University of Helsinki
Dissertationes Universitatis Helsingiensis 89/2026

Department of Public Health
Finnish Institute for Health and Welfare
Finland

Doctoral Programme in Population Health
Faculty of Medicine
University of Helsinki
Finland

Whole grain intake, healthy lifestyles and the gut microbiome in colorectal cancer risk

Rilla Tammi

DOCTORAL DISSERTATION
To be presented for public discussion with the permission of the Faculty of Medicine of the University of Helsinki, in Lecture Room 302, Athena Building, on the 20th of March 2026, at 13 o'clock.

Helsinki 2026

Supervisors

Satu Männistö, Docent, PhD
Department of Public Health
Finnish Institute for Health and Welfare
Helsinki, Finland

Niina Kaartinen, Docent, PhD
Department of Public Health
Finnish Institute for Health and Welfare
Helsinki, Finland

Reviewers

Jyrki Virtanen, Docent, PhD
Institute of Public Health and Clinical Nutrition
University of Eastern Finland
Kuopio, Finland

Cecilie Kyrø, PhD
Danish Cancer Institute
Copenhagen, Denmark

Opponent

Guri Skeie, Professor, PhD
Department of Community Medicine
Arctic University of Norway
Tromsø, Norway

Publisher: University of Helsinki
Series: Dissertationes Universitatis Helsingiensis 89/2026

ISBN 978-952-84-2049-1 (print)
ISBN 978-952-84-2048-4 (online)
ISSN 2954-2898 (print)
ISSN 2954-2952 (online)

PunaMusta, Joensuu 2026

Abstract

Colorectal cancer (CRC) is a major contributor to the cancer burden in Finland and globally, and its aetiology is closely related to unhealthy lifestyles. Of dietary factors, high consumption of red and processed meat and low consumption of plant-based foods, particularly whole grains, have been linked to a higher CRC risk. However, associations between dietary shifts to more plant-based diets and CRC risk remain underexplored. The mechanisms underlying the associations between lifestyle factors and CRC risk, including the role of the gut microbiome in them, are also poorly understood. Furthermore, of CRC-related dietary factors, research on whole grain intake is constrained by methodological challenges reflected in discrepancies in the observed health associations.

The primary aim of the four sub-studies (I–IV) in this thesis was to examine CRC risk and its associations with diet and lifestyle, as well as the role of the gut microbiome in these associations. Of individual CRC risk factors, this thesis focused on whole grain intake, the methodological challenges in its estimation, and its associations with chronic disease risk factors.

This thesis was based on seven study samples of Finnish adults (n=1228–26 915), with five samples from population-based health-examination studies (Health 2000, FinDiet 2002, DILGOM 2007, FinRisk 2012, and FinHealth 2017) and two samples comprised of middle-aged smoking men (ATBC) and older adults born in the Finnish capital area (HBCS). Comprehensive health, lifestyle, and background data were collected through self-administered questionnaires and health examinations. Dietary data was gathered through a food frequency questionnaire (FFQ) or a 48-hour dietary recall. To estimate whole grain intake, a whole grain database was compiled within the Finnish Food Composition Database Fineli®. Diet quality was assessed with the Baltic Sea Diet Score. Data on CRC cases (n=1120) was obtained from the Finnish Cancer Registry. CRC-related lifestyles were modelled with a new CRC lifestyle index, with lower index points reflecting higher-risk lifestyles. The gut microbiome was analysed from faecal samples using metagenomic sequencing. Besides traditional statistical methods, two-stage meta-analysis, Cox regression, substitution analysis and microbiome-specific analysis methods were used.

Of the five potential surrogate measures studied, the combined consumption of rye, oat and barley showed the strongest correlation and highest concordance in cross-classification with whole grain intake in Finnish adults (Sub-study I). Thus, it was used to estimate whole grain intake in the earlier study samples where direct whole grain intake estimation was not feasible (sub-studies III and IV).

Whole grain intake was associated with a lower body mass index, waist circumference and serum total cholesterol concentration in men, whereas no significant associations were observed with the other chronic disease risk factors or in women (II). These associations

were not modified by diet quality, although participants with higher whole grain intake had overall healthier diets.

The modelled partial substitutions of red meat (100 g/week) or processed meat (50 g/week) with vegetables or fruits were associated with 1–3% reductions in CRC risk (III). The substitutions with whole grains were associated with 4–8% risk reductions in participants with whole grain intake below the population median.

Each one-point increase in the CRC lifestyle index was associated with a 19% decrease in CRC risk (IV part A). The index was associated with compositional differences in the gut microbiome, higher-risk lifestyles being associated with lower microbial diversity and a higher relative abundance of several bacterial species previously linked to CRC (part B).

To summarise, this thesis evaluated for the first time the reliability of whole grain surrogate measures and identified the combined consumption of rye, oat and barley a suitable whole grain surrogate in epidemiological research of Finnish adults. Prior research on the associations between whole grain intake and chronic disease risk factors was extended by showing that diet quality did not modify their relationships. The thesis results also demonstrated the potential of dietary shifts to more plant-based diets to lower the population burden of CRC. Finally, using a novel approach, the associations between CRC-related lifestyles and the gut microbiome were explored, the results advancing understanding of microbial traits that may precede CRC onset and play a role in the initiation of its pathogenesis.

Overall, these findings strengthen the knowledge of whole lifestyles and more plant-based diets in improving population health and reducing CRC risk, underscoring the importance of considering factors beyond individual food groups or diet alone. As lifestyle plays a key role in CRC development, greater attention should be given in the future to effective prevention strategies, including identifying high-risk population sub-groups, implementing national nutrition recommendations and introducing policy measures that support healthy lifestyles.

Tiivistelmä

Paksu- ja peräsuolisyövät (engl. colorectal cancer [CRC]) ovat sekä Suomessa että maailmanlaajuisesti yksi yleisimmistä syöpätyypeistä. Epäterveelliset elintavat ovat merkittävä tekijä niiden kehittymisessä; esimerkiksi runsaan punaisen ja prosessoidun lihan kulutuksen ja vähäisen kasviperäisten ruokien, erityisesti täysjyväviljojen, kulutuksen on todettu lisäävän CRC-riskiä. Kasvivoittoisempiin ruokavalioihin siirtymistä suhteessa CRC-riskiin on kuitenkin tutkittu vasta vähän. Lisää tutkimusta tarvitaan myös elintapojen ja CRC-riskin välisten yhteyksien taustamekanismeista, kuten suolistomikrobiston roolista näissä yhteyksissä. CRC:n riskitekijöistä erityisesti täysjyvän saannin tutkimiseen liittyy menetelmällisiä haasteita, jotka voivat osaltaan selittää epäjohdonmukaisia tuloksia täysjyvän ja terveyden välisistä yhteyksistä.

Tämän väitöskirjan neljän osatutkimuksen (I–IV) tavoitteena oli tutkia elintapojen ja CRC-riskin välisiä yhteyksiä sekä suolistomikrobiston merkitystä näissä yhteyksissä. CRC:n riskitekijöistä tutkimus keskittyi erityisesti täysjyvän saantiin, sen arvioinnin haasteisiin ja sen yhteyksiin muihin elintapoihin ja elintapasairauksien riskitekijöihin.

Tutkimuksissa käytettiin seitsemää laajaa suomalaista tutkimusaineistoa (n=1228–26 915), joista viisi koostui kansallisiin terveystarkastustutkimuksiin osallistuneista aikuisista (Terveys 2000, FinRavinto 2002, DILGOM 2007, FinRiski 2012 ja FinTerveys 2017), yksi keski-ikäisistä tupakoivista miehistä (ATBC) ja yksi pääkaupunkiseudulla syntyneistä ikääntyneistä (HBCS). Tietoa tutkittavien terveydestä, elintavoista ja taustamuuttujista kerättiin kyselylomakkeilla ja terveystarkastuksissa. Ruoankäyttöä mitattiin frekvenssityyppisellä ruoankäyttökyselyllä (FFQ) tai 48 tunnin ruoankäyttöhaastattelulla. Täysjyvän saannin arvioimiseksi kansalliseen elintarvikkeiden koostumustietokanta Fineliin® koottiin uusi täysjyvätietokanta. Ruokavalion laatua mitattiin Itämeren ruokavalioindeksillä. Tiedot diagnosoiduista CRC-tapauksista (n=1120) saatiin Suomen Syöpärekisteristä. CRC-riskiin liittyviä elintapoja mallinnettiin uudella CRC-elintapaindeksillä, jossa matalammat pisteet kuvastivat korkeampaan riskiin liittyviä elintapoja. Suolistomikrobisto analysoitiin ulostenäytteistä metagenomisekvensoinnilla. Analyyseissä hyödynnettiin perinteisten tilastomenetelmien lisäksi kaksivaiheista meta-analyysiä, Coxin regressioanalyysiä, substituutioanalyysiä sekä mikrobianalyyseihin kehitettyjä tilastomenetelmiä.

Viidestä tutkitusta täysjyvän saannin vaihtoehtoisesta mittarista rukiin, kauran ja ohran kokonaiskulutus vastasi parhaiten täysjyvän saantia sekä korrelaatioanalyysin että ristiinluokittelun perusteella (osatutkimus I). Näin ollen rukiin, kauran ja ohran kokonaissaantia käytettiin täysjyvän saannin arvioimiseen vanhemmissa tutkimusaineistoissa (osatutkimukset III ja IV), joissa saantia ei voitu suoraan arvioida.

Elintapasairauksien riskitekijöistä täysjyvän saanti oli yhteydessä pienempään painoindeksiin ja vyötärönympärykseen sekä matalampaan seerumin kokonaiskolesterolipitoisuuteen miehillä (II). Naisilla tilastollisesti merkitseviä yhteyksiä

ei havaittu. Vaikka täysjyvän saanti oli yhteydessä parempaan ruokavalion laatuun, ruokavalion laatu ei muokannut täysjyvän saannin ja riskitekijöiden välisiä yhteyksiä.

Punaisen (100 g/viikko) tai prosessoidun (50 g/viikko) lihan osittainen, tilastollisesti mallinnettu korvaaminen kasviksilla tai hedelmillä oli yhteydessä 1–3 % pienempään CRC-riskiin (III). Lihan korvaaminen täysjyvällä oli yhteydessä 4–8 % pienempään CRC-riskiin niillä tutkittavilla, joiden täysjyvän saanti oli alle tutkimusaineiston mediaanin.

Yhden pisteen nousu CRC-indeksissä oli yhteydessä 19 % pienempään CRC-riskiin (IV osa A). Korkeampaan CRC-riskiin liitetyt elintavat (matalammat indeksipisteet) olivat yhteydessä suolistomikrobiston pienempään monimuotoisuuteen sekä useiden aiemmin CRC:n yhdistettyjen bakteerilajien suurempaan suhteelliseen runsauteen (osa B).

Tässä väitöskirjassa tarkasteltiin ensimmäistä kertaa vaihtoehtoisia muuttujia täysjyvän saannin arvioimiseksi, ja osoitettiin rukiin, kauran ja ohran kokonaiskulutuksen olevan luotettava mittari täysjyvän saannille suomalaisessa aikuisväestössä. Aiempaa ymmärrystä täysjyvän saannin ja terveyden välisistä yhteyksistä syvennettiin tutkimalla ruokavalion roolia täysjyvän yhteyksissä elintapasairauksien riskitekijöihin. Tutkimustulokset vahvistivat myös näyttöä kasvivoittoisempiin ruokavalioihin siirtymisen hyödyistä CRC:n tautitaakan vähentämisessä. Lisäksi tässä väitöskirjassa tutkittiin ensimmäistä kertaa CRC:n liittyviä elintapoja kokonaisuutena suhteessa suolistomikrobistoon, tuottaen uutta tietoa mikrobiston ominaisuuksista, jotka voivat osaltaan vaikuttaa CRC:n kehittymiseen.

Kokonaisuutena tämän väitöskirjan tulokset syventävät ymmärrystä terveellisten elintapojen ja kasvivoittoisten ruokavalioiden merkityksestä väestön terveyden edistämisessä ja CRC-riskin pienentämisessä, korostaen tarvetta huomioida elintavat kokonaisuutena. CRC:n ehkäisytoimiin, kuten riskiryhmien tunnistamiseen, kansallisten ravitsemussuositusten toimeenpanoon sekä terveellisiä elintapoja tukeviin politiikkatoimiin tulisi tulevaisuudessa panostaa entistä vahvemmin.

Contents

Abbreviations

48h recall	48-hour dietary recall
AACCI	American Association of Cereal Chemists International
ACS	American Cancer Society
AICR	American Institute for Cancer Research
ANCOM-BC2	Analysis of Compositions of Microbiomes with Bias Correction 2
ATBC	Alpha-Tocopherol Beta-Carotene Cancer Prevention Study
ATC code	Anatomical Therapeutic Chemical code
BMI	Body mass index
CHD	Coronary heart disease
CRC	Colorectal cancer
CRP	C-reactive protein
CVD	Cardiovascular diseases
DILGOM	Dietary, Lifestyle and Genetic Determinants of Obesity and Metabolic Syndrome Study
FFQ	Food frequency questionnaire
FinDiet 2002	FinDiet 2002 Survey
Fineli	Finnish National Food Composition Database
FinHealth 2017	FinHealth 2017 Study
FinRisk	FinRisk Study
FIT	Faecal immunochemical test
HBCS	Helsinki Birth Cohort Study
HCA	Heterocyclic amines
HDL cholesterol	High-density lipoprotein cholesterol
Health 2000	National Health 2000 Health Examination Survey
HR	Hazard ratio
HRT	Hormone-replacement therapy
IARC	International Agency for Research on Cancer
ICD	International Statistical Classification of Diseases and Related Health Problems
IQR	Interquartile range
kcal	Kilocalories
LDL cholesterol	Low-density lipoprotein cholesterol

mBSDS	modified Baltic Sea Diet Score
MJ	Megajoule
NNR23	Nordic Nutrition Recommendations 2023
OR	Odds ratio
PAH	Polycyclic aromatic hydrocarbons
PERMANOVA	Permutational Multivariate Analysis of Variance
PHD	Planetary Health Diet
PPI	Proton pump inhibitor
RCT	Randomised controlled trial
SCFA	Short-chain fatty acid
SD	Standard deviation
SE	Standard error
SSB	Sugar-sweetened beverage
T2D	Type 2 diabetes
TAG	Triacylglycerol
THL	Finnish Institute for Health and Welfare
TMA	Trimetylamine
TMAO	Trimethylamine N-oxide
WC	Waist circumference
WCRF	World Cancer Research Fund
WGI	Whole Grain Initiative
WHO	World Health Organization

List of original publications

This thesis is based on the following publications:

I Tammi R, Männistö S, Reinivuo H, Tapanainen H, Rautanen J, Kaartinen NE. The relation of whole grain surrogate estimates and food definition to total whole grain intake in the Finnish adult population. *European Journal of Nutrition* 62:1821–31, 2023.

II Tammi R, Männistö S, Maukonen M, Kaartinen NE. Whole grain intake, diet quality and risk factors of chronic diseases: results from a population-based study in Finnish adults. *European Journal of Nutrition* 63:397–408, 2024.

III Tammi R, Kaartinen NE, Harald K, Maukonen M, Tapanainen H, Smith-Warner SA, Albanes D, Eriksson JG, Jousilahti P, Koskinen S, Laaksonen MA, Heikkinen S, Pitkäniemi J, Pajari AM, Männistö S. Partial substitution of red meat or processed meat with plant-based foods and the risk of colorectal cancer. *European Journal of Epidemiology* 39:419–28, 2024.

IV Tammi R, Maukonen M, Kaartinen NE, Koponen K, Niiranen T, Méric G, Albanes D, Eriksson JG, Jousilahti P, Koskinen S, Pajari AM, Knight R, Havulinna AS, Salomaa V, Männistö S. Interplay between colorectal cancer-related lifestyles and the gut microbiome: an exploratory analysis of metagenomic data. *Cancer Causes & Control* (under review).

The publications are referred to in the text by their roman numerals.

1 Introduction

Western dietary patterns high in energy and animal-based foods and low in plant-based foods are both unhealthy and environmentally unsustainable [1]. Together with other unhealthy lifestyle behaviours, such as low physical activity and smoking, unhealthy diets are a major cause of various chronic conditions, including obesity, cardiovascular diseases (CVD), type 2 diabetes (T2D) and cancers. Of cancers, particularly the increasing global burden of colorectal cancer (CRC) has been attributed to lifestyle-related factors [2]. In 2022, CRC was the third-most common cancer and the second-leading cause of cancer mortality worldwide, with nearly two million new cases and one million deaths [3]. In Finland in 2023, CRC ranked second in incidence and third in cancer-related mortality [4]. These statistics underscore the urgency of developing effective lifestyle prevention strategies to alleviate the global burden of CRC.

Strong epidemiological evidence indicates that the consumption of red meat, processed meat and alcohol increases the risk of CRC, whereas the consumption of dairy products and intake of whole grains and dietary fibre decreases the risk [5]. Additionally, suggestive evidence supports beneficial associations of the consumption of vegetables, fruits and fish with CRC. Therefore, aligning with the global call to shift toward more plant-based diets [1], replacing red and processed meat with plant-based foods is expected to benefit CRC prevention. However, to date, research on associations between such dietary changes and CRC risk is limited.

In addition to dietary factors, physical inactivity and obesity have been identified as important risk factors for CRC [5]. Nevertheless, the mechanisms underlying the associations of lifestyle factors and obesity with CRC remain poorly understood. One potential factor modifying these associations is the gut microbiome, as variation in its composition has been linked both to CRC risk factors and CRC [6]. Although lifestyle and other environmental factors have been suggested to contribute substantially to the variation in the gut microbiome composition [7], the impact of individual lifestyle factors, such as diet, has appeared to be modest [7–9]. Consequently, a more comprehensive consideration of CRC-related lifestyle factors, including major risk factors and their potential interactions, might capture better the microbial traits underlying the associations between lifestyles and CRC risk.

High whole grain intake is one of the strongest protective dietary factors against CRC [10]. Additionally, it has been associated with a reduced risk of several other chronic diseases, all-cause and chronic disease mortality, as well as improved weight control [11, 12]. Although the evidence of associations between whole grain intake and health outcomes has been relatively consistent, discrepancies remain regarding its associations with chronic disease risk factors in observational and intervention studies [11, 13]. Furthermore, while the putative health effects of whole grains are largely attributed to their high content of dietary fibre, the exact mechanisms underlying their health effects are not yet fully understood. Whole grain intake is also frequently linked to overall healthier diets and lifestyles [14–18], and the extent to which these factors contribute to whole grains' health associations remain uncertain.

Inconsistencies in the associations between whole grain intake and health are likely related to the methodological challenges evident in whole grain research. These include the use of varying whole grain definitions, as well as unstandardised intake estimation methods that may not adequately reflect total whole grain intake from all food sources [19, 20]. Recent efforts to establish a standard whole grain definition and to harmonise intake estimation methods and their reporting have contributed to the adoption of more rigorous and comparable methods [20, 21]. However, owing to the common lack of comprehensive whole grain data in food composition databases, indirect estimation methods are frequently applied, warranting validation.

The aim of this thesis was to examine the associations of diet and whole lifestyles with CRC risk in the Finnish adult population, including the role of the gut microbiome in these associations. Of CRC risk factors, particular focus was placed on whole grain intake, methodological challenges in its estimation, and its relationship with chronic disease risk factors.

2 Literature review

2.1 Whole grains

In recent decades, whole grains have been widely recognised as a key component of healthy diets. Epidemiological studies have provided accumulating evidence of the health benefits of whole grain intake, including lower risk of chronic diseases and improved weight management [11, 12]. Low whole grain intake has also been identified as one of the main dietary contributors to disability-adjusted life years [22]. Based on this evidence, food and nutrition guidelines worldwide have introduced recommendations to increase daily whole grain intake [23–27]. Furthermore, in 2019, the EAT-Lancet Commission launched a global reference diet promoting human and planetary health, which emphasised high whole grain intake [1]. However, notable limitations remain in the estimation of whole grain intake, affecting the comparability and consistency of research findings. These methodological aspects, as well as the current knowledge of the associations between whole grain intake and health, are examined in the following sections.

2.1.1 Grains

Grains refer to the edible seeds or fruits of the plants primarily within the plant family *Poaceae* [19]. Grains belonging to the *Poaceae* family are commonly called cereals, whereas grains within other plant families that are used similarly and have similar nutrient profiles to cereals are often termed pseudocereals (Table 1). Cereals, such as wheat, oat, rye, rice and maize, are staple crops that contribute to a significant part of daily energy intake in the diets of people worldwide. Major use of pseudocereals, such as buckwheat, quinoa and amaranth, is more restricted to certain food cultures. Although pseudocereals are considered whole grains in this thesis, the following sections will primarily focus on cereals given their higher relevance in Finnish diets inspected in this thesis.

Table 1 Cereal and pseudocereal sources of whole grain

Cereals	Pseudocereals[a]
Barley	Amaranth
Canary seed	Buckwheat (common and tartary)
Fonio	Quinoa
Job's tears	Wild rice
Maize	
Millet	
Oat	
Rice	
Rye	
Sorghum	
Teff	
Triticale	
Wheat	

[a] Grains that are used like cereals and have similar nutrient profiles, but do not belong to the plant family *Poaceae*

Cereal grain structure comprises three major anatomical components: the bran, endosperm and germ (Figure 1) [28]. Most of the grain is composed of the endosperm, which primarily consists of starch, but also contains some protein, lipids, non-starchy polysaccharides and phenolic compounds. The other two components, the bran and germ, constitute up to 25% of the grain structure in dry weight. The bran consists of the outer layers of a grain, and contains high amounts of dietary fibre, protein, micronutrients, such as B vitamins and iron, and bioactive compounds, such as phenolic acids. The germ makes up the smallest proportion of the grain, but contains a significant amount of, for example, protein, unsaturated fatty acids, vitamins B and E and minerals. The proportions of the structural parts vary between cereals. Some cereals, such as barley and oat, also contain hull or husk, which are leaflike inedible sheaths covering the grain [28].

The starchy endosperm is the main component of flour, whereas the bran and germ are typically removed in milling processes. The removal of the bran and germ yields refined grain products, which lack the variety of micronutrients and bioactive compounds present in these components.

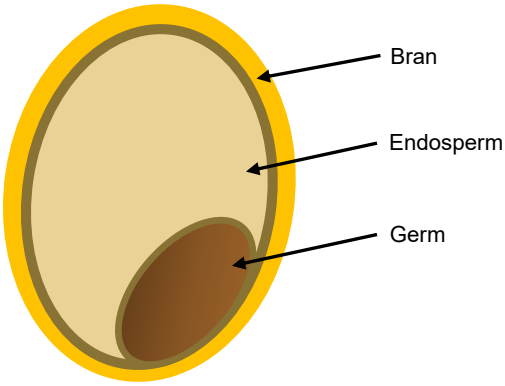


Figure 1 Simplified illustration of the three anatomical components of a cereal grain (created by the author)

2.1.2 Whole grain definition

As noted above, refined grains are cereal grains from which the bran and germ are removed. Conversely, by definition, whole grains contain all (edible) parts of cereal grains. More precisely, defining whole grains requires determining [1] which plants are sources of whole grain, and [2] how their grains can be processed while still being considered as 'whole'. In research, several whole grain definitions have been used with some variation in wording and the level of detail regarding the permitted processing and the losses of grain components. In addition, the inclusion of pseudocereals as a whole grain source vary.

Widely used whole grain definitions include those of the Europe-based HEALTHGRAIN Forum [29] and the American Association of Cereal Chemists International (AACCI; now Cereals & Grains Association) [30]. Additionally, in 2022, the multinational Whole Grain Initiative (WGI), formed by stakeholders from academia, government agencies and industry worldwide (including the HEALTHGRAIN Forum and the Cereals & Grains Association), published a standard whole grain definition with an aim of harmonising whole grain research and the promotion of whole grain intake in different contexts [21]. These three definitions are highly consistent, even though the HEALTHGRAIN Forum definition additionally includes a quantitative limit to the losses of components that may occur in processes that improve the products' safety and quality [29].

Although the use of these definitions have become more standard, several other whole grain definitions have been used in research over the years. The discrepancy in definitions poses challenges in the interpretation and comparability of research findings. Furthermore, the used definition remains frequently unreported, which similarly contributes to inconsistencies.

In this thesis, the HEALTHGRAIN Forum definition was applied as the most recent and up-to-date definition with widespread use at the time of the thesis commence. The definition is presented in detail in Table 2.

Table 2 The HEALTHGRAIN Forum definition for whole grain [29]

Whole grain definition	Included grains
"*Whole grains shall consist of the intact, ground, cracked or flaked kernel after the removal of inedible parts such as the hull and husk. The principal anatomical components − the starchy endosperm, germ and bran − are present in the same relative proportions as they exist in the intact kernel. Small losses of components − that is, less than 2% of the grain/10% of the bran − that occur through processing methods consistent with safety and quality are allowed.*"	Cereals and pseudocereals (including amaranth, buckwheat and quinoa)

2.1.3 Estimation of whole grain intake

As whole grains are essentially defined based on the processing of grains rather than their chemical composition, assessing their quantity in foods requires detailed information on the current processing practices of cereal-based foods. Collecting this information is laborious and time consuming, for which whole grain data is often lacking from food composition databases applied in converting food consumption data into more detailed dietary intake data [20]. Furthermore, regional variation in food processing practices and food items on the market impede broader application of whole grain composition data across countries. Consequently, in the absence of a whole grain database, whole grain intake has been frequently estimated using other approaches. These have primarily been based on estimating consumption of specific foods and food products considered as major sources of whole grain (e.g., "dark bread"), in their whole form [31–34]. Thus, instead of estimating whole grain intake as such, these approaches, in effect, have estimated the consumption of whole grain *foods*. In addition, whole grain intake has been estimated based on, for example, fibre-to-carbohydrate ratio [35].

In addition to intake estimation methods, units used to estimate and report whole grain intake may also vary. Particularly in the US, whole grain servings (or ounce-equivalents) have been frequently used in line with the Dietary Guidelines for Americans [26, 31, 32, 36]. Whole grain servings refer to predefined servings of foods containing approximately 16 g of whole grain (e.g., one slice of whole grain bread or 0.5 cups of cooked whole grains). As the true whole grain content of these foods vary, whole grain servings are not directly comparable to whole grain intake calculated at the ingredient level from all whole grain sources in grams per day.

Similar to the discrepancy in whole grain definitions, the variation in estimation methods contribute to the incomparability of results and inconsistencies in findings regarding, for example, associations between whole grain intake and health.

Nevertheless, as with the standard whole grain definition, efforts have been made to harmonise whole grain intake estimation and to standardise the reporting of the methods used. A paper published in 2015 presented five recommendations for whole grain research, including 1) quantifying whole grain content in foods in grams per dry weight, 2) reporting the whole grain definition used, 3) reporting and separating included grains, 4) reporting the structure of the grains in foods (if relevant), and 5) reporting the main types of grain products and their production methods [20]. Regarding the first recommendation, calculating whole grain content in foods in grams per dry weight is essential to account for the differences in the ways cereal products are prepared and consumed. For example, cooking pasta or oatmeal markedly increases their volume and weight, which can lead to discrepancies depending on whether whole grain content is assessed from dry or cooked foods. In contrast, some cereal products, such as breakfast cereals, are primarily consumed as dry. In addition to water in foods, the residual moisture in ingredients should also be considered in assessing their whole grain content.

In general, the outlined recommendations [20] apply regardless of whether whole grain intake is estimated using a food composition database or another method. Standardising the definitions and methods used, and ensuring their rigorous reporting, will facilitate comparability of research findings and ultimately support efforts to promote whole grain intake in populations.

2.1.4 Whole grain in dietary recommendations

As noted earlier, whole grain intake has received growing emphasis in national food and nutrition recommendations worldwide following accumulating evidence of its health benefits. However, quantitative recommendations (in grams per day [g/d]) are still rare, except for the Nordic countries [23, 37–40]. In 2023, the new Nordic Nutrition Recommendations (NNR23) introduced a recommendation to consume at least 90 grams of whole grains (in dry weight) daily [23]. The same recommendation was subsequently adopted in the updated Finnish, Swedish and Norwegian national recommendations [37–39]. In addition, the Danish recommendation of consuming whole grains at least 75 g/d, issued in 2021, was revised to 90 g/d based on NNR23 [40].

Outside the Nordic countries, a quantitative recommendation has been included in the Dutch Food-Based Dietary Guidelines, although for whole grain foods (90 g/d) [24]. Additionally, serving-based recommendations have been used, for example, in the Republic of Ireland, Australia, Singapore and the US [41, 25, 42, 26]. Otherwise, the recommendations have been more general, advising people to favour whole grains over refined grains. Table 3 displays examples of whole grain recommendations in national dietary guidelines worldwide.

Table 3 Examples of whole grain recommendations in national dietary guidelines worldwide

National dietary guidelines	Whole grain recommendation
Nordic Nutrition Recommendations 2023 [23]	At least 90 g/day
Finnish Nutrition Recommendations 2024[a] [37]	At least 90 g/day
Swedish Dietary Guidelines For Adults 2025[a] [38]	At least 90 g/day
Norwegian Dietary Guidelines 2024[a] [39]	At least 90 g/day
Danish Official Dietary Guidelines 2021 [40]	At least 90 g/day[b]
Estonian National Dietary Recommendations 2025 [43]	Wholegrain products should make up the majority of cereal intake
UK Eatwell Guide 2016 [44]	Choose whole grain varieties
Irish Healthy Eating Guidelines 2016 [41]	3−5[c] servings/day
Dutch Food-Based Dietary Guidelines 2015 [24]	At least 90 g of brown bread, wholemeal bread or other whole-grain products daily
German Food-Based Dietary Guidelines 2024 [45]	Favour whole grain foods
French Recommendations Concerning Diet, Physical Activity and Sedentary Behaviour for Adults 2019 [46]	Consume at least one wholegrain or semi-wholegrain starch per day
Spanish Dietary Recommendations 2022 [47]	Choose whole grain cereals
Dietary Guidelines for Americans 2020−2025 [26]	At least half of total grains[d] should be whole grains
Canada's Dietary Guidelines 2019 [48]	Eat plenty of whole grain foods
Dietary and physical activity guidelines in the context of overweight and obesity in the Mexican population 2015 [27]	At least half of the cereals[e] eaten a day should be whole grains
Australian Dietary Guidelines 2013 [25]	Consume at least four to six serves of grain (cereal) foods, mostly wholegrain and/or high cereal fibre varieties, per day
Eating and Activity Guidelines for New Zealand Adults 2015 [49]	Choose grain foods, mostly whole grain and those naturally high in fibre
Dietary Guidelines for Adult Singaporeans 2003 [42]	At least two servings of rice and alternatives should come from wholegrain food
Dietary Guidelines for Indians 2011 [50]	Use a combination of whole grains, grams and greens
Omani Guide for Healthy Eating 2024 [51]	Favour whole grain foods over refined grains

[a] The Nordic Nutrition Recommendations 2023 [23] adapted for country-specific context
[b] The recommendation was revised based on the Nordic Nutrition Recommendations 2023 [23].
[c] Up to 7 servings per day for men aged 19–50 years
[d] Recommendation for total grains is six ounce equivalents per day at the daily energy intake of 2000 kcal.
[e] Recommendation for total cereal intake is eight servings per day at the daily energy intake of 2000 kcal.

Besides national guidelines, various other stakeholders, such as the European Association of Cardiology [52], the American Heart Association [53] and the World Cancer Research Fund (WCRF) [54], have included whole grains in their recommendations. Furthermore, whole grains were considered one of the key components of healthy and sustainable diets in the global reference diet, the Planetary Health Diet (PHD), launched by the EAT-Lancet Commission [1]. The PHD includes up to 232 g/d of whole grains, with grains covering up to 60% of total energy intake. Given the growing emphasis on sustainability in dietary guidelines [23, 37–40, 47], the PHD recommendation is likely to indicate the direction of future national guidelines, thereby highlighting the role of whole grains in future diets.

2.1.5 Whole grain intake globally and in Finland

Quantitative whole grain recommendations provide an important reference point for assessing the adequacy of intake at a population level. Thus far, whole grain intake has been estimated in adult populations using dietary recalls or food diaries, allowing comparison with the recommendations, in some European countries [55–60], Australia [61] and the US [62] (Table 4). Within these, the average intake levels have been the highest in the Nordic countries, particularly in Finland [55], Sweden [56] and Norway (women) [56], ranging from approximately 40 to 60 g/d. In the UK [57], Republic of Ireland [58] and Australia [61], the average intakes (median in Australia) were estimated at approximately 20–30 g/d, whereas in France [59] and Italy [60], they remained below 10 g/d. In the US, the average intake was estimated at approximately one ounce equivalent per day, corresponding to 16 grams of whole grains [62]. Based on these results, the average whole grain intake across adult populations in Western countries remain well below the Nordic recommendation of 90 g/d. This underscores the global need for improved public health messaging and policies to promote higher whole grain intake, informed by population-level whole grain intake data estimated using standardised methods.

Table 4 Studies estimating whole grain intake based on dietary recalls or food diaries in the adult populations of different countries

Country, Study	Year of data collection	Diet method	Age range (years)	n	Mean whole grain intake (g/d)
Finland, National FinDiet 2017 Survey [55]	2017	2 x 24h recall	18−74	W 875 M 780	47 63
Sweden, HELGA cohort[a] [56]	1992–1998	1 x 24h recall	30, 40, 50 or 60[b]	W 1617 M 1372	41 58
Norway, HELGA cohort[a] [56]	1992–1998	1 x 24h recall	40–55[b]	W 1797	51
Denmark, HELGA cohort[a] [56]	1992–1998	1 x 24h recall	50–65[b]	W 1994 M 1922	37 48
UK, NDNS 2008−2011 [57]	2008−2011	4d food diary	≥18	W+M 1571	26.2
Ireland, NANS [58]	2008−2010	4d food diary	18−90	W 760 M 740	24.1 31.6
France, CCAF 2010 [59]	2010	7d food diary	≥18	W 801 M 588	5.4 3.9
Italy, INRAN-SCAI 2005−2006 [60]	2005−2006	3d food diary	18−64	W 1245 M 1068	4.5 2.9
Australia, Australian Health Survey 2011−2013 [61]	2011−2013	1 x 24h recall	19−85	W 5059 M 4281	median 21.2 median 22.8
US, NHANES 2011−2012 [62]	2011−2012	1 x 24h recall	≥19	W+M 4878	0.97 ounce equivalents

24h recall; 24-hour dietary recall; CCAF, Comportements et Consommations Alimentaires en France; INRAN-SCAI; Italian National Food Consumption Survey; M, men; NANS, National Adult Nutrition Survey; NDNS, National Diet and Nutrition Survey; NHANES, National Health and Nutrition Examination Survey; W, women

[a] HELGA cohort included the Northern Sweden Health and Disease Cohort, the Norwegian Women and Cancer Cohort and the Danish Diet, Cancer and Health Cohort.

[b] At recruitment

2.1.6 Whole grain intake and health

Despite the heterogeneity in whole grain definitions and intake estimation methods, whole grain intake has been fairly consistently associated with a lower risk of several chronic diseases, such as coronary heart diseases (CHD), T2D and CRC [11, 63–65]. For example, in a systematic review of Reynolds et al. (2019), based on prospective studies, higher whole grain intake was associated with a 20% lower CHD risk (6 studies, 7697 cases), 30% lower T2D risk (8 studies, 14 686 cases) and 13% lower CRC risk (7 studies, 8803 cases), compared with lower intakes [11].

Furthermore, each 15 g/d increase in whole grain intake was associated with a reduction of six, seven and three percent in the risk of these diseases, respectively.

In exploring the links between whole grain intake and chronic diseases, several potential mechanisms underlying these associations have been identified (Table 5). These are related to, for example, the regulation of body weight, lipid metabolism, glycaemic response, blood pressure and inflammation, as well as the function of the gut microbiome [5, 11, 12, 63, 64, 66, 67]. Regarding CRC, the mechanisms may also be related to the dilution and reduced transit time of the intestinal contents and binding of carcinogens [5, 66, 67]. These mechanisms may affect the risk of chronic diseases through several pathways. For example, short-chain fatty acids (SCFAs), which are produced in the bacterial fermentation of dietary fibre in the colon, may inhibit endogenous cholesterol synthesis, thus lowering the blood low-density lipoprotein (LDL) cholesterol concentration and the risk of CHD [68]. At the same time, SCFAs may exert anti-proliferative effects on colonic epithelial cells, thus reducing the risk of CRC.

The potential health effects of whole grain intake are primarily thought to arise from their high content of dietary fibre [5,11,69]. However, whole grains contain several other potentially beneficial compounds, such as different vitamins, phenolic acids and other bioactive compounds, which may contribute to their associations with health [70]. In general, the exact mechanisms underlying the protective associations of whole grain intake with the risk of chronic diseases remain incompletely understood.

Table 5 Potential mechanisms underlying the associations between whole grain intake and the risk of coronary heart disease, type 2 diabetes and colorectal cancer

Coronary heart disease	Type 2 diabetes[a]	Colorectal cancer
• Regulation of body weight	• Regulation of body weight	• Regulation of body weight
• Regulation of blood cholesterol concentrations	• Regulation of oxidative stress and inflammation	• Regulation of oxidative stress and inflammation
• Regulation of blood pressure	• Regulation of glycaemic control	• Regulation of glycaemic control
• Regulation of oxidative stress and inflammation	• Prevention of insulin resistance	• Prevention of insulin resistance
• Regulation of glycaemic control		• Fermentation of dietary fibre into SCFAs
• Fermentation of dietary fibre into SCFAs		• Increase of faecal bulk
		• Dilution of intestinal contents
		• Binding of carcinogens
		• Reduction of intestinal transit time

SCFAs, short-chain fatty acids

[a] Type 2 diabetes is a risk factor for both coronary heart disease and colorectal cancer.

Studies examining associations between whole grain intake and chronic disease risk factors (e.g., body mass index [BMI], blood pressure, cholesterol) have only partially supported the presumed mechanisms linking whole grain intake to reduced risk of chronic diseases. Recent systematic reviews of randomised controlled trials and prospective cohorts have reported significant associations primarily between whole grain intake and lower body weight [11] and lower risk of overweight and obesity [12]. The associations with, for example, total and LDL cholesterol, triacylglycerol (TAG), blood pressure and fasting glucose have been non-significant [11]. The findings from cross-sectional studies have been varying, although most studies have reported inverse associations between whole grain intake and obesity measures (Table 6) [14–17, 32, 62, 71–74].

The inconsistency of findings in observational studies is likely related, to some extent, to the methodological discrepancies in whole grain research. A systematic review of 13 observational studies reported that associations of whole grain intake with CVD and its risk factors were less affected by the adjustment for cereal fibre intake when whole grains were defined using a standard definition (e.g., the AACCI definition) compared with other approaches (e.g., estimating the consumption of 'whole grain foods') [75]. This suggests that the contributions of beneficial compounds other than fibre were underestimated when nonstandard definitions were used. The underestimation could result, for example, from refined grains contributing to whole grain intake when the estimation is based on 'whole grain foods', which are rarely fully whole grain. Overall, it is evident that the use of nonstandard methods to estimate whole grain intake may influence the observed associations between whole grain intake and health outcomes.

It is also important to note that whole grain intake has been frequently associated with overall healthier lifestyles, including more physical activity, less smoking and healthier diets [14–18]. The extent to which these factors contribute to whole grains' associations with health remains unclear. Although smoking, physical activity and alcohol consumption are commonly considered in the analyses, adjustment for other dietary factors and diet quality is frequently lacking [11, 75].

Taken together, several research gaps remain regarding the associations between whole grain intake and health. Given the relatively high whole grain intake and wide intake range in the Nordic countries [55, 56], Nordic populations might serve as particularly suitable study samples to examine associations between whole grain intake and health. However, thus far, epidemiological studies in the Nordic countries have frequently examined associations between whole grain intake and health outcomes using nonstandard whole grain definitions and estimation methods [33, 76–80]. Moreover, research on associations between whole grain intake and chronic disease risk factors remains limited.

Table 6 Cross-sectional studies (n>500) on associations between whole grain (WG) intake and chronic disease risk factors in apparently healthy adults

Study, country	Population	Diet method	WG intake estimation method	Adjustments[a]	Results
McKeown et al. 2002 [71] US	Framingham Offspring Study 2941 women and men mean age 54 years	126-item FFQ	WG intake (servings/d) was estimated as the consumption of WG breakfast cereals (≥25% WG or bran by weight), dark bread, popcorn, cooked oatmeal, wheat germ, brown rice and other grains (e.g., couscous).	Sex, age, total energy, multivitamin supplement use, alcohol, BP-lowering medication use, smoking, physical activity, oestrogen use (women), E% PUFA, E% SAFA, meat, fish, fruits, vegetables	Trend across WG quintiles Waist-hip ratio ↓ LDL ↓ Fasting insulin ↓ BMI, SBP, DBP, TC, HDL, TAG, fasting glucose, 2h glucose, 2h insulin, HbA1c: NS
Jensen et al. 2006 [17] US	Nurses' Health Study II (women) & Health Professionals Follow-Up Study (men) 470 women aged 25−42 years 468 men aged 40−75 years	131-item FFQ	WG intake (servings/d) was estimated on a dry weight basis based on the USDA WG classification and using a WG database.	Age, sex, smoking, physical activity, BMI, total energy, alcohol, fruits and vegetables, sucrose, fructose, SAFA, MUFA, PUFA, multivitamin supplement use, folate, choline, betaine, riboflavin, vitamin B6, niacin, vitamin E, glucose, hypercholesterolemia	Trend across WG quintiles BMI (men) ↓ TC ↓ C-peptide ↓ Leptin ↓ Homocysteine ↓ BMI (women), HbA1c, insulin, HDL, LDL, CRP, fibrinogen, IL-6: NS

Table continues

Table 6 continues

Study, country	Population	Diet method	WG intake estimation method	Adjustments[a]	Results
Newby et al. 2007 [72] US	Baltimore Longitudinal Study of Aging 1516 women and men aged 27−88 years	7-day food diary	WG intake (g/d) was estimated using a WG database.	Age, sex, ethnicity, education, decade of visit, physical activity, vitamin supplement use, total energy, E% SAFA, alcohol, refined grains, BMI, lipid-lowering medication use, hypercholesterolemia, BP-lowering medication use, hypertension, oral hypoglycaemic medication use, diabetes diagnosis	Trend across WG quintiles BMI ↓ Weight ↓ WC ↓ TC ↓ LDL ↓ 2h glucose ↓ HDL, TAG, DBP, SBP, fasting glucose, fasting insulin, 2h insulin: NS
Lutsey et al. 2007 [32] US	Multi-Ethnic Study of Atherosclerosis 5496 women and men aged 45−84 years	127-item FFQ	WG intake (servings/d) was estimated as the consumption of WG breakfast cereals (≥3g dietary fibre per 100g dry weight), oatmeal, dark bread, bran muffins and brown or wild rice.	Age, sex, ethnicity, education, survey centre, total energy, smoking, alcohol, fruits, vegetables, refined grains, dairy, fish and poultry, meat, BMI, serum insulin	Trend across WG quintiles BMI ↓ Insulin ↓ Homocysteine ↓ HDL, LDL, SBP, DBP, CRP, IL-6, glucose, urine albumin: NS
van de Vijver et al. 2009 [73] Netherlands	Netherlands Cohort Study 2159 women aged 55−69 years 2078 men aged 55−69 years	150-item FFQ	WG intake (g/d) was estimated as the consumption of muesli, porridge (dry product), brown rice (dry product), and cooked grains (dry product).	Age, smoking, number of cigarettes, physical activity, education, total energy, animal protein, fruits and vegetables	Linear regression BMI (men) ↓ BMI (women) ↓

Table continues

Table 6 continues

Study, country	Population	Diet method	WG intake estimation method	Adjustments[a]	Results
Masters et al. 2010 [81] US	Insulin Resistance Atherosclerosis Study 1015 women and men aged 40−60 years	114-item FFQ	WG intake (servings/d) was estimated as the consumption of dark bread, high-fibre bran or granola cereals, shredded wheat and cooked kernel (weighted by a factor of 0.5).	Age, sex, ethnicity, smoking, total energy, total estimated energy expenditure, alcohol, fruit, vegetable, E% SAFA, E% PUFA, E% oleic acid, refined grains, insulin sensitivity, WC, 2h post-load glucose	Linear regression CRP: NS
O'Neil et al. 2010 [74] US	National Health and Nutrition Examination Survey (NHANES) 1999−2004 13 276 women and men aged ≥19−50 years	1 x 24h recall	WG intake (servings/d) was estimated based on the USDA WG classification and using a WG database.	Age, sex, ethnicity, total energy	Trend across WG intake groups BMI ↓ WC ↓
Kyrø et al. 2011 [14] Sweden, Norway and Denmark	HELGA[b] 5408 women aged 30−65 years 3294 men aged 30−65 years	1 x 24h recall	WG intake (g/d) was estimated based on the AACCI definition from all used cereal-based products.	Age, total energy, country, alcohol, education, smoking, potatoes, vegetables, fruits, nuts, milk, yoghurt, cheese, white bread, red meat, poultry, processed meat, fish and shellfish, eggs, sugar and confectionary, cakes and biscuits, coffee, tea, vegetable oils, butter, margarine	Log-linear trend BMI (men) ↓ BMI (women): NS
Albertson et al. 2016 [62] US	National Health and Nutrition Examination Survey (NHANES) 2001−2012 29 683 women and men aged ≥19 years	1 x 24h recall	WG intake (oz-eq/d) was estimated based on the USDA WG classification and using a WG database.	Age, age^2, sex, ethnicity, total energy, alcohol, physical activity	Linear regression BMI ↓ WC ↓

Table 6 continues

Study, country	Population	Diet method	WG intake estimation method	Adjustments[a]	Results
Barrett et al. 2020 [16] UK	National Diet and Nutrition Survey (NDNS) Rolling Programme 2008–2014 2689 women and men aged ≥18 years	4-day food diary	WG intake (g/d) was estimated per 100g dry weight based on the WGI definition and using a WG database.	Age, sex, total energy, ethnicity, smoking status, alcohol, SAFA, trans fat, non-milk extrinsic sugar, sodium, BMI, BP-lowering medication use, lipid-lowering medication use	Trend across WG quartiles Waist-hip ratio ↓ CRP ↓ Homocysteine ↓ BMI, SBP, TC, HDL, LDL, TAG, fasting glucose, HbA1c: NS
Barrett et al. 2020 [15] Australia	Australian Health Survey 2011–2013 7665 women and men aged ≥18 years	1 x 24h recall	WG intake (g/d) was estimated per 100g dry weight based on the Food Standards Australia New Zealand definition for WG and using a WG database.	Age, sex, total energy, alcohol, smoking, physical activity, fruit and vegetables, sodium, BMI, SAFA, trans fat, PUFA, free sugar, lipid-lowering medication use	Trend across WG tertiles BMI ↓ WC ↓ SBP, DBP, TC, HDL, LDL, Apo-B, CRP, TAG, fasting glucose, HbA1c: NS
Taskinen et al. 2021 [82] Finland	Kuopio Ischaemic Heart Disease Risk Factory Study 365 women aged 53−73 years and 391 men aged 42−60 years	4-day food diary	WG intake (g/d) was estimated based on the HEALTHGRAIN Forum definition from all mixed dishes and ingredients of decomposed foods (recipe-calculation).	Age, sex, examination year, total energy, BMI, smoking, leisure-time physical activity, education, alcohol, fruits, vegetables, berries, ratio of SAFA and trans fats to PUFA and MUFA, red meat, dairy, fish, butter, vegetable oil margarine, eggs	Trend across WG quartiles CRP: NS

24h recall, 24-hour dietary recall; AACCI, American Association of Cereal Chemists International; BMI, body mass index; BP, blood pressure; CRP, C-reactive protein; E%, percent of energy; HbA1c, glycated haemoglobin; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; MUFA, monounsaturated fatty acids; oz-eq, ounce equivalent; PUFA, polyunsaturated fatty acids; SAFA, saturated fatty acids; SBP, systolic blood pressure; TAG, triacylglycerol; USDA, US Department of Agriculture; WC, waist circumference; WG, whole grain; WGI, Whole Grain Initiative

↓ statistically significant inverse association with whole grain intake; ↑ statistically significant direct association with whole grain intake; NS, no statistically significant association with whole grain intake

[a] All adjusted variables for each outcome; the used model may vary by outcome

[b] Including the Northern Sweden Health and Disease Study, the Norwegian Women and Cancer Study and the Danish Diet, Cancer and Health Study

2.1.7 Summary of whole grain research

Although whole grains are recognised as one of the key components in healthy and sustainable diets, whole grain research remains challenged by methodological limitations that hinder the comparability and generalisability of research findings. These limitations are related to the use of heterogenous whole grain definitions and intake estimation methods. Nevertheless, in recent years, several international initiatives have sought to establish a standard definition and a common framework for intake estimation and reporting of the methods used.

Following the accumulating evidence of their potential health benefits, whole grains have been increasingly considered in national dietary guidelines and recommendations of different stakeholders. However, these are predominantly general recommendations to favour whole grains over refined grains, while quantitative recommendations in grams per day have been mainly established in the Nordic countries. The Nordic countries have also reported the highest levels of average whole grain intake, although the intake generally remains well below the Nordic recommendation of 90 g/d in all countries with available data.

Whole grain intake has been associated with a reduced risk of several major chronic diseases. Although various mechanisms have been identified to potentially underlie these associations, uncertainty remains regarding the specific pathways. Studying associations between whole grain intake and chronic disease risk factors contributes to the knowledge of these pathways. Moreover, carefully accounting for other lifestyle factors that may modify the health associations facilitates confirming the independent health benefits of whole grain intake. Given the relatively high whole grain intake and wide intake range, Nordic populations may be particularly suitable for studying associations between whole grain intake and health.

2.2 Colorectal cancer

Colorectal cancer (CRC) is globally the third-most common cancer and the second-most common cause of cancer deaths [83]. The incidence of CRC has increased considerably over the past decades, with several modifiable lifestyle factors identified to promote its pathogenesis. These risk factors are major characteristics of Western lifestyles, including low physical activity, high consumption of red and processed meat and low consumption of plant-based foods [5]. Although CRC incidence has declined in older populations of some high-income countries owing to screening programs and lifestyle changes (e.g., reduced smoking), the incidence has continued to rise in younger populations [84, 85]. Moreover, the incidence is growing particularly in lower-income countries transitioning toward more Western lifestyles. Despite there being strong evidence of association between several

lifestyle factors and CRC, the exact mechanisms underlying these associations remain poorly known [5, 86]. One potential mediator is the gut microbiome, as alterations in its composition and function have been linked both to CRC and CRC-related lifestyle factors [6, 86]. The following sections will examine the current knowledge of CRC, its associations with lifestyle factors, and the role of the gut microbiome in these relationships.

2.2.1 Colorectal cancer definition, classification and screening

CRC refers to cancers of the colon and rectum, which constitute the latter part of the intestinal tract. The colon follows the small intestine, starting from the caecum and extending to the sigmoid colon, which connects to the rectum [87]. Between these are the ascending colon located at the right side of the abdomen, the transverse colon crossing the abdomen to the left side, and the descending colon at the left side of the abdomen. The ascending colon transitions into the transverse colon at the hepatic flexure and the transverse colon into the descending colon at the splenic flexure.

Histologically, most CRCs are adenocarcinomas developing through the adenoma-carcinoma pathway [5, 85]; following a serious of genetic and epigenetic changes, an aberrant crypt develops into a neoplastic precursor lesion and further into colorectal cancer [85]. This development occurs over a long time, taking an average of 10 to 15 years. Other CRC types include mucinous and adenosquamous carcinomas. By tumour location, cancers of the colorectum can be classified as those of the proximal colon (i.e., right-sided, including caecum, ascending colon, hepatic flexure, and part of the transverse colon), distal colon (i.e., left-sided, including part of the transverse colon, splenic flexure, descending colon, and sigmoid and rectosigmoid colon), and rectum. Tumours in different locations of the colorectum differ in their molecular features [85]. At the metastatic phase, the prognosis of proximal cancers is typically worse than that of distal cancers.

CRC remains often asymptomatic until reaching an advanced stage [85]. Therefore, and owing to its slow growth, screening programs have been established in various countries to enable early detection of neoplastic changes. The screening programs primarily rely on stool-based tests, such as the faecal immunochemical test (FIT), which detects occult blood in stool [88]. Colonoscopy is a more invasive diagnostic method, commonly used for further examination following a positive test result. In the US, colonoscopy is, however, often used as a primary screening method. In Finland, nationwide CRC screening programme using FIT was launched in 2022, covering the population aged 60–72 years [89]. By 2031, the biannual screening will extend to cover all age groups from 56 to 74 years.

2.2.2 Colorectal cancer globally and in Finland

As noted earlier, CRC is one of the most common cancer types and causes of cancer deaths worldwide [83]. In 2020, approximately 1.9 million new CRC cases (including anal cancers) were diagnosed globally, accounting for 10% of all cancers. Moreover, the number of CRC deaths was over 900 000, accounting for almost 10% of all cancer deaths. In both women and men, CRC is considerably more common in high-income than lower-income countries. Increase in its incidence is considered a distinct marker of socioeconomic development, characterised by the rise in CRC risk factors. Indeed, following the current progress particularly in lower-income countries, the number of new CRC cases is predicted to increase by 63% from the level of 2020, reaching 3.2 million by 2040 [90].

Northern Europe is among the areas with the highest age-standardised incidence rates of both colon and rectal cancers globally. In Finland in 2023, CRC was the second-most common cancer in women and men, with 2070 new cases diagnosed in women and 2467 in men [4]. CRC was also the third-most common cause of cancer death in both sexes. The incidence of CRC has increased particularly in men, although some of the observed increase is attributable to the initiation of the screening programme in 2022. In contrast, CRC mortality has decreased in Finland since the 1990s.

2.2.3 Colorectal cancer and lifestyle

Lifestyle-related risk factors

Over recent decades, substantial evidence has accumulated linking numerous lifestyle factors to the development of CRC [2, 5]. In response, international organisations, such as WCRF, the American Institute for Cancer Research (AICR) and the International Agency for Research on Cancer (IARC) have conducted comprehensive evaluations to assess the strength of this evidence. In the 2018 WCRF/AICR Third Expert Report of the Continuous Update Project, the evidence of associations between ten lifestyle and anthropometric factors and CRC was considered strong [5]. Of these, five were protective factors, including physical activity, the use of calcium supplements and the consumption of whole grains, foods containing dietary fibre and dairy products. Conversely, five factors were associated with an increased CRC risk, including obesity, adult attained height and the consumption of alcohol, processed meat and red meat. Furthermore, low intake of starchy vegetables and fruits were considered potential risk factors for CRC, with suggestive but limited evidence. In the WCRF/AICR meta-analyses, on which these conclusions were based on, the consumption of, for example, whole grains of 90 g/d or dairy products of 400 g/d were estimated to decrease CRC risk by 17% (6 studies, 8320 cases) and 13% (10 studies, 14 859 cases), respectively [10]. In contrast, the consumption of processed meat of 50 g/d or alcohol of 10 g/d were estimated to

increase CRC risk by 16% (10 studies, 10 738 cases) and 7% (16 studies, 15 896 cases), respectively. In the IARC assessments, processed meat was deemed a cause of CRC with sufficient evidence of its carcinogenicity, whereas the evidence of carcinogenicity for red meat and alcohol were considered limited [91, 92]. Based on the IARC GLOBOCAN2012 data, as much as 13% of colon cancers in men and 7.6% in women were estimated to be attributable to high BMI [93]. Besides diet, anthropometrics and physical activity, smoking has been identified as an important risk factor for CRC [5].

Dietary changes

As discussed above, several dietary factors have been identified as important risk factors for CRC. In general, the current evidence suggests that decreasing the consumption of red and processed meat and increasing the consumption of plant-based foods, particularly whole grains and other fibre-rich foods, would decrease the risk of CRC [5]. Thus, dietary changes toward more plant-based diets, which are frequently recommended for both health and sustainability reasons [1, 23, 37–40, 45, 47, 48], would presumably benefit the prevention of CRC. However, the associations between individual food groups and CRC risk may not directly reflect the summed impact of concurrent dietary changes, such as decreasing the consumption of red and processed meat while increasing the consumption of plant-based foods. Considering the complex interplay within and between foods and nutrients, such changes may influence the risk of diseases to a greater or lesser extent than expected based on their individual associations. Given that dietary guidelines worldwide encourage a shift toward more plant-based diets, it is crucial to understand how such dietary changes would affect the risk of diseases at the population level. For this, modelling studies based on prospective data are required.

To date, few prospective studies have modelled dietary changes toward more plant-based diets in relation to CRC risk. In a study of 490 000 US adults aged 50 to 71 years (9000 CRC cases over a median follow-up of 15.5 years), the highest quintile of substituting total animal *protein* with plant *protein* was associated with a 9% reduction in CRC risk compared with the lowest substitution quintile [94]. Similarly, substituting red meat protein with total plant protein or cereal protein (including bread, cereals and pasta) was associated with 11% and 14% reductions in CRC risk, respectively, when comparing the highest to the lowest substitution quintile. In a study of 44 824 Italian adults (539 CRC cases over a median follow-up of 14 years), substituting 3% of energy intake from red and processed meat *protein* with energy from plant *protein* was associated with a 34% reduction in rectal cancer risk [95]. The number of rectal cancer cases was, however, relatively small (n=101). For colon cancer, the same substitution was associated with a 21% increase in risk, but the result was largely driven by foods with high glycaemic index (e.g., potato, pizza, bread and rice).

Taken together, evidence on the associations between dietary changes to more plant-based diets and CRC risk is limited. More research is warranted particularly at a food-group level, reflecting dietary changes that are promoted in the current food-based dietary guidelines.

CRC indices

Another approach to examine the relationship between lifestyles and CRC risk is to use indices that consider multiple lifestyle factors simultaneously. Such indices may reflect the multifactorial real-world risk environment better than single risk factors, as individuals tend to exhibit multiple risk factors concurrently. Moreover, interactions among lifestyle factors may synergistically modify their associations with disease risk [96, 97].

Several prospective cohort studies have examined associations between *a priori* lifestyle indices, including dietary and other lifestyle factors, and CRC risk (Table 7). *A priori* indices, in which the components and scoring cut-offs are defined based on existing literature, may allow for more consistent comparison across studies and populations compared with data-driven *a posteriori* indices [98, 99]. The majority of the identified studies used indices based on the general cancer prevention recommendations of the 2007 [100–104] or 2018 [105–107] WCRF/AICR Report. Of these, all included fruits and vegetables, red meat and processed meat, sugar-sweetened beverages (SSBs) or sugary drinks, alcohol, BMI and physical activity in some form as index components. In addition, most of the WCRF/AICR-based indices included dietary fibre, sodium and the energy density of foods or diets. Of the other identified studies, one used an index based on the nutrition and physical activity recommendations of the American Cancer Society (ACS) [108], whereas the rest used diverse indices reflecting a healthy lifestyle [97, 107, 109–112]. Similar to the WCRF/AICR-based indices, most of the other indices included BMI or waist circumference (WC), alcohol, physical activity and fruits and vegetables either as such or within an embedded dietary index. Additionally, smoking was frequently included. Overall, only two studies [97, 109] used an index strictly based on CRC-related lifestyle factors. However, both of these failed to consider some of the factors identified as convincing or probable risk factors for CRC in the 2018 WCRF/AICR report (e.g., BMI, whole grain intake, dairy consumption) [5, 97, 109].

Most of the studies examining associations between an *a priori* lifestyle index and CRC risk observed a statistically significant reduction in CRC risk among participants scoring higher index points (more beneficial lifestyles), compared with those scoring lower points. The risk reductions ranged from 27% to 63% with the greatest reductions observed in the Singapore Chinese Health Study using the Protective lifestyle factor index (63% risk reduction, n=50 466, 969 CRC cases) and the US VITAL cohort using an index based on the 2007 WCRF/AICR recommendations (58% risk reduction, n=66 920, 546 CRC cases) [110, 104]. Conversely, the smallest risk reductions were observed in the UK Biobank using an

index based on the 2018 WCRF/AICR recommendations (27% risk reduction, n=94 778, 863 CRC cases), and in the European EPIC cohort (27%, n=386 355, 3880 CRC cases) using an index based on the 2007 WCRF/AICR recommendations [97, 106]. Non-significant associations were mainly observed in women [100, 102, 105].

In summary, prospective cohort studies using *a priori* lifestyle indices suggest a generally consistent association between healthier lifestyles and CRC risk, although the magnitude of the observed risk reduction varies across studies. Few studies have, however, explored associations between an index based on strictly CRC-specific risk factors and CRC risk, and none have incorporated all strong-evidence risk factors for CRC identified in the 2018 WCRF/AICR Report.

Table 7 Prospective cohort studies on associations between *a priori* lifestyle indices (including lifestyle factors and anthropometric measures) and colorectal cancer (CRC) risk

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Kirkegaard et al. 2010 [109] Denmark	Diet, Cancer and Health Cohort Study 55 487 women and men 50–64 years at baseline	9.9 years (median) 678 cases	Healthy lifestyle index[c] Point range: 0–5 <u>Components</u> Waist circumference Smoking Physical activity Vegetables and fruits Red and processed meat Dietary fibre Energy from fatty acids Alcohol	Sex[d], CRC family history, education, NSAID use, HRT use	Per 1-point increase: 0.89 (0.82–0.96) High vs. low points: NS
Romaguera et al. 2012 [103] Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Sweden, UK	European Prospective Investigation into Cancer and Nutrition (EPIC) 386 355 women and men 25–70 years at baseline	11 years (median) 3880 cases	Index based on the 2007 WCRF/AICR general cancer prevention recommendations[e] Point range: 0–7 (0–6 for men) <u>Components</u> BMI Physical activity Diet energy density Sugary drinks Vegetables and fruits Dietary fibre Red and processed meat Alcohol (ethanol) Breastfeeding	Baseline age[d], study centre[d], sex[d], education, chronic diseases at baseline (T2D, heart disease, stroke), smoking, menopausal status, HRT use, oral contraceptive use, age at menarche, parity, age at first full-time pregnancy, total energy	Per 1-point increase: 0.88 (0.84–0.91) High vs. low points: 0.73 (0.65–0.81)

Table continues

Table 7 continues

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Odegaard et al. 2013 [110] Singapore	Singapore Chinese Health Study 50 466 women and men 45−74 years at baseline	Follow-up from 1993–1998 until 2007 969 cases	Protective lifestyle factor index[f] Point range: 0−10 Components BMI Dietary pattern rich in vegetables, fruits and soy Physical activity Sleep Smoking Alcohol	Age, sex, interview year, dialect, education, diabetes status, CRC family history, total energy	High vs. low points: 0.37 (0.22−0.62)
Thomson et al. 2014 [108] US	Women's Health Initiative Observational Study 65 838 women 50−79 years at baseline	12.6 years (mean) 751 cases	Index based on the ACS Nutrition and Physical Activity Cancer Prevention Guidelines[f] Point range: 0−8 Components BMI Physical activity Vegetables, fruits and total carotenoids Ratio of whole grains to total grains Red and processed meat Alcohol	Age, education, smoking, NSAID use, aspirin use, unopposed oestrogen use, oestorgen + progestin use, multivitamin use, race, total energy, parity, colonoscopy or sigmoidoscopy, family history of cancer, having a current healthcare provider	Per 1-point increase: 0.89 (0.85−0.94) High vs. low points: 0.48 (0.32−0.73)
Dartois et al. 2014 [111] France	E3N 64 732 women 43−68 at baseline	8 years (median for those diagnosed with cancer) 481 cases	Health index[g] Point range: 0−5 Components BMI Physical activity Vegetables and fruits Alcohol Smoking	Generation[d], education, residential area, CRC family history, professional activity, oral contraceptive use, menopausal status, HRT use, age at menarch, parity, age at first full-term pregnancy	High vs. low points: 0.66 (0.45−0.97)

Table 7 continues

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Aleksandrova et al. 2014 [97] Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Sweden, UK	European Prospective Investigation into Cancer and Nutrition (EPIC) 347 237 women and men 25−70 years at baseline	12 years (median) 3759 cases	Healthy lifestyle index[h] Point range: 0−5 Components BMI and waist circumference Smoking Physical activity Alcohol Dietary quality index	Study centre[d], age, sex, education	Per 1-point increase: 0.88 (0.86−0.92) High vs. low points: 0.63 (0.54−0.74)
Makarem et al. 2015 [101] US	Framingham Offspring cohort 2983 women and men 66 years (mean) at baseline	11.5 years (mean) 63 cases	Index based on the 2007 WCRF/AICR general cancer prevention recommendations[i] Point range: 0−7 Components BMI Physical activity Energy-dense foods Sugary drinks Vegetables and fruits Refined grains Non-starchy and starchy plant foods Red and processed meat Alcohol (ethanol) Salty foods Sodium	Age, sex, smoking, CVD, diabetes, menopausal status, hormone use	Per 1-point increase: NS

Table continues

Table 7 continues

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Nomura et al. 2016 [102] US	Black Women's Health Study 49 103 women 21−69 years at baseline	15.1 years (mean) 354 cases	Index based on the 2007 WCRF/AICR general cancer prevention recommendations[g] Point range: 0−7 Components BMI Physical activity and sedentary time Sugary beverages Vegetables, fruits and dietary fibre Red and processed meat Alcohol Sodium	Age, region of residence, total energy, smoking, CRC family history, education, menopausal status, insulin use, aspirin use, colonoscopy, sigmoidoscopy	Per 0.5-point increase: NS High vs. low points: NS
Hastert & White 2016 [104] US	VITamins And Lifestyle Study cohort (VITAL) 66 920 women and men 50−76 years at baseline	7.6 years (mean) 546 cases	Index based on the 2007 WCRF/AICR general cancer prevention recommendations[h] Point range: 0−6 Components BMI Physical activity Diet energy density Sugary drinks and fruit juice Vegetables, fruits, whole grains and legumes Red and processed meat Alcohol	Sex, education, race, colonoscopy or sigmoidoscopy, CRC family history, history of cancer other than CRC, smoking, NSAID use, total energy	Per 1-point increase: 0.87 (0.80−0.95) High vs. low points: 0.42 (0.26−0.66)

Table continues

Table 7 continues

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Jones et al. 2018 [100] UK	UK Women's Cohort Study 30 963 women 52 years (mean) at baseline	17.4 years (median) 444 cases	Index based on the 2007 WCRF/AICR general cancer prevention recommendations[e] Point range: 0−8 Components BMI Physical activity Energy-dense foods Sugary drinks Vegetables and fruits Dietary fibre Red and processed meat Alcohol (ethanol) Sodium Breastfeeding	Age, smoking, CRC family history, socioeconomic status	Per 1-point increase: NS High vs. low points: NS
Zhang et al. 2018 [112] China	Shanghai Men's Health Study 59 503 men 40−74 years at baseline	9.3 years (median) 674 cases	Healthy lifestyle index[k] Point range: 0−5 Components Waist-to-hip ratio Physical activity Chinese Food Pagoda score Alcohol Smoking	Age, occupation, education, income, diabetes history, CRC family history	Per 1-point increase: 0.83 (0.78−0.89) High vs. low points: 0.50 (0.39−0.65)

Table continues

Table 7 continues

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Petimar et al. 2019 [105] US	Nurses' Health Study (women) & Health Professionals Follow-up Study (men) 68 977 women 30−55 years at baseline 45 442 men 40−75 years at baseline	Follow-up from 1986 until 2012 2449 cases	Index based on the 2018 WCRF/AICR general cancer prevention recommendations[m] Point range: 0−3 Components Fruits and vegetables Dietary fibre Whole grains and legumes Refined grains and processed foods Red and processed meat SSB Alcohol BMI Waist circumference Weight change Energy expenditure Sedentary activity	Age[d], year[d], total energy, NSAID or aspirin use, CRC family history, CRC screening, history of polyps, smoking, multivitamin use, calcium supplement use, young adult BMI, menopausal status, postmenopausal hormone use	Women, per 1-point increase: NS Women, High vs. low points: NS Men, per 1-point increase: 0.78 (0.72−0.85) Men, High vs. low points: 0.64 (0.52−0.77)
Barrubés et al. 2020 [107] Spain	PREvencion con DIeta MEDiterranea (PREDIMED) 7216 women and men 55−80 years at baseline	6 years (median) 97 cases	Index based on the 2018 WCRF/AICR general cancer prevention recommendations[n] Point range: 0−7 Components BMI Weight change Physical activity Dietary fibre Legumes Fruits and vegetables Fast food and processed food Red and processed meat SSB Alcohol (ethanol)	Age, sex, intervention group, CRC family history, education, diabetes history, total energy, aspirin use, smoking	Per 1-point increase: 0.79 (0.63−0.99) High vs. low points: 0.52 (0.27−0.99)

Table 7 continues

Study, country	Population	Follow-up time and CRC cases	Index[a]	Adjustments	Results, HR (95% CI)[b]
Barrubés et al. 2020 [107] Spain	PREvencion con DIeta MEDiterranea (PREDIMED) 7216 women and men 55−80 years at baseline	6 years (median) 97 cases	Low-Risk Lifestyle Score[h] Point range: 0−5 Components Smoking Alcohol Physical activity AHEI-2010 BMI	Age, sex, intervention group, CRC family history, education, diabetes history, total energy, aspirin use	Per 1-point increase: 0.78 (0.64−0.96) High vs. low points: 0.48 (0.26−0.86)
Malcomson et al. 2023 [106] UK	UK Biobank 94 778 women and men 37−73 years at baseline	8 years (median) 863 cases	Index based on the 2018 WCRF/AICR general cancer prevention recommendations[o] Point range: 0−7 Components BMI Waist circumference Physical activity Fruits and vegetables Total fibre Energy from ultra-processed foods Red and processed meat SSB Alcohol (ethanol)	Age, sex, deprivation index, ethnicity, smoking, total energy, multimorbidity, education, number of 24h recalls, CRC family history	Per 1-point increase: 0.87 (0.82−0.93) High vs. low points: 0.73 (0.62−0.86)

Table continues

Table 7 continues

ACS, American Cancer Society; AHEI, Alternative Healthy Eating Index; BMI, body mass index; CI, confidence interval; CVD, cardiovascular disease; CRC, colorectal cancer; E3N, Étude Épidemiologique auprés des femmes de la Mutuelle Générale de l'Education Nationale; HR, hazard ratio; HRT, hormone-replacement therapy; NS, non-significant; NSAID, nonsteroidal anti-inflammatory drugs; SSB, sugar-sweetened beverages; T2D, type 2 diabetes

[a] In all indices, higher points reflected a healthier lifestyle.

[b] Hazard ratios were derived from Cox proportional hazards regression models.

[c] Waist circumference, smoking, physical activity and alcohol intake were assigned 0 or 1 point. Participants received 1 point for dietary variables (other than alcohol) if they met all the respective recommendations.

[d] Analysis was stratified by the variable.

[e] Each component was assigned 0, 0.5 or 1 point. Points for energy-dense foods and sugary drinks, as well as for fruits and vegetables and dietary fibre were averaged to form one component.

[f] Each component was assigned 0, 1 or 2 points.

[g] Each component was assigned 0, 0.5 or 1 point.

[h] Each component was assigned 0 or 1 point.

[i] Each component was assigned 0, 0.5 or 1 point. Points for energy-dense foods and sugary drinks, as well as for fruits and vegetables, refined grains and starchy and non-starchy plant foods were averaged to form one component.

[k] Each component was assigned 0 or 1 point. Participants in the top three quintiles of the Chinese Food Pagoda Score received 1 point.

[m] Each component was assigned 0, 0.5 or 1 point. The points for dietary, adiposity and physical activity components were averaged within each category.

[n] Each component was assigned 0, 0.5 or 1 point. Points for BMI and weight change, as well as for dietary fibre, legumes and fruits and vegetables were averaged to form one component.

[o] Each component was assigned 0, 0.5 or 1 point. BMI and waist circumference, as well as fruits and vegetables and total fibre received half points (0, 0.25 or 0.5).

Mechanisms

Various mechanisms have been hypothesised to underlie the associations between lifestyle factors and CRC risk. These are related to, for example, inflammation, cell proliferation, differentiation and apoptosis, the volume and transit of faecal bulk through the colorectum, and the function of the gut microbiome [5]. Examples of the potential mechanisms are presented in Figures 2 and 3.

As discussed in sub-section 2.1.6, the beneficial association of whole grain intake with CRC risk has primarily been attributed to the high fibre content of whole grains [5, 11, 69]. However, whole grains also contain various other bioactive compounds, such as polyphenols, that may play a role in their potential beneficial effects [69, 70]. Similarly, the link between dairy consumption and lower CRC risk has largely been attributed to calcium intake [5, 113]. Additionally, dairy products are often fortified with vitamin D, which may have CRC-preventive effects [114]. The adverse association of red and processed meat with CRC risk is thought to arise from their high content of fat, protein and haem iron, as well as heterocyclic amines (HCA) and polycyclic aromatic hydrocarbons (PAH), which are formed when meat is cooked in high temperatures [5, 115, 116]. Alcohol (ethanol) has various detrimental effects on health, which are at least partly mediated through the conversion of ethanol to toxic acetaldehyde [117]. The association between physical activity and lower CRC risk is likely to be primarily related to a reduced risk of obesity [5], which itself is a major risk factor for CRC [118].

Although several putative mechanisms have been identified, uncertainty remains regarding the exact pathways underlying the associations between lifestyle factors and CRC risk. This thesis focuses on elucidating the role of the gut microbiome in these associations. Current knowledge of the interplay between these factors is discussed in the following section.

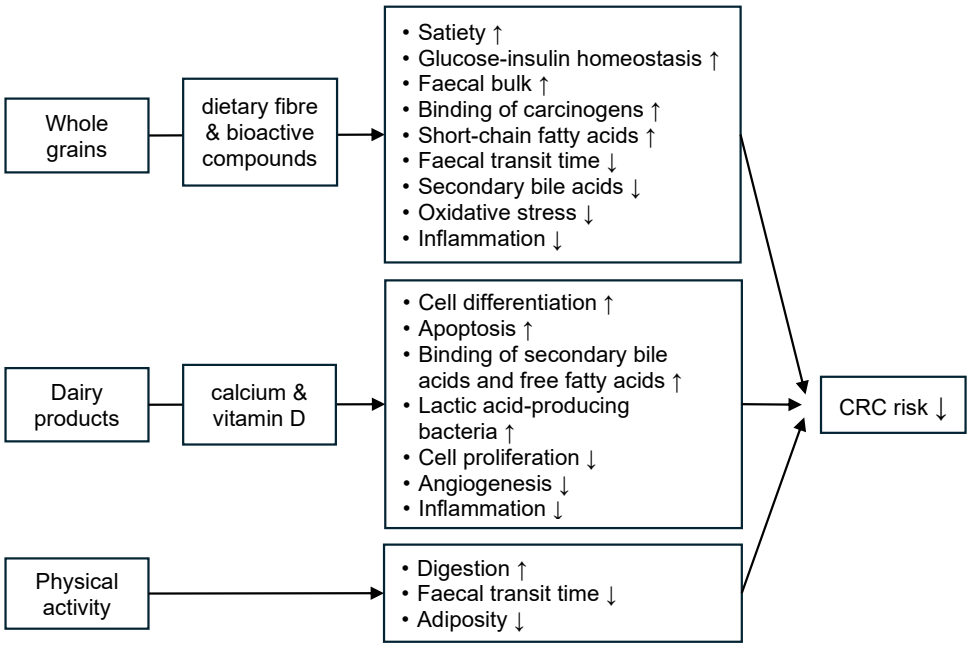


Figure 2 Examples of potential mechanisms underlying the protective associations between lifestyle factors and colorectal cancer (CRC) risk
↑ represents an increase; ↓ represents a decrease

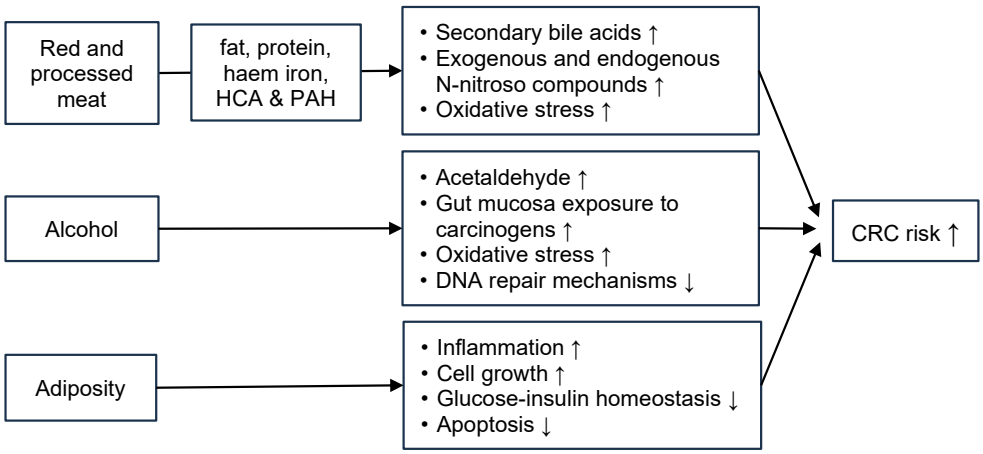


Figure 3 Examples of potential mechanisms underlying the adverse associations between lifestyle factors and colorectal cancer (CRC) risk
HCA, heterocyclic amines; PAH, polycyclic aromatic hydrocarbons
↑ represents an increase; ↓ represents a decrease

2.2.4 Colorectal cancer and the gut microbiome

Gut microbiome

Gut microbiome refers to the diverse ecosystem of microorganisms, such as bacteria, archaea, viruses and fungi, inhabiting the intestinal tract. Among these, the most prevalent are bacteria, with the dominant bacterial phyla being Bacillota (formerly Firmicutes) and Bacteroidota (formerly Bacteroidetes) [119, 120]. The composition of the gut microbiome ecosystem is strongly shaped by environmental factors, including diet and other lifestyle factors, whereas the contribution of host genetics appears more limited [7]. The gut microbiome may affect human health through several pathways, as it has been shown to be involved in, for example, the maturation and education of the immune system, protection against pathogens, elimination of toxins, host-cell growth, proliferation and apoptosis, energy metabolism, nutrient processing, and production of various metabolites, such as SCFAs and secondary bile acids [121–123]. The gut microbiome's contribution to these processes – and ultimately to human health and disease pathogenesis – depends on the composition and function of the ecosystem. Disturbances in these may promote pathogenic processes, such as inflammation, increased production of harmful metabolites and reduced production of beneficial metabolites [6]. Such adverse alterations in the microbiome are called dysbiosis, which has been linked to various diseases, including different cancers.

Colorectal cancer and the gut microbiome

Epidemiological studies have reported differences in the diversity and composition of the gut microbiome between CRC patients and healthy controls across different populations [120, 124–128]. The findings suggest the presence of dysbiosis in CRC patients, including enrichment in several putative pathogenic species, such as those normally inhabiting the oral cavity, and depletion of beneficial species, such as SCFA-producing bacteria. Moreover, CRC has been linked to a shift in the microbial pathways from utilisation of carbohydrates to degradation of amino acids [120, 124].

The bacterial species with the strongest evidence of association with CRC include *Fusobacterium nucleatum*, enterotoxigenic *Bacteroides fragilis* and cytotoxic *Escherichia coli* [129]. Enrichment in *F. nucleatum*, for example, has been observed in cancerous tissue compared with healthy tissue [130], as well as in faecal samples of individuals with CRC compared with healthy controls [120, 124–126, 131]. *F. nucleatum* is a commensal oral bacteria that is rarely found in a healthy colon [129]. It has been implicated particularly in CRC progression, including the promotion of chemoresistance, in later CRC stages [129, 131]. Cytotoxic *E. coli*, in turn, has been shown to induce DNA damage, whereas enterotoxigenic *B. fragilis* appears to contribute to CRC pathogenesis through inflammation particularly in the early disease stages [129].

In general, changes in the gut microbiome composition and function seem to take place throughout CRC development and progression, with certain shifts characterising specific stages of the carcinogenesis [131, 132]. Therefore, studying the microbial traits across the carcinogenesis from precancerous to more advanced stages is imperative.

Interplay between lifestyle, the gut microbiome and colorectal cancer

In addition to its direct link to CRC, the gut microbiome has been associated with major lifestyle-related risk factors for the disease, suggesting it may mediate the relationship between these factors and CRC risk. This sub-section examines the potential mechanisms underlying the interplay between lifestyle, the gut microbiome and CRC risk.

As previously discussed, whole grains are high in dietary fibre, which is an important substrate for microbial fermentation yielding SCFAs, such as butyrate, acetate and propionate. Particularly butyrate has been linked with anticarcinogenic effects related to reduced inflammation and epigenetic regulation of host-cell proliferation and apoptosis [133]. In a recent systematic review of twelve randomised controlled trials (RCTs), most studies reported increased total SCFA levels following fibre interventions in apparently healthy adults [134]. In another meta-analysis of 51 RCTs, fibre interventions increased faecal butyrate concentration and the relative abundance of *Bifidobacterium* and *Lactobacillus* [135]. Although these genera are known SCFA producers, they do not produce butyrate directly, but may promote its production through microbial cross-feeding [68]. *Bifidobacterium* and *Lactobacillus* can also metabolise ferulic acid, one of the main phenolic compounds in whole grains, into dihydroferulic acid, which may contribute to the protective effects of whole grains against CRC [6].

Similarly, consumption of dairy can contribute to the prevention of CRC through the gut microbiome by promoting the growth of beneficial bacterial species and the production of SCFAs [136]. Consumption of fermented dairy products, in particular, has been associated with a higher relative abundance of beneficial bacterial species, such as those of the genera *Lactobacillus* and *Bifidobacterium,* in intervention and observational studies [137, 138]. In an RCT of 28 women with irritable bowel syndrome, the consumption of fermented dairy products enhanced colonic bacterial production of SCFAs and reduced the relative abundance of a potential pathogen *Bilophila wadsworthia* [139]. *B. wadsworthia* is a known producer of hydrogen sulphide, which has been linked to carcinogenesis [140]. The benefits of fermented dairy products may be associated with the bacterial species they contain as starter cultures for fermentation [137]. Another potentially beneficial component of dairy products is lactose, which is a substrate to the microbial production of SCFAs [141]. Conversely, dairy products are often high in saturated fat, which may increase the risk of CRC by promoting the production of

secondary bile acids (discussed further below) [142]. However, the benefits of dairy consumption in CRC prevention appear to exceed their potential adverse effects [5].

The association between physical activity and the gut microbiome remains inconclusive, although some studies have reported greater microbial diversity in more active compared with less active individuals [143]. Moreover, both observational and intervention studies have linked physical activity with a higher abundance of SCFA-producing bacteria, representing a potential pathway in the interplay between physical activity, the gut microbiome and CRC risk.

Several compounds in red and processed meat are substrates to microbial processes that produce potentially carcinogenic metabolites. As illustrated in Figure 3, the consumption of red and processed meat has been found to increase the level of secondary bile acids, which are putative mutagens promoting oxidative stress [5, 144]. Secondary bile acids are produced by anaerobic bacteria from primary bile acids, which are released into the small intestine in response to fat intake [144]. Primary bile acids are predominantly reabsorbed in the small intestine, but a fraction ends up in the colon to be metabolised by microbes. Because red and processed meat are high in saturated fat, their consumption triggers increased release of primary bile acids, resulting in higher levels of them reaching the colon, thereby promoting the production of secondary bile acids. Red and processed meat are also high in choline, carnitine and sulphur, which are substrates to microbial metabolism [6, 144]. Choline and carnitine are metabolised into trimethylamine (TMA), which is further oxidised into trimethylamine N-oxide in the liver (TMAO) [145]. TMAO has been primarily linked to CHD [146], but an association with CRC has also been reported [147, 148]. From sulphur, which is found in red and processed meat both in inorganic (preservative in processed meat) and organic form, the gut microbiome produces hydrogen sulphide [6]. Hydrogen sulphide may promote cancer, for example, by enhancing proliferation, angiogenesis and metastasis [140]. However, in certain levels, it has also been associated with cancer-preventive effects. Observational studies have reported an enrichment in hydrogen sulphide producing bacteria in CRC patients compared with healthy counterparts [131, 132].

The interplay between alcohol consumption, the gut microbiome and CRC risk may involve gut microbiome dysbiosis, including an enrichment of potentially harmful taxa and a reduced abundance of beneficial species [149, 150]. At the phylum level, previous case-control studies have reported a reduced abundance of Bacteroidota and an enrichment in Proteobacteria in individuals with alcoholism compared with healthy controls. One microbiome-mediated pathway linking alcohol consumption to CRC risk is the microbial conversion of ethanol to carcinogenic acetaldehyde, exposing colorectal cells to its mutagenic effects [151, 152].

Obesity has been linked to changes in the gut microbiome composition and function both in human and animal studies [144]. In a systematic review of 22 cross-sectional human studies, most studies reported significantly lower microbial diversity in individuals with obesity compared to those with normal weight [153]. Similar result was reported in a study comparing twins with obesity and normal weight [154]. Regarding specific taxa, both obesity and weight loss in individuals with obesity have been associated with changes in the relative abundance of the phyla Bacillota and Bacteroidota [155]. Obesity has also been linked to a lower abundance of SCFA-producing bacteria, such as *Bifidobacterium* [156]. These obesity-related changes in the gut microbiome may promote carcinogenesis via several pathways, including inflammation [157] and epigenetic remodelling [158].

Although several potential pathways linking lifestyle factors, the gut microbiome and CRC risk have been identified, the precise mechanisms, and the extent to which the microbiome contributes to these relationships, remain uncertain. Moreover, it is unclear how lifestyle factors together interact with the gut microbiome in relation to CRC risk, and whether certain lifestyle factors promote changes in the microbiome that potentiate the effects of others. In observational settings, the combined associations of CRC-related lifestyle factors with the gut microbiome remain largely overlooked. These could be examined, for example, by using an index based on major lifestyle-related risk factors for CRC.

2.2.5 Summary of colorectal cancer research

CRC is one of the most common cancer types worldwide and has a relatively high mortality rate, especially at the metastatic stage. With a substantial proportion of CRC cases being attributable to modifiable lifestyle factors, CRC is considered a highly preventable disease. Consequently, primary prevention through lifestyle modifications is critical in alleviating the burden of CRC in populations worldwide.

Major lifestyle-related risk factors for CRC include obesity, low physical activity and diet. Regarding diet, particularly reducing red and processed meat consumption and increasing whole grain and dietary fibre intake have been emphasised in CRC prevention. In general, dietary characteristics linked to a lower CRC risk align well with the globally advocated shift toward more plant-based diets. However, epidemiological studies that model the replacement of red and processed meat with plant-based foods in relation to CRC risk at the population level are limited.

The gut microbiome has been hypothesised to mediate the associations between lifestyle factors and CRC. Although various observational and intervention studies have explored the associations between individual lifestyle factors, the gut microbiome and CRC risk, the interplay between these is not yet fully understood.

Moreover, research accounting for the whole lifestyle beyond individual factors remains scarce.

2.3 Summary of the literature review

The rising incidence of CRC poses a considerable public health challenge in high-income countries, and is becoming increasingly relevant in lower-income settings. Several modifiable lifestyle factors have been identified as targets for CRC primary prevention, aligning with broader health and sustainability goals. Although evidence has accumulated on associations between individual dietary factors and CRC risk, the potential benefits of more comprehensive dietary changes, including simultaneous decreases in adverse and increases in beneficial food groups, have been less studied. Furthermore, the mechanisms underlying the relationships of dietary and other lifestyle factors with CRC risk, including the contribution of the gut microbiome, remain poorly known.

Strong evidence has shown that whole grain intake lowers CRC risk and protects against several other lifestyle-related chronic diseases. Since cereal grains are staple foods in diets worldwide, promoting higher whole grain intake has substantial potential for improving public health. However, stronger evidence is required particularly on whole grains' associations with chronic disease risk factors to improve understanding of the pathways linking whole grain intake to health outcomes. Achieving this also requires standardisation of methods for estimating whole grain intake, as methodological inconsistencies have thus far limited the comparability and generalisability of results.

Overall, facilitating the lifestyle changes needed to reduce CRC burden and improve population health requires effective measures grounded in rigorous scientific evidence and robust methodologies. Evidence is required not only on the associations between lifestyle factors and CRC risk, but also on the mechanistic pathways underlying these relationships.

3 Thesis aims

The general aim of this thesis was to examine CRC risk in relation to healthy lifestyles, and the role of the gut microbiome in these relationships. Of individual CRC-related lifestyle factors, this thesis focused on whole grain intake, its relationship with health, and the methodological considerations in whole grain research. The specific aims of the four sub-studies were:

1. To evaluate the validity of commonly used or potential surrogate measures for whole grain intake (bread; rye bread; rye, oat and barley combined; rye; dietary fibre) (I)
2. To study associations between whole grain intake, diet quality and chronic disease risk factors (e.g., BMI, blood pressure, cholesterol) (II)
3. To study associations between partial substitutions of red or processed meat with plant-based foods and CRC risk (III)
4. To construct a CRC lifestyle index and examine its association with CRC risk (IV part A)
5. To study associations between the CRC lifestyle index and gut microbiome diversity and composition (IV part B)

4 Methods

4.1 Study populations

This thesis was based on seven study populations of Finnish adults from studies coordinated by the Finnish Institute for Health and Welfare (THL). An overview of each study population is provided in the following paragraphs and illustrated in Figures 4 and 5. Of the thesis sub-studies, I, II and IV part B were cross-sectional studies, whereas III and IV part A were prospective studies. The prospective study design was achieved through following the participants using national registers.

4.1.1 Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (III, IV part A)

The Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (ATBC), coordinated jointly by THL and the US National Cancer Institute, was initially a randomised trial studying the primary prevention of lung cancer through alpha-tocopherol and beta-carotene supplementation [159]. The participants were recruited from the male population aged 50–69 years from 14 adjoining areas in the Southwest Finland. The recruitment continued from April 1985 to June 1988, during which 283 356 men with known addresses in the Central Population Register (98% of the target population) received by mail a study invitation and an initial questionnaire on smoking habits and willingness to participate. Current smokers who smoked at least five cigarettes per day and were willing to participate were invited for a further examination of eligibility. The exclusion criteria included history of cancer (excluding non-melanoma skin cancer), other severe medical condition (e.g., liver cirrhosis), condition potentially limiting participation (e.g., psychiatric disorder), current use of anticoagulant medication and current use of vitamin E, vitamin A or beta-carotene supplements exceeding predefined doses. After these exclusions, 29 133 individuals were enrolled in the trial. Before trial commencement, comprehensive information on lifestyle, health and background characteristics was collected from the participants in health examinations and through self-administered questionnaires, including a food frequency questionnaire (FFQ). The trial ended on 30 April 1993, after which the cohort has been followed through national registers.

4.1.2 National Health 2000 Health Examination Survey (III, IV part A)

The National Health 2000 Health Examination Survey (Health 2000) was a population-based study designed to examine major public health challenges, functional capacity and their determinants in adults aged 30 years and older [160]. In a two-stage stratified clustering, 80 health centre districts were selected from the 249 districts in mainland Finland, from which a random population sample was drawn using the Finnish Population Information System. The final sample comprised 8028 individuals who were invited by mail to a home-visit health interview. During the interview, a health examination was scheduled to take place a few weeks later. Of those invited, 6986 (87%) participated in the interview and 6772 (84%) in the health examination. The study also included self-administered questionnaires, including an FFQ.

4.1.3 Helsinki Birth Cohort Study (III, IV part A)

The Helsinki Birth Cohort Study (HBCS) examined the associations of early-life factors on later health in a birth cohort including 8760 individuals born at the Helsinki University Central Hospital or the Maternity Hospital in the years 1934–1944 [161]. In 2000, the living cohort members (n=7078) received a health questionnaire by mail, to which 4515 (64%) responded. Of those, a random sample of 2902 individuals was drawn and invited to a clinical examination. The examinations took place in the years 2001–2004, during which 2003 (69%) individuals participated and completed self-administered questionnaires, including an FFQ.

4.1.4 FinRisk 2002 (IV part B) and FinRisk 2012 Studies (III, IV part A)

The FinRisk Studies were population-based health examination surveys conducted in Finland every five years between the years 1972 and 2012 [162]. The studies aimed at monitoring health behaviour and risk factors of chronic diseases in the adult population. The FinRisk 2002 Study (FinRisk 2002) covered six and the FinRisk 2012 Study (FinRisk 2012) five large geographical areas of mainland Finland. The study samples were drawn from the Population Information System and stratified by sex, 10-year age group and geographical area. In FinRisk 2002, a random sample of 13 498 individuals aged 25–74 years was invited to the study, including a health examination and self-administered questionnaires. Of those, 8799 (65%) participated. Correspondingly, in FinRisk 2012, 9905 individuals aged

25–74 years were invited and 5827 (59%) participated. In FinRisk 2012, the self-administered questionnaires included an FFQ.

In FinRisk 2002, all participants were asked to donate a faecal sample, which was ultimately received from 7211 individuals (response rate 87%). Additionally, FinRisk 2002 included a sub-study called the FinDiet 2002 Survey (FinDiet 2002), in which detailed dietary data were collected from the participants using a 48-hour dietary recall (48h recall) [163]. The individuals invited to FinDiet 2002 (n=3181) were randomly selected from the original FinRisk 2002 sample (32%), excluding those from the Province of Lapland to match with the FinDiet 1997 Study. Of those invited, 2007 participants (63%) completed the 48h recall acceptably.

FinRisk/FinDiet 2002 was used in Sub-study IV part B as cross-sectional, but prospective data on colorectal cancer cases were additionally obtained to inform interpretation.

4.1.5 Dietary, Lifestyle and Genetic Determinants of Obesity and Metabolic Syndrome 2007 Study (III, IV part A)

The Dietary, Lifestyle and Genetic Determinants of Obesity and Metabolic Syndrome 2007 Study (DILGOM 2007) was a sub-study of the FinRisk 2007 Study (FinRisk 2007), which was conducted following the same principles as FinRisk 2002 and 2012 [164]. In FinRisk 2007, 9958 individuals aged 25–74 years from five large geographical areas of mainland Finland were invited to participate in a health examination and complete self-administered questionnaires. All participants (n=6258, response rate 63%) were invited to the second study phase (DILGOM 2007), which included a second health examination and additional self-administered questionnaires (e.g., FFQ) targeting determinants of obesity and metabolic syndrome. Of those invited, 5024 (80%) participated in the health examination and completed the questionnaires.

4.1.6 FinHealth 2017 Study (I, II)

The FinHealth 2017 Study (FinHealth 2017) was a national health examination survey conducted as a continuation of the Health 2000 and FinRisk studies [165]. It examined the health, health behaviour and functional capacity of the adult population aged ≥18 years residing in mainland Finland. A nationally representative random sample of 10 247 individuals from 50 (of 80) health centre districts was drawn from the Population Information System and invited by mail to participate in a health examination and complete an initial self-administered questionnaire. Of those invited, 5952 (58%) participated in the health examination, during which the remaining self-administered questionnaires, including an FFQ, were distributed.

4.1.7 Exclusion criteria and final study samples

In each sub-study, the final study sample consisted of participants who participated in a health examination and acceptably completed either an FFQ (sub-studies I–IV part A) or a 48h recall (IV part B) (Figures 4 and 5). In Sub-study IV part B, a successfully sequenced faecal sample was additionally required. An FFQ was considered incomplete if it contained multiple empty food item rows, with exclusions made on a case-by-case basis. For the 48h recall, exclusions were made if only one day was recorded or if the recall was otherwise deemed incomplete, for example, due to participant memory problems during the interview [163].

In each study sample, participants with implausible energy intake were excluded. In ATBC and Health 2000, implausible energy intake was identified based on predetermined cut-off values (ATBC: <1000 or >5000 kilocalories [kcal] per day; Health 2000: <600 or >7000 kcal/day). In the other study samples, participants within the 0.5% sex-specific extremes of the energy intake distribution were excluded. Other exclusion criteria were consent withdrawal (all sub-studies), pregnancy (I, II, IV part B), history of cancer (III, IV), use of antimicrobial medication (IV part B) and missing data on the CRC lifestyle index components (IV).

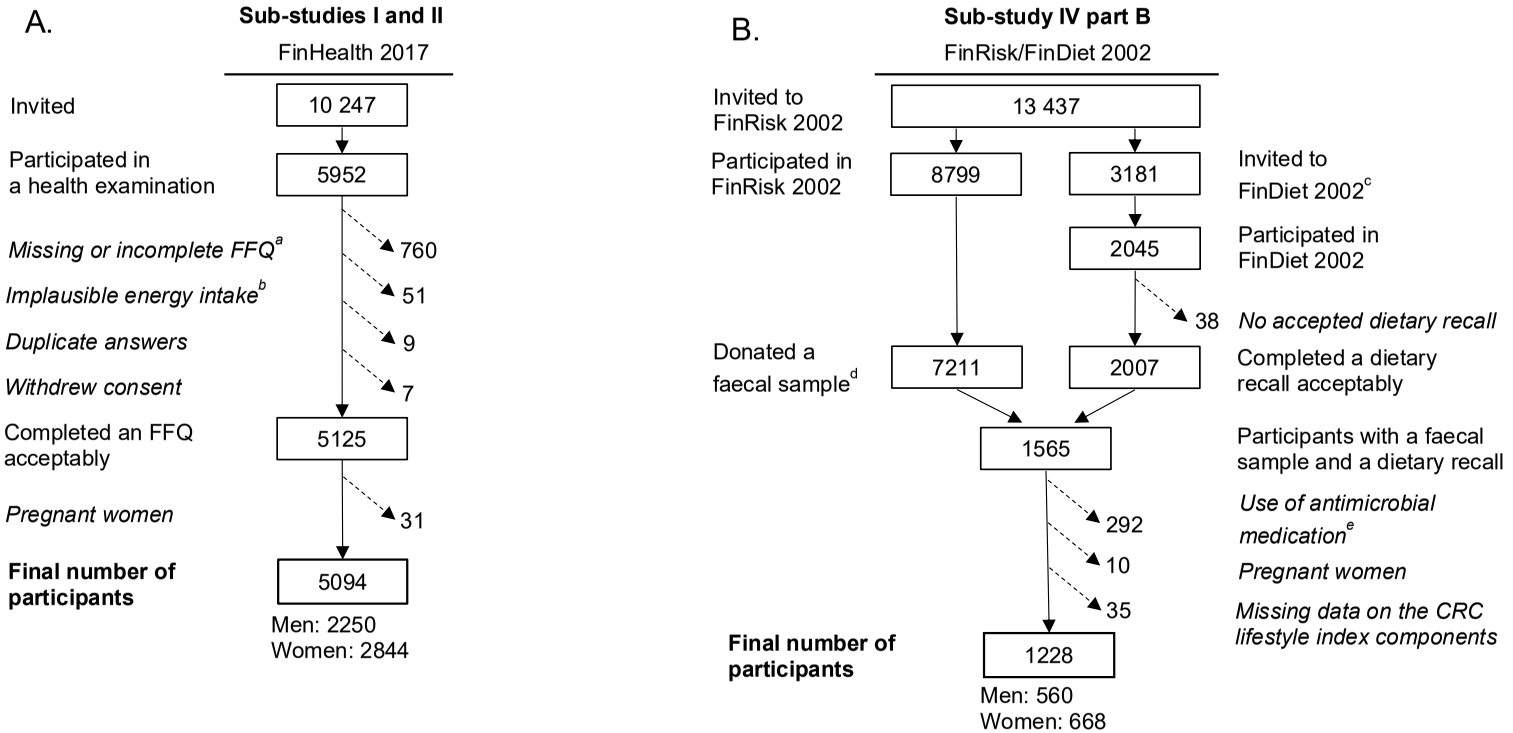


Figure 4 Exclusion criteria and the final study samples in sub-studies A) I and II and B) IV part B

[a] Incomplete questionnaire with multiple empty food item rows (exclusions made on a case-by-case basis)
[b] 0.5% sex-specific extremes of the energy intake distribution
[c] The FinDiet sub-sample (32%) was randomly selected from the original FinRisk 2002 sample
[d] Successfully sequenced faecal sample with sufficient read counts (≥50 000)
[e] Registered purchase of systemic antimicrobial medication (Anatomical Therapeutic Chemical classification code J01) within six months prior to the baseline examination

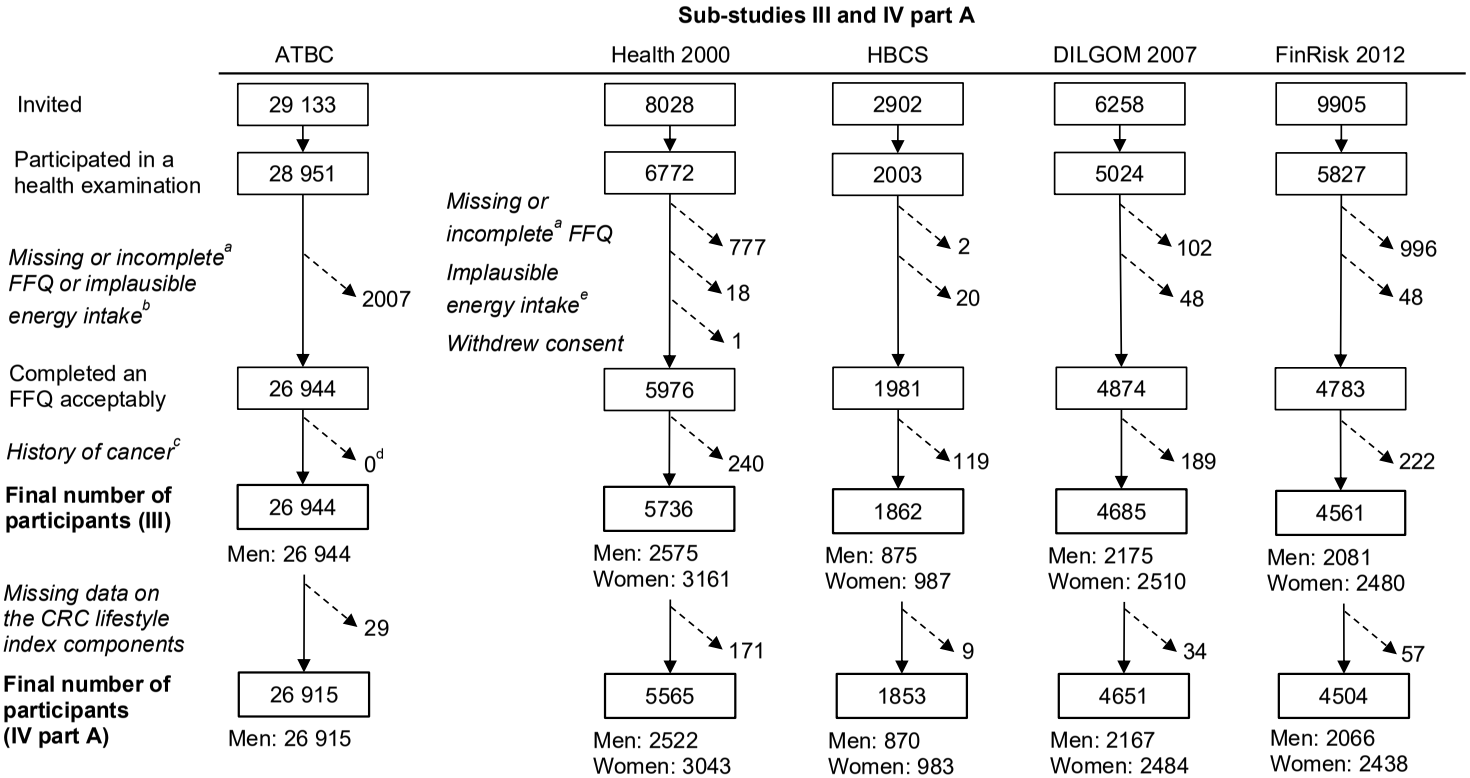


Figure 5 Exclusion criteria and the final study samples in sub-studies III and IV part A

[a] Incomplete questionnaire with multiple empty food item rows (exclusions made on a case-by-case basis)

[b] Energy intake <1000 or >5000 kcal/day

[c] History of cancer other than non-melanoma skin cancer

[d] Individuals with history of cancer (other than non-melanoma skin cancer) were excluded in the recruitment process.

[e] Health 2000: <600 or >7000 kcal/day; HBCS, DILGOM 2007 and FinRisk 2012: 0.5% sex-specific extremes of the energy intake distribution

4.2 Ethical approval

All studies providing the study populations of this thesis were conducted in accordance with the guidelines laid down in the Declaration of Helsinki and following the ethical standards in effect at the time of the study. ATBC was approved by the institutional review boards of both the US National Cancer Institute and THL (at the time National Public Health Institute of Finland). The other studies were approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa. All participants provided written informed consent. The individual-level data was pseudonymised in accordance with the General Data Protection Regulation, and the encryption keys linking the gathered data to each participant were stored separately and accessible only to the dedicated data management staff.

4.3 Dietary assessment

4.3.1 Food frequency questionnaire

Dietary intake was examined in sub-studies I–IV part A using data from a comprehensive semiquantitative food frequency questionnaire (FFQ) measuring habitual food consumption over the past year [80, 166–168]. An FFQ is a standard dietary assessment method in epidemiological research, primarily designed to assess habitual dietary intake and rank individuals by intake levels [169]. In ATBC, the participants completed a validated 276-item FFQ at home [170], recording their usual consumption frequency of each item based on 3–5 portion sizes, with the help of a portion-size picture booklet [171]. The participants then reviewed the FFQ together with a study nurse during a health examination. The FFQ used in the other study samples (Health 2000, HBCS, DILGOM 2007, FinRisk 2012 and FinHealth 2017) was originally developed and validated within the Kuopio Breast Cancer Study [166], and it has subsequently undergone two further validations in the general adult population [167, 168]. The FFQ included approximately 130 items, the number varying slightly between studies due to regular updates based on the National FinDiet Surveys [163, 172–175]. The consumption of each item was recorded using nine frequency categories ranging from 'never or seldom' to 'at least six times a day', with fixed portion sizes defined as household measures, natural units (e.g., glass, slice) or servings. The portion sizes for each item were derived from dietary recall data collected in the FinDiet Surveys [163, 172–175]. Sex-specific portion sizes were used in each study except for Health 2000. In Health 2000 and FinRisk 2012, the participants completed the FFQ at home and returned it by mail to THL, whereas in HBCS and DILGOM 2007, the FFQ was completed at the study

site. In FinHealth 2017, the participants completed the FFQ either electronically or on paper at the study site or at home.

4.3.2 48-hour dietary recall

In Sub-study IV part B, dietary intake was assessed using data from a 48-hour dietary recall [163]. Dietary recalls provide data on absolute dietary intake over shorter time, which makes them particularly suitable for monitoring population dietary intake and for comparisons with dietary recommendations [176]. In FinDiet 2002, trained nutritionists carried out computer-assisted recall interviews during the health examination, documenting all foods and beverages consumed within the two preceding days. Portion sizes were estimated based on a validated portion-size picture booklet [171], household measures and standard food packaging.

As an exception, in Sub-study IV part B, alcohol consumption was estimated from data collected through a self-administered questionnaire within FinRisk 2002. The questionnaire inquired about the number of portions and frequency of consumption of different alcoholic beverages over the past year. Using this data allowed a more accurate estimation of alcohol consumption, which is often underestimated in dietary recalls due to high day-to-day variability.

4.3.3 Dietary intake calculation

To calculate the average daily consumption of foods and intake of energy (kJ/d) and nutrients (g/d), dietary data from the FFQs and 48h recalls were linked to the Finnish National Food Composition Database Fineli® using an in-house calculation software [173]. Weekly consumption (if needed) was calculated by multiplying daily consumption by seven. Within the calculation processes, all mixed dishes and foods were disaggregated into their basic ingredients (e.g., uncooked beef, wheat flour) based on standard recipes in the database. The basic ingredients were then classified into more generic food groups relevant for this thesis (e.g., red meat, including beef, pork and lamb; fruits, including all fruits and berries) [173]. As an exception, in Sub-study I, bread and rye bread were examined as such, without disaggregation into their basic ingredients.

4.3.4 Whole grain database

As previously mentioned, whole grain was defined in this thesis based on the HEALTHGRAIN Forum definition, which was, at the time, the most recent and up-to-date definition with widespread use [29]. The HEALTHGRAIN definition aligns well with the more recent WGI standard definition [21], and describes whole grains as intact, ground, cracked or flaked kernels in which the major anatomical

components remain in the same relative proportions as in an intact kernel (see subsection 2.1.2 Table 2). In addition to cereal grains, the HEALTHGRAIN definition also considers pseudocereals.

Due to the lack of whole grain data in Fineli®, estimating total whole grain intake in the population-based data of Finnish adults required compiling a new whole grain database. To achieve this, whole grain content was first assigned by hand for each basic ingredient in Fineli®. Cereal-containing basic ingredients included, for example, flours, flakes, cereal mixtures and rice, as well as certain foods that are not further disaggregated within the database, such as oat- and rice-based drinks, muesli bars and pasta. Second, the whole grain content of basic ingredients was used to calculate the whole grain content of foods based on their standard recipes. The recipe calculations were conducted using an in-house software [173], and the assigned values were manually verified. To support the database compilation, product labels and information from food manufacturers were used when needed. Whole grain values were assigned as grams in dry weight per 100 g [20].

4.3.5 Diet quality

Diet quality was assessed using the modified Baltic Sea Diet Score (mBSDS) developed to reflect healthy food choices in the Nordic context [177]. The mBSDS comprises eight components, including cereals, fruits, vegetables, low-fat milk, fish, red and processed meat, alcohol (100% ethanol) and the ratio of polyunsaturated fatty acids to saturated and trans fatty acids (fat ratio). As diet quality was studied in relation to whole grain intake in this thesis (Sub-study II), cereals were excluded from the score to avoid artificially strengthening the association. The remaining components were assigned zero to three points based on sex-specific population consumption quartiles; ascending points to fruits, vegetables, low-fat milk, fish and the fat ratio, and descending points to red and processed meat. As an exception, alcohol was scored based on the 2012 Nordic Nutrition Recommendations, with one point assigned for consumption within the recommended limits (women ≤10 g/d, men ≤20 g/d) and zero points for consumption exceeding these limits [178]. As a result, the mBSDS points could range from zero to 19, with higher points indicating a better adherence to a healthy Nordic diet.

4.4 Sociodemographic and lifestyle factors

In each study sample, information on participants' sex and age originated from the sampling frame (Population Information System). Participants reported their educational attainment (total number of school years), leisure-time physical activity and smoking habits in self-administered questionnaires. Based on the total number of school years, participants were categorised into tertiles corresponding to

low, intermediate and high educational attainment, stratified by sex and birth cohort to account for the extension of the basic education system and the increase in average school years over time. As an exception, in ATBC, low educational attainment was defined as elementary or lower education, intermediate as lower or upper secondary education and high as higher than upper secondary education. Based on the reported leisure-time physical activity, participants were categorised into inactive (only light activities, e.g., reading), moderately active (e.g., walking and gardening ≥4 h/wk) and active (e.g., running, swimming and competitive sports ≥3 h/wk) [179]. Based on the reported smoking habits, participants were categorised into current smokers, former smokers and nonsmokers. In Sub-study IV part B, a binomial smoking variable was created by combining current smokers with former smokers who had stopped smoking within the past six months, and nonsmokers with former smokers who had stopped smoking more than six months ago. In ATBC, all participants were current smokers as per the study design [159].

The self-administered questionnaires also included questions on the use of various medications, including hormone-replacement therapy in women (HRT; used ever/never), lipid-lowering medication (current use, yes/no), antihypertensive medication (used never/last time today or yesterday/2–7 days ago/one week to six months ago/6–12 months ago/1–5 years ago/>5 years ago) and diabetes medication (not using medication/using insulin/tablets/insulin and tablets together). In this thesis, current use of antihypertensive medication was defined as use within the past seven days and current use of diabetes medication as use of either insulin or tablets (or both).

4.5 Clinical examinations

4.5.1 Anthropometric measures and blood pressure

In each study sample, weight (kg), height (cm) and WC (cm; not measured in ATBC) were measured in the health examination by trained research staff using international standard protocols [159, 160, 162, 165, 180, 181]. Weight was measured to the nearest 0.1 kg, with participants wearing light clothing and no shoes. Height was measured to the nearest 1 cm in ATBC, 0.5 cm in Health 2000 and 0.1 cm in the other studies, with participants barefoot. WC was measured to the nearest 0.5 cm (0.1 cm in FinHealth 2017) using a measuring tape placed at the midpoint between the lowest rib and the iliac crest, with participants not wearing any upper-body clothing that could interfere with the measurement. BMI (kg/m^2) was calculated as weight divided by height squared. Normal weight (18.5–<25 kg/m^2), overweight (25–<30 kg/m^2) and obesity (≥30 kg/m^2) were defined following the World Health Organization (WHO) guidelines [182].

In FinHealth 2017, blood pressure was measured in the health examination using a mercury sphygmomanometer [165]. Measurements were taken three times from the right arm with the participant seated, and the average of the second and third readings was calculated.

4.5.2 Blood sample collection and analysis

In FinHealth 2017, blood samples were collected in the health examination after a minimum of four hours of fasting and refraining from strenuous exercise [165]. The collection included whole blood, serum and fluoride-citrate plasma. After collection, the samples were kept at room temperature for no longer than 60 minutes before centrifugation. Following appropriate processing steps, the samples were frozen, stored at -20°C and subsequently transferred to the accredited biochemistry laboratory of THL, where whole blood and serum samples were stored at -70°C and fluoride-citrate plasma at -20°C [165]. The samples were analysed at the laboratory using a clinical chemistry analyser Architect ci8200. Total cholesterol, high-density lipoprotein (HDL) cholesterol and TAG were measured from the serum samples using the enzymatic Abbott method (Abell-Kendall verification). C-reactive protein (CRP) was measured from the serum samples using the immunoturbidimetric Abbott method (ERM-DA472/IFCC). Glucose was measured from the fluoride-citrate plasma samples using the enzymatic hexokinase Abbott method (NIST SRM 956). LDL cholesterol was calculated using the Friedewald equation [183].

4.5.3 Faecal samples and microbiome characterisation

In FinRisk 2002, all participants willing to donate a faecal sample received a sampling kit in the health examination with detailed instructions on its use [184]. The participants collected the samples at their earliest opportunity and sent them by mail in a prepaid parcel to the THL laboratory. The samples were collected into 50-millilitre Falcon tubes without a stabilising solution and delivered under typical Finnish winter conditions. At the laboratory, the samples were frozen and stored at -20°C until the transfer to the University of California San Diego for microbiome sequencing in 2017.

The sequencing library was prepared using a miniaturised Kapa HyperPlus library preparation kit (Kapa Biosystems), as described in detail by Salosensaari et al. (2021) [184]. Shallow-shotgun whole-metagenome sequencing was performed using an Illumina HiSeq 4000 instrument (Illumina Inc.), generating 150-bp paired-end reads [185]. On average, each sample had approximately 900 000 reads. Adapter sequences were removed using the fastp preprocessing tool (version 0.23.4) [186], and host sequences were filtered by mapping reads against the

human pangenome, T2T-CHM13v2.0 and GRCh38 using minimap2 [187]. Finally, unmapped reads were extracted using samtools [188]. The raw sequences were then taxonomically annotated using incorporated metagenomes in Greengenes2 microbial reference database [189].

The diversity and composition of the gut microbiome were examined through analyses of alpha diversity, beta diversity and differential relative abundances of individual microbial species. Alpha diversity quantifies microbial diversity within individuals, whereas beta diversity quantifies the variation in microbial community composition between individuals. Alpha diversity was measured based on species-level raw counts using the Shannon index [190], which assesses both microbial richness and evenness within a sample. Beta diversity was measured based on taxa relative abundances using the weighted UniFrac metric, which quantifies compositional dissimilarity between samples, considering the abundance of species and their evolutionary relationships [191]. Taxa relative abundances were calculated from the annotated data by scaling the number of raw read counts of each taxon to the total sum of reads. Differential abundances of individual species were examined based on species-level raw counts, excluding rare species with a prevalence of <1% or a relative abundance of <0.01% [184].

4.6 CRC lifestyle index

A CRC lifestyle index was constructed based on the standardised 2018 WCRF/AICR Score [192]. The standardised score operationalises the general cancer prevention recommendations of the 2018 WCRF/AICR Third Expert Report [54], for which I adapted the score for a CRC-specific approach. The new CRC lifestyle index included nine lifestyle and anthropometric factors with strong evidence of association with CRC, as presented in the 2018 WCRF/AICR report [5]. These were BMI, WC, height, leisure-time physical activity, consumption of whole grains, dairy products, red meat and processed meat, and intake of alcohol (Table 8). Of the strong-evidence risk factors in the 2018 WCRF/AICR report, fibre was omitted to avoid redundancy with whole grains. In previous meta-analyses of prospective studies, cereal fibre, in particular, has been linked to a decreased risk of CRC [11, 193], and whole grains are a major source of both total and cereal fibre in Finnish adults [55, 175]. Furthermore, by focusing on whole grains rather than fibre alone, the potential beneficial effects of other compounds in whole grains could be considered (see 2.1.6) [70]. The use of calcium supplements was also omitted because dietary calcium intake is generally sufficient in Finnish adults [163, 175].

Consistent with the standardised score, each of the included risk factors was assigned 0, 0.5 or 1 point representing not meeting, partially meeting or meeting the target cut-off determined based on the literature (Table 8). As an exception, height was scored using sex-specific tertiles, as standard cut-offs could not be

determined from the literature. In addition, leisure-time physical activity was scored using predefined categories of the variable available in the data used. Following the standardised score, points for BMI and WC were averaged to form one component reflecting body fatness; if data on either were missing for the entire dataset (WC in ATBC), the other was assigned full points [192]. To avoid overemphasising diet, the points for the four dietary components (red and processed meat as one) were also averaged to form a single component. The final CRC lifestyle index was calculated by summing the points for body fatness, height, physical activity and diet, resulting in total points ranging from 0 to 4, with higher points indicating fewer risk factors for CRC (a lower-risk lifestyle for CRC).

Table 8 The components and scoring of the CRC lifestyle index

Score	Body fatness[a]		Height[c]	PA[d]	Diet[e]			
	BMI[b] (kg/m^2)	WC[b] (cm)			WG[f] (g/d)	Dairy[f] (g/d)	RM & PM[g] (g/week)	Alcohol[h] (g/d)
0	≥30	W: ≥88 M: ≥102	T3	Inactive	<45	<200	RM >350 OR PM ≥100	>20
0.5	25–<30	W: 80–<88 M: 94–<102	T2	Mod. active	45–<90	200–<400	RM ≤350 & PM 21–<100	>0–≤20
1	<25	W: <80 M: <94	T1	Active	≥90	≥400	RM ≤350 & PM ≤21	0

BMI, body mass index; dairy, dairy products; M, men; PA, physical activity; PM, processed meat; RM, red meat; T, tertile; W, women; WC, waist circumference; WG, whole grains

[a] The score for body fatness was calculated by averaging the points for BMI and WC. If data on either were missing for the entire dataset, the other was assigned full points.

[b] The cut-offs are based on the 2018 WCRF/AICR Third Expert Report [5] and WHO guidelines [182, 194].

[c] Height was scored based on sex-specific tertiles. Tertile medians (cm) in FinRisk/FinDiet 2002 were in women T1: 157, T2: 163, T3: 169; in men T1: 170, T2: 176, T3: 183.

[d] Leisure-time physical activity was scored based on the predefined categories of the variable available in the data: inactive (only light activities, e.g., reading), moderately active (e.g., walking or gardening ≥4 h/week) or active (e.g., running, swimming, or competitive sports ≥3 h/week).

[e] The score for diet was calculated by averaging the points for whole grains, dairy products, red and processed meat and alcohol.

[f] The cut-off for meeting the target consumption is based on the 2018 WCRF/AICR Third Expert Report [5]. The cut-off for partially meeting the target consumption is defined as consuming at least half of the target amount [192].

[g] The cut-offs are based on the 2018 WCRF/AICR Third Expert Report [5], the standardized 2018 WCRF/AICR Score [192] and the Nordic Nutrition Recommendations 2023 [23].

[h] 100% ethanol; the cut-offs are based on the 2018 WCRF/AICR Third Expert Report [5].

4.7 Register data

National health registers were used to obtain information on participants' current disease status, prior medical history and medication use. The participants were linked to the registers using personal identity codes issued to all Finnish citizens and permanent residents.

Information on CRC cases (International Statistical Classification of Diseases and Related Health Problems [ICD]-9 codes 153, 154.0 and 154.1, or ICD-10 codes C18, C19.9 and C20.9) was obtained from the Finnish Cancer Registry, which upholds comprehensive high-quality nationwide records of all cancer cases diagnosed in Finland since 1953. The coverage of colorectal (and anal) cancers was 97% [195]. The participants in sub-studies III and IV were followed from the date of enrolment until CRC diagnosis, death or the censoring date, which was 31 December 2014 in HBCS, 2015 in Health 2000, 2016 in ATBC, 2019 in DILGOM 2007 and FinRisk 2012, and 2022 in FinRisk/FinDiet 2002.

Information on diabetes status at baseline (IV) was obtained from the Social Insurance Institution of Finland register of prescription medication purchases and medication reimbursements (Anatomical Therapeutic Chemical [ATC] codes A10), THL register of Health Care (ICD-10 codes E10–14) and Statistics Finland register of causes of death (ICD-10 codes E10–14).

Information on the use of potentially microbiome-altering medication was obtained from the Social Insurance Institution of Finland register of prescription medication purchases (IV part B). These included antimicrobial medication (ATC code J01), metformin (A10BA02), psycholeptics (NO5), psychoanaleptics (NO6), proton pump inhibitors (PPI; A02BC) and constipation medication (A06A). In Sub-study IV part B, participants with registered purchase of antimicrobial medication within six months prior to the baseline were excluded. Use of the other medications was included in the analyses as confounders. Participants were categorised as medication users (other than antimicrobial medication) if they had a registered purchase within four months prior to the baseline and at least three consecutive purchases.

4.8 Statistical analyses

In each sub-study, the descriptive statistics of participant characteristics included percentage distributions for categorical variables and means and standard deviations (SD) (sub-studies I and II) or medians and interquartile ranges (IQR) (sub-studies III and IV) for continuous variables. Non-normally distributed variables were log- or cube-root-transformed to meet the assumption of normality. Confounding factors were determined based on the literature. In sub-studies I and II, under-reporting of energy intake was assessed by calculating the ratio of

reported energy intake to predicted basal metabolic rate, with under-reporting defined as a ratio ≤1.14 [196, 197]. In each sub-study, statistical significance was determined as a two-sided $P<0.05$. *P* for trend over quantiles was calculated using the quantile medians as continuous independent variables. Interaction was tested by including an interaction term in the model. The primary statistical analyses and the models used are described in detail in Table 9.

4.8.1 Correspondence of surrogate measures with whole grain intake (I)

In Sub-study I, I examined the correspondence between whole grain intake and five potential whole grain surrogate measures (bread; rye bread; rye, oat and barley combined; rye; dietary fibre) using Spearman rank correlation and partial correlation adjusted for energy intake. The analyses were conducted in the overall population and within subgroups by sex, age (<median, ≥median), educational attainment (low/intermediate, high) and BMI (<30, ≥30 kg/m^2). *P*-values for subgroup differences were calculated using Fisher Z transformation. In addition, I examined cross-classification between the quintiles of whole grain intake and the surrogate measures by assessing the proportion of participants classified into 1) the same quintile, 2) the same or adjacent quintile, 3) the lowest quintile of both whole grain and a surrogate measure and 4) the lowest quintile of whole grain and the highest quintile of a surrogate measure (gross misclassification).

4.8.2 Associations between whole grain intake, diet quality and chronic disease risk factors (II)

In Sub-study II, I examined associations of whole grain intake with background factors (age, educational attainment, physical activity, smoking), diet quality (mBSDS) and chronic disease risk factors (BMI, WC, diastolic and systolic blood pressure, total, HDL and LDL cholesterol, TAG, CRP, glucose) using linear (continuous outcomes) and logistic (categorical outcomes) regression. Whole grain intake was modelled in quintiles and adjusted for energy intake using the residual method [198]. This involved calculating the residuals of a regression analysis between whole grain intake (outcome) and energy intake (exposure), and adding the residuals to the population mean of whole grain intake. All analyses were stratified by sex due to differences in whole grain intake and the outcome variables between women and men. To study whether diet quality modified the associations between whole grain intake and chronic disease risk factors, the analyses were also stratified by the mBSDS (in tertiles). In addition, interaction between whole grain intake and diet quality was tested by including an interaction term (whole grain intake*mBSDS) in the models.

4.8.3 Associations between partial substitutions of red or processed meat with plant-based foods and CRC risk (III)

In Sub-study III, I examined partial substitutions of red meat (100 g/week) or processed meat (50 g/week) with a corresponding amount of plant-based foods (100 or 50 g/week of whole grains, vegetables, fruits or a combination of these) in relation to CRC risk in five pooled study samples. The substituted amounts represented modest dietary changes, corresponding to the substitution of approximately one and a half day of red meat consumption and one day of processed meat consumption in our study population.

As a preliminary analysis, associations between the individual food groups (per 50 g/week for processed meat and 100 g/week for the other food groups) and CRC risk were examined using multivariate Cox proportional hazards regression. The substitution analyses were conducted using multivariate Cox regression and a leave-one-out model, estimating the effect of substituting red or processed meat with plant-based foods, while holding their total consumption constant [199, 200]. The leave-one-out model included the substitute variable (whole grains, vegetables, fruits or a combination of these) and a sum variable of the substitute and the food being substituted (e.g., whole grains + red meat) [199]. Thus, the model can be expressed as

$$f(Y) = \beta_1(substitute) + \beta_2(substitute + food\ being\ substituted) + confounders,$$

where Y represents CRC risk, β_1 is the beta coefficient for the substitution effect (change in CRC risk associated with a simultaneous decrease in red or processed meat consumption and increase in plant-based food consumption), and β_2 is the beta coefficient for the total effect of the substitute and the food being substituted. Hazard ratio (HR) for the substitution effect was calculated by exponentiating β_1.

All analyses were conducted using a two-stage meta-analysis, first calculating the cohort-specific effect estimates and then pooling these estimates using a random-effects model weighted by the inverse of their variances [201]. Between-cohort heterogeneity was tested using Q-statistics. Adjustments for confounders were applied in the cohort-specific analyses before pooling. The analyses were conducted in the overall population and stratifying by sex, age (<median, ≥median), BMI (<25, 25–<30, ≥30 kg/m2), HRT use (in women; never, ever), follow-up time (<median, ≥median) and whole grain intake (<median, ≥median; see 5.2.2). In addition, the analyses were stratified by study sample (ATBC, others) to account for differences arising from the study design of ATBC.

4.8.4 Associations between the CRC lifestyle index and gut microbiome diversity and composition (IV)

In Sub-study IV part A, I examined the association between the CRC lifestyle index (continuous and in quintiles) and CRC risk in five pooled study samples to test the performance of the index in the Finnish adult population. The effect estimates were calculated using multivariate Cox regression and pooled using two-stage meta-analysis as described in Sub-study III (4.8.3).

In Sub-study IV part B, I examined associations of the CRC lifestyle index (and the index components) with microbial alpha diversity, beta diversity and differential abundance of individual species. Continuous variables were standardised using Z-scores. Associations with alpha diversity were assessed using linear regression. Associations with beta diversity were studied using Permutational Multivariate Analysis of Variance (PERMANOVA) [202]. The analysis was run with 999 permutations and using dispersion analysis to check that the differences in beta diversity across the index were not driven by within-group variability. *P* values from the linear regression analyses and PERMANOVA were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) method [203]. Associations with differential relative abundances of all taxa at species level were examined using Analysis of Compositions of Microbiomes with Bias Correction 2 (ANCOM-BC2) [204]. Species that were statistically significantly associated with the index and passed the sensitivity screening for robustness (within ANCOM-BC2) were clustered using Ward's minimum variance method based on proportionality [205]. The optimal number of clusters was determined using Kelley-Gardner-Sutcliffe penalty function. All analyses were conducted in the overall population in accordance with IV part A.

Table 9 Primary statistical analyses and models used by sub-study

	Sub-study				
	I	**II**	**III**	**IV part A**	**IV part B**
Primary analyses	Spearman correlation, cross-classification	linear and logistic regression	Cox, substitution analysis	Cox	linear regression, PERMANOVA, ANCOM-BC2
Exposure(s)	bread; rye bread; rye, oat and barley combined; rye; dietary fibre	whole grain intake	substitution of red meat (100 g/wk) or processed meat (50 g/wk) with plant-based foods	CRC lifestyle index	CRC lifestyle index
Outcome(s)	whole grain intake	mBSDS, BMI, WC, DBP, SBP, TC, HDL, LDL, TAG, CRP, glucose	CRC risk	CRC risk	microbial alpha diversity, beta diversity, differential abundances of species
Stratification/ Interaction	sex, age, education, BMI	sex, mBSDS	study sample, sex, age, BMI, HRT[g], follow-up time, whole grain intake	sex	sex
***Adjustments*[a]**					
Main model	energy intake	age, energy intake, education, smoking, physical activity, BMI, mBSDS, sodium intake[b], added sugar intake[c]	sex, age, energy intake, education, smoking, physical activity, height, BMI, HRT[g], alcohol intake, dairy consumption	sex, age, energy intake, education, smoking	sex, age, energy intake, smoking, use of potentially microbiome-altering medication
Sensitivity analysis 1	excluding energy under-reporters	main model + energy under-reporting	excluding participants consuming red meat <100 g/wk or processed meat <50 g/wk	excluding participants diagnosed with CRC within the first two years of follow-up	

Table continues

Table 9 continues

	Sub-study				
	I	**II**	**III**	**IV part A**	**IV part B**
Sensitivity analysis 2		main model + use of anti-hypertensive[d], lipid-lowering[e] or diabetes[f] medication	excluding participants diagnosed with CRC within the first two years of follow-up time	main model + prevalent diabetes	
Sensitivity analysis 3		main model + fibre intake	main model + CRC family history[h]	main model + HRT[g]	
Statistical software	IBM SPSS v27	IBM SPSS v28	R v3.6	R v4.3	R v4.4

ANCOM-BC2, Analysis of Compositions of Microbiomes with Bias Correction 2; BMI, body mass index; Cox, multivariate Cox proportional hazards regression; CRC, colorectal cancer; DBP, diastolic blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; mBSDS, modified Baltic Sea Diet Score; PERMANOVA, permutational multivariate analysis of variance; SBP, systolic blood pressure; TAG, triacylglycerol; TC, total cholesterol; WC, waist circumference

[a] The outcome variable was excluded from the adjustments.
[b] When analysing blood pressure as the outcome
[c] When analysing TAG or glucose as the outcome
[d] When analysing blood pressure or CRP as the outcome
[e] When analysing cholesterol, TAG or CRP as the outcome
[f] When analysing glucose as the outcome
[g] In women
[h] Data only available in ATBC

5 Results

5.1 Whole grain intake estimation and associations with diet quality and chronic disease risk factors (I, II)

5.1.1 Participant characteristics in FinHealth 2017 (I, II)

In FinHealth 2017, 44% of the participants were men and the mean age was 56 years (Table 10). One fourth of the participants were physically inactive and fewer than one fifth were current smokers. The mean energy intake was 7.9 MJ/d in women and 9.7 MJ/d in men, and the mean whole grain intake 56 g/d (7.0 g/MJ) and 65 g/d (6.6 g/MJ), respectively.

Table 10 Participant characteristics and dietary intake (mean [SD] or %) in FinHealth 2017

	Women n=2844 mean (SD)/%	Men n=2250 mean (SD)/%	Total n=5094 mean (SD)/%
Age, years	56 (16)	56 (16)	56 (16)
BMI, kg/m^2	27 (5)	28 (4)	27 (5)
Low educational attainment[a], %	32	32	32
Physically inactive (leisure time), %	26	22	25
Current smokers, %	13	17	14
Antihypertensive medication use, %	27	28	27
Lipid-lowering medication use, %	15	20	18
Diabetes medication use, %	7	11	9
Energy under-reporters[b], %	36	41	38
Dietary intake			
Energy, MJ/d	7.9 (2.6)	9.7 (3.3)	8.7 (3.1)
Whole grain, g/d	56 (36)	65 (45)	60 (40)
Whole grain, g/MJ	7.0 (3.9)	6.6 (3.9)	6.9 (3.9)
Bread, g/d	82 (55)	91 (63)	86 (59)
Rye bread, g/d	53 (45)	62 (52)	57 (49)
Rye, oat and barley, g/d	57 (39)	64 (46)	60 (42)
Rye, g/d	41 (33)	46 (38)	43 (35)
Dietary fibre, g/d	22 (10)	22 (10)	22 (10)
Added sugar, g/d	36 (22)	48 (29)	41 (26)
Sodium, g/d	2.7 (1.0)	3.5 (1.3)	3.1 (1.2)

BMI, body mass index; SD, standard deviation

[a] Participants were categorised into tertiles by self-reported total years of education, adjusting for sex and birth cohort to account for the extension of the basic education system and increase in average school years over time.

[b] Under-reporting was defined as the ratio of energy intake to predicted basal metabolic rate ≤1.14 [196, 197].

5.1.2 Correspondence of surrogate measures with whole grain intake (I)

Whole grain intake was strongly correlated (energy-adjusted partial Spearman correlation) with the consumption of rye bread, rye, oat and barley combined, and rye alone, the correlations ranging from 0.84 to 0.99 (Table 11). The correlations with bread consumption and dietary fibre intake were below 0.70. The correlations remained similar after excluding energy under-reporters (data not shown). In the subgroup analyses, the correlations were generally stronger in men, older participants and participants with obesity, compared with women, younger participants and participants with normal weight or overweight, respectively. The correlations with rye bread, rye, oat and barley combined, and rye alone remained greater than 0.80 in all subgroups. The differences by sex, as well as by BMI for bread consumption, remained statistically significant after excluding energy under-reporters (data not shown).

Table 11 Energy-adjusted partial Spearman correlations (r_s) between whole grain intake and the potential surrogate measures in the overall population and by subgroups

		Whole grain				
		Bread	**Rye bread**	**Rye, oat and barley**	**Rye**	**Dietary fibre**
Overall	r_s	0.69	0.84	0.99	0.85	0.68
Subgroups						
Women	r_s	0.65	0.84	0.99	0.86	0.64
Men	r_s	0.73	0.83	0.99	0.85	0.75
	P[a]	<0.0001	0.22	<0.0001	0.28	<0.0001
Age <58[b]	r_s	0.67	0.81	0.99	0.83	0.65
Age ≥58[b]	r_s	0.65	0.83	0.99	0.84	0.66
	P[a]	0.25	0.03	0.10	0.04	0.42
BMI <30	r_s	0.67	0.83	0.99	0.84	0.68
BMI ≥30	r_s	0.74	0.86	0.99	0.88	0.71
	P[a]	<0.0001	<0.01	0.43	<0.001	0.05
Low/intermediate education	r_s	0.69	0.83	0.99	0.85	0.70
High education	r_s	0.67	0.84	0.99	0.86	0.66
	P[a]	0.09	0.28	0.20	0.33	0.01

[a] P-values for between-group differences in correlation coefficients were tested using Fisher Z transformation.
[b] Population median age

The proportion of participants categorised into the same quintile as based on whole grain intake was nearly 90% for rye, oat and barley combined, and approximately 50% for the other surrogate measures (Table 12). The proportion categorised into the same or adjacent quintile was 100% for rye, oat and barley combined, and greater than (rye bread and rye) or close to 90% (bread and dietary fibre) for the other measures. Gross misclassification was rare and observed only with bread (4%) and fibre (2%). However, these proportions increased after the exclusion of energy under-reporters (10% and 5%, respectively). In addition, after the exclusion, the proportion categorised into the lowest quintile of whole grain and the lowest quintile of bread or fibre decreased considerably. The results remained otherwise similar.

Overall, of the surrogate measures, rye-based variables, particularly rye, oat and barley combined, showed the best correspondence with whole grain intake. Bread and fibre demonstrated weaker correspondence, both in terms of correlation and cross-classification. Based on these results, the combined consumption of rye, oat and barley was used to estimate whole grain intake in the earlier study samples (sub-studies III and IV) where the whole grain database could not be utilised.

Table 12 The proportion (%) of participants categorised into the same, same or adjacent, or opposite quintile between whole grain intake and the potential surrogate measures in the overall population and after excluding energy under-reporters

Surrogate measure	Lowest quintile of whole grain intake		Same quintile, %	Same or adjacent quintile, %
	Lowest quintile, %	Highest quintile, %		
Bread				
Overall	59	4	49	88
Excl. energy under-reporters[a]	46	10	48	86
Rye bread				
Overall	59	0	48	93
Excl. energy under-reporters[a]	66	0	50	93
Rye, oat and barley				
Overall	92	0	88	100
Excl. energy under-reporters[a]	91	0	87	100
Rye				
Overall	66	0	52	96
Excl. energy under-reporters[a]	69	0	54	96
Dietary fibre				
Overall	65	2	47	88
Excl. energy under-reporters[a]	37	5	47	88

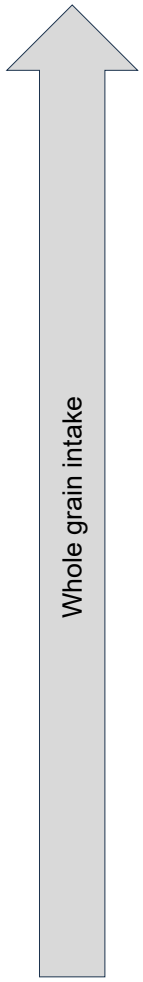
[a] Energy under-reporting was defined as the ratio of energy intake to predicted basal metabolic rate ≤1.14 [196, 197].

5.1.3 Associations between whole grain intake, diet quality and chronic disease risk factors (II)

Participants with higher whole grain intake were on average older (Women Q1: 46 vs. Q5: 59 years; Men Q1: 47 vs. Q5: 62 years; $P<0.001$) and less frequently inactive in leisure time (W: 32 vs. 26%; M: 30 vs. 21%; $P<0.01$) or current smokers (W: 20 vs 9%; M: 23 vs. 11%; P<0.01) compared with those with lower intakes (Table 13). Regarding dietary intake, participants with higher whole grain intake had lower mean energy intake in women (Q1 7.9 vs. Q5 7.3 MJ/d; P<0.001), scored higher mBSDS points in both sexes (P<0.001) and consumed more fruits and berries ($P<0.01$), low-fat milk ($P<0.01$) and unsaturated fatty acids in relation to saturated and trans fatty acids (fat ratio; $P<0.001$) in both sexes compared with those with lower intakes. They also consumed less vegetables ($P<0.05$), fish, red and processed meat and alcohol (all $P<0.001$), compared with participants with lower whole grain intakes. For example, participants in the highest whole grain intake quintile (Q5) consumed approximately 20% more fruits and berries (W: 79 vs. 96 g/d; M: 62 vs. 76 g/d) and 25–33% less red and processed meat (W: 92 vs. 69 g/d; M: 162 vs. 109 g/d), compared with those in the lowest quintile (Q1). The difference in mBSDS points between Q1 and Q5 was, however, only 0.9 in women (9.2 vs. 10.1) and 1.6 (8.9 vs. 10.5) in men. The results on sociodemographic, lifestyle and dietary factors remained similar after adjusting for energy under-reporting (data not shown).

Whole grain intake was statistically significantly associated with chronic disease risk factors only in men, although similar but non-significant trends were detected in women (Table 13). Men with higher whole grain intake had, on average, lower BMI (Q1 27.6 vs. Q5 26.7 kg/m^2; $P<0.001$) and WC (98.4 vs. 95.8 cm; P<0.001), compared with those with lower whole grain intake. These associations were attenuated by the adjustment for fibre intake (data not shown). Whole grain intake was also inversely associated with diastolic blood pressure (Q1 80.5 vs. Q5 79.2 mmHg; P<0.05) and HDL cholesterol (Q1 1.4 vs. Q5 1.3 mmol/L; P<0.05) in men, but these associations attenuated after adjusting for antihypertensive and lipid-lowering medication use (data not shown). In contrast, the adjustment for lipid-lowering medication use strengthened the inverse association with total cholesterol (Q1 5.2 vs. Q5 5.0 mmol/l; $P<0.05$). Similarly, adjusting for fibre intake strengthened the association with total cholesterol in each model (P<0.05; data not shown), although the difference between Q1 and Q5 of whole grain intake remained small (-0.2 mmol/l). The results remained the same after adjusting for energy under-reporting (data not shown). The associations between whole grain intake and the risk factors did not differ across mBSDS tertiles, as assessed by stratified analyses and interaction test ($P_{interaction}>0.05$; data not shown).

Table 13 General direction and significance of associations between whole grain intake (in quintiles) and sociodemographic and lifestyle factors, dietary factors and chronic disease risk factors in the main model. ↑ indicates direct association, ↓ indicates inverse association, and – indicates no statistically significant association.



	Women[a]	Men[a]
Sociodemographic and lifestyle factors		
Age	↑	↑
Education	–	↓
Physical activity	↑	↑
Smoking	↓	↓
Dietary factors		
Energy intake	↓	–
mBSDS	↑	↑
Vegetables	↓	↓
Fruits and berries	↑	↑
Fish	↓	↓
Low-fat milk	↑	↑
Fat ratio[b]	↑	↑
Red and processed meat	↓	↓
Alcohol (100%)	↓	↓
Chronic disease risk factors		
BMI	–	↓
Waist circumference	–	↓
Diastolic blood pressure	–	↓
Systolic blood pressure	–	–
Total cholesterol	–	–
HDL cholesterol	–	↓
LDL cholesterol	–	–
TAG	–	–
CRP	–	–
Glucose	–	–

BMI, body mass index; CRP, C-reactive protein; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol; mBSDS, modified Baltic Sea Diet Score; TAG, triacylglycerol

[a] Adjusted for age, energy intake, education, smoking, leisure time physical activity, BMI, mBSDS, sodium intake (blood pressure) and added sugar intake (TAG and glucose). A confounder was excluded if it was examined as the outcome.
[b] The ratio of unsaturated to saturated and trans fatty acid

5.2 Dietary changes, colorectal cancer-related lifestyles and colorectal cancer risk (III, IV part A)

5.2.1 Participant characteristics in the pooled study samples (III, IV part A)

The pooled study samples used in sub-studies III and IV part A comprised participants from ATBC, Health 2000, HBCS, DILGOM 2007 and FinRisk 2012, with a small variation in the number of participants owing to the different exclusion criteria. In both sub-studies, the median age ranged from 50 to 60 years across study samples, 20–40% of participants were physically inactive in leisure time and approximately 20% were current smokers (100% in ATBC) (Table 14). The median energy intake ranged from 8.7 to 10.8 MJ/d across study samples. In both sub-studies, the median follow-up time was 14 years. During the follow-up, 1124 CRC cases were diagnosed in Sub-study III and 1118 in Sub-study IV part A. The results of sub-studies III and IV part A are reported in the overall population owing to the limited number of CRC cases in women and no evidence of interaction by sex ($P_{interaction}>0.05$).

Table 14 Participant characteristics and dietary intake (median [IQR] or %) by study sample and sub-study

	ATBC[a]	Health 2000	HBCS	DILGOM 2007	FinRisk 2012
Number of men (%)					
Sub-study III	26 944 (100%)	2575 (45%)	875 (47%)	2175 (46%)	2081 (46%)
Sub-study IV part A	26 915 (100%)	2522 (45%)	870 (47%)	2167 (47%)	2066 (46%)
Age, years					
Sub-study III	57 (8)	50 (22)	60 (4)	53 (22)	53 (24)
Sub-study IV part A	57 (8)	50 (21)	60 (4)	53 (22)	53 (24)
BMI, kg/m^2					
Sub-study III	26 (5)	26 (6)	27 (6)	26 (6)	26 (6)
Sub-study IV part A	26 (5)	26 (6)	27 (5)	26 (6)	26 (6)
WC[b,c], cm					
Sub-study IV part A					
Women	-	87 (18)	90 (17)	84 (17)	86 (19)
Men	-	97 (15)	100 (14)	96 (15)	97 (17)
Height[b,c], cm					
Sub-study IV part A					
Women	-	163 (9)	163 (8)	163 (9)	163 (9)
Men	174 (9)	176 (10)	177 (8)	176 (9)	177 (9)
Low educational attainment[d], %					
Sub-study III	78	33	34	30	33
Sub-study IV part A	78	32	33	30	33
Physically inactive (leisure time), %					
Sub-study III	42	28	31	19	20
Sub-study IV part A	42	27	31	19	20
Current smoker, %					
Sub-studies III and IV part A	100	26	24	18	17
Prevalent diabetes[b], %					
Sub-study IV part A	6	4	2	9	12
HRT use (women), ever %					
Sub-study III	-	31	68	16	14
Sub-study IV part A	-	31	69	16	14

Table continues

Table 14 continues

	ATBC[a]	Health 2000	HBCS	DILGOM 2007	FinRisk 2012
CRC lifestyle index[b], points					
Sub-study IV part A	1.9	1.9	1.8	2.1	2.0
Dietary factors					
Energy, MJ/d					
Sub-study III	10.8 (4.0)	9.1 (3.9)	8.7 (4.0)	9.9 (4.5)	8.9 (4.1)
Sub-study IV part A	10.8 (4.0)	9.1 (3.9)	8.7 (4.1)	9.9 (4.5)	8.9 (4.1)
Vegetables[f], g/week					
Sub-study III	658 (588)	1526 (1295)	1722 (1449)	1848 (1533)	1589 (1323)
Fruits[f], g/week					
Sub-study III	756 (819)	1099 (1435)	1512 (1897)	1491 (1687)	1085 (1274)
Whole grains[e]					
Sub-study III, g/week	700 (595)	406 (406)	378 (357)	532 (441)	483 (455)
Sub-study IV part A, g/d	100 (85)	58 (58)	54 (51)	76 (63)	69 (65)
Dairy products, g/d					
Sub-study III	699 (502)	546 (466)	436 (435)	584 (509)	579 (516)
Sub-study IV part A	699 (502)	543 (466)	436 (434)	583 (507)	579 (515)
Red meat, g/week					
Sub-study III	455 (273)	511 (343)	420 (350)	511 (406)	462 (357)
Sub-study IV part A	456 (274)	514 (346)	419 (346)	513 (409)	462 (354)
Processed meat, g/week					
Sub-study III	420 (406)	252 (336)	210 (280)	280 (350)	266 (343)
Sub-study IV part A	421 (403)	253 (338)	211 (279)	281 (352)	268 (341)
Alcohol (100%), g/d					
Sub-studies III and IV part A	11 (23)	2 (7)	5 (11)	4 (9)	4 (9)

BMI, body mass index; CRC, colorectal cancer; HRT, hormone-replacement therapy; IQR, interquartile range; WC, waist circumference

[a] All participants in ATBC were male current smokers as per the study design.

[b] The variable was only used in Sub-study IV part A.

[c] Waist circumference and height are reported by sex due to their sex-specific scoring in the CRC lifestyle index.

[d] In ATBC, low educational attainment corresponded to elementary or lower education. In the other study samples, participants were categorised into those with low, intermediate or high educational attainment according to cohort-specific tertiles of self-reported school years, adjusting for sex and birth year.

[e] Whole grain intake was estimated as combined consumption of rye, oat and barley based on the results of Sub-study I.

[f] The variable was only used in Sub-study III.

5.2.2 Associations between partial substitutions of red or processed meat with plant-based foods and CRC risk (III)

In the preliminary analyses, red meat consumption (per 100 g/week) was associated with a borderline 3% increase (HR 1.03, 95% CI 1.00–1.06) and fruit consumption (per 100 g/week) with a borderline 1% decrease in CRC risk (0.99, 0.98–1.00). No statistically significant associations were observed between the consumption of processed meat (per 50 g/week), whole grains (per 100 g/week) or vegetables (per 100 g/week) and CRC risk (data not shown).

In the substitution analyses, partial substitutions of red meat with vegetables, fruits or a combination of plant-based foods were associated with statistically significant 3% reductions in CRC risk ($P<0.05$; Figure 6). Similarly, partial substitutions of processed meat with vegetables or fruits were associated with 1% reductions in CRC risk ($P<0.05$). Excluding low consumers (<50 g/week of processed meat or <100 g/week of red meat) attenuated the associations between the processed meat substitutions and CRC risk, while the associations of the red meat substitutions remained unchanged (data not shown). The results remained the same also after excluding participants diagnosed with CRC within the first two years of follow-up, and after adjusting for CRC family history (in ATBC; data not shown).

Although no significant heterogeneity was observed between the study samples ($P_{heterogeneity}>0.05$; data not shown), the analyses were repeated excluding ATBC from the pooled data to account for its distinct characteristics; in addition to being male current smokers, the participants in ATBC had a considerably higher median consumption of whole grains and processed meat, and a considerably lower median consumption of vegetables, than participants in the other study samples (Table 14). Excluding ATBC attenuated the inverse associations between the processed meat substitutions and CRC risk (data not shown). Conversely, the inverse associations between the red meat substitutions and CRC risk were slightly strengthened (vegetables: 0.96, 0.93–1.00; fruits: 0.95, 0.92–0.98; plant-based foods combined: 0.97, 0.93–0.99). Furthermore, after excluding ATBC, a statistically significant 7% reduction in CRC risk was observed when red meat was partially substituted with whole grains (0.93, 0.87–0.99). This result suggests that the non-significant association observed in the overall population for the substitution with whole grains was linked to the high whole grain consumption in ATBC (median 700 g/week [100 g/d]). To test this hypothesis, the substitutions with whole grains were repeated, stratified by the median whole grain consumption in the overall population (<, ≥587 g/week [84 g/d]). Following this, significant 8% and 4% reductions in CRC risk were observed when red meat (0.92, 0.86–0.98) or processed meat (0.96, 0.93–0.99) was partially substituted with whole grains in participants with below median whole grain consumption (Figure 7). No significant associations were

observed in participants with greater than median whole grain consumption ($P_{interaction}$=0.001).

All substitutions were also modelled in subgroups by sex, age, BMI, HRT use and follow-up time, but no notable between-group differences were observed (data not shown).

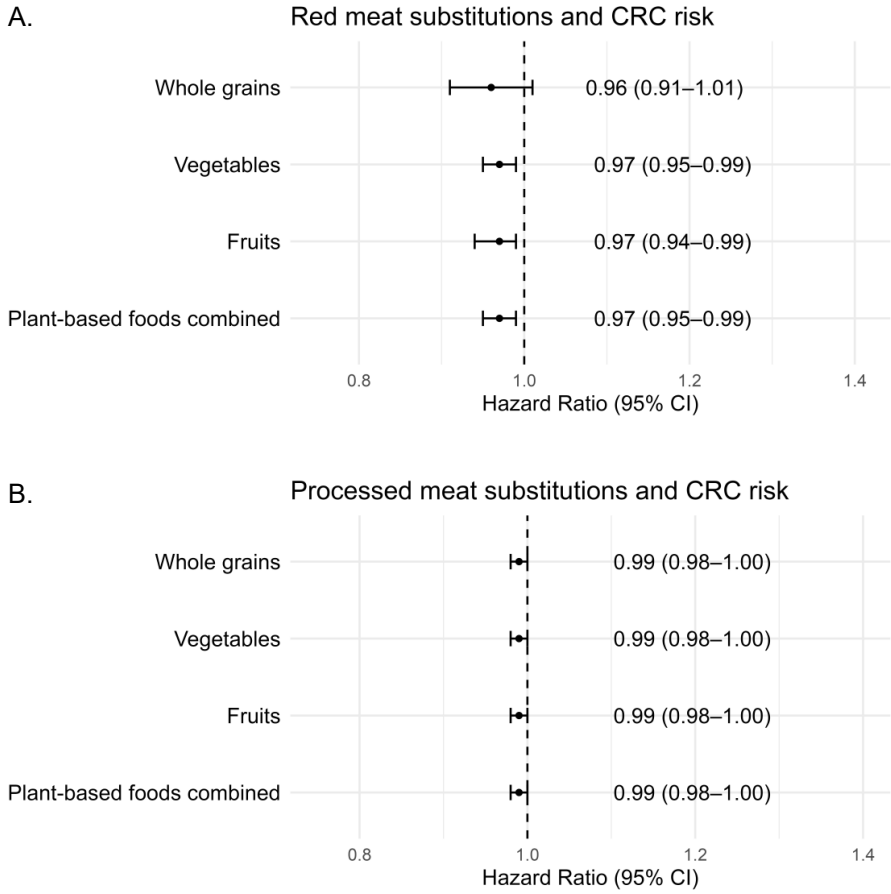


Figure 6 Associations (hazard ratios and 95% confidence intervals [CI]) between partial substitutions of A) red meat (100 g/week) and B) processed meat (50 g/week) with plant-based foods (100 or 50 g/week) and colorectal cancer (CRC) risk in the pooled data of five study samples

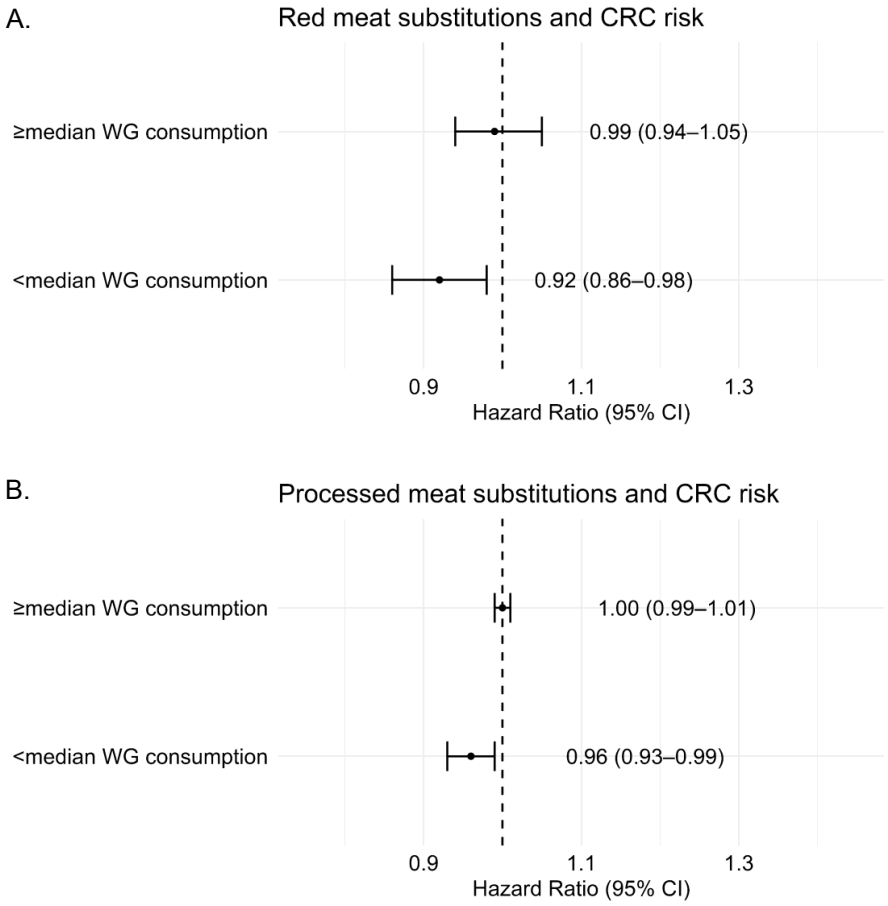


Figure 7 Associations (hazard ratios and 95% confidence intervals [CI]) between partial substitutions of A) red meat (100 g/week) and B) processed meat (50 g/week) with whole grains (WG; 100 or 50 g/week) and colorectal cancer (CRC) risk in the pooled data of five study samples, stratified by the median whole grain consumption (587 g/week).

5.2.3 Association between the CRC lifestyle index and CRC risk (IV part A)

The median CRC lifestyle index points ranged from 1.8 to 2.1 across the pooled study samples (Table 14). Participants who scored higher index points had a higher median energy intake (except in ATBC) and were less frequently current smokers (except for ATBC and HBCS) compared with those who scored lower index points (data not shown).

Participants in the highest quintile of the CRC lifestyle index (Q5; median points 2.6–3.0 across pooled samples; lower-risk lifestyles) had a 31% (HR 0.69, 95% CI 0.58–0.83) lower risk of CRC compared with participants in the lowest quintile (Q1; median points 0.9–1.1 across pooled samples). Moreover, each one-point increase in the index was associated with a 19% reduction in CRC risk (0.81, 0.74–0.88). A one-point increase corresponds to, for example, the difference between physically active vs. inactive participants or participants with both BMI and WC in the highest vs. lowest category. No notable heterogeneity was observed between the study samples ($P_{heterogeneity}$>0.05). The results remained the same after excluding participants diagnosed with CRC within the first two years of follow-up, and after adjusting for prevalent diabetes and HRT use in women (data not shown). Based on these results, the CRC lifestyle index was considered an appropriate tool to distinguish between higher- and lower-risk lifestyles for CRC in Finnish adults.

5.3 Colorectal cancer-related lifestyles and the gut microbiome (IV part B)

5.3.1 Participant characteristics in FinRisk/FinDiet 2002

In FinRisk/FinDiet 2002, 46% of the participants were men and the median age was 48 years (Table 15). The median CRC lifestyle index score was 2.0 points, ranging from 1.1 in Q1 to 3.0 in Q5. One participant scored zero points, while no participant scored full four points. Participants with higher index points (lower-risk lifestyles for CRC) tended to have a higher energy intake and be less frequently current smokers or users of potentially microbiome-altering medication compared with those with lower points. No participant in the final study sample had prevalent CRC at baseline, and only one participant was diagnosed with CRC during the first five years of follow-up. The remaining CRC cases (n=11) were diagnosed more than eight years after the baseline.

Table 15 Participant characteristics, the CRC lifestyle index and its components (medians [IQR] or %) in FinRisk/FinDiet 2002

		CRC lifestyle index quintiles		
	Overall n=1228	**Q1 n=252**	**Q3 n=248**	**Q5 n=209**
Men, %	46	48	42	47
Age, years	48 (18)	49 (17)	47 (18)	47 (21)
Energy intake, MJ/d	7.6 (3.2)	7.3 (2.8)	7.8 (3.6)	7.8 (3.1)
Current smokers, %	24	26	27	17
Medication use[a], %	8.6	11.5	11.3	4.3
CRC lifestyle index and its components				
CRC lifestyle index[b], points	2.0 (0.9)	1.1 (0.4)	2.1 (0.3)	3.0 (0.4)
Body fatness score, points	0.5	0	0.8	1.0
Body mass index, kg/m^2	26 (6)	30 (6)	26 (5)	24 (3)
Waist circumference, cm				
women	82 (16)	96 (16)	80 (14)	75 (8)
men	95 (15)	107 (15)	93 (14)	87 (9)
Height, cm				
women	163 (8)	167 (5)	162 (9)	159 (5)
men	176 (9)	181 (6)	177 (10)	172 (5)
Physically inactive (leisure time), %	19	41	14	0
Dietary score, points	0.4	0.3	0.5	0.5
Whole grains, g/d	49 (58)	35 (48)	52 (56)	67 (65)
Dairy products, g/d	375 (371)	307 (307)	387 (342)	416 (405)
Red meat, g/week	232 (500)	282 (605)	242 (413)	146 (372)
Processed meat, g/week	175 (403)	196 (414)	140 (421)	169 (420)
Alcohol (100%), g/d	6.3 (15)	9 (22)	6 (13)	5 (10)

CRC, colorectal cancer; IQR, interquartile range; Q, quintile

[a] Use of metformin, psycholeptics, psychoanaleptics, proton pump inhibitors or constipation medication

[b] Higher points indicate a lower-risk lifestyle for CRC

Table adapted from the preprint of Sub-study IV.

5.3.2 Associations between the CRC lifestyle index and gut microbiome diversity and composition

Higher points in the CRC lifestyle index (lower-risk lifestyles) were statistically significantly associated with higher microbial alpha diversity (β 0.04, 95% CI 0.01–0.06; Figure 8). Of the individual index components, physical activity was directly (0.09, 0.01–0.17), and BMI (-0.04, -0.07– -0.02) and WC (-0.05, -0.08– -0.02) inversely associated with alpha diversity. The inverse associations of BMI and WC were also reflected in the direct association of the body fatness score (opposite result due to reverse scoring). No significant associations were observed between the other index components and alpha diversity.

The CRC lifestyle index was also associated with compositional differences in the gut microbiome (beta diversity), explaining a small but statistically significant proportion of its variance (R^2 0.002, P=0.010; Figure 9). All index components, except for physical activity and consumption of red meat and dairy products, were associated with beta diversity (P<0.05). The explained proportions were largest for BMI, WC and alcohol consumption (0.5–0.6%).

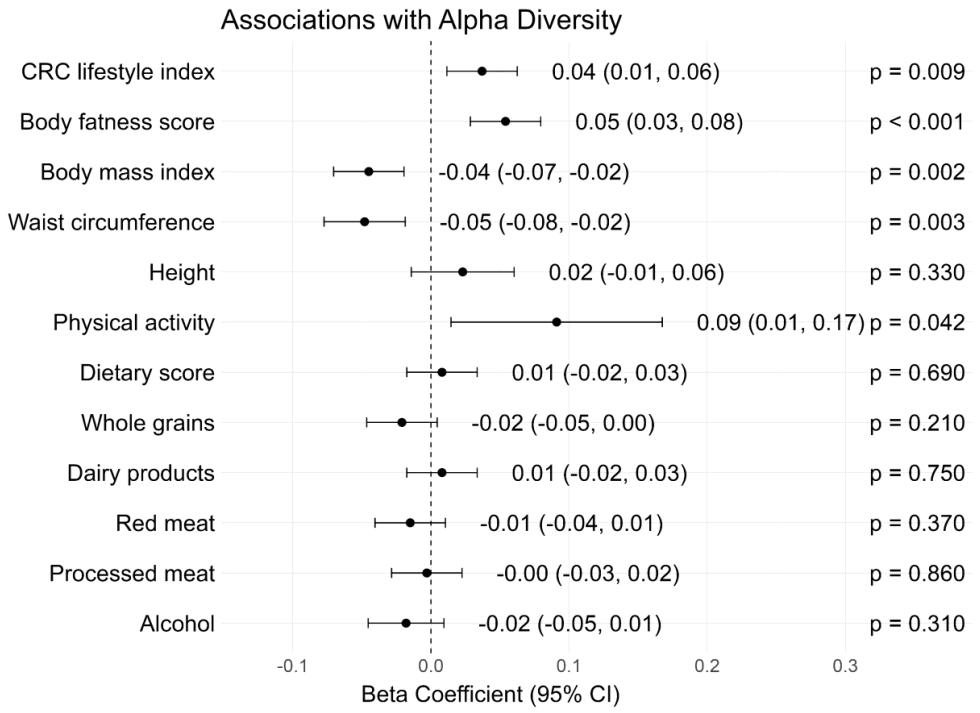


Figure 8 Associations of the CRC lifestyle index and its components with alpha diversity (β coefficients and 95% confidence intervals [CI]). Body fatness score was calculated by averaging the points for body mass index and waist circumference. Dietary score was calculated by averaging the points for the five dietary factors. β coefficient and 95% CI for physical activity were calculated by comparing inactive and active participants.

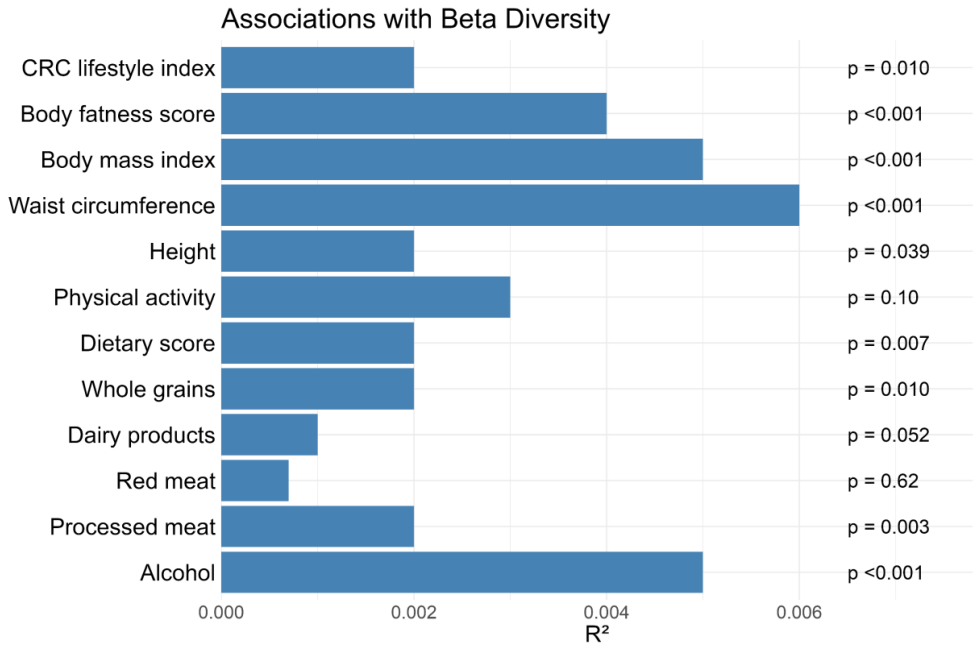


Figure 9 Associations of the CRC lifestyle index and its components with beta diversity (proportion of variance explained, R^2). Body fatness score was calculated by averaging the points for body mass index and waist circumference. Dietary score was calculated by averaging the points for the five dietary factors. R^2 for physical activity was calculated by comparing inactive and active participants.

In the analyses examining differential relative abundance of individual microbial species, the CRC lifestyle index was significantly associated with 103 species. Examples of these associations are displayed in Table 16. Most associations were inverse, indicating a higher relative abundance of microbial species in participants scoring lower index points (higher-risk lifestyles). The inversely associated species represented genera such as *Megasphaera* (two species; family Megasphaeraceae), *Dorea* (two species; Lachnospiraceae), *Mediterraneibacter* (4 species; Lachnospiraceae), *Collinsella* (4 species; Coriobacteriaceae) and *Hungatella* (one species; Lachnospiraceae). Direct associations were observed between the index and species representing genera such as *Bifidobacterium* (three species; Bifidobacteriaceae) and *Haemophilus* (three species; Pasteurellaceae).

In the cluster analysis, the species significantly associated with the CRC index were divided into 13 clusters (Table 16). Of these, five were directly and five inversely associated with the index ($P<0.05$), whereas three clusters had a non-significant association. In general, the directions of the associations were consistent with those of the individual species included in the clusters. In terms of the included species, the clusters were highly diverse, although the inversely associated clusters consisted primarily of species within the family Lachnospiraceae.

Table 16 Examples of associations between the CRC lifestyle index and relative abundances of microbial species and their corresponding clusters (effect size and standard error [SE])

	CRC lifestyle index		
Order/Family/Genus Species	***Effect size*[a,b]**	***SE***	***P*[c]**
Cluster 1 (3 species)	0.098	0.056	0.081
*Bifidobacteriales/Bifidobacteriaceae/**Bifidobacterium longum***	0.141	0.00007	<0.0001
*Bifidobacteriales/Bifidobacteriaceae/**Bifidobacterium saguini***	0.070	0.00005	<0.0001
*Bifidobacteriales/Bifidobacteriaceae/**Bifidobacterium breve***	0.084	0.00006	<0.0001
Cluster 2 (11 species), e.g.,	-0.173	0.056	0.002
*Selenomonadales/Selenomonadaceae/**Selenomonas sputigena***	-0.153	0.00002	<0.0001
*Veillonellales/Megasphaeraceae/**Megasphaera_A_38685 cerevisiae***	-0.307	0.00004	<0.0001
Cluster 3 (18 species), e.g.,	-0.091	0.037	0.015
*Lachnospirales/Lachnospiraceae/**Dorea_A longicatena***	-0.100	0.00004	<0.0001
*Lachnospirales/Lachnospiraceae/**Dorea_A formicigenerans***	-0.080	0.00003	<0.0001
*Lachnospirales/Lachnospiraceae/**Lachnoclostridium_B sp900066555***	-0.079	0.00003	<0.0001
*Lachnospirales/Lachnospiraceae/**Eubacterium_I ramulus***	-0.092	0.00004	<0.0001
*Lachnospirales/Lachnospiraceae/**Mediterraneibacter_A_155507 faecis***	-0.085	0.00004	<0.0001
*Lachnospirales/Lachnospiraceae/**Mediterraneibacter_A_155507 massiliensis***	-0.080	0.00003	<0.0001
*Lachnospirales/Lachnospiraceae/**Mediterraneibacter_A_155590 butyricigenes***	-0.072	0.00003	<0.0001
Cluster 4 (6 species), e.g.,	-0.052	0.044	0.235
*Veillonellales/Megasphaeraceae/**Megasphaera_A 38692 hutchinsoni***	-0.073	0.00003	<0.0001
Peptostreptococcales/Peptostreptococcaceae/ ***Peptostreptococcus anaerobius***	-0.066	0.00003	<0.0001
Cluster 5 (6 species), e.g.,	-0.111	0.039	0.004
*Lachnospirales/Lachnospiraceae/**Mediterraneibacter_A 155507 torques***	-0.263	0.00005	<0.0001
Cluster 6 (8 species), e.g.,	0.074	0.036	0.039
*Erysipelotrichales/Erysipelotrichaceae/**Holdemanella biformis***	-0.111	0.00005	<0.0001
*Lachnospirales/Lachnospiraceae/**Eubacterium_F sp003491505***	0.207	0.00006	<0.0001

Table continues

Table 16 continues

Order/Family/Genus Species	CRC lifestyle index		
	***Effect size*[a,b]**	***SE***	***P*[c]**
Cluster 7 (6 species), e.g.,	-0.155	0.070	0.028
*Coriobacteriales/Coriobacteriaceae/**Collinsella intestinalis***	-0.142	0.00005	<0.0001
*Coriobacteriales/Coriobacteriaceae/**Collinsella stercoris***	-0.124	0.00005	<0.0001
*Coriobacteriales/Coriobacteriaceae/**Collinsella ihuae***	-0.076	0.00004	<0.0001
*Coriobacteriales/Coriobacteriaceae/**Collinsella phocaeensis***	-0.067	0.00004	<0.0001
Cluster 8 (9 species), e.g.,	-0.095	0.045	0.034
*Bacteroidales/Bacteroidaceae/**Prevotella salivae***	-0.089	0.00005	<0.0001
Cluster 9 (10 species), e.g.,	-0.086	0.045	0.053
Lachnospirales/Lachnospiraceae/ ***Hungatella_A_127239 hathewayi_A***	-0.107	0.00004	<0.0001
*Actinomycetales/Actinomycetaceae/**Actinomyces graevenitzii***	-0.058	0.00004	<0.0001
*Anaerovoracaceae/**Eubacterium_T pyruvativorans***	-0.060	0.00004	<0.0001
Cluster 10 (3 species)	0.314	0.096	0.001
*Enterobacterales/Pasteurellaceae/**Haemophilus_A sputorum***	0.139	0.00005	<0.0001
Enterobacterales/Pasteurellaceae/ ***Haemophilus_D_735815 parainfluenzae K 735050***	0.157	0.00007	<0.0001
Enterobacterales/Pasteurellaceae/ ***Haemophilus_D_735815 parainfluenzae_K 735055***	0.233	0.00007	<0.0001
Cluster 11 (6 species), e.g.,	0.338	0.119	0.004
*Christensenellales/CAG-138/**Phil1 sp002069725***	0.083	0.00003	<0.0001
Cluster 12 (5 species), e.g.,	0.284	0.078	0.0003
*Bacteroidales/UBA932/**Cryptobacteroides sp002438635***	0.216	0.00004	<0.0001
Cluster 13 (12 species), e.g.,	0.175	0.058	0.003
*RF39/UBA660/**CAG-533 sp000434495***	0.186	0.00006	<0.0001

[a] The effect size represents log-fold change (natural logarithm of the abundance ratio) for the species and β coefficient for the clusters per one unit change in the CRC lifestyle index. Associations between the CRC lifestyle index and microbial species were assessed using ANCOM-BC2. Species that were significantly associated with the index after Benjamini-Hochberg FDR correction [203], and passed the sensitivity screening for robustness, were included in the cluster analysis.

[c] The analyses were adjusted for sex, age, energy intake (MJ/d), smoking habits (smoker/nonsmoker) and use of potentially microbiome-altering medication (yes/no; metformin, psycholeptics, psychoanaleptics, proton pump inhibitors and constipation medication).

6 Discussion

Cereal grains are staple foods in diets worldwide. When consumed as whole grain, they provide a wide variety of nutrients that benefit human health in various ways [19]. Consequently, high whole grain intake has been linked to a lower risk of major chronic diseases and beneficial changes in their underlying risk factors [11, 12]. Nevertheless, inconsistencies in findings remain regarding several health outcomes, likely arising from methodological discrepancies [11, 19, 75]. To elucidate discrepancies related to whole grain intake estimation, I assessed the validity of five potential whole grain surrogate measures and established high correspondence between rye-based estimates and whole grain intake in Finnish adults (I). In addition, I explored associations between whole grain intake and chronic disease risk factors, as well as the largely overlooked question of whether diet quality modifies these associations (II). Whole grain intake was significantly associated with the risk factors only in men. Although higher whole grain intake was associated with better diet quality, diet quality did not modify its associations with the risk factors.

Whole grain intake has been consistently associated with a lower CRC risk, whereas strong evidence indicates that consumption of red and processed meat increases the risk [5, 11]. Although individual dietary factors have been studied extensively in relation to CRC risk, few studies have modelled dietary changes in this regard. The pooled analysis of five Finnish study samples within this thesis suggest that already small weekly substitutions of either red or processed meat with plant-based foods, particularly whole grains, could reduce CRC risk (III). To elucidate the mechanisms underlying these associations from a whole-lifestyle perspective, I explored the gut microbiome diversity and composition in relation to the CRC lifestyle index (IV). Lower index points, indicating higher-risk lifestyles for CRC, were associated with lower microbial diversity and higher relative abundance of several bacterial species previously linked to CRC.

6.1 Correspondence of surrogate measures with whole grain intake (I)

The discrepancy in whole grain intake estimation methods remains a major challenge in whole grain research, potentially attenuating observed associations with health outcomes [19, 20, 75]. In Sub-study I, I aimed to establish the correspondence between five previously used or potential whole grain surrogate measures and whole grain intake to facilitate more standardised intake estimation when measured whole grain data are unavailable.

The results demonstrated strong correlation between whole grain intake and rye-based variables, including rye bread, rye, and rye, oat and barley combined. Rye is the principal cereal source (>50%) and rye bread the principal food source (≈50%) of whole grains in the diets of Finnish adults, largely explaining the strong correlations [55]. Rye is also primarily consumed as whole grain rather than refined. Correspondingly, oat is the second-most important whole grain source, contributing to more than 20% of whole grain intake, and both oat and barley are predominantly consumed as whole grain in Finland. Consequently, the strongest correlation (r_s=0.99) was observed with the combined consumption of rye, oat and barley in the overall population and in different sub-groups. The magnitude of the correlation suggests that rye, oat and barley together captured almost all variation in whole grain intake in our study population. Additionally, in cross-classification, almost 90% of participants were classified into the same quintile and 100% into the same or adjacent quintile between whole grain intake and rye, oat and barley combined. For the other surrogate measures, these proportions were approximately 50% for the same quintile and between 87% and 96% for the same or adjacent quintile. Thus, the combined consumption of rye, oat and barley appeared robust and the most reliable surrogate measure for whole grain intake in the Finnish context.

Compared with the rye-based variables, bread consumption and dietary fibre intake showed weaker correlations with whole grain intake, with some sub-group differences. Additionally, in cross-classification, the exclusion of energy under-reporters affected more their correspondence to whole grain intake; after the exclusion, fewer participants were similarly categorised into the lowest intake quintile of both whole grain intake and bread consumption or fibre intake, and gross misclassification became more prevalent. These results may reflect disproportionate under-reporting of refined grain products, particularly refined wheat bread. Wheat is the most consumed cereal grain in Finnish adults, but it is predominantly consumed as refined [55, 175]. Furthermore, wheat bread and mixed-grain breads, which are often wheat-based, account for a substantial proportion of bread consumption [175], likely explaining the differing results between rye bread and total bread. In addition, as cereals are the primary source of

dietary fibre in Finnish adults [175], the contribution of wheat, and thus refined grains, to fibre intake is presumably significant. Consequently, disproportionate underreporting of refined grain foods would reflect into the results of both bread consumption and fibre intake.

Given the high prevalence of energy under-reporting across study populations and dietary assessment methods [206], sensitivity to it should be considered in selecting an appropriate surrogate measure for whole grain intake. This may be particularly relevant in populations where greater proportion of whole grain intake is derived from foods that are prone to under-reporting, such as refined grain products and snack foods. In a New Zealand study of forty adults, dietary data gathered using wearable cameras and three 24h recalls demonstrated under-reporting of particularly snack foods, such as biscuits and muesli bars [207]. The consumption of breads, rice and cereals was also under-reported somewhat frequently. In studies of the British (n=1571, ≥18 years), Australian (n=9341, ≥19years) and Swedish (n=1797, 18–80 years) populations, discretionary and other foods containing low amounts of whole grains appeared to contribute significantly to whole grain intake [208–210].

Overall, identifying a surrogate measure with as high correspondence to whole grain intake as the combined consumption of rye, oat and barley in the Finnish population may not be feasible elsewhere. In Finnish adults, rye and oat together contribute to approximately 80% of whole grain intake, and they are both predominantly consumed as whole grain [55]. High consumption of rye bread and oatmeal is a distinctive feature of traditional Finnish diets, and they have maintained their central role in diets to this day. Similar dietary traits exist in some other Nordic countries [56], whereas studies in other Western countries have reported more heterogenous whole grain sources [58–62, 208]. Therefore, finding a single food that is consumed by most of the population and covers most of the whole grain intake may be challenging. Moreover, although wheat is the predominant whole grain source in most Western countries, its frequent consumption in the refined form limits its utility as a proxy for whole grain intake.

Despite being influenced by energy under-reporting, dietary fibre was a reasonably good indicator of whole grain intake in our study population, owing to the predominance of cereals as a fibre source [175]. However, such correspondence may not be observed in populations where larger proportion of fibre is derived from refined grains or non-cereal sources. For example, in a US study of 34 000 individuals aged ≥2 years (NHANES 2003–2010), only 15% of fibre intake originated from foods with some whole grain ingredients, whereas almost 40% of fibre came from foods with no whole grains at all [211].

In conclusion, this Sub-study assessed the validity of five potential surrogate measures for estimating whole grain intake in the absence of measured whole grain data. The results suggest that the combined consumption of rye, oat and barley

provides a robust estimate of whole grain intake in Finnish adults. Employing standardised intake estimation methods, whether based on direct or surrogate measures, would improve comparability across studies and strengthen the evaluation of associations between whole grain intake and health outcomes.

6.2 Associations between whole grain intake, diet quality and chronic disease risk factors (II)

Although whole grain intake has been consistently linked with a lower risk of several chronic diseases, the evidence of its associations with chronic disease risk factors has remained inconclusive [11, 12]. In this Sub-study, I examined associations of whole grain intake with diet quality and chronic disease risk factors in Finnish adults, and established the role of diet quality in modifying these associations.

Whole grain intake was associated with better diet quality assessed by the mBSDS, although the difference in mBSDS points between the first and fifth whole grain intake quintile was relatively small (0.9 points in women and 1.6 points in men). Previous studies in the US (n=13 276, ≥19 years) [212] and Ireland (n=1500, 18–90 years) [58] have reported similar results when diet quality was assessed with the Healthy Eating Index (HEI) and the alternative Mediterranean Diet Index (aMED), respectively. For example, the difference in the average aMED points between non-consumers and the highest whole grain intake category (tertile 3) was 1.9 points, with the possible score ranging between zero and nine points [58]. However, in both studies, the index used included whole grains as a component, and this built-in overlap likely inflated the observed associations with whole grains. Consequently, further research is required to confirm the direct association between whole grain intake and diet quality observed in the present study.

Of the mBSDS components, higher whole grain intake was associated with higher consumption of fruits and berries and low-fat milk, as well as a better fat ratio. Conversely, whole grain intake was associated with lower consumption of red and processed meat, alcohol, vegetables and fish. Similar associations with fruits, low-fat milk, red and processed meat and alcohol have been previously reported in Scandinavian (n=8702, 30–65 years) [14], British (n=1521, ≥18 years) [57], Australian (n=7665, ≥18 years) [15] and US (Nurses' Health Study [NHS] II, n=470, 25–42 years & Health Professionals Follow-up Study [HPFS], n=468, 40–75 years; Baltimore Longitudinal Study on Aging [BLSA], n=1516, 27–88 years) [17, 72] study populations. However, in contrast to our findings, these studies observed direct associations between whole grain intake and vegetable and fish consumption. Although the analyses were adjusted for educational attainment, the inverse associations with vegetables and fish in our study could arise from confounding of other socioeconomic variables, such as income. In a previous study based on a sub-sample of our study population (n=1655), participants with a lower income level

were found to consume more cereals and less vegetables and fruits (one variable) and fish compared to participants with a higher income level [213]. The association with cereal consumption likely reflects a similar trend in the relationship between income level and whole grain intake, because rye and oat, which are predominantly consumed as whole grain, are major contributors to cereal consumption in Finnish adults [175]. The contrasting result between fruits and vegetables may be related to the inclusion of berries in the fruit variable, as forest berries are commonly eaten together with oatmeal (major whole grain food source) in Finland [55].

Regarding chronic disease risk factors, statistically significant associations were observed between whole grain intake and BMI, WC and total cholesterol (after adjusting for medication use) in men, while the trends were similar but non-significant in women. The inverse associations with BMI and WC are in line with previous studies in the Scandinavian countries (n=8702, 30–65 years) [14], Italy (n=2830, ≥18 years) [60], the US (NHANES 2001–2012, n=29 683, ≥19 years; NHS II, n=470, 25–42 years & HPFS, n=468, 40–75 years; BLSA, n=1516, 27–88 years) [17, 62, 72] and Australia (n=7665, ≥18 years) [15]. Of these, the studies in Scandinavia [14] and the US (NHS II & HPFS) [17] examined women and men separately, and both found a significant association with BMI (WC not examined) only in men. In our study, the sex difference may be linked to variation in dietary fibre sources, as fibre appeared to modify the associations, evidenced by attenuation of the results following fibre adjustment. A comprehensive meta-analysis of clinical trials has previously concluded that increased intake of cereal fibre is particularly effective in reducing body weight, while the evidence of other fibre sources is limited [11]. In a previous study using a sub-sample of our study population (n=1655), cereals (and thus presumably whole grains) were found to contribute more to fibre intake in men than in women, while the contribution of vegetables and fruits was higher in women [175]. Thus, greater reliance on cereal fibre among men could partly explain our findings, though caution is warranted in extrapolating trial results to observational studies. However, since similar sex differences have been observed in other study populations, possible contribution of underlying metabolic differences between women and men should be considered.

Besides BMI and WC, after adjusting for medication use, whole grain intake was significantly (inversely) associated only with total cholesterol. The association between whole grain intake and total cholesterol has been previously examined in a cross-sectional setting in two US studies (NHS II, n=470, 25–42 years & HPFS, n=468, 40–75 years; Baltimore Longitudinal Study on Aging, n=1516, 27–88 years) reporting inverse associations [17, 72] and one British (n=2689, aged ≥18 years) [16] and one Australian (n=7665, ≥18 years) [15] study reporting non-significant associations. Of these, the study in 1516 US adults reported an indirect association with total cholesterol after adjusting for the use of lipid-lowering medication [72]. No study examined the association separately in women and men.

Systematic reviews of clinical trials have so far attributed the link between whole grain intake and cholesterol largely on whole grain oat [214, 215]. This is congruent with the consistent evidence showing that β-glucan, the main soluble fibre in oat, reduces serum total and LDL cholesterol [216]. Given that the contribution of oat to whole grain intake is greater in Finnish women than men, the opposite holding true for rye [55], an association with cholesterol would have been expected particularly in women in our study population. However, in addition to oat, also barley and, to a lesser extent, rye, contain β-glucan [217]. Moreover, rye contains another soluble fibre, arabinoxylan, which may have similar beneficial features regarding cholesterol as β-glucan, although this has yet to be confirmed. Nonetheless, since the association between whole grain intake and total cholesterol remained after adjusting for fibre intake, the association may be attributable to other factors. For example, phytosterols found in various plant foods, including whole grains, have been shown to reduce LDL cholesterol [218].

The association between whole grain intake and total cholesterol may also be related to HDL cholesterol, which was likewise inversely associated with whole grain intake. However, the difference in mean HDL cholesterol concentrations between the extreme quintiles of whole grain intake was very small (0.04 mmol/L), and the association was attenuated after adjustment for lipid-lowering medication use. In contrast, the association between whole grain intake and total cholesterol emerged only after adjusting for medication use, suggesting that it may not be attributable to lower HDL cholesterol concentrations among participants with higher whole grain intake.

Previous studies have suggested that the associations of whole grain intake with chronic diseases and their risk factors may reflect the generally healthier diets and lifestyles of whole grain consumers, rather than the effects of whole grains themselves [13, 75, 219]. In this study, participants with higher whole grain intake were less often current smokers, more often physically active and had better diet quality compared to those with lower whole grain intake. Consequently, the analyses were adjusted for smoking and physical activity, and the modifying role of diet quality was explored by first adjusting for the mBSDS, and second, stratifying the analyses by the mBSDS tertiles. However, no indication was observed that diet quality would have modified the associations between whole grain intake and chronic disease risk factors. Nonetheless, it should be considered that, although statistically significant, the variation in mBSDS points between participants with higher and lower whole grain intakes was rather small, which may explain why no differences were observed in the stratification.

To summarise, the evidence of associations between whole grain intake and chronic disease risk factors remains largely inconsistent. Although the observed inverse associations with obesity measures are supported by most previous observational studies, uncertainty remains of the sex-specific nature of these

associations. While no differences were observed in the associations with the risk factors when stratified by diet quality, the role of diet quality and other lifestyle factors should be further explored in other study populations and using different dietary indices.

6.3 Associations between partial substitutions of red or processed meat with plant-based foods and CRC risk (III)

To study whether a shift toward more plant-based diets would benefit CRC prevention, I modelled partial substitutions of 100 g/week of red meat or 50 g/week of processed meat with a corresponding amount of whole grains, vegetables or fruits in relation to CRC risk. Small reductions in CRC risk were observed when red or processed meat was substituted with vegetables or fruits. While the substitutions with whole grains were not associated with CRC risk in the overall population, up to 8% risk reductions were observed in participants with below median whole grain consumption.

Two prospective studies in Italian (n=44 824, 539 CRC cases) [95] and US (n=489 625, 8995 CRC cases) [94] cohorts have previously modelled substitutions of animal-based foods with plant-based foods in relation to CRC risk. However, both of these studies were conducted at a nutrient level, focusing on protein intake. In the Italian study, substituting 3% of energy intake from red and processed meat protein with plant protein from foods with low glycaemic index (GI; e.g., pasta, vegetables, fruits and legumes) was not associated with CRC risk [95]. In contrast, the substitution with plant protein from high-GI foods (e.g., bread, pizza, rice and potatoes) was associated with increased colon cancer and decreased rectal cancer risk. In the US study, the substitution of protein from red meat with protein from bread, cereal and pasta was associated with up to 14% reduction in CRC risk, whereas no significant risk reductions were observed in the substitutions with protein from other plant sources (nuts; beans and legumes; other plant protein) [94]. Although inconclusive, these findings highlight the relevance of a protein source in CRC risk, suggesting that a food-level approach could offer a more applicable perspective for studying the relationship between dietary shifts toward more plant-based diets and CRC risk.

Whole grains, vegetables and fruits represent core food groups in healthy plant-based diets. Of these, the WCRF/AICR has determined the evidence regarding CRC risk to be strong for whole grains and limited but suggestive for non-starchy vegetables and fruits [5]. Accordingly, in the present study, I expected to observe the strongest associations with CRC risk when red or processed meat was partially substituted with whole grains. While not observed in the overall population, the associations between the substitutions with whole grains and CRC risk became

apparent when ATBC was excluded from the study samples, and when participants with whole grain intake below the population median were examined separately from those with above median intakes. The risk reductions in these sub-populations were the largest observed in this study.

Given the particularly high whole grain intake in ATBC (median 100 g/d), these results suggest that increasing whole grain intake beyond a certain threshold may not provide additional benefits in CRC prevention. Such non-linear association between whole grain intake and CRC risk has not been detected in recent comprehensive meta-analyses of prospective cohorts [10, 11]. Nevertheless, in a study of nearly 500 000 older US adults, the decrease in CRC risk was found to plateau at higher, although relatively modest, whole grain intake levels [36]. In this study, the associations between substitutions with whole grains and CRC risk were attenuated in participants with whole grain intakes above 84 g/d (population median). Given that the average whole grain intake in Finland, as well as in most Western adult populations, is well below 84 g/d [55–62], it is reasonable to assume that the partial substitution of red or processed meat with whole grains could benefit CRC prevention in the majority of the adult population.

The partial substitutions of red or processed meat with vegetables or fruits were associated with small reductions in CRC risk. For red meat substitutions, these reductions were likely primarily driven by the decrease in red meat consumption, given the individual associations between the food groups and CRC risk. In contrast, processed meat consumption was not individually associated with CRC risk, for which vegetables and fruits were likely to contribute to the observed risk reductions. However, despite previous compelling evidence of the carcinogenicity of processed meat [2, 5], the risk reductions were modest.

Overall, the modest magnitude of the risk reductions is likely attributable to the small scale of the modelled substitutions. In addition to the substituted amounts, the mode of substitution may also have influenced the results; because the substitutions were modelled on a gram-for-gram basis, the lower energy density of plant-based foods compared with meat could have resulted in minor differences in total energy intake, despite the energy adjustment. Nonetheless, given the small size of the substitutions, these residual differences in energy intake were unlikely to meaningfully affect the observed associations with CRC risk.

Because the substitutions were small, they would be achievable for most adults, increasing their likelihood of adoption in real-life settings compared with more extensive, hypothetical dietary changes. In Finnish adults, the substitutions corresponded to replacing approximately one and a half days of red meat consumption or one day of processed meat consumption per week with plant-based foods, while the rest of the diet remained unchanged. Despite their small scale, in addition to CRC risk, similar substitutions have been linked to a decreased risk of CVD and T2D in the same Finnish study samples [220, 221]. Given the substantial

contribution of CRC, CVD and T2D to the disease burden and mortality in Finnish adults [4, 222], the results of these studies jointly suggest that already small, easily implemented weekly changes toward more plant-based diets could have significant positive effects on the health of the Finnish adult population. As larger shifts toward more plant-based diets are globally called for, the substitutions modelled in this study could serve as an initial step toward healthier and more sustainable diets, facilitating broader dietary shifts with greater potential health benefits.

To conclude, partial substitutions of red or processed meat with plant-based foods, particularly whole grains, were associated with small reductions in CRC risk in Finnish adults. These findings suggest that such dietary shifts toward more plant-based diets could be promoted as part of CRC primary prevention in the population. However, as this is the first study to model food-level substitutions of animal-based with plant-based foods in relation to CRC risk, further research is required to confirm the findings.

6.4 Associations between the CRC lifestyle index and gut microbiome diversity and composition (IV)

CRC lifestyle index and CRC risk (IV part A)

To study the relationship between CRC-related lifestyles and the gut microbiome, I adapted a new CRC lifestyle index from the standardised WCRF/AICR score [192], including nine major lifestyle and anthropometric risk factors for CRC [5]. The index was associated with a 31% lower CRC risk when comparing participants with the highest (Q5) and the lowest index points (Q1). Furthermore, each one-point increase in the index was associated with a 19% decrease in CRC risk.

The standardised WCRF/AICR score was adapted for a CRC-specific approach, because no existing CRC-specific *a priori* index based on the 2018 WCRF/AICR Report was identified [5, 192]. The standardised score or a similar index has been studied in relation to CRC risk in three previous studies, reporting associations of somewhat similar level to our findings. In a study of 94 778 UK adults, a 21% lower CRC risk was observed in participants within the highest versus the lowest score tertile [106]. In a study of 114 419 US adults, a 36% lower CRC risk was detected in the highest versus lowest score quintile in men (n=45 442) [105]. Finally, in a Spanish study of 7216 elderly individuals, a 48% lower CRC risk was observed in the highest versus lowest score quartile [107].

Compared with the standardised score, the CRC lifestyle index constructed in this thesis excluded fruits, vegetables and ultra-processed foods, which have not been convincingly linked to CRC risk, and included height and dairy products as new components [5]. In addition, dietary fibre was replaced with whole grains (see 4.6). Nevertheless, both the standardised score and the new index included BMI, physical activity, red and processed meat consumption and alcohol intake, which

were also considered the most convincing risk factors for CRC in the 2018 WCRF/AICR Report [5]. Moreover, in this thesis, the individual associations with CRC risk were statistically significant for BMI, red meat consumption and alcohol intake. These aspects likely explain the similar findings between this study and the previous studies mentioned above.

To summarise, the findings of Sub-study IV part A suggest that the CRC lifestyle index is an appropriate method to model CRC-related lifestyles in Finnish adults. Consequently, the CRC lifestyle index was used in Sub-study IV part B to explore the associations between higher and lower-risk lifestyles for CRC and the gut microbiome.

CRC lifestyle index and gut microbiome diversity and composition (IV part B)

To elucidate the role of the gut microbiome in the lifestyle-related risk of CRC, I examined associations between the CRC lifestyle index and gut microbiome diversity and composition in Finnish adults. Participants with higher-risk lifestyles for CRC (lower index points) exhibited lower microbial alpha diversity compared to those with lower-risk lifestyles. In addition, the CRC lifestyle index was associated with differences in the gut microbiome composition (beta diversity) between participants. In the species-level analyses, higher-risk lifestyles were associated with enrichment in the relative abundance of several bacterial species previously linked to CRC.

Associations between overall CRC-related lifestyles and the gut microbiome have not been studied before. Furthermore, the current evidence of individual CRC risk factors and the gut microbiome diversity and composition remains inconclusive [136, 143, 153, 223–227]. Observational studies have generally reported non-significant associations between major CRC risk factors and microbial alpha diversity, although higher BMI has been frequently associated with lower alpha diversity [136, 143, 153, 224–227]. Significant associations between the risk factors and beta diversity have been reported more often, although inconsistencies remain. Based on findings of individual risk factors in this thesis, particularly BMI, WC, physical activity and alcohol intake appeared to contribute to the associations observed between the CRC lifestyle index and the gut microbiome.

In the species-level analyses, the CRC lifestyle index was inversely associated with the relative abundance of several bacterial species previously linked to CRC; i.e., participants with higher-risk lifestyles for CRC exhibited enrichment in these species. These species represented predominantly the family Lachnospiraceae, which is one of the dominant bacterial families in the human gastrointestinal tract [228]. For example, of Lachnospiraceae, higher-risk lifestyles were associated with a higher relative abundance of species in genera such as *Dorea*, *Mediterraneibacter* and *Lachnoclostridium*. In previous studies, an enrichment of these genera has been observed in participants with CRC compared with healthy controls and in

more advanced compared with earlier stages of colorectal carcinogenesis [124, 131, 229]. Within Lachnospiraceae, a significant association was also observed between higher-risk lifestyles and a higher relative abundance of *Hungatella hathewayi*, which has been previously linked with TMA-production – one of the key pathways thought to mediate the adverse health effects of red and processed meat consumption (see 2.2.4) [230]. Regarding other bacterial families, significant associations were observed between higher-risk lifestyles for CRC and a higher relative abundance of species within genera, such as *Collinsella*, *Selenomonas*, *Peptostreptococcus* and *Actinomyces*, which have been previously linked to CRC [124, 127, 128, 131, 229]. Some of these species have also been associated with major risk factors for CRC, such as obesity [127, 153].

Along with the enrichment in CRC-related species, higher-risk lifestyles for CRC were also associated with a lower relative abundance of some putatively beneficial species. These included species within the genera *Bifidobacterium* and *Eubacterium*. In previous studies, depletion of *Bifidobacterium* has been associated with obesity, as well as CRC, particularly in the early disease stages [127, 131, 156]. Many species within *Bifidobacterium* are known to ferment dietary fibre and produce SCFAs, and their lower abundance may arise from unhealthy dietary habits, particularly low fibre intake. Species within the genus *Eubacterium* are also capable of fermenting dietary fibre. Nevertheless, in this study, only one species within *Eubacterium* was depleted in participants with higher-risk lifestyles for CRC, whereas the relative abundances of two other species from the same genus were found enriched.

While the findings of this study included several species previously associated with CRC, no statistically significant associations were detected with the species most consistently implicated in elevated CRC risk in prior research, including *Fusobacterium nucleatum,* enterotoxigenic *Bacteroides fragilis* and cytotoxic *Escherichia coli* [129]. *F. nucleatum* has been linked, in particular, with later CRC stages [129, 131], which may explain why no association was observed in the present study of apparently healthy participants. In contrast, enterotoxigenic *B. fragilis* has been implicated in CRC pathogenesis in the early disease stages [129], for which association with it could have been expected also in this study. However, it is important to note that this study did not explore CRC risk, but lifestyle patterns that predispose to or protect against CRC. Thus, while the findings align with previous observations, this study addresses a distinct research question and should be viewed as complementary rather than directly comparable to prior research.

To conclude, the results of this study suggest that higher-risk lifestyles for CRC are associated with lower gut microbial diversity and distinct gut microbiome composition compared with lower-risk lifestyles. Higher-risk lifestyles were also associated with several bacterial species previously linked to CRC. Considering the apparently healthy study population, these findings may reflect microbial traits

characteristic of the early stages of CRC pathogenesis. Consequently, these findings may contribute to research aiming at identifying predictive biomarkers for elevated risk of subsequent cancer development.

6.5 Strengths and limitations

Study samples

The primary strength of this thesis is the use of population-based data of Finnish adults from five large health-examination studies, complemented by two restricted samples in the pooled study samples (sub-studies III and IV part A). In each health-examination study, the participation rate was reasonably high, ranging from 58% in FinHealth 2017 to 84% in Health 2000. However, the higher likelihood of health-conscious individuals to participate in health-examination studies, compared with those with lower health awareness, may have introduced bias to the results [231–233]. This bias would likely have led to an underestimation of the true associations, resulting in more conservative estimates.

Another potential source of bias relates to the inclusion of ATBC within the pooled study samples, as the study comprised exclusively middle-aged male current smokers. ATBC was the largest cohort in the pooled data, and thus, had a significant weight in the analyses. Nonetheless, the other four cohorts in the pooled data were more representative of the general adult population, three of them being population-based. In the meta-analyses, the differences between ATBC and the other study samples were accounted for by using a random-effects model, which assumes variation in true effect sizes across studies and gives more weight to smaller studies compared with a fixed-effect model [234]. Both the random-effects and fixed-effect models ultimately produced similar results, and neither heterogeneity between the cohorts nor interaction by sex were observed in the analyses. As an additional analysis, the study samples were also pooled excluding ATBC, and the results were predominantly similar. Furthermore, although smoking could not be accounted for in ATBC, adjustment for smoking in the other study samples did not materially influence the results. Nevertheless, residual confounding by smoking likely remains.

Overall, although the results of this thesis are based on Finnish adults and may be influenced by certain potential sources of bias, the dietary patterns and risk factors examined (II–IV) are broadly similar to those observed in other Western populations, supporting cautious generalisation of the results.

Study design

All sub-studies in this thesis were observational, and thus, the results cannot establish causation. Additionally, although sub-studies III and IV part A employed a prospective design, exposure data were restricted to baseline, preventing the

consideration of potential changes in diet, lifestyle, anthropometric measures or confounders during follow-up. Such changes may have introduced some imprecision in the results. Nonetheless, given the long latency period of cancer development, studying exposures occurring earlier in life may be more relevant to disease onset than more recent exposures. Thus, the reasonably long follow-up times in the prospective studies represent a major strength of this thesis.

Variables

The universal limitation of dietary assessment methods is their susceptibility to misreporting, whether related to challenges in remembering or estimating past food consumption, or tendency to overestimate the consumption of healthy foods and underestimate the consumption of unhealthy foods. These biases also apply to this thesis. The potential bias related to misreporting of energy intake was addressed by three approaches: 1) excluding participants with extreme energy intakes (low or high), 2) adjusting for total energy intake and 3) adjusting for energy under-reporting or excluding participants determined as energy under-reporters. In each sub-study, at least two of these approaches were applied.

The primary dietary assessment method used in this thesis was the FFQ, which is a widely accepted and applied method in nutritional epidemiology [169]. The FFQ used in sub-studies I–IV part A has been repeatedly validated in the Finnish adult population, demonstrating acceptable performance in ranking participants according to their food consumption and nutrient intake [166–168]. This was also its intended use in this thesis. In Sub-study IV part B, I used dietary data collected through a 48h recall, covering two consecutive days. As the recall days were likely correlated, the food consumption data may not have accurately reflected longer-term absolute intake or captured habitual diet. However, the recall has been previously assessed against a five-day dietary record (48h recall and two-day food diary combined) demonstrating appropriate comparability [163]. Moreover, the food groups studied in Sub-study IV (dietary components within the CRC lifestyle index) are regularly consumed in habitual Finnish diets, for which the 48h recall was deemed to provide an adequate estimate of their habitual consumption. Based on these considerations, the use of two different dietary assessment methods in Sub-study IV (FFQ in part A and 48h recall in part B) was also not expected to introduce any meaningful discrepancy in the index performance, particularly since diet was only one of the index components.

To estimate whole grain intake, a new, up-to-date whole grain database was compiled based on a widely accepted whole grain definition and the current recommendations for standardised whole grain intake estimation [20]. This represents a major strength of this thesis. Furthermore, by evaluating potential surrogate measures for whole grain intake in Sub-study I, reliable estimation of whole grain intake was enabled in the earlier study samples (sub-studies III and IV), where the new database could not be applied. Although the validity of the

surrogate measures was established using more recent food consumption data, rye and oat were likely major contributors to whole grain intake also in the earlier study samples, given their key role in traditional Finnish diets. Overall, our study population may have been particularly well-suited for whole grain research considering the widespread use of whole grain cereals, relatively high average whole grain intake and wide intake range.

Another key strength of this thesis was the availability of comprehensive data on the health status and health behaviour of the participants. In addition to data collected through questionnaires and health examinations, information on health and medication use was obtained from national health registers. For example, information on CRC diagnoses was obtained from the Finnish Cancer Registry with nearly complete case ascertainment [195]. Besides being used as primary study variables, these data allowed for thorough consideration of potential confounders in the analyses. Residual confounding may, however, remain due to variables that were not identified as confounders based on the literature. Moreover, certain potential confounders, such as genetic factors in CRC risk, could not be considered. That said, the primary confounders were accounted for in each sub-study.

The CRC lifestyle index constructed in Sub-study IV is subject to two key limitations that should be considered when interpreting the results. First, although the index was constructed based on previous literature, data-specific cut-offs were used for two components; for height due to the lack of previous data, and for physical activity due to restrictions of our own data. This may affect the comparability of results with future studies. Second, the main components of the index (body fatness, height, physical activity, diet) were assigned equal weights, which may have over- or underestimated the contribution of some components relative to others. Nevertheless, existing literature suggests that the (sub-) components are associated with relatively comparable levels of CRC risk, which may mitigate potential bias arising from equal weighting [5]. However, owing to these limitations, as well as those related to the pooled data used to test the index, further validation of the index is warranted. Moreover, the application of standardised cut-offs for height and physical activity should be further explored.

Finally, in Sub-study IV part B, the gut microbiome was characterised using metagenomic sequencing, which is an important strength of this thesis. Although shallow metagenomic sequencing used in this thesis is less sensitive than deep sequencing for detecting low-abundance taxa, previous research has demonstrated sufficient accuracy also at lower sequencing depths for commonly used diversity metrics and differential abundance analyses, such as those applied in this study [185]. Overall, metagenomic sequencing provides higher taxonomic resolution than 16s rRNA amplicon-based sequencing, which has been used in most previous studies, enabling more accurate identification and classification of microbial taxa [185].

6.6 Future perspectives

Strong evidence of associations between lifestyle factors and CRC risk, as well as of the underlying mechanisms, is essential for public health measures aimed at reducing CRC burden. Although whole grain intake has been fairly consistently linked to lower CRC risk [11], future studies should ensure broader adoption of standardised intake estimation methods. This includes the use of validated surrogate measures when direct estimation is not feasible, to corroborate the evidence and strengthen the basis for public health communication. In addition to CRC, these methodological considerations are equally relevant regarding the associations between whole grains and other chronic diseases, as well as their risk factors, for which the existing evidence remains particularly inconsistent.

Some of the discrepancies in the associations between whole grain intake and health outcomes may arise from compositional differences between whole grain cereals. Therefore, examining whole grain intake by cereal type, alongside total intake, could provide valuable insights into their health associations. Furthermore, while this thesis demonstrated that diet quality did not modify associations between whole grain intake and chronic disease risk factors, further research is needed to confirm these findings and the independent health benefits of whole grains.

The results of this thesis suggest that moderate, partial substitutions of red or processed meat with plant-based foods could reduce the population burden of CRC in Finnish adults. While two previous studies have examined similar dietary changes in relation to CRC risk [94, 95], they focused on protein intake, and further research is warranted to validate our findings regarding food-level substitutions. More research is also required on substitutions of other animal foods beyond red and processed meat. In particular, given the protective association of dairy consumption with CRC risk, their substitution with plant-based foods requires greater attention as it may have adverse implications for CRC risk. Although dairy alternatives are often fortified with nutrients potentially contributing to dairy's protective effects, their effectiveness in replicating these benefits remains uncertain. Consequently, it remains to be determined whether plant-based dairy substitutes can offer comparable protection against CRC.

Previous studies have reported largely inconsistent findings on the associations between individual CRC risk factors and gut microbiome diversity and composition. Since most prior studies have been based on relatively small study samples and 16S rRNA sequencing, the use of larger, population-based samples and metagenomic sequencing, as in this thesis, could facilitate more consistent and generalisable results. The inconsistencies may also reflect the modest effect of individual lifestyle factors on microbiome composition, whereas the cumulative influence of lifestyle is presumed to be substantial [7]. Therefore, in future research, greater emphasis should be placed on exploring multiple risk factors simultaneously in relation to the gut microbiome, better reflecting the real-world risk environment. This approach

was used in this thesis to examine the gut microbiome in participants with high- and low-risk lifestyles for CRC in a healthy study population. Extending this approach to large, prospective cohorts with a sufficient number of incident CRC cases could substantially advance understanding of the gut microbiome's role in mediating lifestyle-related risk of CRC. Moreover, it could help clarify the temporal relationship between compositional changes in the gut microbiome and CRC, thereby facilitating the identification of early risk markers for clinical use.

Based on the findings of this thesis, more emphasis should be placed on primary prevention of CRC through lifestyle changes at the national level, including effective implementation of national nutrition recommendations. Improving population adherence to these recommendations would not only benefit CRC prevention but would also help alleviate the burden of obesity, a major risk factor for CRC, as well as other chronic diseases. Attention should be paid, in particular, to reducing the consumption of red and processed meat and ensuring the availability of replacements that are acceptable, affordable and feasible for the population to integrate into their daily meals. Similarly, higher whole grain intake should be encouraged by increasing awareness of their value as high-quality carbohydrates that form the foundation of healthy diets, along with practical guidance on shifting from refined to whole grain products. The food industry plays a key role in promoting whole grain intake through the availability of foods high in whole grain that also meet consumer preferences.

Overall, facilitating the adoption of recommended dietary changes requires supportive environments. Food services, for example, represent a powerful platform for nutrition education, particularly among children. Furthermore, identifying and reaching population sub-groups at elevated risk of adverse lifestyle behaviours is essential for targeted interventions to promote healthier lifestyles and reduce health disparities. The CRC lifestyle index developed in this thesis offers one potential tool for guiding such efforts.

7 Conclusions

The aim of this doctoral thesis was to examine whole grain intake, healthy lifestyles and the gut microbiome in the context of colorectal cancer risk, also considering their interrelationships. Based on the findings of the four sub-studies, the following conclusions can be drawn:

1. Rye-based variables, particularly the combined consumption of rye, oat and barley, are reliable surrogate measures for whole grain intake in epidemiological research of Finnish adults. In contrast, bread consumption and fibre intake are less suitable surrogate measures due to their lower correlation with whole grain intake and sensitivity to energy under-reporting. These findings facilitate reliable whole grain intake estimation in Finnish adults when direct estimation is not feasible. The findings also contribute to the global efforts to standardise and validate whole grain intake assessment.
2. Whole grain intake was associated with better diet quality and overall healthier lifestyles in Finnish adults. In men, whole grain intake was also associated with lower BMI, WC and total cholesterol concentration, whereas no significant associations were observed in women. Diet quality did not modify these associations, providing further support for independent associations between whole grain intake and health outcomes.
3. Partial substitutions of red or processed meat with vegetables, fruits or whole grains were associated with reduced CRC risk. The associations between substitutions with whole grains and CRC risk were significant only in participants with lower whole grain intakes, suggesting a plateau in benefits at high intake levels. Overall, these findings support the promotion of more plant-based diets as part of CRC primary prevention.
4. Higher points in the CRC lifestyle index, indicating fewer risk factors for CRC, were associated with markedly lower CRC risk both across quintiles and per 1-point increase in the index. These results demonstrate that the index functioned in Finnish adults as intended, and it can be used to distinguish participants with higher- and lower-risk lifestyles for CRC in an epidemiological context.

5. The CRC lifestyle index was associated with compositional differences in the gut microbiome, higher-risk lifestyles being associated with lower microbial diversity and higher relative abundance of several bacterial species potentially related to CRC. Given the apparently healthy study population with only few CRC cases over a long follow-up, these findings suggest that CRC-related microbial traits may already be present before disease onset among individuals at high lifestyle-related risk of CRC.

Acknowledgements

I conducted the research presented in this thesis at the Finnish Institute for Health and Welfare (THL), in the Department of Public Health. I gratefully acknowledge THL and its personnel for the excellent research environment and the support I have received throughout this project. In particular, I am indebted to all current and former THL researchers who contributed to collecting the research data used in this thesis, as well as to all participants involved.

This doctoral thesis would not have been possible without the generous financial support of the Juho Vainio Foundation, the Yrjö Jahnsson Foundation and the Cancer Foundation Finland. I am also grateful to the University of Helsinki Doctoral School, the Finnish Society for Nutrition Research, the Finnish Association of Academic Agronomists, the Nordic Cancer Union and the Cancer Society of Finland for the opportunities to present my research at international conferences and participate in courses abroad.

Although this doctoral thesis is, in name, the achievement of one person, the credit belongs to many. First and foremost, credit is due to Docent Satu Männistö, Docent Niina Kaartinen and Dr Mirkka Maukonen who have guided my work since my master's thesis. With their encouragement, I took the leap to become a doctoral researcher, and I have never regretted that decision. I am deeply grateful for your time, support, enthusiasm and investment in this project—to you I owe the joy of having found my calling.

In addition to having great supervisors, I have also been exceedingly lucky in being part of the Leg4Life research group at THL that has felt like a safety net in the uncertain research world. Our weekly meetings have given me a sense of belonging that has had a profound role in forming my doctoral experience. Overall, it has been a privilege to kick-start my career within the Leg4Life project, which has provided both a supportive research community and exceptional opportunities along the way. In addition to Leg4Life, I have also had the opportunity to participate in another project, SUST-360, which has also been a rewarding experience.

Sharing this journey with fellow doctoral researchers has been one of the greatest joys of my training. I am especially grateful to Tiina, Tuulia, Tuuli and Meri for their peer support throughout these years. Tiina deserves a special shout-out for brightening each workday with her good humour and for always being available at

the other end of a Teams chat, whether for crises and celebrations or for memes and daily conversations.

I sincerely thank all my co-authors for their contributions to each sub-study and for their constructive comments and advise. In particular, I would like to thank Statistical Specialist Heli Tapanainen for her support, both in my research and in her capacity as Team Leader at THL, and Senior Specialist Heli Reinivuo for involving me in the whole grain project that became the foundation for my doctoral research. I am also grateful to IT Systems Analyst Kennet Harald for his help with the data and statistical analyses throughout the project, Associate Professor Maarit Laaksonen for her contribution to the third sub-study and for hosting me during my research visit at the University of Sydney, and Dr Kari Koponen for his pivotal contribution to and expert guidance on the microbiome analyses in the fourth sub-study. I also want to acknowledge my thesis committee members, Professor Marjukka Kolehmainen and Professor Kirsi Laitinen, whose guidance was valuable in ensuring the successful progress of this project.

Finally, I am grateful to Docent Jyrki Virtanen and Dr Cecilie Kyrø for providing valuable feedback on my thesis as pre-examiners. I would also like to sincerely thank Professor Guri Skeie for kindly accepting the invitation to serve as opponent and Professor Anna Keski-Rahkonen for agreeing to serve as custos in my thesis defence.

I have gained a strong foundation for life and a deep belief in the importance of education from my parents and family. All my successes, both as a researcher and as a person, are undoubtedly partly theirs. Moreover, I may never have begun studying nutrition if my mother had not suggested it to me. Small words can echo forever.

I took the earliest steps toward becoming a researcher while living in Melbourne, Australia. In the Parkville student accommodation, I had a neighbour three doors down who, from day one, has walked beside me throughout this journey. This thesis, in its completion, owes much to that neighbour, João, who is now my husband and my first and final reviewer.

Espoo, February 2026

Rilla Tammi-Pereira

References

1. Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, Garnett T, Tilman D, DeClerck F, Wood A, Jonell M, Clark M, Gordon LJ, Fanzo J, Hawkes C, Zurayk R, Rivera JA, De Vries W, Majele Sibanda L, Afshin A, Chaudhary A, Herrero M, Agustina R, Branca F, Lartey A, Fan S, Crona B, Fox E, Bignet V, Troell M, Lindahl T, Singh S, Cornell SE, Srinath Reddy K, Narain S, Nishtar S, Murray CJL. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *Lancet*. 393:447–92, 2019.
2. Wild CP, Weiderpass E, Stewart BW (editors). World Cancer Report: Cancer Research for Cancer Prevention. International Agency for Research on Cancer: Lyon, France; 2020.
3. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 74:229–63, 2024.
4. Seppä K, Lappi-Heikkinen S, Johansson S, Malila N, Pitkäniemi J. Syöpä 2023. Tilastoraportti Suomen syöpätilanteesta. Suomen Syöpäyhdistys: Helsinki, Finland; 2025.
5. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and colorectal cancer. 2018.
6. Song M, Chan AT, Sun J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology*. 158:322–40, 2020.
7. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 555:210–5, 2018.
8. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D'hoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. Population-level analysis of gut microbiome variation. *Science*. 352:560–4, 2016.
9. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, Lucas SK, Beura LK, Thompson EA, Till LM, Batres R, Paw B, Pergament SL, Saenyakul P, Xiong M, Kim AD, Kim G, Masopust D, Martens EC, Angkurawaranon C, McGready R, Kashyap PC, Culhane-Pera KA, Knights D. US Immigration Westernizes the Human Gut Microbiome. *Cell*. 175:962-972.e10, 2018.

10. Vieira AR, Abar L, Chan DSM, Vingeliene S, Polemiti E, Stevens C, Greenwood D, Norat T. Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR Continuous Update Project. *Ann Oncol*. 28:1788–802, 2017.
11. 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*. 393:434–45, 2019.
12. Schlesinger S, Neuenschwander M, Schwedhelm C, Hoffmann G, Bechthold A, Boeing H, Schwingshackl L. Food Groups and Risk of Overweight, Obesity, and Weight Gain: A Systematic Review and Dose-Response Meta-Analysis of Prospective Studies. *Adv Nutr*. 10:205–18, 2019.
13. Buyken AE, Goletzke J, Joslowski G, Felbick A, Cheng G, Herder C, Brand-Miller JC. Association between carbohydrate quality and inflammatory markers: systematic review of observational and interventional studies. *Am J Clin Nutr*. 99:813–33, 2014.
14. Kyrø C, Skeie G, Dragsted LO, Christensen J, Overvad K, Hallmans G, Johansson I, Lund E, Slimani N, Johnsen NF, Halkjær J, Tjønneland A, Olsen A. Intake of whole grains in Scandinavia is associated with healthy lifestyle, socio-economic and dietary factors. *Public Health Nutr*. 14:1787–95, 2011.
15. Barrett EM, Batterham MJ, Beck EJ. Whole grain and cereal fibre intake in the Australian Health Survey: associations to CVD risk factors. *Public Health Nutr*. 23:1404–13, 2020.
16. Barrett EM, Amoutzopoulos B, Batterham MJ, Ray S, Beck EJ. Whole grain intake compared with cereal fibre intake in association to CVD risk factors: a cross-sectional analysis of the National Diet and Nutrition Survey (UK). *Public Health Nutr*. 23:1392–403, 2020.
17. Jensen MK, Koh-Banerjee P, Franz M, Sampson L, Grønbæk M, Rimm EB. Whole grains, bran, and germ in relation to homocysteine and markers of glycemic control, lipids, and inflammation. *Am J Clin Nutr*. 83:275–83, 2006.
18. Andersen JLM, Halkjær J, Rostgaard-Hansen AL, Martinussen N, Lund ASQ, Kyrø C, Tjønneland A, Olsen A. Intake of whole grain and associations with lifestyle and demographics: a cross-sectional study based on the Danish Diet, Cancer and Health—Next Generations cohort. Eur J Nutr. 60:883–95, 2021.
19. Seal CJ, Jones AR, Whitney AD. Whole grains uncovered. *Nutr Bull*. 31:129–37, 2006.
20. Ross AB, Kristensen M, Seal CJ, Jacques P, McKeown NM. Recommendations for reporting whole-grain intake in observational and intervention studies. *Am J Clin Nutr*. 101:903–7, 2015.
21. van der Kamp JW, Jones JM, Miller KB, Ross AB, Seal CJ, Tan B, Beck EJ. Consensus, Global Definitions of Whole Grain as a Food Ingredient and of Whole-Grain Foods Presented on Behalf of the Whole Grain Initiative. *Nutrients*. 14:138, 2021.
22. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, Mullany EC, Abate KH, Abbafati C, Abebe Z, Afarideh M, Aggarwal A, Agrawal S, Akinyemiju T, Alahdab F, Bacha U, Bachman VF, Badali H, Badawi A, Bensenor IM, Bernabe E, Biadgilign SKK, Biryukov SH, Cahill LE, Carrero JJ, Cercy KM, Dandona L, Dandona R, Dang AK, Degefa MG, El Sayed Zaki

M, Esteghamati A, Esteghamati S, Fanzo J, Farinha CSES, Farvid MS, Farzadfar F, Feigin VL, Fernandes JC, Flor LS, Foigt NA, Forouzanfar MH, Ganji M, Geleijnse JM, Gillum RF, Goulart AC, Grosso G, Guessous I, Hamidi S, Hankey GJ, Harikrishnan S, Hassen HY, Hay SI, Hoang CL, Horino M, Ikeda N, Islami F, Jackson MD, James SL, Johansson L, Jonas JB, Kasaeian A, Khader YS, Khalil IA, Khang YH, Kimokoti RW, Kokubo Y, Kumar GA, Lallukka T, Lopez AD, Lorkowski S, Lotufo PA, Lozano R, Malekzadeh R, März W, Meier T, Melaku YA, Mendoza W, Mensink GBM, Micha R, Miller TR, Mirarefin M, Mohan V, Mokdad AH, Mozaffarian D, Nagel G, Naghavi M, Nguyen CT, Nixon MR, Ong KL, Pereira DM, Poustchi H, Qorbani M, Rai RK, Razo-García C, Rehm CD, Rivera JA, Rodríguez-Ramírez S, Roshandel G, Roth GA, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 393:1958–72, 2019.

23. Blomhoff R, Andersen R, Arnesen EK, Christensen JJ, Eneroth H, Erkkola M, Gudanaviciene I, Halldorsson TI, Høyer-Lund A, Lemming EW, Meltzer HM, Pitsi T, Schwab U, Siksna I, Thorsdottir I, Trolle E. Nordic Nutrition recommendations 2023. Nordic Council of Ministers: Copenhagen, Denmark; 2023.
24. Kromhout D, Spaaij CJK, de Goede J, Weggemans RM. The 2015 Dutch food-based dietary guidelines. *Eur J Clin Nutr*. 70:869–78, 2016.
25. National Health and Medical Research Council. Australian Dietary Guidelines. National Health and Medical Research Council: Canberra, Australia; 2013.
26. U.S. Department of Agriculture and U.S. Department of Health and Human Services. Dietary Guidelines for Americans, 2020-2025. 9th Edition. 2020.
27. Bonvecchio Arenas A, Fernández-Gaxiola AC, Plazas Belausteguigoitia M, Kaufer-Horwitz M, Pérez Lizaur AB, Rivera Dommarco J (editors). Guías alimentarias y de actividad física: en contexto de sobrepeso y obesidad en la población mexicana : documento de postura. Academia Nacional de Medicina: México; 2015.
28. Fulcher RG, Rooney Duke TK. Whole-Grain Structure and Organization: Implications for Nutritionists and Processors. In Fulcher RG, Marquart L, Slavin JL. Whole-grain foods in health and disease. American Association of Cereal Chemists : St. Paul, MN, USA; 2002. p. 9–45.
29. van der Kamp JW, Poutanen K, Seal CJ, Richardson DP. The HEALTHGRAIN definition of 'whole grain'. *Food Nutr Res*. 58:22100, 2014.
30. Cereals & Grains Association. Whole grains. Available at: https://www.cerealsgrains.org/Pages/default.aspx (Accessed 24 April 2025).
31. Jacobs DR, Meyer KA, Kushi LH, Folsom AR. Whole-grain intake may reduce the risk of ischemic heart disease death in postmenopausal women: the Iowa Women's Health Study. *Am J Clin Nutr*. 68:248–57, 1998.
32. Lutsey PL, Jacobs DR, Kori S, Mayer-Davis E, Shea S, Steffen LM, Szklo M, Tracy R. Whole grain intake and its cross-sectional association with obesity, insulin resistance, inflammation, diabetes and subclinical CVD: The MESA Study. *Br J Nutr*. 98:397–405, 2007.
33. Larsson SC, Giovannucci E, Bergkvist L, Wolk A. Whole grain consumption and risk of colorectal cancer: a population-based cohort of 60 000 women. *Br J Cancer*. 92:1803–7, 2005.
34. Fisberg RM, Fontanelli MM, Kowalskys I, Gómez G, Rigotti A, Cortés LY, García MY, Pareja RG, Herrera-Cuenca M, Fisberg M, ELANS Study Group,

Fisberg M, Kovalskys I, Salas GG, Rigotti A, Sanabria LYC, García MCY, Torres RGP, Herrera-Cuenca M, Koletzko B, Moreno LA, Pratt M, Fisberg RM, Previdelli AN, Guajardo V, Zimberg IZ, Guajardo V, Amigo MP, Janezic X, Cardini F, Echeverry M, de França NAG, Echeverría G, Landaeta L, Castillo Ó, Vargas LN, Castillo LFTYM, Rojas RM, Cáceres MV, Ocampo MB, Liria MR, Meza K, Abad M, Cuenca MH, Landaeta-Jiménez M, Méndez B, Vásquez M, Ramírez G, Hernández P, Meza C, Rivas O, Morales V, Gonçalves PB, Alberico C, de Moraes Ferrari GL. Total and whole grain intake in Latin America: findings from the multicenter cross-sectional Latin American Study of Health and Nutrition (ELANS). *Eur J Nutr* . 61:489–501, 2022.

35. Huang Q, Hao L, Wang L, Jiang H, Li W, Wang S, Jia X, Huang F, Wang H, Zhang B, Ding G, Wang Z. Differential Associations of Intakes of Whole Grains and Coarse Grains with Risks of Cardiometabolic Factors among Adults in China. *Nutrients*. 14:2109, 2022.
36. Schatzkin A, Mouw T, Park Y, Subar AF, Kipnis V, Hollenbeck A, Leitzmann MF, Thompson FE. Dietary fiber and whole-grain consumption in relation to colorectal cancer in the NIH-AARP Diet and Health Study. *Am J Clin Nutr*. 85:1353– 60, 2007.
37. National Nutrition Council and Finnish Institute for Health and Welfare. Sustainable health from food – National Nutrition Recommendations 2024. PunaMusta Oy: Helsinki, Finland; 2024.
38. Livsmedelsverket. Kostråd för vuxna. 2025. Available at: https://www.livsmedelsverket.se/matvanor-halsa--miljo/kostrad/kostrad-vuxna (Accessed 13 Jun 2025).
39. Helse Norge. Dietary guidelines. 2024. Available at: https://www.helsenorge.no/en/kosthold-og-ernaring/dietary-guidelines/. (Accessed 13 Jun 2025)
40. Ministry of Food, Agriculture and Fisheries of Denmark/The Danish Veterinary and Food Administration. The Official Dietary guidelines – good for health and climate. Ministry of Food, Agriculture and Fisheries: Glostrup, Denmark; 2021.
41. Food Safety Authority of Ireland. Healthy eating, food safety and food legislation – A guide supporting the Healthy Ireland Food Pyramid. Food Safety Authority of Ireland: Dublin, Ireland; 2019.
42. Health Promotion Board. *The eight Dietary Guidelines for Adult Singaporeans* (ages 18-69). Health Promotion Board: Singapore, Singapore; 2003.
43. Tervise Arengu Instituut. Eesti riilikud toitumise, liikumise ja uneaja. Tervise Arengu Instituut: Tallinn, Estonia; 2025.
44. British Nutrition Foundation. The Eatwell Guide. 2022. Available at: https://www.nutrition.org.uk/media/ayth4ma4/eatwel-1.pdf (Accessed 13 Jun 2025).
45. Deutsche Gesellschaft für Ernährung e.V. Eat and drink well – recommendations of the German Nutrition Society (DGE). 2024. Available at: https://www.dge.de/fileadmin/dok/english/dge-recommendations/information-sheet-DGE-recommendations-en.pdf (Accessed 13 Jun 2025).
46. Delamaire C, Escalon E, Noirot L. Recommendations concerning diet, physical activity and sedentary behaviour for adults. Saint-Maurice: Santé publique France: Saint-Maurice, France. 2019.

47. Agencia Española de Seguridad Alimentaria y Nutrición (AESAN). Recomendaciones dietéticas saludables y sostenibles: Come sano, muévete y cuida tu planeta. Recomendaciones dietéticas saludables y sostenibles complementadas con recomendaciones de actividad física para la población española. AESAN: Madrid, Spain; 2022.
48. Health Canada. Canada's Dietary Guidelines for Health Professionals and Policy Makers. Health Canada: Ottawa, ON, Canada; 2019.
49. Ministry of Health. Eating and Activity Guidelines for New Zealand Adults: Updated 2020. Ministry of Health: Wellington, New Zealand; 2020.
50. National Institute of Nutrition. Dietary Guidelines for Indians – A Manual. National Institute of Nutrition: Hyderabad, India; 2011.
51. Sultanate of Oman Ministry of Health. The Omani Guide for Healthy Eating. Sultanate of Oman Ministry of Health: Oman; 2025.
52. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, Benetos A, Biffi A, Boavida JM, Capodanno D, Cosyns B, Crawford C, Davos CH, Desormais I, Di Angelantonio E, Franco OH, Halvorsen S, Hobbs FDR, Hollander M, Jankowska EA, Michal M, Sacco S, Sattar N, Tokgozoglu L, Tonstad S, Tsioufis KP, van Dis I, van Gelder IC, Wanner C, Williams B, ESC Scientific Document Group. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European Association of Preventive Cardiology (EAPC). *Eur Heart J*. 42:3227–337, 2021.
53. American Heart Association. Get to Know Grains: Why You Need Them and What To Look For. Available at: https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/nutrition-basics/whole-grains-refined-grains-and-dietary-fiber (Accessed 13 Jun 2025)
54. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018. Available at dietandcancerreport.org
55. Tammi R, Reinivuo H, Tapanainen H, Rautanen J, Männistö S, Kaartinen N. Täysjyvän saanti Suomen aikuisväestössä - FinRavinto 2017 -tutkimuksen tuloksia. Tutkimuksesta tiiviisti 64/2021. Finnish Institute for Health and Welfare: Helsinki, Finland; 2021.
56. Kyrø C, Skeie G, Dragsted LO, Christensen J, Overvad K, Hallmans G, Johansson I, Lund E, Slimani N, Johnsen NF, HalkjæR J, Tjønneland A, Olsen A. Intake of whole grain in Scandinavia: Intake, sources and compliance with new national recommendations. *Scand J Public Health*. 40:76–84, 2012.
57. Mann KD, Pearce MS, McKevith B, Thielecke F, Seal CJ. Whole grain intake and its association with intakes of other foods, nutrients and markers of health in the National Diet and Nutrition Survey rolling programme 2008–11. *Br J Nutr*. 113:1595–602, 2015.
58. O'Donovan CB, Devlin NF, Buffini M, Walton J, Flynn A, Gibney MJ, Nugent AP, McNulty BA. Whole grain intakes in Irish adults: findings from the National Adults Nutrition Survey (NANS). *Eur J Nutr*. 58:541–50, 2019.
59. Bellisle F, Hébel P, Colin J, Reyé B, Hopkins S. Consumption of whole grains in French children, adolescents and adults. *Br J Nutr*. 112:1674–84, 2014.

60. Sette S, D'Addezio L, Piccinelli R, Hopkins S, Le Donne C, Ferrari M, Mistura L, Turrini A. Intakes of whole grain in an Italian sample of children, adolescents and adults. *Eur J Nutr*. 56:521–33, 2017.
61. Galea LM, Beck EJ, Probst YC, Cashman CJ. Whole grain intake of Australians estimated from a cross-sectional analysis of dietary intake data from the 2011–13 Australian Health Survey. *Public Health Nutr*. 20:2166–72, 2017.
62. Albertson AM, Reicks M, Joshi N, Gugger CK. Whole grain consumption trends and associations with body weight measures in the United States: results from the cross sectional National Health and Nutrition Examination Survey 2001–2012. *Nutr J*. 15:8, 2015.
63. Hauner H, Bechthold A, Boeing H, Brönstrup A, Buyken A, Leschik-Bonnet E, Linseisen J, Schulze M, Strohm D, Wolfram G. Evidence-Based Guideline of the German Nutrition Society: Carbohydrate Intake and Prevention of Nutrition-Related Diseases. *Ann Nutr Metab*. 60:1–58, 2012.
64. Åkesson A, Andersen LF, Kristjánsdóttir ÁG, Roos E, Trolle E, Voutilainen E, Wirfält E. Health effects associated with foods characteristic of the Nordic diet: a systematic literature review. *Food Nutr Res*. 57:22790, 2013.
65. Scientific Advisory Committee on Nutrition. Carbohydrates and Health. Public Health England: London, the UK; 2015.
66. Skeie G, Fadnes LT. Cereals and cereal products – a scoping review for Nordic Nutrition Recommendations 2023. *Food Nutr Res*. 68:10457, 2024.
67. Lourenço S (editor). Whole Grain: definition, evidence base review, sustainability aspects and considerations for a dietary guideline. WholEUGrain: Copenhagen, Denmark; 2021.
68. Fusco W, Lorenzo MB, Cintoni M, Porcari S, Rinninella E, Kaitsas F, Lener E, Mele MC, Gasbarrini A, Collado MC, Cammarota G, Ianiro G. Short-Chain Fatty-Acid-Producing Bacteria: Key Components of the Human Gut Microbiota. *Nutrients*. 15:2211, 2023.
69. Slavin JL, Martini MC, Jacobs DR, Marquart L. Plausible mechanisms for the protectiveness of whole grains. *Am J Clin Nutr*. 70:459S-63S, 1999.
70. Fardet A. New hypotheses for the health-protective mechanisms of whole-grain cereals: what is beyond fibre? *Nutr Res Rev*. 23:65–134, 2010.
71. McKeown NM, Meigs JB, Liu S, Wilson PW, Jacques PF. Whole-grain intake is favorably associated with metabolic risk factors for type 2 diabetes and cardiovascular disease in the Framingham Offspring Study. *Am J Clin Nutr*. 76:390–8, 2002.
72. Newby P, Maras J, Bakun P, Muller D, Ferrucci L, Tucker KL. Intake of whole grains, refined grains, and cereal fiber measured with 7-d diet records and associations with risk factors for chronic disease1–3. *Am J Clin Nutr*. 86:1745–53, 2007.
73. van de Vijver LPL, van den Bosch LMC, van den Brandt PA, Goldbohm RA. Whole-grain consumption, dietary fibre intake and body mass index in the Netherlands cohort study. *Eur J Clin Nutr*. 63:31–8, 2009.
74. O'Neil CE, Zanovec M, Cho SS, Nicklas TA. Whole grain and fiber consumption are associated with lower body weight measures in US adults: National Health and Nutrition Examination Survey 1999-2004. *Nutr Res*. 30:815–22, 2010.

75. Barrett EM, Batterham MJ, Ray S, Beck EJ. Whole grain, bran and cereal fibre consumption and CVD: a systematic review. *Br J Nutr*. 121:914–37, 2019.

76. Egeberg R, Olsen A, Loft S, Christensen J, Johnsen NF, Overvad K, Tjønneland A. Intake of wholegrain products and risk of colorectal cancers in the Diet, Cancer and Health cohort study. *Br J Cancer*. 103:730–4, 2010.

77. Johnsen NF, Frederiksen K, Christensen J, Skeie G, Lund E, Landberg R, Johansson I, Nilsson LM, Halkjær J, Olsen A, Overvad K, Tjønneland A. Whole-grain products and whole-grain types are associated with lower all-cause and cause-specific mortality in the Scandinavian HELGA cohort. *Br J Nutr*. 114:608–23, 2015.

78. Montonen J, Knekt P, Järvinen R, Aromaa A, Reunanen A. Whole-grain and fiber intake and the incidence of type 2 diabetes. *Am J Clin Nutr*. 77:622–9, 2003.

79. Wirström T, Hilding A, Gu HF, Östenson CG, Björklund A. Consumption of whole grain reduces risk of deteriorating glucose tolerance, including progression to prediabetes. *Am J Clin Nutr*. 97:179–87, 2013.

80. Pietinen P, Malila N, Virtanen M, Hartman TJ, Tangrea JA, Albanes D, Virtamo J. Diet and risk of colorectal cancer in a cohort of Finnish men. *Cancer Causes Control*. 10:387-396, 1999.

81. Masters RC, Liese AD, Haffner SM, Wagenknecht LE, Hanley AJ. Whole and Refined Grain Intakes Are Related to Inflammatory Protein Concentrations in Human Plasma. *J Nutr*. 140:587–94, 2010.

82. Taskinen RE, Hantunen S, Tuomainen TP, Virtanen JK. The associations between whole grain and refined grain intakes and serum C-reactive protein. *Eur J Clin Nutr*. 76:544–550, 2022.

83. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 71:209–49, 2021.

84. Araghi M, Soerjomataram I, Bardot A, Ferlay J, Cabasag CJ, Morrison DS, De P, Tervonen H, Walsh PM, Bucher O, Engholm G, Jackson C, McClure C, Woods RR, Saint-Jacques N, Morgan E, Ransom D, Thursfield V, Møller B, Leonfellner S, Guren MG, Bray F, Arnold M. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *Lancet Gastroenterol Hepatol*. 4:511–518, 2019.

85. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet*. 394:1467–1480, 2019.

86. Malcomson FC. Mechanisms underlying the effects of nutrition, adiposity and physical activity on colorectal cancer risk. *Nutr Bull*. 43:400–15, 2018.

87. Umanskiy K, Matthews JB. Colon: Anatomy and Structural Anomalies. In: Podolsky D, Camilleri M, Fitz G, Kalloo AN, Shanahan F, Wang TC. Yamada' s Textbook of Gastroenterology. John Wiley & Sons: Hoboken, NJ, USA; 2015. p. 93–107.

88. Shaukat A, Levin TR. Current and future colorectal cancer screening strategies. *Nat Rev Gastroenterol Hepatol*. 19:521–31, 2022.

89. Finnish Cancer Registry. Colorectal cancer screening. Available at: https://cancerregistry.fi/screening/colorectal-cancer-screening/ (Accessed 20 Jun 2025)

90. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J, Murphy N, Bray F. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 72:338–44, 2023.

91. International Agency for Research on Cancer. Red Meat and Processed meat. IARC monographs on the evaluation of carcinogenic risks to humans. Volume 114. International Agency for Research on Cancer: Lyon, France; 2018.

92. International Agency for Research on Cancer. Reduction or Cessation of Alcoholic Beverages Consumption. IARC Handbooks of Cancer Prevention. Volume 20A. International Agency for Research on Cancer: Lyon, France; 2024.

93. Arnold M, Pandeya N, Byrnes G, Renehan AG, Stevens GA, Ezzati M, Ferlay J, Miranda JJ, Romieu I, Dikshit R, Forman D, Soerjomataram I. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol*. 16:36–46, 2015.

94. Liao LM, Loftfield E, Etemadi A, Graubard BI, Sinha R. Substitution of dietary protein sources in relation to colorectal cancer risk in the NIH-AARP cohort study. *Cancer Causes Control*. 30:1127–35, 2019.

95. Sieri S, Agnoli C, Pala V, Grioni S, Palli D, Bendinelli B, Macciotta A, Ricceri F, Panico S, De Magistris MS, Tumino R, Fontana L, Krogh V. Dietary Intakes of Animal and Plant Proteins and Risk of Colorectal Cancer: The EPIC-Italy Cohort. *Cancers*. 14:2917, 2022.

96. Roos E, Seppä K, Pietiläinen O, Ryynänen H, Heikkinen S, Eriksson JG, Härkänen T, Jousilahti P, Knekt P, Koskinen S, Laaksonen M, Männistö S, Roos T, Rahkonen O, Malila N, Pitkäniemi J, Group TMS. Pairwise association of key lifestyle factors and risk of colorectal cancer: a prospective pooled multicohort study. *Cancer Reports*. e1612, 2022.

97. Aleksandrova K, Pischon T, Jenab M, Bueno-de-Mesquita HB, Fedirko V, Norat T, Romaguera D, Knüppel S, Boutron-Ruault MC, Dossus L, Dartois L, Kaaks R, Li K, Tjønneland A, Overvad K, Quirós JR, Buckland G, Sánchez MJ, Dorronsoro M, Chirlaque MD, Barricarte A, Khaw KT, Wareham NJ, Bradbury KE, Trichopoulou A, Lagiou P, Trichopoulos D, Palli D, Krogh V, Tumino R, Naccarati A, Panico S, Siersema PD, Peeters PH, Ljuslinder I, Johansson I, Ericson U, Ohlsson B, Weiderpass E, Skeie G, Borch KB, Rinaldi S, Romieu I, Kong J, Gunter MJ, Ward HA, Riboli E, Boeing H. Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. *BMC Med*. 12:168, 2014.

98. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 13:3–9, 2002.

99. Kastorini CM, Papadakis G, Milionis HJ, Kalantzi K, Puddu PE, Nikolaou V, Vemmos KN, Goudevenos JA, Panagiotakos DB. Comparative analysis of a-priori and a-posteriori dietary patterns using state-of-the-art classification algorithms: A case/case-control study. *Artif Intell Med*. 59:175–83, 2013.

100. Jones P, Cade JE, Evans CEL, Hancock N, Greenwood DC. Does adherence to the World Cancer Research Fund/American Institute of Cancer Research cancer prevention guidelines reduce risk of colorectal cancer in the UK Women's Cohort Study? *Br J Nutr*. 119:340–8, 2018.

101. Makarem N, Lin Y, Bandera EV, Jacques PF, Parekh N. Concordance with World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) guidelines for cancer prevention and obesity-related cancer

risk in the Framingham Offspring cohort (1991–2008). *Cancer Causes Control*. 26:277–86, 2015.

102. Nomura SJO, Dash C, Rosenberg L, Yu J, Palmer JR, Adams-Campbell LL. Is adherence to diet, physical activity, and body weight cancer prevention recommendations associated with colorectal cancer incidence in African American women? *Cancer Causes Control*. 27:869–79, 2016.

103. Romaguera D, Vergnaud AC, Peeters PH, van Gils CH, Chan DS, Ferrari P, Romieu I, Jenab M, Slimani N, Clavel-Chapelon F, Fagherazzi G, Perquier F, Kaaks R, Teucher B, Boeing H, von Rüsten A, Tjønneland A, Olsen A, Dahm CC, Overvad K, Quirós JR, Gonzalez CA, Sánchez MJ, Navarro C, Barricarte A, Dorronsoro M, Khaw KT, Wareham NJ, Crowe FL, Key TJ, Trichopoulou A, Lagiou P, Bamia C, Masala G, Vineis P, Tumino R, Sieri S, Panico S, May AM, Bueno-de-Mesquita HB, Büchner FL, Wirfält E, Manjer J, Johansson I, Hallmans G, Skeie G, Benjaminsen Borch K, Parr CL, Riboli E, Norat T. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. *Am J Clin Nutr*. 96:150–63, 2012.

104. Hastert TA, White E. Association between meeting the WCRF/AICR cancer prevention recommendations and colorectal cancer incidence: results from the VITAL cohort. *Cancer Causes Control*. 27:1347–1359, 2016.

105. Petimar J, Smith-Warner SA, Rosner B, Chan AT, Giovannucci EL, Tabung FK. Adherence to The World Cancer Research Fund/American Institute for Cancer Research 2018 Recommendations for Cancer Prevention and Risk of Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev*. 28:1469–1479, 2019.

106. Malcomson FC, Parra-Soto S, Ho FK, Lu L, Celis-Morales C, Sharp L, Mathers JC. Adherence to the 2018 World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR) Cancer Prevention Recommendations and risk of 14 lifestyle-related cancers in the UK Biobank prospective cohort study. *BMC Med*. 21:407, 2023.

107. Barrubés L, Babio N, Hernández-Alonso P, Toledo E, Ramírez Sabio JB, Estruch R, Ros E, Fitó M, Alonso-Gómez AM, Fiol M, Lapetra J, Serra-Majem L, Pintó X, Ruiz-Canela M, Corella D, Castañer O, Macías-González M, Salas-Salvadó J. Association between the 2018 WCRF/AICR and the Low-Risk Lifestyle Scores with Colorectal Cancer Risk in the Predimed Study. *J Clin Med*. 9:1215, 2020.

108. Thomson CA, McCullough ML, Wertheim BC, Chlebowski RT, Martinez ME, Stefanick ML, Rohan TE, Manson JE, Tindle HA, Ockene J, Vitolins MZ, Wactawski-Wende J, Sarto GE, Lane DS, Neuhouser ML. Nutrition and Physical Activity Cancer Prevention Guidelines, Cancer Risk, and Mortality in the Women's Health Initiative. *Cancer Prev Res (Phila)*. 7:42–53, 2014.

109. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjonneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. *BMJ*. 341:c5504–c5504, 2010.

110. Odegaard AO, Koh WP, Yuan JM. Combined Lifestyle Factors and Risk of Incident Colorectal Cancer in a Chinese Population. *Cancer Prev Res*. 6:360–367, 2013.

111. Dartois L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. Association between Five Lifestyle Habits and Cancer Risk: Results from the E3N Cohort. *Cancer Prev Res*. 7:516–525, 2014.

112. Zhang QL, Zhao LG, Li HL, Gao J, Yang G, Wang J, Zheng W, Shu XO, Xiang YB. The joint effects of major lifestyle factors on colorectal cancer risk among

Chinese men: A prospective cohort study. *Int J Cancer*. 142:1093–1101, 2018.

113. Baron JA, Beach M, Mandel JS, Stolk RU van, Haile RW, Sandler RS, Rothstein R, Summers RW, Snover DC, Beck GJ, Bond JH, Frankl H, Pearson L, Greenberg ER. Calcium Supplements for the Prevention of Colorectal Adenomas. *N Engl J Med*. 340:101–107, 1999.
114. Fedirko V, Bostick RM, Flanders WD, Long Q, Sidelnikov E, Shaukat A, Daniel CR, Rutherford RE, Woodard JJ. Effects of Vitamin D and Calcium on Proliferation and Differentiation In Normal Colon Mucosa: a Randomized Clinical Trial. *Cancer Epidemiol Biomarkers Prev*. 18:2933–41, 2009.
115. Cross AJ, Sinha R. Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environ Mol Mutagen*. 44:44–55, 2004.
116. Cross AJ, Pollock JRA, Bingham SA. Haem, not Protein or Inorganic Iron, Is Responsible for Endogenous Intestinal N-Nitrosation Arising from Red Meat. *Cances Res*. 63:2358 –2360, 2003.
117. Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat Rev Cancer*. 7:599–612, 2007.
118. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, Qin H. Obesity and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies. Gorlova OY, editor. *PLoS ONE*. 8:e53916, 2013.
119. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, De Vos WM, Brunak S, Doré J, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. *Nature*. 473:174–80, 2011.
120. Avuthu N, Guda C. Meta-Analysis of Altered Gut Microbiota Reveals Microbial and Metabolic Biomarkers for Colorectal Cancer. *Microbiol Spectr*.10:e00013-22, 2022.
121. Lynch SV, Pedersen O. The Human Intestinal Microbiome in Health and Disease. Phimister EG, editor. *N Engl J Med*. 375:2369–79, 2016.
122. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome and the immune system. *Nature*. 474:327–36, 2011.
123. Garrett WS. Cancer and the microbiota. *Cancer Immunol Immunother*. 348:80–86, 2015.
124. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A, Ponnudurai R, Sunagawa S, Coelho LP, Schrotz-King P, Vogtmann E, Habermann N, Niméus E, Thomas AM, Manghi P, Gandini S, Serrano D, Mizutani S, Shiroma H, Shiba S, Shibata T, Yachida S, Yamada T, Waldron L, Naccarati A, Segata N, Sinha R, Ulrich CM, Brenner H, Arumugam M, Bork P, Zeller G. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 25:679–89, 2019.
125. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, Chan FKL, Kristiansen K, Sung JJY, Wong SH, Yu J. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome*. 6:70, 2018.

126. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, Goedert JJ, Hayes RB, Yang L. Human Gut Microbiome and Risk for Colorectal Cancer. *J Natl Cancer Inst*. 105:1907–11, 2013.

127. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer. *PLoS ONE*. 7:e39743, 2012.

128. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S, Zhao L. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J*. 6(2):320–9, 2012.

129. White MT, Sears CL. The microbial landscape of colorectal cancer. *Nat Rev Microbiol*. 22:240–54, 2024.

130. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, Baselga J, Liu C, Shivdasani RA, Ogino S, Birren BW, Huttenhower C, Garrett WS, Meyerson M. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res*. 22:292–8, 2012.

131. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Watanabe H, Masuda K, Nishimoto Y, Kubo M, Hosoda F, Rokutan H, Matsumoto M, Takamaru H, Yamada M, Matsuda T, Iwasaki M, Yamaji T, Yachida T, Soga T, Kurokawa K, Toyoda A, Ogura Y, Hayashi T, Hatakeyama M, Nakagama H, Saito Y, Fukuda S, Shibata T, Yamada T. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med*. 25:968–76, 2019.

132. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, Su L, Li X, Li X, Li J, Xiao L, Huber-Schönauer U, Niederseer D, Xu X, Al-Aama JY, Yang H, Wang J, Kristiansen K, Arumugam M, Tilg H, Datz C, Wang J. Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun*. 6:6528, 2015.

133. Bultman SJ. Molecular Pathways: Gene–Environment Interactions Regulating Dietary Fiber Induction of Proliferation and Apoptosis via Butyrate for Cancer Prevention. *Clin Cancer Res*. 20:799–803, 2014.

134. Vinelli V, Biscotti P, Martini D, Del Bo' C, Marino M, Meroño T, Nikoloudaki O, Calabrese FM, Turroni S, Taverniti V, Unión Caballero A, Andrés-Lacueva C, Porrini M, Gobbetti M, De Angelis M, Brigidi P, Pinart M, Nimptsch K, Guglielmetti S, Riso P. Effects of Dietary Fibers on Short-Chain Fatty Acids and Gut Microbiota Composition in Healthy Adults: A Systematic Review. *Nutrients*. 14:2559, 2022.

135. So D, Whelan K, Rossi M, Morrison M, Holtmann G, Kelly JT, Shanahan ER, Staudacher HM, Campbell KL. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. *Am J Clin Nutr*. 107:965–83, 2018.

136. Aslam H, Marx W, Rocks T, Loughman A, Chandrasekaran V, Ruusunen A, Dawson SL, West M, Mullarkey E, Pasco JA, and Jacka FN. The effects of dairy and dairy derivatives on the gut microbiota: a systematic literature review. *Gut Microbes*. 12:1799533, 2020.

137. Seesaha PK, Chen X, Wu X, Xu H, Li C, Jheengut Y, Zhao F, Liu L, Zhang D. The interplay between dietary factors, gut microbiome and colorectal cancer: a new era of colorectal cancer prevention. *Future Oncol*. 16:293–306, 2020.

138. Maukonen M, Koponen KK, Havulinna AS, Kaartinen NE, Niiranen T, Méric G, Pajari AM, Knight R, Salomaa V, Männistö S. Associations of plant-based

foods, red and processed meat, and dairy with gut microbiome in Finnish adults. *Eur J Nutr*. 63:2247–2260, 2024.

139. Veiga P, Pons N, Agrawal A, Oozeer R, Guyonnet D, Brazeilles R, Faurie JM, Van Hylckama Vlieg JET, Houghton LA, Whorwell PJ, Ehrlich SD, Kennedy SP. Changes of the human gut microbiome induced by a fermented milk product. *Sci Rep*. 4:6328, 2014.
140. Lin H, Yu Y, Zhu L, Lai N, Zhang L, Guo Y, Lin X, Yang D, Ren N, Zhu Z, Dong Q. Implications of hydrogen sulfide in colorectal cancer: Mechanistic insights and diagnostic and therapeutic strategies. *Redox Biol*. 59:102601, 2023.
141. Venema K. Intestinal fermentation of lactose and prebiotic lactose derivatives, including human milk oligosaccharides. *Int Dairy J*. 22:123–40, 2012.
142. Song M, Garrett WS, Chan AT. Nutrients, Foods, and Colorectal Cancer Prevention. *Gastroenterology*. 148:1244-1260.e16, 2015.
143. Pérez-Prieto I, Plaza-Florido A, Ubago-Guisado E, Ortega FB, Altmäe S. Physical activity, sedentary behavior and microbiome: A systematic review and meta-analysis. *J Sci Med Sport*. 27:793–804, 2024.
144. Song M, Chan AT. Environmental Factors, Gut Microbiota, and Colorectal Cancer Prevention. *Clin Gastroenterol Hepatol*. 17:275–89, 2019.
145. Jalandra R, Dalal N, Yadav AK, Verma D, Sharma M, Singh R, Khosla A, Kumar A, Solanki PR. Emerging role of trimethylamine-N-oxide (TMAO) in colorectal cancer. *Appl Microbiol Biotechnol*. 105:7651–60, 2021.
146. Schiattarella GG, Sannino A, Toscano E, Giugliano G, Gargiulo G, Franzone A, Trimarco B, Esposito G, Perrino C. Gut microbe-generated metabolite trimethylamine-N-oxide as cardiovascular risk biomarker: a systematic review and dose-response meta-analysis. *Eur Heart J*. 38:2948–56, 2017.
147. Bae S, Ulrich CM, Neuhouser ML, Malysheva O, Bailey LB, Xiao L, Brown EC, Cushing-Haugen KL, Zheng Y, Cheng TYD, Miller JW, Green R, Lane DS, Beresford SAA, Caudill MA. Plasma Choline Metabolites and Colorectal Cancer Risk in the Women's Health Initiative Observational Study. *Cancer Res*. 74:7442–52, 2014.
148. Byrd DA, Zouiouich S, Karwa S, Li XS, Wang Z, Sampson JN, Loftfield E, Huang WY, Hazen SL, Sinha R. Associations of serum trimethylamine N-oxide and its precursors with colorectal cancer risk in the Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial Cohort. *Cancer*. 130:1982–90, 2024.
149. Mutlu EA, Gillevet PM, Rangwala H, Sikaroodi M, Naqvi A, Engen PA, Kwasny M, Lau CK, Keshavarzian A. Colonic microbiome is altered in alcoholism. *Am J Physiol Gastrointest Liver Physiol*. 302:G966–978, 2012.
150. Tsuruya A, Kuwahara A, Saito Y, Yamaguchi H, Tsubo T, Suga S, Inai M, Aoki Y, Takahashi S, Tsutsumi E, Suwa Y, Morita H, Kinoshita K, Totsuka Y, Suda W, Oshima K, Hattori M, Mizukami T, Yokoyama A, Shimoyama T, Nakayama T. Ecophysiological consequences of alcoholism on human gut microbiota: implications for ethanol-related pathogenesis of colon cancer. *Sci Rep*. 6:27923, 2016.
151. Nosova T, Jokelainen K, Kaihovaara P, Jousimies-Somer H, Siitonen A, Heine R, Salaspuro M. Aldehyde dehydrogenase activity and acetate production by aerobic bacteria representing the normal flora of the human large intestine. *Alcohol Alcohol*. 31:555–64, 1996.

152. Tsuruya A, Kuwahara A, Saito Y, Yamaguchi H, Tenma N, Inai M, Takahashi S, Tsutsumi E, Suwa Y, Totsuka Y, Suda W, Oshima K, Hattori M, Mizukami T, Yokoyama A, Shimoyama T, Nakayama T. Major Anaerobic Bacteria Responsible for the Production of Carcinogenic Acetaldehyde from Ethanol in the Colon and Rectum. *Alcohol Alcohol*. 51:395–401, 2016.

153. Pinart M, Dötsch A, Schlicht K, Laudes M, Bouwman J, Forslund SK, Pischon T, Nimptsch K. Gut Microbiome Composition in Obese and Non-Obese Persons: A Systematic Review and Meta-Analysis. *Nutrients*. 14:12, 2021.

154. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. *Nature*. 457:480–484, 2009.

155. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Human gut microbes associated with obesity. *Nature*. 444:1022–1023, 2006.

156. Ozato N, Saito S, Yamaguchi T, Katashima M, Tokuda I, Sawada K, Katsuragi Y, Kakuta M, Imoto S, Ihara K, Nakaji S. Blautia genus associated with visceral fat accumulation in adults 20–76 years of age. *NPJ Biofilms Microbiomes*. 5:28, 2019.

157. Cani PD, Jordan BF. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. *Nat Rev Gastroenterol Hepatol*. 15:671–82, 2018.

158. Li R, Grimm SA, Chrysovergis K, Kosak J, Wang X, Du Y, Burkholder A, Janardhan K, Mav D, Shah R, Eling TE, Wade PA. Obesity, Rather Than Diet, Drives Epigenomic Alterations in Colonic Epithelium Resembling Cancer Progression. *Cell Metabolism*. 19:702–11, 2014.

159. The ATBC Cancer Prevention Study Group. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. *Ann Epidemiol*. 4:1–10, 1994.

160. Heistaro S (editor). Methodology report: Health 2000 Survey. National Public Health Institute: Helsinki, Finland; 2008.

161. Ylihärsilä H, Kajantie E, Osmond C, Forsén T, Barker DJ, Eriksson JG. Body mass index during childhood and adult body composition in men and women aged 56–70 y. *Am J Clin Nutr*. 87:1769–1775, 2008.

162. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, Kuulasmaa K, Laatikainen T, Männistö S, Peltonen M, Perola M, Puska P, Salomaa V, Sundvall J, Virtanen SM, Vartiainen E. Cohort Profile: The National FINRISK Study. *Int J Epidemiol*. 47:696–696i, 2018.

163. Männistö S, Ovaskainen ML, Valsta L (editors). The national Findiet 2002 study. National Public Health Institute: Helsinki, Finland; 2003.

164. Konttinen H, Silventoinen K, Sarlio-Lähteenkorva S, Männistö S, Haukkala A. Emotional eating and physical activity self-efficacy as pathways in the association between depressive symptoms and adiposity indicators. *Am J Clin Nutr*. 92:1031–9, 2010.

165. Borodulin K, Sääksjärvi K. FinHealth 2017 Study - Methods. Finnish Institute for Health and Welfare: Helsinki, Finland; 2019.

166. Männistö S, Virtanen M, Mikkonen T, Pietinen P. Reproducibility and validity of a food frequency questionnaire in a case-control study on breast cancer. *J Clin Epidemiol*. 49:401–9, 1996.

167. Paalanen L, Männistö S, Virtanen MJ, Knekt P, Räsänen L, Montonen J, Pietinen P. Validity of a food frequency questionnaire varied by age and body mass index. *J Clin Epidemiol*. 59:994–1001, 2006.

168. Kaartinen NE, Tapanainen H, Valsta LM, Similä ME, Reinivuo H, Korhonen T, Harald K, Eriksson JG, Peltonen M, Männistö S. Relative validity of a FFQ in measuring carbohydrate fractions, dietary glycaemic index and load: exploring the effects of subject characteristics. *Br J Nutr*. 107:1367–75.

169. Willett W. Food Frequency Methods. In: Willet W. Nutritional Epidemiology, 3rd edition. Oxford University Press: Oxford, UK; 2013. p. 70–95.

170. Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. *Am J Epidemiol*. 128:655–66, 1988.

171. Ovaskainen ML, Paturi M, Reinivuo H, Hannila ML, Sinkko H, Lehtisalo J, Pynnönen-Polari O, Männistö S. Accuracy in the estimation of food servings against the portions in food photographs. *Eur J Clin Nutr*. 62:674–81, 2008.

172. Anttolainen M (editor). The National FINDIET 1997 Survey. National Public Health Institute: Helsinki, Finland; 1998.

173. Reinivuo H, Hirvonen T, Ovaskainen ML, Korhonen T, Valsta LM. Dietary survey methodology of FINDIET 2007 with a risk assessment perspective. *Public Health Nutr*. 13:915–9, 2010.

174. Helldán A, Raulio S, Kosola M, Tapanainen H, Ovaskainen ML, Virtanen S (editors). The National FINDIET 2012 Survey. National Institute for Health and Welfare: Helsinki, Finland; 2013.

175. Valsta L, Kaartinen N, Tapanainen H, Männistö S, Sääksjärvi K (editors). Nutrition in Finland – The National FinDiet 2017 Survey. Finnish Institute for Health and Welfare: Helsinki, Finland; 2018.

176. Baranowski T. 24-Hour Recall and Diet Record Methods. In: Willet W. Nutritional Epidemiology, 3rd edition. Oxford University Press: Oxford, UK; 2013. p. 49–69.

177. Kanerva N, Kaartinen NE, Schwab U, Lahti-Koski M, Männistö S. The Baltic Sea Diet Score: a tool for assessing healthy eating in Nordic countries. *Public Health Nutr*. 17:1697–705, 2014.

178. Nordic Council of Ministers. Nordic Nutrition Recommendations 2012. Integrating Nutrition and Physical Activity, 5th ed. Norden: Copenhagen, Denmark; 2014.

179. Borodulin K, Laatikainen T, Juolevi A, Jousilahti P. Thirty-year trends of physical activity in relation to age, calendar time and birth cohort in Finnish adults. *Eur J Public Health*. 18:339–44, 2008.

180. Tolonen H. EHES Manual. Part B. Field Work and Procedures, 2nd ed. Finnish Institute of Health and Welfare: Helsinki, Finland; 2016.

181. Eriksson M, Räikkönen K, Eriksson JG. Early life stress and later health outcomes—findings from the Helsinki Birth Cohort Study. *Am J Hum Biol*. 26:111–6, 2014..

182. World Health Organization. Obesity: preventing and managing the global epidemic. World Health Organization Technical Reports Series no. 894. World Health Organization: Geneva, Switzerland; 2000.

183. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 18:499–502, 1972.

184. Salosensaari A, Laitinen V, Havulinna AS, Meric G, Cheng S, Perola M, Valsta L, Alfthan G, Inouye M, Watrous JD, Long T, Salido RA, Sanders K, Brennan C, Humphrey GC, Sanders JG, Jain M, Jousilahti P, Salomaa V, Knight R, Lahti L, Niiranen T. Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nat Commun*. 12:2671, 2021.

185. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems*. 3:e00069-18, 2018.

186. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34:i884–90, 2018.

187. Guccione C, Patel L, Tomofuji Y, McDonald D, Gonzalez A, Sepich-Poore GD, Sonehara K, Zakeri M, Chen Y, Dilmore AH, Damle N, Baranzini SE, Hightower G, Nakatsuji T, Gallo RL, Langmead B, Okada Y, Curtius K, Knight R. Incomplete human reference genomes can drive false sex biases and expose patient-identifying information in metagenomic data. *Nat Commun*. 16:825, 2025.

188. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 10:giab008, 2021.

189. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, Morton JT, Nicolaou G, Parks DH, Karst SM, Albertsen M, Hugenholtz P, DeSantis T, Song SJ, Bartko A, Havulinna AS, Jousilahti P, Cheng S, Inouye M, Niiranen T, Jain M, Salomaa V, Lahti L, Mirarab S, Knight R. Greengenes2 unifies microbial data in a single reference tree. *Nat Biotechnol*. 42:715–8, 2024.

190. Shannon CE. A mathematical theory of communication. *Bell Syst Tec J*. 27:379–423, 1948.

191. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Appl Environ Microbiol*. 73:1576–85, 2007.

192. Shams-White MM, Brockton NT, Mitrou P, Romaguera D, Brown S, Bender A, Kahle LL, Reedy J. Operationalizing the 2018 World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) Cancer Prevention Recommendations: A Standardized Scoring System. *Nutrients*. 11:1572, 2019.

193. Aune D, Chan DSM, Lau R, Vieira R, Greenwood DC, Kampman E, Norat T. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *BMJ*. 343:d6617, 2011.

194. World Health Organization. Waist circumference and waist-hip ratio : report of a WHO expert consultation. World Health Organization: Geneva, Switzerland; 2008.

195. Leinonen MK, Miettinen J, Heikkinen S, Pitkäniemi J, Malila N. Quality measures of the population-based Finnish Cancer Registry indicate sound data quality for solid malignant tumours. *Eur J Cancer*. 77:31–9, 2017.

196. World Health Organization. Energy and Protein Requirements. Report of a Joint WHO/FAO/UNU Expert Consultation. WHO Technical Report Series no. 724. World Health Organization: Geneva, Switzerland; 1985.

197. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr*. 45:569–81, 1991.

198. Willett W, Stampfer MJ. Total energy intake: Implications for epidemiologic analyses. *Am J Epidemiol*. 124:17–27, 1986.

199. Song M, Giovannucci E. Substitution analysis in nutritional epidemiology: proceed with caution. *Eur J Epidemiol*. 33:137–40, 2018.

200. Würtz AML, Hansen MD, Tjønneland A, Rimm EB, Schmidt EB, Overvad K, Jakobsen MU. Substitutions of red meat, poultry and fish and risk of myocardial infarction. *Br J Nutr*. 115:1571–8, 2016.

201. Smith-Warner SA, Spiegelman D, Ritz J, Albanes D, Beeson WL, Bernstein L, Berrino F, van den Brandt PA, Buring JE, Cho E, Colditz GA, Folsom AR, Freudenheim JL, Giovannucci E, Goldbohm RA, Graham S, Harnack L, Horn-Ross PL, Krogh V, Leitzmann MF, McCullough ML, Miller AB, Rodriguez C, Rohan TE, Schatzkin A, Shore R, Virtanen M, Willett WC, Wolk A, Zeleniuch-Jacquotte A, Zhang SM, Hunter DJ. Methods for Pooling Results of Epidemiologic Studies. *Am J Epidemiol*. 163:1053–64, 2006.

202. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol*. 26:32–46, 2001.

203. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 57:289–300, 1995.

204. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nat Methods*. 21:83–91, 2024.

205. White D, Gramacy R. Mapping pruning, and graphing tree models. 2022. Available: https://cran.r-project.org/web/packages/maptree/maptree.pdf (Accessed 24 Jun 2025)

206. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of Dietary Assessment Methods When Compared to the Method of Doubly Labeled Water: A Systematic Review in Adults. *Front Endocrinol*. 10:850, 2019.

207. Gemming L, Ni Mhurchu C. Dietary under-reporting: what foods and which meals are typically under-reported? *Eur J Clin Nutr*. 70:640–1, 2016.

208. Mann KD, Pearce MS, McKevith B, Thielecke F, Seal CJ. Low whole grain intake in the UK: results from the National Diet and Nutrition Survey rolling programme 2008–11. *Br J Nutr*. 113:1643–51, 2015.

209. Kissock KR, Warensjö Lemming E, Axelsson C, Neale EP, Beck EJ. Defining whole-grain foods – does it change estimations of intakes and associations with CVD risk factors: an Australian and Swedish perspective. *Br J Nutr*. 126:1725–36, 2021.

210. Kissock KR, Neale EP, Beck EJ. The relevance of whole grain food definitions in estimation of whole grain intake: a secondary analysis of the National Nutrition and Physical Activity Survey 2011–2012. *Public Health Nutr*. 23:1307–19, 2020.

211. Kranz S, Dodd K, Juan W, Johnson L, Jahns L. Whole Grains Contribute Only a Small Proportion of Dietary Fiber to the U.S. Diet. *Nutrients*. 9:153, 2017.

212. O'Neil CE, Nicklas TA, Zanovec M, Cho S. Whole-Grain Consumption Is Associated with Diet Quality and Nutrient Intake in Adults: The National

Health and Nutrition Examination Survey, 1999-2004. *J Am Diet Assoc*. 110:1461–8, 2010.

213. Valsta LM, Tapanainen H, Kortetmäki T, Sares-Jäske L, Paalanen L, Kaartinen NE, Haario P, Kaljonen M. Disparities in Nutritional Adequacy of Diets between Different Socioeconomic Groups of Finnish Adults. *Nutrients*. 14:1347, 2022.
214. Hollænder PL, Ross AB, Kristensen M. Whole-grain and blood lipid changes in apparently healthy adults: a systematic review and meta-analysis of randomized controlled studies. *Am J Clin Nutr*. 102:556–72, 2015.
215. Marshall S, Petocz P, Duve E, Abbott K, Cassettari T, Blumfield M, Fayet-Moore F. The Effect of Replacing Refined Grains with Whole Grains on Cardiovascular Risk Factors: A Systematic Review and Meta-Analysis of Randomized Controlled Trials with GRADE Clinical Recommendation. *J Acad Nutr Diet*. 120:1859–83.e31.
216. Whitehead A, Beck EJ, Tosh S, Wolever TM. Cholesterol-lowering effects of oat *β*-glucan: a meta-analysis of randomized controlled trials. *Am J Clin Nutr*. 100:1413–21, 2014.
217. Jonsson K, Andersson R, Bach Knudsen KE, Hallmans G, Hanhineva K, Katina K, Kolehmainen M, Kyrø C, Langton M, Nordlund E, Lærke HN, Olsen A, Poutanen K, Tjønneland A, Landberg R. Rye and health - Where do we stand and where do we go? *Trends Food Sci Technol*. 79:78–87, 2018.
218. Abumweis SuhadS, Barake R, Jones P. Plant sterols/stanols as cholesterol lowering agents: a meta-analysis of randomized controlled trials. *Food Nutr Res*. 52:1811, 2008.
219. McRae MP. Health Benefits of Dietary Whole Grains: An Umbrella Review of Meta-analyses. *J Chiropr Med*. 16:10–8, 2017.
220. Simojoki M, Kaartinen NE, Maukonen M, Harald K, Tapanainen H, Albanes D, Eriksson JG, Jousilahti P, Koskinen S, Pajari AM, Männistö S. Partial substitution of red or processed meat with plant-based foods and the risk of cardiovascular disease. *Eur J Epidemiol*. 40:517–25, 2025.
221. Maukonen M, Harald K, Kaartinen NE, Tapanainen H, Albanes D, Eriksson J, Härkänen T, Jousilahti P, Koskinen S, Päivärinta E, Suikki T, Tolonen H, Pajari AM, Männistö S. Partial substitution of red or processed meat with plant-based foods and the risk of type 2 diabetes. *Sci Rep*. 13:5874, 2023.
222. Knudsen AK, Allebeck P, Tollånes MC, Skogen JC, Iburg KM, McGrath JJ, Juel K, Agardh EE, Ärnlöv J, Bjørge T, Carrero JJ, Cederroth CR, Eggen AE, El-Khatib Z, Ellingsen CL, Fereshtehnejad SM, Gissler M, Hadkhale K, Havmoeller R, Johansson L, Juliusson PB, Kiadaliri AA, Kisa S, Kisa A, Lallukka T, Mekonnen T, Meretoja TJ, Meretoja A, Naghavi M, Neupane S, Nguyen TT, Petzold M, Plana-Ripoll O, Shiri R, Sigurvinsdottir R, Skirbekk V, Skou ST, Sigfusdottir ID, Steiner TJ, Sulo G, Truelsen TC, Vasankari TJ, Weiderpass E, Vollset SE, Vos T, Øverland S. Life expectancy and disease burden in the Nordic countries: results from the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. *Lancet Public Health*. 4:e658–69, 2019.
223. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun*. 11:5206, 2020.
224. Zhang Y, Chen H, Lu M, Cai J, Lu B, Luo C, Dai M. Habitual Diet Pattern Associations with Gut Microbiome Diversity and Composition: Results from a Chinese Adult Cohort. *Nutrients*. 14:2639, 2022.

225. Noh H, Jang HH, Kim G, Zouiouich S, Cho SY, Kim HJ, Kim J, Choe JS, Gunter MJ, Ferrari P, Scalbert A, Freisling H. Taxonomic Composition and Diversity of the Gut Microbiota in Relation to Habitual Dietary Intake in Korean Adults. *Nutrients*. 13:366, 2021.

226. Partula V, Mondot S, Torres MJ, Kesse-Guyot E, Deschasaux M, Assmann K, Latino-Martel P, Buscail C, Julia C, Galan P, Hercberg S, Rouilly V, Thomas S, Quintana-Murci L, Albert ML, Duffy D, Lantz O, Touvier M. Associations between usual diet and gut microbiota composition: results from the Milieu Intérieur cross-sectional study. *Am J Clin Nutr*. 109:1472–83, 2019.

227. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, LifeLines cohort study, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*. 352:565–9, 2016.

228. Abdugheni R, Wang W, Wang Y, Du M, Liu F, Zhou N, Jiang C, Wang C, Wu L, Ma J, Liu C, Liu S. Metabolite profiling of human-originated Lachnospiraceae at the strain level. *iMeta*. 1:e58, 2022.

229. Kim DJ, Yang J, Seo H, Lee WH, Ho Lee D, Kym S, Park YS, Kim JG, Jang IJ, Kim YK, Cho JY. Colorectal cancer diagnostic model utilizing metagenomic and metabolomic data of stool microbial extracellular vesicles. *Sci Rep*. 10:2860, 2020.

230. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, Beghini F, Manara S, Karcher N, Pozzi C, Gandini S, Serrano D, Tarallo S, Francavilla A, Gallo G, Trompetto M, Ferrero G, Mizutani S, Shiroma H, Shiba S, Shibata T, Yachida S, Yamada T, Wirbel J, Schrotz-King P, Ulrich CM, Brenner H, Arumugam M, Bork P, Zeller G, Cordero F, Dias-Neto E, Setubal JC, Tett A, Pardini B, Rescigno M, Waldron L, Naccarati A, Segata N. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med*. 25:667–78, 2019.

231. Tolonen H, Laatikainen T, Helakorpi S, Talala K, Martelin T, Prättälä R. Marital status, educational level and household income explain part of the excess mortality of survey non-respondents. *Eur J Epidemiol*. 25:69–76, 2010.

232. Jousilahti P. Total and cause specific mortality among participants and non-participants of population based health surveys: a comprehensive follow up of 54 372 Finnish men and women. *J Epidemiol Community Health*. 59:310–5, 2005.

233. Koponen KK, Salosensaari A, Ruuskanen MO, Havulinna AS, Männistö S, Jousilahti P, Palmu J, Salido R, Sanders K, Brennan C, Humphrey GC, Sanders JG, Meric G, Cheng S, Inouye M, Jain M, Niiranen TJ, Valsta LM, Knight R, Salomaa VV. Associations of healthy food choices with gut microbiota profiles. *Am J Clin Nutr*. 00:1–12, 2021.

234. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis with R: A Hands-On Guide. Chapmann & Hall/CRC Press: Boca Raton, FL, USA and London; 2021.