one of the studies, namely the FinnGen consortium, which may be explained as the Finnish population is quite homogenous and have relatively high milk intake [113].

Intake of dairy products and milk intake have been shown to be positively associated with circulating IGF-I concentrations in cross-sectional studies [3-5] and RCTs [284-286], which may be due to their high protein content. Dairy products are also high in calcium, which may explain the higher prostate cancer risk [262]. The role of circulating IGF-I concentrations in relation to prostate cancer risk is discussed in greater detail below.

Diets high in calcium

In the WCRF meta-analysis of 15 prospective studies, a 400 mg/day increase in calcium intake was associated with a 5% higher risk of total prostate cancer [262]. In countries located in North America and Europe, calcium intake is an approximate marker for dairy product intake and disentangling these two is difficult. In studies assessing non-dairy calcium intake, no association with prostate cancer risk has been observed [287]. Calcium has been hypothesised in studies to influence prostate cancer risk by regulating the formation of 1,25-dihydroxy vitamin D_3, resulting in lower concentrations, which may increase cell proliferation and inhibit differentiation of cells in the prostate [288, 289]. However, in MR analysis, genetic variants for serum calcium concentrations were not significantly associated with prostate cancer risk [290], which may suggest that observational evidence between calcium and prostate cancer is confounded. Similarly, in RCTs assessing calcium supplementation, no effect of taking a calcium supplement was observed in developing prostate cancer [291, 292] although studies may be underpowered due to the few participants (~1800) and small number of prostate cancer cases (48 cases total).

Other dietary factors

Some other dietary factors have been implicated in the risk of prostate cancer, but overall evidence remains inconclusive. The 2014 WCRF meta-analysis determined that there was "limited – no conclusion" for intake of dietary lycopene, an antioxidant found in tomatoes, and

the risk of prostate cancer. However, in a 2017 meta-analysis of dietary intake of lycopene and circulating concentrations of lycopene, an inverse association with prostate cancer was observed [293], although no association was observed for advanced disease. Other studies have suggested that higher levels of some micronutrients, including vitamin D, vitamin E, and selenium, may be associated with a lower risk of prostate cancer, but the findings from RCTs and MR analyses are overall null or inconclusive [262, 294-296].

Dietary patterns

The evidence for following a specific dietary pattern and risk of prostate cancer remains unclear [262]. In EPIC-Oxford, a non-significant inverse association was observed between being a vegetarian or vegan and prostate cancer risk compared to meat-eaters [117], whereas in AHS-2, being vegetarian was not associated with prostate cancer risk, but being vegan was associated with a 34% lower risk of developing prostate cancer compared to meat-eaters [297]. Vegans do not consume any animal products, including meat, dairy, or eggs, and evidence has suggested that they have lower circulating IGF-I concentrations compared to meat-eaters [298, 299], potentially due to excluding dairy from their diet. However, vegetarians and vegans are less likely to have a PSA test [300], which may impose detection bias in prospective cohort studies without adequate adjustment for testing. Besides the possible association with dairy, dietary risk factors for prostate cancer are unclear, and the modest association with dairy intake is unlikely to explain this difference in risk for vegans.

2.4 Circulating insulin-like growth factor-I and risks of colorectal, breast, and prostate cancer

IGF-I belongs to the IGF axis, which consists of a group of proteins with two major ligands (IGF-I and IGF-II) and six binding proteins. The amino acid sequence of IGF-I is similar to that of insulin and insulin and IGF-I share common downstream cellular signalling pathways [6, 78]. Molecularly, IGF-I stimulates the growth of cells by binding to IGF-I receptors, which activates

a protein tyrosine phosphorylation signal cascade, which is also similar to the action of insulin when stimulating the insulin receptor at high levels [78]. Specifically, IGF-I promotes cellular growth, proliferation, and inhibition of apoptosis via the activation of phosphoinositide 3-kinase pathways-Akt-mammalian target of rapamycin (PI3K-Akt-mTOR) and gene transcription and regulation via the mitogen activated protein kinase pathway (**Figure 2.4**) [7, 301]. By IGF-I stimulating IGF-I receptor signalling, which promotes cellular proliferation, this increases the number of cells and potential of damaged cells to be replicated therefore increasing the probability of carcinogenesis [302]. **Figure 2.5** presents a schematic of high and low IGF-I environments over time, where greater cell division and thus the number of cells are produced in a high IGF-I environment, thereby creating a greater potential for carcinogenesis and replication of damaged cells.

IGF-I is primarily produced in the liver (~85%) under the influence of GH [6] but most tissues in the body can also produce IGF-I locally, thus being capable of paracrine and autocrine signalling. Binding proteins partially regulate the action of IGF-I, namely insulin-like growth factor binding proteins (IGFBP), in which most of circulating IGF-I is bound to IGFBP-3. Evidence has shown that higher levels of IGFBP-3 increase the binding capacity of IGF-I and decreases circulating IGF-I bioactivity with approximately 80%-90% of IGF-I being bound to IGFBP-3 [6, 302-304]. IGF-I action is further controlled by IGFBP proteases, which act to cleave IGF-I from binding proteins and make it bioavailable (**Figure 2.4**).

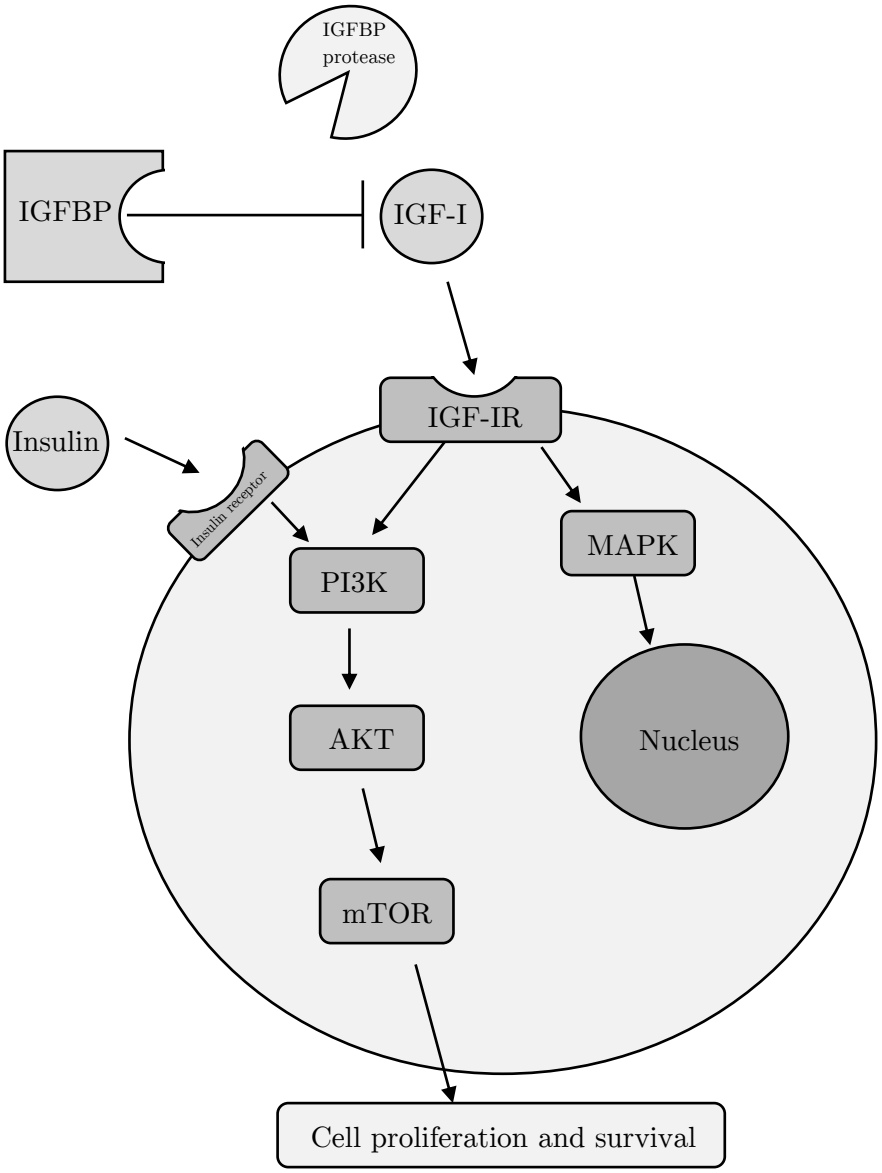


Figure 2.4 IGF-I and insulin cellular signalling leading to cell proliferation and survival.

Figure adapted from [6, 305]

Abbreviations: IGF-I, insulin-like growth factor-I, IGF-IR, insulin-like growth factor-I receptor; IGFBP, insulin-like growth factor binding proteins; MAPK, mitogen activated protein kinase; mTOR, mammalian target of rapamycin; PI3K, phosphoinositide 3-kinase.

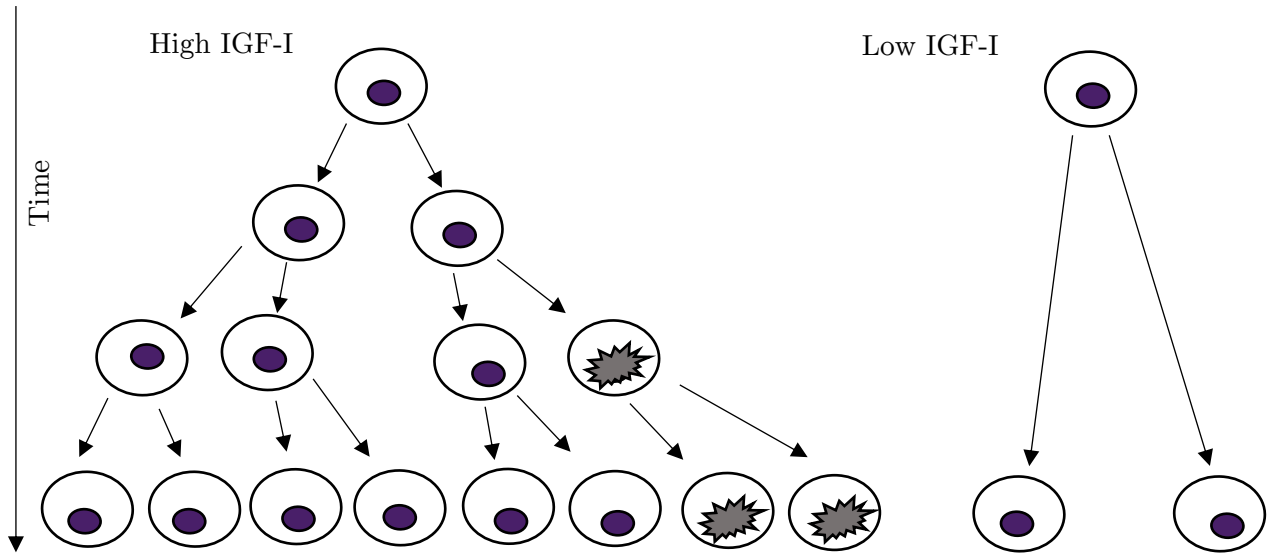


Figure 2.5 IGF-I influence on cellular signalling leading to greater potential of carcinogenesis with proliferation of damaged cells over time.

Figure adapted from Pollak *et al.,* [7]. Cells with grey spikes represent damaged cells. Figure exaggerates the magnitude of the difference between high and low IGF-I environments for illustrative purposes.
Abbreviation: IGF-I, insulin-like growth factor-I.

Circulating IGF-I concentrations increase at puberty and decrease as one ages due to the decrease in GH production [306]. Males are often shown to have higher circulating IGF-I concentrations than females, which may be attributable to sex hormone differences [307]. Individuals of taller height are also shown to have a higher circulating IGF-I concentrations, although being taller could be a result of higher IGF-I concentrations in early-life and during puberty [308]. BMI has been shown to have a U-shape association with IGF-I concentrations with those with very low BMIs ($<$20 kg/m^2) and very high ($\geq$35 kg/m^2) having lower concentrations than individuals with a BMI of 22.5-27.5 kg/m^2 [309, 310]. Other factors that have been shown to influence IGF-I concentrations include alcohol intake, where greater intake of alcohol is inversely associated with IGF-I concentrations [309, 310], ethnicity, where Hispanics have been shown to have lower IGF-I concentrations than white individuals [309] and MHT use, with women currently taking oestrogen MHT having lower concentrations of IGF-I [310, 311] possibly due to oestrogens influence in decreasing IGF-I synthesis in the liver [312].

Circulating IGF-I concentrations may also be modifiable by nutritional factors. Existing evidence suggests that energy restriction and prolonged fasting greatly decrease IGF-I concentrations [313, 314], with one meta-analysis of RCTs showing that reducing energy intake to less than 50% of required daily energy can decrease IGF-I concentrations by ~5 nmol/L [315]. Moreover, the intake of protein and dairy products have been shown to be positively associated with IGF-I concentrations. Multiple RCTs focused on increasing protein intake have found an increase in IGF-I concentrations in the higher protein arm [316, 317]. Similar results from RCTs have shown that greater intake of milk [284, 318] or dairy products [285, 286] elevates IGF-I concentrations. Adequate amino acid availability is required for hepatocytes to express the IGF-I gene, thus resulting in IGF-I synthesis [319]. Insufficient protein and energy intake may decrease the synthesis of IGF-I and greater clearance of IGF-I, thus resulting in decreased circulating IGF-I concentrations [313]. For other nutritional factors, and specific foods, the evidence for modifying IGF-I concentrations remains unclear. To date, no current studies have formally assessed circulating IGF-I concentrations as a mediator for higher cancer risk between dietary factors and IGF-I related cancers.

2.4.1 Insulin-like growth factor-I and colorectal cancer risk

Higher pre-diagnostic circulating IGF-I concentrations have been shown to be positively associated with colorectal cancer risk in nested-case control studies [9], prospective cohort analysis [12], and MR analysis [12] with estimates suggesting a 11% higher risk per standard deviation (SD) increase (~5 nmol/L) in IGF-I concentrations. IGF-I receptors are present in colorectal epithelial cells and overexpression of the IGF-I receptor in colorectal neoplasms is frequently observed [320].

In vitro studies using human colon cancer cells have shown that the blockage of IGF-I receptors by monoclonal antibodies inhibits cell proliferation [321]. However, agents tested in clinical trials to target the IGF system have not been successful in treating patients with colorectal cancer [322]. Another factor that may influence colorectal cancer carcinogenesis through the IGF pathway is insulin. Preclinical studies with individuals who had colorectal

adenomas have shown that insulin not only promotes the growth and survival of colorectal cancer cells directly, but also can promote IGF-I synthesis as well as increase IGF-I bioavailability by inhibiting IGFBP production [323, 324]. Studies have also observed that individuals with obesity-related hyperinsulinemia have higher levels of free IGF-I (unbound IGF-I, which is more bioavailable to receptors) than individuals who are nonobese [325, 326]. Despite the potential implications of insulin-IGF axis in carcinogenesis, how exactly insulin and IGF-I interact to promote tumour formation and how this pathway may be modifiable by lifestyle changes remains unclear [326].

2.4.2 Insulin-like growth factor-I and breast cancer risk

Circulating IGF-I concentrations have been positively associated in observational studies [10, 13] as well as in an MR analysis [13] with premenopausal and postmenopausal breast cancer risk. A recent observational study of 206,000 women showed that a 5 nmol/L increase in circulating IGF-I concentrations was associated with a 13% and 10% higher risk of premenopausal and postmenopausal breast cancer, respectively [13]. Early evidence from case-control studies suggested that IGF-I concentrations may only be associated with premenopausal but not postmenopausal breast cancer [327], but recent evidence from prospective and MR studies has clarified that IGF-I concentrations are associated with breast cancer regardless of menopausal status [10, 13]. However, the association between IGF-I concentrations and breast cancer risk may be exclusively in ER-positive tumours but not ER-negative tumours [10, 13]. In cellular studies, IGF-I and oestrogens have been shown to crosstalk [312, 328] and oestrogen might increase the presence of IGF receptors in breast cancer cells [329]. Moreover, IGF-I may inhibit hepatic synthesis of SHBG, which may make oestradiol more bioavailable. Despite IGFBPs regulating the bioavailability of IGF-I, and potentially having a direct inhibitory effect on cell growth [330], both observational and MR analyses have suggested IGFBP-3 is not associated with breast cancer risk [10, 13]. IGF-I stimulates breast epithelial cell growth and proliferation, resulting in greater potential of carcinogenesis [331] and cellular evidence has shown that a greater number of IGF-I receptors are present on breast cancer cells [332]. Despite IGF

pathways being implicated in breast carcinogenesis and tumour growth, phase II and III RCTs targeting the IGF-I receptor via monoclonal antibodies have not shown any clinical benefit in treating breast cancer [333-335].

2.4.3 Insulin-like growth factor-I and prostate cancer risk

In observational [11, 14] and MR studies [14, 336], circulating IGF-I concentrations have been associated with higher risk of total and aggressive prostate cancer and prostate cancer mortality; a 5 nmol/L increase in IGF-I concentrations was associated with a 9% higher risk of total and aggressive prostate cancer, and 15% higher risk of prostate cancer death. Estimates from MR studies have suggested that one SD increase in circulating IGF-I concentrations (5.4 nmol/L) is associated with a 45% higher risk of prostate cancer and a 2.1-fold higher risk of early onset prostate cancer (≤55 years of age) [336]. *In vitro* studies show an increase in IGF-I receptor expression in prostate tumour tissue [337] with evidence suggesting that expression of the IGF-I receptor elevates the risk of the tumour progressing to metastasize [338]. However, similar to colorectal and breast cancer, phase II clinical trials targeting the IGF-I receptor via monoclonal antibodies have generally shown to not be effective in treating prostate cancer greater than already established treatments such as androgen deprivation therapy [339, 340].

The positive association between dairy intake and prostate cancer risk is hypothesised to be mediated by higher IGF-I concentrations [287], and this may be due to the protein content in dairy products [5]. Evidence in rodent studies has suggested that the amino acid composition of dietary protein may be involved in prostate growth, possibly by IGF-I concentrations and activation of the mammalian target of rapamycin (mTOR) pathway, which ultimately increases proliferation and decreases autophagy [341]. However, these mechanisms are not well characterised and further research understanding the potential mechanisms of how dairy products may influence prostate cancer risk is needed.

2.5 Conclusion

Rates of cancers of the colorectum, breast, and prostate vary across the world suggesting potential environmental factors may be associated with developing these cancers. Circulating IGF-I concentrations have been implicated in the carcinogenesis of the colorectum, breast, and prostate, and IGF-I concentrations may be modifiable through lifestyle factors, such as diet.

Determining causal modifiable risk factors for cancer holds immense public health importance. Research is needed to better understand modifiable determinants of circulating IGF-I, with dietary intake being a plausible candidate for modifying IGF-I concentrations. Further understanding of dietary determinants of common cancers using repeated measures of diet is also needed to reduce measurement error and refine estimates to assess potential diet – cancer associations. Furthermore, nutrients and foods are consumed in combination and intercorrelations between foods may complicate identifying a true association; thus, identifying dietary patterns associated with cancer risk may offer insight into dietary components associated with carcinogenesis.

This thesis will explore dietary determinants of circulating IGF-I concentrations for both food groups and macronutrients, with a particular focus on protein intake (Chapter 4 & 5). Moreover, this thesis will explore how protein sources associated with IGF-I concentrations, namely dairy protein sources, are related to the risks of colorectal, breast, and prostate cancer and explore if IGF-I concentrations mediate these associations (Chapter 6). This thesis will also explore how intakes of carbohydrate sources and types, which influences insulin response, are related to the risk of colorectal cancer, and see if associations are modified by host SCFA genetic variants (Chapter 7). Finally, this thesis will also explore potential mediators for possible relationships between diet and cancer risk and look at dietary patterns, such as being a vegetarian or pescatarian, to further understand if these diets are associated with different risks of common cancer sites and further understand what may mediate these relationships (Chapter 8). **Figure 2.6** presents a schematic overview of the research questions and aims of this thesis.

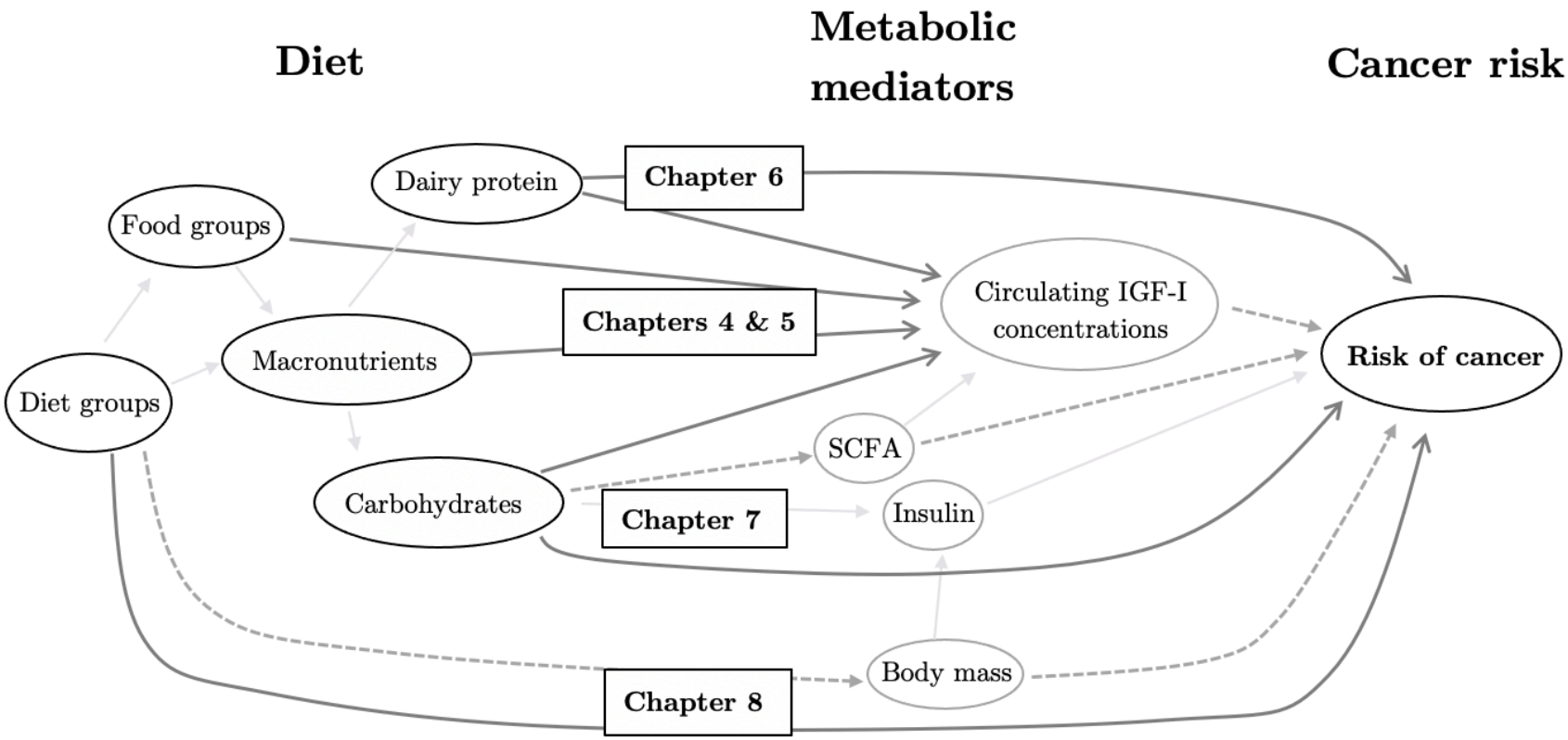


Figure 2.6 Schematic overview of the research questions explored in this thesis and potential hypotheses linking dietary intake with risk of cancer.

Thick arrows represent associations explored in this thesis; dashed arrows represent mediating or moderating factors explored in this thesis; light grey arrows represent hypothetical associations not formally explored in this thesis.

Abbreviations: IGF-I, insulin-like growth factor-I; SCFA, short-chain fatty acids.

Chapter 3

Data source: UK Biobank

3.1 UK Biobank

Data from the UK Biobank, a large prospective cohort of over half a million middle aged individuals from across the UK, were used throughout this thesis. This cohort was established with the aim of investigating various exposures in relation to different health outcomes to better understand and improve the prevention, diagnosis, and treatment of disease in middle age. The UK Biobank is one of the largest prospective cohorts in the world and the goal of recruiting half a million individuals was set to accrue enough cases of diseases over time, and thus obtain sufficient power required to quantify different risk factor-disease associations. This biomedical database contains in-depth information on lifestyle, genetic, biomarker, and health information, which provides the opportunity to conduct epidemiological research to better understand common life-threatening diseases [342].

3.1.1 Design and methods

Recruitment

Participants were recruited to the UK Biobank from March 2006 to September 2010 and were identified from individuals registered with the National Health Service (NHS) in the UK. From March 2006 to June 2006 initial piloting of recruitment to the UK Biobank occurred, which included 3,798 participants who completed a recruitment assessment visit in Stockport, UK. After this, the first assessment centre of the main study opened in April 2007 in Manchester, and other assessment centres were opened on a staggered basis over time. A total of 9,238,453 individuals who were aged 40-69 years and lived within 25 miles (40 kilometres) of an assessment centre were sent a postal invitation to participate in the study and attend one of the 22 assessment centres across England, Wales, and Scotland. A total of 503,317 participants (5.45% response rate) consented to take part in the UK Biobank study and attended an assessment centre over the recruitment period [343].

Data collection at recruitment

Information on participants' lifestyle and other potentially health-related factors were collected at the recruitment visit via a self-administered computer touchscreen questionnaire. The touchscreen questionnaire was utilised to optimise the accuracy and completeness of data as well as to allow for direct data entry, monitoring, and automated coding of responses compared to conventional paper questionnaires. Questions included information on lifestyle such as smoking (never, previous, and current as well as questions on frequency and duration), alcohol consumption (specific consumption of beer, wine, and spirits), and physical activity (through a validated survey instrument [344] as well as assessing sedentary behaviour). The touchscreen questionnaire also included questions on socioeconomic information such as if they owned a house, employment status, how many vehicles they owned, education obtained, total household income, questions on ethnic background, and health history such as diagnosis of disease, or whether they had specific operations such as an oophorectomy or hysterectomy. For women, reproductive factors such as age at menarche, menopause status, number of children birthed were asked in the touchscreen questionnaire. Finally, questions also asked about the use of medications and vitamins, which included if women took MHT or oral contraceptives.

Information that was not easily collected by questions on the touchscreen questionnaire was asked using a subsequent computer-assisted personal interview with a trained member of staff. This included questions that required further details such as the medication(s) they took, family history of disease, and birth information. Before these interviews, an aide was provided to participants to prompt recollection of the types of information they previously indicated on the touchscreen questionnaire. Questions included in this interview were based on the responses of participants in the touchscreen questionnaire. For example, if the participant indicated they had a medical condition in the touchscreen questionnaire, this was prompted further in the personal interview to ask about specific follow-up questions related to the condition.

Anthropometric measurements at recruitment

Trained professionals took anthropometric measurements at the assessment centre during the recruitment visit. Weight and bio-impedance were measured using the Tanita BC-418MA body composition analyser, which provided direct data entry into the assessment centre software. Bioimpedance was not measured if participants were pregnant, using a pacemaker, or were wheelchair-bound to name a few. When participants weight and bio-impedance were measured, they were instructed by the study staff to remove their shoes and heavy outer clothing before being weighed. Standing and sitting height were measured using a Seca 240 cm height measure and hip and waist circumference measurements were taken with a Seca 200 cm tape measure. Height measurements were not taken if participants were unable to bear weight, were wheel-chair bound, or were suffering from neck issues, whereas hip and waist circumferences were not measured if a participant was pregnant, wheel-chair bound, or had a colostomy bag. Height, hip, and waist circumference measurements were manually entered by trained professionals after measurements were taken. Further details regarding anthropometric measurements in the UK Biobank have been described elsewhere [345].

Blood collection at recruitment

A total of 40-50 mL of blood was taken from participants during the recruitment assessment visit using a vacutainer system. A nurse or a phlebotomist collected all blood samples for all participants, apart from a small proportion (0.3%) who declined, were unable to provide a blood sample, or for technical or health reasons the attempt at drawing blood was unsuccessful. Blood was drawn into ethylenediaminetetraacetic acid vacutainers, which were used for plasma and haematological assays, serum vacutainers to measure serum samples, and acid citrate dextrose was used for functional assays.

Upon receiving blood samples from participants at the assessment centre, these were shipped to the central processing laboratory in 4 °C temperature-controlled boxes. Haematological assays were conducted typically within 24-hours of the blood draw at the central

laboratory, whereas all other blood samples were frozen at -80 °C or in a back-up archive where blood samples were stored in liquid nitrogen at -196 °C.

Quality control procedures during the measurement of biomarkers at the central laboratory determined that some serum blood samples collected were inadvertently diluted during the creation of the aliquots from serum vacutainers [346]. This dilution is thought to have occurred due to mixing of participant sample with water due to seals that failed to hold a system vacuum in the automated liquid handling system. Despite this, the vast majority (92%) of assay results were found not to be materially affected by this dilution with an estimated concentration of 1% lower than the unaffected samples [346, 347]. The dilution increased in magnitude with increasing aliquot number, with aliquot 1 being the least affected and aliquot 3 being more affected by dilution (~2% of participants samples); thus aliquot 1 was prioritised for serum blood measurements by the UK Biobank. The UK Biobank discarded all measurements from aliquot 4 due to being unable to characterise the dilution in this aliquot, although the dilution appeared most severe [346].

Genomic data

Participants were genotyped from blood samples collected at the recruitment visit by Affymetrix Research Services Laboratory located in Santa Clara, California. A subset of 49,950 individuals were genotyped at 807,411 genetic markers using the Applied Biosystems UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) Axiom Array produced by Affymetrix. This subset of individuals were selected based on lung function, smoking behaviour reported by participants, and reported European ancestry to further study the genetics of lung health and disease [348]. Otherwise, 438,427 participants were genotyped using the Applied Biosystems UK Biobank Axiom Array, which included 825,927 markers. This array shared 95% of the markers with UK BiLEVE array. The genetic marker content of the UK Biobank Axiom array was chosen to capture genome-wide genetic variations such as SNPs as well as insertions and deletions. Many markers were included in the array because of previous or possible associations with disease. The array also included rare markers, which included minor allele frequencies of $<1\%$ as well as

markers that provided good genome-wide coverage for imputation in European populations. Over 90 million variants were imputed from markers using the Haplotype Reference Consortium and UK10K + 1000 Genomes reference panels. Genotyped data were quality controlled for poor quality markers and poor-quality samples from high missing rate and heterozygosity. In total, genotyped data from 488,377 individuals in the UK Biobank were available. Further details regarding genotyping have been described elsewhere [349, 350].

Data storage and anonymity

All participants included in the UK Biobank remain anonymous to researchers as UK Biobank does not provide identifiable data such as NHS number, name, exact date of birth, or postal or email address of individuals. Data are provided to researchers after the approval of applications to the UK Biobank to access the data and are reverse-anonymised.

3.1.2 Ethical approval

At recruitment, all participants provided informed consent to participate in the study as well as to be followed through their health records. Participants can withdraw from the UK Biobank at any time, in which their data are instructed to be removed from all researchers' datasets. UK Biobank also has an Ethics and Governance council and Ethics Advisory Committee. The Ethics and Governance Council acts as an independent guardian of the UK Biobank Ethics and Governance Framework, monitors ethical conformity, and advises on the interests of research participants. The Ethics Advisory Committee is a committee of members of the UK Biobank board, which provides advice on ethical issues that arise during maintenance, development, and use of UK Biobank data, including reviewing and advising on policies that have ethical relevance. Ethical approval was obtained by North West Multi-Centre Research Ethics Committee and is renewed every five years with the most recent renewal being obtained in June 2021 (21/NW/0157).

3.1.3 Follow-up

Follow-up of participants from the UK Biobank is ongoing primarily through linkage to health records. Some participants have also participated in additional phenotyping, including wearable physical activity monitoring, magnetic resonance imaging of the brain and body, 24-hour dietary assessments, and whole-body DEXA scan, as well as repeating the whole recruitment visit for a second time.

Repeat assessment

Following the completion of the recruitment visit at one of the 22 assessment centres, 103,514 participants who lived within a 35 km radius of the assessment centre in Stockport, UK were invited to complete a repeat recruitment assessment visit from August 2012 to June 2013, of which, 20,3465 participants attended and completed the recruitment assessment visit for a second time (21% response rate). This repeat assessment was conducted to provide another measurement of participants lifestyle, health, and biomarkers over time as well as to correct exposures for regression dilution bias [351].

Disease incidence and mortality

At recruitment, participants consented to have their past and future medical record data available allowing for follow-up through record linkage. Information regarding the incidence and mortality of disease after recruitment is obtained by the UK Biobank study via NHS numbers in England and Wales and Community Health numbers in Scotland. From this, participants can be identified and followed for specific health outcomes from such databases, including hospital admission data [352], cancer registries [353], general practice records [354], and death registries [355].

In this thesis, both cancer registry data and hospital inpatient data were utilised to follow participants for incident cancer diagnoses. The World Health Organization's International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) codes were used to determine whether a participant had an incident cancer diagnosis with codes C00-

97 reflecting any cancer diagnosis (excluding participants with a non-melanoma skin cancer diagnosis (C44)), colorectal cancer (C18-C20), breast cancer (C50), or prostate cancer (C61) [356].

3.1.4 Dietary assessments

Dietary intake was assessed through two different modalities in the UK Biobank, a short dietary assessment completed by all participants and a 24-hour dietary assessment completed by a subsample of participants.

Touchscreen questionnaire

At the recruitment visit, all participants completed 29 questions about dietary intake of commonly consumed foods and 18 questions on alcohol intake using the touchscreen questionnaire. Dietary questions included on the touchscreen questionnaire were designed to characterise participants based on commonly eaten food groups consumed among the British population. These dietary questions asked about frequency of consumption of select foods mostly over the past year (12 months), as well as avoidance of specific foods, last time participants consumed meat, and if they have changed their diet over the past 5 years. For the consumption of vegetables and fresh fruit, participants were asked to enter the number of heaped tablespoons (for both cooked vegetables and salad/raw vegetables) or pieces of fruit (with examples as "one apple, one banana, 10 grapes" constituting one piece) consumed per day. Participants also had the option to select 'less than one', 'do not know' or 'prefer not to answer' for questions on cooked vegetables, raw vegetables, or fresh or dried fruit intake. For beef, lamb, pork, processed meat, poultry, oily fish, non-oily fish, and cheese, no portion size was provided in the question, instead, participants were asked how often each item was consumed per week with possible choices being 'never', 'less than once a week', '2-4 times a week', '5-6 times a week', 'once or more daily', 'do not know', or 'prefer not to answer'. For bread intake, participants were asked to enter the number of slices of bread they consumed per week as well as the type of bread they most consumed. Similarly, participants were asked to report their intake of cereal by entering in

the average number of bowls they consumed per week and report the type of cereal they consumed most.

24-hour dietary assessment – Oxford WebQ

Dietary intake was further assessed in greater detail in a subgroup of participants using a web-based 24-hour dietary assessment tool, namely, the Oxford WebQ [357]. The Oxford WebQ 24-hour dietary assessment questionnaire asks participants to recall the quantity of consumption of 206 types of foods and 32 types of drinks they consumed during the previous 24 hours [358]. Validation of the Oxford WebQ was conducted in a sample of 160 men and women with biomarker measurements for protein, potassium, and sugar, and an accelerometer for energy. The results showed that the Oxford WebQ questionnaire performs well in approximating true dietary intake, with a greater correlation with additional dietary assessments included [359]. The web-based 24-hour dietary assessment was also found to provide similar estimates of energy and nutrient intake compared to an interviewer administered 24-hour dietary recall [358].

The last 70,724 participants recruited to the UK Biobank from April 2009-September 2010 completed the 24-hour dietary assessment at the assessment centre as part of their recruitment visit. Moreover, every 3-4 months for a total of four times between February 2011 and April 2012 (first follow-up, February 2011 to April 2011; second follow-up, June 2011 to September 2011; third follow-up, October 2011 to December 2011; and fourth follow-up April 2012 to June 2012), a link to the 24-hour dietary assessment was emailed to all participants who had provided a valid email address at recruitment (**Figure 3.1**). Invitation emails were sent on variable days of the week and participants in follow-ups 1 and 2 were given 3 days to complete the 24-hour dietary assessment; whereas for follow-ups 3 and 4, 14 days were provided to click the link to complete the dietary assessment before it expired to increase the participation rate. During the follow-up period, the response rate in each follow-up 24-hour dietary assessment varied between 26% and 33% of participants with valid emails responding.

At the start of the 24-hour dietary assessment, participants were provided with instructions on how to complete the questionnaire. Participants were asked to select the type

and amount of food they had consumed in the previous day and questions were grouped by food groups (i.e., beverages, breads, cereals, meat and fish, fruits and vegetables). Participants were also asked at the start of the questionnaire if what they consumed was their 'typical diet' and if they reported it was not, they were asked to provide reason(s) for not eating typically on that day with options being illness, fasting, travel, or other reasons not listed. At completion, participants were provided with a list of all food and beverage items they reported consuming on the previous day and were allowed to make amendments as needed. The Oxford WebQ took an average of 10-15 minutes to complete.

In all analyses in this thesis that utilised the 24-hour dietary assessment to measure diet, individual 24-hour dietary assessments with extreme energy intakes (<3,347 (800 kcal) or >17,573 kJ (4,200 kcal) for men, and <2,092 (500 kcal) or >14,644 kJ (3,500 kcal) for women) [360] were removed, as were 24-hour dietary assessments where participants reported they were ill or fasting on the day of the dietary assessment. Moreover, a minimum of two (or four for the IGF-I concentration analysis) 24-hour dietary assessments were required in order for participants to be included in analyses to reduce random measurement error between day-to-day variations in dietary intake [360].

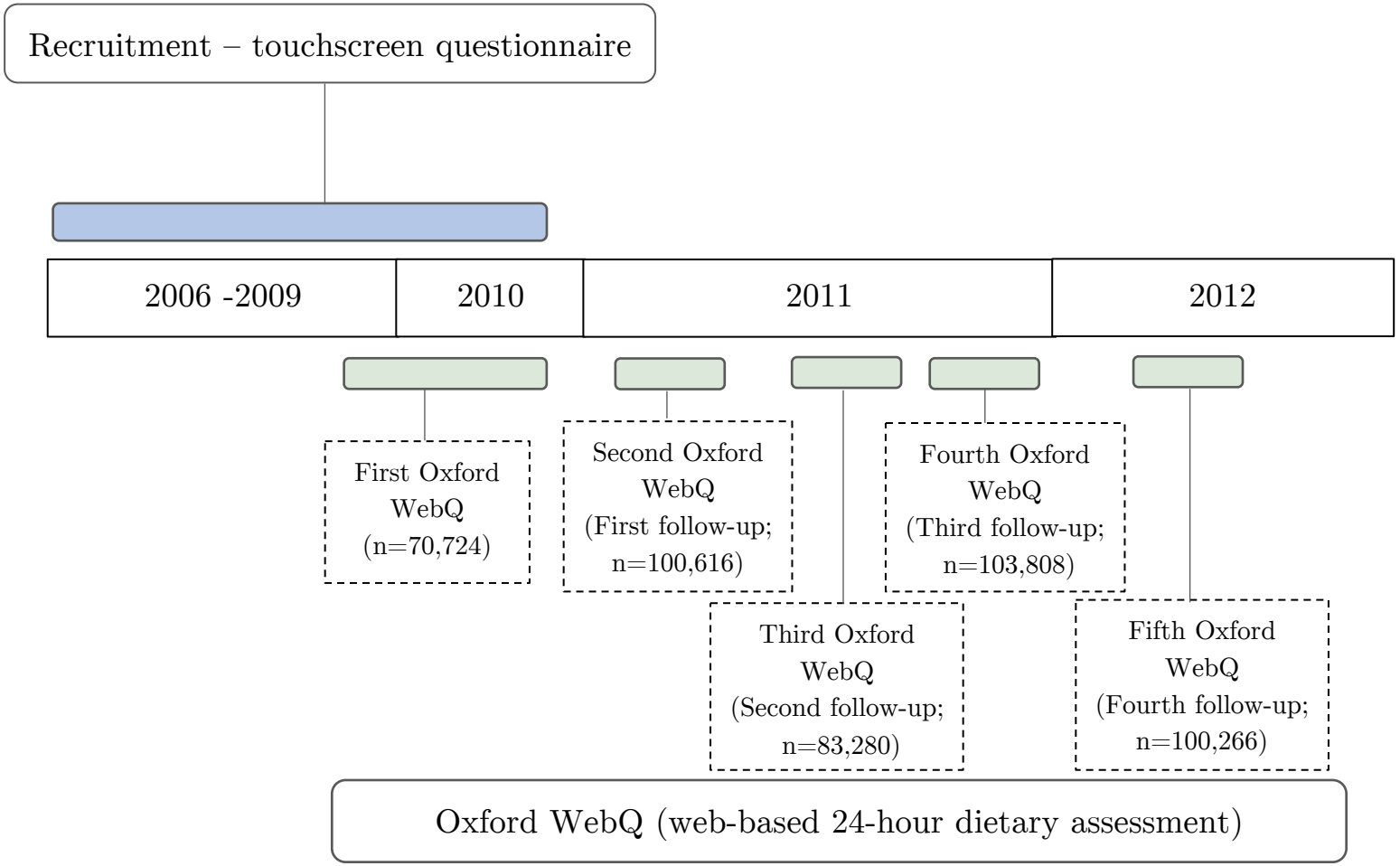


Figure 3.1 Dietary assessments (touchscreen questionnaire and web-based 24-hour dietary assessment) completed by UK Biobank participants over time.

3.1.5 Nutrient calculations from the 24-hour dietary assessment

Macronutrients and their sources, fibre, and energy intake were estimated in each completed web-based 24-hour dietary assessment from each participants' reported food and beverage intake [361, 362]. Specifically, food compositions from the UK Nutrient Databank (2013, survey year 6) were used and 681 food codes were linked to food and drinks in the Oxford WebQ to determine nutrient and energy intakes. The UK Nutrient Databank is commissioned by Public Health England as part of the National Diet and Nutrition Survey and contains widespread information on the nutrient content of foods commonly consumed by individuals in the UK [363]. Each food and beverage items in the 24-hour dietary assessment was assigned one or multiple food codes from the UK Nutrient Databank that reflected the food or beverage item and were classified into food groups [362]. The matching of food codes to food and beverage items in the Oxford WebQ 24-hour dietary assessments was undertaken by researchers in the CEU, University of Oxford and Department of Public Health & Primary Care, University of Cambridge and has been described in greater detail elsewhere [361].

The quantity of each food and beverage consumed was calculated by multiplying the assigned portion size by the reported consumed amount of the food or beverage in each 24-hour dietary assessment. Portion sizes were estimated based on UK standard portion sizes as well as how questions were asked in the 24-hour dietary assessment [361, 364].

The following intakes of macronutrients were determined from all foods consumed in each 24-hour dietary assessment: carbohydrates, total sugars, free sugars, starch, fat, fat from plant sources, fat from animal sources, protein, protein from plant sources, protein from animal sources, fibre, and total energy [361]. Other exposures, namely starch from wholegrains, starch from refined grains, non-free sugars, protein and fat from dairy sources, milk, yogurt, and cheese, protein and fat from animals from non-dairy sources, energy not from alcohol, and alcohol were calculated from specific food groups/items in each 24-hour dietary assessments [362] and are described below.

Non-free sugars

Non-free sugars were calculated by subtracting free sugars from the total sugars calculated for each 24-hour dietary assessment.

Starch from wholegrains

Starch from wholegrains was calculated from the grams of starch from wholegrain bread, wholegrain pasta or brown rice, bran cereal (e.g. bran flakes), biscuit cereal (e.g. Weetabix, Shredded Wheat), oat cereal (with or without sugar) and muesli.

Starch from refined grains

Starch from refined grain was calculated from the grams of starch from white bread, other breads (e.g., naan garlic bread), other cereal, white pasta or rice, pizza, samosa, pakora, grain dishes, savoury snacks, biscuits, and other desserts, cakes, and pastries, and savoury crackers.

Protein and fat from dairy sources

Protein from dairy products and fat from dairy products were calculated from the grams of protein and fat from all types of milk (whole, semi-skimmed, and skimmed), cream, butter (low fat and high fast), yogurt (full fat and low fat), cheese (high fat and medium fat), and dairy drinks.

Protein and fat from animal sources (not dairy)

Protein and fat from animal sources other than dairy were calculated for each participant by subtracting the total amount of grams or energy from animal protein and animal fat from the calculated dairy protein and dairy fat for each 24-hour dietary assessment, respectively.

Protein and fat from milk

Protein from milk and fat from milk were calculated from the reported intake of whole milk >3.6 grams of fat per 100 grams), semi-skimmed milk (1.0-3.6 grams of fat per 100 grams), and skimmed milk (<1.0 grams of fat per 100 grams) in grams in the 24-hour dietary assessment.

Protein and fat from yogurt

Protein from yogurt and fat from yogurt were calculated from the reported intake of full fat yogurt and non-fat yogurt in grams in the 24-hour dietary assessment.

Protein and fat from cheese

Protein from cheese and fat from cheese were calculated from the reported intake of medium or low-fat cheese (≤17.5 grams per 100 grams) and high fat cheese (>17.5 grams per 100 grams) in grams in the 24-hour dietary assessment.

Total energy and energy not from alcohol

For total energy not from alcohol, energy was calculated for each food item excluding alcohol intake.

Alcohol

Alcohol intake was determined from the grams of alcohol in a standard drink asked in the 24-hour dietary assessment. Total grams of alcohol were calculated from wine, spirits, and beer reported in the 24-hour dietary assessment.

Table 3.1 provides a summary of how each type and source of macronutrients utilised in this thesis were derived.

Table 3.1 Summary of the calculations for each macronutrient types and sources from the 24-hour dietary assessments

Macronutrient types and sources	Calculation
Total carbohydrates	UK Nutrient Databank – food compositions
Total sugar	UK Nutrient Databank – food compositions
Total free sugar	Determined from the proportion of sugars that were from free sources (all monosaccharides and disaccharides added to foods, plus sugars naturally present in honey, syrups and unsweetened fruit juices)
Total non-free sugar	Calculated as the difference between total sugars and free sugars
Total starch	UK Nutrient Databank – food compositions
Starch from wholegrains	Starch from wholegrain was calculated from the grams of starch from wholegrain bread, wholegrain pasta or brown rice, bran cereal (e.g., bran flakes), biscuit cereal (e.g., Weetabix, Shredded Wheat), oat cereal (with or without sugar) and muesli.
Starch from refined grains	Starch from refined grains was calculated from the grams of starch from white bread, other breads (e.g., naan garlic bread), other cereal, white pasta or rice, pizza, samosa, pakora, grain dishes, savoury snacks, biscuits, and other desserts, cakes, and pastries, and savoury crackers.
Fibre (non-starch polysaccharides)	UK Nutrient Databank – food compositions
Total protein	UK Nutrient Databank – food compositions
Protein from plant sources	Total protein intake from plant foods derived from UK Nutrient Databank – food compositions
Protein from animal sources	Total protein intake from animal sources from UK Nutrient Databank – food compositions
Protein from dairy products	Protein from dairy products was calculated from the grams of protein from all types of milk, cream, butter, yogurt, cheese and dairy drinks.
Protein from milk	Protein from milk from milk was calculated from the reported intake of whole milk, semi-skimmed milk, and skimmed milk
Protein from yogurt	Protein from full fat yogurt and non-fat yogurt
Protein from cheese	Protein from medium or low-fat cheese and high fat cheese
Protein from animal source (not dairy)	Determined as the difference between protein from animal sources and protein from dairy products
Total fat	UK Nutrient Databank – food compositions
Fat from plant sources	Total fat intake from plant foods derived from UK Nutrient Databank – food compositions
Fat from animal sources	Total fat intake from animal sources from UK Nutrient Databank – food compositions
Fat from dairy products	Fat from dairy products was calculated from the grams of protein from all types of milk, cream, butter, yogurt, cheese and dairy drinks.
Fat from milk	Fat from milk from milk was calculated from the reported intake of whole milk, semi-skimmed milk, and skimmed milk
Fat from yogurt	Fat from full fat yogurt and non-fat yogurt
Fat from cheese	Fat from medium or low-fat cheese and high fat cheese
Fat from animal source (not dairy)	Determined as the difference between fat from animal sources and fat from dairy products
Total alcohol intake	UK Nutrient Databank – food compositions
Total energy	UK Nutrient Databank – food compositions
Total energy not from alcohol	Determined from the difference in total energy minus the energy from alcohol

Data source: [361, 362]

Calculated grams of total carbohydrates and carbohydrate types and sources and total protein and protein sources were multiplied by 16.7 to obtain the kilojoule (kJ) of energy from the specific macronutrient [365]. Total fat and fat sources were multiplied by 37.7 to determine the kJ of energy from fat [365]. Once this was determined from all 24-hour dietary assessments, each macronutrient was divided by the total amount of energy intake for the specific WebQ 24-hour dietary assessment for each participant to determine the percentage of energy from the macronutrient for each individual 24-hour dietary assessment completed.

Chapter 4

Food group intakes and circulating IGF-I concentrations

Chapter 4 summary

Intakes of some foods have been associated with increased circulating IGF-I concentrations; however, evidence remains inconclusive. This chapter examines the cross-sectional associations between food group intakes and circulating IGF-I concentrations in 438,453 UK Biobank participants. At recruitment, the UK Biobank participants reported their intake of commonly consumed foods and blood samples were collected in which serum IGF-I concentrations were measured. From these questions, intakes of total vegetables, fresh fruit, red meat, processed meat, poultry, oily fish, non-oily fish, and cheese were estimated. Compared to never consumers, participants who reported consuming oily fish or non-oily fish ≥2 times/week had 1.25 nmol/L (95% confidence interval:1.19-1.31) and 1.16 nmol/L (1.08-1.24) higher IGF-I concentrations, respectively. Participants who reported consuming poultry ≥2 times/week had 0.87 nmol/L (0.80-0.94) higher IGF-I concentrations than those who reported never consuming poultry. There were no strong associations between other food groups and IGF-I concentrations. Fish and poultry are protein rich foods, however other protein rich foods, such as red meat, were not associated with IGF-I concentrations. Further research is needed to understand the mechanisms which might explain these associations.

4.1 Introduction

IGF-I is a polypeptide hormone, primarily produced in the liver, which stimulates cell growth and proliferation [6]. As described in Chapter 2, prospective and genetic studies have indicated that higher circulating IGF-I concentrations are associated with several cancer outcomes including, higher risks of colorectal, breast, and prostate cancer [10-14, 327].

4.1.1 Previous studies

While there is substantial evidence assessing dietary factors associated with circulating IGF-I concentrations [3-5, 366, 367], to date the evidence seems consistent only for energy restriction, and intakes of protein and dairy products. Restricting energy intake for prolonged periods may decrease circulating IGF-I concentrations [313-315] whereas previous cross-sectional studies [3-5, 366] and RCTs [285, 286, 368, 369], have reported some evidence for a positive association between intake of dairy products and IGF-I concentrations, which has been proposed to be due to the protein content in dairy products [5]. However, for some dairy products, such as cheese, evidence has been inconsistent and in some smaller studies, intake has not associated with IGF-I concentrations [3, 370]. Moreover, it is not well understood whether or how intakes of other protein-rich foods, such as red meat, poultry and fish, and other food groups, such as fruits and vegetables, relate to circulating IGF-I concentrations, with previous studies being limited by relatively small sample sizes [3-5, 366].

4.2 Aim

Using the UK Biobank, which collected dietary information and measured serum IGF-I concentrations from blood collected at the recruitment visit, this chapter investigates the cross-sectional associations of selected food groups with circulating IGF-I concentrations.

4.3 Methods

4.3.1 Study sample & exclusions

Details of the UK Biobank study are described in Chapter 3. At the time of this analysis, 824 participants had withdrawn their informed consent from the study and were excluded. Participants were also excluded if they had a prevalent cancer at recruitment recorded by a cancer registry (excluding non-melanoma skin cancer; N=27,174), were taking medications, which may alter IGF-I concentrations, such as GH (N=4,077; Ancillary Table 4.1) or did not have a measured value for IGF-I at recruitment (N=32,789). In total, a maximum of 438,453 participants were included in this analysis.

4.3.2 Dietary assessment

In this analysis, for each of the following foods and food groups, participants were categorised into four categories based on their reported frequency of intake of specific food groups: vegetables (raw and cooked), fresh fruit, red meat (unprocessed beef, pork, and lamb/mutton), processed meat (e.g. bacon, ham, sausages), poultry, oily fish (e.g. sardines, salmon, mackerel, herring), non-oily fish (e.g. cod, tinned tuna, haddock), and cheese. Cut-offs for categories were chosen based on the data distribution of intakes for each food group. For the consumption of vegetables and fresh fruit, participants were asked to enter the number of heaped tablespoons (for both cooked vegetables and salad/raw vegetables) or pieces of fruit (with examples as "one apple, one banana, 10 grapes" constituting one piece) consumed per day. Participants also had the option to select 'less than one', 'do not know' or 'prefer not to answer' for questions on cooked vegetables, raw vegetables, or fresh fruit intake. Based on this, participants were categorised for the intake of total vegetables (cooked and raw) as 2.0 servings per day, 2.0-2.99 servings per day, 3.0-3.99 servings per day, and ≥4.0 servings per day with two heaped tablespoons of vegetables counting as a serving. For fresh fruit intake, participants were as <1 servings per day, 1.0-1.99 servings per day, 2.0-2.99 servings per day, and ≥3.0 servings per day with one piece of fresh fruit counting as a serving. For oily fish, non-oily fish, processed meats, poultry, beef, lamb, pork and cheese, no portion size was given in the question, instead,

participants were asked how often each item was consumed with possible choices being: 'never, 'less than once a week', '2-4 times a week', '5-6 times a week', 'once or more daily', 'do not know', or 'prefer not to answer'. These frequencies were then converted using the following coding: 'Never' = 0, 'Less than once a week' = 0.5, 'Once a week' = 1, '2-4 times a week' = 3, '5-6 times a week' = 5.5, 'Once or more daily' = 7. From these responses, participants were categorised into intakes for each food group based on their reported consumption. Specifically for total red meat intake, participants were categorised into one of the following categories: <once/week, 1-1.99 times per week, 2-2.99 times per week, and ≥3.0 times per week. For the intake of processed meat, poultry, oily fish, and non-oily fish, participants were categorised into one of four categories: never, <once/week, once a week, and ≥2 times per week. For cheese intake participants were categorised into one of four categories: <once/week, once a week, 2-4 times a week and ≥5 times per week. If a participant selected 'do not know' or 'prefer not to answer' they were excluded from the analysis for that specific food group.

4.3.3 Laboratory analysis

Non-fasted blood samples were collected from nearly all participants (99.7%) at the recruitment visit and were transported at 4°C to the central laboratory for cryopreservation and subsequent biochemical measurements. Further details of blood retrieval from UK Biobank participants are described in Chapter 3. Serum concentrations of IGF-I were measured using the DiaSorin Ltd. LIAISON® XL chemiluminescent immunoassay. The coefficient of variation for circulating IGF-I concentrations between batches was 1.2%. Details about assay methods and quality control procedures for serum blood measurements are available online [371] and in Chapter 3.

4.3.4 Repeat assessment

As described in Chapter 3, approximately 20,000 participants who lived within a 35 km radius participated in a repeat recruitment assessment at the UK Biobank centre in Stockport, an average of 4 years after their initial visit. At this repeat baseline assessment, participants provided a second blood sample. From this follow-up visit, a total of 16,689 participants had a

valid IGF-I concentration measured with 15,419 participants having both (recruitment and follow-up) IGF-I measurements. The reproducibility between the baseline IGF-I measurement and the follow-up IGF-I measurement was determined to be good with an intraclass correlation coefficient of 0.76. Further information on the UK Biobank repeat visit can be found in Chapter 3 and on the UK Biobank website [372].

4.3.5 Statistical analysis

Circulating IGF-I concentrations were logarithmically transformed to minimise the impact of outliers as well to obtain the geometric mean concentrations of IGF-I in line with previous studies. The geometric mean concentration of IGF-I was obtained within each category of intake of food groups from linear regression models based on predictions of the model by using the margins, fixing covariates at their means for each category. To determine relative values, geometric means in the other categories were divided by the geometric mean in the lowest category.

In minimally adjusted linear regression models, adjustments were made for sex and age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years). Multivariable linear regression models were further adjusted for region of recruitment (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland), BMI (<20, 20-22.49, 22.5-24.9, 25-27.49, 27.5-29.9, 30-32.49, 32.5-34.9, ≥35 kg/m^2, and unknown/missing), height (eight sex-specific categories increasing by 5 cm, and unknown/missing), physical activity (low: 0-9.99, medium: 10-49.99, high: ≥50 metabolic equivalent of task (MET)-hours/week, and unknown/missing), Townsend deprivation index (quintiles from most deprived to least deprived, or unknown), education (completion of national exam at 16 years of age, completion of national exam at 17-18 years of age, college or university degree, or unknown/missing), smoking status (never, former, light smoker: <15 cigarettes/day, medium smoker: 15-29 cigarettes/day, heavy smoker: ≥30 cigarettes/day, or missing/unknown), alcohol consumption (non-drinkers, <1, 1-9.99, 10-19.99,

≥20 grams/day or unknown/missing), ethnicity (White, mixed race, Indian/Pakistani/Bangladeshi, Chinese/Asian, Black/Black British, other, or missing/unknown), diabetes (yes, no, unknown), and women specific covariates: MHT use (never, former, current, or unknown), oral contraceptive use (never, former, current, or unknown), and menopausal status (premenopausal, postmenopausal, or unknown). Further information on the classification of covariates can be found in the **Ancillary Methods A.1**. Due to the limited number of foods asked on the recruitment questionnaire, mutual adjustment of dietary variables was not undertaken in these analyses.

Subgroup and sensitivity analyses

Heterogeneity by sex was assessed using a likelihood ratio test (LRT) comparing the multivariable model to a model including an interaction term between the food group and sex. Sensitivity analyses were conducted in participants who had IGF-I concentrations measured at the reassessment visit (mean 4.3 years after recruitment).

All analyses were conducted using Stata version 17.0 (Stata Corp LP, College Station, TX) and figures were produced using "Jasper makes plots" package version 2-266 in R 4.1.0 [373]. P-values were two-sided and, with Bonferroni correction so that p-values$<$0.00625 (0.05/8 exposures) were considered statistically significant. As a result of the large sample size, most results were statistically significant even after correction for multiple testing. As such, only the largest percentage differences of IGF-I concentrations between the highest and lowest categories (~5% or greater difference) have been described in the text. All models were visually assessed to make sure residuals were normally distributed using Q-Q plots, and not heteroscedastic using residual-versus-fitted plots. No assumptions for linear regression were deemed invalid.

4.4 Results

4.4.1 Participant characteristics

Table 4.1 shows participant baseline characteristics by fifths of IGF-I concentrations. Those who had higher IGF-I concentrations were more likely to be men, to be younger and taller, to have lower BMI, and to report that they were never smokers. Circulating IGF-I concentrations were generally higher in Black individuals and lower in participants of South Asian ancestry (**Table 4.1**).

Table 4.1 Characteristics of participants in the UK Biobank by fifths of circulating IGF-I concentrations (N=438,453).

	Circulating IGF-I				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**
Number of participants	87,716	87,686	87,687	87,696	87,668
IGF-I concentration at recruitment, nmol/L	13.9 (2.2)	18.3 (0.9)	21.3 (0.8)	24.1 (0.9)	29.5 (4.0)
IGF-I concentration at follow-up, nmol/L	15.1 (3.6)	18.5 (3.3)	20.8 (3.4)	23.0 (3.6)	26.9 (5.0)
Sex - Male, N (%)	32,843 (37.4%)	38,985 (44.5%)	42,241 (48.2%)	44,267 (50.5%)	43,563 (49.7%)
Age, years	59.1 (7.1)	57.6 (7.6)	56.4 (7.9)	55.2 (8.2)	53.3 (8.4)
Body mass index, kg/m^2	28.6 (5.7)	27.5 (4.9)	27.2 (4.5)	27.0 (4.3)	26.7 (4.1)
Height, cm	166.3 (9.0)	167.9 (9.1)	168.9 (9.2)	169.6 (9.3)	170.0 (9.3)
Physical activity, N (%)					
Low	27,983 (31.9%)	25,238 (28.8%)	24,743 (28.2%)	24,279 (27.7%)	24,268 (27.7%)
Moderate	39,236 (44.7%)	41,613 (47.5%)	42,547 (48.5%)	43,413 (49.5%)	44,324 (50.6%)
High	16,609 (18.9%)	17,518 (20.0%)	17,427 (19.9%)	17,262 (19.7%)	16,355 (18.7%)
Townsend deprivation index, N (%)					
Q1 - Most affluent	15,780 (18.0%)	17,546 (20.0%)	17,959 (20.5%)	18,385 (21.0%)	18,457 (21.1%)
Q5 - Most deprived	20,242 (23.1%)	17,454 (19.9%)	16,698 (19.0%)	16,273 (18.6%)	16,277 (18.6%)
Education, N (%)					
National examination at age 16 years	15,029 (17.1%)	14,887 (17.0%)	14,650 (16.7%)	14,445 (16.5%)	14,337 (16.4%)
National examination at age 17-18 years	4,466 (5.1%)	4,671 (5.3%)	4,750 (5.4%)	4,909 (5.6%)	5,234 (6.0%)
College or University degree	46,604 (53.1%)	50,778 (57.9%)	53,306 (60.8%)	54,863 (62.6%)	56,767 (64.8%)
Smoking, N (%)					
Never	44,447 (50.7%)	46,438 (53.0%)	47,966 (54.7%)	49,401 (56.3%)	51,656 (58.9%)
Previous	32,797 (37.4%)	31,151 (35.5%)	30,139 (34.4%)	28,923 (33.0%)	26,912 (30.7%)
Light smoker: <15 cigarettes/day	2,835 (3.2%)	2,652 (3.0%)	2,638 (3.0%)	2,543 (2.9%)	2,687 (3.1%)
Medium smoker: 15-29 cigarettes/day	3,619 (4.1%)	3,353 (3.8%)	3,003 (3.4%)	2,910 (3.3%)	2,861 (3.3%)
Heavy smoker: ≥30 cigarettes/day	3,357 (3.8%)	3,563 (4.1%)	3,475 (4.0%)	3,496 (4.0%)	3,170 (3.6%)

Table 4.1 continued

	Circulating IGF-I				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**
Alcohol intake, N (%)					
<1 g/day	11,842 (13.6%)	9,684 (11.1%)	8,963 (10.3%)	8,921 (10.2%)	9,550 (11.0%)
1-9.99 g/day	25,849 (29.7%)	26,400 (30.3%)	27,061 (31.1%)	27,666 (31.7%)	30,219 (34.7%)
10-19.99 g/day	15,965 (18.4%)	18,608 (21.4%)	19,360 (22.2%)	20,318 (23.3%)	20,405 (23.4%)
≥20 g/day	23,845 (27.4%)	25,549 (29.3%)	25,420 (29.2%)	24,301 (27.9%)	20,830 (23.9%)
None drinkers	9,483 (10.9%)	6,885 (7.9%)	6,342 (7.3%)	5,983 (6.9%)	6,160 (7.1%)
Ethnicity, N (%)					
White	82,375 (93.9%)	82,788 (94.4%)	82,794 (94.4%)	82,664 (94.3%)	82,060 (93.6%)
Mixed Race	463 (0.5%)	485 (0.6%)	525 (0.6%)	514 (0.6%)	629 (0.7%)
South Asian (Indian/Pakistani)	1,952 (2.2%)	1,424 (1.6%)	1,336 (1.5%)	1,249 (1.4%)	1,124 (1.3%)
Chinese, Asian or other Asian	502 (0.6%)	556 (0.6%)	559 (0.6%)	669 (0.8%)	750 (0.9%)
Black or Black British	1,202 (1.4%)	1,196 (1.4%)	1,284 (1.5%)	1,451 (1.7%)	1,818 (2.1%)
Other	766 (0.9%)	777 (0.9%)	797 (0.9%)	773 (0.9%)	894 (1.0%)
Diabetes - Yes, N (%)	7,068 (8.1%)	4,227 (4.8%)	3,691 (4.2%)	3,431 (3.9%)	3,499 (4.0%)
Women-only covariates:					
Current MHT users, N (%)	7,014 (12.8%)	3,714 (7.6%)	3,060 (6.7%)	2,448 (5.6%)	2,144 (4.9%)
Current oral contraceptive pill users, N (%)	370 (0.7%)	522 (1.1%)	685 (1.5%)	947 (2.2%)	1,892 (4.3%)
Menopause status at recruitment, N (%)					
Premenopausal	4,426 (8.1%)	6,975 (14.3%)	8,945 (19.7%)	11,533 (26.6%)	16,420 (37.2%)
Postmenopausal	46,494 (84.7%)	37,441 (76.9%)	32,244 (70.9%)	27,605 (63.6%)	23,251 (52.7%)

Values are mean (SD) unless otherwise indicated. Percentages include unknown category for missing data and therefore may not add up to 100%.

Abbreviations: g, grams; MHT, menopausal hormone therapy; IGF-I, insulin-like growth factor-I; N, Number of participants; nmol/L, nanomole per litre; Q, quintile; SD, standard deviation.

4.4.2 Food group intake and IGF-I concentrations

Figure 4.1 presents the multivariable adjusted associations between food group intakes and circulating IGF-I concentrations and **Table 4.2** presents absolute and percentage differences in multivariable adjusted geometric mean concentrations of IGF-I between the highest and lowest categories of food group intake. The largest magnitudes of associations with IGF-I were observed for oily fish and non-oily fish, where participants who reported consuming these foods ≥2 times per week had 1.25 nmol/L (95% CI: 1.19-1.31) and 1.16 nmol/L (1.08-1.24) higher circulating IGF-I concentrations than never consumers, respectively (**Figure 4.1** and **Table 4.2**). Participants who reported consuming poultry ≥2 times per week had 0.87 nmol/L (0.80-0.94) higher IGF-I concentrations compared to participants who said they never consumed poultry (**Figure 4.1** and **Table 4.2**). For vegetable and fresh fruit intake, small positive associations were observed for individuals in the highest category compared to the lowest category, while no associations were observed between intakes of red meat, processed meat, or cheese and circulating IGF-I concentrations (**Figure 4.1**).

4.4.3 Subgroup and sensitivity analyses

In subgroup analyses by sex, the directions of the associations remained the same, although the tests for heterogeneity were statistically significant probably due to differences in the magnitudes of the associations where associations were typically stronger in females (**Figure 4.2**). In sensitivity analyses restricted to participants with IGF-I measured ~4 years after recruitment, associations with food group intake measured at baseline were slightly weaker, although differences in intakes of oily fish and non-oily fish remained positively associated with IGF-I concentrations (**Table 4.2** & **Figure 4.3**).

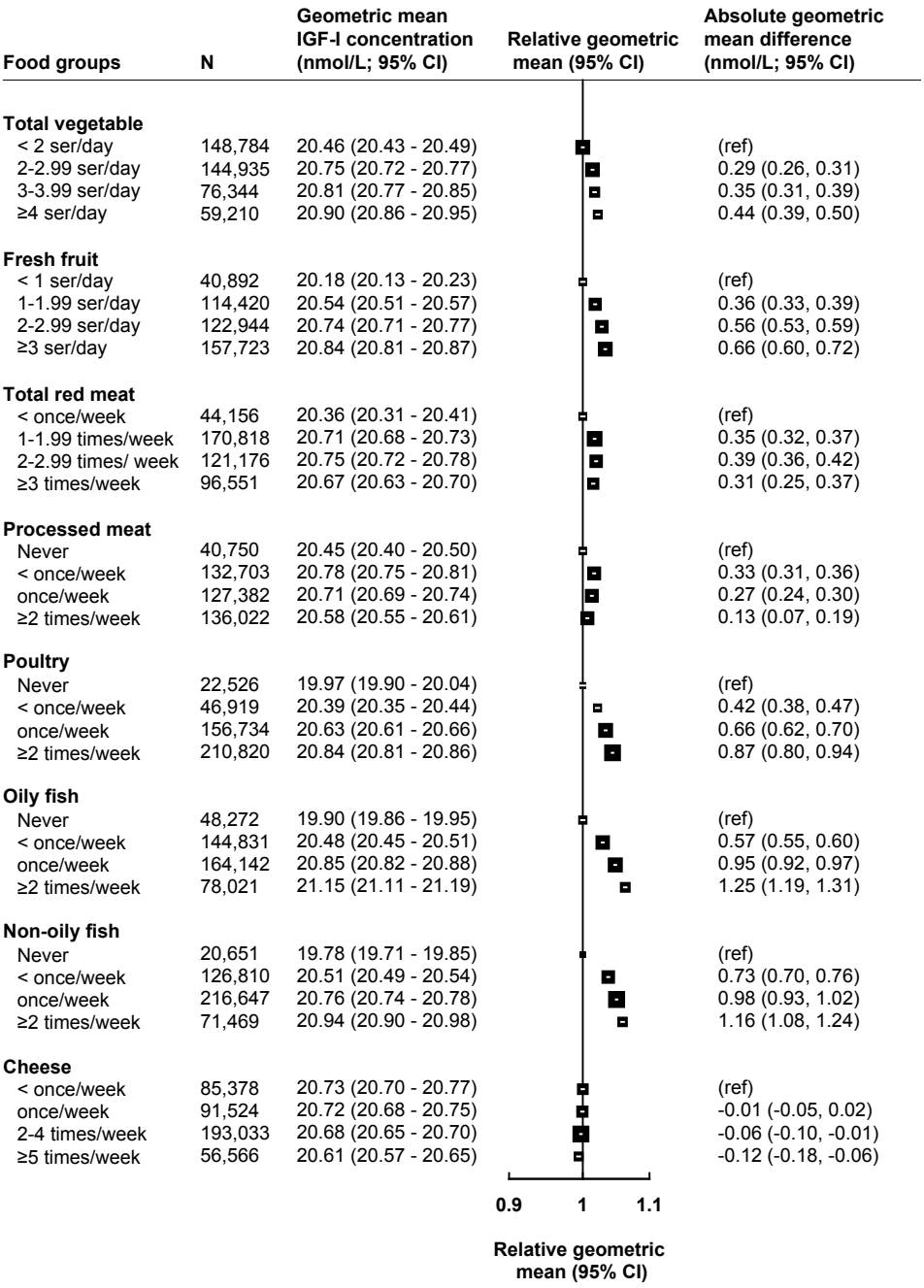


Figure 4.1 Food groups derived from the touchscreen questionnaire in association with geometric mean concentrations of IGF-I (N=438,453).

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status. See Chapter 4 methods for covariate categories.

Abbreviations: CI, confidence intervals; g, grams; IGF-I, insulin-like growth factor-I; N, number of participants; nmol/L, nanomole per litre; ref, reference group; ser, servings.

Table 4.2 Geometric mean difference and percentage difference of IGF-I concentrations comparing the highest category to the lowest category of food group intake using the baseline IGF-I measurement and in sensitivity analyses using the follow-up IGF-I measurement ~4 years after recruitment.

	Baseline IGF-I measurement (n=438,453)		**Follow-up IGF-I measurement (n=16,689)**	
Food groups (categories)	**Absolute geometric mean difference (nmol/L; 95% CI)[1]**	**Percentage geometric mean difference (95% CI)[1]**	**Absolute geometric mean difference (nmol/L; 95% CI) [2]**	**Percentage geometric mean difference (95% CI) [2]**
Total vegetables (≥4 ser/day vs. <2 ser/day)	0.44 (0.39 to 0.50)	2.17% (1.92% to 2.42%)	0.17 (-0.09 to 0.43)	0.84% (-0.47% to 2.16%)
Fresh fruit (≥3 ser/day vs. <1 ser/day)	0.66 (0.60 to 0.72)	3.26% (2.97% to 3.55%)	0.31 (0.00 to 0.63)	1.57% (-0.01% to 3.15%)
Total red meat (≥3 times/week vs. <once/week)	0.31 (0.25 to 0.37)	1.50% (1.21% to 1.80%)	0.26 (-0.05 to 0.56)	1.29% (-0.23% to 2.81%)
Processed meat (≥2 times/week vs. never)	0.13 (0.07 to 0.19)	0.66% (0.36% to 0.95%)	0.08 (-0.22 to 0.38)	0.40% (-1.08% to 1.89%)
Poultry (≥ 2 times/week vs. never)	0.87 (0.80 to 0.94)	4.34% (3.98% to 4.70%)	0.71 (0.38 to 1.05)	3.62% (1.91% to 5.34%)
Oily fish (≥2 times/week vs. never)	**1.25 (1.19 to 1.31)**	**6.26% (5.95% to 6.56%)**	**1.04 (0.73 to 1.36)**	**5.35% (3.72% to 6.99%)**
Non-oily fish (≥2 times/week vs. never)	**1.16 (1.08 to 1.24)**	**5.85% (5.44% to 6.26%)**	0.92 (0.49 to 1.34)	4.73% (2.54% to 6.91%)
Cheese (≥5 times/week vs. <once/week)	-0.12 (-0.18 to -0.06)	-0.59% (-0.86% to -0.31%)	-0.38 (-0.67 to -0.09)	-1.84% (-3.27% to -0.42%)

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status. See Chapter 4 methods for covariate categories.

[1] Baseline IGF-I measurement represents the difference in the highest category of consumption to the lowest category of consumption by food group in Figure 4.1

[2] Follow-up IGF-I measurement represents the difference in the highest category of consumption to the lowest category of consumption by food group reported at the baseline touchscreen questionnaire (see Figure 4.3).

Bolded foods indicate a ≥5% difference in IGF-I concentrations comparing the highest category to the lowest category.

Abbreviations: CI, confidence interval; IGF-I, insulin-like growth factor-I; ser, servings.

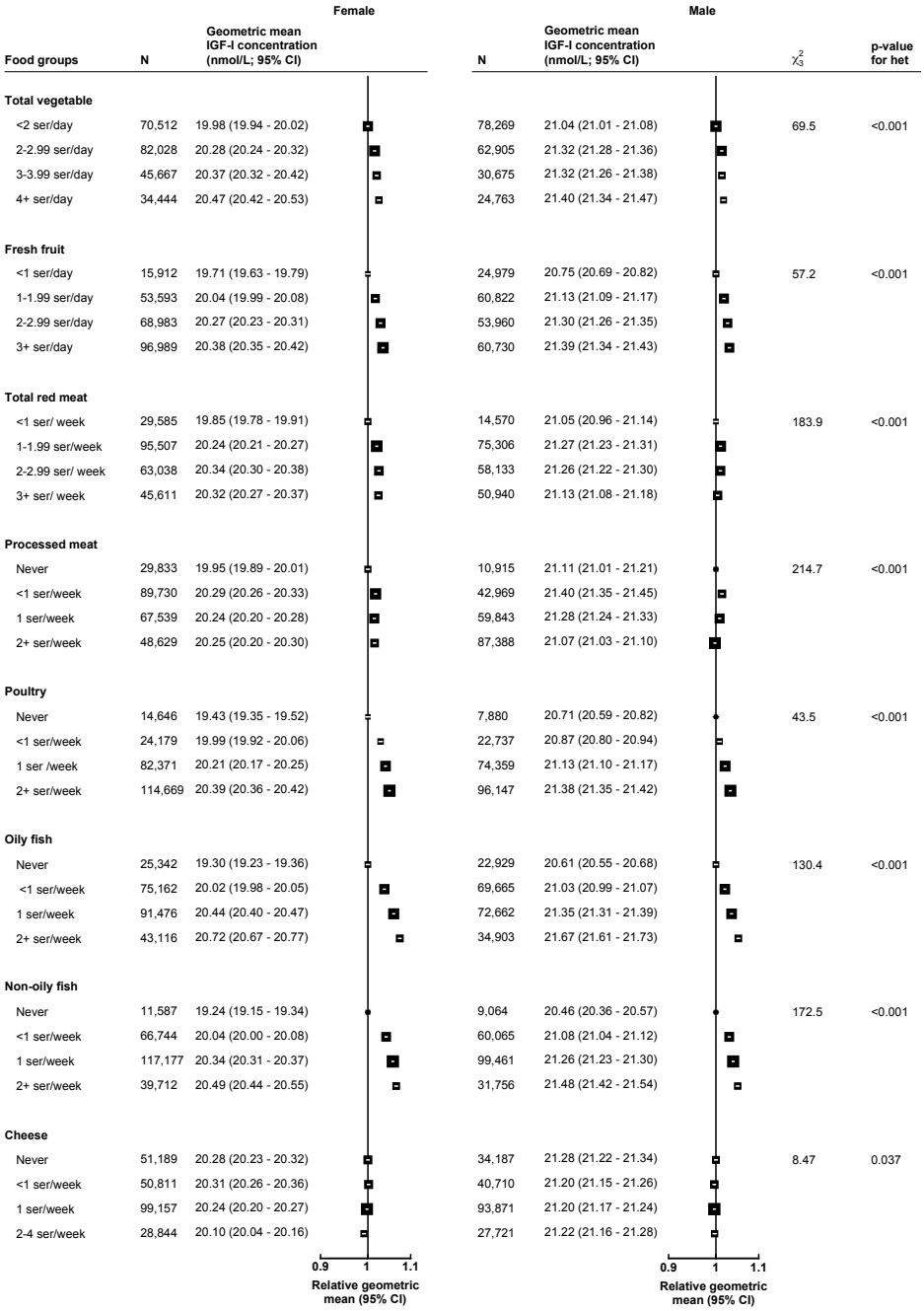


Figure 4.2 Multivariable-adjusted model for food groups derived from the touchscreen questionnaire by sex in association with geometric mean concentrations of IGF-I.

All models are adjusted for age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

Food groups derived from the touchscreen questionnaire completed by all participants at recruitment.

χ^2 and p-values for heterogeneity represent values from likelihood ratio tests for adding an interaction term between sex and food categories and testing for significant model fit.

Abbreviations: CI, confidence intervals; het, heterogeneity; IGF-I, insulin-like growth factor-I; N, number of participants; ser, servings.

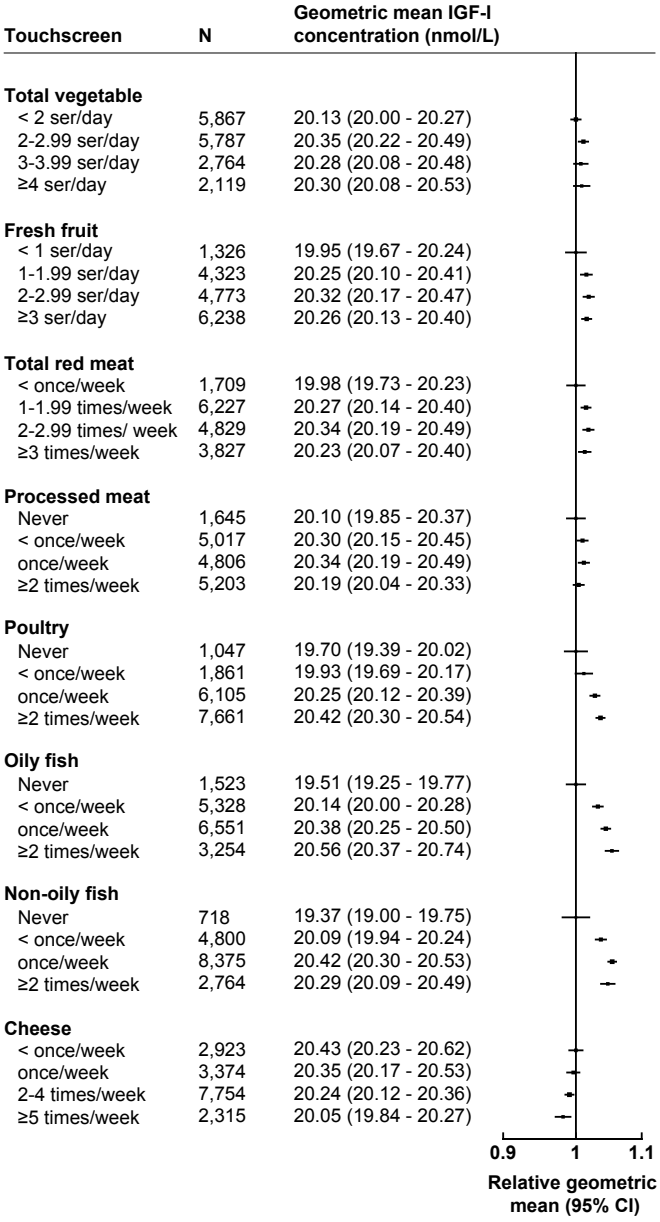


Figure 4.3 Multivariable-adjusted model for food groups derived from the touchscreen at recruitment in association with geometric mean concentrations of follow-up measurement of IGF-I ~4 years after recruitment (N=16,689).

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

Food groups derived from the touchscreen questionnaire completed by all participants at recruitment. Participants are excluded who responded 'prefer not to answer' or 'do not know' for each respective food group therefore not all food groups add up to 16,689.

Participants had IGF-I measured during follow-up. Mean follow-up time from recruitment was 4.3 years.

Abbreviations: CI, confidence intervals; IGF-I, insulin-like growth factor-I; N, number of participants ser, servings.

4.5 Discussion

4.5.1 Main findings

In this cross-sectional analysis of greater than 430,000 individuals in the UK Biobank, positive associations between the intakes of non-oily and oily fish and circulating IGF-I concentrations were found. A modest positive association between intake of poultry and IGF-I concentrations was also observed. No other strong associations were observed for the intakes of fruit, vegetables, red meat, processed meat, or cheese and IGF-I concentrations.

Fish

The consumption of both oily and non-oily fish was positively associated with circulating IGF-I concentrations, which is consistent with some previous cross-sectional studies assessing intake of fish and IGF-I concentrations [3, 4, 374], although one study did not find an association [5]. Both oily and non-oily fish are good sources of protein, essential amino acids, and minerals, such as zinc and potassium, which have been suggested to be positively associated with IGF-I concentrations [3, 5, 366]. Essential amino acids may up-regulate IGF-I mRNA [319] as well as stimulate pathways in the liver necessary for IGF-I synthesis [375]. Some previous studies have also suggested that the intake of polyunsaturated fat, potentially exclusively long chain omega-3 fatty acids [3], may be related to higher IGF-I concentrations [366, 376-378], although the evidence is inconsistent [5]. A relatively strong association for the intake of non-oily fish and IGF-I concentrations was also observed in this analysis, which contains less polyunsaturated fat than oily fish. This may suggest that polyunsaturated fatty acids in oily fish do not explain the association between oily fish intake and IGF-I concentrations, and that other compounds present in fish, such as the high protein content, might explain this association.

Meat

In this analysis, intake of poultry was positively associated with IGF-I concentrations, although this association was weaker than the association with fish intake. The association between poultry intake and IGF-I concentrations has generally been null in previous studies [4, 366, 374].

In contrast, the intake of red and processed meat was not materially associated with IGF-I concentrations in this study, which is consistent with previous studies [3, 4, 367]. It is unclear whether these different associations with IGF-I between protein-rich foods are attributable to different amino acid compositions or if they might be driven by other component(s) in these foods, such as their mineral content [3, 366]. The amino acid profiles of red meat, poultry, and fish do not differ greatly (**Table 4.3**) [379] and therefore may not explain the differences in associations with IGF-I concentrations. However, mineral contents in animal-based foods differ; for example, fish and poultry may have relatively more magnesium than some red meat [379], and some evidence has suggested that magnesium intake may be positively associated with IGF-I concentrations [4, 366]. Despite this, further research is needed to examine how mineral intake may relate to IGF-I concentrations, and whether minerals have independent associations beyond the intake of protein-rich foods.

Table 4.3 Grams and percentage of individual amino acids composition per 100 gram of animal-derived protein source.

Amino Acid	Dairy				Red meat				Poultry			Fish		
	Milk	**Yogurt**	**Blue Cheese**	**Cheddar cheese**	**Beef**	**Veal**	**Pork**	**Lamb**	**Chicken**	**Duck**	**Turkey**	**Salmon**	**Halibut**	**Cod**
Alanine	0.13 (3.4%)	0.16 (4.1%)	0.73 (3.0%)	0.76 (2.6%)	1.69 (6.6%)	1.64 (6.6%)	1.53 (6.0%)	1.29 (6.4%)	-	1.44 (6.3%)	-	1.67 (7.2%)	1.77 (8.1%)	1.42 (6.9%)
Arginine	0.13 (3.4%)	0.13 (3.3%)	1.65 (6.7%)	0.90 (3.1%)	1.54 (6.0%)	1.54 (6.2%)	1.53 (6.0%)	1.39 (6.9%)	0.79 (12.0%)	1.39 (6.1%)	1.21 (12.4%)	1.33 (5.8%)	1.37 (6.3%)	1.21 (5.9%)
Aspartic acid	0.29 (7.5%)	0.28 (7.1%)	0.83 (3.4%)	1.81 (6.2%)	2.34 (9.1%)	2.40 (9.7%)	2.43 (9.6%)	1.89 (9.3%)	-	2.27 (9.9%)	-	2.22 (9.6%)	2.15 (9.9%)	2.01 (9.8%)
Cystine	0.03 (0.7%)	0.03 (0.7%)	0.12 (0.5%)	0.21 (0.7%)	0.28 (1.1%)	0.28 (1.1%)	0.31 (1.2%)	0.17 (0.8%)	0.12 (1.8%)	0.30 (1.3%)	0.27 (2.8%)	0.29 (1.3%)	0.24 (1.1%)	0.25 (1.2%)
Glutamic acid	0.79 (20.5%)	0.70 (17.8%)	4.97 (20.3%)	6.62 (22.7%)	4.13 (16.1%)	3.97 (16.0%)	3.91 (15.4%)	3.05 (15.1%)	-	3.69 (16.1%)	-	3.23 (14.0%)	3.01 (13.9%)	3.13 (15.3%)
Glycine	0.08 (2.0%)	0.09 (2.2%)	0.51 (2.1%)	0.47 (1.6%)	1.56 (6.1%)	1.34 (5.4%)	1.42 (5.6%)	1.09 (5.4%)	-	1.40 (6.1%)	-	1.63 (7.1%)	1.15 (5.3%)	0.94 (4.6%)
Histidine	0.10 (2.5%)	0.10 (2.4%)	0.99 (4.0%)	0.80 (2.7%)	0.85 (3.3%)	0.80 (3.2%)	0.99 (3.9%)	0.60 (3.0%)	0.36 (5.5%)	0.61 (2.7%)	0.52 (5.3%)	0.66 (2.9%)	0.48 (2.2%)	0.52 (2.5%)
Isoleucine*	0.22 (5.7%)	0.22 (5.6%)	1.19 (4.8%)	1.81 (6.2%)	1.25 (4.9%)	1.29 (5.2%)	1.27 (5.0%)	1.02 (5.0%)	0.56 (8.5%)	1.29 (5.6%)	1.01 (10.4%)	1.16 (5.0%)	1.27 (5.8%)	0.99 (4.8%)
Leucine*	0.36 (9.4%)	0.38 (9.7%)	2.14 (8.7%)	2.52 (8.6%)	1.95 (7.6%)	1.89 (7.6%)	1.92 (7.6%)	1.69 (8.3%)	1.03 (15.7%)	1.78 (7.8%)	1.47 (15.1%)	1.77 (7.7%)	1.94 (8.9%)	1.69 (8.3%)
Lysine	0.28 (7.3%)	0.28 (7.1%)	2.38 (9.7%)	2.07 (7.1%)	2.31 (9.0%)	2.05 (8.3%)	2.20 (8.7%)	1.89 (9.3%)	1.01 (15.3%)	2.04 (8.9%)	1.74 (17.9%)	2.02 (8.8%)	1.56 (7.2%)	2.05 (10.0%)
Methionine	0.09 (2.3%)	0.09 (2.3%)	0.52 (2.1%)	0.77 (2.6%)	0.65 (2.5%)	0.60 (2.4%)	0.72 (2.8%)	0.53 (2.6%)	0.33 (5.0%)	0.64 (2.8%)	0.53 (5.5%)	0.70 (3.0%)	0.80 (3.7%)	0.60 (2.9%)
Phenylalanine	0.18 (4.7%)	0.19 (4.8%)	1.22 (5.0%)	1.45 (5.0%)	1.06 (4.1%)	1.02 (4.1%)	0.98 (3.9%)	0.82 (4.0%)	0.54 (8.2%)	0.91 (4.0%)	0.77 (7.9%)	0.91 (3.9%)	0.68 (3.1%)	0.84 (4.1%)
Proline	0.34 (8.8%)	0.42 (10.7%)	2.35 (9.6%)	3.05 (10.4%)	1.28 (5.0%)	1.15 (4.6%)	1.21 (4.8%)	0.97 (4.8%)	-	1.05 (4.6%)	-	1.00 (4.3%)	0.81 (3.7%)	0.82 (4.0%)
Serine	0.21 (5.5%)	0.22 (5.6%)	1.33 (5.4%)	1.58 (5.4%)	1.14 (4.4%)	1.15 (4.6%)	1.12 (4.4%)	0.82 (4.0%)	-	0.92 (4.0%)	-	1.01 (4.4%)	1.27 (5.8%)	0.99 (4.8%)
Threonine	0.16 (4.2%)	0.16 (4.1%)	0.92 (3.7%)	0.98 (3.4%)	1.15 (4.5%)	1.13 (4.6%)	1.25 (4.9%)	0.89 (4.4%)	0.55 (8.4%)	1.01 (4.4%)	0.81 (8.3%)	1.11 (4.8%)	0.99 (4.6%)	0.97 (4.7%)
Tryptophan	0.05 (1.3%)	0.04 (1.1%)	0.21 (0.9%)	0.29 (1.0%)	0.29 (1.1%)	0.30 (1.2%)	0.31 (1.2%)	0.20 (1.0%)	0.12 (1.8%)	0.28 (1.2%)	0.16 (1.6%)	0.26 (1.1%)	0.26 (1.2%)	0.24 (1.2%)
Tyrosine	0.18 (4.7%)	0.18 (4.6%)	1.02 (4.2%)	1.30 (4.5%)	0.89 (3.5%)	0.88 (3.6%)	0.91 (3.6%)	0.82 (4.0%)	0.46 (7.0%)	0.76 (3.3%)	0.28 (2.9%)	0.72 (3.1%)	0.68 (3.1%)	0.71 (3.5%)
Valine*	0.24 (6.2%)	0.27 (6.9%)	1.46 (5.9%)	1.81 (6.2%)	1.32 (5.1%)	1.31 (5.3%)	1.42 (5.6%)	1.12 (5.5%)	0.71 (10.8%)	1.18 (5.1%)	0.95 (9.8%)	1.39 (6.0%)	1.30 (6.0%)	1.09 (5.3%)

Values obtained from Scherz H and Senser F (eds) 2000 Food Composition and Nutrition Tables, 6th edn.

* signifies branched chain amino acid.

Some values for individual amino acids were unavailable for specific foods (chicken and turkey) and therefore were not included.

Vegetables and fresh fruit

In this analysis, small positive associations for intakes of both vegetables and fresh fruit with IGF-I concentrations were observed, however, no significant associations were observed in sensitivity analyses using follow-up measurements of IGF-I. Moreover, although these associations were statistically significant in the main analyses, the differences in IGF-I concentrations between the lowest and highest categories were small (<3.5%); therefore, these results should be interpreted cautiously and could also be due to associations with other foods. Previous evidence has suggested small inverse or null associations of IGF-I with vegetable intake [3, 366], whereas, for fruit intake, small positive [366, 380] or null [3, 381] associations have been reported and further research is needed.

Cheese

No association was observed between cheese intake and IGF-I concentration, which is consistent with previous cross-sectional studies [3, 370, 380]. Despite previous studies showing a positive association between the intake of dairy products and IGF-I concentrations [3, 5, 370], the intake of cheese has not been shown to be associated with IGF-I concentrations [370], suggesting that there may be differences in how dairy products are related with IGF-I concentrations. One possible explanation for the absence of an association for cheese intake is the removal of the whey fraction during cheese production. The whey fraction contains more branched chained amino acids [382], which may be important in stimulating IGF-I production [319], however, further research is needed to explore this and see if the protein content of cheese is associated with IGF-I concentrations. Frequencies of intakes of dairy products other than cheese were not asked in the recruitment questionnaire in UK Biobank, limiting the analyses conducted for dairy foods.

4.5.2 Strengths and limitations

This study has some strengths and limitations that should be considered. This is the largest analysis assessing food group intakes in relation to circulating IGF-I concentrations. The

robustness of the results was also tested using the follow-up measurement of IGF-I, which was from blood samples collected on average 4.3 years after recruitment of participants and results were similar to the main analyses. There are also some limitations that should be considered. Few dietary questions were asked at recruitment, therefore not allowing adjustment for total energy intake and other nutritional factors, as well as limiting the number of foods that could be assessed in relation to IGF-I concentrations in the whole sample. However, adjustment for BMI, height, and physical activity was made to try to control for energy intake. As well, differences in IGF-I concentrations between the highest and lowest categories of the food group were compared, which varied in distribution in the sample and intake amounts between the food groups, thus making it difficult to compare the sizes of the estimates between different foods. There is also measurement error in dietary intake estimates as intakes were only asked once among all participants at the recruitment visit [383]. As well, adjustment for other foods in this analysis was not conducted due to the limited food group questions in the UK Biobank. Due to the observational nature of the study, associations may be subject to unmeasured and residual confounding, and causality cannot be inferred. Moreover, other components of the IGF signalling pathway, such as IGF-II and the IGF binding proteins, which may be important in modulating the effect of IGF-I [6], were not measured in this cohort. The UK Biobank participants are predominantly white and generally healthier than the overall UK population [384], therefore the associations might be influenced by selection bias and may not be generalisable to a wider population. Although some associations were observed between food group intakes and circulating IGF-I concentrations, how the intake of these foods relate to IGF-I associated health outcomes, such as cancer risk and bone health, is unclear [385, 386]. Moreover, intakes of these foods may influence health outcomes through other mechanisms external to the IGF-I pathway, and thus further research is needed before conclusions in relation to disease can be made.

4.6 Conclusions

In conclusion, positive associations between oily fish and non-oily fish and circulating IGF-I concentrations were found. A modest positive association between poultry intake and IGF-I concentrations was also observed, whereas there were no other strong associations with intakes of vegetables, fruit, red meat, processed meat, or cheese. Further research assessing how compounds in these foods, such as individual amino acids and minerals, relate to IGF-I concentrations is warranted. Moreover, studies using methods that may be less susceptible to residual confounding, including large RCTs using isoenergetic methods, are needed to enhance understanding of how dietary components may modulate IGF-I concentrations and potentially impact health outcomes.

Chapter 5

Association of dietary protein and other macronutrients with insulin-like growth factor-I

Chapter 5 summary

Dietary protein intake, particularly dairy protein, may increase circulating IGF-I concentrations; however, associations with different protein sources, other macronutrients, and fibre are inconclusive. Associations between intakes of macronutrients and their sources with IGF-I concentrations were explored among 11,815 participants from the UK Biobank who completed ≥4 24-hour dietary assessments and had serum IGF-I concentrations measured. Circulating IGF-I concentrations were positively associated with intake of total protein (per 2.5% higher energy intake: 0.56 nmol/L, (95% confidence interval: 0.47, 0.66)), milk protein: 1.20 nmol/L (0.90, 1.51), and yogurt protein: 1.33 nmol/L (0.79, 1.86), but not with cheese protein: -0.07 nmol/L (-0.40, 0.25). IGF-I concentrations were also positively associated with intake of fibre (per 5 gram/day higher intake: 0.46 nmol/L (0.35, 0.57)) and starch from wholegrains (Q5 vs. Q1: 1.08 nmol/L (0.77, 1.39)). These results suggest differing associations with IGF-I concentrations depending on the source of dairy protein, with positive associations with milk and yogurt protein but no association with cheese protein, which may be important for the risks of IGF-I related cancers and is further explored in Chapter 6. The positive association of fibre and starch from wholegrains with IGF-I warrants further investigation particularly in relation to colorectal cancer (where these dietary factors have been inversely related with risk) and is further examined in Chapter 7.

5.1 Introduction

Circulating IGF-I concentrations are determined from a combination of both non-modifiable (e.g., age, genetics) and probably also modifiable factors. As described, one potential modifiable factor that has been shown to be positively associated with IGF-I concentrations is intake of protein.

5.1.1 Previous studies

Evidence from small observational studies and RCTs has suggested that a higher intake of protein is associated with higher circulating IGF-I concentrations [3, 4, 313, 316, 387]. However, sources of protein, such as from dairy products or plants, might be associated differently with circulating IGF-I concentrations [5]. As observed in Chapter 4, and in other smaller studies [370], dairy products may be associated differently with IGF-I concentrations and intake of cheese may not be associated with circulating IGF-I concentrations, which has not been well described previously potentially due to studies not having the statistical power to detect associations between intake of dairy product sources and IGF-I concentrations. This may be important as intake of dairy products has been commonly explored together with IGF-I concentrations [5, 310, 367] and different dairy products may have differing associations with IGF-I concentrations and thus have dissimilar associations on cancer risk. For the intake of plant protein, studies have suggested either no association [5, 367] or positive associations with IGF-I concentrations [4, 388]. For dairy protein, evidence suggests a consistent positive association with IGF-I concentrations across studies [5, 310, 367, 388], however, for other sources of animal based protein not from dairy, evidence remains less clear [5].

For other macronutrients, such as carbohydrates and fats, evidence for an association with IGF-I concentrations has generally been null [3-5, 374, 377, 389]. However, previous studies have been limited by relatively small sample sizes and other nutrients have not been well characterised, thus limiting the ability to look at sources of protein and other macronutrients in detail [3-5, 284, 317, 387].

5.2 Aim

The UK Biobank measured IGF-I concentrations in nearly all participants and collected detailed dietary information on a subsample of participants, from which nutrients can be estimated. From this, an observational analysis was conducted to assess the associations of intake of protein from differing sources and other macronutrients with circulating IGF-I concentrations.

5.3 Methods

5.3.1 Study design

Details of the study design, recruitment, and data collection for the UK Biobank are provided in Chapter 3.

5.3.2 Dietary assessment: Oxford WebQ – 24-hour dietary assessment

As previously described, for a subsample of the UK Biobank, the Oxford WebQ, a validated web-based 24-hour dietary assessment was completed multiple times [359]. Details of the Oxford WebQ 24-hour dietary assessments can be found in Chapter 3. Validation of the Oxford WebQ was conducted in a sample of 160 men and women with biomarker measurements and the correlation between protein intake estimated by four 24-hour dietary assessments and recovery biomarker for protein intake (nitrogen excretion) was 0.52 (95% CI: 0.37-0.67) [359].

In this study, participants were eligible to be included if they completed a minimum of four (maximum five) 24-hour dietary assessments, with one being completed at the recruitment visit in order to have one dietary assessment at the same time that the blood sample was taken. A minimum of four 24-hour dietary assessments was chosen to reduce random measurement error due to the day-to-day variation in dietary intake. **Table 5.1** presents the estimated mean and SD of intakes for total carbohydrates, fibre, total protein, protein from dairy products, and protein from milk separated into quartiles calculated from the mean of one, two, three, and four 24-hour dietary assessments among 28,832 participants who

completed a minimum of four WebQs. Using a greater number of 24-hour dietary assessments to estimate intakes results in the mean of quartiles 1 and 4 being higher and lower, respectively, demonstrating potential random measurement error in the estimates with the use of only two WebQs to determine intakes. **Figure 5.1** presents the intakes of total protein, protein from dairy products, protein from milk, and total fibre by quartiles with IGF-I concentrations estimated from two, three, and four 24-hour dietary assessments. As shown in **Figure 5.1**, with a greater number of 24-hour dietary assessments used to determine intakes, there is less random measurement error and thus less regression dilution bias [351]; therefore, for these analyses, a minimum of four 24-hour dietary assessments were required for participants to be included.

Table 5.1 Mean and standard deviation of intakes separated by quartiles derived from the average of one to four 24-hour dietary assessments in 28,832 participants who completed a minimum of four 24-hour dietary assessments (Oxford WebQs).

	Grams				Percentage of energy			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Total carbohydrates								
Mean of 1 WebQ	162.0 (30.5)	224.4 (13.6)	273.6 (15.6)	358.9 (50.0)	38.2 (5.3)	47.4 (1.8)	53.1 (1.6)	61.4 (4.4)
Mean of 2 WebQ	166.9 (29.6)	226.1 (12.6)	271.7 (14.4)	350.7 (46.7)	38.4 (5.3)	47.3 (1.6)	52.7 (1.6)	60.4 (4.1)
Mean of 3 WebQ	174.2 (27.0)	228.7 (11.6)	269.4 (12.5)	338.4 (40.4)	39.7 (4.7)	47.4 (1.4)	52.1 (1.4)	58.7 (3.5)
Mean of 4 WebQ	178.0 (25.5)	229.1 (11.0)	267.5 (12.1)	332.0 (37.5)	40.3 (4.5)	47.6 (1.3)	51.9 (1.3)	58.1 (3.3)
Total fibre								
Mean of 1 WebQ	10.4 (2.47)	15.7 (1.17)	19.9 (1.37)	27.9 (5.16)	-	-	-	-
Mean of 2 WebQ	10.8 (2.37)	15.7 (1.09)	19.7 (1.27)	27.1 (4.78)	-	-	-	-
Mean of 3 WebQ	11.4 (2.14)	15.9 (0.99)	19.5 (1.14)	26.0 (4.23)	-	-	-	-
Mean of 4 WebQ	11.7 (2.00)	16.0 (0.94)	19.3 (1.07)	25.4 (3.96)	-	-	-	-
Total protein								
Mean of 1 WebQ	51.2 (9.39)	71.2 (4.26)	86.1 (4.66)	113.6 (18.6)	11.3 (1.39)	14.3 (0.69)	16.8 (0.77)	21.3 (2.82)
Mean of 2 WebQ	53.4 (9.14)	72.1 (3.93)	85.8 (4.27)	111.1 (17.3)	11.6 (1.37)	14.5 (0.63)	16.8 (0.71)	20.8 (2.70)
Mean of 3 WebQ	56.8 (7.93)	73.0 (3.48)	85.2 (3.80)	106.6 (14.3)	12.2 (1.19)	14.7 (0.55)	16.6 (0.60)	20.1 (2.26)
Mean of 4 WebQ	58.3 (7.43)	73.5 (3.25)	84.7 (3.48)	104.2 (12.7)	12.5 (1.12)	14.8 (0.50)	16.6 (0.56)	19.7 (2.01)
Protein from dairy								
Mean of 1 WebQ	4.7 (2.7)	11.2 (1.5)	16.7 (1.8)	27.2 (7.0)	0.98 (0.55)	2.25 (0.30)	3.35 (0.35)	5.41 (1.32)
Mean of 2 WebQ	5.1 (2.6)	11.3 (1.4)	16.5 (1.7)	26.0 (6.3)	1.04 (0.53)	2.27 (0.28)	3.29 (0.33)	5.18 (1.22)
Mean of 3 WebQ	6.2 (2.6)	11.8 (1.2)	16.3 (1.4)	24.3 (5.1)	1.26 (0.52)	2.38 (0.25)	3.27 (0.28)	4.83 (0.98)
Mean of 4 WebQ	6.6 (2.6)	12.0 (1.2)	16.1 (1.3)	23.5 (4.7)	1.35 (0.52)	2.43 (0.23)	3.26 (0.26)	4.68 (0.89)

Table 5.1 continued

	Grams				Percentage of energy			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Protein from milk								
Mean of 1 WebQ	1.5 (1.6)	5.6 (1.0)	8.4 (0.8)	13.3 (4.1)	0.24 (0.27)	1.03 (0.18)	1.64 (0.19)	2.78 (0.86)
Mean of 2 WebQ	1.5 (1.5)	5.4 (0.8)	8.1 (0.8)	12.9 (3.7)	0.26 (0.28)	1.03 (0.17)	1.62 (0.18)	2.69 (0.80)
Mean of 3 WebQ	1.6 (1.4)	5.4 (0.8)	8.0 (0.7)	12.1 (3.0)	0.32 (0.29)	1.07 (0.16)	1.61 (0.16)	2.54 (0.65)
Mean of 4 WebQ	1.7 (1.5)	5.5 (0.8)	8.0 (0.7)	11.9 (2.9)	0.35 (0.29)	1.11 (0.16)	1.62 (0.15)	2.50 (0.60)

All values are mean (standard deviation)

Abbreviations: Q, quantile.

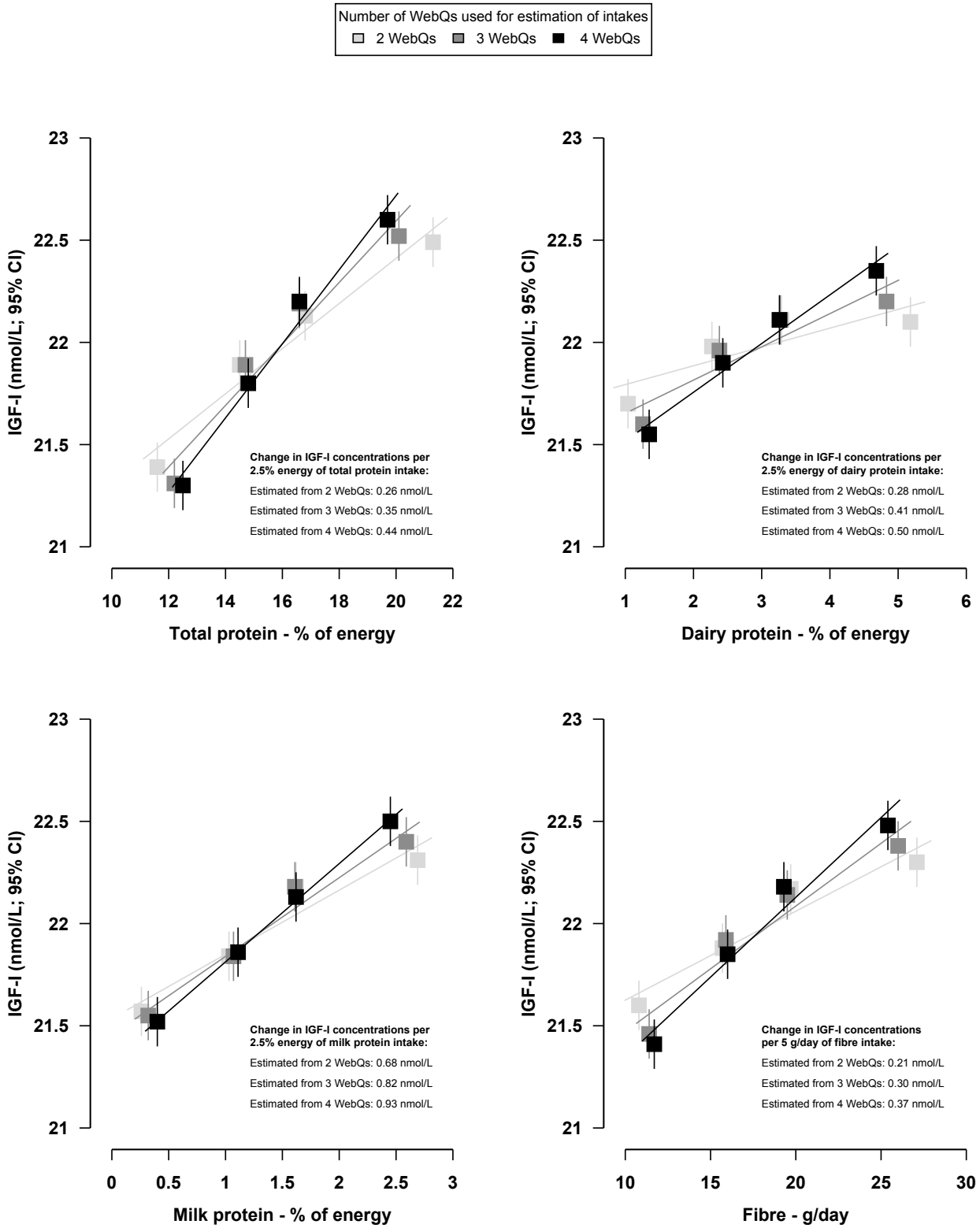


Figure 5.1 Intakes of total protein, dairy protein, milk protein, and fibre derived from two, three, or four WebQs (24-hour dietary assessments) in relation to IGF-I concentrations.

Abbreviations: CI, confidence interval; g/day, grams per day; IGF-I, insulin-like growth factor-I; nmol/L, nanomole per litre.

The percentage of energy from specific macronutrients were calculated for each 24-hour dietary assessment and intake from a minimum of four (maximum five) 24-hour dietary assessments were averaged to estimate the usual percentage of energy intake. Fibre intake was estimated in grams for each 24-hour dietary assessment and averaged from four (maximum five) 24-hour dietary assessments. For alcohol intake, participants were categorised by: <1 gram/day, 1-9.9 g/day, 10-19.9 g/day, 20-39.9 grams/day, and ≥40 g/day from the averaged intake of alcohol. Participants' percentage of energy from macronutrients, fibre in grams, and energy intake were categorised into sex-specific quintiles. The average grams consumed were calculated within each category or quintile of intake of macronutrients, fibre, and alcohol from the 24-hour dietary assessments. For macronutrients, the average percentage of energy consumed per day was calculated within each quintile for the macronutrients.

5.3.3 Laboratory analysis

As described in Chapter 3, non-fasted blood samples were provided by 99.7% of participants at the recruitment visit and transported at 4°C to the central laboratory for cryopreservation and biochemistry measurements. Serum concentrations of IGF-I were measured using DiaSorin Ltd. LIAISON® XL chemiluminescent immunoassay [371]. A total of 16,689 participants had a valid IGF-I follow-up measurement after a subsample of participants repeated the recruitment assessment and provided a second blood sample ~4 years after the initial recruitment visit. Further details of the repeat assessment and repeat measurement of IGF-I concentrations in this subsample are described in Chapters 3 and 4.

5.3.4 Study population and exclusions

The main analyses includes a total of 11,815 participants from UK Biobank who completed four (maximum five) 24-hour dietary assessments and had serum IGF-I concentrations measured at baseline. In secondary analyses, a different sample of participants who had IGF-I re-measured at the reassessment visit (mean 4.3 years after recruitment, SD: 0.9 years) were selected (n=2,724). Participants needed to complete a minimum of four valid 24-hour

dietary assessments at any time to be included in this analysis. A study flowchart of exclusions for each analysis can be found in **Figure 5.2.** At the time of this analysis, 824 participants had withdrawn their informed consent from the UK Biobank and were excluded. Participants were also excluded if they had a prevalent invasive cancer at recruitment recorded in the cancer registry (excluding non-melanoma skin cancer; N=27,174), were taking medications that may alter IGF-I such as GH (N=4,077; Ancillary Table 4.1) or did not have a value for IGF-I concentration at recruitment (N=32,789). Participants were also excluded if they did not complete any 24-hour dietary assessments (N=251,933). As well, individual 24-hour dietary assessments were excluded if they did not report a reliable energy intake (men: >17,575 kJ (4200 kcal) or <3347 kJ (800 kcal); women >14,644 kJ (3500 kcal) or <2092 kJ (500 kcal)) or reported to be ill or fasting on the relevant day. A total of 2,439 and 592 participants were excluded because they did not have a 24-hour dietary assessment with reliable energy intake or they reported to be ill or fasting on the day of the 24-hour dietary assessment, respectively. Participants were also excluded if they did not complete the 24-hour dietary assessment at recruitment (N=123,497), to ensure that at least one valid 24-hour dietary assessment was completed when the participants' blood was drawn. Finally, participants who did not complete a minimum of four valid 24-hour dietary assessments were excluded to reduce random measurement error (N=48,177). Of those who completed a 24-hour dietary assessment at recruitment, 11,815 participants completed four (maximum five) valid 24-hour dietary assessments and had IGF-I concentrations measured at recruitment and were included in this analysis (**Figure 5.2**).

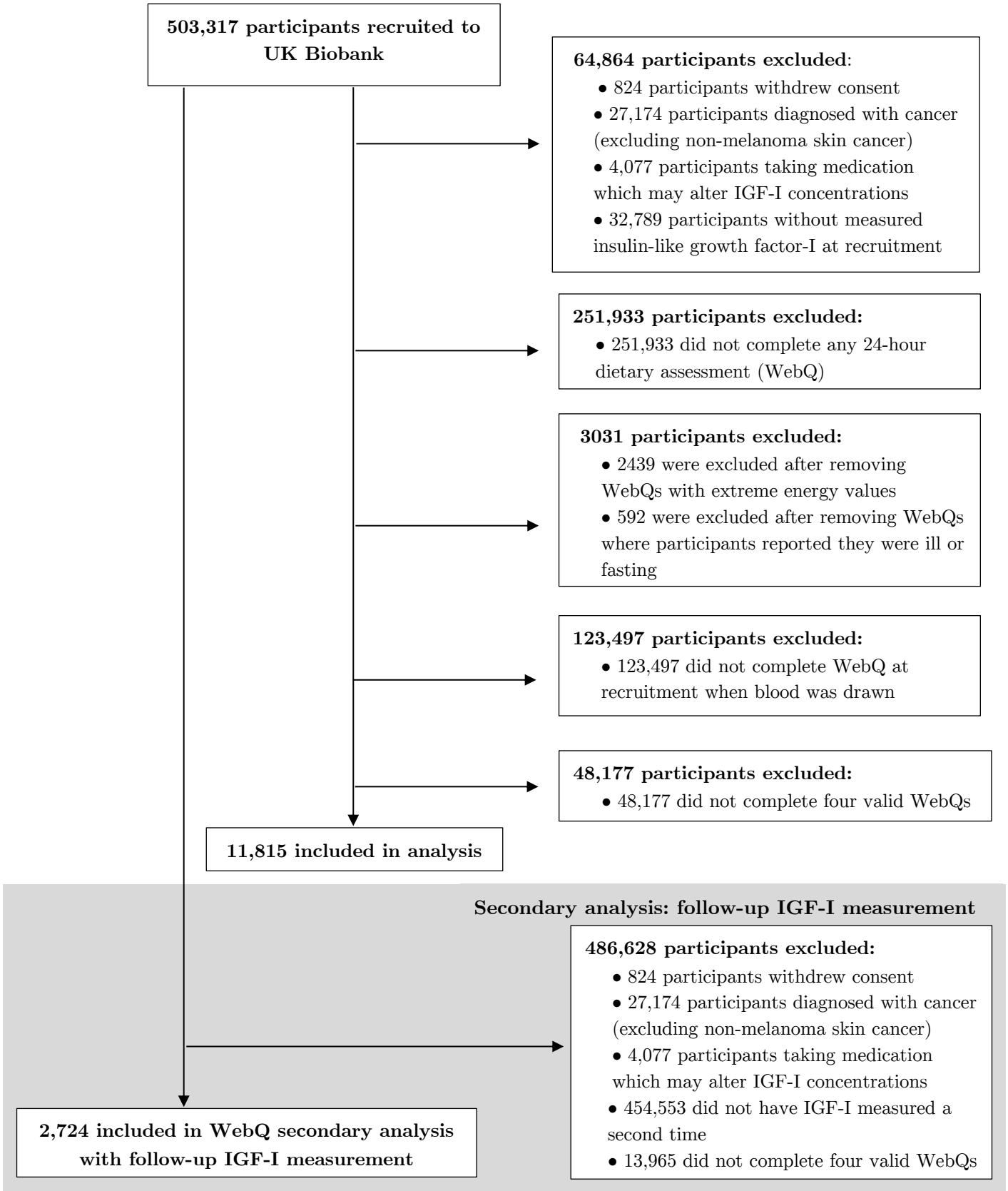


Figure 5.2 Flow chart of exclusion criteria for WebQ 24-hour dietary assessment subsample, and secondary analysis restricting to participants who had a follow-up insulin-like growth factor-I (IGF-I) measurement.

5.3.5 Statistical analysis

IGF-I was logarithmically transformed to obtain the geometric mean concentrations of IGF-I within each category or quintile of macronutrients, fibre, alcohol, and energy intake from linear regression models based on predictions of the model by using the margins, fixing covariates at their means for each category. The geometric means of IGF-I concentrations were estimated rather than the arithmetic mean in order to compare these results to previous studies. To determine relative values, geometric means in the highest category or quintile were divided by the geometric mean in the lowest category or lowest quintile. All macronutrients were also modelled as a continuous variable in increments of 2.5% of energy intake for macronutrients and 5 g/day for fibre intake. A 2.5% higher energy intake was selected to account for the small variability in the macronutrients from specific sources despite the larger variability in total protein, carbohydrate, and fat, and for these analyses, untransformed IGF-I concentrations were used to have a constant slope in the multivariable linear regression models.

All confounders were all selected *a priori* based on probable associations with IGF-I concentrations and dietary intake. Linear regression models were adjusted for sex and age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years), region of recruitment (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland), BMI (<20, 20-22.49, 22.5-24.9, 25-27.49, 27.5-29.9, 30-32.49, 32.5-34.9, ≥35 kg/m^2, and unknown/missing (0.4%)), height (eight sex-specific categories increasing by 5 cm, and unknown/missing (0.5%)), physical activity (low; 0-9.99, medium; 10-49.99, high; ≥50 excess MET-hours/week, and unknown/missing (3.6%)), Townsend deprivation index (quintiles from most deprived to least deprived, and unknown/missing (0.1%)), education (completion of national exam at 16 years of age, completion of national exam at 17-18 years of age, college or university degree, or unknown/missing (17%)), smoking status (never, former, light smoker: <15 cigarettes/day, medium smoker: 15-29 cigarettes/day, heavy smoker: ≥30 cigarettes/day, or missing/unknown

(0.5%)), alcohol consumption from the touchscreen questionnaire (none drinkers, <1, 1-9.99, 10-19.99, ≥20 grams/day or unknown/missing (0.7%); except when alcohol was the exposure of interest), ethnicity (white, mixed race, Indian/Pakistani/Bangladeshi, Chinese/Asian, black/black British, other, or missing/unknown (0.5%)), diabetes status (not diabetic, diabetic, or unknown (0.4%)), energy intake (sex-specific quintiles), and women specific covariates: MHT use (no, former, current, or unknown (0.3%)), oral contraceptive use (no, former, current, or unknown (0.2%)), and menopausal status (premenopausal, postmenopausal, or unknown (4.8%)). Further information on the classification of covariates can be found in the **Ancillary Methods A.1**.

Subgroup and sensitivity analyses

Heterogeneity by sex and by BMI groups (<30, and ≥30 kg/m^2) was assessed using a likelihood ratio test comparing the main model to a model including an interaction term between the macronutrient (modelled per 2.5% of energy intake) and sex or BMI.

To assess the robustness of the results, sensitivity analyses with total percentages of energy from carbohydrates, fat, protein, and alcohol as well as total fibre intake, categorised into sex-specific deciles were conducted. As well, to investigate if the associations observed were independent of other nutrients, analyses were further adjusted for the strongest observed associations, namely fibre intake (sex-specific quintiles), protein from milk (sex-specific quintiles), and protein from yogurt (sex-specific quintiles).

Analyses using IGF-I measurements ~4 years after recruitment (N=2,724)

Finally, secondary analyses were conducted in participants who had IGF-I concentrations measured during the follow-up period at the reassessment visit (mean 4.3 years after recruitment). For this, participants were restricted to those who completed a minimum of four 24-hour dietary assessments at any time and had a follow-up IGF-I measurement (**Figure 5.2**).

All macronutrients were added separately as 2.5% higher energy intake or 5 gram/day fibre intake or modelled as quintiles in multivariable models.

All analyses were conducted using Stata version 15.1 (Stata Corp LP, College Station, TX) and "Jasper makes plots" package version 2-265 in R 3.5.2 was used to make figures. P-values were two-sided and with Bonferroni correction, p-values <0.00185 (0.05/27 exposures) were considered statistically significant. All models were visually assessed to make sure residuals were normally distributed using Q-Q plots, and not heteroscedastic using residual-versus-fitted plots. No assumptions for linear regression were deemed to be broken.

5.4 Results

5.4.1 Participant characteristics

The mean age of participants at recruitment was 56.7 years (SD: 7.7). **Table 5.2** presents participant baseline characteristics by fifths of IGF-I concentrations. Participants who had higher IGF-I concentrations were more likely to be younger, taller, have a lower BMI, report to be never smokers, less likely to be diabetic, and women were less likely to be current users of MHT.

Table 5.2 Characteristics of participants in the UK Biobank WebQ 24-hour dietary assessment subsample by fifths of circulating IGF-I (N=11,815).

	Circulating IGF-I				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**
Number of participants, N	2,363	2,363	2,364	2,363	2,362
IGF-I concentration, nmol/L	14.6 (2.2)	18.9 (0.9)	21.7 (0.8)	24.5 (0.9)	30.1 (4.6)
IGF-I concentration at follow-up, nmol/L[1]	15.7 (3.2)	19.2 (2.9)	21.2 (3.7)	23.2 (3.2)	28.1 (5.0)
Sex - Male, N (%)	859 (36.4%)	984 (41.6%)	1,042 (44.1%)	1,124 (47.6%)	1,120 (47.4%)
Age, years	59.4 (6.5)	58.1 (7.0)	56.6 (7.6)	55.6 (7.8)	53.5 (8.2)
Body mass index, kg/m^2	27.3 (5.3)	26.7 (4.6)	26.2 (4.2)	26.0 (4.0)	25.9 (3.9)
Height, cm	167.5 (8.9)	169.0 (9.0)	169.8 (9.1)	170.2 (9.0)	170.9 (9.1)
Physical activity, N (%)					
Low	661 (28.0%)	615 (26.0%)	608 (25.7%)	609 (25.8%)	585 (24.8%)
Moderate	1,251 (52.9%)	1,262 (53.4%)	1,280 (54.1%)	1,273 (53.9%)	1,346 (57.0%)
High	405 (17.1%)	437 (18.5%)	450 (19.0%)	452 (19.1%)	393 (16.6%)
Townsend deprivation index, N (%)					
Q1 - Most affluent	396 (16.8%)	421 (17.8%)	399 (16.9%)	431 (18.2%)	418 (17.7%)
Q5 - Most deprived	409 (17.3%)	380 (16.1%)	376 (15.9%)	366 (15.5%)	382 (16.2%)
Education, N (%)					
National Examination at age 16 years	358 (15.2%)	331 (14.0%)	300 (12.7%)	303 (12.8%)	286 (12.1%)
National Examination at age 17-18 years	145 (6.1%)	169 (7.2%)	163 (6.9%)	156 (6.6%)	152 (6.4%)
College or University degree	1,707 (72.2%)	1,735 (73.4%)	1,778 (75.2%)	1,792 (75.8%)	1,829 (77.4%)
Smoking, N (%)					
Never	1,290 (54.6%)	1,334 (56.5%)	1,380 (58.4%)	1,437 (60.8%)	1,471 (62.3%)
Previous	914 (38.7%)	877 (37.1%)	842 (35.6%)	798 (33.8%)	756 (32.0%)
Light smoker: <15 cigarettes/day	37 (1.6%)	42 (1.8%)	44 (1.9%)	45 (1.9%)	41 (1.7%)
Medium smoker: 15-29 cigarettes/day	47 (2.0%)	31 (1.3%)	41 (1.7%)	31 (1.3%)	30 (1.3%)
Heavy smoker: 30+ cigarettes/day	71 (3.0%)	75 (3.2%)	55 (2.3%)	48 (2.0%)	60 (2.5%)

Table 5.2 continued

	Circulating IGF-I				
	Q1	**Q2**	**Q3**	**Q4**	**Q5**
Alcohol intake, N (%)					
<1 g/day	526 (22.3%)	489 (20.7%)	498 (21.1%)	518 (21.9%)	520 (22.0%)
1-9.99 g/day	568 (24.0%)	559 (23.7%)	545 (23.1%)	541 (22.9%)	639 (27.5%)
10-19.99 g/day	464 (19.6%)	482 (20.4%)	480 (20.3%)	512 (21.7%)	502 (21.3%)
20-39.9 g/day	477 (20.2%)	517 (21.9%)	580 (24.5%)	535 (22.6%)	517 (21.9%)
≥40 g/day	328 (13.9%)	316 (13.4%)	261 (11.0%)	257 (10.9%)	184 (7.8%)
Ethnicity, N (%)					
White	2,293 (97.0%)	2,300 (97.3%)	2,280 (96.4%)	2,289 (96.9%)	2,261 (95.7%)
Mixed Race	13 (0.6%)	18 (0.8%)	16 (0.7%)	16 (0.7%)	24 (1.0%)
Indian/Pakistani/Bangladeshi	18 (0.8%)	5 (0.2%)	15 (0.6%)	17 (0.7%)	15 (0.6%)
Chinese, Asian, or other Asian	10 (0.4%)	9 (0.4%)	10 (0.4%)	6 (0.3%)	10 (0.4%)
Black or Black British	10 (0.4%)	10 (0.4%)	11 (0.5%)	11 (0.5%)	27 (1.1%)
Other	11 (0.5%)	13 (0.6%)	19 (0.8%)	18 (0.8%)	16 (0.7%)
Diabetic - Yes, N (%)	147 (6.2%)	107 (4.5%)	81 (3.4%)	80 (3.4%)	79 (3.3%)
Women-only covariates					
Current MHT users, N (%)	187 (12.4%)	93 (6.7%)	88 (6.7%)	68 (5.5%)	58 (4.7%)
Current oral contraceptive pill users, N (%)	9 (0.6%)	16 (1.2%)	20 (1.5%)	31 (2.5%)	78 (6.3%)
Menopause status at recruitment, N (%)					
Premenopausal	114 (7.6%)	185 (13.4%)	225 (17.0%)	322 (26.0%)	479 (38.6%)
Postmenopausal	1,297 (86.2%)	1,083 (78.5%)	959 (72.5%)	789 (63.6%)	640 (51.5%)

Values are mean (SD) unless otherwise indicated, percentages include unknown category for missing data.

Percentages calculated including missing values and therefore may not add up to 100%.

[1] Participants include those who had a follow-up blood sample provided and IGF-I concentrations measured (N=2,724).

Abbreviations: MRT, menopausal hormone therapy; IGF-I, insulin-like growth factor-I; N, Number of participants; Q, quintile; SD, standard deviation.

The main results from the multivariable adjusted models are described below if the increment in macronutrient, fibre, alcohol, or energy intake was associated with a ≥5% difference in circulating IGF-I concentrations (**Figures 5.3-5.5** and **Table 5.3**).

5.4.2 Macronutrient intake and circulating IGF-I concentrations

The results for macronutrient intakes by quintiles are shown in **Figures 5.3 & 5.4** and by lowest versus highest quintile in **Table 5.3**. Comparing participants in the highest quintile with participants in the lowest quintile of intake for total protein, protein from animal sources, and protein from milk, IGF-I concentrations were 1.72 nmol/L (95% CI: 1.40, 2.03), 1.19 nmol/L (0.88, 1.50), and 1.18 nmol/L (0.87, 1.49) higher, respectively (**Figure 5.3** & **Table 5.3**). Fat intake was not associated with notable differences in circulating IGF-I concentrations.

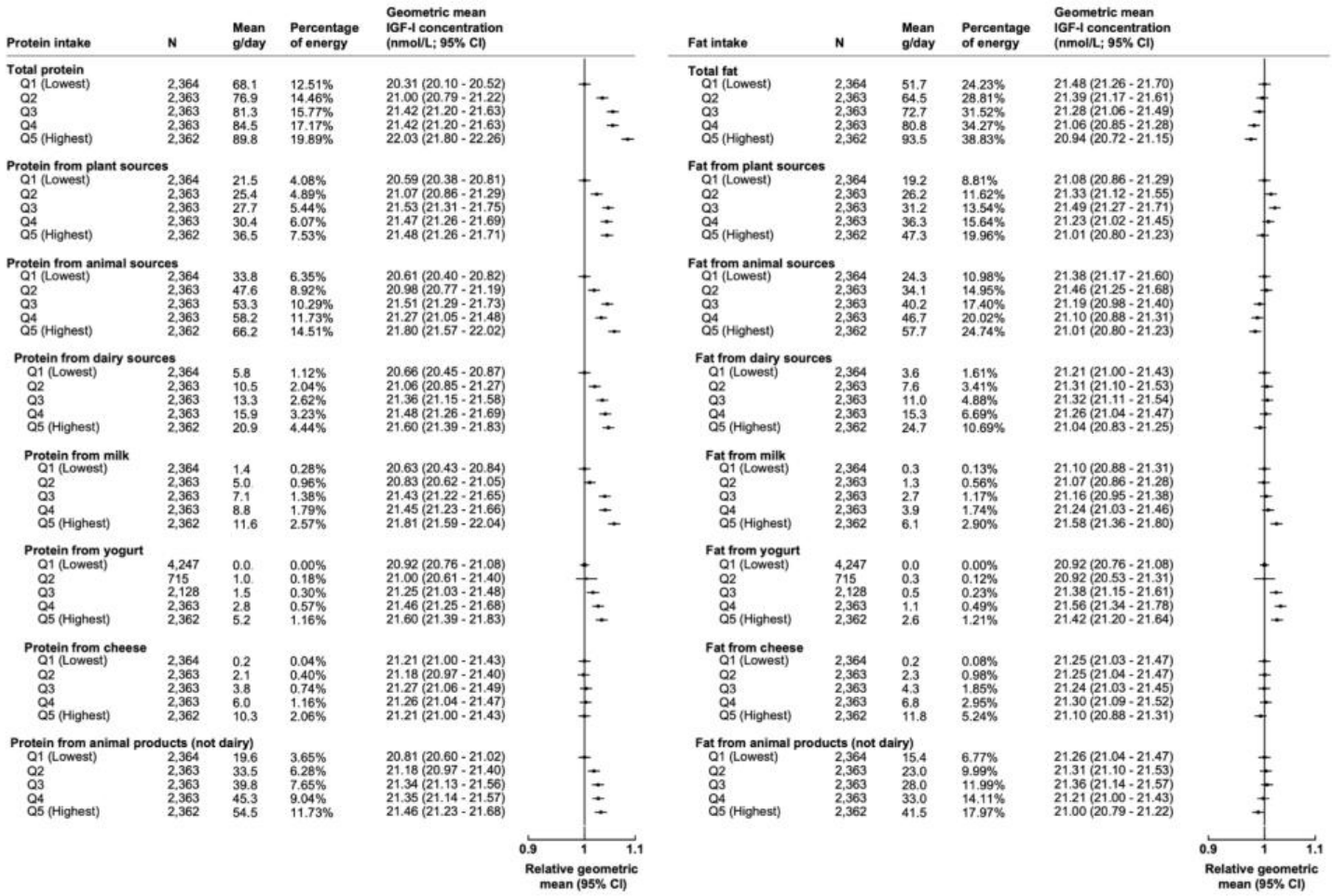


Figure 5.3 Percentage of energy intake from proteins and fats separated by quintiles in association with geometric mean concentrations of IGF-I (N=11,815).

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes status, energy intake and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

Percentage of energy from protein and fat sources calculated from a minimum of four averaged 24-hour web-based dietary assessments with one completed at recruitment.

Grams calculated as mean per day within each quintile. Percentage of energy calculated by mean percentage of energy per day in each quintile.

Abbreviations: CI, confidence intervals; g, grams; IGF-I, insulin-like growth factor-I; N, number of participants; Q, quintile.

Table 5.3 Geometric mean difference (nmol/L) and percentage change of IGF-I comparing the highest quintile to the lowest quintile of nutrients using the baseline IGF-I measurement and in secondary analyses using the follow-up IGF-I measurement ~4 years after recruitment.

	WebQ subsample (N=11,815)		**Secondary analysis: WebQ repeat IGF-I sample (N=2,724)**	
	Absolute change of IGF-I (95% CI)	**Percentage change of IGF-I (95% CI)**	**Absolute change of IGF-I (95% CI)**	**Percentage change of IGF-I (95% CI)**
Total protein	**1.72 (1.40 to 2.03)**	**8.45% (6.92% to 9.98%)**	**1.08 (0.44 to 1.73)**	**5.46% (2.19% to 8.73%)**
Protein from plant sources	0.89 (0.58 to 1.20)	4.32% (2.84% to 5.80%)	0.56 (-0.07 to 1.20)	2.80% (-0.37% to 5.98%)
Protein from animal sources	**1.19 (0.88 to 1.50)**	**5.77% (4.26% to 7.27%)**	0.63 (-0.01 to 1.28)	3.15% (-0.03% to 6.34%)
Protein from dairy sources	0.95 (0.64 to 1.25)	4.59% (3.12% to 6.06%)	**1.39 (0.75 to 2.04)**	**6.98% (3.75% to 10.20%)**
Protein from milk	**1.18 (0.87 to 1.49)**	**5.72% (4.23% to 7.21%)**	**1.38 (0.73 to 2.02)**	**6.90% (3.67% to 10.14%)**
Protein from yogurt	0.69 (0.41 to 0.96)	3.28% (1.98% to 4.58%)	0.95 (0.38 to 1.52)	4.73% (1.90% to 7.55%)
Protein from cheese	0.00 (-0.31 to 0.30)	-0.02% (-1.45% to 1.42%)	0.00 (-0.64 to 0.64)	0.00% (-3.12% to 3.11%)
Protein from animal products (not dairy)	0.65 (0.34 to 0.95)	3.10% (1.62% to 4.58%)	0.05 (-0.59 to 0.69)	0.25% (-2.90% to 3.40%)
Total carbohydrate	0.48 (0.17 to 0.80)	2.33% (0.80% to 3.85%)	0.57 (-0.10 to 1.24)	2.79% (-0.50% to 6.09%)
Total sugars	0.09 (-0.22 to 0.40)	0.42% (-1.05% to 1.88%)	0.07 (-0.58 to 0.72)	0.33% (-2.82% to 3.48%)
Free sugars	-0.96 (-1.27 to -0.66)	-4.46% (-5.87% to -3.05%)	-0.04 (-0.69 to 0.60)	-0.21% (-3.34% to 2.92%)
Non-free sugars	1.02 (0.71 to 1.33)	4.96% (3.45% to 6.48%)	0.84 (0.18 to 1.49)	4.18% (0.92% to 7.44%)
Total starch	0.41 (0.10 to 0.72)	1.95% (0.47% to 3.42%)	0.64 (0.00 to 1.29)	3.18% (-0.02% to 6.37%)
Starch from wholegrains	**1.08 (0.77 to 1.39)**	**5.23% (3.74% to 6.71%)**	**1.26 (0.63 to 1.90)**	**6.41% (3.19% to 9.63%)**
Starch from refined grains	0.29 (-0.01 to 0.60)	1.39% (-0.06% to 2.84%)	0.16 (-0.49 to 0.80)	0.77% (-2.35% to 3.89%)
Fibre	**1.32 (1.00 to 1.65)**	**6.21% (4.87% to 8.06%)**	**1.11 (0.39 to 1.82)**	**5.26% (1.98% to 8.76%)**
Total fat	-0.54 (-0.85 to -0.23)	-2.51% (-3.95% to -1.07%)	-0.61 (-1.26 to 0.05)	-2.89% (-6.00% to 0.22%)
Fat from plant sources	-0.07 (-0.37 to 0.24)	-0.32% (-1.76% to 1.13%)	-0.11 (-0.76 to 0.53)	-0.55% (-3.68% to 2.58%)
Fat from animal sources	-0.37 (-0.67 to -0.07)	-1.73% (-3.16% to -0.30%)	-0.17 (-0.81 to 0.47)	-0.82% (-3.92% to 2.28%)
Fat from dairy sources	-0.17 (-0.47 to 0.13)	-0.81% (-2.24% to 0.62%)	0.10 (-0.54 to 0.75)	0.50% (-2.61% to 3.61%)
Fat from milk	0.48 (0.18 to 0.79)	2.28% (0.83% to 3.73%)	0.66 (0.02 to 1.30)	3.22% (0.09% to 6.35%)
Fat from yogurt	0.50 (0.23 to 0.77)	2.39% (1.10% to 3.67%)	1.00 (0.43 to 1.57)	4.97% (2.15% to 7.80%)

Table 5.3 continued

	WebQ subsample (N=11,815)		Secondary analysis: WebQ repeat IGF-I sample (N=2,724)	
	Absolute change of IGF-I (95% CI)	**Percentage change of IGF-I (95% CI)**	**Absolute change of IGF-I (95% CI)**	**Percentage change of IGF-I (95% CI)**
Fat from cheese	-0.15 (-0.46 to 0.15)	-0.72% (-2.15% to 0.71%)	0.13 (-0.51 to 0.78)	0.65% (-2.47% to 3.77%)
Fat from animal products (not dairy)	-0.25 (-0.56 to 0.05)	-1.19% (-2.62% to 0.24%)	-0.15 (-0.79 to 0.49)	-0.74% (-3.84% to 2.36%)
Alcohol (*comparing ≥40 g/day to <1 g/day)*	**-1.36 (-1.00 to -1.71)**	**-6.32% (-4.68% to -7.97%)**	**-0.50 (-1.24 to 0.24)**	**-2.43% (-6.05% to 1.18%)**
Total energy	0.25 (-0.05 to 0.55)	1.19% (-0.26% to 2.64%)	0.52 (-0.62 to 1.65)	2.55% (-3.06% to 8.17%)
Total energy (not from alcohol)	0.55 (0.24 to 0.85)	2.63% (1.18% to 4.09%)	0.71 (-0.43 to 1.84)	3.53% (-2.14% to 9.21%)

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, energy intake (except when energy intake was the exposure of interest), and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

For baseline IGF-I measurements, quintiles of percentage of energy from carbohydrate sources, percentage of energy from fat sources, and percentage of protein sources, fibre, alcohol categories, and energy intake quintiles calculated from a minimum of four (maximum of five) averaged 24-hour web-based dietary assessments with one assessment completed at recruitment for baseline IGF-I measurement when blood sample was provided. For follow-up IGF-I measurements, quintiles of percentage of energy from carbohydrates, fat sources, protein sources, and intake of fibre, alcohol categories, and energy quintiles calculated from a minimum of four (maximum of five) averaged 24-hour web-based dietary assessments completed at any time.

Percentages and geometric mean difference from highest category or quintile compared to lowest category or quintile of intake are from Figures 5.1 and 5.3 in main text for baseline IGF-I measurement. Quintiles for follow-up IGF-I measurement from highest category or quintile to lowest category. All quintiles of nutrients are not presented for follow-up measurement of IGF-I.

Bolded nutrients represent a ≥5% difference in IGF-I concentrations comparing the highest category or quintile to the lowest category or quintile.

Abbreviations: CI, confidence interval; IGF-I, insulin-like growth factor-I.

Participants in the highest quintile of fibre intake had 1.32 nmol/L (1.00, 1.65) higher IGF-I concentrations compared to those in the lowest quintile (**Figure 5.4**). Circulating IGF-I concentrations were 1.08 nmol/L (0.77, 1.39) higher in the highest quintile of percentage of energy intake from starch from wholegrains compared to the lowest quintile. Circulating IGF-I concentrations were significantly lower for participants who reported consuming ≥40 grams of alcohol a day compared to those who reported drinking <1 gram of alcohol per day (-1.36 nmol/L, 95% CI: -1.00, -1.71; **Figure 5.4**).

A 2.5% higher intake of energy from total protein was associated with higher IGF-I concentrations by 0.56 nmol/L (95% CI: 0.47, 0.66; **Figure 5.5**). When looking at specific sources of protein, a 2.5% higher intake in energy from protein from dairy was associated with 0.73 nmol/L (0.52, 0.94) higher IGF-I concentrations. When dairy sources were looked at in more detail, protein from milk and protein from yogurt were associated with higher IGF-I concentrations of 1.20 nmol/L (0.90, 1.51), and 1.33 nmol/L (0.79, 1.86), respectively; however, there was no association between protein from cheese and IGF-I concentrations (-0.07 nmol/L; -0.40, 0.25; **Figure 5.5** and **Table 5.4**).

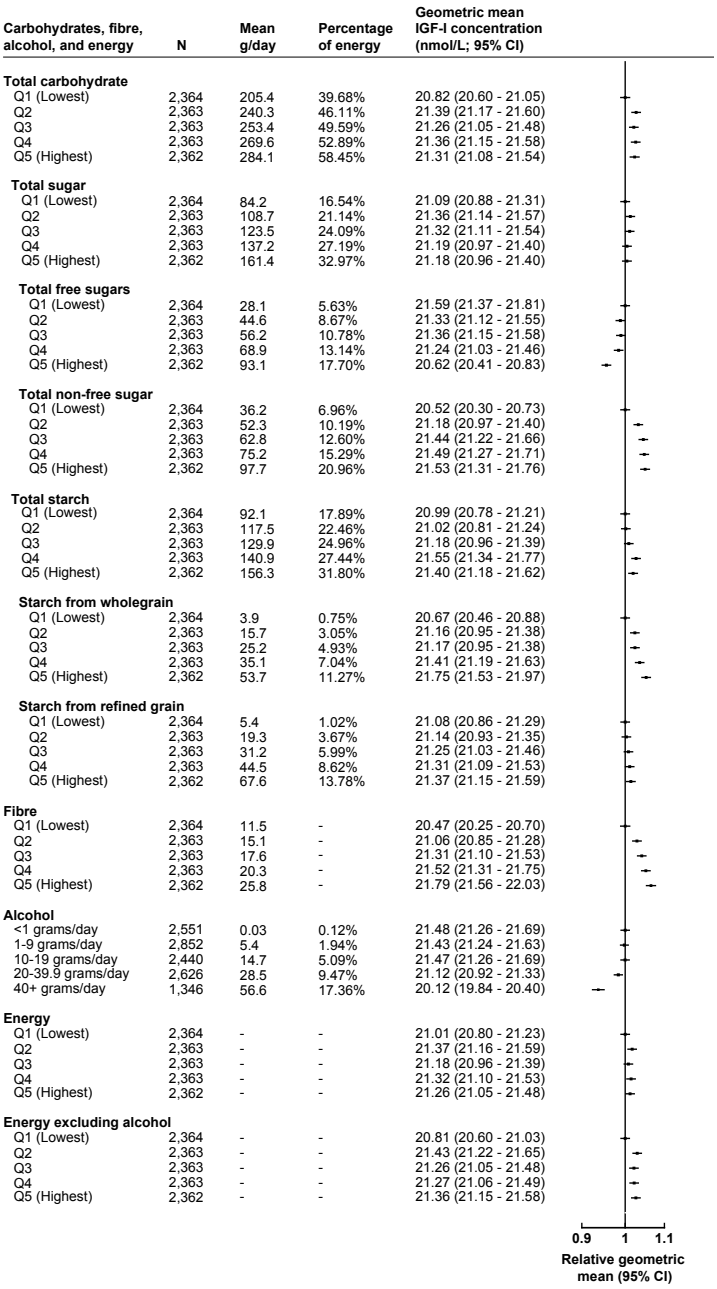


Figure 5.4 Percentage of energy from carbohydrate sources, fibre, alcohol, and energy intake in association with geometric mean concentrations of IGF-I (N=11,815).

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption (except when alcohol was the exposure), ethnicity, diabetes status, energy intake and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

Percentage of energy from carbohydrate sources, fibre quintiles, alcohol categories, and energy intake quintiles calculated from a minimum of four averaged 24-hour web-based dietary assessments with one completed at recruitment.

Grams calculated as mean per day within each quintile and category.

Percentage of energy calculated by mean percentage of energy per day in each quintile.

Abbreviations: CI, confidence interval; g, grams; IGF-I, insulin-like growth factor-I; N, number of participants; nmol/L, nanomole per litre; Q, quintile.

5.4.3 Subgroup and sensitivity analyses

No evidence of heterogeneity was observed between sex and BMI subgroups for macronutrients and fibre intake (Ancillary Figures 5.1 & 5.2).

The patterns of associations remained the same when percentages of energy from macronutrients and fibre were modelled as deciles (Ancillary Figure 5.3). In the multivariable models controlling for fibre intake, the associations of protein from various sources with circulating IGF-I concentrations remained largely the same (data not shown). Similarly, when milk protein and yogurt protein were added to the multivariable models, the association between fibre intake and starch from wholegrains and IGF-I concentrations was not materially affected (data not shown).

Analyses using IGF-I measurements ~4 years after recruitment (N=2,724)

In secondary analyses using participants with IGF-I measured ~4 years after recruitment who had completed a minimum of four 24-hour dietary assessments, the results were largely similar compared to using the recruitment IGF-I measurements; however, the association of IGF-I with protein from milk intake was stronger (per 2.5% energy increase: 1.42 nmol/L (0.81, 2.03); **Figure 5.5** and **Table 5.4**). The associations with protein from yogurt, and fibre intake, became slightly attenuated (per 2.5% energy increase: 0.99 nmol/L (0.06, 2.05) and per 5 g/day increase: 0.23 nmol/L (0.01, 0.45), respectively; **Figure 5.5** and **Table 5.4**).

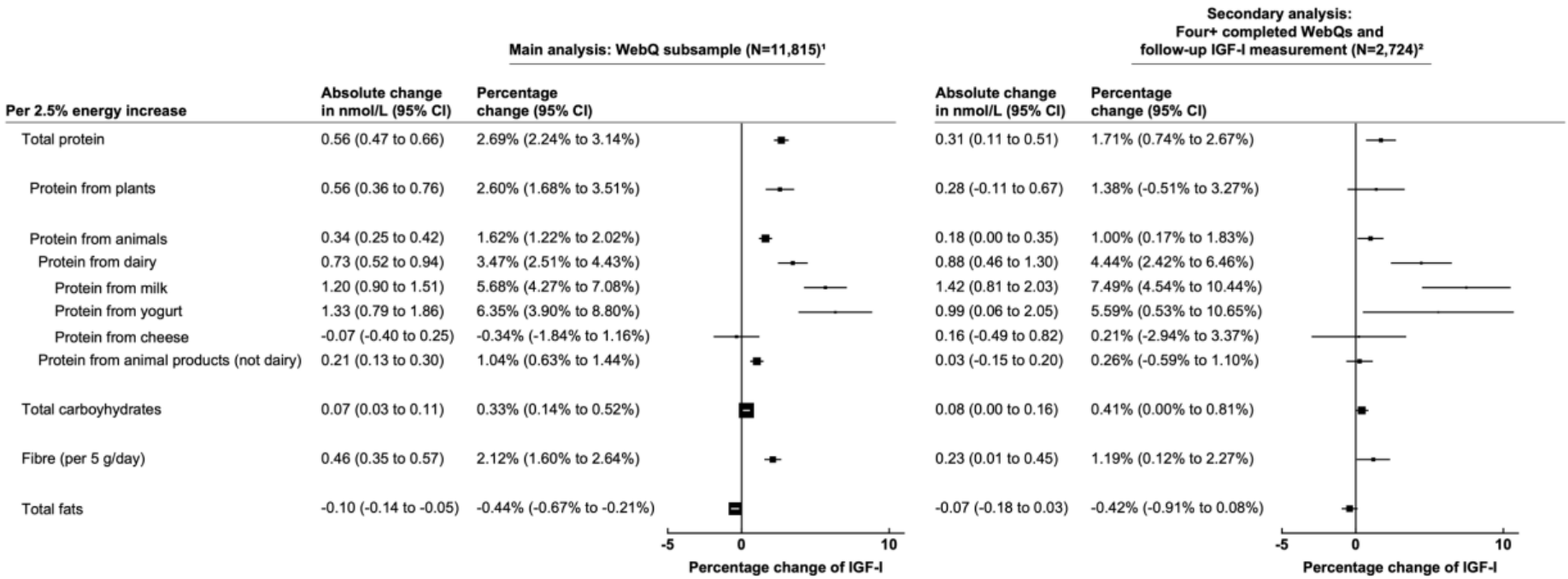


Figure 5.5 Per incremental intake of energy from macronutrients and fibre (N=11,815) and restricting to participants with an IGF-I measurement ~4 years after recruitment (N=2,724) in association with the absolute and percentage change concentration of IGF-I.

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes status, energy intake, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

[1] Macronutrients calculated a minimum of four (maximum of five) averaged 24-hour web-based diet assessments with one completed at recruitment.

[2] Analysis assessing macronutrients and fibre intake calculated for participants with a follow-up IGF-I measurement. Macronutrients and fibre intake were calculated by using a minimum of four (maximum of five) averaged 24-hour web-based dietary assessments completed at any time. IGF-I was measured after all completed 24-hour dietary assessments (mean follow-up time from last 24-hour dietary assessment to follow-up blood measurement: 0.64 years) and ~4 years after recruitment.

Macronutrients are modelled as a 2.5% energy increase whereas fibre is modelled as a 5 gram per day increase.

Abbreviations: CI, confidence interval; g/day, grams per day; IGF-I, insulin-like growth factor-I; nmol/L, nanomole per litre.

Table 5.4 Macronutrient per 2.5% of energy intake in association with absolute difference (nmol/L) and percentage change of IGF-I for the WebQ 24-hour dietary assessment subsample (N=11,815) and in secondary analysis for participants who completed four or more WebQs and had IGF-I measured ~4 years after recruitment (N=2,724).

	WebQ subsample[1] (N=11,815)		**Secondary analysis: Four or more completed WebQs and follow-up IGF-I measurement (N=2,724)[2]**	
Per 2.5% energy increase	**Absolute change of IGF-I (95% CI)**	**Percentage change of IGF-I (95% CI)**	**Absolute change of IGF-I (95% CI)**	**Percentage change of IGF-I (95% CI)**
Total protein	0.56 (0.47 to 0.66)	2.69% (2.24% to 3.14%)	0.31 (0.11 to 0.51)	1.71% (0.74% to 2.67%)
Protein from plant sources	0.56 (0.36 to 0.76)	2.60% (1.68% to 3.51%)	0.28 (-0.11 to 0.67)	1.38% (-0.51% to 3.27%)
Protein from animal sources	0.34 (0.25 to 0.42)	1.62% (1.22% to 2.02%)	0.18 (0.00 to 0.35)	1.00% (0.17% to 1.83%)
Protein from dairy sources	0.73 (0.52 to 0.94)	3.47% (2.51% to 4.43%)	0.88 (0.46 to 1.30)	4.44% (2.42% to 6.46%)
Protein from milk	**1.20 (0.90 to 1.51)**	**5.68% (4.27% to 7.08%)**	**1.42 (0.81 to 2.03)**	**7.49% (4.54% to 10.44%)**
Protein from yogurt	**1.33 (0.79 to 1.86)**	**6.35% (3.90% to 8.80%)**	**0.99 (0.06 to 2.05)**	**5.59% (0.53% to 10.65%)**
Protein from cheese	-0.07 (-0.40 to 0.25)	-0.34% (-1.84% to 1.16%)	0.16 (-0.49 to 0.82)	0.21% (-2.94% to 3.37%)
Protein from animal products (not dairy)	0.21 (0.13 to 0.30)	1.04% (0.63% to 1.44%)	0.03 (-0.15 to 0.20)	0.26% (-0.59% to 1.10%)
Total carbohydrate	0.07 (0.03 to 0.11)	0.33% (0.14% to 0.52%)	0.08 (0.00 to 0.16)	0.41% (0.00% to 0.81%)
Total sugars	0.02 (-0.03 to 0.06)	0.05% (-0.15% to 0.26%)	0.03 (-0.06 to 0.12)	0.14% (-0.28% to 0.55%)
Free sugars	-0.19 (-0.25 to -0.13)	-0.94% (-1.20% to -0.68%)	-0.02 (-0.14 to 0.09)	-0.19% (-0.74% to 0.36%)
Non-free sugars	0.18 (0.13 to 0.23)	0.85% (0.61% to 1.09%)	0.06 (-0.04 to 0.16)	0.34% (-0.15% to 0.83%)
Total starch	0.09 (0.04 to 0.14)	0.44% (0.20% to 0.67%)	0.08 (-0.03 to 0.18)	0.42% (-0.07% to 0.92%)
Starch from wholegrains	0.24 (0.17 to 0.30)	1.15% (0.84% to 1.45%)	0.25 (0.12 to 0.38)	1.26% (0.64% to 1.87%)
Starch from refined grains	0.05 (0.00 to 0.11)	0.25% (0.00% to 0.50%)	0.00 (-0.10 to 0.11)	-0.04% (-0.56% to 0.48%)
Fibre (per 5 grams/day)	0.46 (0.35 to 0.57)	2.12% (1.60% to 2.64%)	0.23 (0.01 to 0.45)	1.19% (0.12% to 2.27%)
Total fat	-0.10 (-0.14 to -0.05)	-0.44% (-0.67% to -0.21%)	-0.07 (-0.18 to 0.03)	-0.42% (-0.91% to 0.08%)
Fat from plant sources	-0.04 (-0.10 to 0.03)	-0.19% (-0.47% to 0.10%)	-0.04 (-0.17 to 0.08)	-0.26% (-0.86% to 0.34%)
Fat from animal sources	-0.08 (-0.13 to -0.03)	-0.34% (-0.57% to -0.11%)	-0.04 (-0.15 to 0.06)	-0.23% (-0.72% to 0.25%)
Fat from dairy sources	-0.07 (-0.14 to 0.01)	-0.33% (-0.67% to 0.01%)	0.00 (-0.15 to 0.15)	-0.13% (-0.85% to 0.58%)
Fat from milk	0.29 (0.06 to 0.52)	1.28% (0.22% to 2.34%)	0.47 (0.01 to 0.93)	2.60% (0.39% to 4.81%)
Fat from yogurt	0.53 (0.05 to 1.00)	2.65% (0.47% to 4.83%)	**0.90 (0.01 to 1.80)**	**5.02% (0.63% to 9.40%)**
Fat from cheese	-0.10 (-0.22 to 0.03)	-0.46% (-1.05% to 0.13%)	0.11 (-0.15 to 0.37)	0.33% (-0.92% to 1.57%)
Fat from animal products (not dairy)	-0.07 (-0.13 to -0.01)	-0.28% (-0.57% to 0.01%)	-0.06 (-0.19 to 0.06)	-0.26% (-0.86% to 0.34%)

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, energy intake, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

[1] WebQ 24-hour dietary assessment subsample represents a 2.5% energy change averaged across a minimum of four (maximum of five) 24-hour web-based diet assessments with one assessment completed at recruitment with baseline IGF-I concentrations. Fibre intake represents a 5 grams per day increase.

[2] Analysis assessing macronutrients and fibre intake calculated for participants with a follow-up IGF-I measurement. Macronutrients and fibre intake were calculated by using a minimum of four (maximum of five) averaged 24-hour web-based dietary assessments completed at any time. IGF-I was measured after all completed WebQs (mean follow-up time from last 24-hour dietary assessment to follow-up blood measurement: 0.64 years, mean follow-up from recruitment to follow-up blood measurement: 4.3 years).

Bolded macronutrients represent ≥5% difference of IGF-I concentrations from the follow-up measurement.

Abbreviations: CI, confidence intervals; g/day, grams per day; IGF-I, insulin-like growth factor-I; nmol/L, nanomole per litre.

5.5 Discussion

5.5.1 Main findings

In this observational analysis in UK Biobank, the association of circulating IGF-I concentrations with protein from dairy products differed by sources of dairy, with protein from milk and yogurt being positively associated with IGF-I concentrations, whereas protein from cheese was not associated with circulating IGF-I concentrations. There were also positive associations between total protein intake, fibre, and starch from wholegrains and circulating IGF-I concentrations, and an inverse association with alcohol consumption. These findings showing different associations with IGF-I depending on the source of dairy protein are potentially important, and together with the positive associations of intake from dietary fibre and starch from wholegrains with IGF-I warrant further investigation.

5.5.2 Protein and IGF-I concentrations

Similar to previous cross-sectional analyses [3-5, 367], higher total protein intake was associated with higher circulating concentrations of IGF-I in these analyses. This association has also been shown in some RCTs of dairy-based protein supplements, which used whey and/or milk protein isolate [316, 317], and in a trial assessing isoenergetic protein restriction where reduction of protein intake resulted in lower IGF-I concentrations [390]. Circulating IGF-I concentrations have also been shown to be lower in individuals who are vegan [298, 299]; a diet that contains no dairy and usually less protein than average[391].

5.5.3 Protein from dairy products and IGF-I concentrations

Although protein from all dairy products combined was associated with higher IGF-I concentrations, protein from milk and yogurt were strongly associated with circulating IGF-I, whereas there was no evidence of an association with protein from cheese, suggesting a potential difference between dairy sources. In analyses using the IGF-I measurement ~4 years after recruitment, the association of protein from milk with IGF-I concentrations became stronger,

whereas protein from yogurt became slightly attenuated, although it was still positively associated with IGF-I. Previous studies have also reported a positive association between total dairy protein and IGF-I [5, 367], as well as a possible null association between cheese intake and IGF-I concentrations [3, 370, 380], although the results are inconsistent [366], and to my knowledge, evidence on protein from different dairy products have not been previously explored. In Chapter 4, there was also no association between cheese intake, measured from the touchscreen questionnaire in the entire UK sample, and circulating IGF-I concentrations. As a possible reason for this difference by dairy subtype, whey protein is found in milk and yogurt, but is removed in most cheese production. Whey contains relatively more branched chained amino acids (leucine, isoleucine, and valine) and is more quickly absorbed into the bloodstream compared to casein, the protein found in cheese [392], although it is unclear whether whey protein increases IGF-I more than casein protein [393]. The supply of essential amino acids, such as tryptophan, may be particularly important in the up-regulation of IGF-I genes [319, 394] and signalling pathways in the liver necessary for IGF-I synthesis [375]. Other components in dairy products, such as calcium, may also positively affect IGF-I concentrations [4, 5, 366, 367], but evidence from a RCT showed no association between calcium supplementation and IGF-I concentrations [395].

5.5.4 Fibre and starch from wholegrains and IGF-I concentrations

This analysis also found that higher fibre and wholegrain intakes were associated with higher circulating concentrations of IGF-I, even after controlling for protein from milk and yogurt. This finding for fibre is consistent with two previous cross-sectional studies of 4,731 individuals and 1,037 women [3, 5], although other smaller cross-sectional studies have not found evidence of an association [377, 396]. Analyses from Chapter 4 also showed small positive associations for vegetables and fruits with IGF-I concentrations, which may be due to their fibre content. Moreover, wholegrains are rich in dietary fibre, and it is possible that the association with IGF-I may be due to their fibre content. Fibre might influence IGF-I concentrations via the gut microbiota; animal studies have shown that anaerobic intestinal microbiota such as *Lactobacillus*

and *Bifidobacterium* ferment fibre in the colon to produce SCFA [397], and some research using mice models suggests that SCFA supplementation increases IGF-I concentrations [397, 398]. Further research is required to determine whether fibre may increase IGF-I concentrations in humans through SCFA or other mechanisms.

In line with previous studies [309, 377], participants consuming the highest amount of alcohol (≥40 g/day) had lower concentrations of IGF-I. This may be due to ethanol toxicity, which impairs the hepatic synthesis of IGF-I [399] and further research should explore this further.

5.5.5 Strengths and limitations

This study has several strengths. To the best of my knowledge, this is currently the largest analysis to be conducted assessing macronutrient intakes in relation to circulating IGF-I concentrations, therefore there is high statistical power to detect associations. Multiple 24-hour dietary assessments were used to estimate the usual macronutrient intake, which reduces random measurement error in the estimates (**Figure 5.1**). The robustness of the results was also tested using the follow-up measurement of IGF-I, which was taken an average of 0.64 years after the last 24-hour dietary assessment was completed.

There are some limitations that should also be considered. Only a subsample of the participants in the UK Biobank completed the 24-hour dietary assessment several times, however, this subsample is larger than any previous study assessing these associations [5]. Although a minimum of four 24-hour dietary assessments were completed by this subsample, dietary intakes estimated using these assessments are still subject to measurement error and inaccuracies in self-report, which may bias these results. As well, the nature of this analysis did not allow us to consider temporality as most 24-hour dietary assessments were completed several months after participants' blood was drawn. This was considered in the secondary analyses restricted to participants with IGF-I measured ~4 years after recruitment, and the results were largely similar and, in some cases, stronger. These analyses also did not consider other related components of

the IGF signalling pathway, such as IGF-II and the IGFBP, as they were not measured in this cohort. Most notably, IGFBP-3, which binds the majority of IGF-I in circulation [400], has been shown to be associated with dietary intake, such as calcium and protein intake, in cross-sectional studies [3-5] and thus, together with possible dietary impacts on IGFBP-1 and IGFBP-2 [298], the bioavailability of IGF-I may be different from the observed associations, which may be important for disease risk. The UK Biobank participants are predominantly white and generally healthier than the overall population [384]. As well, participants in this subsample had to complete a minimum of four 24-hour dietary assessments, which introduces potential selection bias as these individuals are likely to be healthier than those not responding several times [401]; as a result, the estimates may not be generalisable to a wider population. Finally, the associations may be subject to unmeasured and residual confounding, and causality cannot be inferred.

5.6 Conclusions

In conclusion, these results show differing associations with IGF-I depending on the source of dairy protein, with protein from milk and yogurt being associated with IGF-I concentrations, whereas protein from cheese was not associated with circulating IGF-I concentrations. There were also positive associations between total protein intake, fibre, and starch from wholegrains and circulating IGF-I concentrations, and an inverse association with alcohol consumption. The positive association of fibre and starch from wholegrains with IGF-I is not well characterised and warrants further investigation. Further research assessing individual amino acids, as well as using methods that may be less susceptible to residual confounding, including large RCTs and MR studies, are needed to enhance the understanding of how dietary components may modulate IGF-I concentrations.

Chapter 6

Dairy protein intake and risks of colorectal, breast, and prostate cancer

Chapter 6 summary

Intake of protein from dairy products has been associated with higher IGF-I concentrations, a risk factor for colorectal, breast, and prostate cancer. However, evidence regarding the association of intakes of all dairy protein and sources of dairy protein with risks of these common cancers remains unclear. This chapter examines the association between protein from all dairy products and sources in relation to the risks of colorectal, breast, and prostate cancer using a sample of 114,217 UK Biobank participants who completed ≥2 (maximum of 5) 24-hour dietary assessments. After a median of 9.4 years of follow-up, 1,193 colorectal, 2,024 female breast, and 2,422 prostate cancer cases were identified. There were inverse associations of all dairy protein and protein from milk intake with colorectal cancer incidence ($HR_{Q4\ vs\ Q1}$: 0.80; 95% CI: 0.67-0.94 and 0.79; 0.67-0.94, respectively). A borderline positive association between high protein from milk intake and prostate cancer as compared to low intake ($HR_{Q4\ vs\ Q1}$: 1.12; 1.00-1.26) was also observed, although there was no statistically significant linear trend. No significant associations were observed between intake of dairy protein and breast cancer risk. When IGF-I concentrations measured at recruitment were added to the multivariable-adjusted models, associations remained largely unchanged. Other components in dairy products other than protein (i.e., calcium), are likely influencing colorectal cancer risk, however, for prostate cancer, potential explanations for the weak positive association remains unclear. Further research is needed to better understand mechanisms for dairy intake through and separate from the IGF pathway.

6.1 Introduction

Colorectal, breast, and prostate cancer are among the most commonly diagnosed cancers in the world, accounting for approximately 28% of all new cancer cases in 2020 [1]. Dietary intake may influence the risk of developing these types of cancer; however, how the intake of dairy products or sources of dairy influences risk of developing colorectal, breast, or prostate cancer remains unclear.

6.1.1 Previous studies

The most recent WCRF meta-analysis for colorectal cancer, which included studies up until 2017, reported that a higher intake of dairy products is associated with a "probable" lower risk of colorectal cancer [20], possibly due to the high calcium content in dairy products. The findings from the 2017 WCRF meta-analysis on breast cancer classified the evidence for an inverse association between the intake of dairy products and the risk of breast cancer as "limited-suggestive" for premenopausal breast cancer but "limited-no conclusion" for any association for postmenopausal breast cancer [130]. Finally, the latest WCRF meta-analysis for prostate cancer, which included studies until 2014, reported "limited-suggestive" evidence that higher intake of dairy products may be associated with a higher risk of overall prostate cancer [262]. However, the WCRF meta-analyses did not look at protein from dairy products or include separate analyses for different dairy protein sources, and few studies to date have assessed protein from dairy with these cancer sites [388, 402-406]. This may be important as protein, particularly from dairy products, is thought to be an important modifier of circulating IGF-I concentrations [5].

Both observational and MR studies have found that the risks of developing colorectal, breast, and prostate cancer are associated with higher circulating concentrations of IGF-I [11-14, 407]. In Chapter 5, and in other cross-sectional studies and some RCTs, protein intake, particularly from dairy products, has been positively associated with circulating IGF-I concentrations [5, 284, 317]. Moreover, some studies [3, 370] (but not all [366]) that have investigated associations

between dairy product source and IGF-I concentrations suggest that the association might be restricted to protein from milk but not from cheese. Findings from Chapter 5 also suggest that dairy protein, particularly from milk and yogurt, was positively associated with IGF-I concentrations but not protein from cheese. However, it remains unclear if this differing association of dairy product protein intake with IGF-I may be important for the aetiology of IGF-I related cancers.

6.2 Aim

This analysis investigates the associations of intakes of dairy protein and sources of dairy protein with risks of colorectal, breast, and prostate cancer in a subsample of participants who completed at least two 24-hour dietary assessments. A secondary aim was to assess the potential role of IGF-I concentrations in these associations.

6.3 Methods

6.3.1 Study design & participants

Details of the UK Biobank study design, recruitment, and data collection for the UK Biobank are provided in Chapter 3. This analysis included participants from the UK Biobank who completed a minimum of two 24-hour dietary assessments.

6.3.2 Dietary assessment and dairy intake estimation

Protein from dairy products, milk, and cheese were estimated from multiple 24-hour dietary assessment. Details of the Oxford WebQ 24-hour dietary assessments are provided in Chapter 3. Briefly, dairy protein intake was estimated in each 24-hour dietary assessment based on food and beverage intakes reported by participants. From these responses, protein from all dairy products (milk, yogurt, cheese, cream, butter, dairy desserts, and dairy drinks), milk, and cheese were calculated by multiplying the protein content of each respective food and beverage by the frequency of intake using the UK Nutrient Databank food composition tables [361, 362]. Percentages of energy from protein from all dairy sources, and separately from milk and cheese,

were estimated for each 24-hour dietary assessment and then averaged across all available 24-hour dietary assessments. It was not possible to look separately at other dairy sources, such as yogurt, because a large number of participants reported that they did not consume these products on the day(s) of dietary assessment. The estimated absolute g/day of all dairy products, milk, and cheese reported from all 24-hour dietary assessments were also determined for comparison with previous meta-analyses [20, 130, 262]. Total grams of all dairy products, milk, and cheese were calculated based on the reported intake in each 24-hour dietary assessment. The serving size in grams was multiplied by the frequency and amount that was reported in each of the 24-hour dietary assessments to obtain the total grams of dairy products, milk, and cheese

6.3.3 Study sample and exclusions

A total of 922 participants had withdrawn their consent from the UK Biobank and were excluded from this analysis. Participants were also excluded if they had been diagnosed with a prevalent malignant cancer before recruitment (excluding non-melanoma skin cancer (N=29,504)), their genetic sex did not match their reported sex (N=321), or they did not complete a 24-hour dietary assessment (N=251,938). Individual 24-hour dietary assessments were excluded if an unreliable energy intake was reported (men: >17,575 kJ (4200 kcal) or <3347 kJ (800 kcal); women >14,644 kJ (3500 kcal) or <2092 kJ (500 kcal))[360] or the participant reported that they were ill or fasting on the day they completed the questionnaire (N=2,439 reported unreliable energy intake and no longer had a 24-hour dietary assessment, N=592 were ill or fasting and no longer had a valid 24-hour dietary assessment). To reduce random measurement error, participants who did not complete a minimum of two valid 24-hour dietary assessments were excluded (N=100,487). Participants were also excluded if they were censored or had a cancer diagnosis before the completion of their last 24-hour dietary assessment (N=2,897). In total, this left 114,217 participants who completed ≥2 (maximum 5) valid 24-hour dietary assessments, of which 51,278 were men and 62,939 were women. **Figure 6.1** shows the flow diagram of exclusions for these analyses. Of the 114,217 included participants, 44,994

completed two 24-hour dietary assessments, 38,480 completed three, 26,032 completed four and 4,711 completed five.

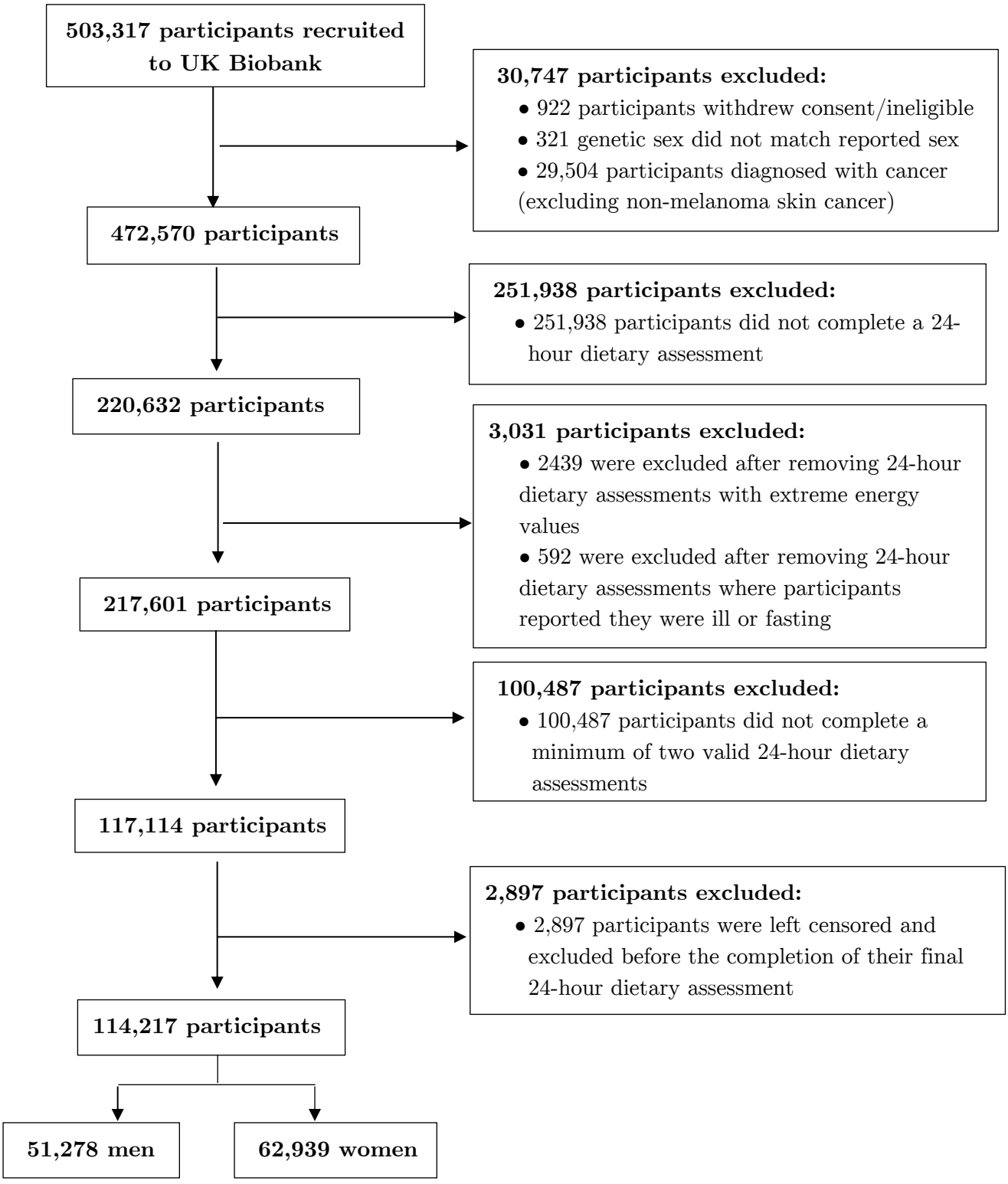


Figure 6.1 Flow diagram showing eligible participants for this study.

6.3.4 Outcome ascertainment: cancer diagnoses

Prevalent and incident cancer diagnoses were determined from a combination of linkage to National Health Service Digital (for participants from England and Wales), HES data for English participants, and NHS Central Register Scotland and Scottish Morbidity Records (SMR) for Scottish participants [353]. Participants contributed follow-up time from the date of completion of their last 24-hour dietary assessment until the date of the first cancer registration (excluding non-melanoma skin cancer (ICD-10: C44)), date of death, or last date of follow-up available from HES and SMR data or the Welsh cancer registry (30th of September 2021 for English participants, 31st of July 2021 for Scottish participants, and 29th of February 2020 for Welsh participants). Cancer registry data were available until 29th of February 2020 for English participants and 31st of January 2021 for Scottish participants, after this time, participants from England and Scotland were followed using HES and SMR databases, respectively. For participants from Wales, hospital episode data were not used because the cancer registry had longer follow-up. Participants were coded as having a diagnosis of cancer based on the ICD-10 codes of their first incident cancer: colorectal cancer (C18-C20), breast cancer (C50), or prostate cancer (C61).

6.3.5 Statistical analysis

Participants were classified into quartiles of percentage of energy intake from protein from all dairy products, protein from milk, and protein from cheese. Intakes of total dairy protein and sources of dairy protein were also modelled as a per 2.5% energy increment (modelled continuously). Baseline characteristics were summarised across quartiles of percentage of energy intake from protein from all dairy products, protein from milk, and protein from cheese.

Cox proportional hazards regressions were used to estimate hazard ratios (HR) and 95% CIs, with age as the underlying time variable. When breast cancer was the outcome, models were restricted to women and when prostate cancer was the outcome, models were restricted to men. The lowest quartile of percentage of energy intake of dairy protein was used as the reference

group for analyses by quartiles of intake. Minimally adjusted models were stratified by sex (only for analyses with colorectal cancer as the outcome) and age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years) and adjusted for region at recruitment (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland).

Multivariable Cox regression analyses were further adjusted for height (six sex-specific categories increasing by 5 cm, and unknown/missing (0.16%)), physical activity (low: 0-9.99, medium: 10-49.99, high: ≥50 excess MET-hours /week, and unknown/missing (1.81%)), Townsend deprivation index (quintiles from most deprived to least deprived, and unknown/missing (0.12%)), education (completion of national exam at 16 years, completion of national exam at 17-18 years, college or university degree, or other/unknown/missing (6.63%)), employment status (employed, retired, not in paid employment, or unknown (0.66%)), smoking status (never, former, light smoker: ≤15 cigarettes/day, medium smoker: 16-29 cigarettes/day, heavy smoker: ≥30 cigarettes/day, or missing/unknown (0.23%)), alcohol consumption (non-drinkers, <1, 1-9.99, 10-19.99, ≥20 grams/day, or unknown/missing (0.36%)), ethnicity (white, mixed race or other, Asian or British Asian, Black or Black British, or missing/unknown (0.5%)), diabetes (yes, no, or unknown (0.33%)), BMI (<20, 20-22.49, 22.5-24.99, 25.0-27.49, 27.5-29.99, 30-32.49, 32.5-34.99, ≥35 kg/m^2, or unknown/missing (0.20%)), and total energy intake measured from the 24-hour dietary assessments (sex-specific quintiles).

For colorectal cancer analyses, models were also adjusted for red and processed meat intake (reported at recruitment: <2 times per week, 2-2.99 times per week, 3-3.99 times per week and 4+ times per week, or unknown (0.41%)), NSAID use (regular user of ibuprofen or aspirin, irregular user, no reported use) and women specific covariates: MHT use (never, former, current, unknown (0.12%)) and menopausal status (premenopausal, postmenopausal, or unknown (5.4%) assigning men to a separate category). For breast cancer analyses, models were additionally adjusted for MHT use (same as above) oral contraceptive use (never, former, current, unknown (0.16%)), age at menarche (≤12 years old, 13 years, ≥14 years, or unknown (21.2%)), parity and age at first birth (nulliparous, 1-2 children <25 years old, 3+ children <25

years old, 1-2 children 25-29.9 years old, 3+ children 25-29.9 years old, 1-2 children 30+ years old, 3+ children 30+ years old, or missing (0.06%)), and menopausal status and BMI (six categories: premenopausal:<25, 25-29.9, 30+ kg/m^2 and postmenopausal: <25, 25-29.9, 30+ kg/m^2 or unknown (5.4%)). For prostate cancer analyses, models were also adjusted for marital status (not living with a partner, living with a partner) [250]. Further information on the classification of covariates can be found in the **Ancillary Methods A.1**. To assess how the addition of confounders changed the association of dairy protein intake with specific cancer sites, χ^2 statistics and p-values for including intake of total dairy protein and dairy protein sources per 2.5% energy increase in the model were estimated using LRT comparing a model without percentage of energy from dairy protein in the model [149]. To assess the potential mediating role of IGF-I, the main multivariable-adjusted Cox regression models with further adjustment for serum IGF-I concentrations (quintiles) measured at recruitment. Assessment of the proportional hazards assumption was evaluated using Schoenfeld residuals and no violations of the assumption were observed ($p>0.05$).

In additional analyses, the associations of intakes of total grams of all dairy products, milk, and cheese with risks of colorectal, breast, and prostate cancer were assessed. Risk analyses looking at a 200 g/day increment in dairy product intake, a 200 g/day increment in milk intake, and a 50 g/day increment in cheese intake were also conducted.

Subgroup and sensitivity analyses

For subgroup analyses, heterogeneity was assessed by BMI (~median; <27 and ≥27 kg/m^2), alcohol intake (<10 g/day and ≥10g/day), and smoking status (ever and never) by using a LRT comparing the main model to a model including an interaction term between dietary intake of dairy protein and the subgroup of interest. For colorectal cancer, heterogeneity by sex was also assessed using a LRT. For breast cancer, heterogeneity by menopausal status was assessed by separating follow-up time at age 55 years and their defined menopause status at recruitment so that women who were premenopausal at recruitment would change into the postmenopausal risk

set at 55 years of age. If a woman was defined as postmenopausal at recruitment, she remained in the postmenopausal risk set regardless of age.

In sensitivity analyses, further adjustments for other components of diet including total fibre intake (sex specific quintiles from ≥2 24-hour dietary assessments) and fruit and vegetable intake (quintiles derived from ≥2 24-hour dietary assessment) as well as red and processed meat intake (from the touchscreen questionnaire) for breast and prostate cancer analyses. Moreover, analyses were restricted to participants who completed a minimum of three 24-hour dietary assessments to reduce random measurement error in the estimation of intake of dairy protein. Finally, to assess for reverse causality, the first two years of follow-up were excluded.

All analyses were conducted using Stata version 17.0 (Stata Corp, TX, United States). P-values were two-sided with $p < 0.05$ being considered statistically significant.

6.4 Results

6.4.1 Participant characteristics and cancer incidence

Over a median of 9.4 years of follow-up, 1,193 incident colorectal cancer, 2,024 breast cancer and 2,422 prostate cancer cases were observed. **Table 6.1** presents baseline characteristics for the highest and lowest quartiles of percentage of energy from protein from all dairy sources, milk, and cheese. **Table 6.2** presents the mean and SD intakes of protein from all dairy products, milk, and cheese by quartiles and separately by men and women. Participants in the highest quartile of total dairy and milk protein intake were more likely to be women, older, be never smokers, consume less alcohol (~8 g/day less), be of white ethnicity, and report a lower energy intake than those in the lowest quartile of intake. Participants in the highest intake of cheese protein were less physically active and more likely to have a university/college degree than those in the lowest quartile. **Figure 6.2** presents the average grams of protein consumed from each dairy, milk, and cheese source.

Table 6.1 Baseline characteristics by lowest and highest quartile of percentage of energy intake from protein from all dairy products, milk, and cheese.

	Protein from dairy products		Protein from milk		Protein from cheese	
	Q1	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**
Number of participants	28,555	28,554	28,555	28,554	30,958	28,554
Sex – Female, N (%)	13,409 (47.0%)	18,423 (64.5%)	15,688 (54.9%)	17,532 (61.4%)	16,162 (52.2%)	16,468 (57.7%)
Age at recruitment - years	55.0 (8.0)	56.6 (7.6)	55.2 (8.0)	56.6 (7.6)	55.7 (7.8)	55.8 (7.8)
Body mass index – kg/m^2	26.9 (4.7)	26.6 (4.6)	26.7 (4.8)	26.6 (4.5)	27.0 (4.6)	26.5 (4.6)
Height - centimetres	170.3 (9.2)	168.1 (8.9)	169.5 (9.2)	168.3 (9.0)	169.3 (9.2)	169.4 (9.1)
Physical activity - High, N (%)	4,944 (17.3%)	4,863 (17.0%)	5,127 (18.0%)	4,758 (16.7%)	5,375 (17.4%)	4,796 (16.8%)
Townsend deprivation index, N (%)						
Q1 - Most affluent	5,745 (20.1%)	6,512 (22.8%)	5,468 (19.1%)	6,792 (23.8%)	6,716 (21.7%)	6,073 (21.3%)
Q5 - Most deprived	5,606 (19.6%)	6,195 (21.7%)	5,447 (19.1%)	6,457 (22.6%)	6,522 (21.1%)	5,798 (20.3%)
Paid employment, N (%)	18,875 (66.1%)	16,780 (58.8%)	18,651 (65.3%)	16,691 (58.5%)	19,345 (62.5%)	18,014 (63.1%)
University/college degree, N (%)	20,793 (72.8%)	20,922 (73.3%)	21,437 (75.1%)	20,365 (71.3%)	21,288 (68.8%)	21,932 (76.8%)
White ethnicity, N (%)	26,923 (94.3%)	27,922 (97.8%)	26,908 (94.2%)	28,021 (98.1%)	29,331 (94.7%)	27,826 (97.5%)
Smoking - Never, N (%)	15,390 (53.9%)	17,192 (60.2%)	15,499 (54.3%)	17,478 (61.2%)	18,010 (58.2%)	16,055 (56.2%)
Diabetes - Yes, N (%)	1,163 (4.1%)	1,040 (3.6%)	987 (3.5%)	1,153 (4.0%)	1,379 (4.5%)	1,072 (3.8%)
NSAID use - regular, N (%)	7,985 (27.9%)	6,950 (24.3%)	7,970 (27.9%)	7,091 (24.8%)	7,654 (24.7%)	7,844 (27.4%)
Living with a partner, N (%)	20,932 (73.3%)	20,914 (73.2%)	20,540 (71.9%)	21,502 (75.3%)	22,647 (73.2%)	20,894 (73.2%)
Women specific variables						
Age at menarche - years	12.6 (2.7)	12.5 (2.6)	12.5 (2.7)	12.6 (2.6)	12.6 (2.7)	12.5 (2.7)
Postmenopausal at recruitment, N (%)	8,308 (62.0%)	13,272 (72.1%)	10,119 (64.5%)	12,621 (72.0%)	10,931 (67.6%)	11,115 (67.5%)
Menopausal hormone therapy use - current, N (%)	1,109 (8.3%)	1,455 (7.9%)	1,334 (8.5%)	1,344 (7.7%)	1,365 (8.4%)	1,301 (7.9%)
Oral contraceptive use - current, N (%)	391 (2.9%)	352 (1.9%)	458 (2.9%)	350 (2.0%)	387 (2.4%)	410 (2.5%)
Nulliparous, N (%)	3,862 (13.5%)	2,371 (8.3%)	3,483 (12.2%)	2,441 (8.5%)	3,595 (11.6%)	2,942 (10.3%)
Diet variables						
Alcohol intake - g/day	20.4 (20.3)	12.2 (12.8)	18.9 (20.3)	12.2 (12.6)	16.2 (18.0)	15.8 (16.2)
Red and processed meat intake – times/week	3.6 (2.3)	3.2 (2.1)	3.3 (2.3)	3.4 (2.0)	3.6 (2.2)	3.2 (2.2)
Vegetable and fruit intake - g/day	374.5 (235.5)	395.4 (224.0)	409.6 (249.6)	365.7 (205.9)	376.1 (235.3)	388.4 (223.0)

Table 6.1 continued

	Protein from dairy products		Protein from milk		Protein from cheese	
	Q1	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**
Total dairy product intake - g/day	169.8 (95.6)	470.5 (175.0)	178.5 (126.5)	456.0 (160.9)	301.7 (172.5)	334.8 (165.5)
Total milk intake - g/day	104.9 (84.0)	279.5 (134.1)	54.6 (48.9)	332.4 (106.3)	195.9 (125.3)	192.3 (121.0)
Total cheese intake - g/day	7.0 (9.4)	29.7 (22.2)	19.1 (19.8)	15.0 (15.9)	0.00 (0.00)	40.4 (17.4)
Total protein intake - % of energy	15.3 (3.1)	16.9 (3.1)	15.4 (3.2)	17.0 (3.0)	16.2 (3.4)	16.1 (3.1)
Dairy product protein - % of energy	1.3 (0.5)	4.8 (1.0)	1.9 (1.2)	4.1 (1.3)	2.2 (1.2)	4.1 (1.3)
Milk protein - % of energy	0.7 (0.5)	2.1 (1.0)	0.3 (0.3)	2.5 (0.7)	1.4 (0.9)	1.3 (0.9)
Cheese protein - % of energy	0.3 (0.4)	1.5 (1.1)	0.9 (0.9)	0.8 (0.8)	0.00 (0.00)	2.0 (0.7)
Total carbohydrate intake - % of energy	48.3 (7.8)	50.3 (7.3)	48.1 (8.2)	51.2 (6.8)	50.8 (7.6)	47.7 (7.5)
Total fat intake - % of energy	31.4 (6.0)	31.3 (5.9)	32.2 (6.2)	30.3 (5.5)	30.0 (5.9)	33.2 (5.8)
Total fibre intake - g/day	18.0 (6.1)	17.4 (5.5)	18.4 (6.3)	16.9 (5.2)	17.3 (5.9)	17.9 (5.7)
Total energy intake - kJ/day	8,860 (2032)	8,094 (1816)	8,775 (2049)	7,847 (1690)	8,224 (1937)	8,607 (1951)

Values are mean (SD) unless otherwise indicated.

Abbreviations: g/day, grams per day; kg/m^2; kilograms per metre squared; kJ/day, kilojoules per day; N, Number of participants; NSAID, non-steroidal anti-inflammatory drug; Q, quantile; y, years.

Table 6.2 Dairy protein intake in grams and percentage of energy per day by quartiles of intake in all participants and separately by males and females.

	All				Male				Female			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Protein from dairy products												
N	28,555	28,554	28,554	28,554	12,820	12,819	12,820	12,819	15,735	15,735	15,735	15,734
g/d	6.96 (3.25)	12.58 (3.09)	16.44 (3.85)	22.42 (6.23)	6.93 (3.15)	12.54 (3.04)	16.63 (3.86)	23.18 (6.55)	6.87 (3.17)	12.49 (2.81)	16.24 (3.51)	22.07 (5.80)
% of total energy intake/d	1.31 (0.51)	2.41 (0.24)	3.25 (0.26)	4.75 (0.96)	1.22 (0.47)	2.23 (0.23)	3.02 (0.25)	4.46 (0.92)	1.41 (0.56)	2.58 (0.25)	3.43 (0.27)	4.95 (0.97)
Protein from milk												
N	28,555	28,554	28,554	28,554	12,820	12,819	12,820	12,819	15,735	15,735	15,735	15,734
g/d	1.87 (1.67)	5.84 (1.55)	8.17 (1.83)	11.38 (3.57)	2.22 (1.76)	6.12 (1.53)	8.39 (1.84)	11.83 (3.93)	1.58 (1.52)	5.55 (1.43)	7.97 (1.68)	11.10 (3.18)
% of total energy intake/d	0.35 (0.29)	1.10 (0.16)	1.62 (0.16)	2.52 (0.64)	0.38 (0.28)	1.07 (0.14)	1.54 (0.14)	2.40 (0.65)	0.32 (0.29)	1.13 (0.17)	1.68 (0.16)	2.60 (0.63)
Protein from cheese												
N												
g/d	30,958	26,151	28,554	28,554	14,796	10,843	12,820	12,819	16,162	15,308	15,735	15,734
% of total energy intake/d	0.00 (0.00)	2.37 (1.03)	5.12 (1.57)	10.33 (4.25)	0.00 (0.00)	2.46 (1.03)	5.32 (1.63)	11.03 (4.51)	0.00 (0.00)	2.29 (0.99)	4.94 (1.42)	9.81 (3.92)

Values represent mean (SD). Grams per day represent the average grams of protein consumed from dairy source within each quantile. % of total energy represents the average percentage of energy consumed from dairy protein source within each quantile.

Abbreviations: g/d, grams per day; Q, quantile; SD, standard deviation.

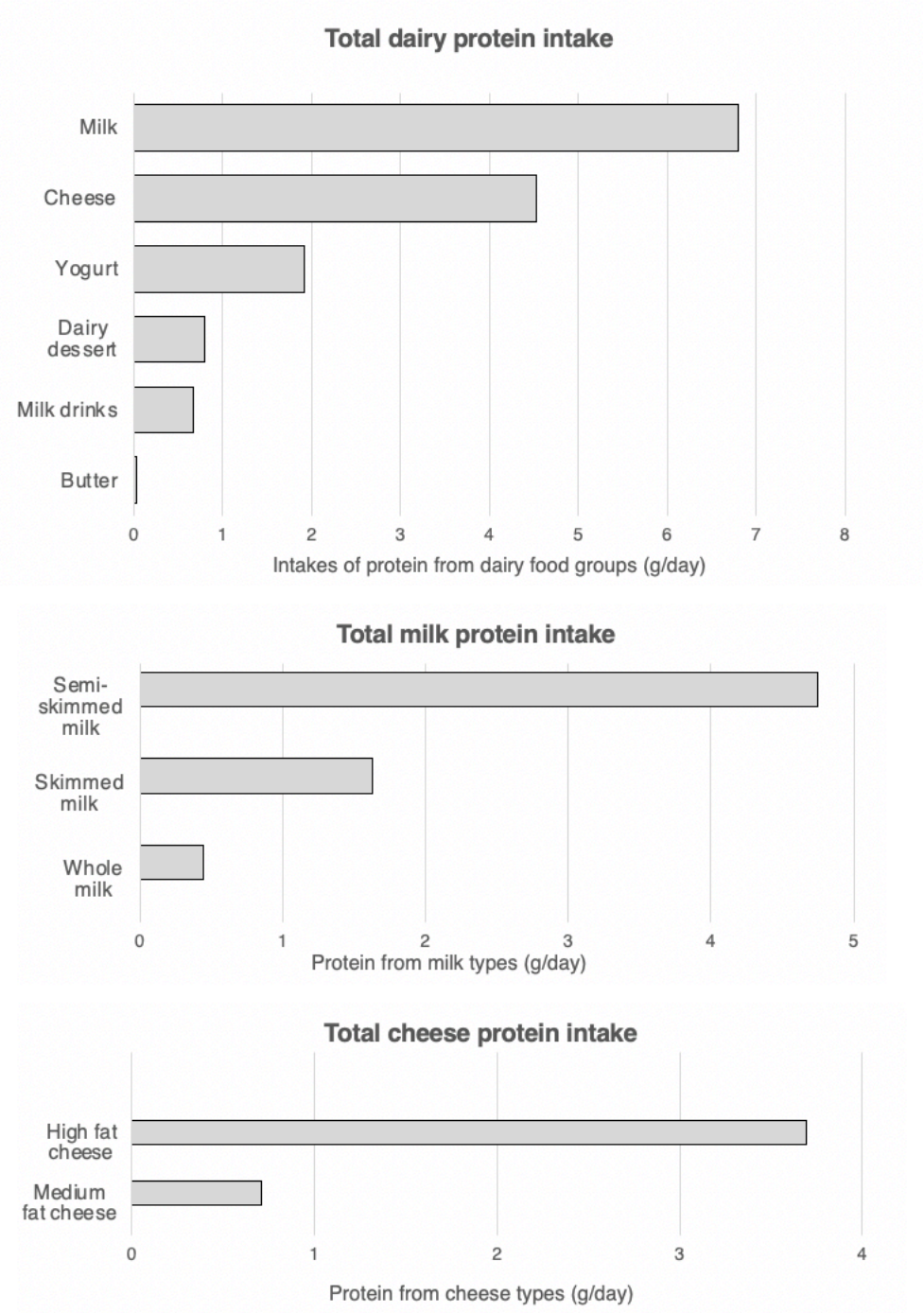


Figure 6.2 Dairy food group (g/day of protein) contributors for total dairy protein, milk protein and cheese protein intake.

Abbreviations: g/day, grams per day.

6.4.2 Associations between dairy protein and cancer risk

Tables 6.3 - 6.5 present the HRs and 95% CIs for the minimally adjusted models with sequential adjustment of potential confounders for dairy protein intake and risks of colorectal, breast, and prostate cancer, while the fully adjusted models (main results) are shown in **Figure 6.3**. Protein from dairy and milk were inversely associated with colorectal cancer risk in minimally adjusted models and only became slightly attenuated with additional adjustment of confounders (**Table 6.3**). No associations were observed between protein from dairy products or dairy sources and risk of breast cancer in minimally adjusted models (**Table 6.4**). For prostate cancer, protein from milk was not associated with prostate cancer risk in minimally adjusted models but with additional adjustment of confounders, the HR between Q4 and Q1 became significant (HR: 1.12, 95% CI: 1.00-1.26; **Table 6.5**). In fully adjusted models, intakes of protein from all dairy sources and from milk were inversely associated with colorectal cancer risk ($HR_{Q4\ vs.\ Q1}$: 0.80, 95% CI: 0.67-0.94, $p\text{-}_{trend}$=0.001 and 0.79, 0.67-0.94, $p\text{-}_{trend}$=0.003, respectively; in **Figure 6.3**) whereas no association was observed for the intake of protein from cheese. For breast cancer risk, no associations were observed across intakes of protein from all dairy products or dairy sources. For prostate cancer, no clear associations were observed for intakes of protein from all dairy products or cheese. However, a positive association was suggested for men in the highest quartile of protein from milk intake compared to the lowest quartile ($HR_{Q4\ vs\ Q1}$: 1.12, 95% CI: 1.00-1.26), although this association was not statistically significant when milk protein was modelled as a continuous variable ($HR_{per\ 2.5\%\ energy\ increment}$: 1.11 95% CI: 0.98-1.26; **Figure 6.3**).

When IGF-I concentrations measured at recruitment were added to the multivariable-adjusted models, associations remained largely unchanged (**Table 6.6**)

In additional analyses looking at the association of intake of total g/day of all dairy products, milk, and cheese with colorectal, breast, and prostate cancer risk, the associations were generally similar to those from analyses looking at protein from all dairy and dairy sources expressed as a

2.5% energy increment. However, there was a suggestion of an inverse association between g/day of total dairy intake and breast cancer risk when comparing the highest quartile to the lowest quartile of intake ($HR_{Q4\ vs\ Q1}$: 0.88, 0.77-1.00), although there was no significant trend ($HR_{\text{per 200 g/day of increase}}$: 0.90, 0.81-1.01; **Figure 6.4**).

Table 6.3 Hazard ratios and 95% confidence intervals for minimally adjusted and sequential adjustment models for protein from all dairy products, milk, and cheese and risk of colorectal cancer.

	Colorectal cancer							
	Q1	**Q2**	**Q3**	**Q4**	**Per 2.5% energy increase**	χ^2	χ^2 - % change	**P-trend**
Protein from dairy products								
Minimally-adjusted model	1 (ref)	0.90 (0.77 - 1.05)	0.85 (0.73 - 1.00)	**0.75 (0.63 - 0.88)**	**0.79 (0.71 - 0.88)**	17.64	-	<0.001
Multivariable-adjusted model 1	1 (ref)	0.89 (0.77 - 1.04)	0.85 (0.73 - 1.00)	**0.76 (0.64 - 0.90)**	**0.80 (0.72 - 0.90)**	15.62	-11.4%	<0.001
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	0.91 (0.78 - 1.07)	0.88 (0.75 - 1.03)	**0.79 (0.67 - 0.94)**	**0.82 (0.74 - 0.92)**	11.74	-33.4%	0.001
Multivariable-adjusted model 3 (+ BMI)	1 (ref)	0.91 (0.78 - 1.07)	0.88 (0.75 - 1.03)	**0.79 (0.67 - 0.94)**	**0.82 (0.74 - 0.92)**	11.84	-32.9%	0.001
Multivariable-adjusted model 4 (+ red and processed meat; fully adjusted)	1 (ref)	0.91 (0.78 - 1.06)	0.88 (0.75 - 1.03)	**0.80 (0.67 - 0.94)**	**0.82 (0.74 - 0.92)**	11.61	-34.2%	0.001
Protein from milk								
Minimally-adjusted model	1 (ref)	0.88 (0.76 - 1.03)	**0.79 (0.67 - 0.92)**	**0.75 (0.64 - 0.89)**	**0.72 (0.61 - 0.86)**	13.60	-	<0.001
Multivariable-adjusted model 1	1 (ref)	0.87 (0.74 - 1.01)	**0.78 (0.67 - 0.92)**	**0.77 (0.65 - 0.91)**	**0.74 (0.62 - 0.88)**	11.66	-14.3%	0.001
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	0.88 (0.75 - 1.02)	**0.80 (0.69 - 0.94)**	**0.80 (0.68 - 0.95)**	**0.77 (0.64 - 0.92)**	8.27	-39.2%	0.004
Multivariable-adjusted model 3 (+ BMI)	1 (ref)	0.88 (0.75 - 1.02)	**0.81 (0.69 - 0.94)**	**0.80 (0.68 - 0.95)**	**0.77 (0.64 - 0.92)**	8.23	-39.5%	0.004
Multivariable-adjusted model 4 (+ red and processed meat; fully adjusted)	1 (ref)	0.87 (0.74 - 1.01)	**0.80 (0.68 - 0.93)**	**0.79 (0.67 - 0.94)**	**0.76 (0.63 - 0.91)**	9.01	-33.7%	0.003
Protein from cheese								
Minimally-adjusted model	1 (ref)	0.94 (0.80 - 1.10)	0.98 (0.84 - 1.15)	0.91 (0.78 - 1.07)	0.91 (0.77 - 1.08)	1.08	-	0.30
Multivariable-adjusted model 1	1 (ref)	0.91 (0.77 - 1.06)	0.95 (0.81 - 1.11)	0.89 (0.75 - 1.04)	0.90 (0.76 - 1.07)	1.44	+32.7%	0.23
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	0.91 (0.77 - 1.06)	0.95 (0.81 - 1.11)	0.88 (0.75 - 1.04)	0.89 (0.75 - 1.06)	1.64	+51.6%	0.20
Multivariable-adjusted model 3 (+ BMI)	1 (ref)	0.91 (0.77 - 1.07)	0.95 (0.81 - 1.11)	0.89 (0.76 - 1.04)	0.90 (0.76 - 1.07)	1.52	+39.9%	0.22
Multivariable-adjusted model 4 (+ red and processed meat; fully adjusted)	1 (ref)	0.91 (0.77 - 1.07)	0.95 (0.82 - 1.12)	0.89 (0.76 - 1.05)	0.91 (0.76 - 1.08)	1.25	+15.7	0.26

Estimates in bold indicate statistically significant results.

Minimally adjusted model had age as the underlying time variable and analyses were stratified by sex and age groups and adjusted for region of recruitment.

Multivariable-adjusted model 1: further adjusted for height, physical activity, Townsend deprivation index, education, employment status, ethnicity, diagnosis of diabetes, non-steroidal anti-inflammatory drug use, menopause status (women only), menopausal hormone therapy use (women only), and energy intake.

Multivariable-adjusted model 2: further adjusted for smoking status and alcohol intake in the multivariable adjusted model 1.

Multivariable-adjusted model 3: further adjusted for BMI from the multivariable adjusted model 2.

Multivariable-adjusted model 4: further adjusted for red and processed meat intake reported at recruitment from the multivariable adjusted model 3.

Abbreviations: BMI, body mass index; Q, quantile.

Table 6.4 Hazard ratios and 95% confidence intervals for minimally adjusted and sequential adjustment models for protein from all dairy products, milk, and cheese and risk of breast cancer.

	Breast cancer							
	Q1	**Q2**	**Q3**	**Q4**	**Per 2.5% energy increase**	**χ^2**	**χ^2 - % change**	**P-trend**
Protein from dairy products								
Minimally-adjusted model	1 (ref)	1.03 (0.91 - 1.16)	0.98 (0.86 - 1.11)	0.92 (0.82 - 1.05)	0.93 (0.86 - 1.01)	3.07	-	0.08
Multivariable-adjusted model 1	1 (ref)	1.03 (0.92 - 1.17)	0.98 (0.87 - 1.11)	0.93 (0.82 - 1.06)	0.94 (0.86 - 1.01)	2.75	-10.4%	0.10
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	1.04 (0.92 - 1.17)	0.99 (0.87 - 1.12)	0.93 (0.82 - 1.06)	0.94 (0.87 - 1.01)	2.67	-12.9%	0.10
Multivariable-adjusted model 3 (+ BMI; fully adjusted)	1 (ref)	1.04 (0.92 - 1.17)	0.99 (0.87 - 1.12)	0.93 (0.82 - 1.06)	0.94 (0.86 - 1.01)	2.72	-11.5%	0.10
Protein from milk								
Minimally-adjusted model	1 (ref)	1.00 (0.89 - 1.13)	0.98 (0.87 - 1.11)	0.92 (0.82 - 1.05)	0.90 (0.79 - 1.02)	2.97	-	0.08
Multivariable-adjusted model 1	1 (ref)	1.00 (0.88 - 1.13)	1.00 (0.88 - 1.13)	0.94 (0.83 - 1.07)	0.92 (0.81 - 1.04)	1.90	-36.1%	0.17
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	1.00 (0.89 - 1.13)	1.00 (0.88 - 1.13)	0.94 (0.83 - 1.07)	0.91 (0.81 - 1.04)	1.95	-34.5%	0.16
Multivariable-adjusted model 3 (+ BMI; fully adjusted)	1 (ref)	1.00 (0.88 - 1.13)	0.99 (0.88 - 1.12)	0.94 (0.83 - 1.07)	0.91 (0.80 - 1.03)	2.09	-29.6%	0.15
Protein from cheese								
Minimally-adjusted model	1 (ref)	1.00 (0.88 - 1.13)	1.04 (0.92 - 1.17)	1.05 (0.93 - 1.18)	1.04 (0.92 - 1.17)	0.34	-	0.56
Multivariable-adjusted model 1	1 (ref)	1.00 (0.88 - 1.13)	1.04 (0.92 - 1.18)	1.04 (0.92 - 1.18)	1.03 (0.91 - 1.16)	0.19	-43.7%	0.66
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	1.00 (0.88 - 1.14)	1.04 (0.92 - 1.18)	1.04 (0.92 - 1.18)	1.02 (0.91 - 1.16)	0.15	-56.4%	0.70
Multivariable-adjusted model 3 (+ BMI; fully adjusted)	1 (ref)	1.00 (0.88 - 1.14)	1.04 (0.92 - 1.18)	1.04 (0.92 - 1.18)	1.03 (0.91 - 1.16)	0.21	-38.9%	0.65

Minimally adjusted model had age as the underlying time variable and analyses were stratified by age groups and adjusted for region of recruitment.

Multivariable-adjusted model 1: further adjusted for height, physical activity, Townsend deprivation index, education, employment, ethnicity, diagnosis of diabetes, menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, and energy intake in the minimally adjusted model.

Multivariable-adjusted model 2: further adjusted for smoking status and alcohol intake in the multivariable adjusted model.

Multivariable-adjusted model 3: further adjusted for BMI and menopausal status from the multivariable adjusted model 2. Abbreviations: BMI, body mass index; Q, quantile.

Table 6.5 Hazard ratios and 95% confidence intervals for minimally adjusted and sequential adjustment models for protein from all dairy products, milk, and cheese and risk of prostate cancer.

	Prostate cancer							
	Q1	**Q2**	**Q3**	**Q4**	**Per 2.5% energy increase**	**χ^2**	**χ^2 - % change**	**P-trend**
Protein from dairy products								
Minimally-adjusted model	1 (ref)	1.01 (0.89 - 1.13)	1.08 (0.96 - 1.21)	1.04 (0.92 - 1.16)	1.03 (0.95 - 1.11)	0.58	-	0.45
Multivariable-adjusted model 1	1 (ref)	1.00 (0.89 - 1.12)	1.08 (0.96 - 1.21)	1.04 (0.93 - 1.17)	1.04 (0.96 - 1.12)	0.80	+37.2%	0.37
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	1.00 (0.89 - 1.12)	1.08 (0.96 - 1.21)	1.04 (0.93 - 1.17)	1.04 (0.96 - 1.12)	0.87	+49.8%	0.35
Multivariable-adjusted model 3 (+ BMI; fully adjusted)	1 (ref)	1.00 (0.89 - 1.12)	1.08 (0.96 - 1.21)	1.04 (0.93 - 1.17)	1.04 (0.96 - 1.12)	0.85	+46.9%	0.36
Protein from milk								
Minimally-adjusted model	1 (ref)	1.04 (0.92 - 1.16)	0.99 (0.89 - 1.12)	1.10 (0.99 - 1.24)	1.09 (0.97 - 1.23)	1.96	-	0.16
Multivariable-adjusted model 1	1 (ref)	1.03 (0.92 - 1.16)	1.00 (0.89 - 1.12)	**1.12 (1.00 - 1.26)**	1.11 (0.98 - 1.26)	2.73	+39.0%	0.10
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	1.03 (0.92 - 1.16)	0.99 (0.88 - 1.12)	**1.12 (1.00 - 1.26)**	1.12 (0.98 - 1.27)	2.88	+46.7%	0.09
Multivariable-adjusted model 3 (+ BMI; fully adjusted)	1 (ref)	1.03 (0.92 - 1.15)	0.99 (0.88 - 1.11)	**1.12 (1.00 - 1.26)**	1.11 (0.98 - 1.26)	2.74	+39.3%	0.10
Protein from cheese								
Minimally-adjusted model	1 (ref)	1.01 (0.90 - 1.13)	1.04 (0.93 - 1.16)	0.97 (0.87 - 1.09)	0.95 (0.85 - 1.07)	0.63	-	0.43
Multivariable-adjusted model 1	1 (ref)	0.99 (0.88 - 1.11)	1.03 (0.92 - 1.15)	0.96 (0.86 - 1.08)	0.95 (0.85 - 1.08)	0.60	-4.8%	0.44
Multivariable-adjusted model 2 (+ smoking and alcohol intake)	1 (ref)	0.99 (0.88 - 1.11)	1.03 (0.92 - 1.15)	0.96 (0.86 - 1.08)	0.96 (0.85 - 1.08)	0.56	-12.0%	0.46
Multivariable-adjusted model 3 (+ BMI; fully adjusted)	1 (ref)	0.99 (0.88 - 1.11)	1.03 (0.92 - 1.15)	0.97 (0.86 - 1.08)	0.96 (0.85 - 1.08)	0.52	-17.3%	0.47

Estimates in bold indicate statistically significant results.

Minimally adjusted model had age as the underlying time variable and analyses were stratified by sex and age groups and adjusted for region of recruitment.

Multivariable-adjusted model 1: further adjusted for height, physical activity, Townsend deprivation index, education, ethnicity, diabetes status, marital status, and energy intake from the minimally adjusted model.

Multivariable-adjusted model 2: further adjusted for smoking status and alcohol intake in the multivariable adjusted model.

Multivariable-adjusted model 3: further adjusted for BMI in the multivariable adjusted model 2.

Abbreviations: BMI, body mass index; Q, quantile.

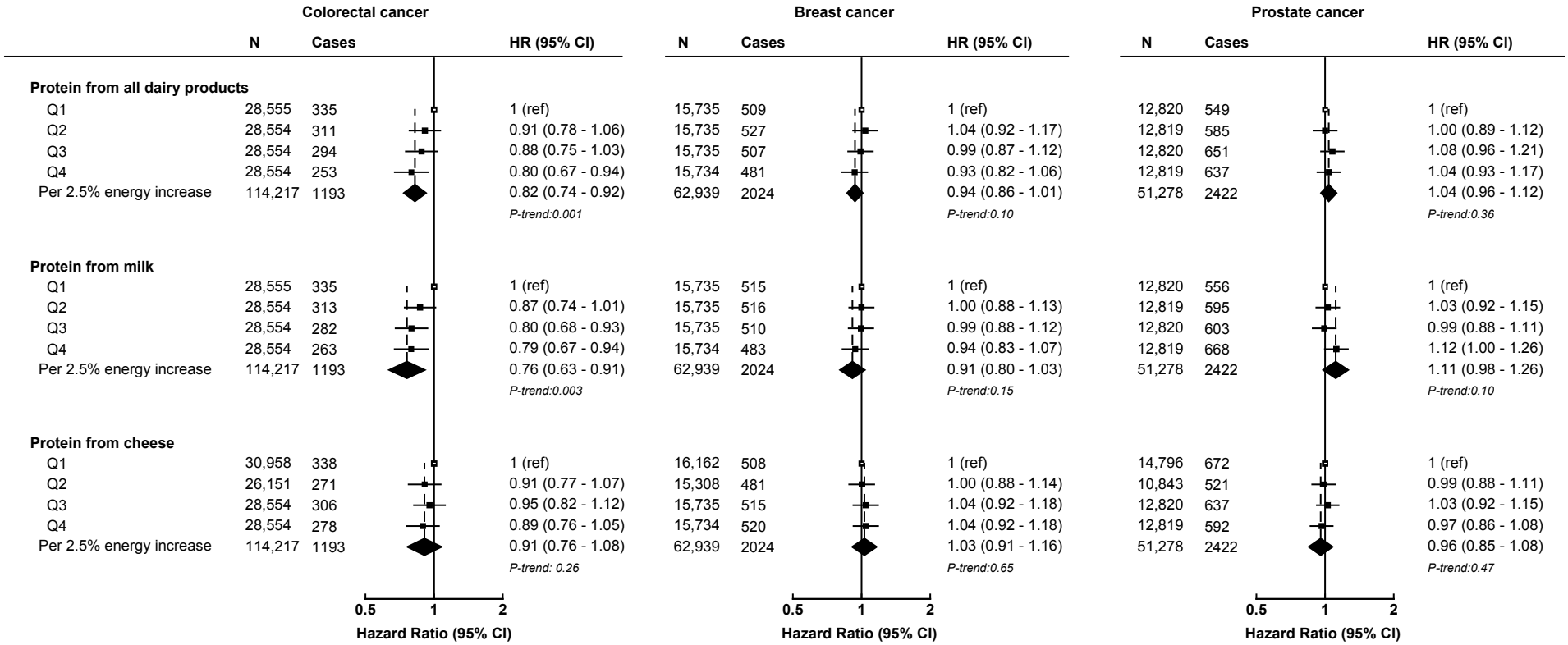


Figure 6.3 Multivariable-adjusted hazard ratios (95% CI) for the association between protein from all dairy products and dairy sources and risks of colorectal, breast, and prostate cancer.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and were further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol intake, ethnicity, diagnosis of diabetes, BMI, energy intake.

For prostate cancer analyses: models were further adjusted for marital status.

For breast cancer analyses: models were further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status.

For colorectal cancer analyses: all models were stratified by sex and adjusted for menopausal status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use.

Abbreviations: BMI, body mass index; HR, hazard ratio; g/day, grams per day; CI, confidence intervals; N, number of participants; Q, quantile.

Table 6.6 Hazard ratios (95% CI) for intake of protein from dairy products and dairy sources with colorectal, breast, and prostate cancer with additional adjustments for circulating insulin-like growth factor-I (IGF-I).

	Colorectal cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from dairy products							
Fully adjusted model	1 (ref)	0.91 (0.78 - 1.06)	0.88 (0.75 - 1.03)	**0.80 (0.67 - 0.94)**	**0.82 (0.74 - 0.92)**	11.61	0.001
Fully adjusted model + IGF-I concentrations	1 (ref)	0.91 (0.78 - 1.06)	0.88 (0.75 - 1.03)	**0.80 (0.67 - 0.94)**	**0.82 (0.74 - 0.92)**	11.58	0.001
Protein from milk							
Fully adjusted model	1 (ref)	0.87 (0.74 - 1.01)	**0.80 (0.68 - 0.93)**	**0.79 (0.67 - 0.94)**	**0.76 (0.63 - 0.91)**	9.01	0.003
Fully adjusted model + IGF-I concentrations	1 (ref)	0.87 (0.74 - 1.01)	**0.80 (0.68 - 0.93)**	**0.79 (0.67 - 0.94)**	**0.76 (0.63 - 0.91)**	9.02	0.003
Protein from cheese							
Fully adjusted model	1 (ref)	0.91 (0.77 - 1.07)	0.95 (0.82 - 1.12)	0.89 (0.76 - 1.05)	0.91 (0.76 - 1.08)	1.25	0.26
Fully adjusted model + IGF-I concentrations	1 (ref)	0.91 (0.77 - 1.07)	0.96 (0.82 - 1.12)	0.89 (0.76 - 1.05)	0.91 (0.76 - 1.08)	1.24	0.26

	Breast cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from dairy products							
Fully adjusted model	1 (ref)	1.04 (0.92 - 1.17)	0.99 (0.87 - 1.12)	0.93 (0.82 - 1.06)	0.94 (0.86 - 1.01)	2.72	0.10
Fully adjusted model + IGF-I concentrations	1 (ref)	1.04 (0.92 - 1.17)	0.99 (0.87 - 1.12)	0.93 (0.82 - 1.05)	0.93 (0.86 - 1.01)	3.02	0.08
Protein from milk							
Fully adjusted model	1 (ref)	1.00 (0.88 - 1.13)	0.99 (0.88 - 1.12)	0.94 (0.83 - 1.07)	0.91 (0.80 - 1.03)	2.09	0.15
Fully adjusted model + IGF-I concentrations	1 (ref)	0.99 (0.88 - 1.12)	0.99 (0.88 - 1.12)	0.92 (0.81 - 1.05)	0.90 (0.79 - 1.02)	2.77	0.10

Table 6.6 continued

	Breast cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from cheese							
Fully adjusted model	1 (ref)	1.00 (0.88 - 1.14)	1.04 (0.92 - 1.18)	1.04 (0.92 - 1.18)	1.03 (0.91 - 1.16)	0.21	0.65
Fully adjusted model + IGF-I concentrations	1 (ref)	0.99 (0.87 - 1.13)	1.04 (0.92 - 1.18)	1.04 (0.92 - 1.18)	1.03 (0.91 - 1.17)	0.27	0.60
	Prostate cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from dairy products							
Fully adjusted model	1 (ref)	1.00 (0.89 - 1.12)	1.08 (0.96 - 1.21)	1.04 (0.93 - 1.17)	1.04 (0.96 - 1.12)	0.85	0.36
Fully adjusted model + IGF-I concentrations	1 (ref)	1.00 (0.89 - 1.12)	1.07 (0.96 - 1.20)	1.03 (0.92 - 1.16)	1.03 (0.96 - 1.12)	0.67	0.41
Protein from milk							
Fully adjusted model	1 (ref)	1.03 (0.92 - 1.15)	0.99 (0.88 - 1.11)	**1.12 (1.00 - 1.26)**	1.11 (0.98 - 1.26)	2.74	0.10
Fully adjusted model + IGF-I concentrations	1 (ref)	1.03 (0.91 - 1.15)	0.99 (0.88 - 1.11)	1.11 (0.99 - 1.25)	1.11 (0.97 - 1.26)	2.40	0.12
Protein from cheese							
Fully adjusted model	1 (ref)	0.99 (0.88 - 1.11)	1.03 (0.92 - 1.15)	0.97 (0.86 - 1.08)	0.96 (0.85 - 1.08)	0.52	0.47
Fully adjusted model IGF-I concentrations	1 (ref)	0.99 (0.88 - 1.11)	1.03 (0.92 - 1.15)	0.97 (0.86 - 1.08)	0.96 (0.85 - 1.08)	0.53	0.47

Estimates in bold indicate statistically significant results.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and further adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol intake, ethnicity, diagnosis of diabetes, BMI, energy intake.

For colorectal cancer analyses: all models were stratified by sex and adjusted for menopausal status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use. For breast cancer analyses: models were further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopausal status. For prostate cancer analyses: models were further adjusted for marital status.

Fully adjusted model + IGF-I concentrations were further adjusted serum circulating insulin-like growth factor-I concentrations measured at recruitment.

Abbreviations: CI, confidence intervals; BMI, body mass index; IGF-I, insulin-like growth factor-I; Q, quantile.

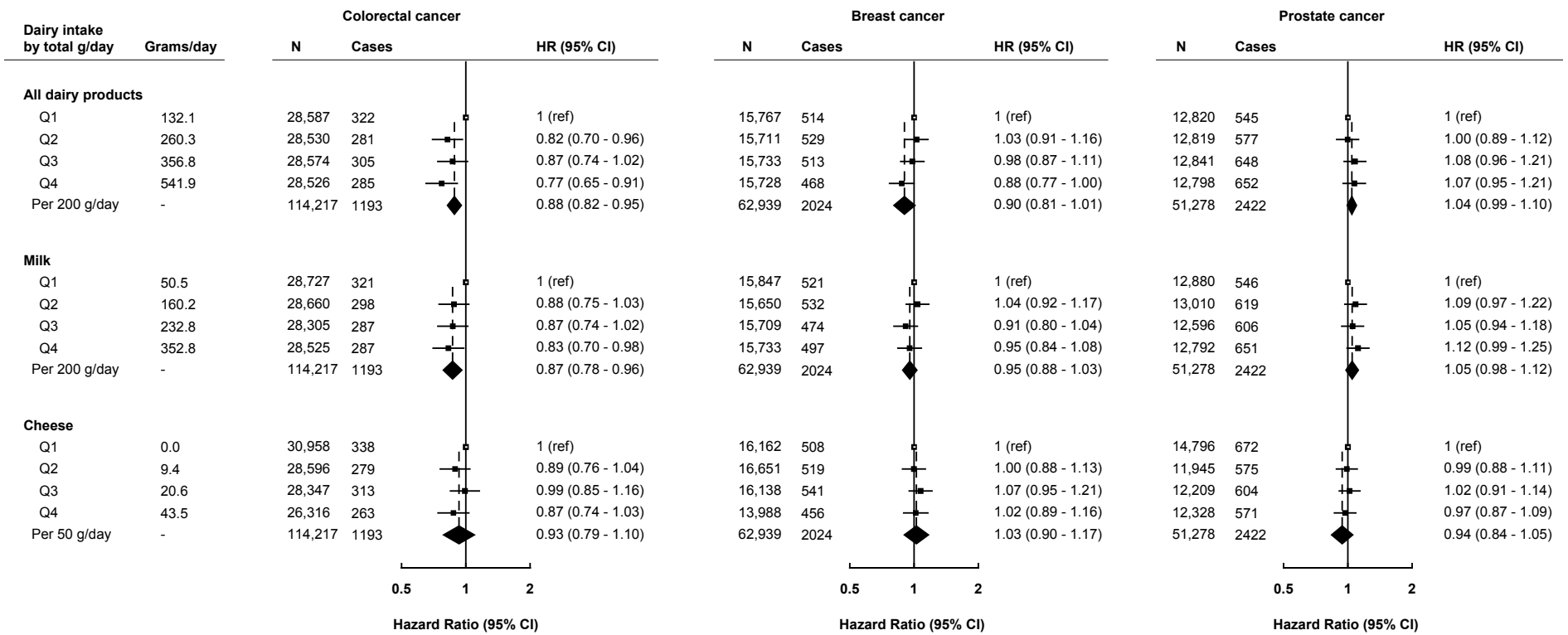


Figure 6.4 Multivariable-adjusted hazard ratios (95% CI) for total grams of dairy products, milk, and cheese with risks of colorectal, breast and prostate cancer.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol intake, ethnicity, diagnosis of diabetes, BMI, energy intake.

For colorectal cancer analyses: all models were stratified by sex and adjusted for menopausal status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use.

For breast cancer analyses: models further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopausal status.

For prostate cancer analyses: models further adjusted for marital status.

Abbreviations: BMI, body mass index; HR, hazard ratio; g/day, grams per day; CI, confidence interval; N, number of participants; Q, quantile.

6.4.3 Subgroup and sensitivity analyses

For colorectal cancer risk, evidence of heterogeneity was observed by alcohol intake for protein from all dairy products; an inverse association was observed only for those consuming ≥10 g/day of alcohol (HR $_{\text{per 2.5\% energy}}$: 0.73, 0.63-0.86) but no association was observed for those consuming <10 g/day of alcohol (0.94, 0.80-1.10; p-$_{\text{het}}$=0.026; **Figure 6.5**). Similar evidence of heterogeneity was observed for intakes of protein from milk by alcohol intake, and also by smoking status; for individuals who consumed ≥10 g/day of alcohol, and for those who were ever smokers, intake of protein from milk was inversely associated with colorectal cancer risk (HR $_{\text{per 2.5\% energy}}$: 0.61, 0.47-0.79; p-$_{\text{het}}$=0.022 and 0.60, 0.46-0.78; p-$_{\text{het}}$=0.009, respectively) but no association was observed for those who consumed <10 g/day of alcohol or were never smokers (**Figure 6.6**). For breast and prostate cancer analyses, no evidence of heterogeneity was observed between subgroups for protein from dairy products or milk (**Figure 6.5 & Figure 6.6**). For breast and prostate cancer analyses, no evidence of heterogeneity by alcohol, smoking status, BMI, or menopause status (only for breast cancer) was observed (**Figure 6.5 & Figure 6.6;** Ancillary Figure 6.1 for protein from cheese intake).

When adjustments were made for other dietary factors in sensitivity analyses, associations remained largely the same with the exception that the association for protein from milk and prostate cancer risk was attenuated and the HR for Q4 vs. Q1 was no longer significant (1.10, 0.98-1.24; **Table 6.7**), which was attributed to red and processed meat in the model being positively associated with prostate cancer risk. In analyses restricted to participants who completed a minimum of three 24-hour dietary assessments, associations remained largely unchanged but with wider CIs probably due to the loss of participants and cases (Ancillary Table 6.1). No material differences in associations were observed in analyses removing the first two years of follow-up (Ancillary Table 6.2).

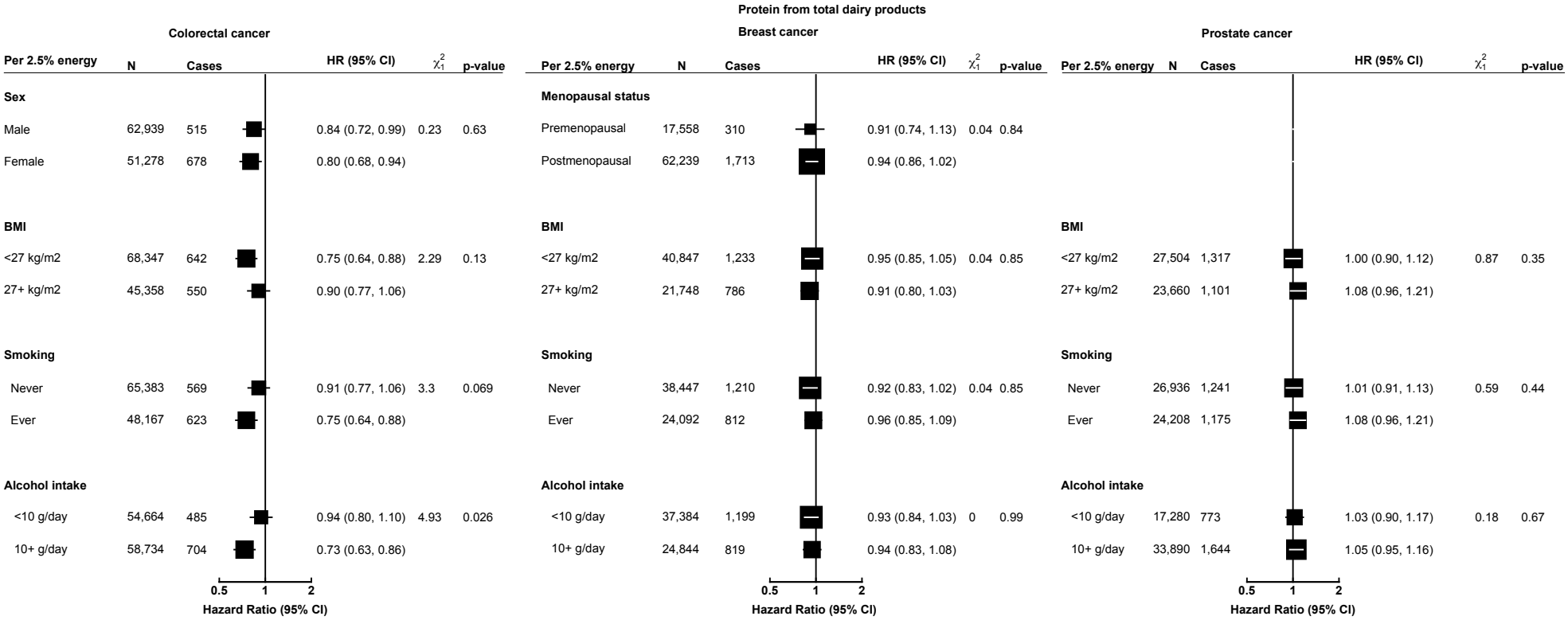


Figure 6.5 Intake of protein from total dairy products per 2.5% energy increment by subgroups and risk of colorectal cancer, breast cancer, prostate cancer.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking status was subgroup if interest), alcohol intake (except when alcohol was the subgroup of interest), ethnicity, diagnosis of diabetes, BMI (except when BMI was subgroup of interest), and energy intake.

For colorectal cancer analyses: all models stratify by sex (except when sex was subgroup of interest) and adjust for menopause status (women only) and menopausal hormone therapy use (women only), red and processed meat intake, non-steroidal anti-inflammatory drug use. For breast cancer analyses: models adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status (except when menopause status was subgroup of interest). For prostate cancer analyses: models further adjusted for marital status. χ^2 and p-value represents improvement of fit obtained from likelihood ratio tests for including an interaction term between subgroup of interest and all dairy protein (modelled as a 2.5% energy increase) into the model.

Abbreviations: BMI, body mass index; CI, confidence interval; g/day, grams per day; HR, hazard ratio; kg/m^2, kilograms per metre squared; N, number of participants.

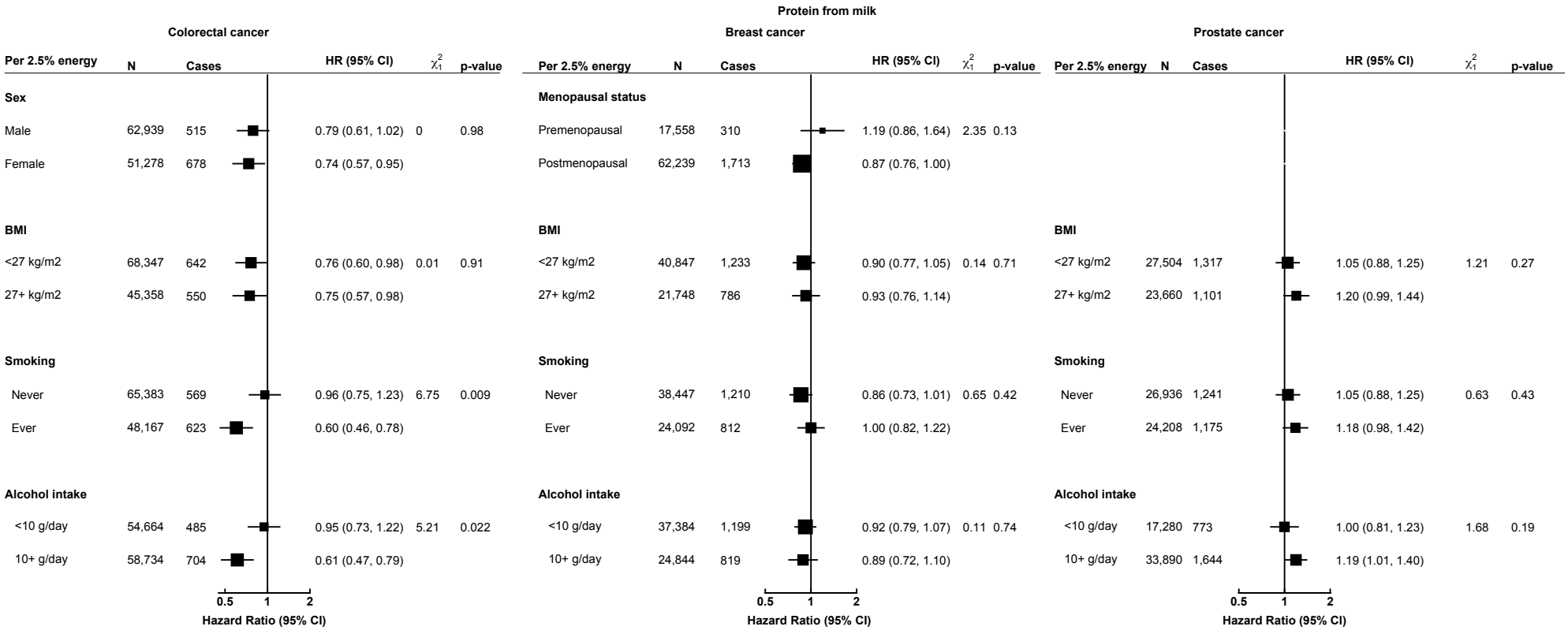


Figure 6.6 Intake of milk protein per 2.5% energy increment by subgroups and risk of colorectal cancer, breast cancer, prostate cancer.

All models used age as the underlying time variable and were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking status was subgroup if interest), alcohol intake (except when alcohol was the subgroup of interest), ethnicity, diagnosis of diabetes, BMI (except when BMI was subgroup of interest), and energy intake. For colorectal cancer analyses: all models were stratified by sex (except when sex was subgroup of interest) and adjusted for menopause status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use. For breast cancer analyses: models were further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status (except when menopause status was subgroup of interest). For prostate cancer analyses: models were further adjusted for marital status. χ^2 and p-value represents improvement of fit obtained from likelihood ratio tests for including an interaction term between subgroup of interest and milk protein (modelled as a 2.5% energy increase) into the model.

Abbreviations: BMI, body mass index; CI, confidence interval; g/day, grams per day; HR, hazard ratio; kg/m^2, kilograms per metre squared; N, number of participants.

Table 6.7 Hazard ratios (95% CI) for intake of protein from dairy products and dairy sources with colorectal, breast, and prostate cancer with additional adjustments for other dietary factors.

	Colorectal cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from dairy products							
Fully adjusted model	1 (ref)	0.91 (0.78 - 1.06)	0.88 (0.75 - 1.03)	**0.80 (0.67 - 0.94)**	**0.82 (0.74 - 0.92)**	11.61	0.001
Fully adjusted model + dietary factors	1 (ref)	0.91 (0.78 - 1.06)	0.88 (0.75 - 1.03)	**0.80 (0.67 - 0.94)**	**0.83 (0.74 - 0.93)**	10.43	0.001
Protein from milk							
Fully adjusted model	1 (ref)	0.87 (0.74 - 1.01)	**0.80 (0.68 - 0.93)**	**0.79 (0.67 - 0.94)**	**0.76 (0.63 - 0.91)**	9.01	0.003
Fully adjusted model + dietary factors	1 (ref)	0.86 (0.74 - 1.01)	**0.79 (0.67 - 0.93)**	**0.79 (0.67 - 0.93)**	**0.77 (0.64 - 0.92)**	8.02	0.005
Protein from cheese							
Fully adjusted model	1 (ref)	0.91 (0.77 - 1.07)	0.95 (0.82 - 1.12)	0.89 (0.76 - 1.05)	0.91 (0.76 - 1.08)	1.25	0.26
Fully adjusted model + dietary factors	1 (ref)	0.91 (0.77 - 1.07)	0.95 (0.82 - 1.12)	0.89 (0.76 - 1.05)	0.90 (0.76 - 1.07)	1.33	0.25

	Breast cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from dairy products							
Fully adjusted model	1 (ref)	1.04 (0.92 - 1.17)	0.99 (0.87 - 1.12)	0.93 (0.82 - 1.06)	0.94 (0.86 - 1.01)	2.72	0.10
Fully adjusted model + dietary factors	1 (ref)	1.03 (0.91 - 1.17)	0.98 (0.87 - 1.12)	0.93 (0.82 - 1.06)	0.93 (0.86 - 1.01)	2.94	0.09
Protein from milk							
Fully adjusted model	1 (ref)	1.00 (0.88 - 1.13)	0.99 (0.88 - 1.12)	0.94 (0.83 - 1.07)	0.91 (0.80 - 1.03)	2.09	0.15
Fully adjusted model + dietary factors	1 (ref)	0.99 (0.87 - 1.12)	0.98 (0.87 - 1.12)	0.92 (0.81 - 1.04)	0.89 (0.79 - 1.01)	3.15	0.08
Protein from cheese							
Fully adjusted model	1 (ref)	1.00 (0.88 - 1.14)	1.04 (0.92 - 1.18)	1.04 (0.92 - 1.18)	1.03 (0.91 - 1.16)	0.21	0.65
Fully adjusted model + dietary factors	1 (ref)	0.99 (0.87 - 1.13)	1.03 (0.91 - 1.17)	1.04 (0.92 - 1.18)	1.03 (0.91 - 1.17)	0.24	0.62

Table 6.7 continued

	Prostate cancer						
	Q1	Q2	Q3	Q4	Per 2.5% energy increase	χ^2	*P-trend*
Protein from dairy products							
Fully adjusted model	1 (ref)	1.00 (0.89 - 1.12)	1.08 (0.96 - 1.21)	1.04 (0.93 - 1.17)	1.04 (0.96 - 1.12)	0.85	0.36
Fully adjusted model + dietary factors	1 (ref)	1.00 (0.88 - 1.12)	1.07 (0.95 - 1.20)	1.04 (0.92 - 1.17)	1.04 (0.96 - 1.12)	0.85	0.36
Protein from milk							
Fully adjusted model	1 (ref)	1.03 (0.92 - 1.15)	0.99 (0.88 - 1.11)	**1.12 (1.00 - 1.26)**	1.11 (0.98 - 1.26)	2.74	0.10
Fully adjusted model + dietary factors	1 (ref)	1.02 (0.91 - 1.14)	0.98 (0.87 - 1.10)	1.10 (0.98 - 1.24)	1.09 (0.96 - 1.24)	1.92	0.17
Protein from cheese							
Fully adjusted model	1 (ref)	0.99 (0.88 - 1.11)	1.03 (0.92 - 1.15)	0.97 (0.86 - 1.08)	0.96 (0.85 - 1.08)	0.52	0.47
Fully adjusted model + dietary factors	1 (ref)	0.98 (0.88 - 1.10)	1.03 (0.92 - 1.15)	0.97 (0.87 - 1.09)	0.97 (0.86 - 1.09)	0.30	0.59

Estimates in bold indicate statistically significant results.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol intake, ethnicity, diagnosis of diabetes, BMI, energy intake.

For colorectal cancer analyses: all models were stratified by sex and adjusted for menopausal status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use.

For breast cancer analyses: models were further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status.

For prostate cancer analyses: models were further adjusted for marital status.

Fully adjusted model + dietary factors were further adjusted for fruit and vegetable intake and fibre intake as well as red and processed meat intake for breast cancer and prostate cancer analyses.

Abbreviations: BMI, body mass index; CI, confidence intervals; Q, quantile.

6.5 Discussion

6.5.1 Main findings

In this analysis using data from the UK Biobank study, higher intakes of protein from total dairy products and milk were associated with a lower risk of colorectal cancer risk. There was suggestive evidence that men with high compared to low intakes of protein from milk may have a slightly higher risk of being diagnosed with prostate cancer, although there was no evidence of a linear trend increasing risk with higher milk protein intake. Dairy protein intake was not associated with breast cancer risk in women.

6.5.2 Colorectal cancer

To date, only two previous prospective studies have looked at the association of dairy protein and its sources with colorectal cancer risk [405, 406], with both suggesting an inverse association with risk. Milk protein intake was the major contributor to total dairy protein in this analysis (~50%) and the observed inverse association for dairy protein intake is probably driven by milk consumption. It is likely that the protein content in dairy products or milk is not the driving force for the observed lower risk of colorectal cancer, as there are some other compounds in milk, such as calcium, that may reduce the risk of colorectal cancer [114].

In the most recent WCRF meta-analysis (2017), including 14 prospective studies, the intake of total grams of all dairy products combined was associated with a 13% lower risk of colorectal cancer risk per 400 g/day intake [20]. Similarly, the WCRF found that in 13 prospective studies, milk intake was associated with a 6% lower risk of colorectal cancer per 200 g/day increment, whereas for cheese intake a non-significant inverse association was observed [20]. In these findings, for comparison, a 12% lower risk of colorectal cancer was observed per 200 g/day of total dairy intake, and a 13% lower risk of colorectal cancer per 200 g/day intake of milk. Moreover, in MR analyses utilising the lactase persistence gene as a proxy for milk intake, which has been shown to predict a 17.1 g/day per allele difference in milk intake [408], a lower risk of colorectal cancer was observed among those who had the lactase persistence genotype and thus

were assumed to consume more milk (odds ratio per allele: 0.95, 95% CI: 0.91-0.99) [113]. However, extrapolating from MR studies, an odds ratio of ~0.55 per 200 g/day intake of milk would be expected, which is a lot stronger than what has been observed in prospective cohort studies [20].

Higher dietary calcium intake has been previously shown to be associated with lower colorectal cancer risk in prospective studies [20, 409], although this may be due to higher dairy intake, as dairy is the primary source of calcium in many countries [410]. Calcium is hypothesized to protect against colorectal cancer risk by locally binding to secondary bile acids in the lumen of the colon [115] and thus inhibiting their potential carcinogenic effect [411] as well as potentially restraining growth and promoting differentiation of colonic epithelium cells through the activation of calcium-sensing receptors [411, 412]. However, other components of dairy products such as conjugated linoleic acid, SCFAs, and lactoferrins have been hypothesized to play a role in colorectal cancer risk [114, 413], but isolating the impacts of these food components, which are consumed together in milk, is challenging.

In these analyses, heterogeneity was observed by both alcohol intake and smoking status in the association of intake of protein from milk with colorectal cancer risk; the inverse association of milk protein intake with colorectal cancer risk was observed only in individuals who consumed ≥10 g/day of alcohol, or were ever smokers, but this was not seen in those consuming <10 g/day of alcohol or those who were never smokers. Previous prospective studies have generally not explored whether the association of dairy product intake with colorectal cancer risk varies by these lifestyle factors [414-417]. Moreover, in agreement with this analysis, most previous prospective studies have observed that those who consume higher amounts of dairy products typically consume less alcohol on average [414-418]. This observed heterogeneity observed could suggest residual confounding by alcohol and smoking, for example, if low milk consumers under-report their alcohol consumption more than high milk consumers, or this finding may be due to chance.

6.5.3 Breast cancer

Intakes of total dairy protein and its sources were not associated with breast cancer risk in this study. To the best of my knowledge, only one previous prospective study has looked at this association; in a recent prospective analysis of 700,000 postmenopausal UK women from the Million Women Study (published after the most recent WCRF meta-analysis), which included 29,000 incident breast cancer cases diagnosed over 12 years of follow-up, intakes of dairy protein, as well as milk, yogurt, or cheese, were not associated with breast cancer risk [404]. In this analysis in UK Biobank that used total grams of all dairy products combined, those in the highest quartile had a suggestive lower risk of breast cancer than those in the lowest quartile of intake, although there was no significant trend. Evidence for the association of dairy intake and breast cancer risk from meta-analyses and recent studies has generally been inconsistent. The 2017 WCRF meta-analysis, including seven prospective studies, suggested that a higher intake of dairy products (per 200g/day) was associated with a 5% lower risk of premenopausal breast cancer, and no association was found with postmenopausal breast cancer [130]. After this meta-analysis was published, two prospective studies, including 2582 and 1057 breast cancer cases among populations with many participants who did not consume dairy or consumed dairy in low amounts, suggested a positive association between dairy intake and breast cancer risk [419-421]. Recently, a pooled analysis of 21 cohort studies with individual level data including 37,000 breast cancer cases found a 5% lower risk of breast cancer for participants in the highest category of milk intake compared to lowest category [422]. However, MR analyses using the lactase persistence gene suggest no association between milk intake and overall breast cancer risk [113].

6.5.4 Prostate cancer

A suggestive positive association was found between the intake of protein from milk and prostate cancer risk. However, this association was attenuated when other dietary components were adjusted for in multivariable models. No other associations between intakes of protein from total dairy products or cheese and prostate cancer risk were observed. To the best of my

knowledge, only three previous prospective studies have looked at the association of protein from dairy products and prostate cancer risk; two studies including men in EPIC where one including 142,251 men of whom 2727 developed prostate cancer, found a 22% higher risk comparing the highest fifth to lowest fifth of dairy protein intake [402] and the other, which included an updated analysis of EPIC including 131,425 men of whom 6939 developed prostate cancer, found a higher risk among the men in the top three fifths compared to the lowest fifth of intake of dairy protein and yogurt protein [403]. The third study to assess this association was a nested case-control study in the Breast and Prostate Cancer Cohort Consortium, which included 4,815 cases and 4,671 controls and found no association between dairy protein intake and prostate cancer risk [388]. When analyses used total grams of all dairy products, milk, and cheese, no association with prostate cancer risk was found, whereas the latest WCRF meta-analysis of prospective studies from 2014 found that the intake of total grams of all dairy products, milk, and cheese were all associated with a higher risk of prostate cancer [262, 287]. MR analyses using the lactase persistence gene as a proxy for milk intake have generally found no association in UK Biobank or the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome consortium [113]; however, using data from a Finnish consortium (3282 prostate cancer cases and 55,968 controls), where milk intake is typically high, the MR analysis showed a significant positive association [113].

The possible positive association between protein from milk and prostate cancer risk has been hypothesized to be mediated by circulating IGF-I concentrations, which have been shown to be associated with both higher dairy protein intake [5] and prostate cancer risk [11, 14, 336], and possibly prostate cancer mortality [14]. However, when circulating IGF-I concentrations measured at baseline (before dietary assessments) were added to multivariable-adjusted models, the association of protein from milk with prostate cancer risk was only slightly attenuated suggesting that circulating IGF-I concentrations do not completely mediate this association. Milk intake may also activate the mTOR pathway and mTOR complex 1 due to the high content of branched chain amino acids, particularly leucine [423], in which milk has been

implicated in carcinogenesis of the prostate [305]. However, the observed results for milk protein may also be subject to detection bias if men who consume more milk are more likely to undergo a PSA test and potentially a prostate cancer diagnosis, and previous evidence has suggested that some lifestyle factors are associated with having a PSA test [424]. Moreover, in this analysis, the risk of total prostate cancer was assessed because data on tumour subtypes were not available; however, risk factors for prostate cancer may vary by tumour characteristics [268] and further research is needed to understand these mechanisms and how they influence prostate carcinogenesis, particularly for aggressive and lethal prostate cancer.

6.5.5 Role of dairy protein intake in relation to circulating IGF-I concentrations

Although dairy protein intake, and more specifically milk protein intake (as observed in Chapter 5), have been associated with higher circulating IGF-I concentrations [5], the magnitude of this relationship, and therefore any impact on cancer risk, may not be large. In cross-sectional analyses from Chapter 5, a 2.5% higher energy intake from milk protein was associated with a 1.20 nmol/L higher circulating IGF-I concentration. In prospective analyses, IGF-I concentrations (per 5 nmol/L) have been associated with ~9-11% higher risks of colorectal, breast, and prostate cancer [12-14]. Similar estimates have been observed utilising MR analyses [12, 13], with the exception that a 5 nmol/L genetically predicted higher IGF-I concentration has been associated with a 34% higher risk of prostate cancer when using a *cis*-SNP [14]. Based on these estimates, one might expect to see a ~2% higher risk of both colorectal and breast cancer and a ~2-8% higher risk of prostate cancer per 2.5% higher energy intake from milk protein (although due to measurement error in dietary assessments these risks might be larger). While protein from dairy products may elevate IGF-I concentrations, other components of dairy products and milk may protect against the risk of cancer at some sites and further research examining other mechanisms and compounds in dairy products is needed to better understand how the intake of dairy products may influence cancer risk.

6.5.6 Strengths and limitations

The strengths of these analyses include the prospective nature of the study, detailed dairy protein intake obtained from 2 to 5 24-hour dietary assessments, which reduces random measurement error, and the use of record linkage limiting loss to follow-up and outcome misclassification. The detailed dietary assessment allowed for the detailed examination of two specific sources of dairy products in association with common cancers. As well, the UK Biobank collected detailed sociodemographic, anthropometric, and lifestyle information from participants at recruitment, allowing adjustment for these potential confounders and other components of diet in the analyses. Participants in the UK Biobank also provided a blood sample at recruitment in which IGF-I concentrations were measured, allowing us to assess how IGF-I concentrations influence the association between dairy products and risk of IGF-I related cancers.

There are also some limitations to consider. Self-reported dietary intake is subject to error; approximately 35% of participants included in these analyses completed only two 24-hour dietary assessments, which increases random measurement error and may therefore bias the estimates towards the null. However, in sensitivity analyses restricted to participants with at least three completed 24-hour dietary assessments, similar results were observed but with wider CIs due to the loss of participants and cases. Associations with all individual dairy sources, such as yogurt, could not be assessed as it was consumed episodically in this sample and therefore a few 24-hour dietary assessments per person cannot accurately estimate usual intake. The UK Biobank is also known to be a generally healthy sample of participants, which may make these results not generalizable to the greater population and may bias the observed results [384], although estimates may remain in the same direction [425]. Moreover, only one measure of IGF-I concentrations was utilised in multivariable-adjusted models, which may underestimate typical long-term concentrations. Finally, these associations are subject to residual and unmeasured confounding due to the observational design of the study, however, assessment of residual

confounding, using the change in χ^2 method, did not suggest that residual confounding greatly influenced the observed results.

6.6 Conclusions

In conclusion, higher intakes of protein from dairy products and protein from milk were inversely associated with colorectal cancer risk, although based on previous evidence it is possible that these associations are driven by other components present in dairy products, such as calcium. A higher intake of protein from milk was weakly positively associated with prostate cancer risk, although this association was only significant when analyses compared the highest with the lowest quartile and there was no significant linear trend. No associations between intake of protein from dairy products sources and breast cancer risk were observed. Further research is needed on both the IGF pathway and other mechanisms as well as other explanations that may account for the possible association between dairy products or milk and prostate cancer risk.

Chapter 7

Associations of carbohydrate types and sources, short-chain fatty acid genetic modifiers, and colorectal cancer risk

Chapter 7 Summary

Wholegrain and fibre intakes may decrease the risk of colorectal cancer, although evidence for potential mechanisms and how intakes of specific types and sources of carbohydrates relate with colorectal cancer remains inconclusive. Moreover, cellular and animal studies have shown that SCFA production in the colon, which increases with wholegrain and fibre intake, may be associated with lower colorectal cancer risk. This chapter examines the associations between carbohydrate types and sources and colorectal cancer risk in 114,217 UK Biobank using ≥2 (maximum of five) 24-hour dietary assessments. A polygenic score was also applied to categorise participants as high or low host genetic score for intra-luminal microbial SCFA production, namely butyrate and propionate. During a median follow-up of 9.4 years, 1,193 participants were diagnosed with colorectal cancer. Intakes of total sugar, non-free sugars, and fibre from wholegrains were inversely associated with colorectal cancer risk (HR per 5% energy: 0.93 95% confidence interval CI: 88-0.98; 0.92, 0.87-0.98; and HR per 5 g/day 0.90, 0.82-1.00, respectively). Evidence of heterogeneity was observed by genetically predicted butyrate synthesis (p-heterogeneity=0.023); participants in the high butyrate genetic score category and consuming higher wholegrains had a lower risk of colorectal cancer (per 5% energy; 0.88, 0.78-0.99). In additional analyses utilising the larger UK Biobank sample (n=343,621) but with less detailed dietary assessment, similar results were observed in which only individuals with a high butyrate genetic score had a lower risk of colorectal cancer per 5 g/day intake of fibre from breads and cereals (0.88, 0.81-0.95; p-heterogeneity=0.021). This chapter suggests that colorectal cancer risk varies by intake of carbohydrate types and sources; total sugar, non-free sugar, and wholegrain intake may lower the risk of colorectal cancer; however, the association for wholegrain intake may be modified by host genetically predicted SCFA production.

7.1 Introduction

Colorectal cancer is the second most common cause of cancer death in both men and women worldwide [1]. Smoking, excess adiposity, processed meat intake, and alcohol consumption are well-established modifiable risk factors for colorectal cancer [20, 102]. Other dietary factors may also influence colorectal cancer risk; the latest WCRF meta-analysis of prospective studies published before 2017 lists red meat as a probable factor for increasing the risk of colorectal cancer and foods containing dietary fibre, wholegrain consumption, dairy products, and calcium supplements as probable factors for decreasing the risk [20]. However, further research is needed to assess if specific sources and types of carbohydrates relate differently to colorectal cancer risk, as well to understand potential mechanisms underlying these associations.

Different types and sources of carbohydrates may influence insulin response; wholegrains and foods rich in fibre may be inversely associated with fasting insulin and positively associated with insulin sensitivity [426], whereas frequent intake of foods high in refined sugars may lead to excess energy intake and weight gain, which may result in hyperinsulinemia [427, 428], a known risk factor for colorectal cancer [74, 108].

One other potential hypothesis that may explain the lower risk of colorectal cancer among individuals who consume more dietary fibre and wholegrains is the fermentation of dietary fibre in the colon, which produces the SCFA butyrate, propionate, and acetate [429]. Evidence from animal and *in vivo* studies has suggested that butyrate, which is used as an energy source by colonocytes, may be protective against carcinogenesis [109, 430-432]. Large genetic studies of up to 18,000 individuals have suggested that the composition of microbiota in the gut is influenced by host genetic factors [433, 434]. As such, synergistic mechanisms may interplay in the colorectum between host genetic factors, colonisation of specific bacteria, production of SCFA, and intake of wholegrains and fibre in reducing colorectal cancer risk; however, further research is needed to understand these potential relationships.

7.2 Aim

This Chapter therefore aims to examine the associations of carbohydrate types and sources with colorectal cancer risk in 114,217 participants from the UK Biobank with detailed dietary data. The secondary aim of this analysis was to examine if host genetic variants related to colonic/microbial SCFA production may modulate the associations of carbohydrate and fibre intakes with the risk of colorectal cancer.

7.3 Methods

7.3.1 Study design & participants

Details of the UK Biobank, including recruitment and data collection are described in Chapter 3. This analysis included a subsample of participants from the UK Biobank who completed a minimum of two 24-hour dietary assessments.

7.3.2 Exclusions

Exclusions for this analysis were the same as in Chapter 6 and can be found in **Section 6.3.3** along with the flow diagram of exclusions (**Figure 6.1**)

7.3.3 Assessment of diet: 24-hour dietary assessment

Dietary intake was estimated in a subsample of participants who completed the Oxford WebQ 24-hour dietary assessment on at least two occasions (maximum of five). When comparing the completion of two 24-hour dietary assessments with energy measured from an accelerometer and urinary sugar, a correlation of 0.38 (95% CI: 0.21, 0.54) for energy and 0.40 (95% CI: 0.24, 0.55) for sugar was observed, with higher correlations when additional 24-hour dietary assessments were completed [359].

Details of the 24-hourd dietary assessment are provided in Chapter 3. Briefly, carbohydrate intakes were estimated in each 24-hour dietary assessment based on the foods and beverages

consumed by participants [361, 435]. The percentage of energy from total carbohydrates, total sugars, free sugars (all monosaccharides and disaccharides added to foods, plus sugars naturally present in honey, syrups and unsweetened fruit juices), non-free sugars (total sugar minus free sugars), total starch, starch from wholegrains, and starch from refined grains were estimated for each 24-hour dietary assessment and were averaged across all available 24-hour dietary assessments for each individual to obtain the average intake of each type and source of carbohydrate. Similarly, total non-starch polysaccharides (Englyst fibre) [435], fibre from vegetables, fibre from fruits, and fibre from wholegrains were estimated in grams/day in each 24-hour dietary assessment and averaged across all available 24-hour dietary assessments for each individual to obtain the average fibre intake from each source.

7.3.4 Colorectal cancer incidence

Ascertainment of colorectal cancer cases from both the cancer registry and HES were utilised for this chapter and was identical to Chapter 6 and are described in detail under section **6.3.4 Outcome ascertainment: cancer diagnoses**. Participants were coded as having an event if they had an incident diagnosis of colorectal cancer (ICD-10 codes: C18 for colon cancer; C19 and C20 for rectal cancer).

7.3.5 Genetic factors and polygenic scores

The genotyping of participants in the UK Biobank has been described in greater detail in Chapter 3. Briefly, the UK Biobank genotyped 488,377 participants using 2 arrays, namely the UK BiLEVE Axiom (n=49,950 participants) including 807,411 SNPs, and the UK Biobank Axiom (n=438,437 participants) which included 825,927 SNPs and shared 95% of marker content with the UK BiLEVE. Over 90 million variants were imputed from this using the Haplotype Reference Consortium and UK10K + 1000 Genomes reference panels [350].

To estimate the potential influence of host genetic factors that may affect microbial SCFA production on the associations of intake of carbohydrate types and sources with colorectal

cancer, a 9 SNP host polygenic score (PGS) for butyrate and a 3 SNP host PGS for propionate were used. These SNPs were previously identified in a GWAS of 952 normoglycemic individuals [436]. For the butyrate PGS, 9 SNPs were used that were associated with the microbial pathway involved in 4-aminobutanoate (GABA) degradation (BioCyc ID: PWY-5022; variance explained=16%; F-statistic=21) where butyrate is a product [436]. Abundance of the PWY-5022 pathway was associated with butyrate producing bacteria namely *Eubacterium rectale* and *Roseburia intestinalis* [436]. For the propionate PGS, 3 SNPs were used that were associated with fecal propionate levels (variance explained = 6.3%; F-statistic=21). The propionate SNPs were independent of the SNPs used for butyrate PWY-5022 abundance PGS. The PGS for butyrate and propionate were calculated from PGS = ($\beta_1 \times SNP_1 + \beta_2 \times SNP_2 + ... \beta_9 \times SNP_9$) where β value was from the estimates of the previous GWAS [436] and SNP represented the number of effect alleles for each SNP (**Table 6.1**). After exclusions (described in detail below), no participants in this subsample of the UK Biobank had any missing genotypes for the butyrate or propionate SNPs. Once the PGS were calculated for all participants, and appropriate exclusions were made, participants were separated by the median PGS (1.88 for butyrate PGS and 0.142 for propionate PGS) among eligible participants, those >50 percentile were considered as having a high PGS whereas those $\leq 50^{th}$ percentile were considered as having a low PGS for both butyrate and propionate. For genetic analyses, participants were excluded if they did not have genetic information (n=1,998) or reported they were not of white British ancestry (n=16,445). Participants were also excluded if they had low call rates (<98% n=1,326), sex chromosome aneuploidy (n=66), as well as related individuals (kinship coefficient >0.0884; n=6,965), leaving a maximum of 87,417 participants and 909 cases of colorectal cancer in these analyses.

Table 7.1 Short-chain fatty acid SNPs included in the polygenic score derived from Sanna et al. [436].

SNP (rs ID)	Effect Allele	Effect allele frequency	Beta (95% CI; standard deviations)	Standard Error	R^2	Effect allele frequency in UKB
Butyrate (PWY-5022 pathway) score						
rs9423658	C	0.86	0.33 (0.19,0.48)	0.07653061	<0.1	0.84
rs881390	C	0.11	0.40 (0.23,0.57)	0.08673469	<0.1	0.12
rs2089222	A	0.04	0.56 (0.32,0.79)	0.11734694	<0.1	0.04
rs9904981	G	0.77	0.25 (0.14,0.36)	0.05612245	<0.1	0.80
rs10483112	T	0.03	0.59 (0.34,0.84)	0.12755102	<0.1	0.03
rs12994030	T	0.27	0.24 (0.14,0.34)	0.05102041	<0.1	0.28
rs2056208	T	0.24	0.24 (0.14,0.34)	0.05612245	<0.1	0.27
rs10019739	C	0.27	0.24 (0.13,0.34)	0.05102041	<0.1	0.27
rs7743827	G	0.80	0.27 (0.15,0.38)	0.05612245	<0.1	0.80
Propionate score						
rs7142308	G	0.61	0.24 (0.14,0.34)	0.05102041	<0.1	0.63
rs12050534	C	0.14	0.31 (0.18,0.44)	0.06632653	<0.1	0.16
rs1400566	G	0.45	-0.22 (-0.31, -0.12)	0.05102041	<0.1	0.56

Abbreviations: SNPs, single nucleotide polymorphisms; UKB, UK Biobank

7.3.6 Statistical analysis

Participants were classified into quartiles for percentage of energy intake from each carbohydrate type and source as well as quartiles by g/day of fibre sources. Intakes of carbohydrates were also modelled as a per 5% energy increment for each source (modelled continuously), and intakes of fibre sources were also modelled as a 5 g/day increment for each fibre source (modelled continuously). Baseline characteristics were summarised across quartiles of carbohydrates and fibre intakes.

Cox proportional hazards regression, with age as the underlying time variable, was used to calculate HRs and 95% CIs for colorectal cancer incidence. For covariates in which responses were unknown or missing, participants were categorised into an unknown/missing category. All covariates added to the model were selected *a priori* based on probable and known risk factors for colorectal cancer. Minimally adjusted models were stratified by sex and age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years) and adjusted for region at recruitment (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland).

Multivariable Cox regression models were further adjusted for height (six sex-specific categories increasing by 5 cm, and unknown/missing (0.16%)), physical activity (low; 0-9.99, medium; 10-49.99, high; ≥50 excess MET-hours/week, and unknown/missing (1.81%)), Townsend deprivation index (quintiles from most deprived to least deprived, and unknown/missing (0.12%)), education (completion of national exam at 16, completion of national exam at 17-18, college or university degree, or other/unknown/missing (6.6%)), employment status (employed, retired, not in paid employment, or unknown (0.66%)), smoking status (never, former, ≤15 cigarettes/day, 16-29 cigarettes/day, ≥30 cigarettes/day, or missing/unknown (0.23%)), alcohol consumption estimated from the recruitment questionnaire (non-drinkers, <1 g/day, 1-9.9 g/day, 10-19.9 g/day, 20+ g/day, or unknown/missing (0.36%)), ethnicity (White, mixed race or other, Asian or British Asian, and Black or Black British, or missing/unknown (0.33%)),

diabetes (yes, no, or unknown (0.15%)), NSAID use (regular user of ibuprofen or aspirin, irregular user, no reported use), BMI (<20, 20-22.49, 22.5-24.99, 25.0-27.49, 27.5-29.99, 30-32.49, 32.5-34.99, ≥35 kg/m^2, and unknown/missing (0.20%)), consumption of red and processed meat (<2 times/week, 2-2.99 times/week, 3-3.99 times/week and 4+ times/week, or unknown (0.41%)) from the touchscreen questionnaire at recruitment, intake of fruit and vegetables estimated from the 24-hour dietary assessments (quintiles; excluded from the model when the exposures were total fibre, fibre from fruit and/or vegetables and non-free sugars), energy intake (sex-specific quintiles), and women specific covariates: MHT use (never, former, current, or unknown (0.12%)), and menopausal status (premenopausal, postmenopausal, or unknown (5.4%)). Further information on classification of covariates can be found in the **Ancillary Methods A.1**.

To assess for residual confounding and the extent to which adjustment for confounders influenced the associations of intakes of carbohydrate types and sources with the risk of colorectal cancer, χ^2 statistics for including the carbohydrate types or sources or fibre source were estimated using LRT by comparing models without the carbohydrate/fibre source to models with the carbohydrate/fibre source (modelled as 5% energy increment for carbohydrates and 5 g/day for fibre intake) [149]. The change in χ^2 values were estimated using the minimally adjusted models χ^2 value as the reference. A threshold of a ~70% reduction in χ^2 values from minimally adjusted models was set to represent the possibility of residual confounding as this represents a substantial change in the χ^2 value with additional adjustment of imperfectly measured confounders, thus suggesting that residual confounding may be present. Tests for departures from linearity were estimated by comparing models with carbohydrates (per 5% energy intake/day) to models with carbohydrate quartiles and models with fibre as a continuous variable (per 5 g/day) with models as fibre quartiles, and no evidence non-linearity was observed (data not presented).

Subgroup and sensitivity analyses

Heterogeneity was assessed by sex or by BMI groups (<27, and ≥27 kg/m^2) by using a LRT comparing the main model to a model including an interaction term between the carbohydrate source (modelled as 5% energy increment) and fibre (modelled as 5 g/day increment) and sex or BMI. Heterogeneity was further explored by tumour site (colon (ICD-10 code: C18; or rectal ICD-10 code: C19/C20) and in 68 instances, the diagnosis of colon and rectal cancer coincided, and these participants were removed from the subgroup analyses by tumour site. For heterogeneity by tumour site, Cox models were stratified using a competing risks approach [437] and compared the risk coefficients and standard errors of each carbohydrate source as a 5% energy increment per day using colon cancer and rectal cancer as separate outcomes. This was also conducted for fibre intake, modelled as a 5 g/day increment in each subgroup.

To assess the robustness of results, sensitivity analyses were conducted restricted to participants who completed a minimum of three 24-hour dietary assessments to further reduce random measurement error in the carbohydrate and fibre intake estimates. As well, to investigate the potential for reverse causality, cases diagnosed, and participants censored in the first two years of follow-up were excluded.

Genetic factors

To assess for an interaction between SCFA production PGS and intakes of carbohydrate types and sources and colorectal cancer risk, a LRT was used comparing the multivariable model to a model with an interaction term between the carbohydrate source (per 5% energy increment) or fibre intake (per 5 g/day increment) with the butyrate PGS or propionate PGS, separating participants using the median PGS for both butyrate and propionate (low or high PGS). The multivariable models included all confounders from the main model and were also adjusted for the first 10 principal components of ancestry to reduce the potential impact of population stratification. Although gene – environment interactions may not be biased by uncontrolled environment confounding when gene – environment are independent and environmental

confounders are not associated with the genetic factor [438], full adjustment for potential confounders between carbohydrate types and sources and colorectal cancer were made to minimise any bias of exposure – outcome confounding.

Additional analyses were conducted assessing intakes of fibre from breads and cereals derived from the few dietary questions asked at the baseline questionnaire [383] with colorectal cancer risk by the butyrate and propionate PGS in the larger UK Biobank sample (n=343,621; 4,191 cases of colorectal cancer). Due to the limited number of dietary questions asked at recruitment, sensitivity analyses further exploring SCFA heterogeneity were restricted to only this variable due to its similarity to the wholegrain variable determined for the 24-hour dietary assessments, as well it was previously associated with a lower risk of colorectal cancer [439]. The fibre from breads and cereals variable was calculated using the frequency and type of bread and cereals consumed reported by individuals at their recruitment visit. The calculation of fibre intake from these dietary responses at recruitment has been previously described in detail elsewhere [383]. Briefly, participants were asked how many slices of bread or bowls of cereal they consumed per week as well as the type of bread and cereal they mainly consumed. From these responses, the non-starch polysaccharide content (in grams) of slices of bread and breakfast cereals were determined based on the frequency of consumption and types of bread or cereal reported in the touchscreen questionnaire. Fibre values were obtained for breads and cereals from McCance and Widdowson's The Composition of Foods (Seventh Summary Edition) [440] and the fibre content was summed across both bread and cereal intakes for all participants. This sample was restricted to individuals who reported they were from white British ancestry and excluded participants with outliers for heterozygosity, low call rates (<98%), sex chromosome aneuploidy, and related individuals (kinship coefficient >0.0884). As well, participants who had a cancer diagnosis at recruitment or withdrew consent were also excluded. Analyses were also adjusted for the first 10 principal components. Adjustment for energy intake was not possible for this analysis due to the limited number of dietary questions asked to the UK Biobank participants at recruitment.

All analyses were conducted using Stata version 17.0 (Stata Corp, TX, United States) and p-values were two-sided with $p < 0.05$ being considered statistically significant. Correction for multiple testing was not undertaken in these analyses as they were primarily exploratory.

7.4 Results

7.4.1 Participant characteristics and colorectal cancer incidence

Over a median of 9.4 years of follow-up and 1,023,396 person-years, 1,193 incident cases of colorectal cancer occurred. **Table 7.2** presents participants' characteristics between those in the lowest and highest quartiles of wholegrains, refined grains, and total fibre intake and **Tables 7.3 & 7.4** present the baseline characteristics by quartiles for total carbohydrates, sugars, and fibre sources and **Figure 7.1** presents the top food group contributors for each carbohydrate type and source. Participants in the highest quartile of fibre intake had a lower BMI, reported lower intake of processed and red meat, consumed less alcohol, and were more likely to have a university or college degree compared to the lowest quartile of fibre intake (**Table 7.2**). Participants who consumed the highest amount of starch or fibre from wholegrains were more likely to have a lower BMI, be never smokers, consume less red and processed meat, and less alcohol, and have a university or college degree compared to those in the lowest quartile (**Table 7.4 & Table 7.6**).

Table 7.2 Baseline characteristics by lowest and highest quartile of intake of wholegrain starch, refined grain starch, and fibre.

	Wholegrain starch		Refined grain starch		Total fibre	
	Q1	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**
Number of participants	28,555	28,554	28,555	28,554	28,555	28,554
Intake of carbohydrate of interest [a]	0.6 (0.8)	10.9 (3.1)	4.4 (2.0)	19.9 (3.9)	11.3 (2.1)	25.5 (4.2)
Sex - Male, N (%)	13,296 (46.6%)	13,392 (46.9%)	12,402 (43.4%)	13,664 (47.9%)	12,208 (42.8%)	14,356 (50.3%)
Age at recruitment - years	55.0 (7.9)	56.6 (7.6)	57.6 (7.3)	53.8 (8.0)	54.8 (7.8)	56.7 (7.8)
Body mass index – kg/m^2	27.4 (4.8)	26.0 (4.3)	26.5 (4.5)	27.0 (4.8)	27.2 (4.7)	26.2 (4.5)
Height - centimetres	169.4 (9.2)	169.7 (9.2)	169.2 (9.1)	169.7 (9.2)	168.4 (9.1)	170.8 (9.3)
Physical activity - High	4,734 (16.6%)	5,110 (17.9%)	5,475 (19.2%)	4,343 (15.2%)	4,154 (14.5%)	6,199 (21.7%)
Townsend deprivation index, N (%)						
Q1 - Most affluent	5,757 (20.2%)	6,398 (22.4%)	6,320 (22.1%)	5,891 (20.6%)	5,913 (20.7%)	6,172 (21.6%)
Q5 - Most deprived	5,325 (18.6%)	4,446 (15.6%)	4,317 (15.1%)	5,184 (18.2%)	5,235 (18.3%)	4,399 (15.4%)
In paid employment, N (%)	18,669 (65.4%)	17,061 (59.7%)	16,134 (56.5%)	19,652 (68.8%)	18,968 (66.4%)	16,717 (58.5%)
University/college degree, N (%)	20,111 (70.4%)	21,618 (75.7%)	21,018 (73.6%)	20,645 (72.3%)	19,570 (68.5%)	21,962 (76.9%)
White ethnicity	27,323 (95.7%)	27,708 (97.0%)	27,908 (97.7%)	26,945 (94.4%)	27,264 (95.5%)	27,650 (96.8%)
Never smoker, N (%)	14,993 (52.5%)	17,547 (61.5%)	16,034 (56.2%)	16,907 (59.2%)	15,322 (53.7%)	17,025 (59.6%)
Diabetes - Yes, N (%)	1,088 (3.8%)	1,150 (4.0%)	1,000 (3.5%)	1,223 (4.3%)	1,033 (3.6%)	1,167 (4.1%)
NSAID - regular user, N (%)	7,585 (26.6%)	6,650 (23.3%)	7,074 (24.8%)	7,291 (25.5%)	7,604 (26.6%)	6,843 (24.0%)
Postmenopausal at recruitment, N (%)	9,626 (63.1%)	10,747 (70.9%)	12,370 (76.6%)	8,329 (56.0%)	10,312 (63.1%)	10,206 (71.9%)
Menopausal hormone therapy use - Current use, N (%)	1,320 (8.7%)	1,090 (7.2%)	1,395 (8.6%)	1,096 (7.4%)	1,518 (9.3%)	1,050 (7.4%)
Diet variables						
Alcohol intake - g/day	19.7 (20.9)	13.5 (13.5)	17.4 (17.6)	14.7 (16.1)	19.0 (19.9)	14.4 (15.3)
Red and processed meat intake - 4+ times/wk, N (%)	13,034 (45.6%)	9,070 (31.8%)	9,918 (34.7%)	11,818 (41.4%)	11,914 (41.7%)	9,546 (33.4%)
Vegetable and fruit intake - g/day	342.0 (226.6)	409.4 (219.0)	460.8 (250.5)	307.5 (193.1)	219.7 (126.1)	578.5 (256.7)
Total carbohydrate intake - % energy	46.6 (8.1)	52.1 (6.5)	47.6 (8.4)	51.5 (6.5)	46.2 (8.2)	52.2 (6.5)
Total sugar intake - % energy	23.0 (7.0)	25.2 (6.1)	26.3 (7.1)	22.1 (5.9)	22.7 (6.8)	26.2 (6.4)
Total starch intake - % energy	23.6 (6.3)	26.8 (5.1)	21.3 (5.6)	29.3 (4.6)	23.4 (6.2)	25.9 (5.4)
Total fibre intake - g/day	14.9 (5.1)	20.5 (5.7)	19.2 (6.2)	16.3 (5.1)	11.3 (2.1)	25.4 (4.1)
Total fibre from wholegrains - g/day	0.5 (0.8)	7.2 (2.8)	4.8 (3.4)	2.4 (2.4)	1.7 (1.7)	5.9 (3.6)
Total energy intake - kJ/day	8,696 (2067)	8,257 (1831)	8,246 (1886)	8,720 (1999)	7,393 (1617)	9,845 (1913)

Values are mean (SD) unless otherwise indicated.

[a] values represent % of energy for wholegrain starch and refined grain starch and g/day for total fibre

Abbreviations: g, grams; g/day, grams per day; kJ, kilojoules; N, Number of participants; NSAID, non-steroidal anti-inflammatory drug; Q, quantile; wk, week; y, years.

Table 7.3 Baseline characteristics by lowest and highest quartile of intake of total carbohydrates, total sugars, non-free sugars, and free sugars.

	Total carbohydrates		Total sugars		Total non-free sugar		Total free sugar	
	Q1	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**
No. of participants	28,555	28,553	28,555	28,554	28,555	28,554	28,555	28,554
Intake of carbohydrate of interest [a]	39.9 (4.6)	58.3 (3.5)	16.5 (2.7)	32.9 (4.1)	6.8 (1.6)	20.5 (4.1)	5.8 (1.7)	17.9 (3.7)
Sex - Male, N(%)	13,810 (48.4%)	12,027 (42.1%)	14,683 (51.4%)	10,706 (37.5%)	17,274 (60.5%)	8,498 (29.8%)	10,586 (37.1%)	15,225 (53.3%)
Age at recruitment - years	55.9 (7.7)	55.8 (8.0)	54.9 (7.9)	56.7 (7.7)	54.1 (8.0)	57.3 (7.4)	56.2 (7.6)	55.4 (8.1)
Body mass index	27.1 (4.6)	26.4 (4.6)	27.2 (4.7)	26.4 (4.5)	27.3 (4.6)	26.4 (4.6)	26.9 (4.8)	26.7 (4.5)
Height	170.1 (9.1)	168.7 (9.1)	170.3 (9.2)	168.2 (9.0)	171.6 (9.1)	167.1 (8.7)	168.2 (9.0)	170.5 (9.2)
Physical activity - High, N (%)	4,573 (16.0%)	5,353 (18.7%)	4,213 (14.8%)	5,839 (20.4%)	4,217 (14.8%)	5,785 (20.3%)	4,755 (16.7%)	5,143 (18.0%)
Townsend deprivation index, N (%)								
Q1 - Most affluent	5,998 (21.0%)	6,015 (21.1%)	5,941 (20.8%)	6,262 (21.9%)	5,775 (20.2%)	6,300 (22.1%)	6,014 (21.1%)	6,072 (21.3%)
Q5 - Most deprived	4,991 (17.5%)	4,671 (16.4%)	5,127 (18.0%)	4,539 (15.9%)	5,316 (18.6%)	4,397 (15.4%)	4,900 (17.2%)	4,898 (17.2%)
In paid employment, N (%)	18,277 (64.0%)	17,397 (60.9%)	19,121 (67.0%)	16,576 (58.1%)	19,609 (68.7%)	16,203 (56.7%)	17,683 (61.9%)	17,896 (62.7%)
University/college degree, N(%)	21,372 (74.8%)	20,510 (71.8%)	21,002 (73.5%)	20,603 (72.2%)	20,469 (71.7%)	20,949 (73.4%)	21,030 (73.6%)	20,305 (71.1%)
White ethnicity, N (%)	27,804 (97.4%)	27,106 (94.9%)	27,592 (96.6%)	27,344 (95.8%)	27,521 (96.4%)	27,479 (96.2%)	27,578 (96.6%)	27,309 (95.6%)
Smoking - Never, N (%)	13,467 (47.2%)	18,486 (64.7%)	14,474 (50.7%)	17,614 (61.7%)	14,450 (50.6%)	17,645 (61.8%)	15,484 (54.2%)	16,500 (57.8%)
Diabetes - Yes, N(%)	1,215 (4.3%)	989 (3.5%)	1,677 (5.9%)	699 (2.4%)	1,043 (3.7%)	1,041 (3.6%)	1,893 (6.6%)	611 (2.1%)
NSAID - regular user, N(%)	7,553 (26.5%)	6,817 (23.9%)	7,636 (26.7%)	7,085 (24.8%)	7,509 (26.3%)	6,794 (23.8%)	7,290 (25.5%)	7,274 (25.5%)
Postmenopausal at recruitment	9,880 (67.0%)	11,177 (67.7%)	8,297 (59.8%)	13,022 (73.0%)	6,044 (53.6%)	15,230 (76.0%)	12,514 (69.7%)	8,496 (63.8%)
Menopausal hormone therapy use - Current, N (%)	1,378 (9.3%)	1,234 (7.5%)	1,145 (8.3%)	1,516 (8.5%)	933 (8.3%)	1,680 (8.4%)	1,502 (8.4%)	1,102 (8.3%)
Diet variables								
Alcohol intake - g/day	24.7 (21.1)	9.6 (11.9)	21.4 (19.3)	11.6 (14.2)	23.4 (21.8)	10.7 (11.9)	15.8 (16.1)	16.8 (19.2)
Red and processed meat intake - 4.00+ times/wk, N (%)	12,961 (45.4%)	8,698 (30.5%)	13,033 (45.6%)	8,776 (30.7%)	14,489 (50.7%)	7,406 (25.9%)	10,084 (35.3%)	12.123 (42.5%)
Vegetable and fruit intake - g/day	325.7 (197.8)	453.2 (255.6)	271.9 (161.2)	506.6 (268.9)	199.8 (117.5)	590.4 (246.1)	449.1 (253.7)	318.7 (197.8)
Total carbohydrate intake - % energy	39.9 (4.6)	58.3 (3.5)	43.0 (6.9)	55.7 (5.5)	44.8 (7.4)	54.1 (6.3)	46.5 (8.2)	52.7 (6.5)

Table 7.3 continued

	Total carbohydrates		Total sugars		Total non-free sugar		Total free sugar	
	Q1	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**
Total sugar intake - % energy	19.0 (4.6)	30.2 (6.2)	16.5 (2.7)	32.9 (4.1)	19.5 (5.6)	30.4 (5.5)	20.4 (6.3)	29.3 (5.7)
Total starch intake - % energy	20.8 (4.9)	28.1 (5.6)	26.5 (6.3)	22.7 (5.0)	25.3 (6.0)	23.6 (5.5)	26.1 (6.4)	23.3 (5.1)
Total fibre intake - g/day	15.4 (5.0)	19.9 (6.1)	16.1 (5.2)	19.2 (6.2)	14.8 (4.7)	20.7 (6.1)	18.6 (6.0)	16.5 (5.3)
Total wholegrain fibre - g/day	2.6 (2.5)	4.5 (3.3)	3.1 (3.0)	3.9 (3.0)	2.7 (2.8)	4.2 (3.0)	3.7 (3.1)	3.3 (2.9)
Total energy intake – kJ/day	8,842 (2137)	8,186 (1919)	8,822 (2082)	8,238 (1978)	9,215 (2153)	7,875 (1816)	8,098 (1919)	8,928 (2103)

Values are mean (SD) unless otherwise indicated.

[a] values represent % of energy from carbohydrates, total sugars, non-free sugars, and free sugars.

Abbreviations: g, grams; g/d, grams per day; kJ, kilojoules; NSAID, non-steroid anti-inflammatory drug; N, Number of participants; Q, quantile; wk, week.

Table 7.4 Baseline characteristics by lowest and highest quartile of intake of fibre from vegetables, fruits, and wholegrains.

	Fibre from vegetables		Fibre from fruits		Fibre from wholegrains	
	Q1	**Q4**	**Q1**	**Q4**	**Q1**	**Q4**
No. of participants	28,555	28,554	28,594	28,553	28,655	27,817
Intake of carbohydrate of interest [a]	0.7 (0.5)	6.1 (2.1)	0.6 (0.4)	5.6 (1.9)	0.4 (0.5)	7.9 (2.3)
Sex - Male, N (%)	16,050 (56.2%)	10,264 (35.9%)	15,269 (53.4%)	11,148 (39.0%)	12,484 (43.6%)	14,640 (52.6%)
Age at recruitment - years	54.5 (7.9)	56.9 (7.7)	54.1 (7.9)	57.4 (7.4)	55.0 (7.9)	56.6 (7.7)
Body mass index – kg/m^2	27.2 (4.7)	26.5 (4.6)	27.2 (4.7)	26.4 (4.6)	27.3 (4.9)	26.0 (4.3)
Height - cm	170.7 (9.3)	168.4 (8.9)	170.4 (9.2)	168.8 (9.1)	168.9 (9.1)	170.8 (9.2)
Physical activity - High	4,220 (14.8%)	6,085 (21.3%)	4,202 (14.7%)	5,939 (20.8%)	4,656 (16.2%)	5,122 (18.4%)
Townsend deprivation index, N (%)						
Q1 - Most affluent	5,934 (20.8%)	6,075 (21.3%)	5,978 (20.9%)	6,226 (21.8%)	5,740 (20.0%)	6,393 (23.0%)
Q5 - Most deprived	5,082 (17.8%)	4,557 (16.0%)	5,074 (17.7%)	4,481 (15.7%)	5,413 (18.9%)	4,048 (14.6%)
In paid employment, N (%)	19,362 (67.8%)	16,519 (57.9%)	19,347 (67.7%)	16,353 (57.3%)	18,813 (65.7%)	16,465 (59.2%)
University/college degree, N (%)	20,146 (70.6%)	21,094 (73.9%)	19,869 (69.5%)	21,732 (76.1%)	20,157 (70.3%)	21,289 (76.5%)
White ethnicity	27,368 (95.8%)	27,618 (96.7%)	27,649 (96.7%)	27,407 (96.0%)	27,334 (95.4%)	27,185 (97.7%)
Never smoker, N (%)	16,246 (56.9%)	16,025 (56.1%)	15,209 (53.2%)	17,046 (59.7%)	15,161 (52.9%)	16,973 (61.0%)
Diabetes - Yes, N (%)	1,224 (4.3%)	1,076 (3.8%)	1,064 (3.7%)	1,135 (4.0%)	1,089 (3.8%)	1,047 (3.8%)
NSAID - regular user, N (%)	7,411 (26.0%)	7,066 (24.7%)	7,547 (26.4%)	6,904 (24.2%)	7,576 (26.4%)	6,543 (23.5%)
Postmenopausal at recruitment, N (%)	7,534 (60.3%)	13,220 (72.3%)	7,453 (56.0%)	13,191 (75.8%)	10,190 (63.0%)	9,280 (70.5%)
Menopausal hormone therapy use - Current use	1,079 (8.6%)	1,498 (8.2%)	1,069 (8.0%)	1,459 (8.4%)	1,397 (8.6%)	943 (7.2%)
Diet variables						
Alcohol intake - g/day	17.3 (18.5)	15.5 (16.6)	20.4 (20.4)	13.1 (14.2)	19.0 (20.4)	14.8 (14.6)
Red and processed meat intake - 4+ times/wk, N (%)	12,362 (43.3%)	9,412 (33.0%)	13,322 (46.6%)	8777 (30.7%)	12,603 (44.0%)	9,811 (35.3%)
Vegetable and fruit intake - g/day	2.8 (1.9)	9.7 (3.5)	2.9 (2.0)	9.6 (3.6)	5.1 (3.5)	6.4 (3.6)
Total carbohydrate intake - % energy	49.7 (7.5)	49.3 (7.7)	46.1 (7.6)	52.8 (6.8)	46.8 (8.2)	51.7 (6.4)
Total sugar intake - % energy	23.5 (6.7)	25.6 (6.7)	20.3 (5.9)	29.1 (6.1)	23.2 (7.2)	25.1 (5.9)
Total starch intake - % energy	26.1 (5.8)	23.5 (5.7)	25.8 (6.1)	23.6 (5.4)	23.6 (6.3)	26.5 (5.0)
Total fibre intake - g/day	14.5 (4.6)	22.2 (5.8)	14.3 (4.5)	22.2 (5.7)	14.4 (4.8)	22.1 (5.4)
Total fibre from wholegrains - g/day	3.4 (3.0)	3.9 (3.1)	2.9 (2.8)	4.3 (3.2)	0.4 (0.5)	7.9 (2.3)
Total energy intake - kJ/day	8,389 (1975)	8,866 (1975)	8,502 (1997)	8,786 (1944)	8,363 (2005)	9,025 (1912)

Values are mean (SD) unless otherwise indicated.

[a] values represent grams of fibre from vegetables, fruits, and wholegrains.

Abbreviations: g, grams; g/day, grams per day; kJ, kilojoules; N, Number of participants; NSAID, non-steroidal anti-inflammatory drug; Q, quantile; wk, week; y, years.

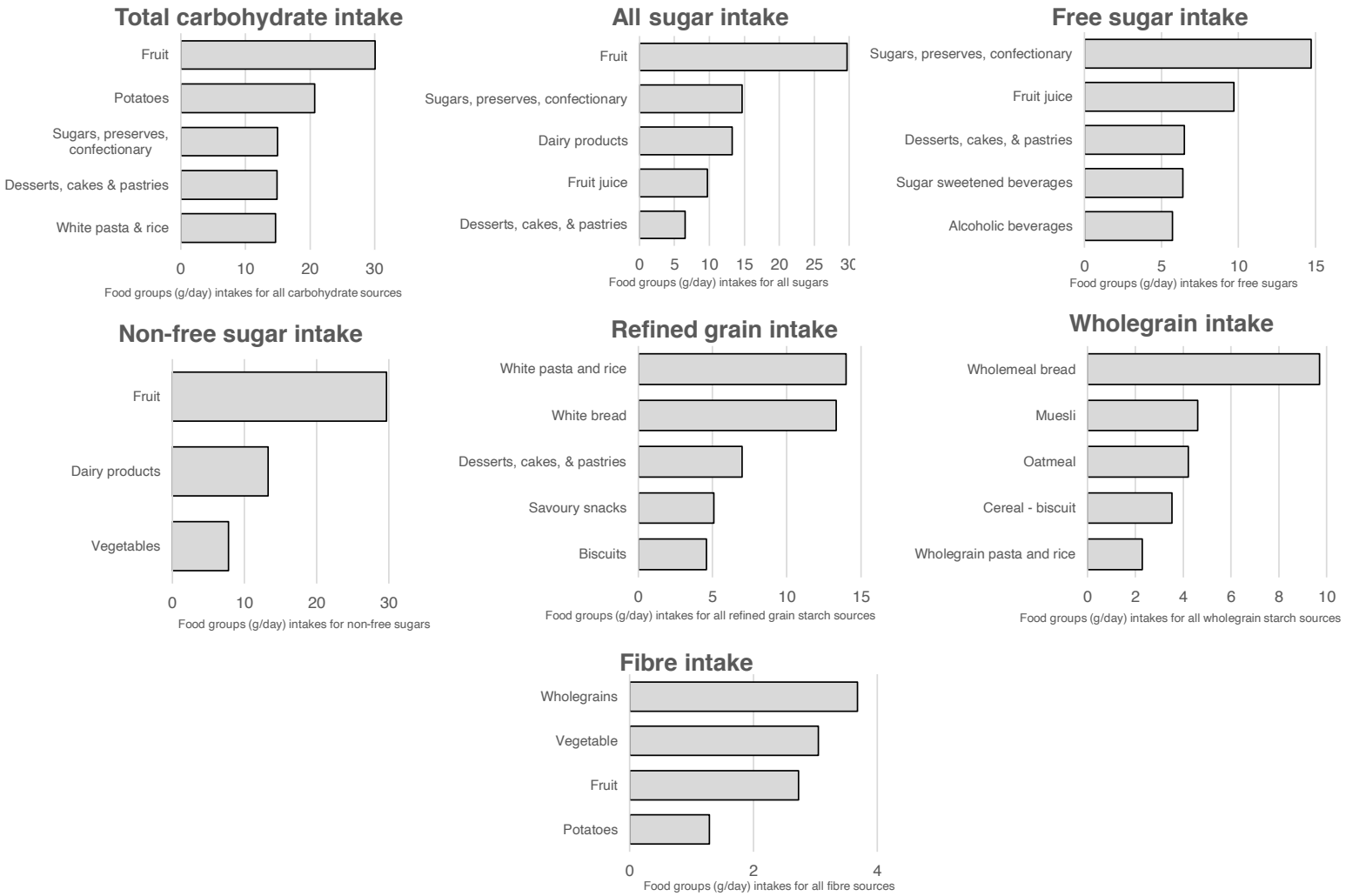


Figure 7.1 Top food group contributors to types and sources of carbohydrates for participants in the 24-hour dietary assessment subsample (n=114,217).

Abbreviation: g/day, grams per day.

7.4.2 Colorectal cancer risk by carbohydrate intake

Multivariable adjusted HRs for intakes of total carbohydrates and carbohydrates from different types and sources and colorectal cancer risk are shown in **Figure 7.2** (see **Table 7.5** for minimally adjusted results and sequential adjustments of potential confounders). Intake of total carbohydrates per 5% energy increase was associated with a 5% lower risk of colorectal cancer (HR: 0.95, 95% CI: 0.91-0.99; **Figure 7.2**); however, the χ^2 value was reduced by more than 70% compared to minimally adjusted models (**Table 7.5**). For every 5% higher energy intakes of total sugars and non-free sugars, an inverse association with colorectal cancer was observed (HR: 0.93, 95% CI: 0.88-0.98 and 0.92, 95% CI: 0.87-0.98, respectively; **Figure 7.2**). No significant association was observed for wholegrain starch and colorectal cancer risk (HR per 5% energy increase: 0.94, 95% CI: 0.87-1.01, *p-trend*=0.09); however, when wholegrain intake was modelled as absolute grams of wholegrain foods instead of starch from wholegrains, an inverse association with colorectal cancer risk was observed (HR per 50 g/day intake: 0.95, 95% CI: 0.91-0.98, p-value=0.004; **Table 7.6**). No other associations between carbohydrate intakes and colorectal cancer risk were observed.

7.4.3 Colorectal cancer risk by fibre intake

Multivariable adjusted HRs for fibre intake and colorectal cancer are presented in **Figure 7.3** (see **Table 7.7** for minimally adjusted results and sequential adjustments of potential confounders). No associations were observed for total fibre intake, however, a borderline inverse association for fibre from wholegrains with colorectal cancer risk was observed (HR per 5/day increase in intake: 0.90, 95% CI: 0.82-1.00, *p-trend*=0.047; **Figure 7.3**) and participants in the highest quartile of intake had a 19% lower risk of colorectal compared to those in the lowest quartile of intake (HR: 0.81, 95% CI: 0.69-0.96; **Figure 7.3**).

Table 7.5 Sequential adjustment hazard ratios and 95% confidence intervals for intakes of carbohydrate sources with colorectal cancer risk.

	Q1	Q2	Q3	Q4	Per 5% energy increment	χ^2	χ^2 – % change	P-trend
Total carbohydrates								
Minimally-adjusted*	1 (ref)	**0.83 (0.71 - 0.96)**	**0.82 (0.70 - 0.96)**	**0.71 (0.60 - 0.83)**	**0.91 (0.88 - 0.95)**	22.3		<0.001
Multivariable-adjusted**	1 (ref)	0.87 (0.74 - 1.01)	0.89 (0.76 - 1.05)	**0.79 (0.66 - 0.95)**	**0.94 (0.90 - 0.98)**	9.44	-57.6%	0.002
+ BMI	1 (ref)	0.87 (0.75 - 1.02)	0.90 (0.77 - 1.07)	**0.81 (0.68 - 0.97)**	**0.94 (0.90 - 0.98)**	7.98	-64.2%	0.005
+ Processed/red meat & vegetable and fruit intake	1 (ref)	0.88 (0.75 - 1.03)	0.92 (0.78 - 1.08)	0.83 (0.70 - 1.00)	**0.95 (0.91 - 0.99)**	5.88	-73.6%	0.015
Total sugar								
Minimally-adjusted*	1 (ref)	**0.81 (0.69 - 0.94)**	**0.74 (0.64 - 0.87)**	**0.67 (0.57 - 0.79)**	**0.90 (0.86 - 0.94)**	22.1		<0.001
Multivariable-adjusted**	1 (ref)	**0.84 (0.72 - 0.98)**	**0.79 (0.67 - 0.93)**	**0.73 (0.62 - 0.87)**	**0.92 (0.88 - 0.97)**	11.4	-48.4%	0.001
+ BMI	1 (ref)	**0.85 (0.72 - 0.99)**	**0.80 (0.68 - 0.94)**	**0.74 (0.63 - 0.88)**	**0.93 (0.88 - 0.97)**	10.3	-53.3%	0.001
+ Processed/red meat & vegetable and fruit intake	1 (ref)	**0.84 (0.72 - 0.99)**	**0.80 (0.68 - 0.94)**	**0.75 (0.63 - 0.90)**	**0.93 (0.88 - 0.98)**	7.58	-65.8%	0.006
Total free sugars								
Minimally-adjusted*	1 (ref)	1.09 (0.93 - 1.28)	0.98 (0.84 - 1.16)	0.90 (0.76 - 1.07)	0.97 (0.91 - 1.03)	1.13		0.29
Multivariable-adjusted**	1 (ref)	1.10 (0.93 - 1.28)	0.99 (0.84 - 1.16)	0.89 (0.76 - 1.06)	0.96 (0.91 - 1.02)	1.45	28.3%	0.23
+ BMI	1 (ref)	1.10 (0.94 - 1.29)	0.99 (0.84 - 1.17)	0.90 (0.76 - 1.07)	0.97 (0.91 - 1.03)	1.27	11.8%	0.26
+ Processed/red meat & vegetable and fruit intake	1 (ref)	1.08 (0.92 - 1.27)	0.97 (0.82 - 1.14)	0.87 (0.73 - 1.04)	0.95 (0.90 - 1.02)	2.24	97.7%	0.13
Total non-free sugars								
Minimally-adjusted*	1 (ref)	**0.82 (0.71 - 0.96)**	**0.72 (0.62 - 0.85)**	**0.68 (0.57 - 0.80)**	**0.87 (0.82 - 0.92)**	23.0		<0.001
Multivariable-adjusted**	1 (ref)	0.87 (0.74 - 1.01)	**0.79 (0.67 - 0.93)**	**0.77 (0.65 - 0.92)**	**0.91 (0.86 - 0.97)**	9.34	-59.4%	0.002
+ BMI	1 (ref)	0.87 (0.75 - 1.02)	**0.79 (0.67 - 0.94)**	**0.78 (0.65 - 0.93)**	**0.92 (0.86 - 0.97)**	8.57	-62.7%	0.003
+ Processed/red meat	1 (ref)	0.87 (0.75 - 1.02)	**0.80 (0.68 - 0.94)**	**0.79 (0.66 - 0.94)**	**0.92 (0.87 - 0.98)**	7.27	-68.4%	0.006

Table 7.5 continued

	Q1	Q2	Q3	Q4	Per 5% energy increment	χ^2	χ^2 – % change	P-trend
Total starch								
Minimally-adjusted*	1 (ref)	0.91 (0.78 - 1.07)	0.93 (0.80 - 1.09)	0.92 (0.78 - 1.08)	0.98 (0.93 - 1.03)	0.80		0.37
Multivariable-adjusted**	1 (ref)	0.93 (0.79 - 1.09)	0.97 (0.83 - 1.14)	0.98 (0.83 - 1.16)	1.00 (0.95 - 1.06)	0.00	-100.0%	0.99
+ BMI	1 (ref)	0.93 (0.80 - 1.09)	0.98 (0.83 - 1.15)	0.99 (0.84 - 1.17)	1.00 (0.95 - 1.06)	0.02	-97.9%	0.90
+ Processed/red meat & vegetable and fruit intake	1 (ref)	0.92 (0.79 - 1.08)	0.96 (0.82 - 1.13)	0.98 (0.82 - 1.16)	1.00 (0.94 - 1.06)	0.00	-99.6%	0.96
Wholegrain starch								
Minimally-adjusted*	1 (ref)	0.88 (0.75 - 1.03)	**0.84 (0.71 - 0.98)**	**0.77 (0.66 - 0.91)**	**0.89 (0.83 - 0.95)**	10.7		0.001
Multivariable-adjusted**	1 (ref)	0.89 (0.76 - 1.05)	0.87 (0.74 - 1.02)	**0.83 (0.70 - 0.97)**	**0.92 (0.85 - 0.99)**	5.26	-51.1%	0.022
+ BMI	1 (ref)	0.90 (0.77 - 1.06)	0.88 (0.75 - 1.04)	0.85 (0.72 - 1.00)	0.93 (0.87 - 1.00)	3.73	-65.3%	0.053
+ Processed/red meat & vegetable and fruit intake	1 (ref)	0.90 (0.77 - 1.06)	0.89 (0.76 - 1.05)	0.86 (0.73 - 1.02)	0.94 (0.87 - 1.01)	2.90	-73.0%	0.089
Refined grain starch								
Minimally-adjusted*	1 (ref)	0.93 (0.80 - 1.10)	1.07 (0.91 - 1.25)	1.09 (0.93 - 1.29)	1.04 (0.99 - 1.09)	2.10		0.15
Multivariable-adjusted**	1 (ref)	0.92 (0.78 - 1.08)	1.05 (0.89 - 1.23)	1.09 (0.92 - 1.28)	1.04 (0.99 - 1.09)	2.06	-1.6%	0.15
+ BMI	1 (ref)	0.92 (0.78 - 1.08)	1.04 (0.89 - 1.22)	1.08 (0.91 - 1.27)	1.03 (0.98 - 1.08)	1.77	-15.4%	0.18
+ Processed/red meat & vegetable and fruit intake	1 (ref)	0.91 (0.77 - 1.06)	1.02 (0.87 - 1.20)	1.05 (0.89 - 1.25)	1.03 (0.98 - 1.08)	1.06	-49.7%	0.30

*Minimally-adjusted model stratified for sex, age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years) and region (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland)

**Multivariable adjusted model further adjusting for height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, female specific covariates: menopausal hormone therapy use and menopausal status.

+ BMI further adjusting for body mass index categories to multivariable adjusted models.

+ processed/red meat & total vegetable and fruit intake further adjusting for red and processed meat consumption reported at recruitment and fruit and vegetables intake measured by averaged 24-hour dietary assessments to multivariable adjusted models + BMI.

χ^2 and p-trend represent improvement of fit obtained from likelihood ratio tests for including the carbohydrate type/source (modelled as a 5% energy increase).

χ^2 - % change calculated from the percentage difference in χ^2 value using the minimally adjusted χ^2 value as the reference group.

Abbreviations: CI, confidence intervals; BMI, body mass index; HR, hazard ratio; N, number of participants; Q, quartile; ref, reference group.

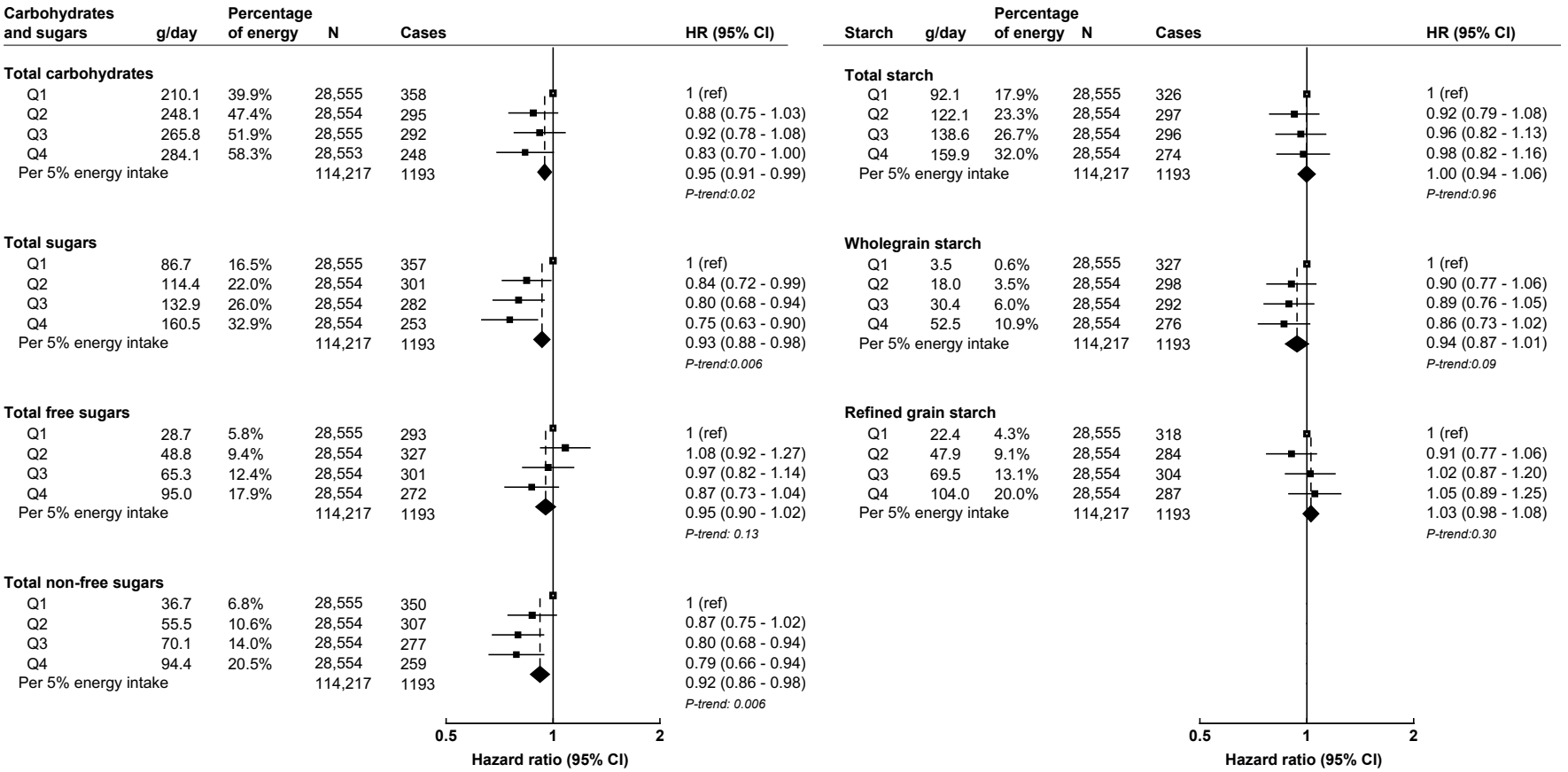


Figure 7.2 Multivariable-adjusted hazard ratios (95% CI) for percentage of energy intakes of carbohydrates, sugars and starches with colorectal cancer risk.

All models are stratified by sex, age at recruitment and adjusted for region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, red and processed meat intake, fruit and vegetables intake, energy intake, and female specific covariates: menopausal hormone therapy use, and menopausal status.

Grams and percentage of energy from carbohydrate, sugars, and starch calculated as mean per day within each quartile.

Abbreviations: CI, confidence intervals; HR, hazard ratio; N, number of participants; Q, quartile; ref, reference group.

Table 7.6 Multivariable adjusted hazard ratios and 95% confidence intervals for wholegrain and refined grain intake measured from total food weight of wholegrain and refined grain sources.

Wholegrains	**N/Cases**	**Grams**	**HR (95% CI)**
Q1	28,927 / 329	8.71 (10.8)	1 (ref)
Q2	28,184 / 301	52.0 (12.3)	0.93 (0.79 - 1.08)
Q3	28,604 / 294	101.2 (17.2)	0.88 (0.745- 1.03)
Q4	28,502 / 269	204.5 (61.8)	**0.79 (0.67 - 0.94)**
Per 50 g/day	-	-	**0.95 (0.91 - 0.98)**
p-value	-	-	0.004

Refined grains	**N/Cases**	**Grams**	**HR (95% CI)**
Q1	28,560 / 299	54.9 (25.2)	1 (ref)
Q2	28,565 / 290	120.5 (16.3)	0.95 (0.81 - 1.12)
Q3	28,546 / 302	180.6 (19.5)	1.00 (0.85 - 1.19)
Q4	28,546 / 302	293.5 (70.8)	1.05 (0.87 - 1.26)
Per 50 g/day	-	-	1.02 (0.98 - 1.05)
p-value	-	-	0.37

Models stratified for sex and age at recruitment, and further adjusted for region, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, BMI, processed and red meat intake, total vegetable fruit intake, and female specific covariates: menopausal hormone therapy use and menopausal status.

Grams of wholegrains and refined grains is the mean and standard deviation of reported intakes in each quartile.

P-values represent the p for trend for wholegrains and refined grains modelled as 50 g/day increase in the model.

Abbreviations: CI, confidence intervals; g/day, grams per day; HR, hazard ratio; N, number of participants; Q, quartile; ref, reference group.

Table 7.7 Sequential adjustment hazard ratios and 95% confidence intervals for intakes of fibre sources with colorectal cancer risk.

Fibre and fibre sources	Q1	Q2	Q3	Q4	Per 5 g/day increment	χ^2	χ^2 – % change	P-trend
Total fibre								
Minimally-adjusted*	1 (ref)	1.11 (0.94 - 1.30)	1.01 (0.86 - 1.19)	0.99 (0.84 - 1.17)	0.97 (0.92 - 1.02)	1.37		0.24
Multivariable-adjusted**	1 (ref)	1.07 (0.90 - 1.27)	0.96 (0.81 - 1.15)	0.92 (0.76 - 1.11)	**0.94 (0.89 - 1.00)**	3.98	189.3%	0.05
+ BMI	1 (ref)	1.08 (0.91 - 1.27)	0.98 (0.82 - 1.17)	0.95 (0.78 - 1.14)	0.95 (0.89 - 1.01)	2.81	104.2%	0.09
+ Processed/red meat intake	1 (ref)	1.08 (0.91 - 1.28)	0.99 (0.83 - 1.18)	0.97 (0.80 - 1.17)	0.96 (0.90 - 1.02)	1.81	32.0%	0.18
Total fibre from vegetables & fruit								
Minimally-adjusted*	1 (ref)	0.95 (0.81 - 1.12)	0.99 (0.84 - 1.16)	0.91 (0.77 - 1.07)	0.93 (0.85 - 1.01)	3.03		0.08
Multivariable-adjusted**	1 (ref)	0.97 (0.82 - 1.14)	1.01 (0.85 - 1.19)	0.92 (0.78 - 1.09)	0.93 (0.85 - 1.02)	2.59	-14.6%	0.11
+ BMI	1 (ref)	0.98 (0.83 - 1.15)	1.02 (0.86 - 1.20)	0.93 (0.79 - 1.11)	0.94 (0.86 - 1.02)	2.21	-27.1%	0.14
+ Processed/red meat intake	1 (ref)	0.98 (0.83 - 1.15)	1.02 (0.87 - 1.20)	0.94 (0.80 - 1.12)	0.94 (0.86 - 1.03)	1.82	-40.0%	0.18
Total fibre from vegetables								
Minimally-adjusted*	1 (ref)	1.09 (0.92 - 1.28)	1.03 (0.87 - 1.21)	1.03 (0.87 - 1.21)	0.97 (0.86 - 1.11)	0.16		0.69
Multivariable-adjusted**	1 (ref)	1.09 (0.92 - 1.28)	1.02 (0.87 - 1.21)	1.01 (0.85 - 1.19)	0.96 (0.84 - 1.09)	0.45	186.4%	0.50
+ BMI	1 (ref)	1.09 (0.93 - 1.29)	1.03 (0.88 - 1.22)	1.02 (0.86 - 1.20)	0.96 (0.84 - 1.10)	0.35	121.8%	0.55
+ Processed/red meat intake	1 (ref)	1.09 (0.93 - 1.29)	1.04 (0.88 - 1.22)	1.02 (0.86 - 1.21)	0.96 (0.84 - 1.10)	0.31	93.9%	0.58
Total fibre from fruits								
Minimally-adjusted*	1 (ref)	0.96 (0.82 - 1.13)	0.85 (0.72 - 1.00)	**0.83 (0.70 - 0.97)**	**0.84 (0.73 - 0.97)**	5.99		0.01
Multivariable-adjusted**	1 (ref)	0.99 (0.85 - 1.16)	0.89 (0.75 - 1.05)	0.87 (0.73 - 1.03)	**0.87 (0.75 - 1.00)**	3.70	-38.3%	0.05
+ BMI	1 (ref)	1.00 (0.85 - 1.17)	0.90 (0.76 - 1.06)	0.88 (0.74 - 1.04)	0.88 (0.76 - 1.01)	3.27	-45.4%	0.07
+ Processed/red meat intake	1 (ref)	1.00 (0.85 - 1.18)	0.90 (0.76 - 1.06)	0.89 (0.75 - 1.05)	0.89 (0.77 - 1.03)	2.62	-56.3%	0.11

Table 7.7 continued

	Q1	Q2	Q3	Q4	Per 5 g/day increment	χ^2	χ^2 – % change	P-trend
Total fibre from wholegrains								
Minimally-adjusted*	1 (ref)	0.87 (0.74 - 1.02)	0.90 (0.77 - 1.06)	**0.77 (0.65 - 0.90)**	**0.87 (0.79 - 0.96)**	7.82		0.01
Multivariable-adjusted**	1 (ref)	0.88 (0.75 - 1.04)	0.93 (0.79 - 1.08)	**0.78 (0.66 - 0.92)**	**0.88 (0.80 - 0.97)**	6.53	-16.5%	0.01
+ BMI	1 (ref)	0.89 (0.76 - 1.05)	0.94 (0.80 - 1.10)	**0.80 (0.68 - 0.95)**	**0.90 (0.81 - 0.99)**	4.84	-38.1%	0.03
+ Processed/red meat & vegetable and fruit intake	1 (ref)	0.89 (0.76 - 1.05)	0.94 (0.81 - 1.11)	**0.81 (0.69 - 0.96)**	**0.90 (0.82 - 1.00)**	3.94	-49.6%	0.047

*Minimally-adjusted model stratified for sex, age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years) and region (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland).

**Multivariable adjusted model further adjusted for height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, female specific covariates: menopausal hormone therapy use and menopausal status.

+ BMI further adjusted for body mass index categories to multivariable adjusted models.

+ processed/red meat & total vegetable fruit intake further adjusted for red and processed meat consumption reported at recruitment and fruit and vegetables intake measured by averaged 24-hour dietary assessments to multivariable adjusted models + BMI.

χ^2 and p-trend represent improvement of fit obtained from likelihood ratio tests for including fibre source (modelled as a 5 gram/day increase) variable into the model.

χ^2 - % change calculated from the percentage difference in χ^2 value using the minimally adjusted χ^2 value as the reference group.

Abbreviations: g/day, grams per day; Q, quartile; ref, reference group.

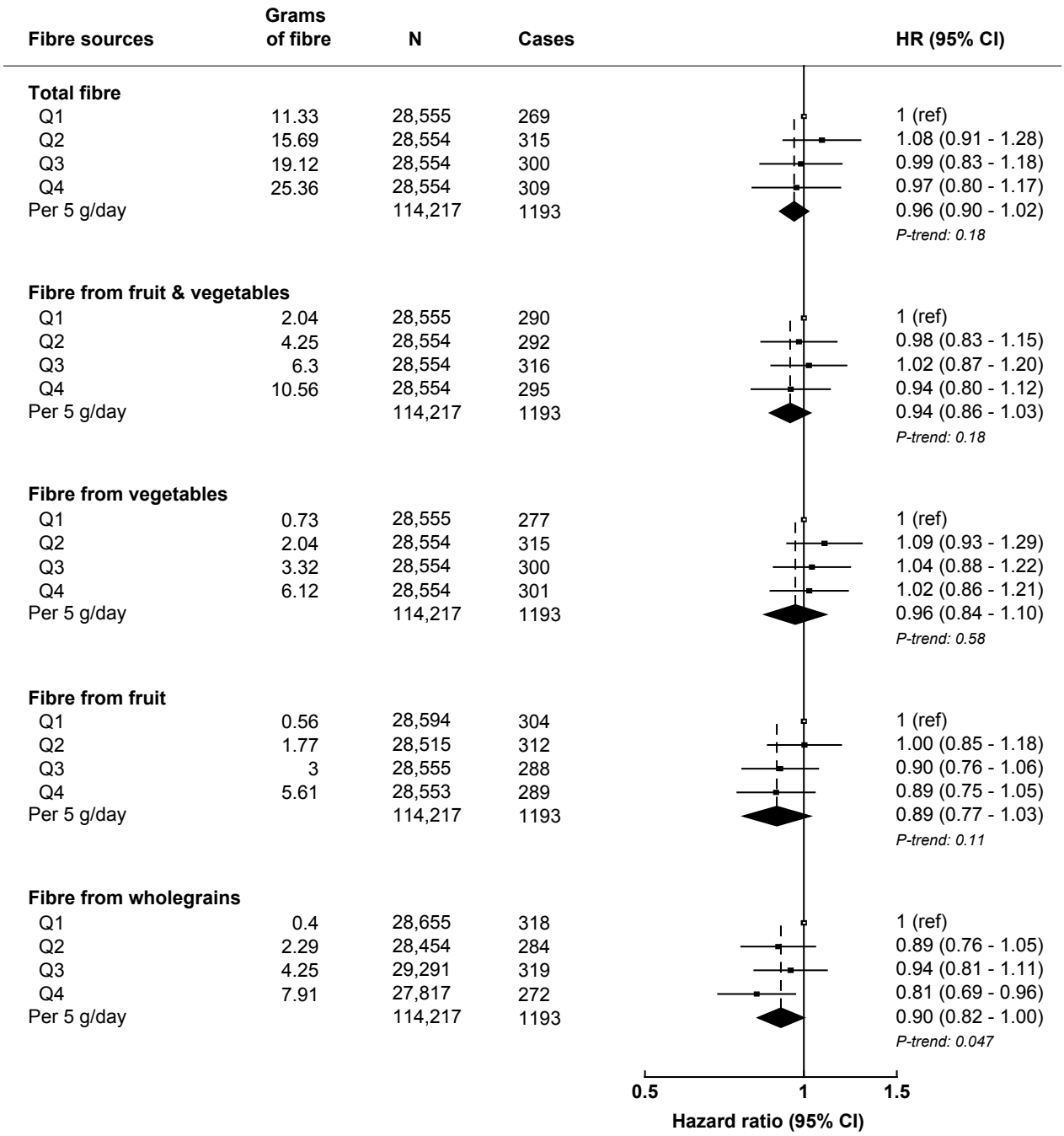


Figure 7.3 Multivariable-adjusted hazard ratios (95% CI) for intakes of fibre and fibre sources with colorectal cancer risk.

All models are stratified by sex, age at recruitment, and adjusted for region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, red and processed meat intake, fruit and vegetable intake (except for total fibre, and fibre from fruits and/or vegetables), energy intake, and female specific covariates: menopausal hormone therapy use, and menopausal status.

Grams of fibre calculated as mean intake per day within each quartile.

Abbreviations: CI, confidence intervals; g/day, grams per day; HR, hazard ratio; N, number of participants.

7.4.4 Short-chain fatty acid genetic modifiers

In participants with a high PGS for butyrate, intake of wholegrain starch was inversely associated with colorectal cancer risk (HR per 5% energy increase: 0.88, 95% CI: 0.78-0.99) whereas no association was observed between intake of wholegrain starch and colorectal cancer risk for individuals with low host butyrate PGS (HR: 1.09, 95% CI: 0.96-1.22; $p\text{-}_{heterogeneity}$ = 0.023; **Figure 7.4**). Participants consuming higher amounts of total starch and having a low butyrate PGS had a higher risk of colorectal cancer (HR per 5% energy increase: 1.14; 95% CI: 1.04-1.25) whereas no association was observed between starch intake and risk among individuals with a high butyrate PGS (HR: 0.93, 95% CI: 0.85-1.02; $p\text{-}_{heterogeneity}$ = 0.012; **Figure 7.4**).

Replication in a larger UK Biobank sample

In additional analyses utilising the fibre from bread and cereals derived from dietary questions asked in the touchscreen questionnaire completed by the larger UK Biobank sample (n=343,621), there was evidence of heterogeneity observed by the butyrate PGS: a 5 g/day higher intake of fibre from breads and cereals was inversely associated with colorectal cancer risk for individuals with a high butyrate PGS (HR per 5 g/day increase: 0.88; 95% CI: 0.81-0.95) whereas no association was observed for individuals in the low butyrate PGS (HR: 1.00 95% CI: 0.92-1.08; $p\text{-}_{heterogeneity}$=0.021; **Figure 7.5**); similar associations observed for the butyrate PGS were also observed for the propionate PGS, however the test for heterogeneity was non-significant ($p\text{-}_{heterogeneity}$ = 0.26; **Figure 7.5**).

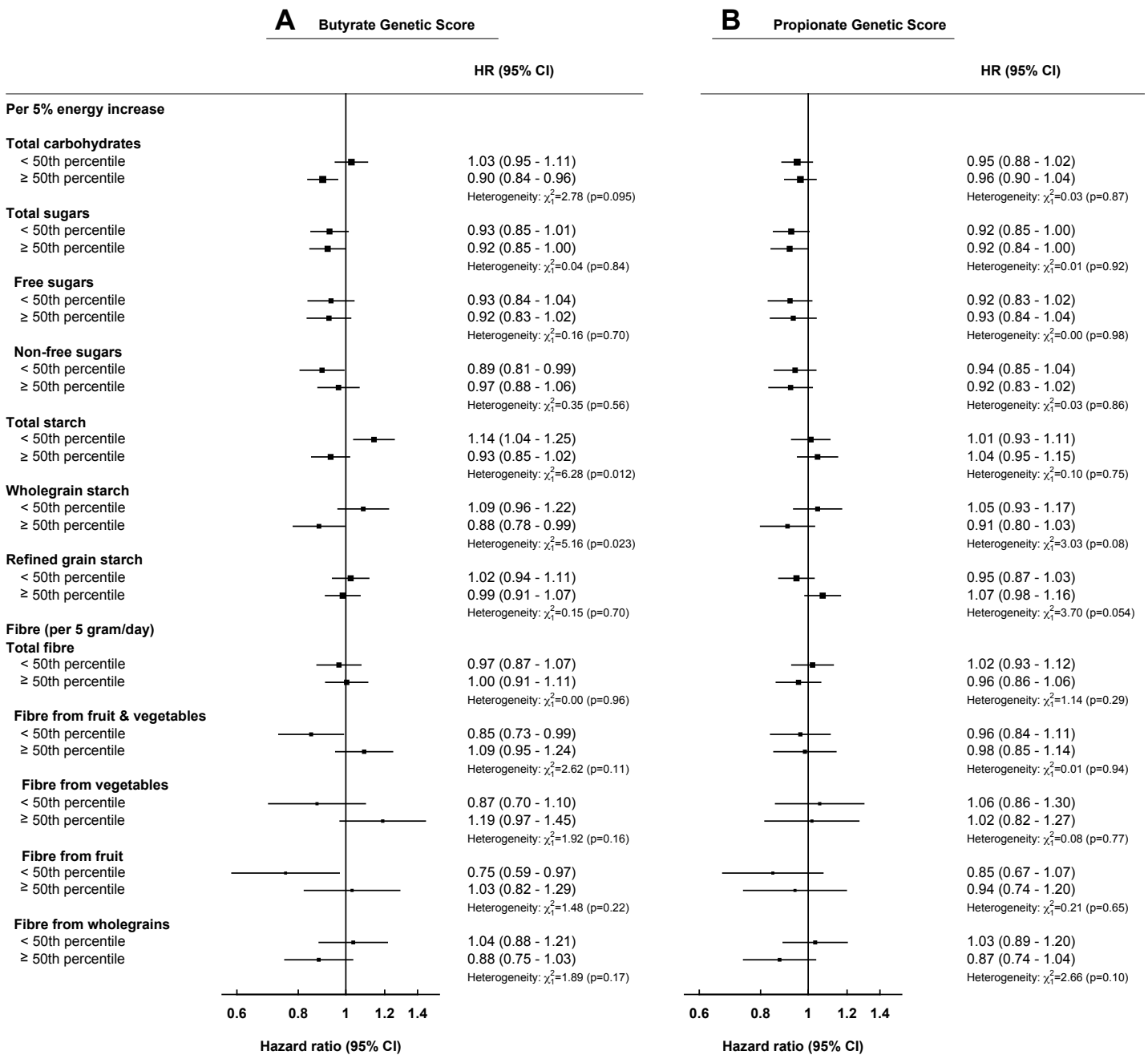


Figure 7.4 Multivariable-adjusted hazard ratios (95% CI) for intake of carbohydrates and fibre and colorectal cancer risk by genetically predicted host short-chain fatty acid production for A) butyrate and B) propionate and intake of carbohydrates and fibre (n=87,453).

All models are stratified by sex, age at recruitment, and adjusted region of recruitment, first 10 genetic principal components, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, diabetes status, non-steroidal anti-inflammatory drug use, red and processed meat intake, fruit and vegetable intake (except when fibre from vegetables and/or fruits, and non-free sugar intake was the exposure), serum IGF-I concentrations, energy intake, and female specific covariates: menopausal hormone therapy use, and menopausal status. Analyses are restricted to white British participants.

χ^2 and p-value represents improvement of fit obtained from likelihood ratio tests for including an interaction term between butyrate or propionate polygenic score and carbohydrate type/source (modelled as a 5% energy increase) or fibre source (modelled as a 5 gram/day increase) into the model. Abbreviations: CI, confidence intervals; HR, hazard ratio.

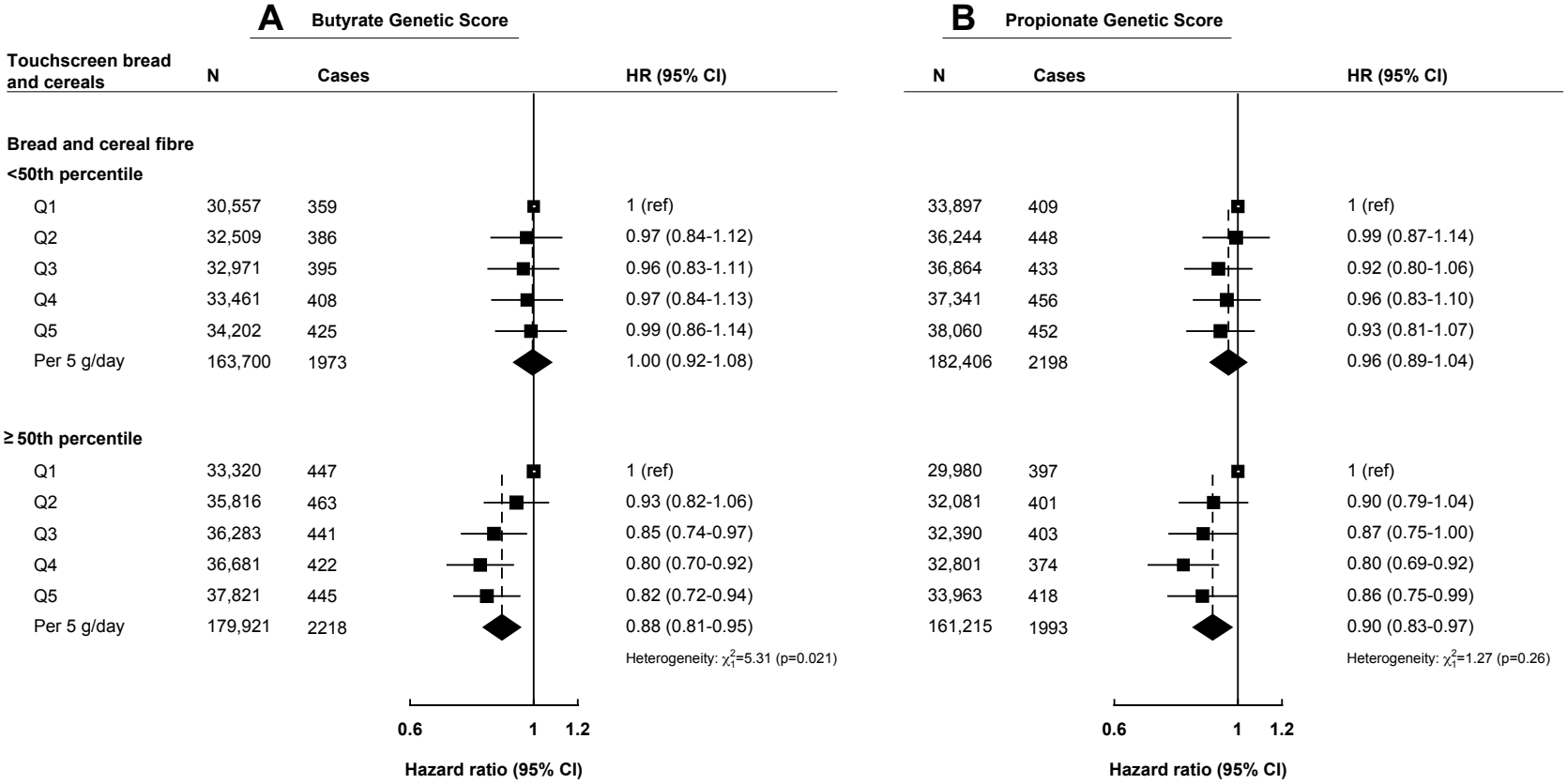


Figure 7.5 Multivariable-adjusted hazard ratios (95% CI) between intake of fibre from breads and cereals from the touchscreen questionnaire and colorectal cancer risk by genetically predicted host short-chain fatty acid production for A) butyrate and B) propionate (n=343,621).

Models stratified for sex and age at recruitment, and further adjusted for region, first 10 principal components, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, diabetes status, non-steroidal anti-inflammatory drug use, body mass index, processed and red meat intake, and female specific covariates: menopausal hormone therapy use and menopausal status.

χ^2 and p-value represents improvement of fit obtained from likelihood ratio tests for including an interaction term between butyrate or propionate polygenic score and fibre from breads and cereals (modelled as a 5 gram/day increase) into the model. Abbreviations: CI, confidence intervals; HR, hazard ratio; ref, reference group.

7.4.5 Subgroup and sensitivity analyses

No evidence of heterogeneity across sex and BMI categories was observed for the main analyses (24-hour dietary assessments; Ancillary Figures 6.1 & 6.2). Some evidence of heterogeneity was observed when looking at tumour sites (i.e., colon and rectal); non-free sugar intake was inversely associated with rectal cancer risk (HR: 0.81, 95% CI: 0.73-0.90) but was not associated with colon cancer risk (HR: 0.98, 95% CI: 0.91-1.05; $p_{\text{-heterogeneity}}$=0.003; **Figure 7.6**). Similarly, evidence of heterogeneity was observed for intake of fibre from fruit, which was inversely associated with rectal cancer risk (HR: 0.62, 95% CI: 0.47-0.83) but was not associated with colon cancer risk (HR: 1.04 95% CI: 0.87-1.25; $p_{\text{-heterogeneity}}$=0.003; **Figure 7.6**).

In sensitivity analyses restricting to participants who completed a minimum of three 24-hour dietary assessments (total n=69,223 and n of cases=734) and removing the first two years of follow-up (total N=111,724; and n of cases = 978), associations remained largely unchanged with CIs widening probably due to a smaller number of cases and number of participants, thus reduced statistical power (Ancillary Tables 7.1 & 7.2).

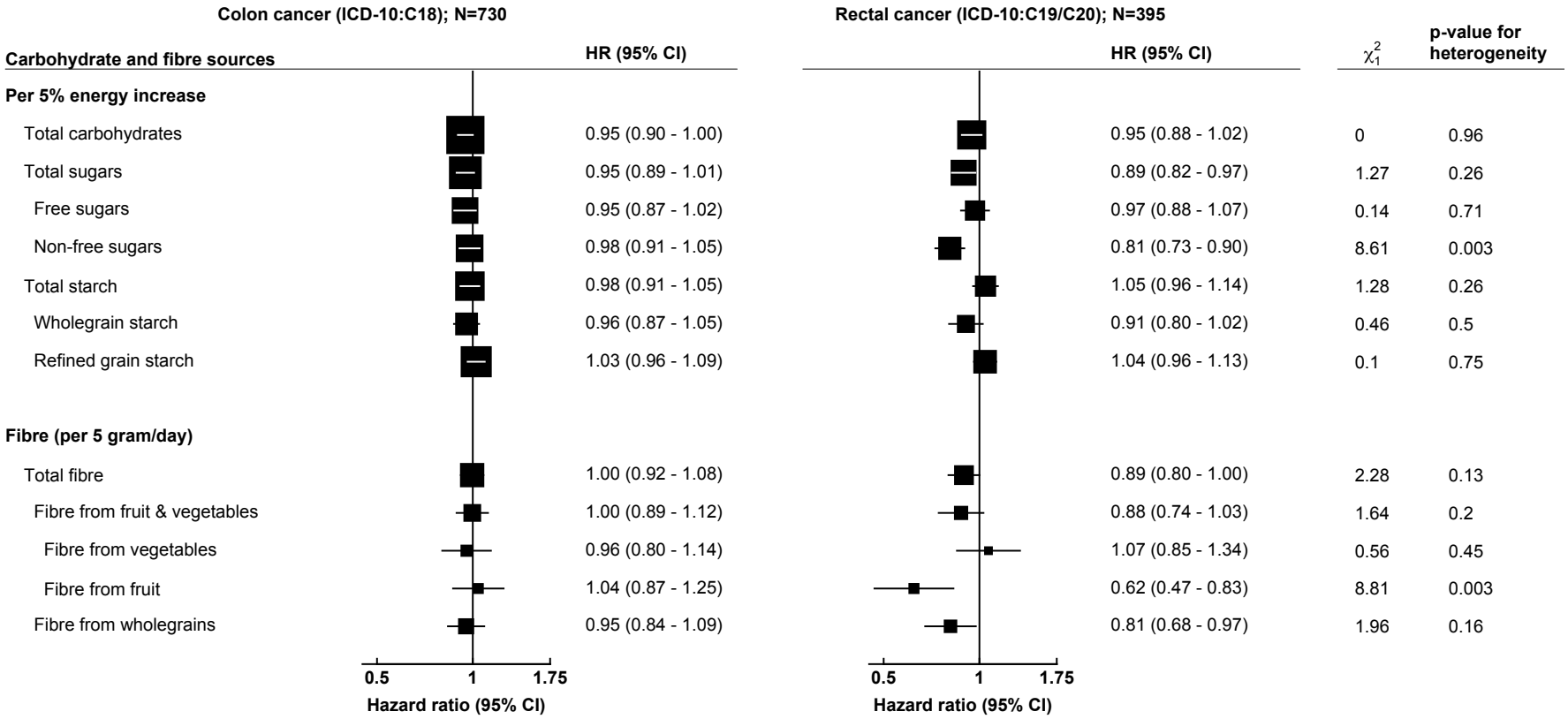


Figure 7.6 Multivariable hazard ratios and 95% confidence intervals for intake of carbohydrates and fibre with colorectal cancer risk separated by tumour site (colon and rectal).

Models stratified for sex and age at recruitment, and further adjusted for region, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, body mass index, processed and red meat intake, total vegetable fruit intake (except for total fibre, fibre from vegetables and/or fruit, or non-free sugars), and female specific covariates: menopausal hormone therapy use and menopausal status.

χ^2 and p-value for heterogeneity represent Wald's test for heterogeneity between tumour subgroups and carbohydrate types/sources (per 5% energy increase) or fibre source (per 5 g/day increment).

Abbreviations: CI, confidence interval; g/day, grams per day; HR, hazard ratio; N, number of participants.

7.5 Discussion

7.5.1 Main findings

In this analysis of 114,217 participants from the UK Biobank, intakes of carbohydrates, total sugar, non-free sugars, total wholegrains, and fibre from wholegrains were inversely associated with colorectal cancer risk. Moreover, suggestive evidence of effect modification was found by host genetically predicted SCFA production, with higher intakes of wholegrain starch being inversely associated with colorectal cancer risk only in those with a high host genetically predicted butyrate PGS. Similar evidence of heterogeneity was observed for intake of fibre from cereals and breads in the larger UK Biobank sample (n=343,621) with less detailed dietary assessment.

7.5.2 Carbohydrate types and sources

In this study, intakes of total and non-free sugars were inversely associated with colorectal cancer risk. The majority of intake of total sugars in this cohort was from non-free sugar sources and this is likely why an inverse association was also observed for total sugar intake. Participants consuming a higher percentage of energy from non-free sugars in this cohort consumed greater amounts of wholegrains as well as fruits and vegetables, for which limited evidence has suggested that they are protective factors for colorectal cancer [20]. The association of non-free sugars with colorectal cancer may also be partly due to intake of dairy products, a source of non-free sugars (~23% of all non-free sugars) in this population [362]. Dairy product consumption has been suggested as a probable factor for reducing the risk of colorectal cancer [20, 441], potentially due to the high calcium content [409]. However, associations remained when sugars from dairy products were removed from this exposure although the association were of a smaller magnitude (data not shown). In contrast, there was no association between free sugar intake and colorectal cancer risk, which is similar to previous findings from prospective cohort studies looking at intake of sugar (fructose) and sugar sweetened beverages [442, 443]. Evidence of heterogeneity was observed by tumour site with an inverse association for non-free

sugars for rectal cancer, but not for colon cancer. Existing evidence does not suggest differences in colorectal cancer risk by tumour sites for dietary factors such as fibre, wholegrains, and processed meat [20] and mechanisms for any potential difference are unclear; therefore, these results may be a result of chance. However, some evidence does suggest that certain risk factors for colorectal cancer do vary by anatomical sites [21], with potential differences in risk being attributable to microbial and physiological variation through the large intestine [444] but whether this varies for dietary factors remains unknown.

Higher intakes of total carbohydrates were inversely associated with colorectal cancer risk, however, the χ^2 value was reduced significantly from minimally adjusted models with additional adjustment of confounders, which suggests residual confounding may operate [149]. Higher intakes of carbohydrate sources of low nutritional quality, such as refined grains and free sugars, have been associated with higher risks of hyperinsulinemia and obesity [428], which are well-established risk factors for colorectal cancer [20, 108]. However, in this analysis no associations between refined grain and free sugar intake and colorectal cancer risk were observed, for both analyses that did and did not adjust for BMI. Previous prospective studies have assessed the associations of intakes of refined grains and sugars with colorectal cancer risk, with some results showing a positive association between sucrose or fructose intake and colorectal cancer risk [445, 446], while meta-analyses and other prospective studies have found no associations between refined grain or sugar intake and colorectal cancer risk [20, 443, 447-449].

In the latest WCRF meta-analysis from 2017, a higher intake of dietary fibre was associated with a 7% lower risk of colorectal cancer per 10 g/day increment. Total fibre intake was not associated with colorectal cancer risk in the main analyses, although an inverse association with fibre from wholegrains was observed. Tests for heterogeneity also suggested differing associations by tumour site and an inverse association for intake of fibre from fruit with risk of rectal cancer but no association with the risk of colon cancer. In a meta-analysis of prospective studies, no evidence of heterogeneity by anatomical site was observed for the association of intake of dietary

fibre from fruit with risk [110]. The current WCRF meta-analysis assessing the intake of fibre from fruit and colorectal cancer risk suggested a non-significant inverse association; however, tumour site heterogeneity was not reported possibly due to the absence of studies examining differences in risk by tumour site [20].

In these analyses participants with higher intake of fibre from wholegrains had a lower risk of colorectal cancer, and there was also suggestion of an inverse association with wholegrain starch. In the latest WCRF meta-analysis from 2017 that included six prospective cohort studies, intake of wholegrain foods (per 90 g/day) was associated with a 17% reduction in colorectal cancer risk [20]; a slightly weaker associations for intake of wholegrain foods (5% lower colorectal cancer risk per 50 g/day intake, which is equivalent to 9% lower per 90 g/day) was observed. In a previous analysis of the UK Biobank study using the short food frequency questionnaire (FFQ) with broad questions and few food items that were completed by all participants, intake of fibre from breads and cereals was associated with a 10% reduction of risk of colorectal cancer when modelled as a 5 g/day increase [439]. In this subsample with 24-hour dietary assessment data, fibre from wholegrains was infrequently consumed, which may introduce potential confounding if individuals who are consuming the most were more health conscious. However, the intake of wholegrains and fibre from wholegrains has been suggested to reduce the risk of colorectal cancer through several potential mechanisms including reduced bowel transit time due to their high fibre content, their hypothesised anti-carcinogenic bioactive compounds (such as phenolic compounds and minerals), as well as the possible prevention of insulin resistance [20, 324].

7.5.3 Short-chain fatty acids genetic modifiers

One hypothesis that has garnered interest recently is the potential lower risk of colorectal cancer due to the production of SCFAs from fibre and wholegrain fermentation by microbiota in the colon. Butyrate may have an anti-carcinogenic effect [109, 430-432], however, as well as diet, other factors can regulate the gut microbiome and the production of SCFA [450-452], including human genetic factors. In these analyses, heterogeneity by host genetically predicted butyrate

production was observed where only participants with a high PGS for butyrate production and consuming high amounts of wholegrain starch had a significant lower risk of colorectal cancer, and there was a non-significant positive association between wholegrain starch and risk of colorectal cancer for participants in the low butyrate PGS category. The non-significant positive association in the low butyrate PGS category contributed to the significant heterogeneity observed by wholegrains starch intake, which was unexpected as evidence has suggested that higher wholegrain intake is inversely associated with colorectal cancer risk [20]. Evidence of heterogeneity was not observed for fibre from wholegrains by the butyrate PGS, although similar trends were observed as wholegrain starch. This discrepancy could be due to fibre intake in the Nutrient Databank being calculated from non-starch polysaccharides, and resistant starch, which produces SCFA via microbiota fermentation in the colon [453], is excluded from this calculation. Thus, future work should examine the association between wholegrain fibre calculated from the total amount of non-digestible polysaccharides and colorectal cancer by butyrate PGS. When sensitivity analyses were performed using the larger UK Biobank cohort with the fibre from bread and cereals estimated from the few relevant dietary questions completed at recruitment, significant heterogeneity was observed by butyrate PGS, with an inverse association between the intake of fibre from breads and cereals and colorectal cancer only observed in those in the high butyrate PGS category. SCFAs, such as butyrate, may modulate cell histone deacetylases in the colorectum, and function to reinforce cellular junctions, which may reduce low-grade inflammation, and lead to a lower risk of colorectal cancer [92, 432, 454, 455]. Several studies have also shown that there is a lower abundance of butyrate producing bacteria among individuals with colorectal cancer compared to control participants [109, 456]. Some RCTs have shown that a higher wholegrain intake is associated with a higher abundance of SCFA-producing bacteria [429, 457] and higher levels of SCFA in fecal samples compared to at recruitment to the trial [458], however, other randomised trials have found no differences in the gut microbiota composition with higher wholegrain intake [459, 460], but have observed higher levels of fecal butyrate [459]. Evidence from mouse models has suggested that fibre and wholegrains are not protective against colorectal cancer without microbiota that can produce

SCFA, such as butyrate [92, 461]. Thus, the heterogeneity observed for wholegrain intake and colorectal cancer risk by butyrate PGS (which was shown to be associated with butyrate producing bacteria) may underline the importance of butyrate production in reducing colorectal carcinogenesis, and further research in humans is needed to confirm these findings.

Heterogeneity by the butyrate PGS was also observed for the association of total starch intake and colorectal cancer risk; participants in the low butyrate PGS category and consuming higher amounts of total starch had a higher risk of colorectal cancer, whereas no association was observed for those in the high butyrate PGS. Higher intake of certain sources of starch, such as potatoes, may promote a higher insulin response [462] and the butyrate PGS has been associated with an improved insulin response [436]; however, it remains unclear if this is the reason why there is a difference between butyrate PGS categories and starch intake, and these results may be due to chance because of multiple comparisons. Future research is needed to attempt to replicate these findings and better understand how SCFA production and human genetics may interact with dietary intake and influence the risk of colorectal cancer.

7.5.4 Strengths and limitations

Strengths of this analysis include the use of multiple 24-hour dietary assessments to reduce random measurement error and the prospective nature of this study using linkage to health records, thus minimising loss to follow-up. As well, due to the design of the 24-hour dietary assessments, analyses could include sources and types of carbohydrates in detail to assess if associations with colorectal cancer risk differed. The UK Biobank also collected detailed information from participants regarding socioeconomic and lifestyle factors and took standardised physical measurements, allowing us to adjust for these potential confounders in the analysis. In addition, genotyping was conducted in nearly the entire cohort allowing to separate participants by host SCFA genetic factors.

Several limitations need to be considered. Numerous tests were performed, and if correction for multiple testing was undertaken no significant findings would be observed, thus these results

need to be interpreted cautiously. All self-reported dietary intake is subject to error and some participants (~40,000) only completed two 24-hour dietary assessments and will therefore be affected by random error which biases the results to the null. Although analyses adjusted for multiple confounders, unmeasured and residual confounding may still influence the associations observed. However, formal assessment of residual confounding using the change in χ^2 method did not suggest residual confounding apart from total carbohydrate intake where the χ^2 value decreased more than 73.6% in comparison to minimally adjusted models. Reverse causality is also plausible due to colorectal cancer symptom onset potentially altering the consumption of specific foods. However, in sensitivity analyses, the first two years of follow-up were removed, and estimates remained the same although with wider CIs. The UK Biobank is generally a healthy sample, with the subsample of participants completing the 24-hour dietary assessments being healthier and due to this selection, results may be biased. Moreover, as the UK Biobank is predominantly white, the results may not be generalisable to a wider UK population; however, estimates are likely to be in the same direction [425]. As well, with just over 1,000 cases of colorectal cancer in the main analyses, results may also be underpowered to detect modest associations.

For analyses assessing modification by SCFA genetic variants, the genetic scores used were derived from one sample of 952 individuals [436], limiting the power of these genetic variants in predicting differences in SCFA. Moreover, the butyrate PGS was found to predict the abundance of the PWY-5022 pathway in stool samples, which was deemed to act as a proxy for butyrate production [436]; however, this PGS has not been assessed with regard to the actual amount of butyrate produced in the colon or absorbed by the host, but was associated with the abundance of butyrate producing bacteria (*Eubacterium rectale* and *Roseburia intestinalis*) in faecal samples [436]. Nevertheless, further research including GWAS is needed to confirm whether these genetic variants are reproducibly related to butyrate and propionate production. Moreover, the genetic analyses were restricted to white British, individuals limiting the generalisability and power of these analyses. Considering this, these results need to be

replicated using larger, more diverse samples to understand how genetics may interplay with carbohydrate intake, SCFA production, and colorectal cancer risk.

7.6 Conclusions

In summary, in this analysis of participants from the UK Biobank, intakes of total carbohydrates, total sugars, non-free sugars, wholegrain foods, and fibre from wholegrains were inversely associated with colorectal cancer risk. Some evidence of heterogeneity was observed by genetically predicted butyrate production, where the intake of starch from wholegrains was inversely associated with colorectal cancer risk only in participants with high host genetically predicted butyrate synthesis. Although these results may support the possible importance of SCFA in these associations, further research in humans is needed to replicate these findings.

Chapter 8

Risk of cancer among regular and low meat-eaters, fish-eaters, and vegetarians: a prospective analysis

Chapter 8 Summary

Some evidence suggests that being vegetarian may be associated with a lower risk of cancer overall. However, for specific cancer sites, the evidence is limited, and potential mechanisms are unclear. This chapter investigates the prospective associations of vegetarian and non-vegetarian diets with risks of all cancer, colorectal cancer, postmenopausal breast cancer, and prostate cancer, as well as explores the role of potential mediators among 472,377 UK Biobank participants. After an average follow-up of 11.4 years, 54,961 incident cancers were identified, including 5,882 colorectal, 7,537 postmenopausal breast, and 9,501 prostate cancers. Compared to regular meat-eaters, being a low meat-eater, fish-eater, or vegetarian were all associated with a lower risk of all cancer (HR:0.98, 95% CI:0.96-1.00; 0.90, 0.84-0.96; 0.86, 0.80-0.93, respectively). Being a low meat-eater was associated with a lower risk of colorectal cancer compared to regular meat-eaters (0.91, 0.86-0.96); however, there was heterogeneity in this association by sex (p=0.007), with an inverse association across diet groups in men, but not in women. Vegetarian postmenopausal women had a lower risk of breast cancer (0.82, 0.68-0.99), which was attenuated and non-significant after adjusting for BMI (0.87, 0.72-1.05); in mediation analyses, BMI was found to possibly mediate the observed association. In men, being a fish-eater or a vegetarian was associated with a lower risk of prostate cancer (0.80, 0.65-0.99 and 0.69, 0.54-0.89, respectively). The lower risk of postmenopausal breast cancer in vegetarian women may be explained by their lower BMI. It is not clear whether the other differences observed for all cancers and for prostate cancer, reflect any causal relationships or are due to other factors such as residual confounding or differences in cancer detection.

8.1 Introduction

Colorectal, breast, and prostate cancer collectively account for 39% of all new cancer diagnoses in the UK [463], and it has been estimated that nearly 40% of cancer cases may be preventable through modifiable factors [464, 465]. Although several dietary factors have been suggested to influence cancer risk, it remains unclear whether dietary patterns are related to the risk of developing cancer [386, 466].

It has been hypothesised that vegetarian diets, which exclude the consumption of all meat and fish, may be associated with a lower cancer risk. In addition to excluding red and processed meat, which are associated with an increased risk of colorectal cancer [386], vegetarians also generally consume higher amounts of plant foods such as fruits, vegetables, and wholegrains compared to meat-eaters [391, 467], which might also lower the risk of some site-specific cancers [386] and was suggested in analyses in Chapter 7 to be associated with the lower risk of colorectal cancer. Dietary patterns, such as vegetarianism, may also offer insight into dietary components associated with cancer risk as foods and nutrients are consumed together, which may complicate identifying true associations between foods.

8.1.1 Previous studies

Evidence from two large cohorts, which include a large proportion of vegetarians, the EPIC-Oxford and the AHS-2, has suggested that vegetarians may have a lower risk of developing cancer (all types combined) compared to meat-eaters [117, 118], but the evidence remains unclear for individual cancer sites [117, 119, 212, 297, 468]. Moreover, the risk of cancer in those who do not consume meat but do eat fish (fish-eaters or pescatarians) may differ from that of meat-eaters; some evidence has suggested that fish-eaters may have a lower overall risk of cancer [117], and a lower risk of colorectal cancer [117, 119] than meat-eaters, but no differences have been reported for breast [212] or prostate cancer risk [117, 297]. Despite the substantial number of vegetarians and fish-eaters in these cohorts (18,000-25,000 vegetarian participants and ~8,000 pescatarians), power to detect an association for specific cancer sites may be limited

due to the relatively small numbers of cancer cases (~5000 total cases) in each of these individual studies [117, 119, 212, 297, 468].

8.1.2 Mechanisms for cancer risk

Any difference in cancer risk between diet groups may be due to differences in physiological characteristics, including adiposity. In western populations, vegetarians and fish-eaters have been shown to have lower body mass indices compared to the BMI of meat-eaters [469-471], which is important for cancer risk because obesity is a known risk factor for several cancer sites [386]. Another hypothesized explanation for the lower risk of cancer observed among vegetarians and fish-eaters is the possible differences in hormone levels [472], such as IGF-I and testosterone, which may be related to their dietary intakes [298, 299, 472]. As described in Chapters 4 and 5, different food groups and macronutrients were associated with IGF-I concentrations, and vegetarians have been shown to consume less protein than meat-eaters and potentially less milk [391, 467], which may confer a lower risk of IGF-I related cancers. As such, hormone differences may be important as higher levels of IGF-I have been associated with higher risks of colorectal, breast, and prostate cancer [407] and higher levels of free testosterone have been associated with prostate cancer [14] and postmenopausal breast cancer [138].

8.2 Aim

To further understand these relationships, this chapter assesses the associations of diet groups with risks of all, colorectal, postmenopausal breast, and prostate cancer in the UK Biobank, which includes approximately 10,000 fish-eaters, 8,000 vegetarians, and nearly 55,000 total incident cancer cases. This chapter additionally assesses the roles of BMI, circulating IGF-I, and calculated free testosterone as potential mediators of the observed associations between diet groups and cancer risk.

8.3 Methods

8.3.1 Study design and participants

Details of recruitment, ethical approval, and methods of the UK Biobank are described in detail in Chapter 3.

8.3.2 Exclusions

Participants were excluded from this analysis if they withdrew consent over the study period (n=871), had a prevalent cancer diagnosis at recruitment reported in the cancer registry or in hospital admission data (excluding non-melanoma skin cancer ICD-10 code: C44; n=29,504), their genetic sex was different from their reported sex (n=321), or they did not contribute any follow-up time (n=2; **Figure 8.1**). Participants who responded as 'do not know' or 'prefer not to say' for all dietary questions regarding meat intake were also excluded from the analyses (n=282). This left a total of 472,337 participants, of whom 217,937 were males and 254,400 were females. For prostate cancer analyses women were excluded, and for postmenopausal breast cancer analyses, women who were premenopausal at recruitment and did not reach the age of 55 over the follow-up time (n=16,222), and men, were excluded.

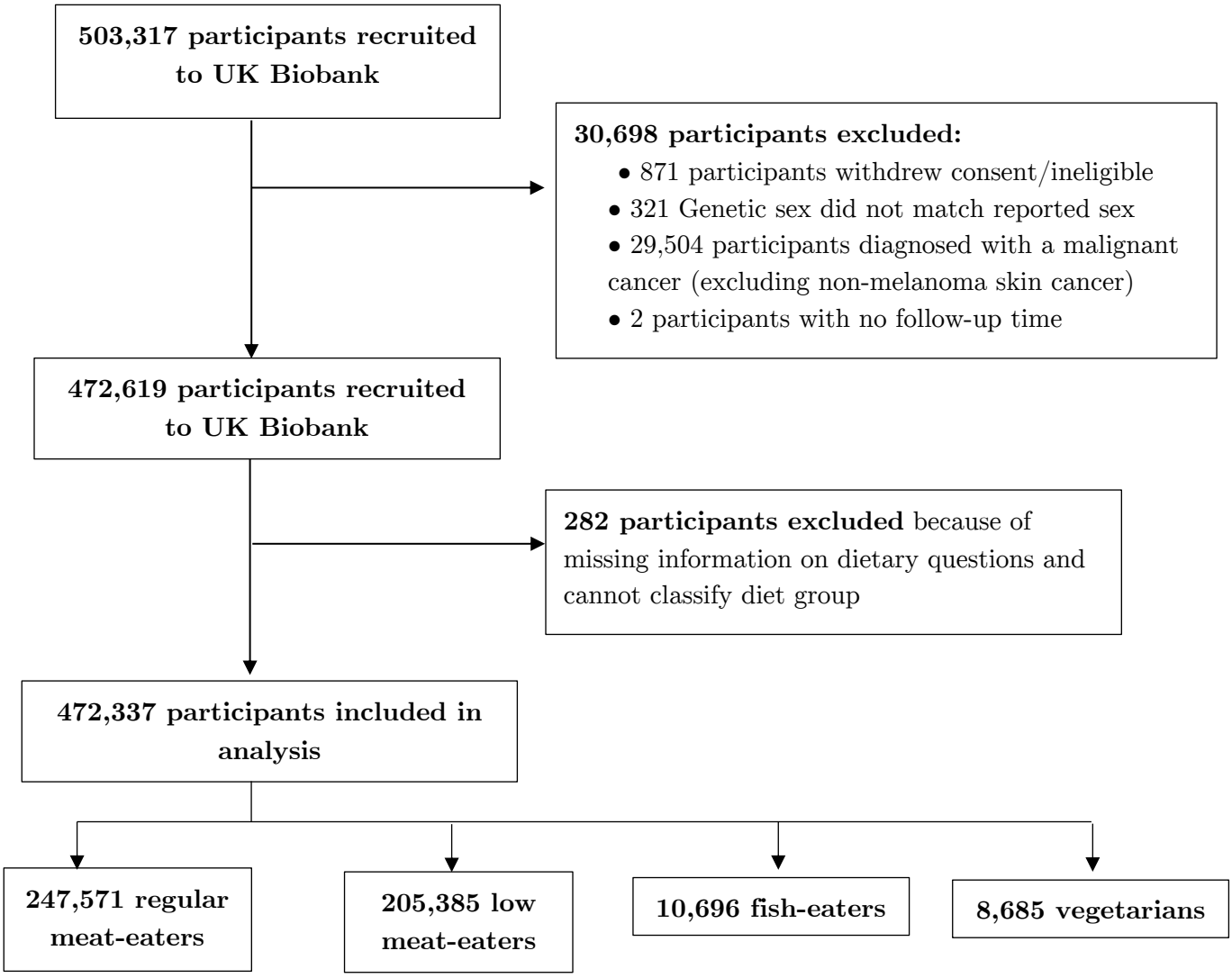


Figure 8.1 Flow diagram of the study exclusion criteria.

8.3.3 Diet group classification

Diet groups were categorised using the touchscreen questionnaire completed at recruitment, which asked participants about their frequency of consumption of processed meat, beef, lamb or mutton, pork, chicken, turkey or other poultry, and oily and non-oily fish. Participants chose a frequency of intake ranging from "Never" to "Once or more daily". From these responses, participants were categorised into four diet groups (regular meat-eaters; low meat-eaters; fish-eaters; and vegetarians). Regular meat-eaters were participants who said they consumed processed, red meat (beef, pork, lamb), or poultry >5 times a week. Low meat-eaters were participants who reported consuming processed, red meat, or poultry ≤5 times a week. Fish-eaters were participants who reported that they never consumed red meat, processed meat, or poultry but ate oily and/or non-oily fish. Vegetarians were defined as participants who reported that they never consumed any meat or fish. The vegetarian group also included vegans who reported not consuming any meat, fish, dairy, or eggs (n=446).

8.3.4 Covariates and biomarkers

The baseline touchscreen questionnaire also asked participants about sociodemographic, reproductive, and lifestyle factors, which are explained in detail in Chapter 3. Additionally, all participants had their blood drawn and anthropometric measurements, including height and weight, taken by a trained professional.

Non-fasting blood samples were provided by 99.7% of participants at recruitment and were shipped to the central processing laboratory at 4℃ prior to serum preparation, aliquoting, and cryopreservation in the central working archive. Biochemistry markers were measured including IGF-I and testosterone, as well as SHBG which was used to calculate an estimate of free testosterone [473]. Further description of the UK Biobank biomarker measurements can be found online [474] and in Chapter 3 and further details of the calculation of free testosterone can be found in **Ancillary Methods A.2**.

8.3.5 Follow-up and outcome ascertainment

Data on cancer diagnosis were ascertained using a combination of records from the NHS Digital (cancer registry) and Public Health England for participants from England and Wales, NHS Central Register for participants from Scotland [353] as well as the HES data for English participants and SMR for Scottish participants (please see details in the Additional File 1: Supplementary Methods). Using the World Health Organization's ICD-10 codes, participants were classified as having an event if they had an incident diagnosis of cancer recorded as: all cancer (C00-97 excluding non-melanoma skin cancer: C44), colorectal cancer (C18-C20), breast cancer (C50), or prostate cancer (C61), or if no prior incident diagnosis was reported their primary underlying cause of death was the respective cancer. Participants contributed follow-up time from the date of recruitment until the date of the first cancer registration or cancer first recorded on death certificate, date of death, or last day of follow-up available from HES and SMR data (28th February 2021 for England & Scotland). Cancer registry data were available until 31st July 2019 for England and Wales, and 31st October 2015 for Scotland, after this time only HES and SMR data were used for the follow-up of participants. For Welsh participants, hospital episode data did not extend past the cancer registry censoring date and therefore were not used. For breast cancer, analyses were restricted to postmenopausal breast cancer and women contributed follow-up time beginning when they turned 55 years of age (for women who were categorised as premenopausal or unknown menopause status at recruitment), or their date at recruitment if they were categorised as being postmenopausal from questions asked at baseline. The age of 55 was used as ~98% of females would have undergone menopause by this age [145].

8.3.6 Statistical analyses

Baseline characteristics of UK Biobank participants were summarised across diet groups for all participants, and separately for men and women.

Cox proportional hazards regressions were used, with age as the underlying time variable, to estimate HRs and 95% CIs. Minimally adjusted models were stratified by sex (for all cancer and colorectal cancer analyses only) and age at recruitment (<45, 45-49, 50-54, 55-59, 60-64, ≥65 years) and adjusted for region at recruitment (North-West England, North-Eastern England, Yorkshire & the Humber, West Midlands, East Midlands, South-East England, South-West England, London, Wales, and Scotland).

Multivariable-adjusted Cox regression models for all analyses were further adjusted for height (eight sex-specific categories increasing by 5 cm, and unknown/missing (0.51%)), physical activity (low: 0-9.99, medium: 10-49.99, high: ≥50 excess MET-hours /week, and unknown/missing (4.04%)), Townsend deprivation index (quintiles from most deprived to least deprived, and unknown/missing (0.13%)), education (completion of national exam at 16, completion of national exam at 17-18, college or university degree, or other/unknown/missing (18.7%)), employment status (employed, retired, not in paid employment, or unknown (1.15%)), smoking status (never, former, light smoker: ≤15 cigarettes/day, medium smoker: 16-29 cigarettes/day, heavy smoker: ≥30 cigarettes/day, or missing/unknown (0.65%)), alcohol consumption (none drinkers, <1, 1-9.99, 10-19.99, ≥20 grams/day, or unknown/missing (0.73%)), ethnicity (White, Mixed race or other, Asian or British Asian, and Black or Black British, or missing/unknown (0.56%)), and diabetes status (no, yes, or unknown (0.53%)).

For colorectal cancer and for all cancer sites, multivariable models were further adjusted for female specific covariates: MHT use (no, former, current, or unknown (0.58%)), and menopausal status at recruitment (premenopausal, postmenopausal, or unknown (9.0%)). Moreover, for colorectal cancer, multivariable models were adjusted for NSAID use (no reported use, irregular use, regular use of aspirin/ibuprofen). For prostate cancer, models were additionally adjusted for marital status (not living with a partner, living with a partner) [250]. For postmenopausal breast cancer, models were additionally adjusted for MHT use (same as above), age at menarche (≤12 years, 13 years old, ≥14 years, or unknown (22.5%)), parity and age at first birth

(nulliparous, 1-2 children <25 years old, 3+ children <25 years old, 1-2 children 25-29.9 years old, 3+ children 25-29.9 years old, 1-2 children 30+ years old, 3+ children 30+ years old, or missing (0.3%)). Further information on the classification of covariates can be found in the **Ancillary Methods A.1**. In all models, the proportional hazards assumption was evaluated using Schoenfeld residuals, and no violations were observed.

BMI was considered a potential confounder as well as a mediator. When BMI was considered as a potential confounder, BMI measured at recruitment was added to multivariable models (multivariable adjusted + BMI; <20, 20-22.49, 22.5-24.9, 25.0-27.49, 27.5-29.9, 30-32.49, 32.5-34.9, ≥35 kg/m^2, or unknown/missing (0.57%)). Models assessing BMI as a mediator are explained below in the mediation analyses section.

To determine whether there was heterogeneity in the associations of diet groups with cancer risk, and to assess the influence of confounder adjustments [149, 475], χ^2 statistics and p-values for including the diet group in the model were estimated using LRTs comparing a model without the diet groups variable to the model with the diet groups variable.

Absolute rate difference between diet groups and cancer incidence

The absolute rate differences were calculated for all cancer incidence and colorectal, postmenopausal breast, and prostate cancer incidences across the diet groups from multivariable adjusted HRs. Predicted incidence and absolute rate difference were calculated per 10,000 individuals over 10 years and were estimated using floating absolute risks [476]. Floating absolute risks assign 95% CIs to all groups including the reference group allowing for an estimate of uncertainty in the effect size in the reference group. For regular meat-eaters, predicted 10-year cancer-free survival was calculated using the equation Survival (Sr) = (1-observed incidence in regular meat-eaters)^{10} [477]. This was then used to determine the predicted incidence over time by using (1-Sr) x 10,000, which equates to the incidence per 10,000 individuals over a 10-year time frame assuming a constant rate over time. For the other diet

groups, predicted incidence was calculated by $(1 - Sr^{HR})$ where HR is the multivariable-adjusted hazard ratio for each diet group for the cancer outcome and thus Sr^{HR} estimates the predicted 10-year cancer-free survival rate in each of the other diet groups. The absolute rate differences were then calculated as the difference between the predicted incidence per 10,000 individuals over 10 years between the other diet groups (low meat-eaters, fish-eaters, and vegetarians) and regular meat-eaters.

Subgroup and sensitivity analyses

For all analyses, heterogeneity was assessed by BMI (~median: $<$27.5 and $\geq$27.5 kg/m^2) and smoking status (ever and never) by using a LRT comparing the main model to a model including an interaction term between diet groups and the subgroup variable (BMI and smoking status). For colorectal cancer heterogeneity, by sex was also assessed. For all cancer sites combined, additional heterogeneity was explored by smoking status and censoring participants at baseline who were diagnosed with lung cancer.

In sensitivity analyses, participants and cases were excluded who had less than 2 years of follow-up and all participants with missing data on covariates. Associations were also examined separately in white participants because a large proportion of the vegetarians in this cohort are of South Asian ethnicity (~17.5%). Furthermore, fruit and vegetable intake was added to multivariable adjusted models ($<$3 servings/day, 3-3.99 servings/day, 4-5.99 servings/day, $\geq$6 servings/day, unknown) to control for this component of dietary intake as a proxy for a healthy diet. For prostate cancer analyses, PSA testing was included in the multivariable adjusted model (no PSA testing, had PSA test, or unknown) reported at baseline in all men and during follow-up from general practice records in a subsample (n=99,412 males; records available for participants until 31st of May 2016 for England, 31st of March 2017 for Scotland, and 31st of August 2017 for Wales).

8.3.7 Mediation analyses

If a significant association was observed between a diet group and a cancer outcome in the main analyses, potential mediators shown to be different across diet groups [470, 472] and were previously related to the cancer site of interest (BMI, IGF-I, and free testosterone) [14, 138] were further explored. To determine whether differences in mediators were observed by diet group, multivariable linear regression was used to compare the selected biomarker measurements (IGF-I and free testosterone [473]) and BMI across dietary groups, adjusting for potential confounders. Multivariable linear regression models comparing biomarkers or BMI with diet groups were adjusted for age, sex, region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, diabetes status, MHT use, menopausal status, and BMI (except when BMI was the outcome of interest). Mediation was not explored if there was no significant difference in cancer risk between each diet group and regular meat-eaters, or if the biomarker concentrations were not significantly different between diet groups. Potential mediation for all cancer risk for vegetarians versus regular meat-eaters, fish-eaters versus regular meat-eaters, and low meat-eaters versus regular meat-eaters was explored via BMI [386, 470].

For colorectal cancer risk, potential mediation via BMI was explored for low meat-eaters versus regular meat-eaters. For postmenopausal breast cancer risk, mediation was explored for vegetarians versus regular meat-eaters via BMI, IGF-I concentrations, and free testosterone (estimated using a formula based on the law of mass action from measured total testosterone, SHBG, and albumin concentrations [473]; **Ancillary Methods A.2**), among postmenopausal women. If a woman was missing a value for testosterone due to a very low serum concentration reported, the concentration for these participants was set to three-quarters of the minimum reportable value (0.35 nmol/L; n=23,494) and free testosterone was calculated from this. For prostate cancer risk, potential mediation for vegetarians versus regular meat-eaters was explored via IGF-I and free testosterone, as these hormones have been associated with higher prostate cancer risk [14]. Mediation via free testosterone was also explored for the fish-eater versus

regular meat-eater analysis but not IGF-I as no differences in IGF-I concentrations were observed between fish-eaters and regular meat-eaters. BMI was not explored as a potential mediator for prostate cancer as BMI is heterogeneously associated with prostate cancer risk [268]. Specifically, BMI may be a risk factor for prostate cancer death and aggressive disease but has been inversely associated with total prostate cancer and non-aggressive disease risk [262, 268] and data on stage and grade are not available in the UK Biobank. Biomarker mediation was not explored for the all cancer - diet group associations as these biomarkers have not been associated with all cancer risk.

To assess for mediation, the inverse odds ratio weighting (IORW) method was used [478, 479]. This method is able to decompose associations between exposure and outcome that is (1) mediated by the potential mediator (natural indirect effect [NIE]) and (2) not mediated by the mediator of interest (natural direct effect [NDE]). The term "effect" is used here in concordance with the causal mediation literature but should not be interpreted as implying causality. To determine the proportion of the association between diet groups and cancer outcome mediated by the mediator of interest (e.g., BMI) the log of the indirect effect HR was divided by the log of the total effect HR. The IORW method was used as it is generally more flexible in estimating mediation as it does not require specification of regression models for each mediator of interest (like other traditional methods used for mediation [480, 481]), as well, it does not assume there is no exposure-mediator interactions, and can easily be applied to survival analyses, specifically Cox regression. The IORW evade these issues by condensing the relationship between exposure and mediators of interest into a single odds ratio (OR), thus an OR for the exposure conditional on all mediators is equivalent to the OR for the mediators, conditional on exposure. This OR can then be applied as a weight to a regression of the outcome on the exposure, thus estimating the natural direct effect (NDE) estimate as the pathway through which the mediators are deactivated. As with all mediation analyses, key assumptions are made using the IORW method, namely 1) there is no unmeasured exposure – mediator confounding, 2) there is no unmeasured mediator - outcome confounding, 3) there is no unmeasured exposure – outcome

confounding, 4) there are no mediator-outcome confounders that is itself affected by the exposure [480].

The weight of mediators was estimated from a logistic regression model where the binary diet group (i.e. vegetarians – regular meat-eaters; fish-eaters – regular meat-eaters; or low meat-eater – regular meat-eaters) was the response variable and mediators, and potential confounders were the predictors. In each respective mediation model, the mediators of interest were modelled as continuous variables. The estimated coefficients of this model were then used to predict the OR of the binary diet group for each participant and constructed weights for binary diet groups as the inverse of these predicted ORs. Here the reference category (regular meat-eaters) was assigned the weight of 1 and the opposing diet group (vegetarians, fish-eaters, or low-meat eaters) were assigned the value of the inverse odds of the logistic models modelled with the mediators and adjusted confounders. Cox proportional hazards regressions were then conducted using age as the underlying time variable with all cancer sites or specific cancer site, adjusting for confounders, weighting each observation by the weights derived from the logistic regression. From this Cox regression, the direct effect was obtained as the HR for the binary diet group (e.g., vegetarians – regular meat-eaters) of interest in this model and the indirect effect was calculated by taking the log HR of the total effect minus the log HR of the direct effect and then was exponentiated to obtain the indirect HR. For the total, direct, and indirect effects, the resulting HRs were estimated and bootstrapped standard errors (300 replications) to determine 95% CIs. The IORW method was conducted separately for each outcome (all cancer, prostate cancer, postmenopausal breast cancer, or colorectal cancer) and binary diet group of interest and excluding participants not included in the binary diet group of interest, and women for the prostate cancer analyses, and men for the postmenopausal breast cancer analyses. Participants with missing values for the potential mediator (i.e., BMI, IGF-I, free testosterone) were excluded from the respective mediation analyses.

All analyses were conducted using Stata version 17.0 (Stata Corp LP, College Station, TX). P-values were two-sided with $p < 0.05$ being considered statistically significant.

8.4 Results

8.4.1 Participant characteristics

Of the participants included in the analysis, 247,571 (52.4%) were classified as regular meat-eaters, 205,385 (43.5%) as low meat-eaters, 10,696 (2.3%) as fish-eaters and 8,685 (1.8%) as vegetarians. After an average of 11.4 years of follow-up, 54,961 incident cases of any type of cancer were diagnosed; 5,882 participants were diagnosed with colorectal cancer, 7,537 women were diagnosed with postmenopausal breast cancer, and 9,501 men were diagnosed with prostate cancer.

Table 8.1 presents participants' baseline characteristics across diet groups. Vegetarians and fish-eaters had a lower BMI, were younger, more likely to be never smokers, have a university/college degree, report consuming less alcohol at recruitment, and were more likely to be Asian compared to regular meat-eaters. Vegetarian men were also less likely to have had a PSA test compared to meat-eaters (**Table 8.1**). **Table 8.2** presents the baseline characteristics across diet groups stratified by sex. Both men and women fish-eaters and vegetarians had lower BMIs and were younger at recruitment compared to regular meat-eaters.

Table 8.1 Baseline characteristics across diet groups in UK Biobank.

	Diet groups			
	Regular meat-eaters	**Low meat-eaters**	**Fish-eaters**	**Vegetarians**
Number of participants	247,571	205,385	10,696	8,685
Age at recruitment- years, mean (SD)	56.0 (8.2)	56.9 (8.0)	54.0 (8.0)	53.0 (7.9)
Sex - female	114849 (46.4%)	126165 (61.4%)	7664 (71.7%)	5722 (65.9%)
BMI - kg/m^2, mean (SD)	27.9 (4.9)	27.0 (4.7)	25.3 (4.3)	25.7 (4.7)
Male height - cm, mean (SD)	175.7 (6.8)	175.4 (6.9)	176.4 (6.9)	175.5 (7.2)
Female height - cm, mean (SD)	162.4 (6.3)	162.5 (6.3)	163.5 (6.4)	162.1 (6.8)
Physical activity				
Low	72,811 (29.4%)	59,752 (29.1%)	2,430 (22.7%)	2,371 (27.3%)
Moderate	116,591 (47.1%)	98,692 (48.1%)	5,710 (53.4%)	4,346 (50.0%)
High	48,012 (19.4%)	38,704 (18.8%)	2,273 (21.3%)	1,690 (19.5%)
Townsend deprivation index				
Q1- Most affluent	51,117 (20.6%)	40,433 (19.7%)	1,777 (16.6%)	1,258 (14.5%)
Q5 - Most deprived	48,227 (19.5%)	41,627 (20.3%)	2,385 (22.3%)	2,216 (25.5%)
Ethnicity				
White	233,959 (94.5%)	193,033 (94.0%)	9,922 (92.8%)	6,903 (79.5%)
Mixed other	3,576 (1.4%)	3,294 (1.6%)	172 (1.6%)	152 (1.8%)
Asian or British Asian	4,114 (1.7%)	5,054 (2.5%)	369 (3.4%)	1524 (17.5%)
Black or Black British	4,218 (1.7%)	3,295 (1.6%)	167 (1.6%)	48 (0.6%)
Education				
National exam at 16 years	41,764 (16.9%)	34,271 (16.7%)	1,180 (11.0%)	1,099 (12.7%)
National exam at 17-18 years	13,750 (5.6%)	10,805 (5.3%)	578 (5.4%)	551 (6.3%)
Degree or college	146,214 (59.1%)	119,791 (58.3%)	8,015 (74.9%)	6,109 (70.3%)
Employment				
In paid employment	146,078 (59.0%)	115,579 (56.3%)	7,338 (68.6%)	6,065 (69.8%)
Retired	77,483 (31.3%)	70,640 (34.4%)	2,341 (21.9%)	1,582 (18.2%)
Not in paid employment	21,068 (8.5%)	17,028 (8.3%)	877 (8.2%)	921 (10.6%)
Living with a partner - Yes	187,545 (75.8%)	141,711 (69.0%)	6,930 (64.8%)	5,771 (66.4%)
Smoking status				
Never	132,294 (53.4%)	114,385 (55.7%)	6,075 (56.8%)	5,561 (64.0%)
Previous	85,319 (34.5%)	69,642 (33.9%)	3,800 (35.5%)	2,480 (28.6%)
Light smoker <15 cig/day	7,594 (3.1%)	6,299 (3.1%)	290 (2.7%)	210 (2.4%)
Medium smoker 15-29 cig/day	10,101 (4.1%)	6,644 (3.2%)	157 (1.5%)	139 (1.6%)
Heavy smoker 30+ cig/day	10,418 (4.2%)	7,432 (3.6%)	333 (3.1%)	252 (2.9%)
Alcohol intake g/day, mean (SD)	19.9 (21.3)	14.9 (16.6)	13.6 (14.5)	13.0 (16.1)
Diabetic - Yes	15,603 (6.3%)	10,748 (5.2%)	290 (2.7%)	465 (5.4%)
Prostate specific antigen test reported at baseline or in follow-up – Yes, Male only	51,555 (38.8%)	33,394 (42.2%)	1,125 (37.1%)	929 (31.4%)

Table 8.1 continued

	Diet groups			
	Regular meat-eaters	**Low meat-eaters**	**Fish-eaters**	**Vegetarians**
Female specific covariates				
Age at menarche – years, mean (SD)	12.6 (2.8)	12.5 (2.9)	12.5 (2.9)	12.4 (3.3)
Menopausal status				
Premenopausal	24,939 (21.7%)	23,360 (18.5%)	2,232 (29.1%)	1,843 (32.2%)
Postmenopausal	78,626 (68.5%)	92,413 (73.2%)	4,717 (61.5%)	3,342 (58.4%)
MHT use				
Never	70,830 (61.7%)	76,747 (60.8%)	5,468 (71.3%)	4,386 (76.7%)
Former	34,149 (29.7%)	38,960 (30.9%)	1,590 (20.7%)	966 (16.9%)
Current	8,993 (7.8%)	9,988 (7.9%)	584 (7.6%)	320 (5.6%)
Parity				
Nulliparous	17,671 (15.4%)	25,569 (20.3%)	2,330 (30.4%)	1,736 (30.3%)
1-2 children	67,306 (58.6%)	70,915 (56.2%)	3,910 (51.0%)	2,770 (48.4%)
3+ children	29,343 (25.5%)	29,526 (23.4%)	1,415 (18.5%)	1,203 (21.0%)
Age at first birth – years, mean (SD)	25.4 (4.6)	25.2 (4.6)	26.4 (5.1)	26.0 (4.9)

Values are N (%) unless otherwise indicated.

Percentages include missing values and therefore may not add up to 100%.

Abbreviations: BMI, body mass index; cig, cigarette; g/day, grams per day; Q, quintile; MHT, menopausal hormone therapy, SD, standard deviation.

Table 8.2 Baseline characteristics of UK Biobank participants across diet groups separated by sex.

	Male				Female			
	Regular meat-eater	**Low meat-eaters**	**Fish-eater**	**Vegetarian**	**Regular meat-eater**	**Low meat-eaters**	**Fish-eater**	**Vegetarian**
Number of participants	132,722	79,220	3,032	2,963	114,849	126,165	7,664	5,722
Age – years, mean (SD)	56.2 (8.2)	57.3 (8.1)	54.3 (8.1)	53.1 (8.0)	55.9 (8.1)	56.6 (7.9)	53.9 (8.0)	53.0 (7.9)
BMI - kg/m^2, mean (SD)	28.2 (4.3)	27.5 (4.1)	25.8 (3.6)	25.9 (3.9)	27.7 (5.4)	26.7 (5.0)	25.1 (4.5)	25.6 (5.0)
Height - cm, mean (SD)	175.7 (6.8)	175.4 (6.9)	176.4 (6.9)	175.5 (7.2)	162.4 (6.3)	162.5 (6.3)	163.5 (6.4)	162.1 (6.8)
Physical activity								
Low	36,801 (27.7%)	21,972 (27.7%)	628 (20.7%)	777 (26.2%)	36,010 (31.4%)	37,780 (29.9%)	1,802 (23.5%)	1,594 (27.9%)
Moderate	62,156 (46.8%)	37,698 (47.6%)	1,636 (54.0%)	1,505 (50.8%)	54,435 (47.4%)	60,994 (48.3%)	4,074 (53.2%)	2,841 (49.7%)
High	28,947 (21.8%)	16,599 (21.0%)	704 (23.2%)	610 (20.6%)	19,065 (16.6%)	22,105 (17.5%)	1,569 (20.5%)	1,080 (18.9%)
Townsend deprivation index								
Q1- Most affluent	27,202 (20.5%)	15,766 (19.9%)	458 (15.1%)	430 (14.5%)	23,915 (20.8%)	24,667 (19.6%)	1,319 (17.2%)	828 (14.5%)
Q5 - Most deprived	26,819 (20.2%)	16,416 (20.7%)	740 (24.4%)	827 (27.9%)	21,408 (18.6%)	25,211 (20.0%)	1,645 (21.5%)	1,389 (24.3%)
Ethnicity								
White	125,463 (94.5%)	74,035 (93.5%)	2,792 (92.1%)	2,321 (78.3%)	108,496 (94.5%)	118,998 (94.3%)	7,130 (93.0%)	4,582 (80.1%)
Mixed or other	1,776 (1.3%)	1,112 (1.4%)	45 (1.5%)	45 (1.5%)	1,800 (1.6%)	2,182 (1.7%)	127 (1.7%)	107 (1.9%)
Asian or British Asian	2,422 (1.8%)	2,615 (3.3%)	125 (4.1%)	558 (18.8%)	1,692 (1.5%)	2,439 (1.9%)	244 (3.2%)	966 (16.9%)
Black or Black British	2,092 (1.6%)	1,101 (1.4%)	46 (1.5%)	12 (0.4%)	2,126 (1.9%)	2,194 (1.7%)	121 (1.6%)	36 (0.6%)
Education								
National exam at 16 years	18,401 (13.9%)	10,397 (13.1%)	273 (9.0%)	323 (10.9%)	23,363 (20.3%)	23,874 (18.9%)	907 (11.8%)	776 (13.6%)
National exam at 17-18	6,811 (5.1%)	3,807 (4.8%)	145 (4.8%)	188 (6.3%)	6,939 (6.0%)	6,998 (5.5%)	433 (5.6%)	363 (6.3%)
Degree or college	83,092 (62.6%)	48,766 (61.6%)	2,340 (77.2%)	2,144 (72.4%)	63,122 (55.0%)	71,025 (56.3%)	5,675 (74.0%)	3,965 (69.3%)
Employment								
In paid employment	83,026 (62.6%)	46,243 (58.4%)	2,145 (70.7%)	2,146 (72.4%)	63,052 (54.9%)	69,336 (55.0%)	5,193 (67.8%)	3,919 (68.5%)
Retired	38,471 (29.0%)	26,228 (33.1%)	633 (20.9%)	514 (17.3%)	39,012 (34.0%)	44,412 (35.2%)	1,708 (22.3%)	1,068 (18.7%)
Not in paid employment	9,691 (7.3%)	5,887 (7.4%)	206 (6.8%)	274 (9.2%)	11,377 (9.9%)	11,141 (8.8%)	671 (8.8%)	647 (11.3%)

Table 8.2 continued

	Male				Female			
	Regular meat-eater	**Low meat-eaters**	**Fish-eater**	**Vegetarian**	**Regular meat-eater**	**Low meat-eaters**	**Fish-eater**	**Vegetarian**
Living with a partner - Yes	102,443 (77.2%)	59,313 (74.9%)	2,163 (71.3%)	2,104 (71.0%)	67,306 (58.6%)	70,915 (56.2%)	3,910 (51.0%)	2,770 (48.4%)
Smoking status								
Never	63,688 (48.0%)	39,608 (50.0%)	1,658 (54.7%)	1,727 (58.3%)	68,606 (59.7%)	74,777 (59.3%)	4,417 (57.6%)	3,834 (67.0%)
Previous	50,320 (37.9%)	30,201 (38.1%)	1,096 (36.1%)	951 (32.1%)	34,999 (30.5%)	39,441 (31.3%)	2,704 (35.3%)	1,529 (26.7%)
Light smoker <15 cigarettes/day	3,887 (2.9%)	2,048 (2.6%)	72 (2.4%)	79 (2.7%)	3,707 (3.2%)	4,251 (3.4%)	218 (2.8%)	131 (2.3%)
med smoker 15-29 cigarettes/day	6,093 (4.6%)	2,715 (3.4%)	51 (1.7%)	53 (1.8%)	4,008 (3.5%)	3,929 (3.1%)	106 (1.4%)	86 (1.5%)
heavy smoker 30+ cigarettes/day	7,738 (5.8%)	4,208 (5.3%)	141 (4.7%)	134 (4.5%)	2,680 (2.3%)	3,224 (2.6%)	192 (2.5%)	118 (2.1%)
Alcohol intake g/day, mean (SD)	26.4 (24.7)	21.4 (21.1)	20.1 (18.9)	18.5 (21.1)	11.7 (11.9)	10.4 (10.6)	11.0 (11.3)	9.9 (11.2)
Diabetes - Yes	10,377 (7.8%)	5,943 (7.5%)	121 (4.0%)	207 (7.0%)	5,226 (4.6%)	4,805 (3.8%)	169 (2.2%)	258 (4.5%)
PSA test at baseline or during follow-up - Yes	51,555 (38.8%)	33,394 (42.2%)	1,125 (37.1%)	929 (31.4%)	-	-	-	-
Age-adjusted proportion of PSA test at baseline or during follow-up - Yes	39.5%	40.5%	41.6%	37.2%	-	-	-	-
Insulin-like growth factor-I (nmol/L), mean (SD)[1]	21.9 (5.5)	21.9 (5.5)	22.0 (5.7)	21.2 (5.5)	21.1 (5.8)	21.0 (5.7)	20.6 (5.8)	19.7 (5.8)
Free testosterone (pmol/L), mean (SD)[1]	220.0 (65.9)	219.7 (63.5)	218.7 (61.5)	217.4 (62.3)	15.3 (11.4)	14.9 (10.9)	14.0 (8.6)	14.4 (9.4)

Values are N (%) unless otherwise indicated.

Percentages include missing values and therefore may not add up to 100%.

[1] Biomarker levels are adjusted for age at recruitment, region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, diabetes status, menopausal hormone therapy use and menopausal status, body mass index and season of blood collection.

Abbreviations: BMI, body mass index; g/day, grams per day; Q, quintile; PSA, prostate specific antigen; SD, standard deviation.

8.4.2 Diet group and cancer risk associations

The minimally adjusted models and sequential adjustments for the associations between diet groups and cancer risks are presented in **Table 8.3**, and **Figure 8.2** shows the multivariable-adjusted models. In multivariable-adjusted models there was evidence of heterogeneity across diet groups for risk of all cancer (p-value for heterogeneity=<0.001), colorectal cancer (p-value for heterogeneity=0.001), and prostate cancer (p-value for heterogeneity=0.003), but not postmenopausal breast cancer (p-value for heterogeneity=0.084; **Figure 8.2A**). In the multivariable-adjusted models (not including BMI), a vegetarian diet was associated with a lower risk of all cancer (HR: 0.86, 95% CI: 0.80-0.93), postmenopausal breast cancer (0.82, 0.68-0.99) and prostate cancer (0.69, 0.54-0.89; **Figure 8.2A**) compared to regular meat-eaters. Furthermore, compared to regular meat-eaters, fish-eaters had a lower risk of all cancers (0.90, 0.84-0.96) and prostate cancer (0.80, 0.65-0.99), and low meat-eaters had a lower risk of colorectal cancer (0.91, 0.86-0.96; **Figure 8.2A**). When including BMI as a potential confounder, associations were slightly attenuated apart from prostate cancer, which did not change (**Figure 8.2B**). For postmenopausal breast cancer, after adjustment for BMI, the risk for vegetarians compared to regular meat-eaters was no longer statistically significant (0.87, 0.72-1.05; **Figure 8.2B**).

8.4.3 Absolute rate difference

Table 8.4 presents the absolute difference in cancer cases by diet groups. Being vegetarian or fish-eater was associated with 135 (95% CI: 66-199) and 101 (95% CI: 42-156) fewer cases of any cancer diagnosis per 10,000 population over 10 years compared to regular meat-eaters, respectively. For colorectal cancer, being a low meat-eater was associated with 11 fewer cases per 10,000 population over 10 years compared to regular meat-eaters (95% CI: 5-17). For postmenopausal breast cancer, being a vegetarian was associated with 59 fewer cases of postmenopausal breast cancer compared to regular meat-eaters (95% CI: 3-106). Vegetarians and fish-eaters also had 117 (95% CI: 38-168) and 75 (5.2-132) fewer cases of prostate cancer per 10,000 population over 10 years compared to regular meat-eaters.

Table 8.3 Hazard ratios and 95% confidence intervals for sequential adjustment between association of diet groups and risk of all cancer, colorectal cancer, breast cancer, and prostate cancer.

All cancer	Regular meat-eaters	Low meat-eaters	Fish-eaters	Vegetarians	χ^2	P-value for heterogeneity
Minimally adjusted model	1 (ref)	0.97 (0.95 - 0.99)	0.87 (0.82 - 0.93)	0.80 (0.74 - 0.86)	56.00	<0.001
Multivariable-adjusted model	1 (ref)	0.98 (0.96 - 1.00)	0.90 (0.84 - 0.96)	0.86 (0.80 - 0.93)	26.88	<0.001
Multivariable-adjusted model + BMI	1 (ref)	0.99 (0.97 - 1.01)	0.92 (0.86 - 0.99)	0.88 (0.82 - 0.96)	15.53	0.001
Colorectal cancer						
Minimally adjusted model	1 (ref)	0.89 (0.84 - 0.94)	0.81 (0.66 - 1.00)	0.72 (0.56 - 0.93)	26.12	<0.001
Multivariable-adjusted model	1 (ref)	0.91 (0.86 - 0.96)	0.84 (0.68 - 1.03)	0.78 (0.61 - 1.01)	16.80	0.001
Multivariable-adjusted model + BMI	1 (ref)	0.92 (0.87 - 0.97)	0.86 (0.70 - 1.06)	0.81 (0.63 - 1.04)	12.69	0.005
Postmenopausal breast cancer						
Minimally adjusted model	1 (ref)	0.97 (0.92 - 1.01)	0.92 (0.80 - 1.07)	0.80 (0.67 - 0.97)	7.43	0.059
Multivariable-adjusted model	1 (ref)	0.96 (0.92 - 1.01)	0.92 (0.79 - 1.06)	0.82 (0.68 - 0.99)	6.65	0.084
Multivariable-adjusted model + BMI	1 (ref)	0.99 (0.94 - 1.03)	0.97 (0.84 - 1.12)	0.87 (0.72 - 1.05)	2.46	0.483
Prostate cancer						
Minimally adjusted model	1 (ref)	1.00 (0.96 - 1.04)	0.80 (0.65 - 0.98)	0.61 (0.48 - 0.79)	22.54	<0.001
Multivariable-adjusted model	1 (ref)	1.00 (0.96 - 1.05)	0.80 (0.65 - 0.99)	0.69 (0.54 - 0.89)	14.08	0.003
Multivariable-adjusted model + BMI	1 (ref)	1.00 (0.96 - 1.04)	0.80 (0.65 - 0.98)	0.69 (0.54 - 0.89)	14.28	0.003
Multivariable-adjusted + PSA test	1 (ref)	1.00 (0.96 - 1.04)	0.80 (0.65 - 0.98)	0.70 (0.55 - 0.90)	13.47	0.004

Regular meat-eaters: consumed red or processed meat or poultry >5 times a week. Low meat-eaters: consumed red and processed meat or poultry ≤5 times per week. Fish-eaters: do not consume red, processed meat, or poultry but consumed fish. Vegetarians (including vegans): do not consume any meat or fish.

Minimally adjusted model used age as the underlying time variable and are stratified by sex (for all cancer and colorectal cancer), age groups, and controlled for region of recruitment. Models are restricted to only men for prostate cancer and only women for postmenopausal breast cancer.

Multivariable-adjusted models for all cancer outcomes: Minimally adjusted model + height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, and diabetes status. Full details for each covariate are provided in the statistical analysis section in the main text.

For all cancer and colorectal cancer multivariable-adjusted model additionally adjust for menopausal hormone therapy (MHT) use and menopausal status. For colorectal cancer, multi-variable-adjusted was also adjusted for NSAID use.

Postmenopausal breast cancer models additionally adjust for MHT use, age at menarche, and parity.

Prostate cancer multivariable-adjusted models additionally adjust for marital status.

Multivariable-adjusted model + BMI adds body mass index (<20, 20-22.49,22.5-24.9, 25.0-27.49, 27.5-29.9, 30-32.49, 32.5-34.9, ≥35 kg/m^2 or unknown) to multivariable models.

Multivariable-adjusted model + PSA test adds prostate specific antigen testing reported at baseline and from follow-up general practice records available in a subsample of men.

χ^2 (df=3) and p-values from likelihood ratio tests for model fit comparing a model without diet groups, to a model including diet groups.

Abbreviations: BMI:, body mass index; df, degrees of freedom; MHT, menopausal hormonal therapy; NSAID, non-steroid anti-inflammatory drugs; PSA, prostate specific antigen test; ref, reference group.

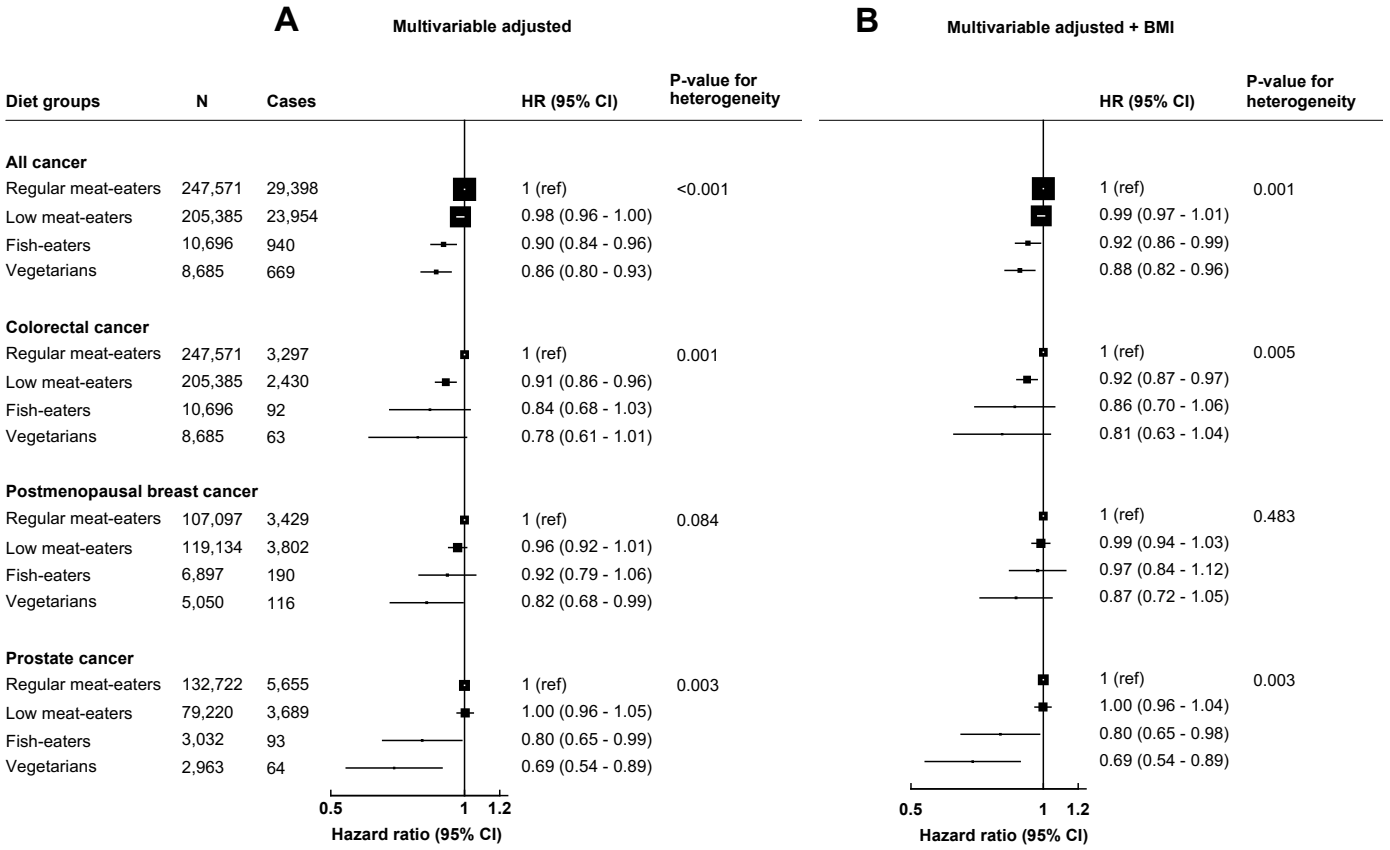


Figure 8.2 Multivariable adjusted hazard ratios (95% CI) for diet groups and risk of all cancer, prostate cancer, postmenopausal breast cancer and colorectal cancer not adjusting for BMI (A) and adjusting for BMI (B).

All models used age as the underlying time variable and are stratified by sex (for only all cancer and colorectal cancer), age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, and diabetes status. Full details for each covariate are provided in the statistical analysis section in the main text. For all cancer and colorectal cancer analyses, models were further adjusted for menopausal hormone therapy use and menopausal status and colorectal cancer models are adjusted for non-steroid anti-inflammatory drug use. Postmenopausal breast cancer models are further adjusted for menopausal hormone therapy use, age at menarche, and age at first birth/parity. Prostate cancer models are further adjusted for marital status.

Multivariable + BMI models further adjusts for BMI.

P-value for heterogeneity from likelihood ratio tests for model fit comparing a model without diet groups, to a model including diet group.

Abbreviations: BMI, body mass index; CI, confidence intervals; HR, hazard ratio; N, number of participants; ref, reference group.

Table 8.4 Absolute rate difference (per 10,000 individuals per 10 years) for all cancer incidence, colorectal, postmenopausal breast, and prostate cancer across diet groups.

Cancer and diet groups	Predicted incidence per 10,000 per 10 years	Absolute rate difference per 10,000 population per 10 years
All cancer		
Regular meat-eaters	1,033 (1,015 to 1,050)	1 (ref)
Low meat-eaters	1,012 (995 to 1,029)	-20.7 (-37.2 to -3.9)
Fish-eaters	931 (875 to 991)	-101.2 (-157.5 to -41.5)
Vegetarians	898 (834 to 967)	-134.5 (-198.7 to -65.7)
Colorectal cancer		
Regular meat-eaters	120.9 (114.8 to 127.4)	1 (ref)
Low meat-eaters	109.8 (104.1 to 115.8)	-11.1 (-16.8 to -5.2)
Fish-eaters	101.2 (82.2 to 124.5)	-19.7 (-38.7 to 3.6)
Vegetarians	95.0 (73.9 to 122.1)	-25.9 (-47.0 to 1.2)
Postmenopausal breast cancer		
Regular meat-eaters	337.1 (323.1 to 351.7)	1 (ref)
Low meat-eaters	325.5 (310.9 to 340.7)	-11.6 (-26.2 to 3.6)
Fish-eaters	309.0 (267.3 to 357.0)	-28.2 (-69.8 to 19.9)
Vegetarians	277.9 (230.8 to 334.5)	-59.2 (-106.3 to -2.6)
Prostate cancer		
Regular meat-eaters	311.3 (254.2 to 380.9)	1 (ref)
Low meat-eaters	387.5 (371.9 to 403.7)	1.4 (-14.2 to 17.6)
Fish-eaters	311.3 (254.2 to 380.9)	-74.8 (-131.9 to -5.2)
Vegetarians	269.6 (211.0 to 344.4)	-116.5 (-175.1 to -41.7)

Regular meat-eaters: red or processed meat or poultry >5 times a week. Low meat-eaters: red and processed meat or poultry ≤5 times per week. Fish-eaters: do not consume red, processed meat, or poultry but consumed fish. Vegetarians (including vegans): do not consume any meat or fish.

Incidence per 10,000 individuals per 10 years was calculated for regular meat-eaters by (1-Sr) ×10,000, where Sr= (1-observed incidence in regular meat-eaters)^{10} which represents the predicted 10-year survival rate in regular meat-eaters. Regular meat-eaters 95% CI were determined using floating absolute risks. For other diet groups, predicted incidence was determined by (1-Sr^{HR}) ×10,000, where the HR represents the hazard ratio or confidence interval of the hazard ratio for the diet group where Sr^{HR} represents the predicted 10-year survival (non-incidence) rate in the diet groups.

Absolute risk difference was calculated as the difference between the predicted incidence per 10,000 individuals over 10 years between regular meat-eaters as the reference group and the other diet groups.

8.4.4 Subgroup analyses

No evidence of heterogeneity was observed across BMI subgroups in the associations between diet groups and risk of all cancer, colorectal, postmenopausal breast, and prostate cancer (**Table 8.5 & Table 8.6**; Ancillary Tables 8.1 & 8.2). For smoking status, some evidence of heterogeneity was observed in the association between diet groups and all cancer risk (p_{het}=0.056); among ever smokers, the low meat-eaters, fish-eaters, and vegetarians had lower risks of all cancer sites than regular meat-eaters (0.97, 0.94-0.99; 0.86, 0.78-0.95; 0.79, 0.70-0.90, respectively), whereas these associations were non-significant for non-smokers (**Table 8.5**). However, when censoring participants who developed lung cancer during follow-up, the p-value for heterogeneity by smoking status became greater (p_{het}=0.22; **Table 8.5**), although the associations with diet group for all cancer sites was still only significant among smokers. For colorectal cancer, there was evidence of heterogeneity by sex (P_{het}=0.007), with male low meat-eaters, fish-eaters, and vegetarians having a lower risk of colorectal cancer (0.89, 0.83-0.95; 0.69, 0.47-1.01; 0.57, 0.36-0.91, respectively) compared to regular meat-eaters, whereas no significant association was observed across diet groups for females (**Table 8.6**).

Table 8.5 Subgroup analyses for diet groups with risk of all cancers.

	BMI							
	<27.5 kg/m^2			≥27.5 kg/m^2				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	**χ^2**	**P-value**
Regular meat-eaters	127,785	14,174	1 (ref)	118,297	15,055	1 (ref)	1.07	0.79
Low meat-eaters	123,610	13,585	0.99 (0.97 - 1.02)	80,720	10,244	0.98 (0.95 - 1.00)		
Fish-eaters	8,001	690	0.94 (0.87 - 1.01)	2,644	241	0.85 (0.75 - 0.97)		
Vegetarians	6,152	459	0.87 (0.80 - 0.96)	2,475	208	0.91 (0.79 - 1.04)		
	Smoking status							
	Never			**Ever**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	**χ^2**	**P-value**
Regular meat-eaters	132,294	13,154	1 (ref)	113,432	15,995	1 (ref)	7.55	0.056
Low meat-eaters	114,385	11,699	1.01 (0.98 - 1.03)	90,017	12,102	0.97 (0.94 - 0.99)		
Fish-eaters	6,075	498	0.95 (0.87 - 1.04)	4,580	439	0.86 (0.78 - 0.95)		
Vegetarians	5,561	405	0.94 (0.85 - 1.04)	3,081	262	0.79 (0.70 - 0.90)		
	Smoking status – censoring lung cancer cases							
	Never			**Ever**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	**χ^2**	**P-value**
Regular meat-eaters	132,050	12,911	1 (ref)	111,542	14,110	1 (ref)	4.38	0.22
Low meat-eaters	114,105	11,419	1.00 (0.98 - 1.03)	88,648	10,735	0.98 (0.96 - 1.01)		
Fish-eaters	6,062	485	0.95 (0.87 - 1.04)	4,536	395	0.89 (0.80 - 0.98)		
Vegetarians	5,550	394	0.93 (0.84 - 1.03)	3,057	238	0.82 (0.72 - 0.93)		

All models used age as the underlying time variable and are stratified by sex and age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking status was the subgroup of interest), alcohol consumption, ethnicity, diabetes status, menopausal hormone therapy use, and menopausal status. Full details for each covariate are provided in the statistical analysis section in the main text.

χ^2 and p-values from likelihood ratio tests for model fit comparing a model without an interaction term between subgroup and diet groups, to a model including an interaction between subgroup and diet groups.

Abbreviations: BMI: body mass index; CI: confidence intervals; HR: hazard ratios; ref: reference group.

Table 8.6 Subgroup analyses for diet groups with risk of colorectal cancer.

	Sex							
	Female			**Male**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	χ^2	**P-value**
Regular meat-eaters	114,849	1,168	1 (ref)	132,722	2,129	1 (ref)	12.17	0.007
Low meat-eaters	126,165	1,280	0.96 (0.89 - 1.04)	79,220	1,150	0.89 (0.83 - 0.95)		
Fish-eaters	7,664	66	0.97 (0.75 - 1.24)	3,032	26	0.69 (0.47 - 1.01)		
Vegetarians	5,722	45	0.97 (0.72 - 1.31)	2,963	18	0.57 (0.36 - 0.91)		
	BMI							
	<27.5 kg/m^2			**≥27.5 kg/m^2**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	χ^2	**P-value**
Regular meat-eaters	127,785	1,553	1 (ref)	118,297	1,728	1 (ref)	2.79	0.43
Low meat-eaters	123,610	1,358	0.90 (0.84 - 0.97)	80.720	1,058	0.92 (0.86 - 1.00)		
Fish-eaters	8,001	69	0.87 (0.68 - 1.11)	2,644	20	0.69 (0.44 - 1.07)		
Vegetarians	6,152	48	0.86 (0.64 - 1.15)	2,475	15	0.65 (0.39 - 1.09)		
	Smoking status							
	Never			**Ever**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	χ^2	**P-value**
Regular meat-eaters	132,294	1,472	1 (ref)	113,432	1,799	1 (ref)	2.17	0.54
Low meat-eaters	114,385	1,180	0.92 (0.85 - 0.99)	90,017	1,243	0.92 (0.86 - 1.00)		
Fish-eaters	6,075	51	0.91 (0.69 - 1.21)	4,580	41	0.82 (0.60 - 1.11)		
Vegetarians	5,561	31	0.68 (0.47 - 0.97)	3,081	32	0.99 (0.70 - 1.41)		

All models used age as the underlying time variable and are stratified by sex (except when sex was the subgroup of interest) and age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking was subgroup of interest), alcohol consumption, ethnicity, diabetes status, menopausal hormone therapy use, menopausal status, non-steroid anti-inflammatory drug use. Full details for each covariate are provided in the statistical analysis section in the main text.

χ^2 and p-values from likelihood ratio tests for model fit comparing a model without an interaction term between subgroup and diet groups, to a model including an interaction between subgroup and diet groups.

Abbreviations: BMI: body mass index; CI: confidence intervals; HR: hazard ratios; ref: reference group.

8.4.5 Sensitivity analyses

Associations remained largely the same when analyses were restricted to participants of white European ancestry, and when participants with missing data were excluded (Ancillary Figure 8.1). When the participants who had an event or were censored in the first two years of follow-up were excluded, results remained mostly the same except that being a fish-eater was more strongly associated with a lower risk of prostate cancer (HR 0.69, 0.55-0.88) compared to regular meat-eaters (Ancillary Figure 8.2). In analyses additionally adjusted for intake of fruit and vegetables in the multivariable models, no changes in associations were observed (Ancillary Figure 8.2). For prostate cancer risk, when PSA testing was added to multivariable models the associations were not materially changed (**Table 8.3**).

8.4.6 Mediation analyses

Adjusted and relative means of BMI, IGF-I, and free testosterone across diet groups are shown in **Table 8.7**. Explorations of potential mediators for significant diet-cancer associations are shown in **Table 8.8**. When the potential of mediation via BMI in the associations of diet groups and risk of all cancer was considered, this was not found to substantially mediate the observed associations (**Table 8.8**). For colorectal cancer risk, BMI was not found to mediate the observed lower risk in low meat-eaters compared with regular meat-eaters (**Table 8.8**); hormonal biomarkers were not explored because no differences in concentrations were observed between regular and low meat-eaters (**Table 8.7**). For postmenopausal breast cancer risk, BMI was found to be a potential mediator for the observed difference in the risk between vegetarians and regular meat-eaters, with a decomposed HR^{NIE} of 0.83 (95% CI: 0.63-1.08) implying that BMI may explain nearly 93% of the lower risk observed in vegetarian women although this was not statistically significant (**Table 8.8**). When IGF-I was explored independently and after adjusting for BMI, a HR^{NIE} of 0.91 (95% CI: 0.73-1.15) was observed. For prostate cancer risk, IGF-I and free testosterone concentrations did not appear to mediate the observed difference in risk between vegetarians and regular meat-eaters, and free testosterone was not found to mediate the difference in risk between fish-eaters and regular meat-eaters (**Table 8.8**).

Table 8.7 Adjusted and relative means (95% CI) of BMI, IGF-I, and free testosterone concentrations measured at recruitment across diet groups.

	BMI (kg/m^2)		**IGF-I (nmol/L)**		**Free testosterone (pmol/L)**	
All	**Adjusted mean**	**Relative Mean**	**Adjusted mean**	**Relative Mean**	**Adjusted mean**	**Relative Mean**
Regular meat-eater	27.90 (27.88-27.92)	1 (ref)	21.49 (21.47-21.51)	1 (ref)	119.67 (119.47-119.87)	1 (ref)
Low meat-eater	27.03 (27.01-27.05)	0.97 (0.97 - 0.97)	21.43 (21.40-21.45)	1.00 (1.00 - 1.00)	119.19 (118.97-119.42)	1.00 (0.99 - 1.00)
Fish-eater	25.68 (25.60-25.77)	0.92 (0.92 - 0.92)	21.20 (21.10-21.31)	0.99 (0.98 - 0.99)	117.84 (116.85-118.82)	0.98 (0.98 - 0.99)
Vegetarian	25.87 (25.77-25.97)	0.93 (0.92 - 0.93)	20.29 (20.17-20.41)	0.94 (0.94 - 0.95)	117.69 (116.58-118.80)	0.98 (0.97 - 0.99)
Females						
Regular meat-eater	27.71 (27.68-27.74)	1 (ref)	21.16 (21.12-21.19)	1 (ref)	15.28 (15.21-15.35)	1 (ref)
Low meat-eater	26.69 (26.66-26.72)	0.96 (0.96 – 0.96)	21.00 (20.97-21.03)	0.99 (0.99 - 0.99)	14.88 (14.81-14.95)	0.97 (0.97 - 0.98)
Fish-eater	25.43 (25.32-25.54)	0.92 (0.91 - 0.92)	20.66 (20.53-20.78)	0.98 (0.97 - 0.98)	13.98 (13.70-14.27)	0.92 (0.90 - 0.93)
Vegetarian	25.61 (25.48-25.74)	0.92 (0.91 - 0.93)	19.69 (19.54-19.84)	0.93 (0.92 - 0.94)	14.39 (14.05-14.73)	0.94 (0.92 - 0.96)
Males						
Regular meat-eater	28.14 (28.12-28.17)	1 (ref)	21.91 (21.88-21.94)	1 (ref)	220.00 (219.65-220.36)	1 (ref)
Low meat-eater	27.47 (27.45-27.50)	0.98 (0.98 - 0.98)	21.92 (21.88-21.96)	1.00 (1.00 - 1.00)	219.77 (219.31-220.23)	1.00 (1.00 - 1.00)
Fish-eater	26.05 (25.91-26.20)	0.93 (0.93 - 0.93)	22.01 (21.81-22.20)	1.00 (1.00 - 1.01)	218.65 (216.32-220.99)	0.99 (0.98 - 1.00)
Vegetarian	26.10 (25.95-26.25)	0.93 (0.92 - 0.93)	21.17 (20.97-21.37)	0.97 (0.96 - 0.98)	217.41 (215.01-219.81)	0.99 (0.98 – 1.00)

All biomarkers are adjusted for: age groups, sex (for all participants), region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, diabetes status, menopausal hormone therapy (except for males) use and menopausal status (except for males), and body mass index (except when BMI was the outcome).

Abbreviations: BMI, body mass index; CI, confidence intervals; IGF-I, insulin-like growth factor-I; ref, reference group.

Table 8.8 Summary of estimated direct effect, indirect effect, and total effect using potential mediators for the association of diet groups in comparison to regular meat-eaters and risk of all cancer, colorectal cancer, postmenopausal breast cancer, and prostate cancer risk.

	Potential mediators (Hazard ratio; 95% CI)		
All cancer	**Mediation through BMI**	**Mediation through IGF-I [a]**	**Mediation through free testosterone [a]**
Low meat-eaters versus regular meat-eaters	(n=450,412)		
Total effect	0.99 (0.96 – 1.00)		
Natural indirect effect	0.99 (0.98 – 1.00)		
Natural direct effect	0.99 (0.95 – 1.00)		
	Mediation through BMI (n=256,727)		
Fish-eaters versus regular meat-eaters			
Total effect	0.90 (0.83 – 0.97)		
Natural indirect effect	0.99 (0.90 – 1.12)		
Natural direct effect	0.90 (0.83 – 0.98)		
	Mediation through BMI (n=254,709)		
Vegetarians versus regular meat-eaters			
Total effect	0.86 (0.78 – 0.96)		
Natural indirect effect	0.94 (0.81 – 1.08)		
Natural direct effect	0.92 (0.77 – 1.09)		
Colorectal cancer	Mediation through BMI (n=450,412)		
Low meat-eaters versus regular meat-eaters			
Total effect	0.91 (0.85 – 0.97)		
Natural indirect effect	1.00 (0.98 – 1.02)		
Natural direct effect	0.91 (0.86 – 0.97)		

Table 8.8 continued

Postmenopausal breast cancer	Mediation through BMI (n=111,574)	Mediation through IGF-I (n=103,853)	Mediation through free testosterone (n=93,662)
Vegetarians versus regular meat-eaters			
Total effect	0.82 (0.68 – 0.99)	0.86 (0.71 – 1.05)	0.86 (0.71 – 1.05)
Natural indirect effect	0.83 (0.63 – 1.08)	0.91 (0.73 – 1.15)	1.06 (0.76 – 1.38)
Natural direct effect	0.99 (0.79 – 1.23)	0.94 (0.70 – 1.21)	0.81 (0.62 – 1.05)

Prostate cancer [b]		Mediation through IGF-I (n=126,538)	Mediation through free testosterone (n=116,087)
Vegetarians versus regular meat-eaters			
Total effect		0.71 (0.56 - 0.92)	0.71 (0.56 - 0.92)
Natural indirect effect		1.10 (0.77 - 1.56)	0.99 (0.67 - 1.48)
Natural direct effect		0.64 (0.50 - 1.01)	0.71 (0.51- 1.01)
Fish-eaters versus regular meat-eaters			Mediation through free testosterone (n=116,186)
Total effect			0.80 (0.65-0.99)
Natural indirect effect			0.95 (0.70-1.29)
Natural direct effect			0.86 (0.56-1.32)

All models used age as the underlying time variable and are stratified by sex (for only all cancer and colorectal cancer) and age groups at recruitment, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, diabetes status, and body mass index (except when it was considered a potential mediator). For all cancer and colorectal cancer, models are further adjusted for menopausal hormone therapy use and menopausal status. Colorectal cancer models are adjusted for non-steroid anti-inflammatory drug use. Postmenopausal breast cancer models are further adjusted for menopausal hormone therapy use, age at menarche, and age at first birth/ parity. Full details for each covariate are provided in the statistical analysis section in the main text. Prostate cancer models are further adjusted for marital status.

Mediation analyses restricted to significant associations between diet-cancer in the main analyses (Figure 8.2) and if there was a significant difference in biomarker concentrations between diet group (Table 8.7).

Natural indirect effect represents the estimated association of diet group and cancer outcome through the potential mediator.

Natural direct effect represents the estimated association of diet group and cancer outcome not through the potential mediator.

Models exclude participants with missing values for mediator(s).

[a] Models are adjusted for BMI.

[b] BMI not assessed as a mediator with total prostate cancer risk. Association of IGF-I and free testosterone presented as both hormones have been associated with prostate cancer risk. IGF-I concentrations not assessed for fish-eaters as no difference in concentrations in comparison to regular meat-eaters was observed.

Abbreviations: BMI: body mass index; CI, confidence intervals; IGF-I, insulin like growth factor-I.

8.5 Discussion

In this large British cohort, being a low meat-eater, fish-eater, or vegetarian was associated with a lower risk of all cancer sites compared to regular meat-eaters. This study also observed a lower risk of colorectal cancer among low meat-eaters, a lower risk of postmenopausal breast cancer risk in vegetarian women, and a lower risk of prostate cancer among vegetarian men. The lower risk of postmenopausal breast cancer in vegetarians may be largely a result of vegetarians having a lower BMI than regular meat-eaters, with possibly some further impact due to vegetarian women in this population having slightly lower circulating IGF-I concentrations.

8.5.1 All cancer

In this study, vegetarians, fish-eaters, and low meat-eaters all had a lower risk of developing all cancer compared to regular meat-eaters. It is important to consider that although some cancers may have similar aetiologies, some cancer sites may not be associated with dietary or nutritional factors and that using all cancer incidence as an outcome may crudely capture other lifestyle factors, external to diet, that may be associated with cancer risk and may confound the associations observed; therefore, these results should be interpreted carefully. In the two largest previous prospective studies following vegetarians, EPIC-Oxford and AHS-2 found that being a vegetarian was associated with a 10% and 8% lower risk of all cancer than being a meat-eater, respectively, after adjusting for lifestyle risk factors and BMI [117, 118]. Fish-eaters in EPIC-Oxford had a lower risk of developing all cancer [117] but no association with risk for all cancer was observed for fish-eaters compared to meat-eaters in AHS-2 [118]. In the current analysis, there was potential evidence of heterogeneity by smoking status, and in analyses which removed lung cancer from all cancer cases, significant associations were observed only across diet groups within the ever smoker subgroup. Therefore, the differences observed between diet groups for all cancer outcomes combined may not be due to diet and might be due to residual confounding by differences in other lifestyle factors, such as smoking.

8.5.2 Colorectal cancer

The risk of colorectal cancer was lower in low meat-eaters compared to regular meat-eaters whereas there was no significant difference for fish-eaters and vegetarians, potentially due to lack of power as the point estimates suggested lower risks in both these non-meat-eating diet groups. In both EPIC-Oxford and AHS-2, being a fish-eater was associated with a lower risk of colorectal cancer compared to meat-eaters, whereas no association was observed for being vegetarian and risk of colorectal cancer compared to regular meat-eaters [117, 119]. Heterogeneity was also observed by sex, in that significant inverse associations were observed with risk across diet groups in men, when compared to regular meat-eaters, but not for women. This may in part be due to dietary differences between sexes, however, the number of colorectal cancer cases in some diet groups was too small to draw a clear conclusion. The intake of processed meat has been evaluated by the World Health Organization and WCRF to be a definite cause of colorectal cancer [102] and red meat as a probable cause of colorectal cancer [20, 102]. This is likely to, at least, in part explain the lower risk of colorectal cancer in low meat-eaters, and mechanisms suggested include chemicals in processed meat such as nitrites [102, 439]. Overweight and obesity also increase the risk of colorectal cancer [81, 482] but in mediation analyses, BMI did not appear to mediate the difference observed between low meat-eaters and regular meat-eaters.

8.5.3 Postmenopausal breast cancer

A borderline significantly lower risk of postmenopausal breast cancer was observed for vegetarian women, which appeared to be largely due to their lower BMI as evidenced in mediation analyses and the attenuation of estimates when analyses were adjusted for BMI. A small potential effect for mediation for lower risk of postmenopausal breast cancer for vegetarians through lower IGF-I concentrations was also observed, perhaps influenced by the inclusion of vegans in this group [298]. To date, studies have reported a non-significantly lower risk of breast cancer for women following a vegetarian or pescatarian diet with or without adjustment for BMI [117, 212, 468, 483], which may be due to the lack of power to detect

modest associations in individual studies. Breast cancer is a heterogeneous disease, with differing risk factors by menopausal status and hormone receptor status [130]. BMI is robustly associated with higher postmenopausal breast cancer risk, probably due to higher circulating oestrogen derived from aromatisation of androgens in the adipose tissue [130]. As such, being vegetarian would be expected to confer a lower risk of postmenopausal breast cancer compared to meat-eaters because vegetarians generally have a lower BMI, but whether BMI is a confounder or a mediator for this association is unclear; if vegetarians have a lower BMI because of their diet then BMI would be a mediator, but if vegetarians have a lower BMI that is not due to their dietary intake but rather due to other non-dietary lifestyle factors (e.g. physical activity), then BMI would be deemed a confounder.

Previous research has also suggested that vegetarian women are less likely to use MHT or to attend breast cancer screening [300]. In this analysis MHT use at baseline was adjusted for, but residual confounding due to differences in use of MHT during follow-up is still possible. Data on breast cancer screening during follow-up were not available in this cohort, therefore full adjustment for screening attendance was not possible and differences between diet groups in screening may have influenced these findings.

8.5.4 Prostate cancer

The risk of prostate cancer was lower in men who were vegetarians or fish-eaters compared to regular meat-eaters, but no difference in risk was observed for low meat-eaters. Previous analyses in the EPIC-Oxford cohort found a non-significantly lower risk of prostate cancer for British vegetarians and fish-eaters compared to meat-eaters [117]. In the AHS-2 study, no difference was found for vegetarians or fish-eaters, whereas being vegan was associated with a 35% lower risk of prostate cancer (based on 1079 cases in the cohort of which only 59 were in vegans)[297]. To date, no established dietary risk factor has been found in relation to prostate cancer risk, although some evidence suggests that higher intake of dairy products, and possibly milk specifically, may increase the risk of prostate cancer [262]. This association has been

proposed to be possibly mediated through IGF-I [5, 299], a hormone shown to be positively associated with both milk intake and prostate cancer risk [11, 14]. In this cohort, slightly lower IGF-I concentrations were observed in vegetarians compared to regular meat-eaters [472], and IGF-I has also been associated with prostate cancer risk [14]; however, the difference in IGF-I concentrations between these diet groups is small and may not confer a substantial difference in prostate cancer risk. As might be expected, in mediation analyses, the estimates were imprecise and there was no evidence that the difference in IGF-I concentrations between diet groups mediates the observed associations with cancer risk.

In this cohort, vegetarian men were less likely than meat-eaters to have had a PSA screening test at recruitment, therefore vegetarians may have a lower risk of having prostate cancer diagnosed following a PSA test. Similarly, two other cohorts have also reported that vegetarian men were less likely to have had a PSA test [300, 484]. When data at recruitment and available general practice records during follow-up were assessed for PSA testing in the UK Biobank, there was only a small difference with 40% of regular meat-eaters and 37% of vegetarians reporting having had a PSA test (although general practice records were only available for half of the participants) after adjusting for age differences. Adding PSA screening in the multivariable-adjusted model did not attenuate the estimates, suggesting the differences in PSA screening in vegetarians or fish-eaters compared to regular meat-eaters does not explain the observed associations, but other differences in attendance for medical examinations could possibly also contribute. Due to unavailable data in UK Biobank, analyses could not be stratified to assess associations by tumour subtypes, which may be aetiologically different [268]. Considering the substantial difference in risk observed for vegetarian men, differences in detection and residual confounding, as well as chance, may contribute to this observed difference.

8.5.5 Interpretation of results: role of confounding and mediation

The role of residual and unmeasured confounding must be considered when interpreting the findings from this study. Vegetarians and fish-eaters differ from meat-eaters in many non-dietary lifestyle factors such as lower smoking and alcohol consumption, and higher physical activity [215]. Although relevant potential confounders were added to the multivariable models to adjust for these differences, imperfect measurements and/or changes in these confounders over time may have resulted in incomplete adjustment for these variables. For example, the evidence of heterogeneity by smoking status when looking at all cancer as an outcome suggested that residual confounding by smoking may be present.

Differences in BMI between diet groups have also been suggested to explain the lower cancer incidence observed amongst vegetarians [469], however, when BMI was considered a potential confounder and mediator, the difference between BMI by diet groups only slightly attenuated the estimates, with the exception of postmenopausal breast cancer. Whether differences in BMI by diet group is due solely to their diet or other lifestyle factors remains unclear, making it difficult to tease out whether BMI mediates or confounds the associations between diet group and cancer risk.

8.5.6 Strengths and limitations

The strengths of this study include the prospective nature and moderately long follow-up time of participants. Data-linkage to health records was used to determine cancer diagnoses, which minimises misclassification and loss to follow-up of participants. The UK Biobank study also gathered data on an array of potential confounders and biochemical biomarkers among participants; thus, the models could be adjusted for potential confounding as well as mediation analysis exploring potential mediators between diet groups and cancer risk could be investigated. When analyses excluded the first two years of follow-up, the results remained largely the same, reducing the chance that these associations are due to reverse causality.

There are some limitations to consider in these analyses. Although there were many cancer cases accrued during the follow-up period, these analyses may still be underpowered to detect moderate associations due to the relatively small numbers of cancer cases among vegetarians and fish-eaters in this cohort. Hospital admission data was also used to follow-up participants after 2015 in Scotland and 2019 in England because cancer registry data were not available in UK Biobank after this date, which may result in some missing cancer cases and relatively later dates of diagnosis. Adjustment for total energy intake was not possible as this could not be calculated due to the limited number of dietary questions asked at recruitment. As detailed above, the results may be influenced by unmeasured and residual confounding, as well as chance with numerous comparisons, and causality cannot be confirmed. Misclassification of diet may also be possible, as participants may have underreported their intake or changed their diet over the follow-up period, possibly resulting in attenuation of the risk estimates. Vegetarian diets are characterised by not consuming meat; however, this does not necessarily mean that all vegetarians follow a healthy diet, which may influence their risk of cancer and these results. The mediation analyses only explored three potential mediators, and other possible mediating factors, such as other biomarkers relevant for cancer (e.g., oestradiol), were not available. Moreover, baseline BMI and biomarkers were used to assess mediation, and therefore these measures may not represent BMI during the follow-up and long-term biomarker concentrations, although correlations with repeat measures of BMI and biomarkers showed high agreement [347]. As well, the IORW method was used to explore mediation and this method bootstraps CIs and therefore may make the estimates less statistically efficient compared to parametric methods. However, the IORW has the advantage that it can be applied in survival analysis and provides estimates of the proportion mediated for the mediators of interest. The UK Biobank has a healthier risk profile than the UK population [384] and only included British participants, most of whom are of white European ancestry (94%); therefore, these results may be limited in generalisability to other populations. However, the risks estimated may still be valid to estimate relative differences for risk-factor disease associations [425].

8.6 Conclusions

In conclusion, this analysis found that being a low meat-eater, fish-eater, or vegetarian was associated with a lower risk of all cancer, which may be a result of dietary factors and/or non-dietary differences in lifestyle such as smoking. Low meat-eaters had a lower risk of colorectal cancer, vegetarian women had a lower risk of postmenopausal breast cancer, and men who were vegetarians or fish-eaters had a lower risk of prostate cancer. BMI was found to potentially mediate or confound the association between vegetarian diets and postmenopausal breast cancer. It is not clear if the other associations are causal or a result of differences in detection between diet groups or unmeasured and residual confounding. Future research assessing cancer risk in cohorts with large numbers of vegetarians is needed to provide more precise estimates of the associations and to explore other possible mechanisms or explanations for the observed differences.

Chapter 9

Summary and recommendations for future research

9.1 Summary of findings

In this thesis, several dietary factors were explored in relation to IGF-I concentrations and IGF-I related cancers, namely, colorectal, breast, and prostate cancer, in the UK Biobank cohort. The main results in each chapter are summarised below.

Chapter 4: Cross-sectional analyses exploring food group intakes in relation to IGF-I concentrations showed that the intakes of oily fish and non-oily fish were positively associated with circulating IGF-I concentrations. A weaker positive association between poultry intake and IGF-I was also observed, whereas intakes of red meat, processed meat, and cheese were not associated with IGF-I concentrations.

Chapter 5: Observational analyses assessing intakes of macronutrient types and sources in relation to IGF-I concentrations showed that intake of protein from dairy products was positively associated with IGF-I concentrations. However, sources of dairy protein related differently to IGF-I; protein intakes from milk and yogurt were positively associated with IGF-I concentrations, whereas there was no association for intake of protein from cheese. This analysis also suggested that higher intakes of fibre and wholegrains were positively associated with circulating IGF-I concentrations.

Chapter 6: Prospective analyses assessing intakes of protein from dairy products and risks of colorectal, breast, and prostate cancer showed that higher intake of protein from dairy products, mostly from milk, was inversely associated with colorectal cancer risk, although this may not be due to protein content but other components of dairy products such as calcium. No associations were observed between protein from dairy products and breast cancer risk, whereas a borderline positive association with prostate cancer risk was observed for men in the highest quartile of intake of protein from milk in comparison to men in the lowest quartile. When IGF-I concentrations were added to multivariable adjusted models assessing the possible role of IGF-I in this association, estimates were slightly attenuated.

Chapter 7: Prospective analyses investigating the association of intakes of carbohydrate types and sources with colorectal cancer risk suggested that intakes of total sugars, non-free sugars, and fibre from wholegrains were inversely associated with colorectal cancer risk. When analyses further explored if associations were modified by host SCFA genetic modifiers, some evidence of heterogeneity was observed; an inverse association between wholegrain intake and colorectal cancer was found in participants in the highest butyrate PGS but not those in the lowest butyrate PGS category.

Chapter 8: Prospective analyses looking at diet groups, specifically regular meat-eaters, low meat-eaters, fish-eaters, and vegetarians, in relation to cancer risk showed that low meat-eaters, fish-eaters, and vegetarians all had a lower risk of overall cancer (all cancer sites combined) compared to regular meat-eaters. For specific cancer sites, low meat-eaters had a lower risk of colorectal cancer compared to regular-meat eaters. Vegetarian women had a lower risk of postmenopausal breast cancer compared to regular meat-eaters, which may be due to their lower BMI. Men who were vegetarians or fish-eaters had a lower risk of prostate cancer compared to regular meat-eaters, however this may have been due to other non-dietary factors.

9.1.1 Findings across the analyses presented in this thesis

These analyses, taken together, contribute to further understanding dietary intake in relation to IGF-I concentrations and risks of common cancers.

Protein intake, IGF-I concentrations, and cancer risk

Cross-sectional analyses in Chapter 5 provide corroborative evidence that intake of dietary protein and protein from dairy products are positively associated with IGF-I concentrations. However, dietary protein from different sources showed different associations with IGF-I concentrations; intake of protein from both milk and yogurt was positively associated with IGF-I whereas protein from cheese was not associated with IGF-I concentrations, which is in line

with previous cross-sectional analyses [3, 370]. Additionally, Chapter 4 also suggested that there was no association between cheese intake and IGF-I.

Despite positive associations between total dairy, milk and yogurt protein and IGF-I concentrations, prospective associations for colorectal cancer showed that protein from dairy products and milk were inversely associated with risk. Similarly, previous prospective and MR evidence has suggested an inverse association with intake of dairy and milk with colorectal cancer risk [20, 113], which may be due to the calcium content in dairy products. It is likely that the beneficial effect of other compounds in dairy, such as calcium, may play a more important role in preventing colorectal cancer than the potential harmful effect of increasing IGF-I concentrations. Therefore, foods or dietary components that are associated with higher IGF-I concentrations may not necessarily lead to higher cancer risk as there are many compounds in foods, and thus potentially multiple mechanisms that interplay in the prevention or initiation of carcinogenesis. There was, however, a suggestion of a positive association between protein from milk and prostate cancer risk among men in the highest quartile of intake in prospective analyses, which may suggest IGF-I concentrations are important mediators in this relationship. However, minimal attenuation was observed when IGF-I concentrations were added to multivariable models, potentially due to only using one measurement of IGF-I. Moreover, circulating IGF-I concentrations are regulated partly through binding proteins [6] and these were not measured in the UK Biobank. The bioactivity of circulating IGF-I may therefore differ by dietary intake and this needs to be further characterised in future studies to determine if associations change. Additionally, there may be other mechanisms external to the IGF-I pathway that may mediate the milk protein – prostate cancer association, such as stimulation of mTOR by branched-chain amino acids [305], and this should be explored further.

The changes in IGF-I concentrations observed in relation to dietary intake have led some researchers to propose cancer prevention strategies by modifying IGF-I concentrations through diet and other lifestyle factors [485]. However, as observed in this thesis, dietary factors that are associated with higher IGF-I concentrations, such as intake of milk, may be inversely associated with certain cancer types. As such, the potential adverse impact of intake of milk on higher

IGF-I concentrations, and prostate cancer risk, may be in opposition to the beneficial influence of milk intake on colorectal cancer risk. Moreover, considering the magnitude of the difference in IGF-I concentrations associated with higher intake of protein from milk, the influence of milk increasing cancer risk at specific cancer sites may be small, although this might be underestimated due to measurement error in dietary intakes which is discussed in greater detail below. Further research describing how different foods influence cancer risk at different sites is needed to provide overall dietary recommendations for cancer prevention.

Carbohydrate intakes and colorectal cancer risk

The insulin-IGF axis may be particularly important for carcinogenesis as hyperinsulinemia may result in greater bioavailability of IGF-I concentrations [324] and insulin has downstream signalling (PI3K–AKT–mTOR) common to that of IGF-I through the insulin receptor [8, 78]. Previous RCTs have shown that intakes of carbohydrate types and sources have different effects on insulin response [486], and results from Chapter 5 suggested that intake of fibre and wholegrains were positively associated with IGF-I concentrations. SCFA production through fibre fermentation has been implicated in lowering colorectal cancer risk [108, 431], however, SCFAs may elevate IGF-I concentrations [397, 398] thus several mechanisms may work simultaneously and the beneficial effect of SCFAs on the colon may be greater than the slightly deleterious effect of elevated IGF-I concentrations.

Findings from these analyses suggested that sugars, mostly from non-free sugar sources (i.e., fruit, vegetables, and dairy) and fibre from wholegrains were inversely associated with colorectal cancer risk. However, exploring the potential role of host genetic factors for SCFA production in modulating carbohydrate types and sources and colorectal cancer risk suggested that wholegrain intake was more clearly inversely associated with risk among those who have genetic variants for higher butyrate synthesis. This finding potentially suggests the importance of butyrate production from wholegrain fermentation and the presence of butyrate producing bacteria in modulating colorectal carcinogenesis, however further work is needed to replicate this finding. Moreover, how intakes of wholegrain or fibre, and potentially SCFA, influence IGF-I

concentrations needs to be further characterised as multiple mechanistic pathways may interplay in influencing colorectal carcinogenesis.

Dietary groups, cancer risk, and potential mediators

Dietary patterns may better capture nutritional factors that are associated with cancer risk as foods and nutrients are consumed together and may act synergistically on different metabolic pathways. Vegetarians and vegans have been shown to have lower circulating IGF-I concentrations compared to meat-eaters [298, 299, 472], which may be due to differences in dietary intake. However, previous evidence has not fully clarified why different dietary groups that consume low, or no meat (i.e., fish-eaters or vegetarians) may have a lower risk of developing any cancer, or whether these individuals have a lower risk of specific cancer sites. In this thesis, vegetarians and fish-eaters were observed to have a lower risk of developing cancer overall, which is in line with previous evidence [117, 118]. Differences in risks of colorectal, postmenopausal breast, and prostate cancer by diet groups were also observed, and potential mediators for these associations were explored. Mediation analyses did not suggest that IGF-I concentrations measured at recruitment were a major mediator between diet group and cancer risk associations; however, for the lower risk of postmenopausal breast among vegetarian women, BMI was suggested to be an important mediator, or potentially a confounder, in this association. For prostate cancer, inverse associations were observed for both vegetarians and fish-eaters, however, neither IGF-I concentrations or free testosterone appeared to explain this association, and these risk differences could be a result of differences in non-dietary factors such as PSA testing.

Ultimately, understanding the mechanisms between dietary foods, nutrients, or patterns and cancer risk remains difficult as the same foods contain different compounds that may be involved in numerous and differing metabolic pathways modulating cancer risk. Although this thesis potentially elucidated some novel associations, there are some methodological considerations that must be contemplated regarding the findings.

9.2 Methodological considerations

9.2.1 Sample size and statistical power

All analyses included in this thesis contained relatively large sample sizes, and in some instances, the largest number of participants in a similar study to date (e.g., Chapters 4 & 5). For prospective analyses, around 1,000-9,000 incident cancer cases were observed for each cancer site over a relatively long follow-up period (>9 years). However, some of these analyses may still remain under powered. For example, relatively few participants were vegetarians (and vegans) and fish-eaters and although more than 5,000 cancer cases were ascertained for each site, in some instances only 60-90 cases were observed for these diet groups, thus there may not be sufficient power to observe modest associations (less than 10% difference in risk) between diet groups and cancer risk. The limited statistical power may also apply to prospective analyses in the subsample of participants who completed the 24-hour dietary assessments, as fewer cases of colorectal cancer were observed over the follow-up period due to a smaller sample size and less follow-up time. Many associations were borderline significant or non-significant in both Chapters 6 & 7, which may be due to the lack of cases and power.

Multiple testing also needs to be considered as numerous tests were conducted in each chapter, which may result in a type I error. Although corrections for multiple testing were conducted in Chapters 4 & 5 and all associations remained statistically significant, formal correction for multiple testing was not undertaken in Chapters 6-8 but was considered when interpreting the findings. Chapters 6-8 assessed prospective associations with cancer risk, thus the number of cancer cases limited the power of each analysis. Moreover, in Chapter 6, multiple testing was not corrected for as this would not change the conclusion of the results as significant inverse associations between dairy or milk protein and colorectal cancer risk would remain even after correcting for multiple testing whereas other findings from this chapter were borderline/not significant (e.g., protein from milk and prostate cancer risk) without correcting for multiple comparisons. In Chapter 7, if correction for multiple testing had been applied, none of the main results would be significant, which may be due to the modest associations between carbohydrate types and sources and colorectal cancer risk, and relatively few colorectal cancer cases. Taking

multiple testing into consideration, results for Chapter 7 should be interpreted cautiously with further work needed to assess these associations with a greater number of colorectal cancer cases. Additionally, in Chapter 8, although multiple testing was not explicitly outlined, all significant associations met Bonferroni correction and therefore, such correction would not alter the conclusions of the results.

9.2.2 Selection bias

The UK Biobank consists of middle-aged individuals, among whom most participants (~94%) are of white European ancestry. Evidence has also shown that participants in the UK Biobank are generally healthier compared to the general UK population according to national statistics [384]. Moreover, only 5.5% of participants who were invited to the UK Biobank participated in the study, with only a quarter of these participants completing multiple 24-hour dietary assessments therefore this subsample may be even healthier. Taking this into consideration, selection bias must be acknowledged as a limitation of these analyses and could influence the findings. Men included in the UK Biobank also have a greater incidence rate of prostate cancer than the UK general population, which may be due to higher PSA testing [384]. This is particularly important as this can introduce collider bias; if the exposure of interest and the outcome are both associated with participating in the study, this can lead to spurious associations between exposure and outcome [487]. However, recent analyses assessing risk factor-disease associations in the UK Biobank as well as among a representative sample from the UK have found directionally consistent results and similar magnitudes in associations between explored cancer associations such as smoking and lung cancer risk, and obesity and risk of obesity-related cancers [425]. One possibility to control for the potential selection bias in the UK Biobank would be to include population weights based on the greater UK population. The use of statistical methods such as inverse probability weighting can account for the unequal probability of taking part in the study and thus adjust the sample to be more representative of the greater population of interest [488]. Some researchers have recently developed inverse probability weights for the UK Biobank and these should be potentially used for future analyses [489].

Moreover, how generalisable findings are to other countries and individuals of other ethnicities remains unclear. Further research including individuals from other countries and diverse populations is needed to determine if associations remain similar.

9.2.3 Measurement error

Measurement error must be considered in all epidemiological studies. Random measurement error (non-differential) will result in reduced precision of risk estimates and therefore reduce power, which may attenuate findings towards the null, whereas systematic measurement error (differential) can bias risk estimates in either direction, depending on the nature of the error. These types of measurement errors and the implications in these findings are described in greater detail below.

Dietary measurement error

Measurement error in dietary assessment remains the greatest challenge in nutritional epidemiology. Measuring diet has been especially difficult due to methodological issues such as reporting biases, limited number of foods in questionnaires, different participant interpretations of questions (i.e., portion size, identity of foods), and not accounting for changes in diet over time. To date, most epidemiological studies on diet-disease associations have measured dietary intake using FFQs, which ask participants to report the frequency of intake of a limited number of foods over the previous 12-month period. Limitations of FFQs include recall bias and under or over reporting, which may result in greater systematic error, thus biasing estimates in either direction from the true value. The analyses included in this thesis primarily used multiple web-based 24-hour dietary assessments to estimate the dietary intake of macronutrients, and some evidence has suggested that multiple 24-hour dietary assessments and internet based 24-hour dietary assessments may provide more accurate estimates, particularly for protein intake, than other forms of dietary measurements such as FFQs [490-492]. This may be particularly important for milk intake, which is often consumed daily, and in these analyses, milk intake was estimated from multiple sources such as added to coffee (including different types of coffee such

as lattes), tea, and cereals, as well as consumed on its own, from each 24-hour dietary assessment, and it is likely this more detailed dietary collection may decrease measurement error and more accurately define intake. However, for foods that are consumed episodically, such as fish, even multiple 24-hour dietary assessments may inadequately measure usual dietary intake. This issue was observed in Chapter 6 for yogurt intake, where many participants reported not consuming yogurt on the days they completed their 24-hour dietary assessments; however, whether participants do not consume yogurt at all, or just did not consume it on the assessment days remains unknown.

Advantages of this 24-hour dietary assessment include the multiple foods and beverages included in the questionnaire, which are often greater than FFQs, as well as calculation of amounts of specific nutrients and foods consumed daily. In contrast, FFQs are typically better at assessing the composition of diet and foods consumed episodically rather than nutrient amounts. However, it should be noted that the Oxford WebQ 24-hour dietary assessment does differ from a typical 24-hour dietary recall and the way this questionnaire asks about specific foods consumed the previous day is similar to a FFQ. Moreover, the Oxford WebQ is also grouped by food groups, whereas a typical 24-hour recall allows a participant to report every food they consumed (not specific food or food groups) and asks about meals separately. In the biomarker validation study assessing estimated intake from the Oxford WebQ 24-hour dietary assessment, the correlation between average protein intake estimated from two 24-hour dietary assessments and a recovery biomarker for protein from nitrogen excretion was 0.47 (95% CI: 0.33-0.61) [359], whereas for FFQs correlation between estimated protein intake and the biomarker for protein has been observed to be anywhere from 0.07 to 0.41 [492-498]. Similar evidence has been observed for energy intake in which FFQs inadequately estimates energy intakes [492-498]. Although the 24-hour dietary assessment is not a perfect measurement of energy intake, some evidence has suggested that this assessment performs better compared to FFQs [359]. Regardless, because of this imperfect measurement of dietary intake, estimates will be partially biased towards the null.

Due to the nature of the 24-hour dietary assessments, random measurement error is inevitable due to the within-person day-to-day variation in dietary intake. Effort was made to minimise random measurement error and thus regression dilution bias by excluding participants who only completed one 24-hour dietary assessment from all analyses; however, as most participants in prospective analyses only completed two 24-hour dietary assessments and dietary intake does vary day-to-day, random measurement will still be present, particularly for foods not consumed daily. In Chapter 5, analyses were further restricted to participants who completed a minimum of four 24-hour dietary assessments to further minimize random measurement error, however, this was not possible in prospective analyses due to the limited number of participants who completed four or more assessments as this would have decreased the statistical power significantly; thus, analyses were restricted to individuals who completed two or more dietary assessments in prospective analyses using dietary intake estimates derived from the Oxford WebQ as the exposure. Systematic error may also still be likely as some individuals may under- or over-report their dietary intake or not report foods they consume and if these foods or nutrients are related to the outcome of interest, this may bias the estimates in either direction. To reduce measurement error related to under-or-over-reporting, nutrient estimates were adjusted for energy intake, and nutrient densities were used thus, minimising error associated with reporting biases.

For Chapters 4 and 8, food group intakes reported in the touchscreen baseline questionnaire were used. Food group intakes reported at recruitment compared to intakes reported by a subsample of participants who completed the repeat recruitment visits showed moderate reproducibility in that 70-90% of participants reported the same or adjacent category of intakes for the same food groups [383]. Moreover, when comparing the touchscreen to the intakes reported in the 24-hour dietary assessments, the touchscreen was found to adequately rank participants according to their reported intakes of the food groups [383]. Regardless, for Chapters 4 & 8, misclassification of diet is possible, as participants may have under- or over-reported their intake or changed their diet over the follow-up period possibly resulting in attenuation of the risk estimates. However, using Chapter 8 as an example, 92% of participants

with follow-up data (~10% of vegetarians) remained categorised as vegetarians during the repeat recruitment assessment 5 years later, 83% remained categorised as fish-eaters, and 98% of participants remained in the regular or low-meat category.

For the touchscreen analyses, due to the limited number of dietary questions, adjustment for total energy intake was not possible, and therefore, results could not be at least partly adjusted for over or under estimation of specific intakes of food groups. However, adjustment for predictors of energy intake such as height, sex, physical activity, and BMI were made to try and control for energy requirements, which should be similar to total energy intake in participants with stable weight [360].

IGF-I concentration & other biomarkers

Although IGF-I concentrations were measured in a central laboratory (minimising differences in processing and measurement between samples) and the UK Biobank used standardised methods for blood collection and storage, random measurement error between batches may still be possible due to laboratory drift; however, IGF-I concentrations were adjusted for the date of assay, which may minimise any differences in concentrations. As described in Chapter 3, some of the serum aliquots were inadvertently diluted and although most participants' aliquots were not affected (~92%), this may have resulted in further attenuation of results. However, this dilution was corrected for, and aliquot 3 (aliquot most influenced by dilution) was used only for ~1% of participants in analyses and thus is unlikely to materially influence the observed results. Only one measurement of circulating IGF-I concentrations was used in most analyses, which may not reflect long-term circulating concentrations. This also applies to the mediation analyses conducted in Chapter 8, as only one measure was used for all biomarker measurements, which will underestimate typical or long-term associations and the extent of mediation analyses.

9.2.4 Confounding

All observational findings are subject to residual and unmeasured confounding and analyses included in this thesis are not immune to this. Residual confounding arises due to imperfect

measurement of confounders, whereby including imperfectly measured confounders in statistical models only partially adjusts for these factors rather than completely controlling for them. Although the UK Biobank collected a large amount of data from participants on lifestyle, socioeconomic, and anthropometric information, allowing for multiple variables to be adjusted for in each analysis to reduce the influence of potential confounding, not every possible confounder may be adjusted for (as data may not be available) and measurement error in confounding variables may influence estimates. For example, in the analyses presented in Chapter 7, the χ^2 statistics from the LRT decreased greatly (~72%) for some exposures when additional confounders were added to the model, suggesting the estimates for the dietary exposure of interest may be subject to residual confounding [149]. Although the UK Biobank used standardised measurements and, in some instances, had trained professionals take measurements, confounding factors can change over time resulting in imperfect adjustment. This is particularly relevant for Chapter 8 as both vegetarians and fish-eaters are very different in multiple ways (e.g., physical activity, alcohol intake, BMI, smoking, screening attendance) in comparison to meat-eaters. Moreover, some participants were categorised into missing or unknown categories if information on a specific confounding variable was not available. As such, whether some associations observed in this thesis are due to confounding by other lifestyle factors remains unknown because not all confounders may have been adequately adjusted for.

9.2.5 Reverse causality

For all analyses, results may be subject to reverse causality in which the outcome precedes and causes the exposure, thus altering the temporality of the association. In Chapter 5, IGF-I concentrations were measured in blood samples collected before most dietary assessments were completed, in which temporality cannot be considered; however, secondary analyses using the repeat IGF-I measurement in a subsample (on average 0.6 years after the last 24-hour dietary assessment was completed) showed similar associations to the main analyses, and it is unlikely that IGF-I concentrations cause changes in dietary intake. Reverse causality may be particularly important to consider in assessing the association between dietary intake and cancer risk. For

example, a slow growing colorectal tumour can cause discomfort for numerous months and sometimes years before it is diagnosed. During this time, an individual may stop consuming certain foods, thus if they are to limit the intake of a specific food or nutrient, it may appear that that food or nutrient is inversely associated with risk. To assess for reverse causality, all analyses excluded individuals with a previous cancer diagnosis and sensitivity analyses for prospective analyses removed individuals with a cancer diagnosis within the first two years of follow-up to see whether results changed, and no noticeable differences were observed. All prospective analyses included in this thesis had relatively long follow-up time (median 9.4-11.4 years), thus making reverse causality less plausible, but this still cannot be ruled out.

9.2.6 Follow-up of participants and cancer subtypes

Participants in the UK Biobank were followed via record linkage to national health records, thus minimising the chance of loss to follow-up, which can bias associations and result in less statistical power due to the loss of cases. In prospective analyses, cancer cases were ascertained using multiple methods through the cancer registries and hospital admission data. This method was employed to extend the follow-up time of participants as the cancer registry data generally ended earlier than the hospital admission data. This method also allowed for a greater number of cases, thus increasing statistical power in analyses. However, on average, the cancer registry will contain the earliest possible date of cancer diagnosis as it utilises multiple sources of information, including hospital admission data, to determine the date of diagnosis. Using only the hospital admissions data may slightly bias estimates as cases of cancer could be missing, particularly for cancers that may not require a hospital admission (e.g., early-stage prostate cancer) and individuals defined as a case from only the hospital admission data may have a later date of diagnosis than those from the cancer registry. However, analyses assessing cancer ascertainment from hospital admissions and the cancer registry have found comparable results in the number and types of cancers, and dates of diagnosis [499].

At the time of analyses and writing this thesis, the UK Biobank did not have information on the stage and grade of cancer, nor any information on molecular subtypes. This

may be particularly important as cancer subtypes can be heterogeneous, particularly for prostate cancer, as detection via PSA testing may result in greater diagnosis of indolent non-aggressive prostate cancers, which could influence the observed results. Previous evidence has indicated that men with higher levels of education are generally more health conscious and more likely to have a PSA test [424], and thus potentially be diagnosed with prostate cancer. Moreover, vegetarians may be less likely to attend cancer screening (i.e., breast mammography) and have a PSA test [300]. Considering this, detection bias in prospective analyses may bias estimates and should be considered a potential limitation in this thesis.

9.3 Recommendations for future research

Understanding what dietary factors are causally related to cancer risk holds immense public health importance. However, research assessing the effects of diet and nutrition on cancer risk is difficult due to measurement error, reporting biases, correlations between foods and nutrients, limited ranges of intakes within populations, and the observational nature of studies [500, 501]. To date, only a few dietary factors, such as processed meat and alcohol intake, have been established as risk factors for specific cancer types and future research is needed to better understand the dietary factors and mechanisms that may be causally associated with developing different types of cancer. Future research would benefit by conducting studies among diverse populations with better measurement of diet, including triangulation or results from different study designs, exploring potential mechanisms, and tumour subtypes, as well as collaborating and systematically analysing findings.

9.3.1 Diverse populations

Further research including diverse populations is needed to improve understanding of diet-cancer associations. Dietary intake varies considerable between countries, and intake has been shifting among countries as economies grow [502]. This thesis uses data among predominantly health-conscious white British participants, and future cohorts must include participants from different ethnocultural backgrounds and from different underrepresented countries to better

understand if similar or different associations are observed among these populations. There are already several large cohorts with more diverse populations, such as China Kadoorie Biobank and Mexico City Prospective Study, but further extension of research to more populations worldwide should be a priority. Continually advancing research knowledge amongst predominantly white individuals may only exacerbate health disparities. This is particularly true for genetic studies that have created PGS based on white populations to predict disease outcomes or to assess for potential heterogeneity between risk factors and genetics, which have the potential to be applied clinically [503]. In Chapter 7, genetic analyses were restricted to white participants of British ancestry to control for population stratification and because the SCFA PGS was derived from a sample of white participants. As such, analyses assessing genetic variants with disease outcomes and possible mediators need to be replicated and conducted with participants from other regions, such as Asia and Africa.

9.3.2 Better and repeated measurements of dietary intake

A major criticism that nutritional epidemiology has faced is inadequately measuring long-term dietary intake. Future efforts should be made to measure intake of diet using multiple repeated methods, such as FFQs and 24-hour dietary recalls as well as new tools during the follow-up period, to adequately measure long-term diet. Leveraging current and future internet technologies and assessments such as the Oxford WebQ (utilised in this thesis), Automated Self-Assessment 24-hour dietary recall, and myfood24 [504], will allow for future cohorts to have dietary intake well characterised and easily repeated during follow-up. These methods allow for anyone with internet access to report their dietary intake multiple times over time therefore estimating long-term dietary intake, which offers the potential to decrease random measurement error in dietary intake and cancer associations. Although this may seem ideal, large attrition is often observed with repeated measures of diet, thus future work may require incentives to retain participants to complete numerous dietary assessments [505]. Moreover, new technologies that leverage metabolomics and novel biomarkers of dietary intake may provide improved estimates due to being less prone to self-reporting bias, which may therefore elucidate potential

associations between diet and cancer risk; however, more research is needed as very few biomarkers have been discovered that are related to dietary intake, and these are still prone to confounding and reverse causality. Self-reported dietary intake can also be calibrated using recovery biomarkers (e.g., doubly labelled water for energy and nitrogen excretion for protein), which may correct for misreporting of intakes and provide more accurate estimates [501].

The analyses included in this thesis modelled estimates as a linear dose-response, and although there was no evidence of departures from linearity, some dietary factors may not have linear relationships with cancer sites [506]. Future analyses using restricted cubic spline regression might show differing relationships between dietary factors and cancer risk and should be considered when associations are not linear.

Some evidence has also suggested that early-life exposures have important consequences for cancer risk later in life, and to date, most diet-cancer associations have measured dietary intake in adulthood to determine if adult dietary intake is important for cancer risk [507]. Future research should focus on measuring diet and following participants throughout their life course as dietary and nutrient intake during certain key developmental periods, such as puberty, may be important in modifying cancer risk later in life [507]. For example, breast cancer is a disease that is strongly associated with reproductive and hormonal factors, and dietary intake may have a modest, if any, relationship with sex hormone levels in adulthood. However, breast cancer is associated with age at menarche [145] and dietary intake may also be associated with menarche timing [195] thus future research is needed to better characterise this. Following participants throughout their lives would be financially costly, require multiple re-measurements throughout timepoints, and would require decades of follow-up to obtain results on cancer outcomes, therefore conducting these types of studies is challenging although in theory increasingly feasible with modern electronic methods.

Dietary patterns, rather than individual components of dietary intake, may also be better to capture diet-cancer associations as foods are consumed together and comprise of many components which may interact, thus influencing cancer risk [501]. However, other lifestyle factors such as physical activity, body weight, and smoking are often associated with following

dietary patterns, therefore adequately controlling for these factors and teasing out the individual components of dietary intake or pattern is often difficult [501]. Despite this, efforts should be made to characterise dietary patterns that may be associated with lower cancer risk to better understand what dietary factors may be beneficial in reducing risk of cancer types as well as be more useful for policy makers.

9.3.3 Study designs

In this thesis, observational study designs, specifically cross-sectional and prospective analyses, were utilised to assess dietary intake in relation to IGF-I concentrations and risk of common cancers. Using different study designs, such as RCTs and MR, will be important in the triangulation of findings to determine potential causal dietary associations [508]. Identifying genetic variants that predict intakes of different foods or nutrients is needed to conduct MR studies of dietary intake and nutrition status, which are less susceptible to limitations of observational evidence such as confounding. However, finding variants associated with specific foods or dietary intake has been proven difficult and may suffer from weak instrument bias. As well, other drawbacks of utilising MR for diet - cancer associations include dietary intake changing over time, food and nutrients being interconnected with other behaviours and the environment, and increasing the consumption of one food will displace another, which may be associated with risk [509]. Very few genetic variants have been identified as markers for dietary intake thus making MR studies on diet difficult to conduct. Both the *LCT* gene and *ALDH2* and *ADH1B* genes have been found to be reliable genetic predictors of milk and alcohol intake, respectively. However, even utilising these genetic variants in MR analysis has proven difficult as other foods and nutrients have been determined to be associated with lactase persistence [408, 510], while genetic variants for alcohol intake have been associated with smoking [511], both thus violating the key assumptions of MR [509, 512].

Relatively few RCTs, such as the Women's Health Initiative [187] and another Canadian trial [188], have been conducted to assess dietary changes and cancer risk as RCTs focused on dietary intake face a unique set of challenges, including nonadherence to dietary interventions, contamination from controls if they change their diet, and needing a large sample of participants

following dietary interventions over a long follow-up time. Some trials have been conducted assessing specific vitamin or mineral supplementation with cancer risk, however findings have generally been null or inconclusive [513-515]; however, supplementation is not the same as dietary intake of foods and may not reflect how specific foods or overall diet influences cancer risk. As such, RCTs may not always be feasible to conduct to determine if there is an effect of a dietary component or specific diet on cancer risk.

In contrast, RCTs that assess the effects of isocaloric intakes of different protein sources on IGF-I may be helpful to determine how different intakes influence IGF-I concentrations over weeks and months. RCTs would better characterise what protein sources are associated with IGF-I concentrations and therefore provide more accurate and precise estimates for each food or protein source while avoiding potential confounding factors. If there are true differences between protein sources and IGF-I concentrations, as indicated in this thesis and other cross-sectional studies [4, 5], this will be important for determining the hypothetical effect estimate that the protein source has on IGF-I related cancers based on the difference in concentration.

9.3.4 Exploration of mechanisms and tumour heterogeneity

Understanding the mechanisms behind diet - cancer associations remains a priority for nutrition research [386]. There are many potential mechanisms in which foods and dietary intake influence cancer risk and further research is needed to understand how multiple metabolic pathways work concurrently. To do this, future cohorts must consider enhancing phenotyping of participants, including genetic and epigenetic markers, and information on the gut microbiome, metabolic profiles, as well as other biological measures to better understand and characterise associations and potential mechanisms.

Tumours are aetiologically and clinically heterogeneous with different molecular and hormone markers as well as potential for progression and metastasis. Heterogeneity in tumour subtypes may be a result of different aetiological factors and carcinogenesis pathways, underscoring the importance of looking at different subtypes of cancer separately with dietary intake. This may be particularly important for anatomical location in colorectal cancer, ER

receptor and menopausal status for breast cancer, and stage and grade of prostate cancer tumours. As some subtypes are less common than others, large studies with long duration are needed to accrue enough cases to obtain sufficient statistical power. To date, some evidence does suggest that dietary factors differ by tumour subtypes [386] and future efforts must be made to characterise tumours and assess if associations differ.

9.3.5 Collaboration and systematic analyses of existing studies

Researchers must work collectively to bring together all individual level data from prospective cohort studies to systematically analyse potential diet - cancer associations. Using individual level data offer advantages over aggregating data from publications, such as standardising analyses across studies, using the most up-to-date data regarding follow-up and cancer cases, standardised adjustment for confounders, and including unpublished data, thus limiting the chance of publication bias [516]. Bringing together data from all existing studies may be particularly important for studies assessing associations with unusual diet groups such as vegetarians and vegans, as most cohorts contain relatively few numbers of participants who follow these dietary patterns and thus limit the power to detect modest associations in individual studies. As well, for uncommon cancer sites, individual cohorts may not have the power to assess these associations due to very few number of cancer cases.

9.4 Conclusion

This thesis adds important findings to existing research on dietary factors in relation to the risk of common cancers by providing large-scale studies on dietary intake in relation to IGF-I concentrations, and risks of colorectal, breast, and prostate cancer. The main findings from the cross-sectional analyses suggested that foods and macronutrients from different sources, namely milk and yogurt protein and fibre, were positively associated with IGF-I concentrations. However, milk may be differently associated with risks of common sites of cancer having an inverse association with colorectal cancer risk and potentially a positive association with prostate cancer risk. The relationship between wholegrain intake and colorectal cancer risk may

also be modified by genetic factors for butyrate production and future work is needed to replicate this finding. As well, following different dietary patterns, which include little or no meat may be inversely related to the risk of colorectal, postmenopausal breast, and prostate cancer, and this may be due to other factors external to the IGF pathway; the lower risk of postmenopausal breast cancer in vegetarian women may be explained by their lower BMI, but it remains unclear whether differences observed for prostate cancer reflect any causal relationships or are due to other factors such as residual confounding or differences in cancer detection.

Future research assessing dietary intake and cancer risk can be advanced with enhanced measurement of diet and phenotyping of participants, conducting large prospective studies in underrepresented populations, as well as collaborative efforts to combine all existing studies to be able to assess if associations vary by tumour characteristics.

Ancillary Methods

A.1 Covariate classification

Region

The region of participants was grouped based on the recruitment centre they attended. A total of 10 regions were used corresponding approximately to the areas covered by the assessment centre: London (assessment centres: St Bartholomew's Hospital, Hounslow, Croydon), Wales (assessment centres: Swansea, Wrexham, Cardiff), North-West England (assessment centres: Stockport, Manchester, Liverpool, Bury), North-East England (assessment centres: Newcastle, Middlesbrough), Yorkshire (assessment centres: Leeds, Sheffield), West Midlands (assessment centres: Stoke, Birmingham), East Midlands (assessment centre: Nottingham), South-East England (assessment centres: Oxford, Reading), South-West England (assessment centre: Bristol), Scotland (assessment centres: Glasgow, Edinburgh).

Height

Participants were grouped into eight sex-specific categories for height. For females, categories were: <150, 150-154.9, 155-159.9, 160-164.9, 165-169.9, 170-174.9, 175-179.9, ≥180 cm. For males, categories were: <160, 160-164.9, 165-169.9, 170-174.9, 175-179.9, 180-184.9, 185-189.9, ≥190 cm. From this, males and females were combined into one height variable with categories going from 1 to 8, and a missing category for participants with missing data.

Body mass index (BMI)

Both height and weight were measured at the baseline visit and were used to determine participant's BMI. BMI was calculated by taking the participants measured weight in kilograms and dividing it by the participants squared standing height in metres. Individuals with missing data were coded into a missing category. Participants were categorised as follows: <20, 20-22.49, 22.5-24.99, 25.0-27.49, 27.5-29.99, 30-32.49, 32.5-34.99, ≥35 kg/m^2 and unknown/missing category.

Alcohol intake reported from the touchscreen questionnaire

Participants were asked on the baseline questionnaire at recruitment how often they drank alcohol with the possible responses being: "daily or almost daily", "three or four times a week", "once or twice a week", "one to three times a month", "special occasions only", "never", or "prefer not to answer". Participants were also asked about their weekly and monthly intake of pints of beer, glasses of red wine, glasses of white wine/champagne, glasses of fortified wine, measures of spirits/liqueurs and glasses of other alcohol. A pint of beer was assumed to contain 20 grams of alcohol, and all other drinks contained 10 grams of alcohol. Participants total weekly and monthly consumption of alcohol was then summed accordingly. If the participant reported "do not know" or "prefer not to answer" to one of these questions on weekly or monthly consumption, they were coded as missing, except for "other alcohol", in which case participants were assigned 0 grams from other alcohol. Participants reported weekly consumption of alcohol was used, if this was unknown (due to the participant reporting "do not know" or "prefer not to answer" for one or more of the relevant questions, except for "other alcohol") monthly consumption was used, if available. To get an estimated daily total, weekly consumption was divided by 7 (or monthly consumption by 30.4375). Alcohol consumption was categorised as <1 g/day, 1-9.99 g/day, 10-19.99 g/day, and ≥ 20 g/day, never or unknown. Participants who had unknown grams/day of alcohol intake but reported consuming alcohol intake on "special occasions" were assigned to the category of "<1 g/day".

Physical activity

Physical activity was determined from questions on the touchscreen questionnaire which asked about walking, moderate physical activity, and vigorous physical activity. These were used to estimate excess metabolic equivalent (MET)-hours/week of physical activity during work and leisure time. For each of the three activity categories (walking, moderate physical activity and vigorous physical activity), participants were asked how many days in a typical week they did each of the activities for 10 minutes or more. For each category, participants who entered one or more days were then asked how many minutes they spent doing those activities on a typical

day. For each activity category, the number of reported days was multiplied by the number of reported minutes on a typical day to generate duration of activity in minutes per week. Activity on a typical day of 1260 min per week (equivalent to an average of 3 hours per day) were truncated at 1260. Total MET values for each category from the International Physical Activity Questionnaire short form were: 3.3 for walking, 4.0 for moderate physical activity and 8.0 for vigorous physical activity. Excess MET values were therefore 2.3 for walking, 3.0 for moderate physical activity and 7.0 for vigorous physical activity. Excess MET-hours per week were calculated by multiplying the excess MET value for each activity by the duration of activity in hours per week.

Townsend deprivation index

Townsend deprivation index was based on the preceding national census output areas. Each participant was assigned a score in correspondence to the output area in which their postcode was located. From this, participants were categorised into quintiles from most deprived to least deprived and a missing category where postcode information was not provided.

Smoking status

Smoking was determined from questions from the recruitment questionnaire. Participants were asked "Do you smoke tobacco now?" and "in the past, how often have you smoked tobacco?" to determine their smoking status. Smokers were further divided based on how many cigarettes they said they smoked on average per day from the question "About how many cigarettes do you smoke on average each day?". If participants did not respond to how many cigarettes they had on average each day, they were categorised into the missing group as they could not be accurately categorised.

Ethnicity

Ethnicity of participants was determined from questions in the touchscreen questionnaire "What is your ethnic group?". Options included: White, mixed race or other, Asian or Asian British,

Black or Black British, Chinese, and other ethnic group. From this, participants were grouped into five categories: White, Mixed race or other, Asian or British Asian, and Black or Black British, or missing/unknown.

Education

For education, participants were asked "Which of the following qualifications do you have?" being able to select more than one. Possible answers were: College or University degree; A levels/AS levels or equivalent (General Certificate of Education (GCE) Advanced Level); O levels/ General Certificate of Secondary Education (GCSEs) or equivalent; Certificate of Secondary Education (CSEs) or equivalent; National vocational qualification (NVQ) or Higher National Diploma (HND) or Higher National Certificate (HNC) or equivalent; Other professional qualifications example: nursing, teaching; None of the above; Prefer not to answer. Participants were grouped into the following categories, based on their highest reported level of education: (College or University degree, vocational qualifications (other professional qualifications/NVQ or HND or HNC), optional national exams at ages 17 to 18 years (A levels/AS levels), national exams at age 16 years (O levels/GCSEs/CSEs), none of the above, unknown (prefer not to answer)).

Employment status

Employment status at recruitment was assessed by asking participants "Which of the following describes your current situation?" in which they selected answers which were applicable to them including "In paid employment or self-employed", "retired", "looking after home and/or family", "unable to work because of sickness or disability", "unemployed", "doing unpaid or voluntary work", "full or part time student", "none of the above" or "prefer not to answer". Participants were defined as being "in paid employment" if they responded they were in paid employment or self-employed, "retired" if they responded they were retired, and "not in paid employment" if they reported being unemployed, inability to work, being a student, or having caring responsibilities for their family. Finally, an unknown/missing category consisted of

participants who did not respond, said they prefer not to answer, or answered none of the above to the options in the question.

Diabetes status

Participant's diabetes status was determined using multiple variables from recruitment. First, from the question "Has a doctor ever told you that you have diabetes?" participants were classified as "yes", "no" or "unknown" based on their response. As well, participants who reported using metformin or insulin at recruitment were considered having diabetes and included in the "yes" category. Finally, if a participant had a measured glycated hemoglobin (HbA1c) of $\geq$ 48 mmol/mol at recruitment, they were defined as having diabetes and included in the "yes" category.

Marital status

Living with a partner was derived from a question asked at recruitment in which participants answered if they lived with anyone else in their household. If the participants reported to be living with a husband, wife, or partner in their household they were classified as "living with a partner" if the participant reported to be living with any other person or alone, they were categorised as "not living with a partner".

Non-steroid anti-inflammatory drugs

Non-steroid anti-inflammatory drug (NSAID) use was determined based on the medications reported at recruitment. Participants were categorised into three groups "Non-users", "irregular NSAID users", and "regular users of aspirin or ibuprofen". Participants were categorised as "regular users of aspirin or ibuprofen" if they responded to taking ibuprofen or aspirin regularly. Participants were categorised into irregular users if they responded to taking any classification of NSAID at recruitment. If no use of NSAIDs were reported at recruitment, participants were categorised as non-users.

Vegetable and fruit intake from the touchscreen questionnaire (Chapter 8)

For total vegetable intake, participants were asked to enter the number of heaped tablespoons of cooked vegetables and salad/raw vegetables eaten per day or select "less than one", "do not know" or "prefer not to answer". Two heaped tablespoons of vegetables were counted as a serving. For fresh fruit, participants were asked to enter the number of pieces of fresh fruit and dried fruit (with examples given as to what constitutes a piece eaten per day) or select "less than one", "do not know" or "prefer not to answer". One piece of fresh fruit (one apple, one banana, 10 grapes etc.) was counted as a serving. From this, servings of both vegetables and fruit were combined and participants were categorised as consuming <3 servings/day, 3.00-3.99 servings/day, 4.00-5.99 servings/day, ≥6.00 servings/day.

Prostate specific antigen testing (Chapter 8)

Prostate specific antigen (PSA) testing was determined from baseline reports from male participants as well as follow-up through general practice records. At recruitment men were asked if they ever had a PSA test in the past. From this, they answered "yes", "no" and "do not know". Additionally, 99,412 men had information from general practice records during follow-up (records available for participants until 31st of May 2016 for England, 31st of March 2017 for Scotland, and 31st of August 2017 for Wales). From the general practice records, men with a recorded PSA test in the general practice record were coded as having a PSA test. From both reported at baseline and the subsample of men with follow-up general practice records, men were coded as "no record of having a PSA" test, "had a PSA test", or "unknown" if men reported "do not know" at recruitment and no general practice records were available.

Other dietary covariates

Red and processed meat intake

Total red and processed meat intake was estimated from participants' responses to how often they consumed beef, lamb/mutton, pork, and processed meat from the options of: 'never, '<once a week', 'once a week', '2-4 times a week', '5-6 times a week' in the baseline questionnaire. This

was estimated from the recruitment questionnaire rather than the 24-hour dietary assessment due to red and processed meat being consumed more episodically by these participants. The frequency of consumption of beef, pork, lamb, and processed meat were summed by assigning participants' intakes as never as a 0, less than once a week a 0.5, once a week 1.0, 2-4 times a week a value of 3, and 5-6 times a week a value of 5.5. From this, intake of red and processed meat was categorised as <2.0 times per week, 2.0-2.9 times per week, 3.0-3.9 times per week, and ≥4.0 times per week and an unknown/missing category if participants reported 'do not know' or 'prefer not to say'.

Intake of fruit and vegetables

Intake of fruit and vegetables were derived from each 24-hour dietary assessment. Intake of all fruits and vegetables in grams from: apples/pears, berries, citrus, dried fruit, other fruit, stewed fruit, allium vegetables, leafy greens, raw salad, root vegetables, tomatoes, other vegetables, and peas and corn were summed, and participants were categorised into quintiles of total grams of fruit and vegetables estimated from the average of all available 24-hour dietary assessments.

Intake of fibre

Fibre intake was estimated from each 24-hour dietary assessment using nutrient profiles from the reported foods consumed. The mean of fibre intake was determined across all available 24-hour dietary assessments and from this, participants were categorised into sex-specific quintiles of fibre intake.

<u>Women specific covariates</u>

All men were put into a separate category for all women-specific covariates when analyses included both men and women.

Menopausal hormone therapy

For women, use of menopausal hormone therapy (MHT) was categorised as "current user", "former user" and "never user" or "unknown/missing" based on the questions asked about MHT use in the touchscreen questionnaire at recruitment. Women were asked "Have you ever used hormone replacement therapy?" and if they answered yes: "How old were you when you last used HRT?". Women were asked to enter their age when they last used HRT or could choose "Still taking HRT", or they could select "prefer not to answer" or "do not know".

Menopausal status at recruitment

Menopausal status was first determined by multiple questions asked in the baseline questionnaire. Women were defined as being pre-menopausal if they:

- Answered "no" to the question regarding having gone through menopause, or
- Reported they were "not sure" or did not respond to if they had gone through menopause and:
    - Were <50 years of age, did not have a bilateral oophorectomy/hysterectomy, and reported they were not using menopausal hormone therapy.
    - Were <50 years of age, reported they were menstruating today, and did not have a bilateral oophorectomy/hysterectomy.

Women were defined as postmenopausal if they:

- Answered "yes" to having gone through menopause
- Answered "not sure" or did not answer if they had gone through menopause and:
    - were ≥55 years of age, or
    - had a bilateral oophorectomy

Women were defined as their menopausal status being unknown if:

- Answered "no" to having gone through menopause and:
    - Did not answer no to using HRT, or
    - Did not answer no to having a bilateral oophorectomy, or

- Did not answer no to having a hysterectomy, or
- Were 50-54.9 years of age.

Menopause status defined at recruitment was used as the covariate in models for all cancer and colorectal cancer analyses.

Parity and age at first birth

Parity was defined by the recruitment question of "how many children have you given birth to?". Women who said they had given birth were asked how old they were in years when they gave birth to their first child. Based on these responses, women were categorised into groups of 0 children, 1-2 children <25 years of age, or 3+ children <25 years of age, 1-2 children 25-29.9 years of age, or 3+ children 25-29.9 years of age, 1-2 children 30+years of age, or 3+ children 30+ years of age or unknown if the participants responded, "do not know" or "prefer not to answer".

Age at menarche

Women were categorised based on the recruitment question "How old were you when your period started". From this, women were categorised into age groups as: ≤12 years, 13 years old, ≥14 years old. If a participant responded "prefer not to answer" or "do not know" they were categorised into an unknown group.

<u>Breast cancer analyses (Chapter 6):</u>

Interaction between BMI and menopause status

To account for the heterogeneity between BMI and menopause on breast cancer risk, women were categorised into 6 categories. Women who were categorised as premenopausal:<25, 25-29.9, ≥30 kg/m^2 and postmenopausal: <25, 25-29.9, ≥30 kg/m^2; due to the limited number of premenopausal women, only three categories were made. If a woman had an unknown

menopause status and/or BMI they were categorised into an unknown category. This variable replaced menopause in the model for breast cancer analyses.

A.2 Free testosterone calculation

Free testosterone (pmol/L) [473, 517]:

$$[\mathrm{FT}] = 1.0\times1012\times(\sqrt{}((\mathrm{N}/\mathrm{ks}-\mathrm{ks}+[\mathrm{S}])2/(4\mathrm{N}2)+\mathrm{ks}/(\mathrm{Nks}))-(\mathrm{N}/\mathrm{ks}-[\mathrm{T}]+[\mathrm{S}])/(2\mathrm{N}))$$

Where:

$\mathrm{N} = 1 + [\mathrm{A}]\mathrm{kA}$

[T] = total testosterone concentration (mol/L)

[S] = SHBG concentration (mol/L)

[A] = albumin concentration (assuming 6.2 x 10^{-4} mol/L)

ks = association constant of testosterone to SHBG (assuming 1.0 x 10^{9})

kA= association constant of testosterone to albumin (assuming 3.6 x 10^{4})

Ancillary Tables

Ancillary Table 4.1 List of medications and UK Biobank codes used to exclude participants taking medication at recruitment which may modify insulin-like growth factor-I concentrations.

Medication code	Medication
1140868644	somatropin
1140857838	somatrem
1140857840	somatonorm 4iu injection
1141167490	somatuline la 30mg injection (pdr)+diluent+syringe
1140857748	genotropin 12iu multidose injection
1140857750	somatropin 12iu injection
1140868646	humatrope(rbe) 4iu(1.3mg) injection (pdr for recon)+diluent
1140868648	saizen(rmc) 4iu(1.33mg) injection (pdr for recon)+diluent
1140868650	norditropin(epr) 12iu(4mg) injection (pdr for recon)+diluent
1140884544	leuprorelin
1141157394	goserelin product
1140870194	goserelin
1140870196	zoladex 3.6 mg implant
1140921100	triptorelin
1141189852	decapeptyl sr 3mg injection (pdr for recon)+diluent
1141189772	gonapeptyl depot 3.75mg inj (pdr for recon)+solv p/f syringe
1140870248	buserelin
1141157392	buserelin product
1140870252	suprefact 100micrograms nasal spray
1140868490	gestanin 5mg tablet
1140870084	depostat 200mg/2ml oily injection
1140876638	cyproterone acetate+ethinyloestradiol
1141192344	cyproterone acetate+ethinylestradiol
1140884634	cyproterone
1140868524	androcur 50mg tablet
1140869270	medroxyprogesterone
1141190580	conjugated oestrogens 0.3mg / medroxyprogesterone 1.5mg tab
1140864232	provera 2.5mg tablet
1140857620	depo-provera 50mg/1ml injection
1140870274	flutamide
1140917306	bicalutamide
1140917310	casodex 50mg tablet
1141179886	propecia 1mg tablet
1140928222	andropatch 2.5mg/24hours transdermal patch
1140910802	androstanazol
1141193272	testogel 50mg gel 5g sachet
1141166354	testoderm 6mg/24hours transdermal patch
1140868534	primoteston depot 250mg/1ml oily injection
1140868536	restandol 40mg capsule
1140868538	sustanon 100 oily injection

Ancillary Table 4.1 continued

Medication code	Medication
1140864502	testotop tts 15mg transdermal patch
1141167430	lanreotide
1140870200	octreotide
1140870208	sandostatin 50micrograms/1ml injection
1141195128	pegvisomant
1141195032	somavert 10mg injection (pdr for recon) +solvent
1140857656	methyltestosterone product
1140857668	virormone-oral 5mg tablet
1140857736	virormone 10mg/1ml injection
1140865136	yohimbine/pemoline/methyltestosterone
1140868532	testosterone product
1140868528	pro-viron 25mg tablet
1140868526	mesterolone
1140868524	androcur 50mg tablet
1140868534	primoteston depot 250mg/1ml oily injection
1140868536	restandol 40mg capsule
1140868538	sustanon 100 oily injection
1140868550	finasteride
1140868608	proscar 5mg tablet
1140868614	deca-durabolin 25mg/1ml oily injection
1140868618	stanozolol
1140868620	stromba 5mg tablet
1141179886	propecia 1mg tablet
1141192000	dutasteride
1141192004	avodart 500micrograms capsule
1140868968	danazol
1140870284	prostap sr 3.75mg injection (pdr for recon)+diluent+kit
1141201718	nebido 1000mg/4ml solution for injection
1140923018	anastrozole
1141171100	exemestane
1141145896	letrozole
1140870164	tamoxifen
1140888684	diazoxide
1140928276	humatrope(rbe) 18iu(6mg) inj cartridge(pdr for recon)+dil
1141189090	rosiglitazone 1mg
1140874686	glucophage 500mg tablet
1141190802	nutropinaq 10mg(30iu)/2ml injection cartridge
1140923890	zomacton(rbe) 12iu(4mg) injection (pdr for recon)+diluent
1140868810	geref 50 injection (pdr for recon)+solvent
1140909918	biosynthetic human growth hormone
1140882976	growth hormone product

Ancillary Table 6.1 Hazard ratios (95% CI) for intake of protein from dairy and dairy sources with risk of prostate, breast, and colorectal cancer for participants who completed a minimum of three 24-hour dietary assessments.

	Colorectal cancer (734 cases, 69,223 participants)				
	Q1	Q2	Q3	Q4	Per 2.5% energy increase
Protein from dairy	1 (ref)	0.88 (0.72 - 1.08)	0.89 (0.73 - 1.09)	**0.75 (0.61 - 0.94)**	**0.79 (0.68 - 0.92)**
Protein from milk	1 (ref)	0.85 (0.69 - 1.03)	**0.74 (0.61 - 0.91)**	**0.79 (0.64 - 0.98)**	**0.75 (0.59 - 0.95)**
Protein from cheese	1 (ref)	0.96 (0.79 - 1.18)	0.91 (0.74 - 1.11)	0.92 (0.74 - 1.13)	0.87 (0.69 - 1.10)

	Breast cancer (1,187 cases, 38,282 participants)				
	Q1	Q2	Q3	Q4	Per 2.5% energy increase
Protein from dairy	1 (ref)	1.06 (0.90 - 1.25)	1.01 (0.86 - 1.19)	0.99 (0.84 - 1.17)	0.97 (0.87 - 1.08)
Protein from milk	1 (ref)	0.99 (0.85 - 1.17)	0.93 (0.79 - 1.10)	0.96 (0.81 - 1.13)	0.91 (0.77 - 1.08)
Protein from cheese	1 (ref)	0.94 (0.79 - 1.11)	1.08 (0.92 - 1.26)	1.03 (0.88 - 1.21)	1.03 (0.87 - 1.22)

	Prostate cancer (1,451 cases, 30,941 participants)				
	Q1	Q2	Q3	Q4	Per 2.5% energy increase
Protein from dairy	1 (ref)	0.99 (0.85 - 1.15)	1.05 (0.90 - 1.22)	1.06 (0.91 - 1.23)	1.07 (0.96 - 1.19)
Protein from milk	1 (ref)	1.04 (0.89 - 1.21)	1.05 (0.90 - 1.22)	1.11 (0.95 - 1.29)	1.12 (0.95 - 1.32)
Protein from cheese	1 (ref)	0.90 (0.77 - 1.04)	0.99 (0.86 - 1.15)	0.96 (0.83 - 1.12)	0.99 (0.84 - 1.17)

Estimates in bold indicate statistically significant results.

All models used age as the underlying time variable and were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol intake, ethnicity, diagnosis of diabetes, BMI, and energy intake.

For colorectal cancer analyses: all models stratify by sex and adjust for menopause status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use.

For breast cancer analyses: models adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status.

For prostate cancer analyses: models further adjusted for marital status.

Abbreviations: BMI, body mass index; Q, quantile; ref, reference group.

Ancillary Table 6.2 Hazard ratios (95% CI) for intake of protein from dairy products and dairy sources with risk of colorectal, breast, and prostate cancer removing the first two years of follow-up.

	Colorectal cancer (978 cases, 111,724 participants)				
	Q1	Q2	Q3	Q4	Per 2.5% energy increase
Protein from dairy	1 (ref)	0.89 (0.75 - 1.05)	0.87 (0.73 - 1.04)	**0.80 (0.66 - 0.96)**	**0.83 (0.74 - 0.94)**
Protein from milk	1 (ref)	0.89 (0.75 - 1.05)	**0.78 (0.66 - 0.94)**	**0.80 (0.66 - 0.96)**	**0.76 (0.62 - 0.92)**
Protein from cheese	1 (ref)	0.89 (0.75 - 1.07)	0.93 (0.78 - 1.11)	0.90 (0.75 - 1.07)	0.92 (0.76 - 1.11)

	Breast cancer (1,598 cases, 61,756 participants)				
	Q1	Q2	Q3	Q4	Per 2.5% energy increase
Protein from dairy	1 (ref)	1.04 (0.91 - 1.20)	0.99 (0.86 - 1.14)	0.95 (0.82 - 1.09)	0.95 (0.87 - 1.04)
Protein from milk	1 (ref)	0.95 (0.83 - 1.10)	1.02 (0.89 - 1.18)	0.93 (0.81 - 1.08)	0.92 (0.80 - 1.06)
Protein from cheese	1 (ref)	0.99 (0.86 - 1.15)	1.05 (0.92 - 1.21)	1.07 (0.93 - 1.23)	1.05 (0.91 - 1.21)

	Prostate cancer (1,976 cases, 49,968 participants)				
	Q1	Q2	Q3	Q4	Per 2.5% energy increase
Protein from dairy	1 (ref)	0.97 (0.85 - 1.10)	1.07 (0.94 - 1.21)	1.02 (0.90 - 1.16)	1.04 (0.95 - 1.13)
Protein from milk	1 (ref)	1.02 (0.90 - 1.16)	0.99 (0.87 - 1.13)	1.12 (0.98 - 1.27)	1.09 (0.95 - 1.26)
Protein from cheese	1 (ref)	0.99 (0.87 - 1.13)	1.02 (0.91 - 1.16)	0.96 (0.85 - 1.09)	0.95 (0.83 - 1.09)

Estimates in bold indicate statistically significant results.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol intake, ethnicity, diagnosis of diabetes, BMI, energy intake.

For colorectal cancer analyses: all models were stratified by sex and adjusted for menopause status (women only), menopausal hormone therapy use (women only), red and processed meat intake, and non-steroidal anti-inflammatory drug use. For breast cancer analyses: models were further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status. For prostate cancer analyses: models were further adjusted for marital status.

Abbreviations: BMI, body mass index; Q, quantile; ref, reference group.

Ancillary Table 7.1 Multivariable hazard ratios and 95% confidence intervals for carbohydrate and fibre sources with colorectal cancer risk restricting to participants who completed ≥3 24-hour dietary assessments (N= 69,223; n cases = 734).

Carbohydrates, sugars, & starch	**Q1**	**Q2**	**Q3**	**Q4**	**Per 5% energy increment**	**P-trend**
Total carbohydrates	1 (ref)	0.89 (0.72 - 1.08)	0.90 (0.73 - 1.11)	0.88 (0.69 - 1.11)	0.95 (0.89 - 1.01)	0.079
Total sugars	1 (ref)	0.82 (0.67 - 1.00)	0.87 (0.71 - 1.07)	**0.67 (0.53 - 0.86)**	**0.90 (0.84 - 0.96)**	0.002
Total free sugars	1 (ref)	1.12 (0.91 - 1.36)	0.87 (0.70 - 1.08)	**0.78 (0.62 - 0.97)**	**0.90 (0.82 - 0.98)**	0.01
Total non-free sugars	1 (ref)	1.00 (0.82 - 1.22)	0.86 (0.69 - 1.07)	0.87 (0.69 - 1.10)	**0.92 (0.85 - 1.00)**	0.04
Total starch	1 (ref)	0.98 (0.80 - 1.21)	1.03 (0.84 - 1.27)	1.14 (0.92 - 1.42)	1.04 (0.97 - 1.13)	0.26
Wholegrain starch	1 (ref)	0.90 (0.73 - 1.10)	0.93 (0.76 - 1.14)	0.84 (0.68 - 1.04)	0.93 (0.84 - 1.03)	0.15
Refined grain starch	1 (ref)	0.96 (0.78 - 1.18)	1.04 (0.84 - 1.28)	1.22 (0.99 - 1.51)	1.07 (1.00 - 1.14)	0.070
Fibre	**Q1**	**Q2**	**Q3**	**Q4**	**Per 5 g/day increment**	**P-trend**
Total fibre	1 (ref)	1.08 (0.87 - 1.33)	0.94 (0.75 - 1.17)	0.89 (0.70 - 1.14)	0.96 (0.88 - 1.04)	0.33
Total fibre from vegetables & fruit	1 (ref)	1.02 (0.83 - 1.25)	0.91 (0.74 - 1.13)	0.93 (0.75 - 1.15)	0.94 (0.83 - 1.05)	0.26
Total fibre from vegetables	1 (ref)	1.10 (0.90 - 1.35)	0.91 (0.74 - 1.13)	0.91 (0.74 - 1.14)	0.89 (0.74 - 1.07)	0.20
Total fibre from fruit	1 (ref)	1.07 (0.87 - 1.31)	1.01 (0.82 - 1.25)	0.92 (0.74 - 1.15)	0.95 (0.79 - 1.15)	0.63
Total fibre from wholegrains	1 (ref)	0.94 (0.76 - 1.15)	0.92 (0.75 - 1.13)	0.85 (0.69 - 1.06)	0.89 (0.78 - 1.02)	0.09

Models stratified for sex and age at recruitment, and further adjusted for region, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, BMI, processed and red meat intake, total vegetable fruit intake (except for total fibre, fibre from vegetables and/or fruit, or non-free sugars is the exposure), and female specific covariates: menopausal hormone therapy use and menopausal status.

P-trend values represent the p-values for carbohydrate types/source modelled as 5% energy increase per day in the model and fibre by 5 g/day increase in the model.

Abbreviations: BMI, body mass index; g/day, grams per day; Q, quartile; ref, reference group.

Ancillary Table 7.2 Multivariable hazard ratios and 95% confidence intervals for carbohydrate and fibre sources with colorectal cancer risk removing participants with <2 years of follow-up (N=111,724; n cases = 978).

Carbohydrates, sugars, & starch	**Q1**	**Q2**	**Q3**	**Q4**	**Per 5% energy increment**	**P-trend**
Total carbohydrates	1 (ref)	0.89 (0.75 - 1.06)	0.93 (0.78 - 1.12)	0.83 (0.67 - 1.01)	**0.94 (0.90 - 0.99)**	0.02
Total sugars	1 (ref)	0.84 (0.70 - 1.00)	**0.83 (0.69 - 0.99)**	**0.72 (0.59 - 0.89)**	**0.94 (0.88 - 0.99)**	0.02
Total free sugars	1 (ref)	1.08 (0.91 - 1.29)	0.94 (0.78 - 1.13)	0.87 (0.72 - 1.06)	0.96 (0.90 - 1.03)	0.28
Total non-free sugars	1 (ref)	0.87 (0.73 - 1.04)	**0.79 (0.66 - 0.95)**	**0.78 (0.64 - 0.95)**	**0.91 (0.85 - 0.98)**	0.008
Total starch	1 (ref)	0.90 (0.75 - 1.07)	0.96 (0.81 - 1.15)	0.94 (0.78 - 1.14)	0.99 (0.93 - 1.05)	0.64
Wholegrain starch	1 (ref)	0.90 (0.75 - 1.07)	0.91 (0.77 - 1.09)	0.84 (0.70 - 1.01)	0.94 (0.86 - 1.02)	0.12
Refined grain starch	1 (ref)	0.91 (0.76 - 1.09)	1.02 (0.85 - 1.21)	1.04 (0.86 - 1.25)	1.01 (0.96 - 1.07)	0.68
Fibre	**Q1**	**Q2**	**Q3**	**Q4**	**Per 5 g/day increment**	**P-trend**
Total fibre	1 (ref)	1.07 (0.89 - 1.29)	1.00 (0.82 - 1.21)	0.95 (0.77 - 1.17)	0.95 (0.89 - 1.02)	0.13
Total fibre from vegetables & fruit	1 (ref)	0.95 (0.79 - 1.14)	1.02 (0.85 - 1.22)	0.92 (0.76 - 1.11)	0.93 (0.84 - 1.02)	0.13
Total fibre from vegetables	1 (ref)	1.10 (0.91 - 1.31)	1.04 (0.87 - 1.25)	1.01 (0.84 - 1.22)	0.94 (0.82 - 1.09)	0.44
Total fibre from fruit	1 (ref)	0.94 (0.79 - 1.12)	0.86 (0.72 - 1.04)	0.83 (0.69 - 1.00)	0.88 (0.75 - 1.03)	0.10
Total fibre from wholegrains	1 (ref)	0.90 (0.75 - 1.08)	0.97 (0.81 - 1.15)	**0.78 (0.65 - 0.94)**	**0.89 (0.80 - 1.00)**	0.04

Models stratified for sex and age at recruitment, and further adjusted for region, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, BMI, processed and red meat intake, total vegetable fruit intake (except for total fibre, fibre from vegetables and/or fruit, or non-free sugar), and female specific covariates: menopausal hormone therapy use and menopausal status.

P-trend values represent the carbohydrate types/source modelled as 5% energy increase per day in the model and fibre by 5 g/day increase in the model.

Abbreviations: BMI, body mass index; g/day, grams per day; Q, quartile; ref, reference group.

Ancillary Table 8.1 Subgroup analyses for diet groups on risk of postmenopausal breast cancer in women.

	BMI							
	<27.5 kg/m^2			≥27.5 kg/m^2				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	**χ^2**	**P-value**
Regular meat-eaters	59198	1735	1 (ref)	47360	1680	1 (ref)	5.23	0.156
Low meat-eaters	75163	2198	0.96 (0.90 – 1.02)	43399	1583	1.00 (0.94 – 1.08)		
Fish-eaters	5201	148	1.03 (0.87 – 1.22)	1663	42	0.75 (0.55 – 1.01)		
Vegetarians	3560	79	0.85 (0.68 – 1.07)	1456	37	0.86 (0.61 – 1.20)		
	Smoking status							
	Never			**Ever**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	**χ^2**	**P-value**
Regular meat-eaters	63753	1909	1 (ref)	42559	1501	1 (ref)	1.32	0.725
Low meat-eaters	70440	2159	0.98 (0.93 – 1.05)	48173	1631	0.94 (0.88 – 1.01)		
Fish-eaters	3993	104	0.93 (0.77 – 1.14)	2879	86	0.90 (0.72 – 1.12)		
Vegetarians	3401	76	0.86 (0.68 – 1.09)	1625	39	0.75 (0.55 – 1.04)		

All models used age as the underlying time variable and are stratified, age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking status was the subgroup of interest), alcohol consumption, ethnicity, diabetes status, menopausal hormone therapy use, age at menarche, and age at first birth/parity. Full details for each covariate are provided in the statistical analysis section in the main text.

Models are restricted to only women.

X^2 and p-values from likelihood ratio tests for model fit comparing a model without an interaction term between subgroup and diet groups, to a model including an interaction between subgroup and diet groups.

Abbreviations: BMI: body mass index; CI: confidence intervals; HR: hazard ratios; ref: reference group.

Ancillary Table 8.2 Subgroup analyses across diet groups on risk of prostate cancer in men.

	BMI							
	<27.5 kg/m^2			≥27.5 kg/m^2				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	χ^2	**P-value**
Regular meat-eaters	63771	2750	1 (ref)	68053	2873	1 (ref)	4.02	0.260
Low meat-eaters	43570	2098	1.02 (0.96 - 1.08)	35201	1583	0.98 (0.92 - 1.05)		
Fish-eaters	2174	60	0.70 (0.54 - 0.91)	844	32	1.04 (0.73 - 1.48)		
Vegetarians	2087	49	0.72 (0.54 - 0.95)	858	15	0.62 (0.37 - 1.04)		
	Smoking status							
	Never			**Ever**				
	N	**Cases**	**HR (95% CI)**	**N**	**Cases**	**HR (95% CI)**	χ^2	**P-value**
Regular meat-eaters	63688	2593	1 (ref)	68038	3009	1 (ref)	0.99	0.804
Low meat-eaters	39608	1809	1.00 (0.95 - 1.07)	39172	1861	1.00 (0.94 - 1.06)		
Fish-eaters	1658	53	0.83 (0.63 - 1.09)	1360	40	0.79 (0.58 - 1.08)		
Vegetarians	1727	40	0.76 (0.56 - 1.05)	1217	24	0.61 (0.41 - 0.92)		

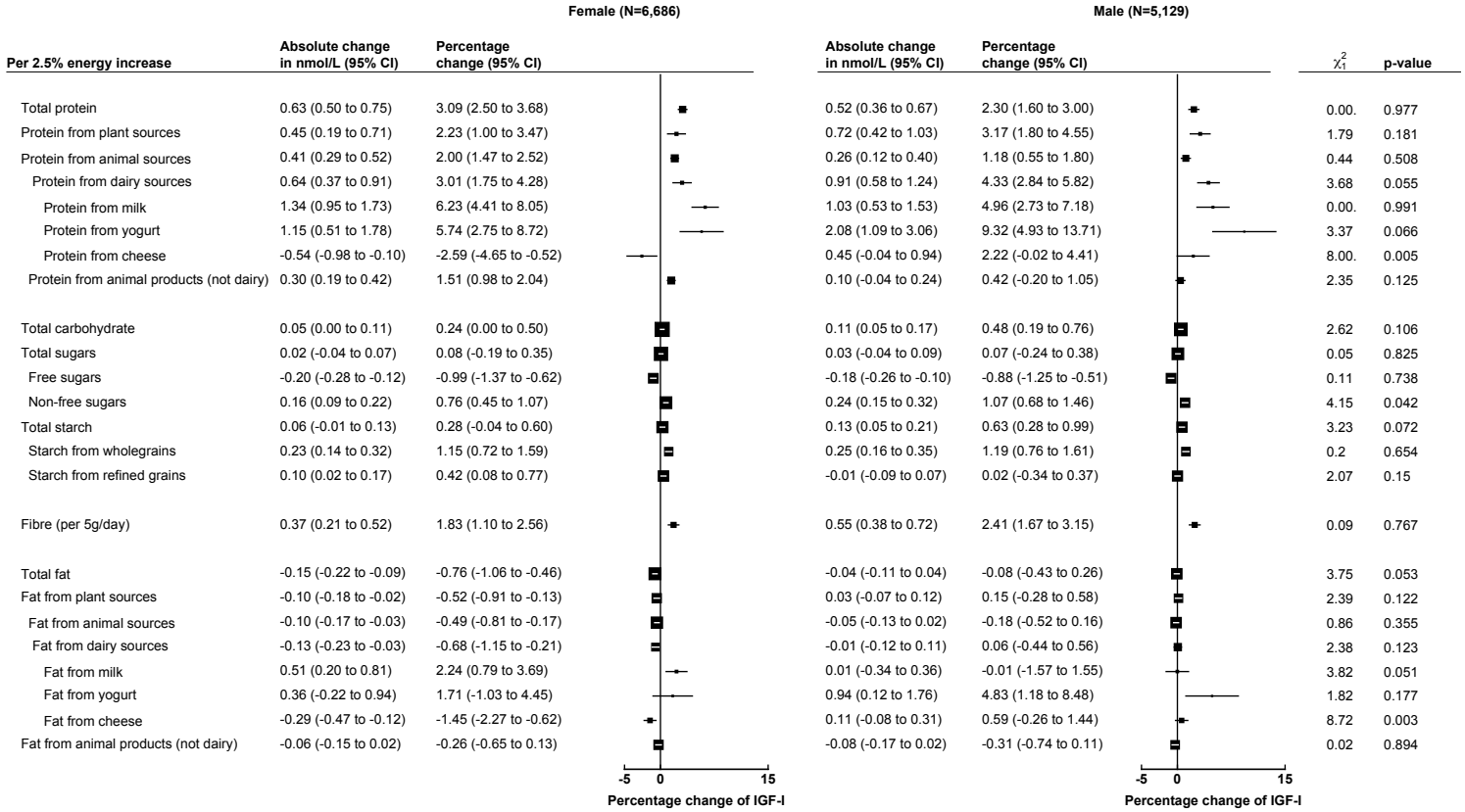
All models used age as the underlying time variable and are stratified by age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking status was the subgroup of interest), alcohol consumption, ethnicity, diabetes status, and marital status. Full details for each covariate are provided in the statistical analysis section in the main text.

Models are restricted to only men.

χ^2 and p-values from likelihood ratio tests for model fit comparing a model without an interaction term between subgroup and diet groups, to a model including an interaction between subgroup and diet groups.

Abbreviations: BMI: body mass index; CI: confidence intervals; HR: hazard ratios; ref: reference group.

Ancillary Figures

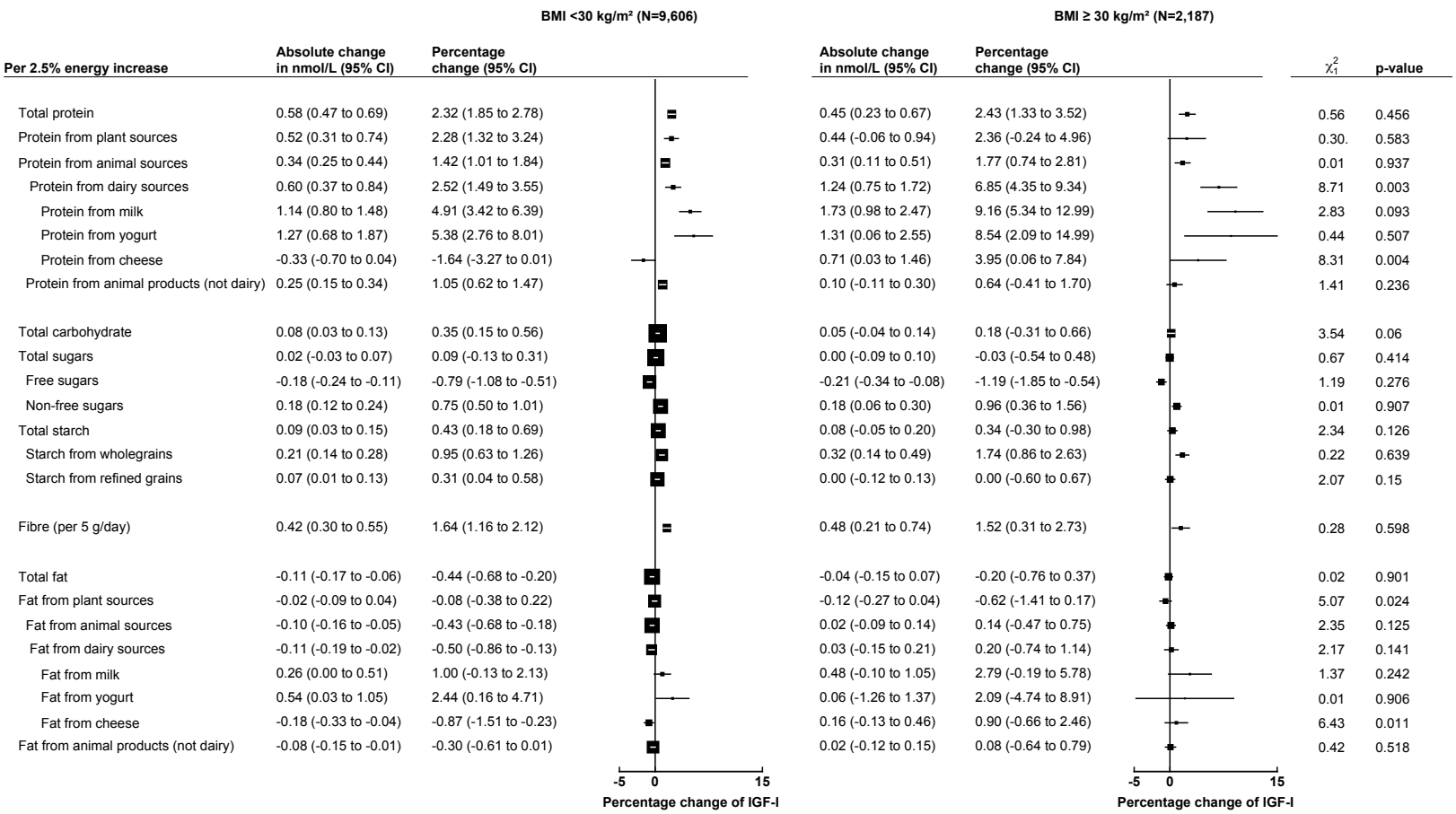


Ancillary Figure 5.1 Multivariable adjusted models for nutrient intake per 2.5% increase of energy from carbohydrates, fats, and proteins and 5 grams per day increase in fibre by sex in association with IGF-I concentrations.

All models are adjusted for age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, energy intake and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status.

χ^2 and p-values for heterogeneity represent values from likelihood ratio tests for adding an interaction term between sex and 2.5% energy increase for macronutrients or 5 g/day increase of fibre and testing for significant model fit. Bonferroni correction: p-values <0.00185 considered statistically significant.

Abbreviations: CI, confidence intervals; g/day, grams per day; IGF-I, insulin-like growth factor-I.

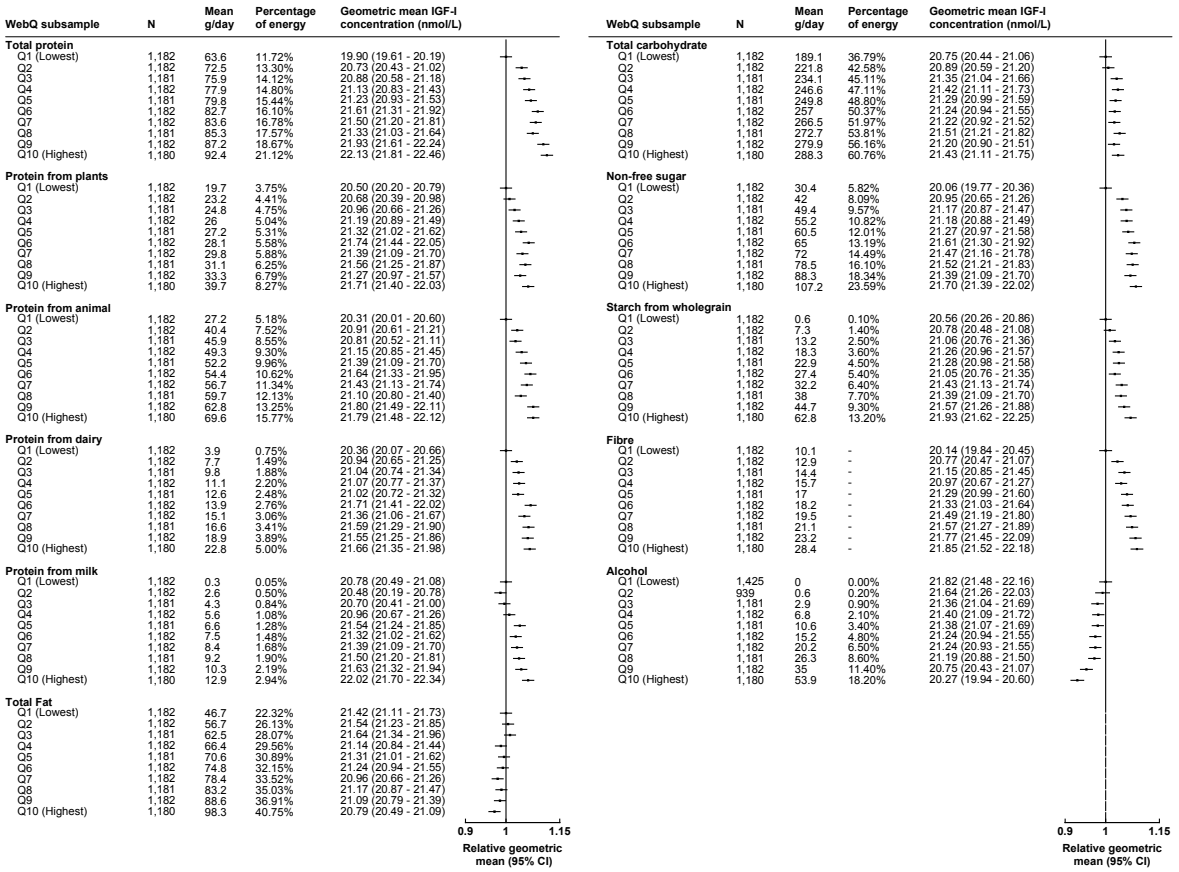


Ancillary Figure 5.2 Multivariable-adjusted model for nutrient intake per 2.5% increase of energy from carbohydrates, fats, and proteins and 5 grams per day increase in fibre by body mass index (<30 and ≥ 30 kg/m^2) in association with IGF-I.

All models are adjusted for sex, age at recruitment, region of recruitment, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption, ethnicity, diabetes, energy intake, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status. Participants with unknown BMI excluded from analyses.

χ^2 and p-values for heterogeneity represent values from likelihood ratio tests for adding an interaction term between BMI subgroups and 2.5% energy increase for macronutrients or 5g/day increase of fibre and testing for significant model fit. Bonferroni correction: p-values <0.00185 considered statistically significant.

Abbreviations: BMI, body mass index; CI, confidence intervals; g/day, grams per day; IGF-I, insulin-like growth factor-I.

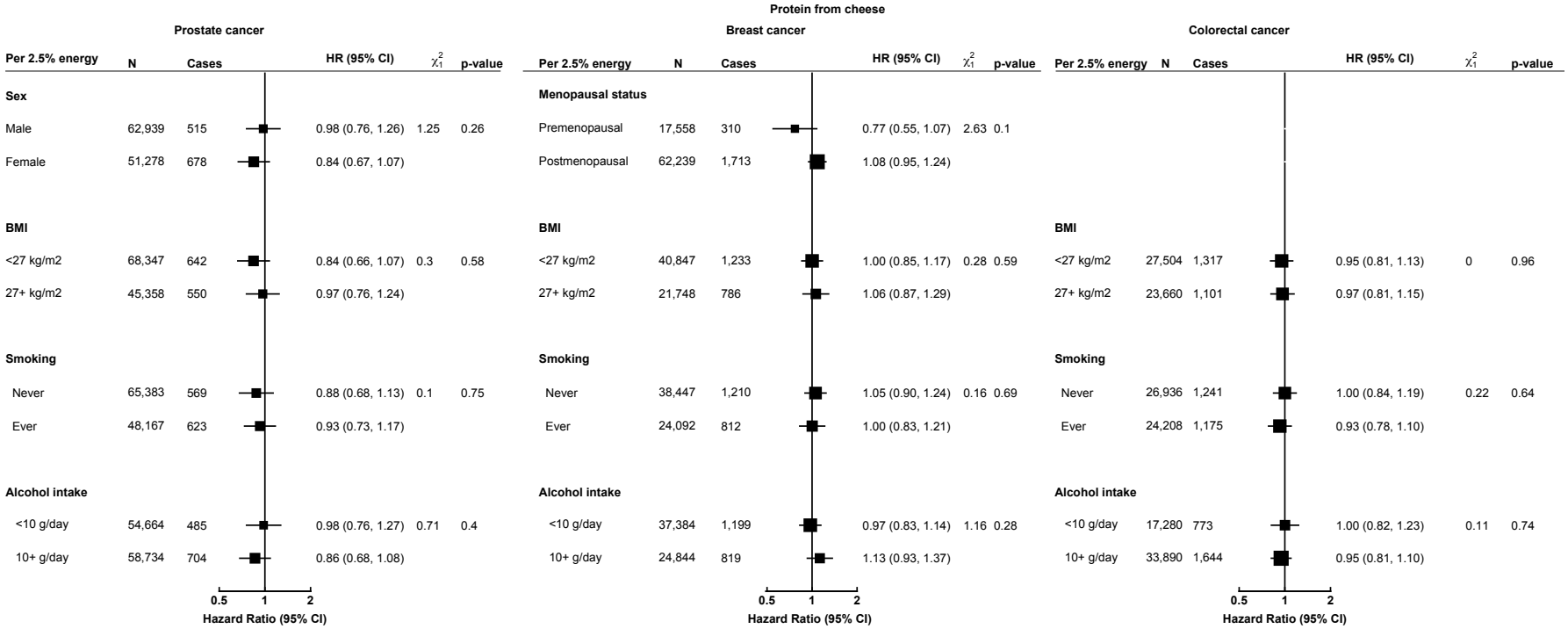


Ancillary Figure 5.3 Multivariable-adjusted model for percentage of energy in deciles from several macronutrients and fibre from the WebQ 24-hour dietary assessment subsample in association with the geometric mean concentration of IGF-I (N=11,815).

All models are adjusted for sex, age at recruitment, region of recruitment, body mass index, height, physical activity, Townsend deprivation index, education, smoking, alcohol consumption (except when alcohol was the exposure), ethnicity, diabetes, energy intake, and women specific covariates: hormone replacement therapy use, oral contraceptive use, and menopausal status. Quantiles of percentage of energy from carbohydrate sources, percentage of energy from fat sources, percentage of protein sources, and percentage of energy from alcohol, and fibre quantiles calculated from a minimum of four (maximum of five) averaged 24-hour web-based dietary assessments with one assessment completed at recruitment.

Mean grams per day calculated as average gram intake per day within each category. Percentage of energy calculated by mean percentage of energy per day in each quintile.

Abbreviations: g/day, grams per day; IGF-I, insulin-like growth factor-I; N, number of participants; Q, quantile.

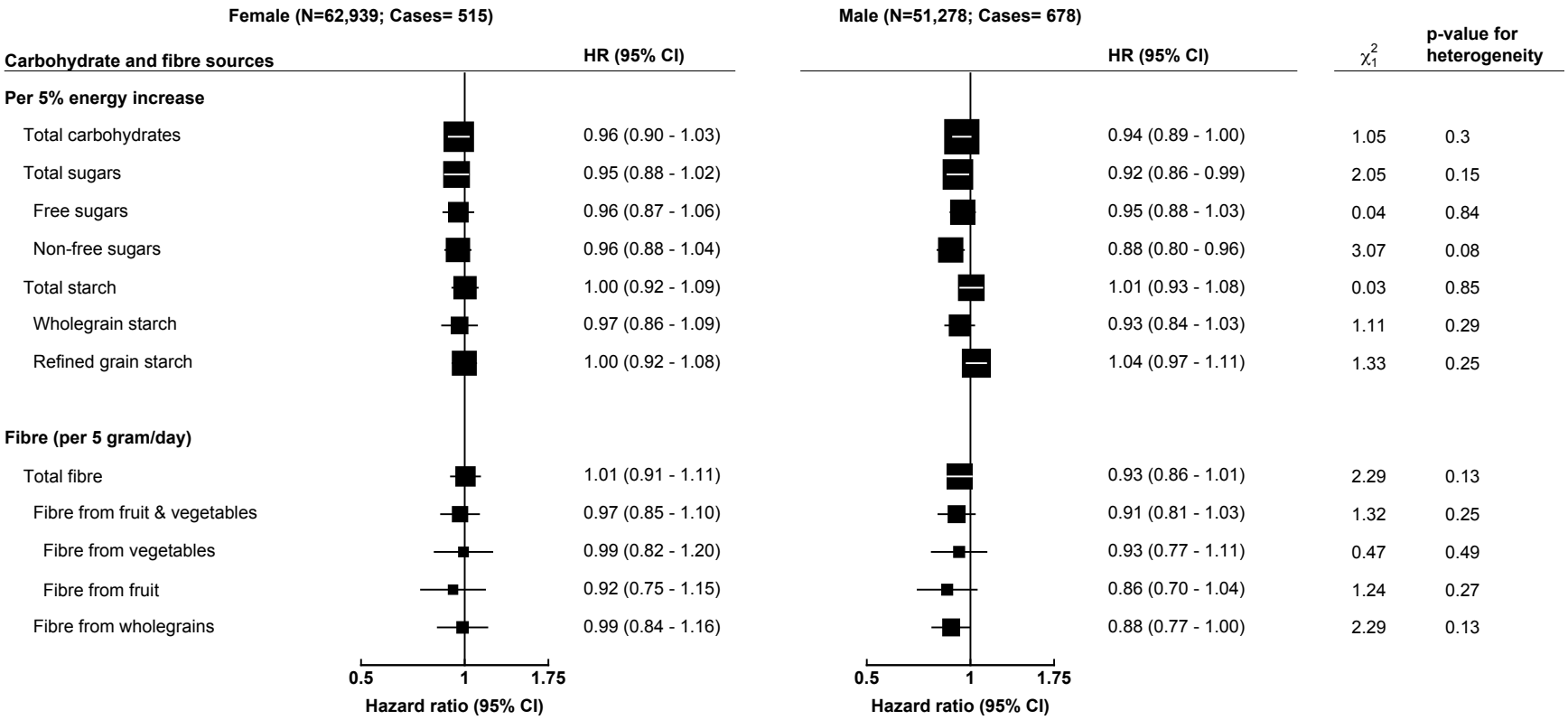


Ancillary Figure 6.1 Intake of cheese protein per 2.5% energy increase by subgroups risk of colorectal cancer, breast cancer, prostate cancer.

All models used age as the underlying time variable, were stratified by age groups at recruitment, and further adjusted for height, physical activity, Townsend deprivation index, education, employment status, smoking status (except when smoking status was subgroup if interest), alcohol intake (except when alcohol was the subgroup of interest), ethnicity, diagnosis of diabetes, BMI (except when BMI was subgroup of interest), and energy intake. For colorectal cancer analyses: all models were stratified by sex (except when sex was subgroup of interest) and adjusted for menopause status (women only) and menopausal hormone therapy use (women only), red and processed meat intake, non-steroidal anti-inflammatory drug use. For breast cancer analyses: models were further adjusted for menopausal hormone therapy use, oral contraceptive use, parity and age at first birth, age at menarche, BMI and menopause status (except when menopause status was subgroup of interest). For prostate cancer analyses: models were further adjusted for marital status.

χ^2 represents improvement of model fit and p-value represents heterogeneity between subgroups obtained from likelihood ratio tests for including an interaction term between subgroup of interest and cheese protein (modelled as a 2.5% energy increase) into the model.

Abbreviations: BMI, body mass index; CI, confidence interval; g/day, grams per day; HR, hazard ratio; kg/m^2, kilograms per metre squared; N, number of participants.

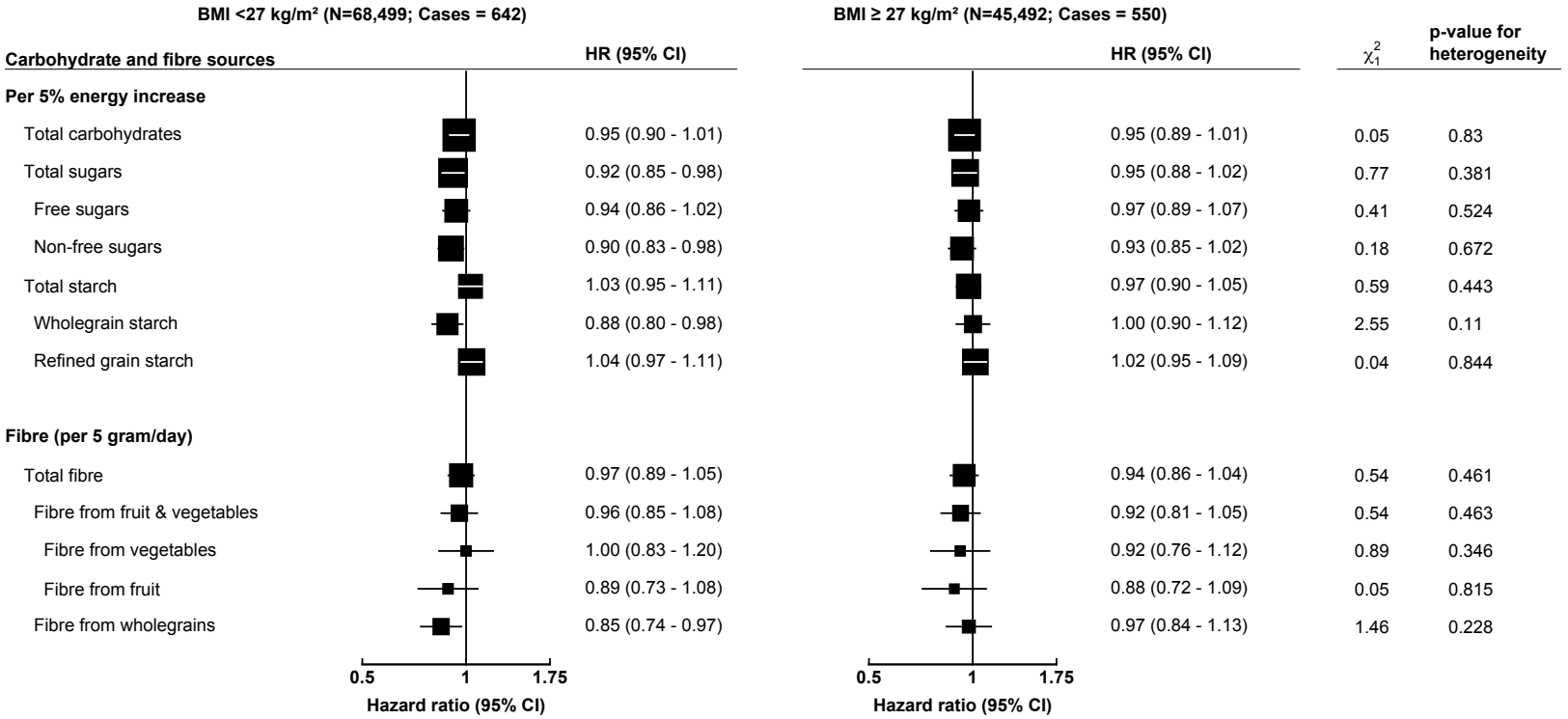


Ancillary Figure 7.1 Multivariable hazard ratios and 95% confidence intervals for carbohydrate and fibre sources with colorectal cancer risk stratified by sex.

Models stratified for age at recruitment, and further adjusted for region, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, BMI, processed and red meat intake, total vegetable fruit intake (except for total fibre, fibre from vegetables and/or fruit, or non-free sugars), and female specific covariates: menopausal hormone therapy use and menopausal status.

χ^2 and p-value for heterogeneity represents improvement of fit obtained from likelihood ratio tests for including an interaction term between sex and carbohydrate type/source (modelled as a 5% energy increase) or fibre source (modelled as a 5 gram/day increase) into the model.

Abbreviations: CI, confidence interval; HR, hazard ratio, N, number of participants.

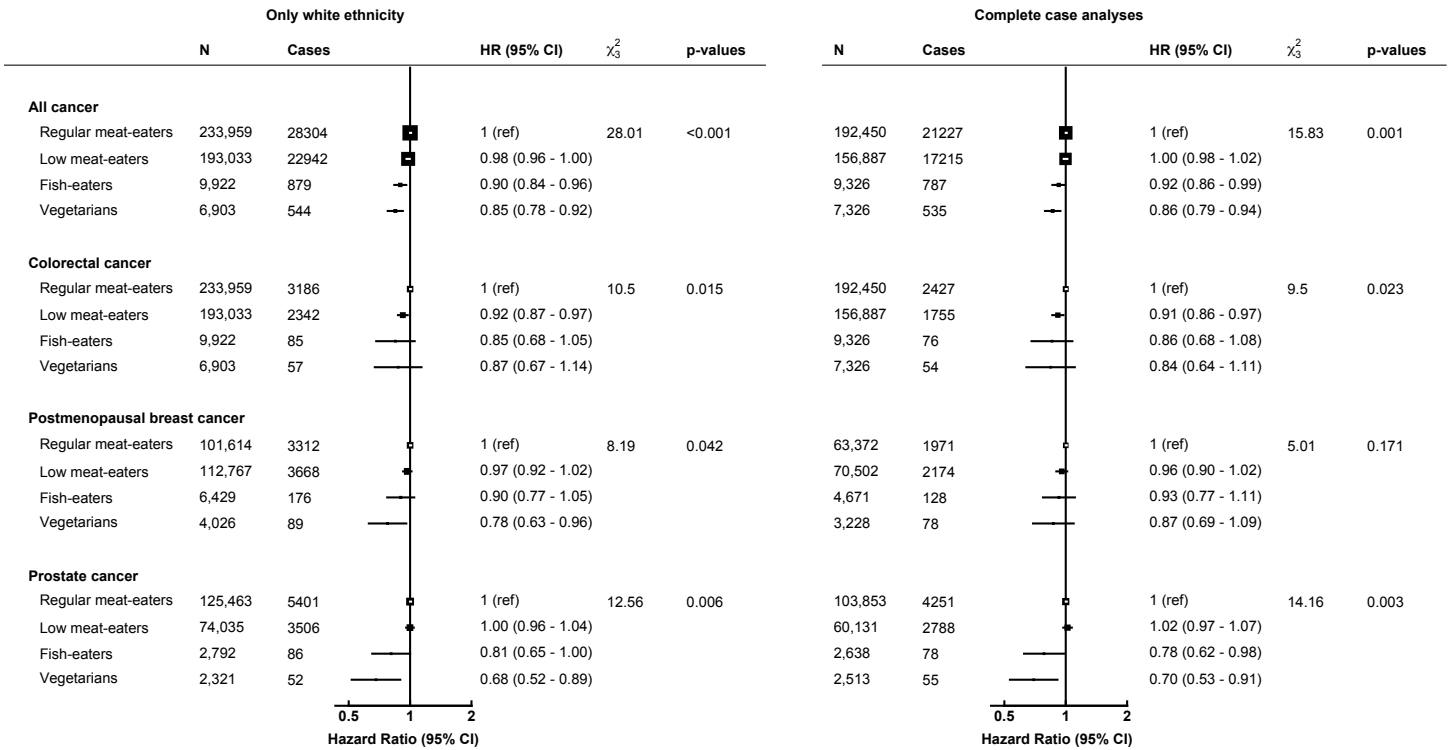


Ancillary Figure 7.2 Multivariable hazard ratios and 95% confidence intervals for carbohydrate and fibre sources with colorectal cancer risk stratified by body mass index (<27 kg/m^2 vs. ≥ 27 kg/m^2).

Models stratified for sex and age at recruitment, and further adjusted for region, height, physical activity, Townsend deprivation index, education, employment, smoking, alcohol consumption measured at recruitment, ethnicity, diabetes status, non-steroidal anti-inflammatory drug use, energy intake, processed and red meat intake, total vegetable fruit intake (except for total fibre, fibre from vegetables and/or fruit, or non-free sugars), and female specific covariates: menopausal hormone therapy use and menopausal status.

χ^2 and p-value for heterogeneity represent improvement of fit obtained from likelihood ratio tests for including an interaction term between BMI subgroups and carbohydrate type/source (modelled as a 5% energy increase) or fibre source (modelled as a 5 gram/day increase) into the model.

Abbreviations: CI, confidence interval; HR, hazard ratio, N, number of participants.



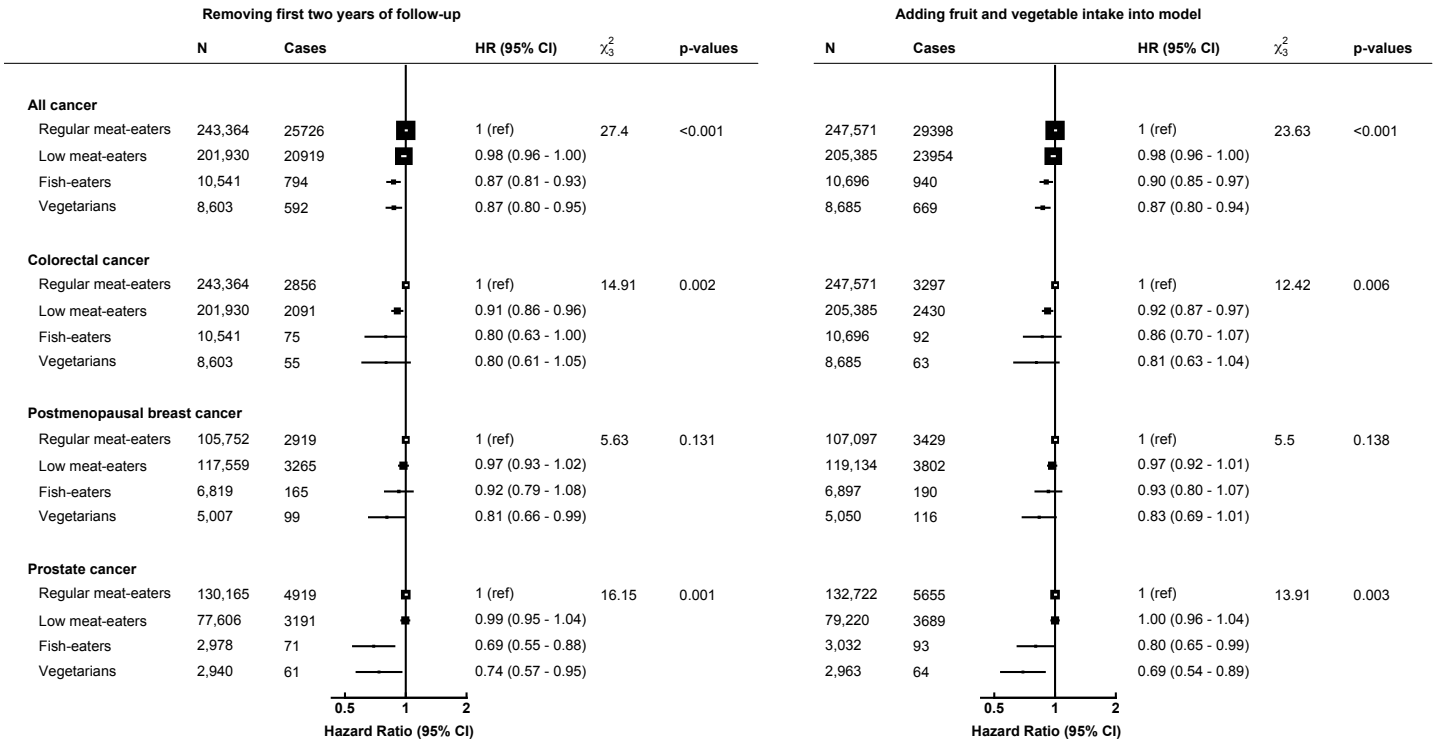
Ancillary Figure 8.1 Hazard ratios and 95% confidence intervals for sensitivity analyses including only participants of white ethnicity and complete cases analyses on associations between diet groups and risk of all cancer, prostate, postmenopausal breast, or colorectal cancer.

All models used age as the underlying time variable and are stratified by sex (for only all cancer and colorectal cancer) and age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, and diabetes status. Full details for each covariate are provided in the statistical analysis section in the main text.

For all cancer and colorectal cancer models are further adjusted for: menopausal hormone therapy use, and menopausal status. For colorectal cancer, models were also adjusted for NSAID use. For prostate cancer models are further adjusted for: marital status. For postmenopausal breast cancer models are further adjusted for: menopausal hormone therapy use, age at menarche, and age at first birth/parity. Models are restricted to only men for prostate cancer and only women for postmenopausal breast cancer.

χ^2 (degrees of freedom in subscript) and p-values from likelihood ratio tests for model fit comparing a model without diet groups, to a model including diet groups.

Abbreviations: CI, confidence intervals; HR, hazard ratios; N, number of participants; NSAID, non-steroid anti-inflammatory drug; ref, reference group.



Ancillary Figure 8.2 Hazard ratios and 95% confidence intervals for sensitivity analyses removing first two years of follow-up and adjusting for fruit and vegetable intake on associations between diet groups and risk of all cancer, prostate, postmenopausal breast, or colorectal cancer.

All models used age as the underlying time variable and are stratified by sex (for only all cancer and colorectal cancer), age groups, and adjusted for region of recruitment, height, physical activity, Townsend deprivation index, education, employment status, smoking status, alcohol consumption, ethnicity, and diabetes status. Full details for each covariate are provided in the statistical analysis section in the main text.

For all cancer and colorectal cancer models are further adjusted for: menopausal hormone therapy use, and menopausal status. For colorectal cancer, models were also adjusted for NSAID use. For prostate cancer models are further adjusted for: marital status. For postmenopausal breast cancer model are further adjusted for: menopausal hormone therapy (MHT) use, age at menarche, and age at first birth/parity. Models are restricted to only men for prostate cancer and only women for postmenopausal breast cancer.

χ^2 (degrees of freedom in subscript) and p-values from likelihood ratio tests for model fit comparing a model without diet groups, to a model including diet groups.

Abbreviations: CI, confidence intervals; HR, hazard ratios; N, number of participants; NSAID, non-steroid anti-inflammatory drug; ref, reference group.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71: 209-49. https://doi.org/https://doi.org/10.3322/caac.21660
2. World Health Organization (WHO). Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. 2020. https://who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death. Accessed June 20 2022.
3. Holmes MD, Pollak MN, Willett WC, Hankinson SE. Dietary correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev. 2002;11: 852-61.
4. Giovannucci E, Pollak M, Liu Y, Platz EA, Majeed N, Rimm EB, et al. Nutritional predictors of insulin-like growth factor I and their relationships to cancer in men. Cancer Epidemiol Biomarkers Prev. 2003;12: 84-9.
5. Crowe FL, Key TJ, Allen NE, Appleby PN, Roddam A, Overvad K, et al. The association between diet and serum concentrations of IGF-I, IGFBP-1, IGFBP-2, and IGFBP-3 in the European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev. 2009;18: 1333-40. https://doi.org/10.1158/1055-9965.EPI-08-0781
6. Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. Nat Rev Cancer. 2008;8: 915-28. https://doi.org/10.1038/nrc2536
7. Pollak MN, Schernhammer ES, Hankinson SE. Insulin-like growth factors and neoplasia. Nat Rev Cancer. 2004;4: 505-18. https://doi.org/10.1038/nrc1387
8. Pollak M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. Nat Rev Cancer. 2012;12: 159-69. https://doi.org/10.1038/nrc3215
9. Rinaldi S, Cleveland R, Norat T, Biessy C, Rohrmann S, Linseisen J, et al. Serum levels of IGF-I, IGFBP-3 and colorectal cancer risk: results from the EPIC cohort, plus a meta-analysis of prospective studies. Int J Cancer. 2010;126: 1702-15. https://doi.org/10.1002/ijc.24927
10. Endogenous Hormones and Breast Cancer Collaborative Group. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. Lancet Oncol. 2010;11: 530-42. https://doi.org/10.1016/S1470-2045(10)70095-4
11. Travis RC, Appleby PN, Martin RM, Holly JMP, Albanes D, Black A, et al. A meta-analysis of individual participant data reveals an association between circulating levels of IGF-I and prostate cancer risk. Cancer Res. 2016;76: 2288-300. https://doi.org/10.1158/0008-5472.CAN-15-1551
12. Murphy N, Carreras-Torres R, Song M, Chan AT, Martin RM, Papadimitriou N, et al. Circulating levels of insulin-like growth factor 1 and insulin-like growth factor binding protein 3

associate with risk of colorectal cancer based on serologic and mendelian randomization analyses. Gastroenterology. 2020;158: 1300-12.e20. https://doi.org/10.1053/j.gastro.2019.12.020

13. Murphy N, Knuppel A, Papadimitriou N, Martin RM, Tsilidis KK, Smith-Byrne K, et al. Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and Mendelian randomization analyses with ~430 000 women. Ann Oncol. 2020;31: 641-9. https://doi.org/10.1016/j.annonc.2020.01.066

14. Watts EL, Fensom GK, Smith Byrne K, Perez-Cornago A, Allen NE, Knuppel A, et al. Circulating insulin-like growth factor-I, total and free testosterone concentrations and prostate cancer risk in 200 ,000 men in UK Biobank. Int J Cancer. 2021;148: 2274-88. https://doi.org/https://doi.org/10.1002/ijc.33416

15. World Health Organization. Cancer. https://www.who.int/cancer/prevention/en/. Accessed 26 May 2022.

16. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144: 646-74. https://doi.org/10.1016/j.cell.2011.02.013

17. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015;347: 78-81. https://doi.org/10.1126/science.1260825

18. WHO Classification of Tumours Editorial Board. Digestive System Tumours: Lyon (France): International Agency for Research on Cancer; 2019. (WHO classification of tumours series, 5th ed.; vol. 1).

19. Lee YC, Lee YL, Chuang JP, Lee JC. Differences in survival between colon and rectal cancer from SEER data. PLoS One. 2013;8: e78709. https://doi.org/10.1371/journal.pone.0078709

20. World Cancer Research Fund/ American Institute for Cancer Research. Continuous Update Project Expert Report. Diet, nutrition, physical activity and colorectal cancer. 2018.

21. Murphy N, Ward HA, Jenab M, Rothwell JA, Boutron-Ruault M-C, Carbonnel F, et al. Heterogeneity of colorectal cancer risk factors by anatomical subsite in 10 European countries: a multinational cohort study. Clin Gastroenterol Hepatol. 2019;17: 1323-31.e6. https://doi.org/10.1016/j.cgh.2018.07.030

22. Wang L, Lo C-H, He X, Hang D, Wang M, Wu K, et al. Risk factor profiles differ for cancers of different regions of the colorectum. Gastroenterology. 2020;159: 241-56.e13. https://doi.org/10.1053/j.gastro.2020.03.054

23. Sharma R, Abbasi-Kangevari M, Abd-Rabu R, Abidi H, Abu-Gharbieh E, Acuna JM, et al. Global, regional, and national burden of colorectal cancer and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Gastroenterol Hepatol. 2022;7: 627-47. https://doi.org/10.1016/S2468-1253(22)00044-9

24. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F,. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. 2020. https://gco.iarc.fr/today. Accessed 20 June 2022.

25. Fidler MM, Soerjomataram I, Bray F. A global view on cancer incidence and national levels of the human development index. Int J Cancer. 2016;139: 2436-46. https://doi.org/10.1002/ijc.30382

26. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72: 7-33. https://doi.org/https://doi.org/10.3322/caac.21708

27. Cardoso R, Guo F, Heisser T, Hackl M, Ihle P, De Schutter H, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. Lancet Oncol. 2021;22: 1002-13. https://doi.org/https://doi.org/10.1016/S1470-2045(21)00199-6

28. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66: 683-91. https://doi.org/10.1136/gutjnl-2015-310912

29. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020;70: 145-64. https://doi.org/10.3322/caac.21601

30. Vuik FE, Nieuwenburg SA, Bardou M, Lansdorp-Vogelaar I, Dinis-Ribeiro M, Bento MJ, et al. Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. Gut. 2019;68: 1820-6. https://doi.org/10.1136/gutjnl-2018-317592

31. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, et al. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. J Natl Cancer Inst. 2017;109. https://doi.org/10.1093/jnci/djw322

32. Araghi M, Soerjomataram I, Bardot A, Ferlay J, Cabasag CJ, Morrison DS, et al. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. Lancet Gastroenterol Hepatol. 2019;4: 511-8. https://doi.org/10.1016/S2468-1253(19)30147-5

33. O'Sullivan DE, Sutherland RL, Town S, Chow K, Fan J, Forbes N, et al. Risk factors for early-onset colorectal cancer: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2022;20: 1229-40.e5. https://doi.org/10.1016/j.cgh.2021.01.037

34. Brenner DR, Heer E, Sutherland RL, Ruan Y, Tinmouth J, Heitman SJ, et al. National trends in colorectal cancer incidence among older and younger adults in Canada. JAMA Netw Open. 2019;2: e198090. https://doi.org/10.1001/jamanetworkopen.2019.8090

35. Eng C, Hochster H. Early-onset colorectal cancer: the mystery remains. J Natl Cancer Inst. 2021;113: 1608-10. https://doi.org/10.1093/jnci/djab127

36. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin. 2018;68: 250-81. https://doi.org/10.3322/caac.21457

37. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343: 78-85. https://doi.org/10.1056/nejm200007133430201

38. Kachuri L, Graff RE, Smith-Byrne K, Meyers TJ, Rashkin SR, Ziv E, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. Nature Commun. 2020;11: 6084. https://doi.org/10.1038/s41467-020-19600-4

39. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. Int J Cancer. 2002;99: 260-6. https://doi.org/10.1002/ijc.10332

40. Lynch HT, de la Chapelle A. Hereditary Colorectal Cancer. N Engl J Med. 2003;348: 919-32. https://doi.org/10.1056/NEJMra012242

41. Burt RW. Colon cancer screening. Gastroenterology. 2000;119: 837-53. https://doi.org/10.1053/gast.2000.16508

42. Lubbe SJ, Di Bernardo MC, Chandler IP, Houlston RS. Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J Clin Oncol. 2009;27: 3975-80. https://doi.org/10.1200/jco.2008.21.6853

43. Bonadona V, Bonaïti B, Olschwang S, Grandjouan S, Huiart L, Longy M, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA. 2011;305: 2304-10. https://doi.org/10.1001/jama.2011.743

44. Møller P, Seppälä T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. Gut. 2017;66: 464-72. https://doi.org/10.1136/gutjnl-2015-309675

45. Hampel H, de la Chapelle A. How do we approach the goal of identifying everybody with Lynch Syndrome? Fam Cancer. 2013;12: 313-7. https://doi.org/10.1007/s10689-013-9611-5

46. Win AK, Jenkins MA, Dowty JG, Antoniou AC, Lee A, Giles GG, et al. Prevalence and penetrance of major genes and polygenes for colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2017;26: 404-12. https://doi.org/10.1158/1055-9965.Epi-16-0693

47. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet. 2009;76: 1-18. https://doi.org/10.1111/j.1399-0004.2009.01230.x

48. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51: 76-87. https://doi.org/10.1038/s41588-018-0286-6

49. Schmit SL, Edlund CK, Schumacher FR, Gong J, Harrison TA, Huyghe JR, et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst. 2019;111: 146-57. https://doi.org/10.1093/jnci/djy099

50. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nature Commun. 2015;6: 7138. https://doi.org/10.1038/ncomms8138

51. Jiao S, Peters U, Berndt S, Brenner H, Butterbach K, Caan BJ, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet. 2014;23: 3898-905. https://doi.org/10.1093/hmg/ddu087

52. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol. 2001;96: 2992-3003. https://doi.org/10.1111/j.1572-0241.2001.04677.x

53. Hemminki K, Chen B. Familial risk for colorectal cancers are mainly due to heritable causes. Cancer Epidemiol Biomarkers Prev. 2004;13: 1253-6. https://doi.org/10.1158/1055-9965.1253.13.7

54. Graff RE, Möller S, Passarelli MN, Witte JS, Skytthe A, Christensen K, et al. Familial risk and heritability of colorectal cancer in the Nordic twin study of cancer. Clin Gastroenterol Hepatol. 2017;15: 1256-64. https://doi.org/https://doi.org/10.1016/j.cgh.2016.12.041

55. Wan Q, Zhao R, Xia L, Wu Y, Zhou Y, Wang Y, et al. Inflammatory bowel disease and risk of gastric, small bowel and colorectal cancer: a meta-analysis of 26 observational studies. J Cancer Res Clin Oncol. 2021;147: 1077-87. https://doi.org/10.1007/s00432-020-03496-0

56. Jess T, Rungoe C, Peyrin–Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clin Gastroenterol Hepatol. 2012;10: 639-45. https://doi.org/https://doi.org/10.1016/j.cgh.2012.01.010

57. Jess T, Gamborg M, Matzen P, Munkholm P, Sørensen TI. Increased risk of intestinal cancer in Crohn's disease: a meta-analysis of population-based cohort studies. Am J Gastroenterol. 2005;100: 2724-9. https://doi.org/10.1111/j.1572-0241.2005.00287.x

58. Sugiyama H, Misumi M, Brenner A, Grant EJ, Sakata R, Sadakane A, et al. Radiation risk of incident colorectal cancer by anatomical site among atomic bomb survivors: 1958-2009. Int J Cancer. 2020;146: 635-45. https://doi.org/10.1002/ijc.32275

59. Kamran SC, Berrington de Gonzalez A, Ng A, Haas-Kogan D, Viswanathan AN. Therapeutic radiation and the potential risk of second malignancies. Cancer. 2016;122: 1809-21. https://doi.org/https://doi.org/10.1002/cncr.29841

60. Wallis CJ, Mahar AL, Choo R, Herschorn S, Kodama RT, Shah PS, et al. Second malignancies after radiotherapy for prostate cancer: systematic review and meta-analysis. BMJ. 2016;352: i851. https://doi.org/10.1136/bmj.i851

61. Richardson DB, Cardis E, Daniels RD, Gillies M, Haylock R, Leuraud K, et al. Site-specific solid cancer mortality after exposure to ionizing radiation: a cohort study of workers (INWORKS). Epidemiology. 2018;29: 31-40. https://doi.org/10.1097/ede.0000000000000761

62. Cogliano VJ, Baan R, Straif K, Grosse Y, Lauby-Secretan B, El Ghissassi F, et al. Preventable exposures associated with human cancers. J Natl Cancer Inst. 2011;103: 1827-39. https://doi.org/10.1093/jnci/djr483

63. Botteri E, Borroni E, Sloan EK, Bagnardi V, Bosetti C, Peveri G, et al. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. Am J Gastroenterol. 2020;115: 1940-9. https://doi.org/10.14309/ajg.0000000000000803

64. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA. 2008;300: 2765-78. https://doi.org/10.1001/jama.2008.839

65. Dimou N, Yarmolinsky J, Bouras E, Tsilidis KK, Martin RM, Lewis SJ, et al. Causal effects of lifetime smoking on breast and colorectal cancer risk: mendelian randomization study. Cancer Epidemiol Biomarkers Prev. 2021;30: 953-64. https://doi.org/10.1158/1055-9965.Epi-20-1218

66. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21: 7435-51. https://doi.org/10.1038/sj.onc.1205803

67. Ruder EH, Laiyemo AO, Graubard BI, Hollenbeck AR, Schatzkin A, Cross AJ. Non-steroidal anti-inflammatory drugs and colorectal cancer risk in a large, prospective cohort. Am J Gastroenterol. 2011;106: 1340-50. https://doi.org/10.1038/ajg.2011.38

68. Rothwell PM, Wilson M, Elwin CE, Norrving B, Algra A, Warlow CP, et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet. 2010;376: 1741-50. https://doi.org/10.1016/s0140-6736(10)61543-7

69. Baron JA, Cole BF, Sandler RS, Haile RW, Ahnen D, Bresalier R, et al. A randomized trial of aspirin to prevent colorectal adenomas. N Engl J Med. 2003;348: 891-9. https://doi.org/10.1056/NEJMoa021735

70. Flossmann E, Rothwell PM. Effect of aspirin on long-term risk of colorectal cancer: consistent evidence from randomised and observational studies. Lancet. 2007;369: 1603-13. https://doi.org/https://doi.org/10.1016/S0140-6736(07)60747-8

71. Arber N, Eagle CJ, Spicak J, Rácz I, Dite P, Hajer J, et al. Celecoxib for the Prevention of Colorectal Adenomatous Polyps. N Engl J Med. 2006;355: 885-95. https://doi.org/10.1056/NEJMoa061652

72. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ. 2002;324: 71-86. https://doi.org/10.1136/bmj.324.7329.71

73. Pearson-Stuttard J, Papadimitriou N, Markozannes G, Cividini S, Kakourou A, Gill D, et al. Type 2 diabetes and cancer: an umbrella review of observational and Mendelian randomization studies. Cancer Epidemiol Biomarkers Prev. 2021;30: 1218-28. https://doi.org/10.1158/1055-9965.epi-20-1245

74. Murphy N, Song M, Papadimitriou N, Carreras-Torres R, Langenberg C, Martin RM, et al. Associations between glycemic traits and colorectal cancer: a Mendelian randomization analysis. J Natl Cancer Inst. 2022;114: 740-52. https://doi.org/10.1093/jnci/djac011

75. Yuan S, Kar S, Carter P, Vithayathil M, Mason AM, Burgess S, et al. Is type 2 diabetes causally associated with cancer risk? Evidence from a two-sample Mendelian randomization study. Diabetes. 2020;69: 1588-96. https://doi.org/10.2337/db20-0084

76. Goto A, Yamaji T, Sawada N, Momozawa Y, Kamatani Y, Kubo M, et al. Diabetes and cancer risk: A Mendelian randomization study. Int J Cancer. 2020;146: 712-9. https://doi.org/10.1002/ijc.32310

77. Giovannucci E. Insulin and colon cancer. Cancer Causes Control. 1995;6: 164-79.
78. Gallagher EJ, LeRoith D. Hyperinsulinaemia in cancer. Nat Rev Cancer. 2020;20: 629-44. https://doi.org/10.1038/s41568-020-0295-5
79. Hu Y, Zhang X, Ma Y, Yuan C, Wang M, Wu K, et al. Incident type 2 diabetes duration and cancer risk: A prospective study in two US cohorts. J Natl Cancer Inst. 2021;113: 381-9. https://doi.org/10.1093/jnci/djaa141
80. Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. Gut. 2013;62: 933-47. https://doi.org/10.1136/gutjnl-2013-304701
81. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer-viewpoint of the IARC working group. N Engl J Med. 2016;375: 794-8. https://doi.org/10.1056/NEJMsr1606602
82. Bull CJ, Bell JA, Murphy N, Sanderson E, Davey Smith G, Timpson NJ, et al. Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study. BMC Med. 2020;18: 396. https://doi.org/10.1186/s12916-020-01855-9
83. Jarvis D, Mitchell JS, Law PJ, Palin K, Tuupanen S, Gylfe A, et al. Mendelian randomisation analysis strongly implicates adiposity with risk of developing colorectal cancer. Br J Cancer. 2016;115: 266-72. https://doi.org/10.1038/bjc.2016.188
84. Yoon YS, Keum N, Zhang X, Cho E, Giovannucci EL. Hyperinsulinemia, insulin resistance and colorectal adenomas: A meta-analysis. Metabolism. 2015;64: 1324-33. https://doi.org/https://doi.org/10.1016/j.metabol.2015.06.013
85. Petimar J, Tabung FK, Valeri L, Rosner B, Chan AT, Smith-Warner SA, et al. Mediation of associations between adiposity and colorectal cancer risk by inflammatory and metabolic biomarkers. Int J Cancer. 2019;144: 2945-53. https://doi.org/10.1002/ijc.32047
86. Murphy N, Jenab M, Gunter MJ. Adiposity and gastrointestinal cancers: epidemiology, mechanisms and future directions. Nat Rev Gastroenterol Hepatol. 2018;15: 659-70. https://doi.org/10.1038/s41575-018-0038-1
87. Dashti SG, Viallon V, Simpson JA, Karahalios A, Moreno-Betancur M, English DR, et al. Explaining the link between adiposity and colorectal cancer risk in men and postmenopausal women in the UK Biobank: A sequential causal mediation analysis. Int J Cancer. 2020;147: 1881-94. https://doi.org/https://doi.org/10.1002/ijc.32980
88. Boyle T, Keegel T, Bull F, Heyworth J, Fritschi L. Physical Activity and Risks of Proximal and Distal Colon Cancers: A Systematic Review and Meta-analysis. J Natl Cancer Inst. 2012;104: 1548-61. https://doi.org/10.1093/jnci/djs354
89. Papadimitriou N, Dimou N, Tsilidis KK, Banbury B, Martin RM, Lewis SJ, et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. Nature Commun. 2020;11: 597. https://doi.org/10.1038/s41467-020-14389-8
90. Mahmood S, MacInnis RJ, English DR, Karahalios A, Lynch BM. Domain-specific physical activity and sedentary behaviour in relation to colon and rectal cancer risk: a systematic review and meta-analysis. Int J Epidemiol. 2017;46: 1797-813. https://doi.org/10.1093/ije/dyx137

91. McTiernan A. Mechanisms linking physical activity with cancer. Nat Rev Cancer. 2008;8: 205-11. https://doi.org/10.1038/nrc2325

92. Song M, Chan AT. Environmental factors, gut microbiota, and colorectal cancer prevention. Clin Gastroenterol Hepatol. 2019;17: 275-89. https://doi.org/10.1016/j.cgh.2018.07.012

93. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer. 1975;15: 617-31. https://doi.org/10.1002/ijc.2910150411

94. McCredie M. Cancer epidemiology in migrant populations. Recent Results Cancer Res. 1998;154: 298-305. https://doi.org/10.1007/978-3-642-46870-4_21

95. Shimizu H, Mack TM, Ross RK, Henderson BE. Cancer of the gastrointestinal tract among japanese and white immigrants in Los Angeles county. J Natl Cancer Inst. 1987;78: 223-8. https://doi.org/10.1093/jnci/78.2.223

96. Tominaga S, editor Cancer incidence in japanese in Japan, Hawaii, and western United States. Fourth Symposium on Epidemiology and Cancer Registries in the Pacific Basin: Proceedings of a Symposium Held in Kona, Hawaii, January 16-20, 1984; 1985: US Department of Health and Human Services, Public Health Service, National.

97. Rumgay H, Shield K, Charvat H, Ferrari P, Sornpaisarn B, Obot I, et al. Global burden of cancer in 2020 attributable to alcohol consumption: a population-based study. Lancet Oncol. 2021;22: 1071-80. https://doi.org/10.1016/S1470-2045(21)00279-5

98. Zhou X, Wang L, Xiao J, Sun J, Yu L, Zhang H, et al. Alcohol consumption, DNA methylation and colorectal cancer risk: Results from pooled cohort studies and Mendelian randomization analysis. Int J Cancer. 2022;151: 83-94. https://doi.org/10.1002/ijc.33945

99. Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. Nat Rev Cancer. 2007;7: 599-612. https://doi.org/10.1038/nrc2191

100. Albano E. Alcohol, oxidative stress and free radical damage. Proc Nutr Soc. 2006;65: 278-90. https://doi.org/10.1079/PNS2006496

101. Erol A, Ho AM, Winham SJ, Karpyak VM. Sex hormones in alcohol consumption: a systematic review of evidence. Addict Biol. 2019;24: 157-69. https://doi.org/10.1111/adb.12589

102. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi FE, Benbrahim-Tallaa L, et al. Carcinogenicity of consumption of red and processed meat. Lancet Oncol. 2015;16: 1599-600. https://doi.org/10.1016/S1470-2045(15)00444-1

103. Cross AJ, Sinha R. Meat-related mutagens/carcinogens in the etiology of colorectal cancer. Environ Mol Mutagen. 2004;44: 44-55. https://doi.org/10.1002/em.20030

104. Burkitt DP. Editorial: Large-bowel cancer: an epidemiologic jigsaw puzzle. J Natl Cancer Inst. 1975;54: 3-6. https://doi.org/10.1093/jnci/54.1.3

105. Song M, Wu K, Meyerhardt JA, Ogino S, Wang M, Fuchs CS, et al. Fiber intake and survival after colorectal cancer diagnosis. JAMA Oncology. 2018;4: 71-9. https://doi.org/10.1001/jamaoncol.2017.3684

106. Lipkin M, Reddy B, Newmark H, Lamprecht SA. Dietary factors in human colorectal cancer. Annu Rev Nutr. 1999;19: 545-86. https://doi.org/10.1146/annurev.nutr.19.1.545

107. Reddy BS, Engle A, Simi B, Goldman M. Effect of dietary fiber on colonic bacterial enzymes and bile acids in relation to colon cancer. Gastroenterology. 1992;102: 1475-82. https://doi.org/10.5555/uri:pii:0016508592917048

108. Giovannucci E. Metabolic syndrome, hyperinsulinemia, and colon cancer: a review. Am J Clin Nutr. 2007;86: s836-42. https://doi.org/10.1093/ajcn/86.3.836S

109. Song M, Chan AT, Sun J. Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. Gastroenterology. 2020;158: 322-40. https://doi.org/10.1053/j.gastro.2019.06.048

110. Aune D, Chan DSM, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ. 2011;343: d6617. https://doi.org/10.1136/bmj.d6617

111. Reynolds A, Mann J, Cummings J, Winter N, Mete E, Te Morenga L. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet. 2019;393: 434-45. https://doi.org/10.1016/s0140-6736(18)31809-9

112. Slavin JL. Mechanisms for the impact of whole grain foods on cancer risk. J Am Coll Nutr. 2000;19: 300s-7s. https://doi.org/10.1080/07315724.2000.10718964

113. Larsson SC, Mason AM, Kar S, Vithayathil M, Carter P, Baron JA, et al. Genetically proxied milk consumption and risk of colorectal, bladder, breast, and prostate cancer: a two-sample Mendelian randomization study. BMC Med. 2020;18: 370. https://doi.org/10.1186/s12916-020-01839-9

114. Norat T, Riboli E. Dairy products and colorectal cancer. A review of possible mechanisms and epidemiological evidence. Eur J Clin Nutr. 2003;57: 1-17. https://doi.org/10.1038/sj.ejcn.1601522

115. Kühn T, Stepien M, López-Nogueroles M, Damms-Machado A, Sookthai D, Johnson T, et al. Prediagnostic plasma bile acid levels and colon cancer risk: a prospective study. J Natl Cancer Inst. 2020;112: 516-24. https://doi.org/10.1093/jnci/djz166

116. Zhong Y, Zhu Y, Li Q, Wang F, Ge X, Zhou G, et al. Association between Mediterranean diet adherence and colorectal cancer: a dose-response meta-analysis. Am J Clin Nutr. 2020;111: 1214-25. https://doi.org/10.1093/ajcn/nqaa083

117. Key TJ, Appleby PN, Crowe FL, Bradbury KE, Schmidt JA, Travis RC. Cancer in British vegetarians: updated analyses of 4998 incident cancers in a cohort of 32,491 meat eaters, 8612 fish eaters, 18,298 vegetarians, and 2246 vegans. Am J Clin Nutr. 2014;100: 378S-85S. https://doi.org/10.3945/ajcn.113.071266

118. Tantamango-Bartley Y, Jaceldo-Siegl K, Fan J, Fraser G. Vegetarian diets and the incidence of cancer in a low-risk population. Cancer Epidemiol Biomarkers Prev. 2013;22: 286-94. https://doi.org/10.1158/1055-9965.Epi-12-1060

119. Orlich MJ, Singh PN, Sabaté J, Fan J, Sveen L, Bennett H, et al. Vegetarian dietary patterns and the risk of colorectal cancers. JAMA Intern Med. 2015;175: 767-76. https://doi.org/10.1001/jamainternmed.2015.59

120. WHO Classification of Tumours Editorial Board. Breast Tumours: Lyon (France): International Agency for Research on Cancer; 2019. (WHO classification of tumours series, 5th ed.; vol. 2).

121. Bray F, McCarron P, Parkin DM. The changing global patterns of female breast cancer incidence and mortality. Breast Cancer Res. 2004;6: 229-39. https://doi.org/10.1186/bcr932

122. Heer E, Harper A, Escandor N, Sung H, McCormack V, Fidler-Benaoudia MM. Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. Lancet Glob Health. 2020;8: e1027-e37. https://doi.org/10.1016/S2214-109X(20)30215-1

123. Joko-Fru WY, Jedy-Agba E, Korir A, Ogunbiyi O, Dzamalala CP, Chokunonga E, et al. The evolving epidemic of breast cancer in sub-Saharan Africa: Results from the African Cancer Registry Network. Int J Cancer. 2020;147: 2131-41. https://doi.org/10.1002/ijc.33014

124. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. CA Cancer J Clin. 2019;69: 438-51. https://doi.org/https://doi.org/10.3322/caac.21583

125. Jedy-Agba E, McCormack V, Adebamowo C, Dos-Santos-Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. Lancet Glob Health. 2016;4: e923-e35. https://doi.org/10.1016/s2214-109x(16)30259-5

126. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380: 1778-86. https://doi.org/10.1016/s0140-6736(12)61611-0

127. Malvezzi M, Carioli G, Bertuccio P, Boffetta P, Levi F, La Vecchia C, et al. European cancer mortality predictions for the year 2019 with focus on breast cancer. Ann Oncol. 2019;30: 781-7. https://doi.org/10.1093/annonc/mdz051

128. Gutnik LA, Matanje-Mwagomba B, Msosa V, Mzumara S, Khondowe B, Moses A, et al. Breast cancer screening in low- and middle-income countries: a perspective From Malawi. J Glob Oncol. 2016;2: 4-8. https://doi.org/10.1200/jgo.2015.000430

129. Black E, Richmond R. Improving early detection of breast cancer in sub-Saharan Africa: why mammography may not be the way forward. Global Health. 2019;15: 3. https://doi.org/10.1186/s12992-018-0446-6

130. World Cancer Research Fund/ American Institute for Cancer Research. Continuous Update Project Expert Report. Diet, nutrition, physical activity and breast cancer. 2018.

131. Anderson WF, Rosenberg PS, Prat A, Perou CM, Sherman ME. How many etiological subtypes of breast cancer: two, three, four, or more? J Natl Cancer Inst. 2014;106. https://doi.org/10.1093/jnci/dju165

132. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. The Lancet Oncology. 2001;2: 133-40. https://doi.org/https://doi.org/10.1016/S1470-2045(00)00254-0

133. McPherson K, Steel C, Dixon J. Breast cancer—epidemiology, risk factors, and genetics. BMJ. 2000;321: 624-8.

134. Balmaña J, Díez O, Rubio IT, Cardoso F. BRCA in breast cancer: ESMO Clinical Practice Guidelines. Ann Oncol. 2011;22 Suppl 6: vi31-4. https://doi.org/10.1093/annonc/mdr373

135. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015;47: 373-80. https://doi.org/10.1038/ng.3242

136. Endogenous Hormones and Breast Cancer Collaborative Group. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol. 2013;14: 1009-19. https://doi.org/10.1016/S1470-2045(13)70301-2

137. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst. 2002;94: 606-16. https://doi.org/10.1093/jnci/94.8.606

138. Tin Tin S, Reeves GK, Key TJ. Endogenous hormones and risk of invasive breast cancer in pre- and post-menopausal women: findings from the UK Biobank. Br J Cancer. 2021;125: 126-34. https://doi.org/10.1038/s41416-021-01392-z

139. Larsson SC, Kar S, Perry JRB, Carter P, Vithayathil M, Mason AM, et al. Serum estradiol and 20 site-specific cancers in women: mendelian randomization study. J Clin Endocrinol Metab. 2021;107: e467-e74. https://doi.org/10.1210/clinem/dgab713

140. Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, et al. Using human genetics to understand the disease impacts of testosterone in men and women. Nat Med. 2020;26: 252-8. https://doi.org/10.1038/s41591-020-0751-5

141. Dimou NL, Papadimitriou N, Gill D, Christakoudi S, Murphy N, Gunter MJ, et al. Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study. Int J Epidemiol. 2019;48: 807-16. https://doi.org/10.1093/ije/dyz107

142. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer. 2002;87: 1234-45. https://doi.org/10.1038/sj.bjc.6600596

143. Assi N, Rinaldi S, Viallon V, Dashti SG, Dossus L, Fournier A, et al. Mediation analysis of the alcohol-postmenopausal breast cancer relationship by sex hormones in the EPIC cohort. Int J Cancer. 2020;146: 759-68. https://doi.org/10.1002/ijc.32324

144. Cleary MP, Grossmann ME. Obesity and breast cancer: the estrogen connection. Endocrinology. 2009;150: 2537-42. https://doi.org/10.1210/en.2009-0070

145. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13: 1141-51. https://doi.org/10.1016/S1470-2045(12)70425-4

146. Reeves GK, Pirie K, Green J, Bull D, Beral V. Reproductive factors and specific histological types of breast cancer: prospective study and meta-analysis. Br J Cancer. 2009;100: 538-44. https://doi.org/10.1038/sj.bjc.6604853

147. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. Nat Genet. 2017;49: 834-41. https://doi.org/10.1038/ng.3841

148. Chen F, Wen W, Long J, Shu X, Yang Y, Shu X-O, et al. Mendelian randomization analyses of 23 known and suspected risk factors and biomarkers for breast cancer overall and by molecular subtypes. Int J Cancer. 2022;151: 372-80. https://doi.org/https://doi.org/10.1002/ijc.34026

149. Floud S, Balkwill A, Moser K, Reeves GK, Green J, Beral V, et al. The role of health-related behavioural factors in accounting for inequalities in coronary heart disease risk by education and area deprivation: prospective study of 1.2 million UK women. BMC Med. 2016;14: 145. https://doi.org/10.1186/s12916-016-0687-2

150. Lambe M, Hsieh C, Trichopoulos D, Ekbom A, Pavia M, Adami HO. Transient increase in the risk of breast cancer after giving birth. N Engl J Med. 1994;331: 5-9. https://doi.org/10.1056/nejm199407073310102

151. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet. 1996;347: 1713-27. https://doi.org/10.1016/s0140-6736(96)90806-5

152. Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. Lancet. 2019;394: 1159-68. https://doi.org/10.1016/S0140-6736(19)31709-X

153. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Combined estrogen-progestogen contraceptives and combined estrogen-progestogen menopausal therapy. IARC Monogr Eval Carcinog Risks Hum. 2007;91: 1-528.

154. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15: 1159-69.

155. Modan B, Chetrit A, Alfandary E, Katz L. Increased risk of breast cancer after low-dose irradiation. Lancet. 1989;1: 629-31. https://doi.org/10.1016/s0140-6736(89)92140-5

156. Shuryak I, Sachs RK, Brenner DJ. Cancer Risks After Radiation Exposure in Middle Age. J Natl Cancer Inst. 2010;102: 1628-36. https://doi.org/10.1093/jnci/djq346

157. Schubauer-Berigan MK, Anderson JL, Hein MJ, Little MP, Sigurdson AJ, Pinkerton LE. Breast cancer incidence in a cohort of U.S. flight attendants. Am J Ind Med. 2015;58: 252-66. https://doi.org/10.1002/ajim.22419

158. El Ghissassi F, Baan R, Straif K, Grosse Y, Secretan B, Bouvard V, et al. A review of human carcinogens —Part D: radiation. Lancet Oncol. 2009;10: 751-2. https://doi.org/10.1016/S1470-2045(09)70213-X

159. Boyle P, Boniol M, Koechlin A, Robertson C, Valentini F, Coppens K, et al. Diabetes and breast cancer risk: a meta-analysis. Br J Cancer. 2012;107: 1608-17. https://doi.org/10.1038/bjc.2012.414

160. Shu X, Wu L, Khankari NK, Shu XO, Wang TJ, Michailidou K, et al. Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. Int J Epidemiol. 2019;48: 795-806. https://doi.org/10.1093/ije/dyy201

161. Schoemaker MJ, Nichols HB, Wright LB, Brook MN, Jones ME, O'Brien KM, et al. Association of Body Mass Index and Age With Subsequent Breast Cancer Risk in Premenopausal Women. JAMA Oncology. 2018;4: e181771. https://doi.org/10.1001/jamaoncol.2018.1771

162. Tworoger SS, Eliassen AH, Missmer SA, Baer H, Rich-Edwards J, Michels KB, et al. Birthweight and body size throughout life in relation to sex hormones and prolactin concentrations in premenopausal women. Cancer Epidemiol Biomarkers Prev. 2006;15: 2494-501. https://doi.org/10.1158/1055-9965.Epi-06-0671

163. Colditz GA, Frazier AL. Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. Cancer Epidemiol Biomarkers Prev. 1995;4: 567-71.

164. De Pergola G, Silvestris F. Obesity as a major risk factor for cancer. J Obes. 2013;2013: 291546. https://doi.org/10.1155/2013/291546

165. Iyengar NM, Arthur R, Manson JE, Chlebowski RT, Kroenke CH, Peterson L, et al. Association of body fat and risk of breast cancer in postmenopausal women with normal body mass index: a secondary analysis of a randomized clinical trial and observational study. JAMA Oncology. 2019;5: 155-63. https://doi.org/10.1001/jamaoncol.2018.5327

166. Key TJ, Appleby PN, Reeves GK, Roddam AW, Helzlsouer KJ, Alberg AJ, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. Br J Cancer. 2011;105: 709-22. https://doi.org/10.1038/bjc.2011.254

167. Guo Y, Warren Andersen S, Shu XO, Michailidou K, Bolla MK, Wang Q, et al. Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of European descent. PLoS Med. 2016;13: e1002105. https://doi.org/10.1371/journal.pmed.1002105

168. Ooi BNS, Loh H, Ho PJ, Milne RL, Giles G, Gao C, et al. The genetic interplay between body mass index, breast size and breast cancer risk: a Mendelian randomization analysis. Int J Epidemiol. 2019;48: 781-94. https://doi.org/10.1093/ije/dyz124

169. Song M, Zheng Y, Qi L, Hu FB, Chan AT, Giovannucci EL. Longitudinal analysis of genetic susceptibility and BMI throughout adult life. Diabetes. 2018;67: 248-55. https://doi.org/10.2337/db17-1156

170. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. BMJ. 2020;369: m1203. https://doi.org/10.1136/bmj.m1203

171. Fang Z, Song M, Lee DH, Giovannucci EL. The role of Mendelian randomization studies in deciphering the effect of obesity on cancer. J Natl Cancer Inst. 2021;114: 361-71. https://doi.org/10.1093/jnci/djab102

172. Zhang B, Shu X-O, Delahanty RJ, Zeng C, Michailidou K, Bolla MK, et al. Height and breast cancer risk: evidence from prospective studies and Mendelian randomization. J Natl Cancer Inst. 2015;107: djv219. https://doi.org/10.1093/jnci/djv219

173. Albanes D, Winick M. Are cell number and cell proliferation risk factors for cancer? J Natl Cancer Inst. 1988;80: 772-4. https://doi.org/10.1093/jnci/80.10.772

174. He M, Xu M, Zhang B, Liang J, Chen P, Lee J-Y, et al. Meta-analysis of genome-wide association studies of adult height in East Asians identifies 17 novel loci. Hum Mol Genet. 2015;24: 1791-800. https://doi.org/10.1093/hmg/ddu583

175. Hojman P, Gehl J, Christensen JF, Pedersen BK. Molecular Mechanisms Linking Exercise to Cancer Prevention and Treatment. Cell Metab. 2018;27: 10-21. https://doi.org/10.1016/j.cmet.2017.09.015

176. Campbell KL, Foster-Schubert KE, Alfano CM, Wang C-C, Wang C-Y, Duggan CR, et al. Reduced-calorie dietary weight loss, exercise, and sex hormones in postmenopausal women: randomized controlled trial. J Clin Oncol. 2012;30: 2314-26. https://doi.org/10.1200/jco.2011.37.9792

177. Bertone-Johnson ER, Tworoger SS, Hankinson SE. Recreational physical activity and steroid hormone levels in postmenopausal women. Am J Epidemiol. 2009;170: 1095-104. https://doi.org/10.1093/aje/kwp254

178. Zhang X, Ashcraft KA, Betof Warner A, Nair SK, Dewhirst MW. Can exercise-induced modulation of the tumor physiologic microenvironment improve antitumor immunity? Cancer Res. 2019;79: 2447-56. https://doi.org/10.1158/0008-5472.Can-18-2468

179. Woods JA, Vieira VJ, Keylock KT. Exercise, inflammation, and innate immunity. Neurol Clin. 2006;24: 585-99. https://doi.org/10.1016/j.ncl.2006.03.008

180. Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. J Natl Cancer Inst. 2013;105: 515-25. https://doi.org/10.1093/jnci/djt023

181. Hecht SS. Tobacco smoke carcinogens and breast cancer. Environ Mol Mutagen. 2002;39: 119-26. https://doi.org/10.1002/em.10071

182. Petrakis NL, Gruenke LD, Beelen TC, Castagnoli N, Craig JC. Nicotine in Breast Fluid of Nonlactating Women. Science. 1978;199: 303-5. https://doi.org/doi:10.1126/science.619458

183. Tankó LB, Christiansen C. An update on the antiestrogenic effect of smoking: a literature review with implications for researchers and practitioners. Menopause. 2004;11: 104-9. https://doi.org/10.1097/01.Gme.0000079740.18541.Db

184. Jensen J, Christiansen C, Rødbro P. Cigarette smoking, serum estrogens, and bone loss during hormone-replacement therapy early after menopause. N Engl J Med. 1985;313: 973-5. https://doi.org/10.1056/nejm198510173131602

185. Willett WC. Dietary fat and breast cancer. Toxicol Sci. 1999: 127-46.

186. Michels KB, Mohllajee AP, Roset-Bahmanyar E, Beehler GP, Moysich KB. Diet and breast cancer. Cancer. 2007;109: 2712-49. https://doi.org/https://doi.org/10.1002/cncr.22654

187. Prentice RL, Caan B, Chlebowski RT, Patterson R, Kuller LH, Ockene JK, et al. Low-fat dietary pattern and risk of invasive breast cancer: the women's health initiative randomized controlled dietary modification trial. JAMA. 2006;295: 629-42. https://doi.org/10.1001/jama.295.6.629

188. Martin LJ, Li Q, Melnichouk O, Greenberg C, Minkin S, Hislop G, et al. A randomized trial of dietary intervention for breast cancer prevention. Cancer Res. 2011;71: 123-33. https://doi.org/10.1158/0008-5472.Can-10-1436

189. Fung TT, Hu FB, Barbieri RL, Willett WC, Hankinson SE. Dietary patterns, the Alternate Healthy Eating Index and plasma sex hormone concentrations in postmenopausal women. Int J Cancer. 2007;121: 803-9. https://doi.org/10.1002/ijc.22728

190. Boyd NF, Lockwood GA, Greenberg CV, Martin LJ, Tritchler DL, for the Canadian D, et al. Effects of a low-fat high-carbohydrate diet on plasma sex hormones in premenopausal women: results from a randomized controlled trial. Br J Cancer. 1997;76: 127-35. https://doi.org/10.1038/bjc.1997.348

191. Berrino F, Bellati C, Secreto G, Camerini E, Pala V, Panico S, et al. Reducing Bioavailable Sex Hormones through a Comprehensive Change in Diet: the Diet and Androgens (DIANA) Randomized Trial. Cancer Epidemiol Biomarkers Prev. 2001;10: 25-33.

192. Heber D, Ashley JM, Leaf DA, Barnard RJ. Reduction of serum estradiol in postmenopausal women given free access to low-fat high-carbohydrate diet. Nutrition 1991;7: 137-40.

193. Wu AH, Pike MC, Stram DO. Meta-analysis: dietary fat intake, serum estrogen levels, and the risk of breast cancer. J Natl Cancer Inst. 1999;91: 529-34. https://doi.org/10.1093/jnci/91.6.529

194. Günther AL, Karaolis-Danckert N, Kroke A, Remer T, Buyken AE. Dietary protein intake throughout childhood is associated with the timing of puberty. J Nutr. 2010;140: 565-71. https://doi.org/10.3945/jn.109.114934

195. Cheng TS, Sharp SJ, Brage S, Emmett PM, Forouhi NG, Ong KK. Longitudinal associations between prepubertal childhood total energy and macronutrient intakes and subsequent puberty timing in UK boys and girls. Eur J Nutr. 2022;61: 157-67. https://doi.org/10.1007/s00394-021-02629-6

196. Berkey CS, Gardner JD, Lindsay Frazier A, Colditz GA. Relation of childhood diet and body size to menarche and adolescent growth in girls. Am J Epidemiol. 2000;152: 446-52. https://doi.org/10.1093/aje/152.5.446

197. Dunneram Y, Greenwood DC, Burley VJ, Cade JE. Dietary intake and age at natural menopause: results from the UK Women's Cohort Study. J Epidemiol Community Health. 2018;72: 733-40. https://doi.org/10.1136/jech-2017-209887

198. Nagel G, Altenburg H-P, Nieters A, Boffetta P, Linseisen J. Reproductive and dietary determinants of the age at menopause in EPIC-Heidelberg. Maturitas. 2005;52: 337-47. https://doi.org/10.1016/j.maturitas.2005.05.013

199. Grisotto G, Farago JS, Taneri PE, Wehrli F, Roa-Díaz ZM, Minder B, et al. Dietary factors and onset of natural menopause: A systematic review and meta-analysis. Maturitas. 2021.

200. Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, et al. Carcinogenicity of alcoholic beverages. Lancet Oncol. 2007;8: 292-3. https://doi.org/10.1016/S1470-2045(07)70099-2

201. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol consumption and ethyl carbamate. IARC Monogr Eval Carcinog Risks Hum. 2010;96: 3-1383.

202. Ong JS, Derks EM, Eriksson M, An J, Hwang LD, Easton DF, et al. Evaluating the role of alcohol consumption in breast and ovarian cancer susceptibility using population - based cohort studies and two - sample Mendelian randomization analyses. Int J Cancer. 2021;148: 1338-50. https://doi.org/10.1002/ijc.33308

203. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. JAMA. 2001;286: 2143-51. https://doi.org/10.1001/jama.286.17.2143

204. Boffetta P, Hashibe M. Alcohol and cancer. Lancet Oncol. 2006;7: 149-56. https://doi.org/10.1016/S1470-2045(06)70577-0

205. Papadimitriou N, Dimou N, Gill D, Tzoulaki I, Murphy N, Riboli E, et al. Genetically predicted circulating concentrations of micronutrients and risk of breast cancer: A Mendelian randomization study. Int J Cancer. 2021;148: 646-53. https://doi.org/10.1002/ijc.33246

206. Cui Y, Rohan TE. Vitamin D, calcium, and breast cancer risk: a review. Cancer Epidemiol Biomarkers Prev. 2006;15: 1427-37. https://doi.org/10.1158/1055-9965.Epi-06-0075

207. Cashman KD. Calcium intake, calcium bioavailability and bone health. Br J Nutr. 2002;87: S169-S77. https://doi.org/10.1079/BJN/2002534

208. Aune D, Chan DS, Vieira AR, Navarro Rosenblatt DA, Vieira R, Greenwood DC, et al. Dietary compared with blood concentrations of carotenoids and breast cancer risk: a systematic review and meta-analysis of prospective studies. Am J Clin Nutr. 2012;96: 356-73. https://doi.org/10.3945/ajcn.112.034165

209. Elliott R. Mechanisms of genomic and non-genomic actions of carotenoids. Biochim Biophys Acta. 2005;1740: 147-54. https://doi.org/10.1016/j.bbadis.2004.12.009

210. Sánchez-Zamorano LM, Flores-Luna L, Angeles-Llerenas A, Ortega-Olvera C, Lazcano-Ponce E, Romieu I, et al. The Western dietary pattern is associated with increased serum concentrations of free estradiol in postmenopausal women: implications for breast cancer prevention. Nutr Res. 2016;36: 845-54. https://doi.org/10.1016/j.nutres.2016.04.008

211. Brinkman MT, Baglietto L, Krishnan K, English DR, Severi G, Morris HA, et al. Consumption of animal products, their nutrient components and postmenopausal circulating steroid hormone concentrations. Eur J Clin Nutr. 2010;64: 176-83. https://doi.org/10.1038/ejcn.2009.129

212. Penniecook-Sawyers JA, Jaceldo-Siegl K, Fan J, Beeson L, Knutsen S, Herring P, et al. Vegetarian dietary patterns and the risk of breast cancer in a low-risk population. Br J Nutr. 2016;115: 1790-7. https://doi.org/10.1017/S0007114516000751

213. Stone EM, Williams JA, Grover PL, Gusterson BA, Phillips DH. Interindividual variation in the metabolic activation of heterocyclic amines and their N-hydroxy derivatives in primary cultures

of human mammary epithelial cells. Carcinogenesis. 1998;19: 873-9. https://doi.org/10.1093/carcin/19.5.873

214. Wei Y, Lv J, Guo Y, Bian Z, Gao M, Du H, et al. Soy intake and breast cancer risk: a prospective study of 300,000 Chinese women and a dose-response meta-analysis. Eur J Epidemiol. 2020;35: 567-78. https://doi.org/10.1007/s10654-019-00585-4

215. Davey GK, Spencer EA, Appleby PN, Allen NE, Knox KH, Key TJ. EPIC–Oxford:lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public Health Nutr. 2003;6: 259-68. https://doi.org/10.1079/PHN2002430

216. Orlich MJ, Fraser GE. Vegetarian diets in the Adventist Health Study 2: a review of initial published findings. Am J Clin Nutr. 2014;100 Suppl 1: 353s-8s. https://doi.org/10.3945/ajcn.113.071233

217. Farvid MS, Spence ND, Holmes MD, Barnett JB. Fiber consumption and breast cancer incidence: A systematic review and meta-analysis of prospective studies. Cancer. 2020;126: 3061-75. https://doi.org/10.1002/cncr.32816

218. Holger Moch PAH, Thomas M. Ul bright, Victor E. Reuter (Eds):. WHO Classification of Tumours of the Urinary System and Male Genital Organs (4th edition). IARC: Lyon 2016.

219. Hurwitz LM, Agalliu I, Albanes D, Barry KH, Berndt SI, Cai Q, et al. Recommended definitions of aggressive prostate cancer for etiologic epidemiologic research. J Natl Cancer Inst. 2020;113: 727-34. https://doi.org/10.1093/jnci/djaa154

220. Platz EA, De Marzo AM, Giovannucci E. Prostate cancer association studies: pitfalls and solutions to cancer misclassification in the PSA era. J Cell Biochem. 2004;91: 553-71. https://doi.org/10.1002/jcb.10700

221. Gleason DF. Histologic grading and clinical staging of prostate carcinoma. Urologic pathology: the prostate. 1977.

222. Gleason DF. Histologic grading of prostate cancer: a perspective. Hum Pathol. 1992;23: 273-9. https://doi.org/10.1016/0046-8177(92)90108-f

223. Wong MCS, Goggins WB, Wang HHX, Fung FDH, Leung C, Wong SYS, et al. Global incidence and mortality for prostate cancer: analysis of temporal patterns and trends in 36 countries. Eur Urol. 2016;70: 862-74. https://doi.org/https://doi.org/10.1016/j.eururo.2016.05.043

224. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med. 1991;324: 1156-61. https://doi.org/10.1056/nejm199104253241702

225. Carter HB. Prostate-specific antigen (PSA) screening for prostate cancer: revisiting the evidence. JAMA. 2018;319: 1866-8. https://doi.org/10.1001/jama.2018.4914

226. Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and overtreatment of prostate cancer. Eur Urol. 2014;65: 1046-55.

227. Potosky AL, Miller BA, Albertsen PC, Kramer BS. The role of increasing detection in the rising incidence of prostate cancer. JAMA. 1995;273: 548-52. https://doi.org/10.1001/jama.1995.03520310046028

228. Zhou CK, Check DP, Lortet-Tieulent J, Laversanne M, Jemal A, Ferlay J, et al. Prostate cancer incidence in 43 populations worldwide: An analysis of time trends overall and by age group. Int J Cancer. 2016;138: 1388-400. https://doi.org/10.1002/ijc.29894

229. Tsodikov A, Gulati R, Heijnsdijk EAM, Pinsky PF, Moss SM, Qiu S, et al. Reconciling the effects of screening on prostate cancer mortality in the ERSPC and PLCO Trials. Ann Intern Med. 2017;167: 449-55. https://doi.org/10.7326/m16-2586

230. Moyer VA. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;157: 120-34. https://doi.org/10.7326/0003-4819-157-2-201207170-00459

231. Tikkinen KAO, Dahm P, Lytvyn L, Heen AF, Vernooij RWM, Siemieniuk RAC, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. BMJ. 2018;362: k3581. https://doi.org/10.1136/bmj.k3581

232. Bradley SH, Funston G, Jones D, Watson J. Diagnosing prostate cancer in asymptomatic patients. BMJ. 2022;377: e071076. https://doi.org/10.1136/bmj-2022-071076

233. Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, Davidson KW, et al. Screening for Prostate Cancer. JAMA. 2018;319: 1901. https://doi.org/10.1001/jama.2018.3710

234. Young GJ, Harrison S, Turner EL, Walsh EI, Oliver SE, Ben-Shlomo Y, et al. Prostate-specific antigen (PSA) testing of men in UK general practice: a 10-year longitudinal cohort study. BMJ Open. 2017;7: e017729. https://doi.org/10.1136/bmjopen-2017-017729

235. SEER Cancer Stat Facts: Prostate Cancer. National Cancer Institute. Bethesda MD. https://seer.cancer.gov/statfacts/html/prost.html. Accessed June 20 2022.

236. Bleyer A, Spreafico F, Barr R. Prostate cancer in young men: An emerging young adult and older adolescent challenge. Cancer. 2020;126: 46-57. https://doi.org/https://doi.org/10.1002/cncr.32498

237. Bell KJ, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. Int J Cancer. 2015;137: 1749-57. https://doi.org/10.1002/ijc.29538

238. Pettersson A, Robinson D, Garmo H, Holmberg L, Stattin P. Age at diagnosis and prostate cancer treatment and prognosis: a population-based cohort study. Ann Oncol. 2018;29: 377-85. https://doi.org/10.1093/annonc/mdx742

239. Knipper S, Pecoraro A, Palumbo C, Rosiello G, Luzzago S, Deuker M, et al. The effect of age on cancer-specific mortality in patients with prostate cancer: a population-based study across all stages. Cancer Causes Control. 2020;31: 283-90. https://doi.org/10.1007/s10552-020-01273-5

240. Hjelmborg JB, Scheike T, Holst K, Skytthe A, Penney KL, Graff RE, et al. The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev. 2014;23: 2303-10. https://doi.org/10.1158/1055-9965.Epi-13-0568

241. Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet. 2014;46: 1103-9. https://doi.org/10.1038/ng.3094

242. Kiciński M, Vangronsveld J, Nawrot TS. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. PLoS One. 2011;6: e27130.

243. Eeles R, Goh C, Castro E, Bancroft E, Guy M, Al Olama AA, et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat Rev Urol. 2014;11: 18-31. https://doi.org/10.1038/nrurol.2013.266

244. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E, Habel LA, Graff RE, et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. Cancer Discov. 2015;5: 878-91. https://doi.org/10.1158/2159-8290.Cd-15-0315

245. Benafif S, Kote-Jarai Z, Eeles RA. A review of prostate cancer genome-wide association studies (GWAS). Cancer Epidemiol Biomarkers Prev. 2018;27: 845-57. https://doi.org/10.1158/1055-9965.Epi-16-1046

246. Shang Z, Zhu S, Zhang H, Li L, Niu Y. Germline momeobox B13 (HOXB13) G84E mutation and prostate cancer risk in european descendants: a meta-analysis of 24 213 cases and 73 631 controls. Eur Urol. 2013;64: 173-6. https://doi.org/https://doi.org/10.1016/j.eururo.2013.03.007

247. Nyberg T, Tischkowitz M, Antoniou AC. BRCA1 and BRCA2 pathogenic variants and prostate cancer risk: systematic review and meta-analysis. Br J Cancer. 2022;126: 1067-81. https://doi.org/10.1038/s41416-021-01675-5

248. Ryan S, Jenkins MA, Win AK. Risk of prostate cancer in Lynch syndrome: a systematic review and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2014;23: 437-49. https://doi.org/10.1158/1055-9965.Epi-13-1165

249. DeSantis CE, Miller KD, Goding Sauer A, Jemal A, Siegel RL. Cancer statistics for African Americans, 2019. CA Cancer J Clin. 2019;69: 211-33. https://doi.org/https://doi.org/10.3322/caac.21555

250. Perez-Cornago A, Key TJ, Allen NE, Fensom GK, Bradbury KE, Martin RM, et al. Prospective investigation of risk factors for prostate cancer in the UK Biobank cohort study. Br J Cancer. 2017;117: 1562-71. https://doi.org/10.1038/bjc.2017.312

251. McGinley KF, Tay KJ, Moul JW. Prostate cancer in men of African origin. Nat Rev Urol. 2016;13: 99-107. https://doi.org/10.1038/nrurol.2015.298

252. Karakas C, Wang C, Deng F, Huang H, Wang D, Lee P. Molecular mechanisms involving prostate cancer racial disparity. Am J Clin Exp Urol. 2017;5: 34-48.

253. Han Y, Rand KA, Hazelett DJ, Ingles SA, Kittles RA, Strom SS, et al. Prostate Cancer Susceptibility in Men of African Ancestry at 8q24. J Natl Cancer Inst. 2016;108: djv431. https://doi.org/10.1093/jnci/djv431

254. Platz EA, Giovannucci E. The epidemiology of sex steroid hormones and their signaling and metabolic pathways in the etiology of prostate cancer. J Steroid Biochem Mol Biol. 2004;92: 237-53. https://doi.org/https://doi.org/10.1016/j.jsbmb.2004.10.002

255. Sharifi N, Gulley JL, Dahut WL. Androgen deprivation therapy for prostate cancer. JAMA. 2005;294: 238-44. https://doi.org/10.1001/jama.294.2.238

256. Andriole GL, Bostwick DG, Brawley OW, Gomella LG, Marberger M, Montorsi F, et al. Effect of dutasteride on the risk of prostate cancer. N Engl J Med. 2010;362: 1192-202. https://doi.org/10.1056/NEJMoa0908127

257. Dunn JF, Nisula BC, Rodbard D. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma. J Clin Endocrinol Metab. 1981;53: 58-68. https://doi.org/10.1210/jcem-53-1-58

258. Watts EL, Appleby PN, Perez-Cornago A, Bueno-de-Mesquita HB, Chan JM, Chen C, et al. Low free testosterone and prostate cancer risk: a collaborative analysis of 20 prospective studies. Eur Urol. 2018;74: 585-94. https://doi.org/10.1016/j.eururo.2018.07.024

259. Morgentaler A, Traish AM. Shifting the paradigm of testosterone and prostate cancer: the saturation model and the limits of androgen-dependent growth. Eur Urol. 2009;55: 310-20. https://doi.org/10.1016/j.eururo.2008.09.024

260. Watts EL, Perez-Cornago A, Fensom GK, Smith-Byrne K, Noor U, Andrews CD, et al. Circulating free testosterone and risk of aggressive prostate cancer: Prospective and Mendelian randomisation analyses in international consortia. Int J Cancer. 2021;151: 1033-46. https://doi.org/10.1002/ijc.34116

261. Watts EL, Appleby PN, Albanes D, Black A, Chan JM, Chen C, et al. Circulating sex hormones in relation to anthropometric, sociodemographic and behavioural factors in an international dataset of 12,300 men. PLoS One. 2017;12: e0187741. https://doi.org/10.1371/journal.pone.0187741

262. World Cancer Research Fund/ American Institute for Cancer Research. Continuous Update Project Expert Report. Diet, nutrition, physical activity and prostate cancer. 2018.

263. Khankari NK, Shu X-O, Wen W, Kraft P, Lindström S, Peters U, et al. Association between adult height and risk of colorectal, lung, and prostate cancer: results from meta-analyses of prospective studies and mendelian randomization analyses. PLoS Med. 2016;13: e1002118. https://doi.org/10.1371/journal.pmed.1002118

264. Davies NM, Gaunt TR, Lewis SJ, Holly J, Donovan JL, Hamdy FC, et al. The effects of height and BMI on prostate cancer incidence and mortality: a Mendelian randomization study in 20,848 cases and 20,214 controls from the PRACTICAL consortium. Cancer Causes Control. 2015;26: 1603-16. https://doi.org/10.1007/s10552-015-0654-9

265. Cohen P, Bright GM, Rogol AD, Kappelgaard AM, Rosenfeld RG. Effects of dose and gender on the growth and growth factor response to GH in GH-deficient children: implications for efficacy and safety. J Clin Endocrinol Metab. 2002;87: 90-8. https://doi.org/10.1210/jcem.87.1.8150

266. Kriström B, Lundberg E, Jonsson B, Albertsson-Wikland K, group obots. IGF-1 and growth response to adult height in a randomized GH treatment trial in short non-GH-deficient children. J Clin Endocrinol Metab. 2014;99: 2917-24. https://doi.org/10.1210/jc.2014-1101

267. Harrison S, Tilling K, Turner EL, Martin RM, Lennon R, Lane JA, et al. Systematic review and meta-analysis of the associations between body mass index, prostate cancer, advanced prostate cancer, and prostate-specific antigen. Cancer Causes Control. 2020;31: 431-49. https://doi.org/10.1007/s10552-020-01291-3

268. Perez-Cornago A, Appleby PN, Pischon T, Tsilidis KK, Tjønneland A, Olsen A, et al. Tall height and obesity are associated with an increased risk of aggressive prostate cancer: results from the EPIC cohort study. BMC Med. 2017;15: 115. https://doi.org/10.1186/s12916-017-0876-7

269. Wright ME, Chang SC, Schatzkin A, Albanes D, Kipnis V, Mouw T, et al. Prospective study of adiposity and weight change in relation to prostate cancer incidence and mortality. Cancer. 2007;109: 675-84. https://doi.org/10.1002/cncr.22443

270. Perez-Cornago A, Dunneram Y, Watts EL, Key TJ, Travis RC. Adiposity and risk of prostate cancer death: a prospective analysis in UK Biobank and meta-analysis of published studies. BMC Med. 2022;20: 143. https://doi.org/10.1186/s12916-022-02336-x

271. Gao C, Patel CJ, Michailidou K, Peters U, Gong J, Schildkraut J, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. Int J Epidemiol. 2016;45: 896-908. https://doi.org/10.1093/ije/dyw129

272. Allott EH, Masko EM, Freedland SJ. Obesity and prostate cancer: weighing the evidence. Eur Urol. 2013;63: 800-9. https://doi.org/10.1016/j.eururo.2012.11.013

273. Parekh N, Lin Y, Dipaola RS, Marcella S, Lu-Yao G. Obesity and prostate cancer detection: insights from three national surveys. Am J Med. 2010;123: 829-35. https://doi.org/10.1016/j.amjmed.2010.05.011

274. Jian Gang P, Mo L, Lu Y, Runqi L, Xing Z. Diabetes mellitus and the risk of prostate cancer: an update and cumulative meta-analysis. Endocr Res. 2015;40: 54-61. https://doi.org/10.3109/07435800.2014.934961

275. Tsilidis KK, Allen NE, Appleby PN, Rohrmann S, Nöthlings U, Arriola L, et al. Diabetes mellitus and risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2015;136: 372-81. https://doi.org/10.1002/ijc.28989

276. Au Yeung SL, Schooling CM. Impact of glycemic traits, type 2 diabetes and metformin use on breast and prostate cancer risk: a Mendelian randomization study. BMJ Open Diabetes Res Care. 2019;7: e000872. https://doi.org/10.1136/bmjdrc-2019-000872

277. Teppala S, Shankar A. Association between serum IGF-1 and diabetes among U.S. adults. Diabetes Care. 2010;33: 2257-9. https://doi.org/10.2337/dc10-0770

278. Grossmann M. Low testosterone in men with type 2 diabetes: significance and treatment. J Clin Endocrinol Metab. 2011;96: 2341-53. https://doi.org/10.1210/jc.2011-0118

279. Wright JL, Stanford JL. Metformin use and prostate cancer in Caucasian men: results from a population-based case-control study. Cancer Causes Control. 2009;20: 1617-22. https://doi.org/10.1007/s10552-009-9407-y

280. Huncharek M, Haddock KS, Reid R, Kupelnick B. Smoking as a risk factor for prostate cancer: a meta-analysis of 24 prospective cohort studies. Am J Public Health. 2010;100: 693-701. https://doi.org/10.2105/ajph.2008.150508

281. Islami F, Moreira DM, Boffetta P, Freedland SJ. A systematic review and meta-analysis of tobacco use and prostate cancer mortality and incidence in prospective cohort studies. Eur Urol. 2014;66: 1054-64. https://doi.org/10.1016/j.eururo.2014.08.059

282. Vasconcelos A, Santos T, Ravasco P, Neves PM. Dairy products: is there an impact on promotion of prostate cancer? A review of the literature. Front Nutr. 2019;6: 62. https://doi.org/10.3389/fnut.2019.00062

283. Downer MK, Batista JL, Mucci LA, Stampfer MJ, Epstein MM, Håkansson N, et al. Dairy intake in relation to prostate cancer survival. Int J Cancer. 2017;140: 2060-9. https://doi.org/https://doi.org/10.1002/ijc.30642

284. Heaney RP, McCarron DA, Dawson-Hughes B, Oparil S, Berga SL, Stern JS, et al. Dietary changes favorably affect bone remodeling in older adults. J Am Diet Assoc. 1999;99: 1228-33. https://doi.org/10.1016/s0002-8223(99)00302-8

285. Manios Y, Moschonis G, Trovas G, Lyritis GP. Changes in biochemical indexes of bone metabolism and bone mineral density after a 12-mo dietary intervention program: the Postmenopausal Health Study. Am J Clin Nutr. 2007;86: 781-9. https://doi.org/10.1093/ajcn/86.3.781

286. Iuliano S, Poon S, Robbins J, Bui M, Wang X, De Groot L, et al. Effect of dietary sources of calcium and protein on hip fractures and falls in older adults in residential care: cluster randomised controlled trial. BMJ. 2021;375: n2364. https://doi.org/10.1136/bmj.n2364

287. Aune D, Navarro Rosenblatt DA, Chan DS, Vieira AR, Vieira R, Greenwood DC, et al. Dairy products, calcium, and prostate cancer risk: a systematic review and meta-analysis of cohort studies. Am J Clin Nutr. 2015;101: 87-117. https://doi.org/10.3945/ajcn.113.067157

288. Giovannucci E. Dietary influences of 1,25(OH)2 vitamin D in relation to prostate cancer: a hypothesis. Cancer Causes Control. 1998;9: 567-82. https://doi.org/10.1023/a:1008835903714

289. Chen TC, Holick MF. Vitamin D and prostate cancer prevention and treatment. Trends Endocrinol Metab. 2003;14: 423-30. https://doi.org/10.1016/j.tem.2003.09.004

290. Yarmolinsky J, Berryman K, Langdon R, Bonilla C, Davey Smith G, Martin RM, et al. Mendelian randomization does not support serum calcium in prostate cancer risk. Cancer Causes Control. 2018;29: 1073-80. https://doi.org/10.1007/s10552-018-1081-5

291. Baron JA, Beach M, Wallace K, Grau MV, Sandler RS, Mandel JS, et al. Risk of prostate cancer in a randomized clinical trial of calcium supplementation. Cancer Epidemiol Biomarkers Prev. 2005;14: 586-9. https://doi.org/10.1158/1055-9965.Epi-04-0319

292. Bristow SM, Bolland MJ, MacLennan GS, Avenell A, Grey A, Gamble GD, et al. Calcium supplements and cancer risk: a meta-analysis of randomised controlled trials. Br J Nutr. 2013;110: 1384-93. https://doi.org/10.1017/S0007114513001050

293. Rowles JL, Ranard KM, Smith JW, An R, Erdman JW. Increased dietary and circulating lycopene are associated with reduced prostate cancer risk: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2017;20: 361-77. https://doi.org/10.1038/pcan.2017.25

294. Dimitrakopoulou VI, Tsilidis KK, Haycock PC, Dimou NL, Al-Dabhani K, Martin RM, et al. Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. BMJ. 2017;359: j4761. https://doi.org/10.1136/bmj.j4761

295. Yarmolinsky J, Bonilla C, Haycock PC, Langdon RJQ, Lotta LA, Langenberg C, et al. Circulating selenium and prostate cancer risk: a Mendelian randomization analysis. J Natl Cancer Inst. 2018;110: 1035-8. https://doi.org/10.1093/jnci/djy081

296. Xin J, Jiang X, Ben S, Yuan Q, Su L, Zhang Z, et al. Association between circulating vitamin E and ten common cancers: evidence from large-scale Mendelian randomization analysis and a longitudinal cohort study. BMC Med. 2022;20: 168. https://doi.org/10.1186/s12916-022-02366-5

297. Tantamango-Bartley Y, Knutsen SF, Knutsen R, Jacobsen BK, Fan J, Beeson WL, et al. Are strict vegetarians protected against prostate cancer? Am J Clin Nutr. 2015;103: 153-60. https://doi.org/10.3945/ajcn.114.106450

298. Allen NE, Appleby PN, Davey GK, Kaaks R, Rinaldi S, Key TJ. The associations of diet with serum insulin-like growth factor I and its main binding proteins in 292 women meat-eaters, vegetarians, and vegans. Cancer Epidemiol Biomarkers Prev. 2002;11: 1441-8.

299. Allen NE, Appleby PN, Davey GK, Key TJ. Hormones and diet: low insulin-like growth factor-I but normal bioavailable androgens in vegan men. Br J Cancer. 2000;83: 95-7. https://doi.org/10.1054/bjoc.2000.1152

300. Tong TYN, Appleby PN, Bradbury KE, Key TJ. Cross-sectional analyses of participation in cancer screening and use of hormone replacement therapy and medications in meat eaters and vegetarians: the EPIC-Oxford study. BMJ Open. 2017;7: e018245. https://doi.org/10.1136/bmjopen-2017-018245

301. Valentinis B, Baserga R. IGF-I receptor signalling in transformation and differentiation. Mol Pathol. 2001;54: 133-7. https://doi.org/10.1136/mp.54.3.133

302. Fürstenberger G, Senn H-J. Insulin-like growth factors and cancer. Lancet Oncol. 2002;3: 298-302. https://doi.org/https://doi.org/10.1016/S1470-2045(02)00731-3

303. Deal C, Ma J, Wilkin F, Paquette J, Rozen F, Ge B, et al. Novel promoter polymorphism in insulin-like growth factor-binding protein-3: Correlation with serum levels and interaction with known regulators. J Clin Endocrinol Metab. 2001;86: 1274-80. https://doi.org/10.1210/jc.86.3.1274

304. Kelley KM, Oh Y, Gargosky SE, Gucev Z, Matsumoto T, Hwa V, et al. Insulin-like growth factor-binding proteins (IGFBPs) and their regulatory dynamics. Int J Biochem Cell Biol. 1996;28: 619-37. https://doi.org/https://doi.org/10.1016/1357-2725(96)00005-2

305. Melnik BC, John SM, Carrera-Bastos P, Cordain L. The impact of cow's milk-mediated mTORC1-signaling in the initiation and progression of prostate cancer. Nutr Metab (Lond). 2012;9: 74. https://doi.org/10.1186/1743-7075-9-74

306. Cohen P, Ocrant I, Fielder PJ, Neely EK, Gargosky SE, Deal CI, et al. Insulin-like growth factors (IGFs): implications for aging. Psychoneuroendocrinology. 1992;17: 335-42. https://doi.org/10.1016/0306-4530(92)90039-a

307. Kaklamani VG, Linos A, Kaklamani E, Markaki I, Mantzoros C. Age, sex, and smoking are predictors of circulating insulin-like growth factor 1 and insulin-like growth factor-binding protein 3. J Clin Oncol. 1999;17: 813-7. https://doi.org/10.1200/jco.1999.17.3.813

308. Soliman A, De Sanctis V, Elalaily R, Bedair S. Advances in pubertal growth and factors influencing it: Can we increase pubertal growth? Indian J Endocrinol Metab. 2014;18: S53-62. https://doi.org/10.4103/2230-8210.145075

309. Watts EL, Perez-Cornago A, Appleby PN, Albanes D, Ardanaz E, Black A, et al. The associations of anthropometric, behavioural and sociodemographic factors with circulating concentrations of IGF-I, IGF-II, IGFBP-1, IGFBP-2 and IGFBP-3 in a pooled analysis of 16,024 men from 22 studies. Int J Cancer. 2019;145: 3244-56. https://doi.org/10.1002/ijc.32276

310. Bradbury KE, Balkwill A, Tipper SJ, Crowe FL, Reeves GK, Green J, et al. The association of plasma IGF-I with dietary, lifestyle, anthropometric, and early life factors in postmenopausal women. Growth Horm IGF Res. 2015;25: 90-5. https://doi.org/https://doi.org/10.1016/j.ghir.2015.01.001

311. Holmes MD, Pollak MN, Hankinson SE. Lifestyle correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev. 2002;11: 862-7.

312. Bleach R, Sherlock M, O'Reilly MW, McIlroy M. Growth hormone/insulin growth factor axis in sex steroid associated disorders and related cancers. Front Cell Dev Biol. 2021;9. https://doi.org/10.3389/fcell.2021.630503

313. Thissen J-P, Ketelslegers J-M, Underwood LE. Nutritional regulation of the insulin-like growth factors. Endocr Rev. 1994;15: 80-101. https://doi.org/10.1210/edrv-15-1-80

314. Estívariz CF, Ziegler TR. Nutrition and the insulin-like growth factor system. Endocrine. 1997;7: 65-71. https://doi.org/10.1007/bf02778066

315. Rahmani J, Kord Varkaneh H, Clark C, Zand H, Bawadi H, Ryan PM, et al. The influence of fasting and energy restricting diets on IGF-1 levels in humans: A systematic review and meta-analysis. Ageing Res Rev. 2019;53: 100910. https://doi.org/10.1016/j.arr.2019.100910

316. Zhu K, Meng X, Kerr DA, Devine A, Solah V, Binns CW, et al. The effects of a two-year randomized, controlled trial of whey protein supplementation on bone structure, IGF-1, and urinary calcium excretion in older postmenopausal women. J Bone Miner Res. 2011;26: 2298-306. https://doi.org/10.1002/jbmr.429

317. Ballard TL, Clapper JA, Specker BL, Binkley TL, Vukovich MD. Effect of protein supplementation during a 6-mo strength and conditioning program on insulin-like growth factor

I and markers of bone turnover in young adults. Am J Clin Nutr. 2005;81: 1442-8. https://doi.org/10.1093/ajcn/81.6.1442

318. Hidayat K, Chen J-S, Wang T-C, Liu Y-J, Shi Y-J, Su H-W, et al. The effects of milk supplementation on bone health indices in adults: a meta-analysis of randomized controlled trials. Adv Nutr. 2021. https://doi.org/10.1093/advances/nmab136

319. Thissen JP, Pucilowska JB, Underwood LE. Differential regulation of insulin-like growth factor I (IGF-I) and IGF binding protein-1 messenger ribonucleic acids by amino acid availability and growth hormone in rat hepatocyte primary culture. Endocrinology. 1994;134: 1570-6. https://doi.org/10.1210/endo.134.3.7509741

320. Sekharam M, Zhao H, Sun M, Fang Q, Zhang Q, Yuan Z, et al. Insulin-like Growth Factor 1 Receptor Enhances Invasion and Induces Resistance to Apoptosis of Colon Cancer Cells through the Akt/Bcl-xL Pathway. Cancer Res. 2003;63: 7708-16.

321. Lahm H, Amstad P, Wyniger J, Yilmaz A, Fischer JR, Schreyer M, et al. Blockade of the insulin-like growth-factor-I receptor inhibits growth of human colorectal cancer cells: evidence of a functional IGF-II-mediated autocrine loop. Int J Cancer. 1994;58: 452-9. https://doi.org/10.1002/ijc.2910580325

322. Vigneri PG, Tirrò E, Pennisi MS, Massimino M, Stella S, Romano C, et al. The insulin/IGF system in colorectal cancer development and resistance to therapy. Front Oncol. 2015;5. https://doi.org/10.3389/fonc.2015.00230

323. Keku TO, Lund PK, Galanko J, Simmons JG, Woosley JT, Sandler RS. Insulin resistance, apoptosis, and colorectal adenoma risk. Cancer Epidemiol Biomarkers Prev. 2005;14: 2076-81. https://doi.org/10.1158/1055-9965.Epi-05-0239

324. Sandhu MS, Dunger DB, Giovannucci EL. Insulin, insulin-like growth factor-I (IGF-I), IGF binding proteins, their biologic interactions, and colorectal cancer. J Natl Cancer Inst. 2002;94: 972-80. https://doi.org/10.1093/jnci/94.13.972

325. Nam SY, Lee EJ, Kim KR, Cha BS, Song YD, Lim SK, et al. Effect of obesity on total and free insulin-like growth factor (IGF)-1, and their relationship to IGF-binding protein (BP)-1, IGFBP-2, IGFBP-3, insulin, and growth hormone. Int J Obes Relat Metab Disord. 1997;21: 355-9. https://doi.org/10.1038/sj.ijo.0800412

326. Renehan AG, Frystyk J, Flyvbjerg A. Obesity and cancer risk: the role of the insulin–IGF axis. Trends Endocrinol Metab. 2006;17: 328-36. https://doi.org/https://doi.org/10.1016/j.tem.2006.08.006

327. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet. 2004;363: 1346-53. https://doi.org/https://doi.org/10.1016/S0140-6736(04)16044-3

328. Dupont J, Le Roith D. Insulin-like growth factor 1 and oestradiol promote cell proliferation of MCF-7 breast cancer cells: new insights into their synergistic effects. Mol Pathol. 2001;54: 149-54. https://doi.org/10.1136/mp.54.3.149

329. Maor S, Mayer D, Yarden RI, Lee AV, Sarfstein R, Werner H, et al. Estrogen receptor regulates insulin-like growth factor-I receptor gene expression in breast tumor cells: involvement of transcription factor Sp1. J Endocrinol. 2006;191: 605-12. https://doi.org/10.1677/joe.1.07016

330. Perks CM, Holly JMP. IGF Binding Proteins (IGFBPs) and Regulation of Breast Cancer Biology. J Mammary Gland Biol Neoplasia. 2008;13: 455-69. https://doi.org/10.1007/s10911-008-9106-4

331. Khandwala HM, McCutcheon IE, Flyvbjerg A, Friend KE. The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. Endocr Rev. 2000;21: 215-44. https://doi.org/10.1210/edrv.21.3.0399

332. Pandini G, Vigneri R, Costantino A, Frasca F, Ippolito A, Fujita-Yamaguchi Y, et al. Insulin and insulin-like growth factor-I (IGF-I) receptor overexpression in breast cancers leads to insulin/IGF-I hybrid receptor overexpression: evidence for a second mechanism of IGF-I signaling. Clin Cancer Res. 1999;5: 1935-44.

333. Yee D. Insulin-like growth factor receptor inhibitors: baby or the bathwater? J Natl Cancer Inst. 2012;104: 975-81. https://doi.org/10.1093/jnci/djs258

334. Langer CJ, Novello S, Park K, Krzakowski M, Karp DD, Mok T, et al. Randomized, phase III trial of first-line figitumumab in combination with paclitaxel and carboplatin versus paclitaxel and carboplatin alone in patients with advanced non-small-cell lung cancer. J Clin Oncol. 2014;32: 2059-66. https://doi.org/10.1200/jco.2013.54.4932

335. Ryan P, Neven P, Blackwell K, Dirix L, Barrios C, Miller W, Jr, et al. Figitumumab plus exemestane versus exemestane as first-line treatment of postmenopausal hormone receptor-positive advanced breast cancer: A randomized, open-label phase II trial. Cancer Res2011.

336. Watts EL, Perez-Cornago A, Fensom GK, Smith-Byrne K, Noor U, Andrews CD, et al. Circulating insulin-like growth factors and risks of overall, aggressive and early-onset prostate cancer: a collaborative analysis of 20 prospective studies and Mendelian randomization analysis. Int J Epidemiol. 2022. https://doi.org/10.1093/ije/dyac124

337. Cox ME, Gleave ME, Zakikhani M, Bell RH, Piura E, Vickers E, et al. Insulin receptor expression by human prostate cancers. Prostate. 2009;69: 33-40. https://doi.org/10.1002/pros.20852

338. Ahearn TU, Peisch S, Pettersson A, Ebot EM, Zhou CK, Graff RE, et al. Expression of IGF/insulin receptor in prostate cancer tissue and progression to lethal disease. Carcinogenesis. 2018;39: 1431-7. https://doi.org/10.1093/carcin/bgy112

339. Yu EY, Li H, Higano CS, Agarwal N, Pal SK, Alva A, et al. SWOG S0925: A randomized phase II study of androgen deprivation combined with cixutumumab versus androgen deprivation alone in patients with new metastatic hormone-sensitive prostate cancer. J Clin Oncol. 2015;33: 1601-8. https://doi.org/10.1200/jco.2014.59.4127

340. Wu J, Yu E. Insulin-like growth factor receptor-1 (IGF-IR) as a target for prostate cancer therapy. Cancer Metastasis Rev. 2014;33: 607-17. https://doi.org/10.1007/s10555-013-9482-0

341. Fontana L, Adelaiye RM, Rastelli AL, Miles KM, Ciamporcero E, Longo VD, et al. Dietary protein restriction inhibits tumor growth in human xenograft models. Oncotarget. 2013;4: 2451-61. https://doi.org/10.18632/oncotarget.1586

342. UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007. https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf. Accessed 25 June 2022.

343. UK Biobank. Background. https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us. Accessed 19 January 2022.

344. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35: 1381-95. https://doi.org/10.1249/01.Mss.0000078924.61453.Fb

345. UK Biobank. Anthropometry 2014. http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/Anthropometry.pdf. Accessed May 30 2020.

346. UK Biobank. Biomarker assay quality procedures: approaches used to minimise systematic and random errors (and the wider epidemiological implications) 2019. 1.2:[https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/biomarker_issues.pdf. Accessed 20 June 2022.

347. Allen N, Arnold M, Parish S, Hill M, Sheard S, Callen H, et al. Approaches to minimising the epidemiological impact of sources of systematic and random variation that may affect biochemistry assay data in UK Biobank. Wellcome Open Res. 2021;5. https://doi.org/10.12688/wellcomeopenres.16171.2

348. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med. 2015;3: 769-81. https://doi.org/10.1016/s2213-2600(15)00283-0

349. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203-9. https://doi.org/10.1038/s41586-018-0579-z

350. UK Biobank. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource 2015. https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping_qc.pdf. Accessed March 20 2022.

351. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. Am J Epidemiol. 1999;150: 341-53. https://doi.org/10.1093/oxfordjournals.aje.a010013

352. UK Biobank. Hospital inpatient data 2020. https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/HospitalEpisodeStatistics.pdf. Accessed.

353. UK Biobank. Cancer data: linkage from national cancer registries https://biobank.ndph.ox.ac.uk/crystal/crystal/docs/CancerLinkage.pdf. Accessed 15 September 2020.

354. UK Biobank. Primary care linked data 2019. https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf. Accessed 20 May 2022.

355. UK Biobank. Mortality data: linkage to death registries 2020. https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/DeathLinkage.pdf. Accessed 20 May 2022.

356. World Health Organization. (2004). ICD-10: international statistical classification of diseases and related health problems: tenth revision, 2nd ed. World Health Organization. https://apps.who.int/iris/handle/10665/42980. Accessed 20 May 2022.

357. UK Biobank. 24-hour dietary recall questionnaire. 2012. https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/DietWebQ.pdf. Accessed March 20 2022.

358. Liu B, Young H, Crowe FL, Benson VS, Spencer EA, Key TJ, et al. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. Public Health Nutr. 2011;14: 1998-2005. https://doi.org/10.1017/S1368980011000942

359. Greenwood DC, Hardie LJ, Frost GS, Alwan NA, Bradbury KE, Carter M, et al. Validation of the Oxford WebQ online 24-hour dietary questionnaire using biomarkers. Am J Epidemiol. 2019;188: 1858-67. https://doi.org/10.1093/aje/kwz165

360. Willett W. Nutritional Epidemiology. 3rd edition ed: Oxford University Press 2013.

361. Perez-Cornago A, Pollard Z, Young H, van Uden M, Andrews C, Piernas C, et al. Description of the updated nutrition calculation of the Oxford WebQ questionnaire and comparison with the previous version among 207,144 participants in UK Biobank. Eur J Nutr. 2021;60: 4019-30. https://doi.org/10.1007/s00394-021-02558-4

362. Piernas C, Perez-Cornago A, Gao M, Young H, Pollard Z, Mulligan A, et al. Describing a new food group classification system for UK biobank: analysis of food groups and sources of macro- and micronutrients in 208,200 participants. Eur J Nutr. 2021;60: 2879-90. https://doi.org/10.1007/s00394-021-02535-x

363. National Diet and Nutrition Survey Rolling Programme (NDNS RP) Years 5-6 (2012/13-2013/14): User Guide. http://doc.ukdataservice.ac.uk/doc/6533/mrdoc/pdf/6533_ndns_yrs5-6_uk_user_guide.pdf. Accessed 20 May 2022.

364. Crawley H. Food portion sizes: HM Stationery Office 1988.

365. Fellows PJ. Properties of food processing. In: Fellows PJ, editor. Food Processing Technology (Third Edition): Woodhead Publishing; 2009. p. 11-95.

366. Norat T, Dossus L, Rinaldi S, Overvad K, Grønbæk H, Tjønneland A, et al. Diet, serum insulin-like growth factor-I and IGF-binding protein-3 in European women. Eur J Clin Nutr. 2007;61: 91-8. https://doi.org/10.1038/sj.ejcn.1602494

367. Young NJ, Metcalfe C, Gunnell D, Rowlands MA, Lane JA, Gilbert R, et al. A cross-sectional analysis of the association between diet and insulin-like growth factor (IGF)-I, IGF-II, IGF-binding protein (IGFBP)-2, and IGFBP-3 in men in the United Kingdom. Cancer Causes Control. 2012;23: 907-17. https://doi.org/10.1007/s10552-012-9961-6

368. Trombetti A, Carrier E, Perroud A, Lang F, Herrmann FR, Rizzoli R. Influence of a fermented protein-fortified dairy product on serum insulin-like growth factor-I in women with anorexia nervosa: A randomized controlled trial. Clin Nutr. 2016;35: 1032-8. https://doi.org/10.1016/j.clnu.2015.10.014

369. Harrison S, Lennon R, Holly J, Higgins JPT, Gardner M, Perks C, et al. Does milk intake promote prostate cancer initiation or progression via effects on insulin-like growth factors (IGFs)? A systematic review and meta-analysis. Cancer Causes Control. 2017;28: 497-528. https://doi.org/10.1007/s10552-017-0883-1

370. Romo Ventura E, Konigorski S, Rohrmann S, Schneider H, Stalla GK, Pischon T, et al. Association of dietary intake of milk and dairy products with blood concentrations of insulin-like growth factor 1 (IGF-1) in Bavarian adults. Eur J Nutr. 2019;59: 1413-20. https://doi.org/10.1007/s00394-019-01994-7

371. UK Biobank. UK Biobank Biomarker Project - Companion Document to Accompany Serum Biomarker Data 2019. https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf. Accessed April 20 2020.

372. UK Biobank. Repeat assessment. https://biobank.ctsu.ox.ac.uk/~bbdatan/Repeat_assessment_doc_v1.0.pdf. Accessed March 1 2022.

373. Matt Arnold. Jasper: Jasper makes plots. R package version 2-266.

https://github.com/arnhew99/Jasper. 2020.

374. Larsson SC, Wolk K, Brismar K, Wolk A. Association of diet with serum insulin-like growth factor I in middle-aged and elderly men. Am J Clin Nutr. 2005;81: 1163-7. https://doi.org/10.1093/ajcn/81.5.1163

375. Wan X, Wang S, Xu J, Zhuang L, Xing K, Zhang M, et al. Dietary protein-induced hepatic IGF-1 secretion mediated by PPARγ activation. PLoS One. 2017;12: e0173174. https://doi.org/10.1371/journal.pone.0173174

376. Gunnell D, Oliver SE, Peters TJ, Donovan JL, Persad R, Maynard M, et al. Are diet–prostate cancer associations mediated by the IGF axis? A cross-sectional analysis of diet, IGF-1 and IGFBP-3 in healthy middle-aged men. Br J Cancer. 2003;88: 1682-6. https://doi.org/10.1038/sj.bjc.6600946

377. Baibas N, Bamia C, Vassilopoulou E, Sdrolias J, Trichopoulou A, Trichopoulos D. Dietary and lifestyle factors in relation to plasma insulin-like growth factor I in a general population sample. Eur J Cancer Prev. 2003;12: 229-34. https://doi.org/10.1097/00008469-200306000-00010

378. Damsgaard CT, Mølgaard C, Matthiessen J, Gyldenløve SN, Lauritzen L. The effects of n-3 long-chain polyunsaturated fatty acids on bone formation and growth factors in adolescent boys. Pediatr Res. 2012;71: 713-9. https://doi.org/10.1038/pr.2012.28

379. U.S. Department of Agriculture. Agricultural Research Service. FoodData Central, 2019. fdc.nal.usda.gov. Accessed 15 April 2022.

380. Maruyama K, Iso H, Ito Y, Watanabe Y, Inaba Y, Tajima K, et al. Associations of food and nutrient intakes with serum IGF-I, IGF-II, IGFBP-3, TGF-b1, total SOD activity and sFas levels among middle-aged Japanese: the Japan Collaborative Cohort study. Asian Pac J Cancer Prev. 2009;10 Suppl: 7-22.

381. Tran CD, Diorio C, Bérubé S, Pollak M, Brisson J. Relation of insulin-like growth factor (IGF) I and IGF-binding protein 3 concentrations with intakes of fruit, vegetables, and antioxidants. Am J Clin Nutr. 2006;84: 1518-26. https://doi.org/10.1093/ajcn/84.6.1518

382. Gorissen SHM, Crombag JJR, Senden JMG, Waterval WAH, Bierau J, Verdijk LB, et al. Protein content and amino acid composition of commercially available plant-based protein isolates. Amino Acids. 2018;50: 1685-95. https://doi.org/10.1007/s00726-018-2640-5

383. Bradbury KE, Young HJ, Guo W, Key TJ. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. J Nutr Sci. 2018;7: e6. https://doi.org/10.1017/jns.2017.66

384. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol. 2017;186: 1026-34. https://doi.org/10.1093/aje/kwx246

385. Farina EK, Kiel DP, Roubenoff R, Schaefer EJ, Cupples LA, Tucker KL. Protective effects of fish intake and interactive effects of long-chain polyunsaturated fatty acid intakes on hip bone mineral density in older adults: the Framingham Osteoporosis Study. Am J Clin Nutr. 2011;93: 1142-51. https://doi.org/10.3945/ajcn.110.005926

386. World Cancer Research Fund/ American Institute for Cancer Research. Continuous Update Project Expert Report. Diet, nutrition, physical activity and cancer risk. 2018.

387. Khalil DA, Lucas EA, Juma S, Smith BJ, Payton ME, Arjmandi BH. Soy protein supplementation increases serum insulin-like growth factor-I in young and old men but does not affect markers of bone metabolism. J Nutr. 2002;132: 2605-8. https://doi.org/10.1093/jn/132.9.2605

388. Tsilidis KK, Travis RC, Appleby PN, Allen NE, Lindström S, Albanes D, et al. Insulin-like growth factor pathway genes and blood concentrations, dietary protein and risk of prostate cancer in the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). Int J Cancer. 2013;133: 495-504. https://doi.org/10.1002/ijc.28042

389. Colangelo LA, Chiu BCH, Liu K, Kopp PA, Gann PH, Gapstur SM. IGF-1, IGFBP-3, and nutritional factors in young black and white men: the CARDIA male hormone study. Nutr Cancer. 2005;53: 57-64. https://doi.org/10.1207/s15327914nc5301_7

390. Fontana L, Weiss EP, Villareal DT, Klein S, Holloszy JO. Long-term effects of calorie or protein restriction on serum IGF-1 and IGFBP-3 concentration in humans. Aging Cell. 2008;7: 681-7. https://doi.org/10.1111/j.1474-9726.2008.00417.x

391. Papier K, Tong TY, Appleby PN, Bradbury KE, Fensom GK, Knuppel A, et al. Comparison of major protein-source foods and other food groups in meat-eaters and non-meat-eaters in the EPIC-Oxford cohort. Nutrients. 2019;11: 824. https://doi.org/10.3390/nu11040824

392. Cintineo HP, Arent MA, Antonio J, Arent SM. Effects of protein supplementation on performance and recovery in resistance and endurance training. Front Nutr. 2018;5: 83. https://doi.org/10.3389/fnut.2018.00083

393. Hoppe C, Mølgaard C, Dalum C, Vaag A, Michaelsen KF. Differential effects of casein versus whey on fasting plasma levels of insulin, IGF-1 and IGF-1/IGFBP-3: results from a randomized 7-day supplementation study in prepubertal boys. Eur J Clin Nutr. 2009;63: 1076-83. https://doi.org/10.1038/ejcn.2009.34

394. Melnik BC, John SM, Schmitz G. Milk is not just food but most likely a genetic transfection system activating mTORC1 signaling for postnatal growth. Nutr J. 2013;12: 103-. https://doi.org/10.1186/1475-2891-12-103

395. Um CY, Fedirko V, Flanders WD, Höflich C, Wirthgen E, Bostick RM. Circulating insulin-like growth factor-related biomarkers: correlates and responses to calcium supplementation in colorectal adenoma patients. Mol Carcinog. 2017;56: 2127-34. https://doi.org/10.1002/mc.22669

396. Probst-Hensch NM, Wang H, Goh VH, Seow A, Lee HP, Yu MC. Determinants of circulating insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations in a cohort of Singapore men and women. Cancer Epidemiol Biomarkers Prev. 2003;12: 739-46.

397. Yan J, Charles JF. Gut Microbiota and IGF-1. Calcif Tissue Int. 2018;102: 406-14. https://doi.org/10.1007/s00223-018-0395-3

398. Yan J, Herzog JW, Tsang K, Brennan CA, Bower MA, Garrett WS, et al. Gut microbiota induce IGF-1 and promote bone formation and growth. Proc Natl Acad Sci. 2016: 201607235. https://doi.org/10.1073/pnas.1607235113

399. Ronis MJ, Wands JR, Badger TM, de la Monte SM, Lang CH, Calissendorff J. Alcohol-induced disruption of endocrine signaling. Alcohol Clin Exp Res. 2007;31: 1269-85. https://doi.org/10.1111/j.1530-0277.2007.00436.x

400. Jones JI, Clemmons DR. Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev. 1995;16: 3-34. https://doi.org/10.1210/edrv-16-1-3

401. Galante J, Adamska L, Young A, Young H, Littlejohns TJ, Gallacher J, et al. The acceptability of repeat internet-based hybrid diet assessment of previous 24-h dietary intake: administration of the Oxford WebQ in UK Biobank. Br J Nutr. 2016;115: 681-6. https://doi.org/10.1017/S0007114515004821

402. Allen NE, Key TJ, Appleby PN, Travis RC, Roddam AW, Tjønneland A, et al. Animal foods, protein, calcium and prostate cancer risk: the European Prospective Investigation into Cancer and Nutrition. Br J Cancer. 2008;98: 1574-81. https://doi.org/10.1038/sj.bjc.6604331

403. Schmidt JA, Huybrechts I, Overvad K, Eriksen AK, Tjønneland A, Kaaks R, et al. Protein and amino acid intakes in relation to prostate cancer risk and mortality—A prospective study in the European Prospective Investigation into Cancer and Nutrition. Cancer Med. 2022. https://doi.org/https://doi.org/10.1002/cam4.5289

404. Key TJ, Balkwill A, Bradbury KE, Reeves GK, Kuan AS, Simpson RF, et al. Foods, macronutrients and breast cancer risk in postmenopausal women: a large UK cohort. Int J Epidemiol. 2018;48: 489-500. https://doi.org/10.1093/ije/dyy238

405. Tantamango-Bartley Y, Knutsen SF, Jaceldo-Siegl K, Fan J, Mashchak A, Fraser GE. Independent associations of dairy and calcium intakes with colorectal cancers in the Adventist Health Study-2 cohort. Public Health Nutr. 2017;20: 2577-86. https://doi.org/10.1017/s1368980017001422

406. Pietinen P, Malila N, Virtanen M, Hartman TJ, Tangrea JA, Albanes D, et al. Diet and risk of colorectal cancer in a cohort of Finnish men. Cancer Causes Control. 1999;10: 387-96. https://doi.org/10.1023/a:1008962219408

407. Knuppel A, Fensom GK, Watts EL, Gunter MJ, Murphy N, Papier K, et al. Circulating insulin-like growth factor-I concentrations and risk of 30 cancers: prospective analyses in UK Biobank. Cancer Res. 2020;80: 4014-21. https://doi.org/10.1158/0008-5472.Can-20-1281

408. Vissers LET, Sluijs I, van der Schouw YT, Forouhi NG, Imamura F, Burgess S, et al. Dairy product intake and risk of type 2 diabetes in EPIC-InterAct: a Mendelian randomization study. Diabetes Care. 2019;42: 568-75. https://doi.org/10.2337/dc18-2034

409. Keum N, Aune D, Greenwood DC, Ju W, Giovannucci EL. Calcium intake and colorectal cancer risk: Dose–response meta-analysis of prospective observational studies. Int J Cancer. 2014;135: 1940-8. https://doi.org/https://doi.org/10.1002/ijc.28840

410. Cormick G, Belizán JM. Calcium intake and health. Nutrients. 2019;11. https://doi.org/10.3390/nu11071606

411. Govers MJAP, Termont DSML, Lapré JA, Kleibeuker JH, Vonk RJ, Van der Meer R. Calcium in milk products precipitates intestinal fatty acids and secondary bile acids and thus inhibits colonic cytotoxicity in humans. Cancer Res. 1996;56: 3270-5.

412. Lamprecht SA, Lipkin M. Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms. Nat Rev Cancer. 2003;3: 601-14. https://doi.org/10.1038/nrc1144

413. Pufulete M. Intake of dairy products and risk of colorectal neoplasia. Nutr Res Rev. 2008;21: 56-67. https://doi.org/10.1017/S0954422408035920

414. Larsson SC, Bergkvist L, Rutegård J, Giovannucci E, Wolk A. Calcium and dairy food intakes are inversely associated with colorectal cancer risk in the Cohort of Swedish Men. Am J Clin Nutr. 2006;83: 667-73.

415. Park SY, Murphy SP, Wilkens LR, Nomura AM, Henderson BE, Kolonel LN. Calcium and vitamin D intake and risk of colorectal cancer: the Multiethnic Cohort Study. Am J Epidemiol. 2007;165: 784-93. https://doi.org/10.1093/aje/kwk069

416. McCullough ML, Robertson AS, Rodriguez C, Jacobs EJ, Chao A, Carolyn J, et al. Calcium, vitamin D, dairy products, and risk of colorectal cancer in the Cancer Prevention Study II Nutrition Cohort (United States). Cancer Causes Control. 2003;14: 1-12. https://doi.org/10.1023/a:1022591007673

417. Park Y, Leitzmann MF, Subar AF, Hollenbeck A, Schatzkin A. Dairy food, calcium, and risk of cancer in the NIH-AARP Diet and Health Study. Arch Intern Med. 2009;169: 391-401. https://doi.org/10.1001/archinternmed.2008.578

418. Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G, et al. Consumption of dairy products and colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). PLoS One. 2013;8: e72715. https://doi.org/10.1371/journal.pone.0072715

419. Fraser GE, Jaceldo-Siegl K, Orlich M, Mashchak A, Sirirat R, Knutsen S. Dairy, soy, and risk of breast cancer: those confounded milks. Int J Epidemiol. 2020;49: 1526-37. https://doi.org/10.1093/ije/dyaa007

420. Perez-Cornago A. Commentary: Dairy milk intake and breast cancer risk: does an association exist, and what might be the culprit? Int J Epidemiol. 2020;49: 1537-9. https://doi.org/10.1093/ije/dyaa199

421. Kakkoura MG, Du H, Guo Y, Yu C, Yang L, Pei P, et al. Dairy consumption and risks of total and site-specific cancers in Chinese adults: an 11-year prospective study of 0.5 million people. BMC Med. 2022;20: 134. https://doi.org/10.1186/s12916-022-02330-3

422. Wu Y, Huang R, Wang M, Bernstein L, Bethea TN, Chen C, et al. Dairy foods, calcium, and risk of breast cancer overall and for subtypes defined by estrogen receptor status: a pooled analysis of 21 cohort studies. Am J Clin Nutr. 2021;114: 450-61. https://doi.org/10.1093/ajcn/nqab097

423. Kimball SR, Jefferson LS. Signaling pathways and molecular mechanisms through which branched-chain amino acids mediate translational control of protein synthesis. J Nutr. 2006;136: 227s-31s. https://doi.org/10.1093/jn/136.1.227S

424. Littlejohns TJ, Travis RC, Key TJ, Allen NE. Lifestyle factors and prostate-specific antigen (PSA) testing in UK Biobank: Implications for epidemiological research. Cancer Epidemiol. 2016;45: 40-6. https://doi.org/10.1016/j.canep.2016.09.010

425. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. BMJ. 2020;368: m131. https://doi.org/10.1136/bmj.m131

426. Liese AD, Schulz M, Fang F, Wolever TM, D'Agostino RB, Jr., Sparks KC, et al. Dietary glycemic index and glycemic load, carbohydrate and fiber intake, and measures of insulin sensitivity, secretion, and adiposity in the Insulin Resistance Atherosclerosis Study. Diabetes Care. 2005;28: 2832-8. https://doi.org/10.2337/diacare.28.12.2832

427. Macdonald IA. A review of recent evidence relating to sugars, insulin resistance and diabetes. Eur J Nutr. 2016;55: 17-23. https://doi.org/10.1007/s00394-016-1340-8

428. Te Morenga L, Mallard S, Mann J. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ. 2013;346: e7492. https://doi.org/10.1136/bmj.e7492

429. So D, Whelan K, Rossi M, Morrison M, Holtmann G, Kelly JT, et al. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. Am J Clin Nutr. 2018;107: 965-83. https://doi.org/10.1093/ajcn/nqy041

430. Sengupta S, Muir JG, Gibson PR. Does butyrate protect from colorectal cancer? J Gastroenterol Hepatol. 2006;21: 209-18. https://doi.org/10.1111/j.1440-1746.2006.04213.x

431. Hajjar R, Richard CS, Santos MM. The role of butyrate in surgical and oncological outcomes in colorectal cancer. Am J Physiol Gastrointest Liver Physiol. 2021;320: G601-G8. https://doi.org/10.1152/ajpgi.00316.2020

432. Mirzaei R, Afaghi A, Babakhani S, Sohrabi MR, Hosseini-Fard SR, Babolhavaeji K, et al. Role of microbiota-derived short-chain fatty acids in cancer development and prevention. Biomed Pharmacother. 2021;139: 111619. https://doi.org/https://doi.org/10.1016/j.biopha.2021.111619

433. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53: 156-65. https://doi.org/10.1038/s41588-020-00763-1

434. Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sánchez S, Chen L, et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet. 2022;54: 143-51. https://doi.org/10.1038/s41588-021-00992-y

435. Englyst HN, Quigley ME, Hudson GJ. Determination of dietary fibre as non-starch polysaccharides with gas-liquid chromatographic, high-performance liquid chromatographic or spectrophotometric measurement of constituent sugars. Analyst. 1994;119: 1497-509. https://doi.org/10.1039/an9941901497

436. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019;51: 600-5. https://doi.org/10.1038/s41588-019-0350-x

437. Lunn M, McNeil D. Applying cox regression to competing risks. Biometrics. 1995;51: 524-32. https://doi.org/10.2307/2532940

438. Vanderweele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment interaction studies. Am J Epidemiol. 2013;178: 144-52. https://doi.org/10.1093/aje/kws439

439. Bradbury KE, Murphy N, Key TJ. Diet and colorectal cancer in UK Biobank: a prospective study. Int J Epidemiol. 2019;49: 246-58. https://doi.org/10.1093/ije/dyz064

440. Finglas PM RM, Pinchen HM, et al. McCance and Widdowson's The Compostion of Foods, Seventh Summary Edition. Cambridge: Royal Society of Chemistry. 2015.

441. Aune D, Lau R, Chan DSM, Vieira R, Greenwood DC, Kampman E, et al. Dairy products and colorectal cancer risk: a systematic review and meta-analysis of cohort studies. Ann Oncol. 2012;23: 37-45. https://doi.org/10.1093/annonc/mdr269

442. Chazelas E, Srour B, Desmetz E, Kesse-Guyot E, Julia C, Deschamps V, et al. Sugary drink consumption and risk of cancer: results from NutriNet-Santé prospective cohort. BMJ. 2019;366: l2408. https://doi.org/10.1136/bmj.l2408

443. Yuan C, Giovannucci EL. Epidemiological evidence for dietary sugars and colorectal cancer. Curr Colorectal Cancer Rep. 2020;16: 55-63. https://doi.org/10.1007/s11888-020-00453-4

444. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16: 713-32. https://doi.org/10.1038/s41575-019-0189-8

445. Joh H-K, Lee DH, Hur J, Nimptsch K, Chang Y, Joung H, et al. Simple sugar and sugar-sweetened beverage intake during adolescence and risk of colorectal cancer precursors. Gastroenterology. 2021;161: 128-42.e20. https://doi.org/10.1053/j.gastro.2021.03.028

446. Yuan C, Joh H-K, Wang Q-L, Zhang Y, Smith-Warner SA, Wang M, et al. Sugar-sweetened beverage and sugar consumption and colorectal cancer incidence and mortality according to anatomic subs. Am J Clin Nutr. 2022;115: 1481-9. https://doi.org/10.1093/ajcn/nqac040

447. Huang J, Pan G, Jiang H, Li W, Dong J, Zhang H, et al. A meta-analysis between dietary carbohydrate intake and colorectal cancer risk: evidence from 17 observational studies. Biosci Rep. 2017;37. https://doi.org/10.1042/bsr20160553

448. McCarl M, Harnack L, Limburg PJ, Anderson KE, Folsom AR. Incidence of colorectal cancer in relation to glycemic index and load in a cohort of women. Cancer Epidemiol Biomarkers Prev. 2006;15: 892-6. https://doi.org/10.1158/1055-9965.Epi-05-0700

449. Schwingshackl L, Schwedhelm C, Hoffmann G, Knüppel S, Laure Preterre A, Iqbal K, et al. Food groups and risk of colorectal cancer. Int J Cancer. 2018;142: 1748-58. https://doi.org/https://doi.org/10.1002/ijc.31198

450. Holscher HD. Dietary fiber and prebiotics and the gastrointestinal microbiota. Gut Microbes. 2017;8: 172-84. https://doi.org/10.1080/19490976.2017.1290756

451. Sonnenburg JL, Bäckhed F. Diet–microbiota interactions as moderators of human metabolism. Nature. 2016;535: 56-64. https://doi.org/10.1038/nature18846

452. Goodrich Julia K, Waters Jillian L, Poole Angela C, Sutter Jessica L, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159: 789-99. https://doi.org/10.1016/j.cell.2014.09.053

453. Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol Rev. 2001;81: 1031-64. https://doi.org/10.1152/physrev.2001.81.3.1031

454. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. Science. 2013;341: 569-73. https://doi.org/10.1126/science.1241165

455. Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. Cell. 2016;165: 1332-45. https://doi.org/10.1016/j.cell.2016.05.041

456. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 2012;6: 320-9. https://doi.org/10.1038/ismej.2011.109

457. Martínez I, Lattimer JM, Hubach KL, Case JA, Yang J, Weber CG, et al. Gut microbiome composition is linked to whole grain-induced immunological improvements. ISME J. 2013;7: 269-80. https://doi.org/10.1038/ismej.2012.104

458. Vitaglione P, Mennella I, Ferracane R, Rivellese AA, Giacco R, Ercolini D, et al. Whole-grain wheat consumption reduces inflammation in a randomized controlled trial on overweight and obese subjects with unhealthy dietary and lifestyle behaviors: role of polyphenols bound to cereal dietary fiber. Am J Clin Nutr. 2015;101: 251-61. https://doi.org/10.3945/ajcn.114.088120

459. Vuholm S, Nielsen DS, Iversen KN, Suhr J, Westermann P, Krych L, et al. Whole-grain rye and wheat affect some markers of gut health without altering the fecal microbiota in healthy overweight adults: a 6-week randomized trial. J Nutr. 2017;147: 2067-75. https://doi.org/10.3945/jn.117.250647

460. Roager HM, Vogt JK, Kristensen M, Hansen LBS, Ibrügger S, Mærkedahl RB, et al. Whole grain-rich diet reduces body weight and systemic low-grade inflammation without inducing major changes of the gut microbiome: a randomised cross-over trial. Gut. 2019;68: 83-93. https://doi.org/10.1136/gutjnl-2017-314786

461. Donohoe DR, Holley D, Collins LB, Montgomery SA, Whitmore AC, Hillhouse A, et al. A gnotobiotic mouse model demonstrates that dietary fiber protects against colorectal tumorigenesis in a microbiota- and butyrate-dependent manner. Cancer Discov. 2014;4: 1387-97. https://doi.org/10.1158/2159-8290.Cd-14-0501

462. Atkinson FS, Brand-Miller JC, Foster-Powell K, Buyken AE, Goletzke J. International tables of glycemic index and glycemic load values 2021: a systematic review. The American Journal of Clinical Nutrition. 2021;114: 1625-32. https://doi.org/10.1093/ajcn/nqab233

463. Cancer Research UK. Cancer incidence statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared. Accessed 8 January 2021.

464. Brown KF, Rumgay H, Dunlop C, Ryan M, Quartly F, Cox A, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br J Cancer. 2018;118: 1130-41. https://doi.org/10.1038/s41416-018-0029-6

465. Key TJ, Bradbury KE, Perez-Cornago A, Sinha R, Tsilidis KK, Tsugane S. Diet, nutrition, and cancer risk: what do we know and what is the way forward? BMJ. 2020;368: m511. https://doi.org/10.1136/bmj.m511

466. Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. J Nutr. 2020;150: 663-71. https://doi.org/10.1093/jn/nxz268

467. Bradbury KE, Tong TYN, Key TJ. Dietary intake of high-protein foods and other major foods in meat-eaters, poultry-eaters, fish-eaters, vegetarians, and vegans in UK Biobank. Nutrients. 2017;9: 1317.

468. Cade JE, Taylor EF, Burley VJ, Greenwood DC. Common dietary patterns and risk of breast cancer: analysis from the United Kingdom Women's Cohort Study. Nutr Cancer. 2010;62: 300-6. https://doi.org/10.1080/01635580903441246

469. Appleby PN, Key TJ. The long-term health of vegetarians and vegans. Proc Nutr Soc. 2016;75: 287-93. https://doi.org/10.1017/S0029665115004334

470. Tong TY, Key TJ, Sobiecki JG, Bradbury KE. Anthropometric and physiologic characteristics in white and British Indian vegetarians and nonvegetarians in the UK Biobank. Am J Clin Nutr. 2018;107: 909-20. https://doi.org/10.1093/ajcn/nqy042

471. Rosell M, Appleby P, Spencer E, Key T. Weight gain over 5 years in 21 966 meat-eating, fish-eating, vegetarian, and vegan men and women in EPIC-Oxford. Int J Obes. 2006;30: 1389-96. https://doi.org/10.1038/sj.ijo.0803305

472. Tong T.Y.N., Perez-Cornago A., Bradbury K.E., Key T.J. Biomarker concentrations in white and British Indian vegetarians and non-vegetarians in the UK Biobank. J Nutr. 2021. https://doi.org/https://doi.org/10.1093/jn/nxab192

473. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab. 1999;84: 3666-72. https://doi.org/10.1210/jcem.84.10.6079

474. UK Biobank. Analysis of biological samples. https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/biomarker-data. Accessed March 25 2021.

475. Parish S, Peto R, Palmer A, Clarke R, Lewington S, Offer A, et al. The joint effects of apolipoprotein B, apolipoprotein A1, LDL cholesterol, and HDL cholesterol on risk: 3510 cases of acute myocardial infarction and 9805 controls. Eur Heart J. 2009;30: 2137-46. https://doi.org/10.1093/eurheartj/ehp221

476. Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. Stat Med. 1991;10: 1025-35. https://doi.org/10.1002/sim.4780100703

477. Tong TYN, Appleby PN, Bradbury KE, Perez-Cornago A, Travis RC, Clarke R, et al. Risks of ischaemic heart disease and stroke in meat eaters, fish eaters, and vegetarians over 18 years of follow-up: results from the prospective EPIC-Oxford study. BMJ. 2019;366: l4897. https://doi.org/10.1136/bmj.l4897

478. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. Stat Med. 2013;32: 4567-80. https://doi.org/https://doi.org/10.1002/sim.5864

479. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen Tchetgen EJ. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. Am J Epidemiol. 2015;181: 349-56. https://doi.org/10.1093/aje/kwu278

480. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health. 2016;37: 17-32. https://doi.org/10.1146/annurev-publhealth-032315-021402

481. VanderWeele TJ, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiologic methods. 2014;2: 95-115. https://doi.org/10.1515/em-2012-0010

482. Ortega LS, Bradbury KE, Cross AJ, Morris JS, Gunter MJ, Murphy N. A prospective investigation of body size, body fat composition and colorectal cancer risk in the UK Biobank. Sci Rep. 2017;7: 17807. https://doi.org/10.1038/s41598-017-17997-5

483. Gilsing AM, Weijenberg MP, Goldbohm RA, Dagnelie PC, van den Brandt PA, Schouten LJ. Vegetarianism, low meat consumption and the risk of lung, postmenopausal breast and prostate cancer in a population-based cohort study. Eur J Clin Nutr. 2016;70: 723-9. https://doi.org/10.1038/ejcn.2016.25

484. Ibrayev Y, Oda K, Fraser GE, Knutsen SF. Utilization of prostate cancer screening according to dietary patterns and other demographic variables. The Adventist Health Study-2. J Cancer. 2013;4: 416-26. https://doi.org/10.7150/jca.6442

485. Voskuil DW, Vrieling A, van't Veer LJ, Kampman E, Rookus MA. The insulin-like growth factor system in cancer prevention: potential of dietary intervention strategies. Cancer Epidemiol Biomarkers Prev. 2005;14: 195-203. https://doi.org/10.1158/1055-9965.195.14.1

486. Ludwig DS, Aronne LJ, Astrup A, de Cabo R, Cantley LC, Friedman MI, et al. The carbohydrate-insulin model: a physiological perspective on the obesity pandemic. Am J Clin Nutr. 2021;114: 1873-85. https://doi.org/10.1093/ajcn/nqab270

487. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 2017;47: 226-35. https://doi.org/10.1093/ije/dyx206

488. Mansournia MA, Altman DG. Inverse probability weighting. BMJ. 2016;352: i189. https://doi.org/10.1136/bmj.i189

489. van Alten S, Domingue BW, Galama T, Marees AT. Reweighting the UK Biobank to reflect its underlying sampling population substantially reduces pervasive selection bias due to volunteering. medRxiv. 2022: 2022.05.16.22275048. https://doi.org/10.1101/2022.05.16.22275048

490. Schatzkin A, Subar AF, Moore S, Park Y, Potischman N, Thompson FE, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). Cancer Epidemiol Biomarkers Prev. 2009;18: 1026-32. https://doi.org/10.1158/1055-9965.Epi-08-1129

491. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. Am J Epidemiol. 2012;175: 340-7. https://doi.org/10.1093/aje/kwr317

492. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. Am J Epidemiol. 2014;180: 172-88. https://doi.org/10.1093/aje/kwu116

493. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol. 2003;158: 1-13. https://doi.org/10.1093/aje/kwg092

494. Arab L, Wesseling-Perry K, Jardack P, Henry J, Winter A. Eight self-administered 24-hour dietary recalls using the internet are feasible in African Americans and Whites: the energetics study. J Am Diet Assoc. 2010;110: 857-64. https://doi.org/10.1016/j.jada.2010.03.024

495. Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler WV, et al. The US department of agriculture automated multiple-pass method reduces bias in the collection of energy intakes. Am J Clin Nutr. 2008;88: 324-32. https://doi.org/10.1093/ajcn/88.2.324

496. Neuhouser ML, Tinker L, Shaw PA, Schoeller D, Bingham SA, Horn LV, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. Am J Epidemiol. 2008;167: 1247-59. https://doi.org/10.1093/aje/kwn026

497. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SA, Caan B, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. Am J Epidemiol. 2011;174: 591-603. https://doi.org/10.1093/aje/kwr140

498. Wark PA, Hardie LJ, Frost GS, Alwan NA, Carter M, Elliott P, et al. Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. BMC Med. 2018;16: 136. https://doi.org/10.1186/s12916-018-1113-8

499. Arhi CS, Bottle A, Burns EM, Clarke JM, Aylin P, Ziprin P, et al. Comparison of cancer diagnosis recording between the Clinical Practice Research Datalink, Cancer Registry and Hospital Episodes Statistics. Cancer Epidemiol. 2018;57: 148-57. https://doi.org/https://doi.org/10.1016/j.canep.2018.08.009

500. Mozaffarian D, Rosenberg I, Uauy R. History of modern nutrition science—implications for current research, dietary guidelines, and food policy. BMJ. 2018;361: k2392. https://doi.org/10.1136/bmj.k2392

501. Mayne ST, Playdon MC, Rock CL. Diet, nutrition, and cancer: past, present and future. Nat Rev Clin Oncol. 2016;13: 504-15. https://doi.org/10.1038/nrclinonc.2016.24

502. Amuna P, Zotor FB. Epidemiological and nutrition transition in developing countries: impact on human health and development: The epidemiological and nutrition transition in developing countries: evolving trends and their impact in public health and human development. Proc Nutr Soc. 2008;67: 82-90. https://doi.org/10.1017/S0029665108006058

503. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51: 584-91. https://doi.org/10.1038/s41588-019-0379-x

504. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet. 2012;112: 1134-7. https://doi.org/10.1016/j.jand.2012.04.016

505. Booker CL, Harding S, Benzeval M. A systematic review of the effect of retention methods in population-based cohort studies. BMC Public Health. 2011;11: 249. https://doi.org/10.1186/1471-2458-11-249

506. Aune D, Lau R, Chan DS, Vieira R, Greenwood DC, Kampman E, et al. Nonlinear reduction in risk for colorectal cancer by fruit and vegetable intake based on meta-analysis of prospective studies. Gastroenterology. 2011;141: 106-18. https://doi.org/10.1053/j.gastro.2011.04.013

507. Mahabir S, Aagaard K, Anderson LM, Herceg Z, Hiatt RA, Hoover RN, et al. Challenges and opportunities in research on early-life events/exposures and cancer development later in life. Cancer Causes Control. 2012;23: 983-90. https://doi.org/10.1007/s10552-012-9962-5

508. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2017;45: 1866-86. https://doi.org/10.1093/ije/dyw314

509. Merino J, Tobias DK. The unique challenges of studying the genetics of diet and nutrition. Nat Med. 2022;28: 221-2. https://doi.org/10.1038/s41591-021-01626-w

510. Vissers LET, Sluijs I, Burgess S, Forouhi NG, Freisling H, Imamura F, et al. Milk intake and incident stroke and CHD in populations of European descent: a Mendelian randomisation study. Br J Nutr. 2021: 1-9. https://doi.org/10.1017/S0007114521004244

511. Rosoff DB, Davey Smith G, Mehta N, Clarke T-K, Lohoff FW. Evaluating the relationship between alcohol consumption, tobacco use, and cardiovascular disease: A multivariable Mendelian randomization study. PLoS Med. 2020;17: e1003410. https://doi.org/10.1371/journal.pmed.1003410

512. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362: k601. https://doi.org/10.1136/bmj.k601

513. Jiang L, Yang K-h, Tian J-h, Guan Q-l, Yao N, Cao N, et al. Efficacy of antioxidant vitamins and selenium supplement in prostate cancer prevention: a meta-analysis of randomized controlled trials. Nutr Cancer. 2010;62: 719-27. https://doi.org/10.1080/01635581.2010.494335

514. Vollset SE, Clarke R, Lewington S, Ebbing M, Halsey J, Lonn E, et al. Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50 000 individuals. Lancet. 2013;381: 1029-36. https://doi.org/10.1016/s0140-6736(12)62001-7

515. Lippman SM, Klein EA, Goodman PJ, Lucia MS, Thompson IM, Ford LG, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA. 2009;301: 39-51. https://doi.org/10.1001/jama.2008.864

516. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340: c221. https://doi.org/10.1136/bmj.c221

517. Bartsch W. Interrelationships between sex hormone-binding globulin and testosterone, 5 alpha-dihydrotestosterone and oestradiol-17 beta in blood of normal men. Maturitas. 1980;2: 109-18. https://doi.org/10.1016/0378-5122(80)90044-4