

Indicators of energy balance and risk of epigenetic instability in colorectal cancer

Citation for published version (APA):

Hughes, L. (2011). *Indicators of energy balance and risk of epigenetic instability in colorectal cancer*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.20110519lh>

Document status and date:

Published: 01/01/2011

DOI:

[10.26481/dis.20110519lh](https://doi.org/10.26481/dis.20110519lh)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Indicators of energy balance and risk of epigenetic instability in colorectal cancer

Indicators of energy balance and risk of epigenetic instability in colorectal cancer
Laura Hughes

ISBN 978-90-9026124-9

Cover picture: © WarrenGoldswain | Dreamstime.com

Printed by: Ipskamp Drukkers, Enschede, the Netherlands

© 2011, Laura Hughes

All rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage or retrieval system, without permission in writing from the author, or, when appropriate, from the publishers of the publications.

Indicators of energy balance and risk of epigenetic instability in colorectal cancer

DISSERTATION

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. G.P.M.F. Mols in
accordance with the decision of the Board of Deans, to be defended in public
on
Thursday, 19th of May 2011 at 14:00 hours

by

Laura Anne Elizabeth Hughes

Born 3 June 1982, Vancouver, Canada

Supervisor

Prof. dr. ir. P.A. van den Brandt

Co-supervisors

Dr. ir. M.P. Weijenberg

Dr. M. van Engeland

Assessment Committee

Prof. dr. F.C.S. Ramaekers (Chair)

Prof. dr. P.H.M. Peeters, Julia Centrum, University Medical Centre Utrecht

Prof. dr. M.A. van Baak

Dr. B. Heijmans, Leiden University Medical Centre

Dr. K. Meijer

“Cancer begins and ends with people”

June Goodfield

For my Dad

Table of Contents

Chapter 1 Introduction	9
Chapter 2 Self-reported clothing size as a proxy measure for body size	35
Chapter 3 Childhood and adolescent energy restriction and subsequent colorectal cancer risk: results from the Netherlands Cohort Study	51
Chapter 4 Early life exposures to famine and colorectal cancer risk: a role for epigenetic mechanisms?	75
Chapter 5 Body size and colorectal cancer risk after 16.3 years follow-up: an analysis from the Netherlands Cohort Study	95
Chapter 6 Physical activity, sitting time and colorectal cancer risk in the Netherlands Cohort Study: Regular long-term physical activity may matter	124
Chapter 7 Body size, physical activity and risk of colorectal cancer with or without the CpG island methylator phenotype (CIMP): an analysis from the Netherlands Cohort Study	153
Chapter 8 Body size and risk for colorectal cancers showing BRAF mutations or microsatellite instability: a pooled analysis of the Netherlands Cohort Study and Melbourne Collaborative Cohort Study	177
Chapter 9 General discussion	203

English summary	231
Nederlandse samenvatting (summary in Dutch)	237
Acknowledgements	243
Curriculum Vitae	247
List of publications	251

1

Introduction

Globally, more than 1 billion adults are overweight, and 300 million of these are clinically obese.¹ Perhaps even more alarming is that 22 million children under the age of five years are thought to be overweight, and this number is increasing.¹ It is also clear that these conditions are no longer restricted to industrialised nations; the World Health Organization (WHO) reports that the observed increase in overweight and obese individuals is occurring most rapidly in developing countries.¹

That overweight and obesity have reached pandemic proportions² is not surprising. Over the past half-century, there has been a shift towards diets that are high in fat and refined sugar, and low in fibre. Simultaneously, average daily physical activity rates have dropped as a consequence of improvements in technology, automated transport, and fewer laborious demands in the home and work place.¹ In essence, overweight is a reflection of energy imbalance.³

Next to the increase in overweight and obesity, the incidence of colorectal cancer (CRC) has also been on the rise.⁴ In the Netherlands, for example, the age standardised incident rate for colon cancer rose from 31.4 to 41.1 per 100,000 person years during the period of 1989-2007, and for rectal cancer, from 15.5 to 21.4. During the same period, the age standardised incidence rate for colon cancer in women increased from 26.7 to 31.3 per 100,000 person years, and for rectal cancer from 8.6 to 12.1.⁷ Similar increases have been observed in other Western countries.^{8, 9} Furthermore, according to recent global estimates, CRC is the third most common cancer in men, and the second most common cancer in women.⁴ Approximately 608,000 deaths from CRC occur each year, making it the fourth most common cause of death from cancer in 2008.⁴ Within western Europe, these rates worsen: CRC is the second major incident cancer, and the second major cause of cancer death.¹⁰ Migration studies show that environmental lifestyle factors are associated with an increased risk of CRC, especially with respect to populations who have transitioned to a 'Western lifestyle'.¹²⁻¹⁴ It is therefore timely to ask how, and to what extent, are indicators of energy balance linked to the risk of developing CRC.

It is now well established that CRC is not a single disease. Rather, CRC encompasses a heterogeneous complex of diseases characterized by numerous genetic and epigenetic abnormalities.¹⁵ By distinguishing underlying pathways responsible for these differences, the exposure factors

that are thought to cause CRC and differential risk between individuals and populations may be better understood. Furthermore, establishing whether lifestyle factors can prevent the occurrence of a specific molecular change in CRC would add considerable scientific basis to prevention strategies.¹⁵

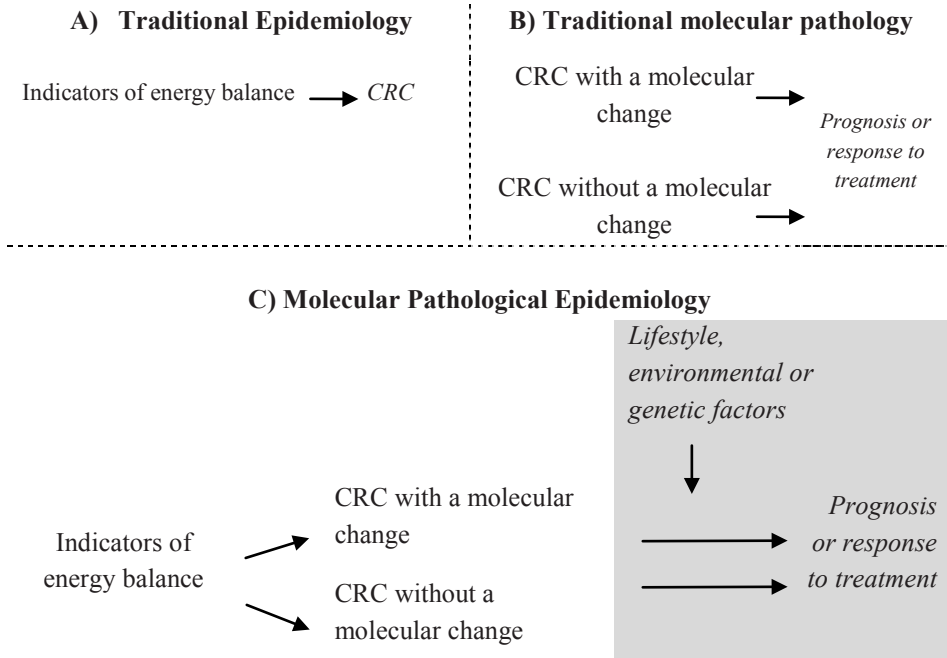


Figure 1:

Molecular pathological epidemiology (figure adapted from Ogino *et al.*¹⁶). The shaded section of C) indicates an aspect of molecular pathological epidemiology that was not considered in the present thesis.

Traditionally, epidemiological research has been used to investigate how an exposure such as overweight may increase or decrease the risk of developing cancer, and pathological research has been used to explore molecular characteristics of tumours to predict prognosis and response to treatment. By combining these two disciplines, a new field of scientific investigation has emerged: molecular pathological epidemiology.^{15, 16} With energy balance as the risk factor, **Figure 1** illustrates the concepts behind a) traditional epidemiology, b) traditional molecular pathology and c) how

fusing the two disciplines creates a molecular pathological epidemiology approach.^{15, 16} As indicated in this figure, molecular pathological epidemiology may also provide insights with respect to prognosis or response to treatment, however, this aspect was not investigated in the current thesis.

This thesis utilizes molecular pathological epidemiology to investigate how indicators of energy balance influence the risk of CRC, and the risk of CRC specifically characterised by molecular abnormalities associated with epigenetic instability, one of the known pathways to CRC development. Prior to delving into this research, it is first important to introduce what is meant by indicators of energy balance, CRC, and why epigenetic instability is of such interest with respect to these exposures.

INDICATORS OF ENERGY BALANCE

Multiple lifestyle factors play a role in the development of overweight and obesity. To maintain a stable energy balance and maintain one's body weight, the calories consumed (from foods, drinks and alcohol) must be balanced by the calories used (in normal body functions and basal metabolic rate, daily activities, and exercise). If one consumes more calories than the body uses, then weight gain will occur as the body stores the excess calories as fat.³ In this thesis, we consider various indicators of energy balance as risk factors for CRC.

Body Size

Body Mass Index (BMI)

In population based studies, it is important that measurements of body fat correlate well with actual body fat in both men and women, and across all ages and ethnic groups.¹⁷ Individuals with different heights or body builds may have similar fat masses, but very different proportions of total body fat.¹⁷ To account for such issues, it is common for epidemiological studies use body mass index (BMI) as a marker of body fatness. BMI is a measure of weight adjusted for height, and is calculated as: weight in kilograms divided by height squared (kg/m^2).²

There are methodological advantages of using BMI in epidemiological research. First, it is not unusual for studies to enrol thousands of participants. Body weight and height are easy to collect by self-report, and have been shown to correlate well with technician measured

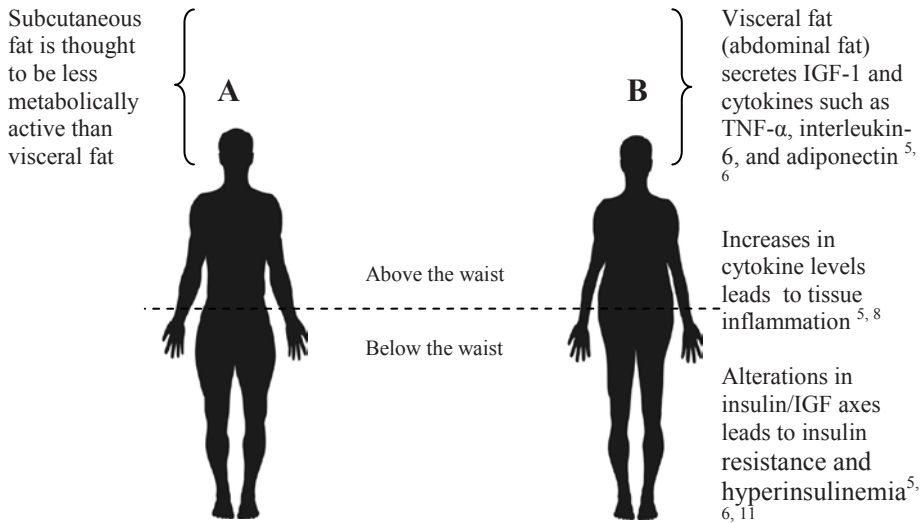
values.^{18, 19} BMI can be subsequently calculated by the researchers and used as an indicator of exposure. Second, the WHO has created BMI cut-off points based on the risks associated with being underweight, overweight or obese. These ranges make it relatively easy to compare adiposity within and between populations.¹⁷

Although BMI is reliably linked to body fatness,²⁰ its use is not without limitations. Very muscular or lean people can have a relatively high BMI, even if they have relatively little body fat.² Moreover, there is evidence suggesting that fat carried around organs in the abdominal region of the body is more dangerous than subcutaneous fat, in terms of disease risk;^{5, 6, 11} an individual may have a normal BMI, but be carrying fat in the abdominal region. In this instance, an indicator reflective of central obesity is more informative for predicting risk.

Waist circumference

Body fat can be more or less metabolically active, depending on where it is stored in the body (**Figure 2**). This may explain why not all clinically overweight or obese individuals develop diseases associated with such conditions, such as type 2 diabetes and cardiovascular disease.¹¹ Intra-abdominal fat, or visceral fat, is considered to be especially dangerous, because it is known to influence several important hormones involved in inflammation and insulin responses.^{2, 21, 22} It has been shown that abdominal fat can predict risk of metabolic and cardiovascular diseases better than BMI or other indicators of body fat.^{2, 23} Research also suggests that indicators of central obesity predict risk of CRC independently of BMI.^{24, 25}

In epidemiological studies, the most common proxy measure for abdominal fat is waist circumference. Usually, this is a direct measure that is reported in centimetres (cm), and obtained either by self-report, or measured by a study technician. In this thesis, we add to the body of evidence suggesting that lower clothing size, e.g. trousers or skirt, can also be used as a proxy measure for central adiposity.²⁶

**Figure 2:**

The distribution of body fat is important for disease risk. Fat distributed below the waist (A) is primarily subcutaneous and less metabolically active than fat distributed at or above the waist (B).

Adult attained height

Depending on the nutritional contrasts in a population, up to 90% of the variation in height is determined by a combination of multiple inherited factors.²⁷ Genome wide association studies (GWAS) are underway to identify genes associated with stature, so that we may gain a deeper understanding of human growth and how it influences later disease risk.^{27, 28} Independent of its use in calculating BMI, height is informative as an indicator of energy balance because it is a reflection of environmental and nutritional exposures that occurred *in utero* and in early childhood.²⁹ In this way it is also linked to birth weight, rate of growth, and age of puberty.² Additionally, taller individuals have a longer colorectal tract and undergo more cell divisions stimulated by IGF-1 and growth hormones, creating more potential for error during DNA replication.²

Populations become collectively taller as undernutrition decreases and food supplies become more secure.² With many developing and transitioning countries now facing such a scenario, investigating height as a

risk factor for CRC may offer informative insights about what such populations may face in the future, as well as potential mechanisms that may be at play in CRC development.

Energy intake and expenditure

Caloric intake

The physiological cause of weight gain is consuming more energy from food and drinks than is used by the body.² The unit of measure for dietary energy is the kilocalorie (kcal), and this can vary widely between foods and individual requirements. As one may imagine, accurately collecting such information in large studies is a challenge. Many studies utilise self-administered food-frequency questionnaires (FFQ) to estimate the energy associated with food consumption and dietary patterns in their population. However, this is prone to problems. A large part of energy intake is determined by energy requirements, which in turn is determined by body size, and further by physical activity. Furthermore, underreporting of food intake, especially for overweight and obese individuals, is a major limitation of self-reported food intake.^{30, 31} As a consequence, the variation in energy intake observed in a population may not actually be reflective of the number of calories consumed. Although the FFQ utilized in this thesis has been validated,³² we do not consider total energy intake as an independent risk factor in this thesis or as an interaction term with body fat and/or physical activity for the reasons just described.

With respect to cancer risk, it is informative to study both positive and negative energy balance. Energy restriction appears to be one of the most convincing interventions for lowering cancer risk in laboratory animals, through mechanisms such as inhibiting cell proliferation, accelerating apoptosis, reducing IGF-1, enhancing DNA repair, and lowering sex steroid hormones.³³⁻³⁹ Interestingly, energy restriction early in life appears to be especially protective against this risk.^{40, 41} When investigating risk in humans, epidemiological researchers often depend on historical data, for example exposure to a period of known energy restriction, such as a famine or war, as a proxy measure for energy restriction earlier in life.

Physical activity

Physical activity is a major aspect of energy expenditure, and is defined as any movement that uses skeletal muscles.² Currently, there is no standard classification of physical activity level, however, for conceptualisation, the World Cancer Research Fund (WCRF) considers physical activity according to four categories: occupational physical activity (e.g. manual labour), household physical activity (e.g. household chores), transport (e.g. walking/biking to work), and recreational physical activity (e.g. sports participation and leisure pursuits). Within these categories, intensity and duration can also be distinguished. Generally, recreational activity accounts for the greatest proportion of exercise in developed countries, however, overall physical activity levels are lower in industrialised countries than developing countries.^{2, 42}

As with energy intake, measuring physical activity is complex. It is especially difficult to measure in large observational studies because these are not equipped or designed to record the frequency, duration and intensity of everyone's physical activity over an extended period. As a result, self-administered questionnaires about jobs, activities in daily life and history of sport participation are most often used.² However, an accurate measurement of lifetime patterns of physical activity is still difficult to achieve. In terms of cancer prevention, physical activity may act directly by improving immune function, reducing free radicals, increasing prostaglandins, increasing gut motility, and increasing insulin sensitivity.⁴³⁻⁴⁶ Indirectly, it protects against overweight and obesity.

EPIDEMIOLOGICAL EVIDENCE

In 2007, the World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AIRC) released a report entitled: *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*.² To date, it is the most comprehensive review on such risk factors for cancer.

Briefly, independently conducted literature reviews were commissioned from academic institutions in the United States, United Kingdom, and Europe. Based on the subsequent meta-analyses, an expert panel reviewed and judged this evidence and ranked the probability of a

causal relationship as convincing, probable, limited, or unlikely.² With respect to CRC and indicators of energy balance, it was concluded that:

- there is convincing evidence that subcutaneous fat (as indicated by BMI) and abdominal fat (as indicated by waist circumference) **increases** risk of CRC;
- there is convincing evidence that adult-attained height **increases** risk of CRC;
- there is probable evidence that low energy dense foods (low kcal) foods **decreases** risk and energy-dense (high kcal) foods **increases** risk of CRC;
- there is convincing evidence that physical activity **decreases** risk and a sedentary life style **increases** risk of CRC.

Since the publication of this report, additional studies have been published which support the consensus that body fat increases the risk of CRC⁴⁷⁻⁵² and physical activity decreases risk of CRC.^{43, 53-58} Energy restriction early in life was not evaluated in the WCRF/AICR report, however, evidence suggests that energy restriction early in life **decreases** the risk of CRC later in life.⁵⁹⁻⁶¹

COLORECTAL CANCER (CRC)

When studying CRC, it is common to consider risk either according to the entire large intestine, or risk according to the colon, where extraction of water and salts take place, and the rectum, which functions as a temporary storage compartment for faeces. However, as depicted in **Figure 3**, further classifications within the colon and rectum can be made.

It may be more informative to consider risk according to the proximal colon, distal colon, rectosigmoid and rectum because several anatomic, embryologic, and physiologic differences are known to exist between sub-anatomic locations.⁶²⁻⁶⁶ As a consequence of such differences, subsequent tumours likely arise via different etiologic pathways.

There also appear to be differences in risk according to sex, both for overall CRC and individual tumour sub-sites. In general, the incidence rate of CRC is higher in men than in women. However, absolute number of cases are similar between men and women, because risk of CRC increases for older age groups in which women outnumber men.⁶⁷ The incidence of left sided (distal colon) is greater in men, whereas women appear more at risk for right

sided (proximal colon) tumours.⁶⁷⁻⁶⁹ Reasons for these differences remain unclear, but differences in exposure and response to risk factors between men and women may explain the varying sub-site distribution observed.⁶² Many studies do not distinguish between tumours of the rectosigmoid and rectum, however, a higher incidence of rectal tumours in males than females has been reported.^{67, 70}

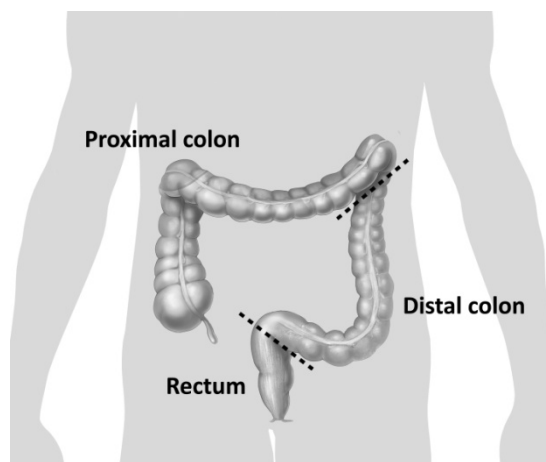


Figure 3:

The anatomy of the large intestine, indicating anatomical sub-sites of the colon (proximal colon, distal colon) and rectum

CRC is the most studied and best described cancer in terms of molecular events involved. Even though each tumour is (epi)genetically complex, and arises and behaves in a unique manner, it is common to classify tumours according to a limited number of phenotypes, because it is assumed that tumours with similar molecular characteristics have arisen through common mechanisms.¹⁶ Two parallel pathways to CRC are frequently described.

Approximately 85% of CRC tumours are thought to arise via the adenoma pathway, and are characterised by mutations in the *KRAS* oncogene and the *APC* tumour suppressor gene, as well as chromosomal instability (CIN).⁷¹ CIN describes a condition of aneuploidy that is caused by an accelerated rate of gains and losses of entire or large portions of the

chromosome during cell division.^{72, 73} CIN is also characterized by an increased rate of loss of heterozygosity, which may contribute to the inactivation of tumour suppressor genes.⁷⁴ These tumours are frequently associated with the male sex and are observed often in the distal colon.⁶⁴ Although this phenotype makes up the majority of CRCs, of interest to this thesis is the remaining 15% of tumours, which arise via the serrated neoplasia pathway and are characterised in part by a high degree of epigenetic instability and microsatellite instability (MSI). These will be explained in the following section, but briefly, we started this line of research with investigating this pathway because, in contrast to CIN, there is a clear biological mechanism and an established definition for MSI. In addition, it is easily measured in paraffin embedded tissue and therefore suitable to study in molecular pathological epidemiology studies.

EPIGENETIC INSTABILITY IN CRC

Epigenetics refers to heritable modifications on chromatin that do not change the nucleotide sequence of the DNA. These modifications can be imagined as a layer of information superimposed on the DNA sequence that ultimately influences gene expression.^{75, 76} Multiple epigenetic processes have been identified, including histone modification, chromatin remodelling, and DNA methylation. In recent years, considerable attention has been devoted to the role of DNA methylation in CRC development.

DNA methylation is a biochemical modification that, in human cells, primarily affects cytosines when they are part of the symmetrical dinucleotide CpG.⁷⁷⁻⁷⁹ When areas of DNA are particularly rich in CpG sequences, they are referred to as CpG islands. About half of all genes have a CpG island in their promoter region, which are normally unmethylated.⁷⁸ Adding methyl groups to these CpG islands often results in transcriptional silencing of the gene in question.⁸⁰ It is now known that there are many CRC specific methylation events associated with the pathways/process involved in CRC development (**Table 1**). It is also clear that epigenetic aberrations occur much more frequently than genetic aberrations.⁸¹

Table 1: Promoter CpG Island Hypermethylated Genes

Signalling Pathways/Cancer hallmarks	Promoter CpG island methylated genes
WNT	<i>APC, SFRP1, SFRP2, SFRP4, SFRP5, SOX17, WNT5a, DKK1, DKK3, WIF1, AXIN2</i>
RTK	<i>RASSF1A, RASSF2A, EPHB2, RAB32, NORE1</i>
NOTCH	<i>NEURL</i>
TP53	<i>P14^{ARF}, HIC1, DFNA5</i>
PI3K	<i>PIK3CG</i>
ER	<i>ER</i>
Retinoid acid	<i>RAR, CRABP1</i>
IGF	<i>IGF^{BP3}, IGFBP7</i>
Cell cycle regulation	<i>P16INK4A, KLF4</i>
Transcription regulation	<i>GATA4, GATA5, RUNX3, CDX1, HLF, FOXL2, ALX4</i>
DNA repair/stability	<i>MLH1, MGMT, WRN, CHFR</i>
Apoptosis	<i>BNIP3, IRF8, DAPK1, HRK</i>
Adhesion	<i>CDH4, CDH13</i>
Angiogenesis	<i>THBS1</i>
Invasion and metastasis	<i>TIMP3, RECK, CXCL12, TFPI2</i>
Axon guidance	<i>SLIT2</i>
Transmembrane (glycol)proteins	<i>MUC1, NMDAR2A, HPP1, SPARC, DCC, CD133</i>
Peptide hormones	<i>SST, TAC1</i>
Chromatin organization	<i>PRDM2</i>
Unknown function/pathway	<i>VIM, NDRG4</i>
Gene Panels	
Classic CIMP panel	<i>MINT1, MINT2, MINT31, P16^{INK4A}, MLH1</i>
Weisenberger panel	<i>CACNA1G, IGF2, NEUROG1, RUNX3, SOCS1</i>

NOTE. This Table, which is not exhaustive, represents genes of major signaling pathways/cancer hallmarks which are epigenetically silenced by promoter CpG island hypermethylation (adapted from van Engeland *et al*⁸¹).

Abbreviations: ER, estrogen receptor; IGF, insulin-like growth factor; CIMP, CpG island methylator phenotype.

It has been observed that a distinct sub-set of CRC tumours displays widespread promoter methylation, which is now referred to as the ‘CpG island methylator phenotype’ (CIMP).^{78, 82-84} While there is no standard definition for CIMP, the distinction between ‘CIMP’ and ‘non-CIMP’ tumours is usually based on the number of methylated markers examined.⁸⁵ In **Table 1**, it can be seen that two gene panels are frequently utilized to define CIMP. These markers were found via cluster analyses in semi-epigenome wide studies.^{82, 86}

Tumours characterized by CIMP are thought to arise via the serrated neoplasia pathway,⁸⁷ and have distinctly different histology than tumours of the traditional adenoma pathway.⁸⁸⁻⁹¹ The initiating incident appears to be a mutation in the *BRAF* proto-oncogene, which inhibits normal apoptosis of colonic epithelial cells.⁹⁰ In subsequent lesions, the most well characterized epigenetic silencing is that of the *MLH1* gene, which is thought to cause microsatellite instability (MSI), a form of genetic instability characterised by length alterations within simple repeated microsatellite sequences of DNA.^{92, 93} Once *MLH1* is inactivated, the rate of progression to malignant transformation is rapid.⁹⁰ Clinically, understanding this pathway is important because arising carcinomas are not prevented by removing conventional adenomas and may therefore be missed in routine screening programs. In addition, because a subset of these lesions may potentially progress rapidly to carcinoma, it is likely that these lesions will require a different screening strategy from that used for conventional adenoma.⁹⁰ Descriptively, tumours of the serrated neoplasia pathway are associated with older age, female sex, and tumours of the proximal colon.^{82, 94-98}

What is particularly intriguing about epigenetic instability is that timing of exposure to a given risk factor may be very important for establishing mechanisms. Animal and human studies indicate that certain transient environmental influences can produce persistent changes in epigenetic marks that have life-long phenotypic consequences.^{75, 99, 100} These observations add another layer of complexity to studying risk of CRC. Could body size, energy intake, and physical activity influence CRC risk differently depending on the period of life in which the exposure occurred? If so, could the epigenetic mechanisms just described play a role?

RATIONALE AND AIMS OF THIS THESIS

As outlined in the previous sections, a large body size is associated with an increased risk of CRC, whereas physical activity and energy restriction, especially energy restriction early in life, are associated with a protective effect. However, it is now well accepted that CRC is not a single disease, and the importance of these exposures on specific sub-sites of the large bowel, and according to sex, is not well defined. Furthermore, previously reported associations may have been masked or attenuated because etiologic heterogeneity between sub-groups with distinct molecular features was not accounted for.

To date, epidemiologic data for associations between indicators of energy balance and risk of molecular abnormalities associated with epigenetic instability (CIMP, MSI, *BRAF* mutations) remains rare, and limited to case-control studies.^{47, 101, 102} Ideally, such associations should be investigated in a prospective cohort setting, as this design is less susceptible to bias and therefore generally has higher validity. The aim of this thesis was to elucidate potential differences in risk factors for the sub-sites of the large bowel and sex, and for molecular sub-types of these tumours, which may lead to a better understanding of CRC aetiology and prevention.

At the onset of this thesis, it was hypothesised that body fat and adult attained height would increase the risk of CRC, and physical activity and energy restriction in adolescence would decrease the risk of CRC in our study population. Differential risk between tumour sub-sites was expected, and it was hypothesised that these differences may be linked to molecular phenotypes of the tumours, which in turn, may be associated with timing of exposure to the risk factors.

STUDY DESIGN

All studies in this thesis were completed within the Netherlands Cohort Study on diet and cancer (NLCS). The NLCS includes 120,852 men and women who were between the ages of 55 and 69 at baseline (1986), and completed a self-administered questionnaire on dietary habits, lifestyle, and other risk factors for cancer. The cohort was followed-up for cancer incidence through linkage to the Netherlands Cancer Registry, and PALGA (the nationwide registry for histopathology and cytopathology in the Netherlands). Further, a sub-cohort of 5000 individuals was randomly

selected after baseline exposure measurement, to estimate accumulation of person time at risk in the cohort through biennial follow-up of vital status.^{103, 104}

Overall CRC could be investigated among 3360 cases from a follow-up period of 16.3 years after baseline. Tumour material was collected from CRC patients identified within the first 7.3 years of follow-up, excluding the first 2.3 years, and of these, 734 had sufficient DNA for molecular analyses. This scheme is outlined in **Figure 4**.

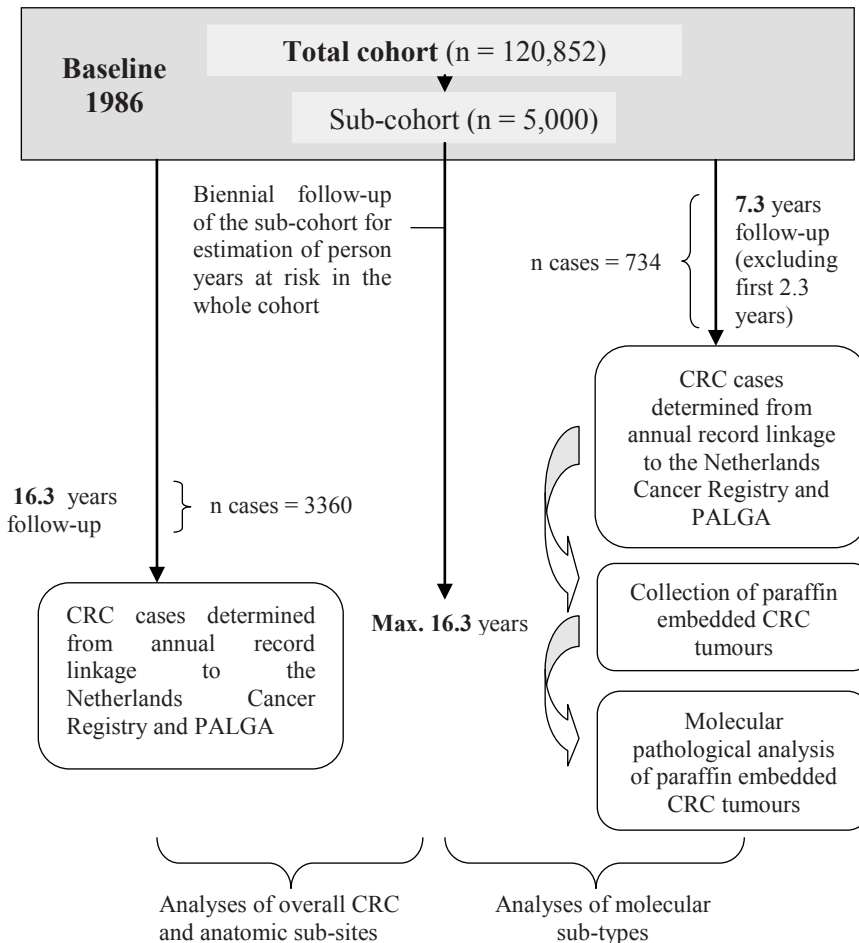


Figure 4:

Study design of the NLCS, including the collection of CRC tumours and molecular information

OUTLINE OF THIS THESIS

This thesis begins with a methodological study in which we evaluate the utility of self-reported clothing size as a predictor of cancer in the NLCS (**Chapter 2**).

We then focus our attention on energy restriction during childhood and adolescence as a risk factor for CRC. In **Chapter 3**, we describe sex-specific associations between energy restriction during this period of growth and risk of CRC, as well as risk of specific tumour sub-sites. In **Chapter 4**, we investigate whether this exposure influences risk of developing a tumour with or without CIMP or MSI.

Next, we consider adult aspects of energy balance as risk factors for CRC. In **Chapters 5 and 6**, the sex-specific relationship between body size and physical activity and CRC risk are described, respectively, as well as associations with respect to specific tumour sub-sites. We continue in **Chapter 7** by investigating whether heterogeneity of CIMP and non-CIMP tumours plays a major role with respect to adult body size, physical activity, and CRC risk. In **Chapter 8**, we pooled data from the NLCS with data from a large Australian cohort, the Melbourne Collaborative Cohort Study (MCCS), to describe associations between body size and risk of tumours with or without *BRAF* mutations and MSI.

Finally, **Chapter 9** concludes the thesis with a discussion of the main findings, as well as the challenges of and insights offered by molecular pathological epidemiology and implications and suggestions for future research.

REFERENCES

1. World Health Organization. 2010. Information sheet on obesity and overweight factsheet.
http://www.who.int/dietphysicalactivity/media/en/gsfs_obesity.pdf
<accessed 04/11/2010>
2. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AIRC, 2007.
3. Overweight and Obesity: Causes and consequences. CDC Centers for Disease Control and Prevention.
<http://www.cdc.gov/obesity/causes/index.html> <accessed 17/11/2010>
4. International Agency for Research on Cancer (IARC). 2008. *GLOBOCAN 2008: Colorectal Cancer Incidence and Mortality Worldwide in 2008*.
<http://globocan.iarc.fr/factsheets/cancers/colorectal.asp> <accessed 17/11/2010>
5. Kang HW, Kim D, Kim HJ, *et al*. Visceral obesity and insulin resistance as risk factors for colorectal adenoma: a cross-sectional, case-control study. *Am J Gastroenterol*. 2010; 105: 178-87.
6. Kissebah AH, Krakower GR. Regional adiposity and morbidity. *Physiol Rev*. 1994; 74: 761-811.
7. Netherlands Cancer Registry. Age standardised incidence rates per 100,000 person-years by site and sex, 1989-2007. Available at
http://www.ikcnet.nl/page.php?id=2755&nav_id=97 <accessed 03/12/10>.
8. Center MM, Jemal A, Smith RA, *et al*. Worldwide variations in colorectal cancer. *CA Cancer J Clin*. 2009; 59: 366-78.
9. Haggard FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in colon and rectal surgery*. 2009; 22: 191-7.
10. Ferlay J, Autier P, Boniol M, *et al*. Estimates of the cancer incidence and mortality in Europe in 2006. *Ann Oncol*. 2007; 18: 581-92.
11. Lebovitz HE, Banerji MA. Point: visceral adiposity is causally related to insulin resistance. *Diabetes care*. 2005; 28: 2322-5.

12. Haenszel W, Berg JW, Segi M, *et al.* Large-bowel cancer in Hawaiian Japanese. *J Natl Cancer Inst.* 1973; 51: 1765-79.
13. Kolonel LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat Rev Cancer.* 2004; 4: 519-27.
14. Kolonel LN. Cancer patterns of four ethnic groups in Hawaii. *J Natl Cancer Inst.* 1980; 65: 1127-39.
15. Ogino S, Chan AT, Fuchs CS, *et al.* Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut.* 2010; doi:10.1136/gut.2010.217182.
16. Ogino S, Stampfer M. Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. *J Natl Cancer Inst.* 2010; 102: 365-7.
17. Dietz WH, Bellizzi MC. Introduction: the use of body mass index to assess obesity in children. *Am J Clin Nutr.* 1999; 70: 123S-5S.
18. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, *et al.* Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health.* 2000; 54: 143-8.
19. Spencer EA, Appleby PN, Davey GK, *et al.* Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr.* 2002; 5: 561-5.
20. Hannan WJ, Wrate RM, Cowen SJ, *et al.* Body mass index as an estimate of body fat. *Int J Eat Disord.* 1995; 18: 91-7.
21. Bigaard J, Tjonneland A, Thomsen BL, *et al.* Waist circumference, BMI, smoking, and mortality in middle-aged men and women. *Obes Res.* 2003; 11: 895-903.
22. Dagenais GR, Yi Q, Mann JF, *et al.* Prognostic impact of body weight and abdominal obesity in women and men with cardiovascular disease. *Am Heart J.* 2005; 149: 54-60.
23. Arner P. Regional adiposity in man. *J Endocrinol.* 1997; 155: 191-2.
24. Pischon T, Lahmann PH, Boeing H, *et al.* Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst.* 2006; 98: 920-31.
25. Pischon T, Nothlings U, Boeing H. Obesity and cancer. *Proc Nutr Soc.* 2008; 67:128-45.

26. Han TS, Gates E, Truscott E, *et al.* Clothing size as an indicator of adiposity, ischaemic heart disease and cardiovascular risks. *J Hum Nutr Diet.* 2005; 18: 423-30.
27. Hirschhorn JN, Lettre G. Progress in genome-wide association studies of human height. *Horm Res.* 2009; 71 Suppl 2: 5-13.
28. Soranzo N, Rivadeneira F, Chinappan-Horsley U, *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS genetics.* 2009; 5:e1000445.
29. Okasha M, Gunnell D, Holly J, *et al.* Childhood growth and adult cancer. *Best Pract Res Clin Endocrinol Metab.* 2002; 16:225-41.
30. Johnson RK. Dietary intake--how do we measure what people are really eating? *Obes Res.* 2002; 10 Suppl 1: 63S-8S.
31. Briefel RR, Sempos CT, McDowell MA, *et al.* Dietary methods research in the third National Health and Nutrition Examination Survey: underreporting of energy intake. *Am J Clin Nutr.* 1997; 65: 1203S-9S.
32. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr.* 1994; 48: 253-65.
33. Albanes D. Total calories, body weight, and tumor incidence in mice. *Cancer Res.* 1987; 47: 1987-92.
34. Albanes D, Salbe AD, Levander OA, *et al.* The effect of early caloric restriction on colonic cellular growth in rats. *Nutr Cancer.* 1990; 13: 73-80.
35. Kritchevsky D. Colorectal cancer: the role of dietary fat and caloric restriction. *Mutat Res.* 1993; 290: 63-70.
36. Kritchevsky D, Klurfeld DM. Influence of caloric intake on experimental carcinogenesis: a review. *Adv Exp Med Biol.* 1986; 206: 55-68.
37. Kritchevsky D, Klurfeld DM. Interaction of fiber and energy registration in experimental colon carcinogens. *Cancer Lett.* 1997; 114: 51-2.
38. Rous P. The Influence of Diet on Transplanted and Spontaneous Mouse Tumors. *The J Exp Med.* 1914; 20: 433-51.

39. Tannenbaum A. The dependence of tumor formation on the composition of the calorie-restricted diet as well as on the degree of restriction. 1945. *Nutrition*. 1996; 12: 653-4.
40. Engelman RW, Day NK, Good RA. Calorie intake during mammary development influences cancer risk: lasting inhibition of C3H/HeOu mammary tumorigenesis by peripubertal calorie restriction. *Cancer Res*. 1994; 54: 5724-30.
41. Silverman J, Powers J, Stromberg P, *et al*. Effects on C3H mouse mammary cancer of changing from a high fat to a low fat diet before, at, or after puberty. *Cancer Res*. 1989; 49: 3857-60.
42. Food and Agriculture Organization of the United Nations. 2004. Human Energy Requirements. Report of a joint FAO/WHO/UNU Expert Consultation. In: FAO Food and Nutrition Technical Report, Series 1. .
43. Friedenreich CM, Neilson HK, Lynch BM. State of the epidemiological evidence on physical activity and cancer prevention. *Eur J Cancer*. 2010; 46: 2593-604.
44. Harriss DJ, Cable NT, George K, *et al*. Physical activity before and after diagnosis of colorectal cancer: disease risk, clinical outcomes, response pathways and biomarkers. *Sports Med*. 2007; 37: 947-60.
45. McTiernan A. Mechanisms linking physical activity with cancer. *Nat Rev Cancer*. 2008; 8: 205-11.
46. Slattery ML. Physical activity and colorectal cancer. *Sports Med*. 2004; 34: 239-52.
47. Campbell PT, Jacobs ET, Ulrich CM, *et al*. Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. *J Natl Cancer Inst*. 2010; 102: 391-400.
48. Dai Z, Xu YC, Niu L. Obesity and colorectal cancer risk: a meta-analysis of cohort studies. *World J Gastroenterol*. 2007; 13: 4199-206.
49. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr*. 2007; 86: 556-65.

50. Ning Y, Wang L, Giovannucci EL. A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies. *Obes Rev.* 2010; 11: 19-30.
51. Oxentenko A, Bardia A, Vierkant RA, *et al.* Body Size and Incident Colorectal Cancer: A Prospective Study of Older Women. *Cancer Prev Res.* 2010; doi: 10.1158/1940-6207.CAPR-10-0116.
52. Wang Y, Jacobs EJ, Patel AV, *et al.* A prospective study of waist circumference and body mass index in relation to colorectal cancer incidence. *Cancer Causes Control.* 2008; 19: 783-92.
53. Harriss DJ, Atkinson G, George K, *et al.* Lifestyle factors and colorectal cancer risk: a systematic review and meta-analysis of associations with body mass index. *Colorectal Dis.* 2009; 11: 547-63.
54. Moradi T, Gridley G, Bjork J, *et al.* Occupational physical activity and risk for cancer of the colon and rectum in Sweden among men and women by anatomic subsite. *Eur J Cancer Prev.* 2008; 17: 201-8.
55. Parent ME, Rousseau MC, El-Zein M, *et al.* Occupational and recreational physical activity during adult life and the risk of cancer among men. *Cancer Epidemiol.* 2010; doi:10.1016/j.canep.2010.09.004
56. Spence RR, Heesch KC, Brown WJ. A systematic review of the association between physical activity and colorectal cancer risk. *Scan J Med Sci Sports.* 2009; 19: 764-81.
57. Wolin KY, Lee IM, Colditz GA, *et al.* Leisure-time physical activity patterns and risk of colon cancer in women. *Int J Cancer.* 2007; 121: 2776-81.
58. Nilsen TI, Romundstad PR, Petersen H, *et al.* Recreational physical activity and cancer risk in subsites of the colon (the Nord-Trondelag Health Study). *Cancer Epidemiol Biomarkers Prev.* 2008; 17: 183-8.
59. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Energy restriction early in life and colon carcinoma risk: results of The Netherlands Cohort Study after 7.3 years of follow-up. *Cancer.* 2003; 97: 46-55.
60. Svensson E, Grotmol T, Hoff G, *et al.* Trends in colorectal cancer incidence in Norway by gender and anatomic site: an age-period-cohort analysis. *Eur J Cancer Prev.* 2002; 11: 489-95.

61. Svensson E, Moller B, Tretli S, *et al.* Early life events and later risk of colorectal cancer: age-period-cohort modelling in the Nordic countries and Estonia. *Cancer Causes Control.* 2005; 16: 215-23.
62. Cancer Care Ontario. Colorectal cancer subsite distribution differs by sex. April 2010. Available at <http://www.cancercare.on.ca/cancerfacts/> <accessed 03/12/10>.
63. Morois S, Mesrine S, Josset M, *et al.* Anthropometric Factors in Adulthood and Risk of Colorectal Adenomas: The French E3N-EPIC Prospective Cohort. *Am J Epidemiol.* 2010; 172: 1166-80.
64. Li FY, Lai MD. Colorectal cancer, one entity or three. *Journal of Zhejiang University Science.* 2009; 10: 219-29.
65. Breivik J, Lothe RA, Meling GI, *et al.* Different genetic pathways to proximal and distal colorectal cancer influenced by sex-related factors. *Int J Cancer.* 1997; 74: 664-9.
66. Lindblom A. Different mechanisms in the tumorigenesis of proximal and distal colon cancers. *Curr Opin Oncol.* 2001; 13: 63-9.
67. Gao RN, Neutel CI, Wai E. Gender differences in colorectal cancer incidence, mortality, hospitalizations and surgical procedures in Canada. *J Public Health.* 2008; 30:194-201.
68. Larsen IK, Bray F. Trends in colorectal cancer incidence in Norway 1962-2006: an interpretation of the temporal patterns by anatomic subsite. *Int J Cancer.* 2010; 126: 721-32.
69. Wu X, Cheng X, Chen VW, *et al.* Subsite-specific incidence rate and stage of disease in colorectal cancer by race, gender, and age group in the United States, 1992-1997. *Cancer.* 2001; 92: 2547-54.
70. Wu X, Chen VW, Martin J, *et al.* Subsite-specific colorectal cancer incidence rates and stage distributions among Asians and Pacific Islanders in the United States, 1995 to 1999. *Cancer Epidemiol Biomarkers Prev.* 2004; 13: 1215-22.
71. Derks S, Postma C, Carvalho B, *et al.* Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis.* 2008; 29: 434-9.
72. Sieber OM, Heinimann K, Tomlinson IP. Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer.* 2003; 3: 701-8.

73. Rajagopalan H, Nowak MA, Vogelstein B, *et al.* The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer.* 2003; 3: 695-701.
74. Hermsen M, Postma C, Baak J, *et al.* Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology.* 2002; 123: 1109-19.
75. Heijmans BT, Tobi EW, Stein AD, *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008; 105: 17046-9.
76. Huang YW, Kuo CT, Stoner K, *et al.* An overview of epigenetics and chemoprevention. *FEBS letters.* 2010; doi:10.1016/j.febslet.2010.11.002.
77. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002; 16: 6-21.
78. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer.* 2004; 4: 988-93.
79. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003; 33 Suppl: 245-54.
80. Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* 2000; 16: 168-74.
81. van Engeland M, Derks S, Smits KM, *et al.* Colorectal Cancer Epigenetics: Complex Simplicity. *JCO.* 2010; In press.
82. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006; 38: 787-93.
83. Wong JJ, Hawkins NJ, Ward RL. Colorectal cancer: a model for epigenetic tumorigenesis. *Gut.* 2007; 56: 140-8.
84. Ogino S, Cantor M, Kawasaki T, *et al.* CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut.* 2006; 55: 1000-6.
85. Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. *J Mol Diagn.* 2008; 10: 13-27.

86. Toyota M, Ahuja N, Ohe-Toyota M, *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A.* 1999; 96: 8681-6.
87. Noffsinger AE. Serrated polyps and colorectal cancer: new pathway to malignancy. *Ann Rev Pathol.* 2009; 4: 343-64.
88. East JE, Saunders BP, Jass JR. Sporadic and syndromic hyperplastic polyps and serrated adenomas of the colon: classification, molecular genetics, natural history, and clinical management. *Gastroenterol Clin North Am.* 2008; 37: 25-46.
89. Snover DC. Serrated polyps of the large intestine. *Semin Diagn Pathol.* 2005; 22: 301-8.
90. Snover DC. Update on the serrated pathway to colorectal carcinoma. *Hum Pathol.* 2011; 42: 1-10.
91. Snover DC, Jass JR, Fenoglio-Preiser C, *et al.* Serrated polyps of the large intestine: a morphologic and molecular review of an evolving concept. *Am J Clin Pathol.* 2005; 124: 380-91.
92. Imai K, Yamamoto H. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. *Carcinogenesis.* 2008; 29: 673-80.
93. Herman JG, Umar A, Polyak K, *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A.* 1998; 95: 6870-5.
94. Barault L, Charon-Barra C, Jooste V, *et al.* Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res.* 2008; 68: 8541-6.
95. Samowitz WS, Albertsen H, Herrick J, *et al.* Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology.* 2005; 129: 837-45.
96. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology.* 2007; 50: 113-30.
97. Ogino S, Kawasaki T, Kirkner GJ, *et al.* Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn.* 2007; 9: 305-14.
98. Samowitz WS. The CpG island methylator phenotype in colorectal cancer. *J Mol Diagn.* 2007; 9: 281-3.

99. Sinclair KD, Allegrucci C, Singh R, *et al.* DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proc Natl Acad Sci U S A.* 2007; 104: 19351-6.
100. Waterland RA, Jirtle RL. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol.* 2003; 23: 5293-300.
101. Slattery ML, Curtin K, Sweeney C, *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer.* 2007; 120: 656-63.
102. Slattery ML, Curtin K, Wolff RK, *et al.* Diet, physical activity, and body size associations with rectal tumor mutations and epigenetic changes. *Cancer Causes Control.* 2010; 21: 1237-45.
103. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol.* 1990; 43: 285-95.
104. Van den Brandt PA, Schouten LJ, Goldbohm RA, *et al.* Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol.* 1990; 19: 553-8.

1

2

Self-reported clothing size as a proxy measure for body size

Laura AE Hughes
Leo J Schouten
R Alexandra Goldbohm
Piet A van den Brandt
Matty P Weijnenberg

Epidemiology 2009; 20(5): 673-6

ABSTRACT

Background: Few studies have considered the potential utility of clothing size as a predictor of diseases associated with body weight.

Methods: We used data on weight-stable men and women from the Netherlands Cohort Study to assess the correlation of clothing size with other anthropometric variables. Cox regression using the case-cohort approach was performed to establish whether clothing size can predict cancer risk after 13.3 years of follow-up, and if additionally considering body mass index (BMI) in the model improves the prediction.

Results: Trouser and skirt size correlated well with circumference measurements. Skirt size predicted endometrial cancer risk, and this effect was slightly attenuated when BMI was added to the model. Trouser size predicted risk of renal cell carcinoma, regardless if BMI was present in the model.

Conclusions: Clothing size appears to predict cancer risk independently of BMI, suggesting that clothing size is a useful measure to consider in epidemiological studies when waist circumference is not available.

INTRODUCTION

Waist and hip circumferences are useful in epidemiological research because they are stronger predictors of total body fat than body mass index (BMI).¹ Additionally, waist and hip measurements reflect intra-abdominal fat accumulation, which is an important risk factor for a number of diseases.²⁻⁴ It has been hypothesized that clothing size is related to physique,⁵ and it was recently reported that clothing size appears to be a strong surrogate for estimating obesity and intra-abdominal fat.⁶

In the Netherlands Cohort Study on diet and cancer (NLCS), clothing size as well as weight and height were self-reported by all individuals at baseline. Waist and hip circumferences were only reported by a small sub-cohort of this population, and therefore, BMI has been used to assess the association between anthropometric indices and cancer risk. The present study had two objectives. One, among the sub-cohort of persons for whom both clothing size and waist and hip circumferences are available, we estimated the correlation of self-reported clothing size and self-reported waist circumference, hip circumference, and BMI. Two, we explored the predictive power of clothing size in cancer risk, and the degree to which the addition of clothing size added to the predictive power of a more complex measure such as BMI.

METHODS

The study design of the NLCS has been described in detail elsewhere.⁷⁻⁹ This prospective cohort study on diet and cancer was initiated in the Netherlands in 1986. The cohort includes 58,279 men and 62,573 women who were aged 55-69 years at baseline. It was designed as a case-cohort, using all cases and a random sample of 5000 individuals who were randomly selected from the main cohort at baseline. The sub-cohort has been followed-up biennially from baseline in 1986 for vital status to estimate person time at risk for the cohort, and to accumulate prospective data regarding a number of dietary and lifestyle factors. Collection of anthropometric information is depicted in **Figure 1**, and described in detail in the Appendix available at the end of this chapter.

The correlation study was conducted using only sub-cohort members. Individuals were selected if they reported their body weight both at baseline in 1986 and follow-up in 1992 (N=4009), and if their reported weights did

not differ more than 5 kg between these two time points (N=3392). Individuals missing one or more anthropometric measurements of interest (height [N=159], hip circumference [N=1246], waist circumference [N=1252], and trouser or skirt size [N=258] were dropped, resulting in 1158 men and 1334 women included in the correlation analysis. We used Spearman rank correlation was used to assess the sex-specific association between trouser or skirt size and the other anthropometric variables.

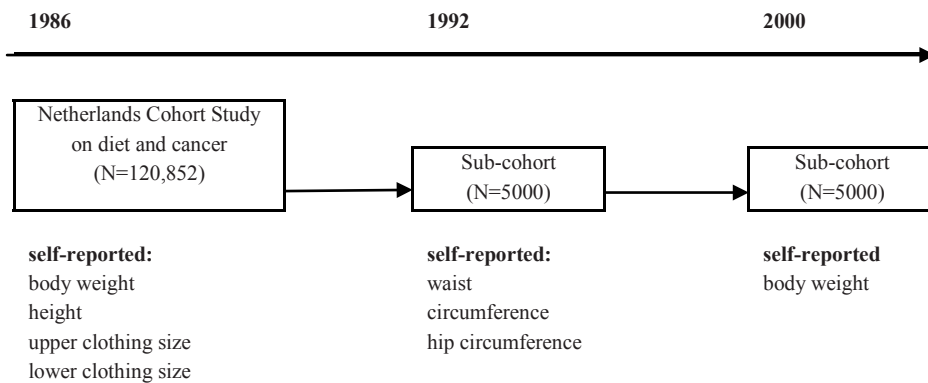


Figure 1:

A timeline of self-reported anthropometric variables in the Netherlands Cohort Study. All persons reported body weight, height, and clothing size at baseline. A random sub-cohort was selected (at baseline) and followed up with additional questionnaires.

We next aimed to establish whether clothing size can predict cancers known to be associated with intra-abdominal fat or a large waist circumference.^{2,3,10-13} The association between BMI and both endometrial cancer in women, and renal cancer in men, were investigated previously in this cohort.^{3,13} We hypothesized that using clothing size as a proxy measure for waist circumference would predict risk of these cancers and that adding BMI to this model would not improve the prediction.

Cancer incidence during follow-up was assessed by record linkage of the full cohort with the Netherlands cancer registry and a nationwide pathology database. Multivariate Cox regression using the case-cohort approach was performed on the whole cohort to obtain hazard rate ratios (HRs) and 95% confidence intervals after 13.3 years of follow-up. We

created models to assess the association between trouser or skirt size and the incidence of endometrial cancer (women) and renal cell carcinoma (men). To evaluate whether clothing size was predicting cancer risk independently of BMI, we compared models with and without BMI.

Cases and sub-cohort members were excluded from the endometrial cancer analyses if they reported cancer other than non-melanoma skin cancer, reported a hysterectomy at baseline (N=77), or had incomplete data for BMI, trouser or skirt size, or confounders (N= 350). Models were adjusted for age, age at menarche, use of oral contraceptives, age at menopause, parity, cigarette smoking, and physical activity.³ A total of 1708 sub-cohort members and 297 endometrial case patients were included.

Cases and sub-cohort members were excluded from the renal cell carcinoma analyses if they were female (N=2575), reported cancer other than non-melanoma skin cancer, and were missing variables for BMI, trouser size, or confounders (N= 523). Models were adjusted for age, cigarette smoking, total energy intake, and physical activity.¹³ A total of 1841 sub-cohort members and 195 renal cell carcinoma cases were included.

RESULTS

Baseline characteristics for the sub-cohort members used in the correlation analysis are presented in **Table 1**.

Table 1: Characteristics of sub-cohort members in the NLCS who met the inclusion criteria ^a for the correlation analysis

Variable ^{b,c}	Men (N=1158)		Women (N=1334)	
	Mean (SD)	Range	Mean (SD)	Range
age (years)	61.1 (4.1)	51.0-128.0	61.3 (4.2)	54.0-70.0
weight (kg)	77.4 (8.9)	46.0-110.0	67.5 (9.8)	43.0-115.0
height (cm)	176.6 (6.6)	147.0-199.0	165.1 (6.3)	140.0-185.0
body mass index (kg/m ²)	24.8 (2.2)	17.6-37.0	24.8 (3.3)	16.8-40.7
waist circumference (cm)	95.2 (8.4)	54.0-70.0	88.1 (10.7)	47.0-129.0
hip circumference (cm)	101.0 (6.5)	49.0-131.0	104.1 (8.4)	54.0-142.0
Waist-Hip ratio	0.94 (0.06)	0.55-1.76	0.85 (0.08)	0.47-1.65
Dutch trouser/skirt size	51.7 (2.5)	40.0-65.0	43.3 (3.0)	36.0-54.0
Dutch shirt size	40.6 (2.5)	30.0-58.0	42.8 (2.9)	36.0-54.0

^a Individuals did not change more than 5kg between the two weight reports in 1986 and 1992, and were not missing measurements for weight, height, waist circumference, hip circumference, pant/skirt size, or shirt size

^b All variables were self-reported, with the exception of BMI and waist-hip ratio which were calculated from self-reported variables

^c All variables were reported at baseline (1986), with the exception of waist and hip circumference which was reported in 1988

In men, trouser size correlated slightly better with waist circumference (0.64) and hip circumference (0.63) than BMI (0.53). In women, the correlations between skirt size and hip circumference, waist circumference, and BMI were similar (0.71-0.78). Correlations in women were generally higher than in the men (**Table 2**). Stratifying correlations by age and BMI did not yield different findings

Table 2: Spearman Rank Correlations between self-reported trouser and skirt size and self-reported anthropometric variables in men and women of sub-cohort members of the NLCS

	Correlations with trouser and skirt size ^a	
	Men (N=1158)	Women (N=1334)
	Correlation (95% CI)	Correlation (95% CI)
BMI (kg/m ²)	0.53 (0.49-0.57)	0.76 (0.74-0.78)
waist circumference (cm)	0.64 (0.60-0.67)	0.71 (0.68-0.73)
hip circumference (cm)	0.63 (0.60-0.67)	0.78 (0.75-0.80)
waist-hip ratio	0.28 (0.23-0.34)	0.29 (0.24-0.33)

^a Spearman rank correlation (95% CI)



Skirt size predicted endometrial cancer risk for women in the two largest size categories (for size 46-48), HR: 1.60, 95%CI: 1.11-2.34 and (for size 50+), HR: 5.4, 95%CI: 2.97-9.71, and when size was entered as a continuous variable (HR: 1.35, 95% CI: 1.22-1.51). When BMI was added to the model, all associations were attenuated (**Table 3**).

In men, HRs for the association between trouser size and renal cell carcinoma for each size category were similar regardless if BMI was present in the model. This was also true when size was considered as a continuous variable ((for trouser size alone HR: 1.05, 95%CI: 0.99-1.11); (for pant size plus BMI HR:1.05, 95%CI: 0.98-1.13)) (Table 3).

Table 3: Associations of skirt size with endometrial cancer ^a and trouser size with renal cell carcinoma ^b in the NLCS after 13.3 years of follow-up, and the effect of BMI on these associations

Variable	N cases	PY at risk	HR (95% CI)	HR (95% CI) ^c
<i>Endometrial cancer (skirt size)^d</i>				
<40	30	3828	0.73 (0.45-1.17)	0.88 (0.53-1.43)
42	64	5486	1 (referent)	1 (referent)
44	72	5798	1.06 (0.73-1.54)	0.90 (0.60-1.32)
46-48	101	5499	1.60 (1.12-2.29)	1.14 (0.74-1.75)
>50	30	598	5.38 (2.97-9.71)	2.98 (1.39-6.35)
continuous	297	21,200	1.35 (1.22-1.51)	1.11 (0.93-1.32)
<i>P</i> for trend			0.00	0.06
<i>Renal cell carcinoma (trouser size)^e</i>				
<50	28	3182	1.28 (0.75-2.18)	1.29 (0.76-2.21)
50-51	35	4899	1 (referent)	1 (referent)
52-53	66	7379	1.23 (0.80-1.93)	1.24 (0.79-1.95)
54-55	41	4091	1.41 (0.87-2.29)	1.41 (0.83-2.39)
>56	25	2030	1.63 (0.93-2.84)	1.55 (0.83-2.89)
continuous	195	21,582	1.05 (0.99-1.11)	1.05 (0.98-1.13)
<i>P</i> for trend			0.19	0.34

^a Analyses have been adjusted for age, age at menarche, use of oral contraceptives, age at menopause, parity, cigarette smoking (never, past, current), and non-occupational physical activity.

^b Analyses have been adjusted for age, cigarette smoking (current smoking (yes/no), number of cigarettes/day, number of years smoking), total energy intake, and non-occupational physical activity.

^c Analyses have been additionally adjusted for BMI (kg/m²) at baseline.

^d Skirt size corresponds to the following standard waist circumferences: <40= <74 cm; 40= 74cm; 42= 78cm; 44= 82 cm; 46= 86cm; >50= >96 cm.

^e Trouser size corresponds to the following standard waist circumferences: <50= <88cm; 50-51= 93cm; 52-53= 93cm; 54-55= 98cm; >56= >103 cm.

DISCUSSION

Although the idea of using clothing sizes in epidemiologic studies is not new,^{5,6,14,15} few studies have focused on establishing the potential of this measurement as a predictor of diseases such as cancer that are associated with body weight.^{2,10-12}

Clothing size correlated relatively well with waist and hip circumferences in both men and women, consistent with past findings.⁶ With respect to cancer risk, mutual adjustment of clothing size and BMI did not improve the prediction of renal cell carcinoma in men compared to clothing size alone. Although the association was not entirely removed in women when BMI was added to the endometrial cancer model, the strength of the association was attenuated. When we repeated the analyses using shirt size, similar associations were observed (see appendix at the end of this chapter), suggesting that both upper- and lower- clothing size reflect the same fat distribution.

The difference observed between men and women may in part be explained in part by the fact that men tend to have a greater percentage of abdominal fat.¹⁶ Etiologically, abdominal fat may be more relevant than peripheral fat for many types of cancer.¹⁰ Clothing size may be more reflective of the fat distribution associated with cancer risk, at least in men.

Measurement bias may have influenced our findings. All body measures were obtained by self-report. While self-reported weight, height, waist circumference and hip circumference correlate highly with technician measured values,¹⁷⁻²⁴ study participants (especially women with a high BMI) are more likely to under-estimate their waist circumference than over-estimate.^{24,25} This might lead to an underestimation of actual disease risk.^{20,24} Secondly, clothing sizes and waist and hip circumferences were reported two years apart. We attempted to control for potential bias by limiting our correlation analyses to individuals who did not report an extreme weight change over six years; however, misclassification is still possible.

Misclassification on the basis of clothing size may have occurred for additional reasons. Some individuals may prefer clothes with a very loose, or a very tight fit. It could be that individuals have a large waist circumference yet wear a smaller size trouser that sits below the waist. Similarly, women may wear dresses with an elastic waist, which would also underestimate their size.⁶ Despite these shortcomings, a recent British study reported that in a

study population of 362 people, only 3% all of subjects reported one clothing size differently from their labels.⁶

In recent years, clothing sizes have drifted away from original standards; differences among manufactures have been reported in Europe and the United States, and numerical sizing systems differ globally.^{6,26} Researchers interested in assessing the association between clothing size measurements and disease risk should consider that with international travel, internet shopping, and the concept of vanity sizing,^{6,26} individuals may report sizes that are not standard in their population.

Clothing size appears to predict cancer risk independently of BMI, suggesting that clothing size reflects a fat distribution different from that indicated by weight and height. Therefore, it may be useful for future epidemiological studies to collect clothing sizes in addition to weight and height, especially in populations where obtaining waist circumference is culturally problematic, or challenging because of subjects with extreme obesity.⁶ There is evidence to suggest clothing size indicates adiposity,⁶ but more studies are needed to validate this association.

In conclusion, self-reported clothing sizes correlates well with self-reported waist circumference and hip circumference in weight stable men and women in a Dutch cohort. Furthermore, clothing size can be used in addition to BMI to predict cancer risk.

APPENDIX

Ascertainment of anthropometric variables

All cohort members were asked to report their height (cm) and body weight (kg) on the baseline questionnaire in 1986. BMI was calculated using the reported weight at baseline, divided by height at baseline squared. At this time point individuals were also asked to generally report their upper body (shirt) and lower body (trouser or skirt) clothing size from their clothing label (Dutch sizes). Clothing variables were re-coded if they were reported in American sizes (shirt size: 1 woman, 25 men; trouser size: 1 woman, 38 men). Trouser sizes in men were reported as even and odd numbers. Odd numbered trouser sizes in men are identical in waist circumference to the even number preceding it, and differ only in that they are longer. We were interested in clothing size as a proxy for waist circumference, so odd numbered sizes were recoded to the even size preceding it (i.e. a size 51 was recoded to a size 50). In some cases (N=16), women reported more than one size, for example 40 and 42. To account for this, the average size was entered (i.e. 41) during data entry. For the present analysis, we recoded odd numbered sizes in woman to the next highest even size (i.e. 41 to 42).

Biennial follow-up of the sub-cohort has allowed for the accumulation of prospective data regarding a number of factors related to body weight and lifestyle habits. At these time points, self-administered questionnaires were mailed to all sub-cohort members given they were still alive and reachable at the same address. If necessary, new addresses were obtained and individuals were contacted there. Two years after baseline (1988), self-reported waist and hip measurements were requested using such a follow-up questionnaire. The instructions were: measure while standing, do not measure over clothing (underwear may be kept on), measure the waist at the level of the belly button, and measure the hip where it is largest. A cartoon of a body depicting the waist and hip measurement was included with the boxes to enter the measurement (cm) next to the waist and hip, respectively. We calculated the waist-hip ratio by dividing the reported waist circumference by the reported hip circumference. Six years after baseline (1992), individuals of the sub-cohort were asked again to report their current body weight (kg).

Tables

Appendix Table 1: Spearman rank correlation between self-reported shirt size and self-reported anthropometric variables in men and women of the sub-cohort of the NLCS

	Correlations with shirt size ^a	
	Men (N=1158) Correlation (95% CI)	Women (N= 1334) Correlation (95% CI)
BMI (kg/m ²)	0.58 (0.54-0.62)	0.73 (0.70-0.75)
Waist circumference (cm)	0.54 (0.50-0.58)	0.67 (0.64-0.70)
Hip circumference (cm)	0.45 (0.41-0.50)	0.70 (0.68-0.73)
Waist-hip ratio	0.33 (0.27-0.38)	0.30 (0.25-0.34)
^a Spearman rank correlation (95% CI)		

Appendix Table 2: Associations of shirt size with endometrial cancer ^a and with renal cell carcinoma ^b in the NLCS after 13.3 years of follow-up, and the effect of BMI on these associations

Variable	N cases	PY at risk	HR (95% CI)	HR (95% CI) ^c
<i>Endometrial cancer^d</i>				
<40	37	4975	0.65 (0.43-0.99)	0.75 (0.46-1.21)
42	78	6259	1 (referent)	1 (referent)
44	70	5309	1.03 (0.72-1.48)	0.92 (0.63-1.34)
46-48	89	4120	1.75 (1.24-2.49)	1.32 (0.86-2.04)
>50	23	536	4.56 (2.45-8.49)	2.62 (1.21-5.68)
continuous	297	21,200	1.36 (1.25-1.51)	1.14 (0.98-1.34)
<i>P</i> for trend			0.00	0.02
<i>Renal cell carcinoma^e</i>				
<40	60	6462	1.29 (0.81-2.05)	1.34 (0.83-2.16)
40	30	3977	1 (referent)	1 (referent)
41	35	4544	1.03 (0.61-1.72)	1.03 (0.61-1.73)
42	32	3598	1.19 (0.70-2.01)	1.13 (0.65-1.98)
≥43	38	2999	1.67 (0.99-2.79)	1.52 (0.88-2.64)
continuous	195	21,582	1.01 (0.96-1.06)	1.00 (0.95-1.05)
<i>P</i> for trend				

^a Analyses have been adjusted for age, age at menarche, use of oral contraceptives, age at menopause, parity, cigarette smoking (never, past, current), and non-occupational physical activity.

^b Analyses adjusted for age, cigarette smoking (current smoking (yes/no), number of cigarettes/day, number of years smoking), total energy intake, and physical activity.

^c Analyses have been additionally adjusted for BMI (kg/m²) at baseline.

^d Shirt size corresponds to the following standard waist circumferences: <40 = <74 cm; 40= 74 cm; 42=78 cm; 44= 82 cm; 46= 86cm; >50 = >96 cm.

^e Shirt size corresponds to neck circumference in cm.

REFERENCES

1. Lean ME, Han TS, Deurenberg P. Predicting body composition by densitometry from simple anthropometric measurements. *Am J Clin Nutr.* 1996; 63: 4-14.
2. Pischon T, Lahmann PH, Boeing H, *et al.* Body size and risk of renal cell carcinoma in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer.* 2006; 118: 728-38.
3. Schouten LJ, Goldbohm RA, van den Brandt PA. Anthropometry, physical activity, and endometrial cancer risk: results from the Netherlands Cohort Study. *J Natl Cancer Inst.* 2004; 96: 1635-8.
4. Vazquez G, Duval S, Jacobs DR, *et al.* Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiol Rev.* 2007; 29: 115-28.
5. Heady JA, Morris JN, Raffle PA. Physique of London busmen; epidemiology of uniforms. *Lancet.* 1956; 271: 569-70.
6. Han TS, Gates E, Truscott E, *et al.* Clothing size as an indicator of adiposity, ischaemic heart disease and cardiovascular risks. *J Hum Nutr Diet.* 2005; 18: 423-30.
7. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol.* 1990; 43: 285-95.
8. Goldbohm RA, van 't Veer P, van den Brandt PA, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr.* 1995; 49: 420-9.
9. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr.* 1994; 48: 253-65.
10. Pischon T, Nothlings U, Boeing H. Obesity and cancer. *Proc Nutr Soc.* 2008; 67: 128-45.
11. Adams KF, Leitzmann MF, Albanes D, *et al.* Body size and renal cell cancer incidence in a large US cohort study. *Am J Epidemiol.* 2008; 168: 268-77.
12. Friedenreich C, Cust A, Lahmann PH, *et al.* Anthropometric factors and risk of endometrial cancer: the European prospective

- investigation into cancer and nutrition. *Cancer Causes Control*. 2007; 18: 399-413.
13. van Dijk BA, Schouten LJ, Kiemeneij LA, *et al*. Relation of height, body mass, energy intake, and physical activity to risk of renal cell carcinoma: results from the Netherlands Cohort Study. *Am J Epidemiol*. 2004; 160: 1159-67.
 14. Morris JN, Heady JA, Raffle PA, *et al*. Coronary heart-disease and physical activity of work. *Lancet*. 1953; 265: 1111-20.
 15. Morris JN, Heady JA, Raffle PA, *et al*. Coronary heart-disease and physical activity of work. *Lancet*. 1953; 265: 1053-7.
 16. Executive summary of the clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. *Arch Intern Med*. 1998; 158: 1855-67.
 17. Freudenheim JL, Darrow SL. Accuracy of self-measurement of body fat distribution by waist, hip, and thigh circumferences. *Nutr Cancer*. 1991; 15: 179-86.
 18. Han TS, Seidell JC, Currall JE, *et al*. The influences of height and age on waist circumference as an index of adiposity in adults. *Int J Obes Relat Metab Disord*. 1997; 21: 83-9.
 19. Kushi LH, Kaye SA, Folsom AR, *et al*. Accuracy and reliability of self-measurement of body girths. *Am J Epidemiol*. 1988; 128: 740-8.
 20. Rimm EB, Stampfer MJ, Colditz GA, *et al*. Validity of self-reported waist and hip circumferences in men and women. *Epidemiology*. 1990; 1: 466-73.
 21. Sonnenschein EG, Kim MY, Pasternack BS, *et al*. Sources of variability in waist and hip measurements in middle-aged women. *Am J Epidemiol*. 1993; 138: 301-9.
 22. Spencer EA, Roddam AW, Key TJ. Accuracy of self-reported waist and hip measurements in 4492 EPIC-Oxford participants. *Public Health Nutr*. 2004; 7: 723-7.
 23. Weaver TW, Kushi LH, McGovern PG, *et al*. Validation study of self-reported measures of fat distribution. *Int J Obes Relat Metab Disord*. 1996; 20: 644-50.
 24. Hall TR, Young TB. A validation study of body fat distribution as determined by self-measurement of waist and hip circumference. *Int J Obes*. 1989; 13: 801-7.

25. Bigaard J, Spanggaard I, Thomsen BL, *et al.* Self-reported and technician-measured waist circumferences differ in middle-aged men and women. *J Nutr.* 2005; 135: 2263-70.
26. Kinley T. Size variation in women's pants. *Clothing and Textiles Research Journal.* 2003; 21: 19-31.

3

Childhood and adolescent energy restriction and subsequent colorectal cancer risk:

Results from the Netherlands Cohort Study

Laura AE Hughes
Piet A van den Brandt
R Alexandra Goldbohm
Anton FPM de Goeij
Adriaan P de Bruïne
Manon van Engeland
Matty P Weijnenberg

ABSTRACT:

Background: Energy restriction during childhood and adolescence is suggested to lower colorectal cancer (CRC) risk. We investigated this in the Netherlands Cohort Study.

Methods: Information on diet and other risk factors was collected by a baseline questionnaire in 1986 when cohort members were 55-69 years of age (n=120,852). Three indicators of early life exposure to energy restriction were assessed: father's employment status during the Economic Depression (1932-40), and place of residence during World War 2 years (1940-44) and the 'Hunger Winter' (1944-45); a severe famine. Using the case-cohort approach, hazard ratios (HR) and 95% confidence intervals (CI) were calculated for total colorectal, proximal colon, distal colon, rectosigmoid, and rectal cancers, according to the three time periods of energy restriction. After 16.3 years of follow-up, 2573 cases were available for multivariate analyses.

Results: Men who lived in a western city during the Hunger Winter and therefore exposed to the highest degree of energy restriction, had a lower risk of developing CRC (HR: 0.81, 95% CI: 0.68-0.98), and tumours of the proximal colon (HR: 0.72, 95% CI: 0.54-0.96) and rectum (HR: 0.71, 95% CI: 0.53-0.96). In women, non-statistically significant inverse associations were observed for tumours of the distal colon, rectosigmoid, and rectum. Inverse associations were also observed between the other two exposure times and studied endpoints, though not statistically significant.

Conclusions: This unique observational evidence suggests that severe energy restriction during childhood and adolescence may lower CRC risk, especially in men, thus providing insight regarding the role of energy intake during early life in CRC development.

INTRODUCTION

Cancers of the colon and rectum are the third most common type of cancer worldwide.¹ According to the World Cancer Research Fund (WCRF), there is convincing evidence that components of energy balance such as body and abdominal fatness, and factors leading to a greater adult attained height increase the risk of colorectal cancer (CRC).¹ Numerous animal studies suggest that energy restriction protects against the development of CRC,²⁻⁶ and observational research suggests that energy restriction during childhood and adolescence may especially be protective against CRC cancer development and or mortality later in life.⁷⁻¹⁰

For ethical and practical reasons, it is difficult to initiate a study to test the hypothesis of childhood energy restriction and future cancer risk. However, information derived from a unique situation in the Netherlands has enabled this hypothesis to be tested. During the early to mid 20th century, a large part of the Dutch population experienced three subsequent periods of impaired nutrition: the Economic Depression of 1932-40, World War 2 (WW2) (1940-44), and a severe famine: the Hunger Winter of 1944-45.

During the Economic Depression, there was a high rate of unemployment in the Netherlands. Surveys document that while the amount of total calories available to families with an unemployed father was sufficient, they consumed fewer calories (3000 vs. 3400 kcal/day) and had less variation in their food patterns than families with an employed father. The daily menu of the unemployed families consisted primarily of boiled potatoes and some fat, and lacked cheese, meat and confectionary.^{7, 11} During WW2, food rationing was introduced and because of poor food availability in large cities, nutritional differences emerged between people living in urban versus rural areas. Individuals living in cities consumed fewer calories daily; the average allowance was ~1600 kcal, which was maintained with difficulty. Furthermore, urban diets consisted of a higher percentage of carbohydrates (70% vs 65%) and a lower percentage of fat (10% vs 15%).^{7, 12} The southern part of The Netherlands was liberated from German occupation in September 1944, whereas the western part of the country remained under German rule. The Hunger Winter occurred primarily in the western cities of the Netherlands as the end result of a German imposed embargo on food transport into the major Dutch cities. While the diet remained nutritionally balanced, individuals living in western urban areas experienced rationing of

<700 kcal per day.^{13, 14} The famine was relatively short; it abruptly ended in May 1945 when the western and northern part of the Netherlands was liberated as well.^{13, 14} This situation provides a rare opportunity to study the association between severe energy restriction during childhood and adolescence, and CRC risk later in life.

The Netherlands Cohort Study on diet and cancer (NLCS) is comprised of individuals who were children or going through the adolescent growth spurt during the years spanning the Economic Depression, WW2, and the Hunger Winter. Although individual food intake data is not available for these time periods, a number of proxy measures have been collected from cohort members that reflect their exposure to energy restriction. While the association between energy restriction during childhood and colon carcinomas was studied in the NLCS at an earlier follow-up point,⁷ that study lacked the power to consider risk at multiple tumour sub-sites. The present study investigated the association between population based energy restriction in childhood and adolescence and the subsequent risk of total CRC, and for tumours of the proximal and distal colon, rectosigmoid, and rectum.

METHODS

Study population and design

The NLCS is a prospective cohort study that was initiated in 1986 with the purpose of investigating the association between diet and the development of cancer. The cohort included 58,279 men and 62,573 women between the ages of 55-69 years at baseline who completed a self-administered questionnaire on dietary habits and lifestyle, health, and demographic information. The questionnaire requested information about the residence(s) of cohort members, including their residence during WW2 and the winter of 1944-45, and also asked for their father's employment status during the Economic Depression. Municipal registries from throughout the Netherlands were used to constitute an efficient sampling frame.¹⁵⁻¹⁷ A case – cohort approach was used for data processing and analysis; case subjects were derived from the entire cohort, and the number of person-years at risk for the entire cohort was estimated from a sub-cohort of 5000 men and women who were randomly sampled from the full cohort at baseline. Further

details of the design of the NLCS have been described in previous publications.¹⁵⁻¹⁷

Incident CRC cases from the entire cohort were identified by annual record linkage to nine regional cancer registries and a national pathology database (PALGA). This method of record linkage has been described elsewhere.¹⁸ The completeness of cancer follow-up has been estimated to be ~100%.¹⁹ CRCs were classified as proximal colon cancer (International Classification of Diseases for Oncology, first edition codes 153.0, 153.1, 153.4, 153.5 and 153.6), distal colon cancer (codes 153.2, 153.3 and 153.7), rectosigmoid (code 154.0) and rectal cancer (code 154.1). All prevalent cancer cases at baseline other than non-melanoma skin cancer and subjects with incomplete or inconsistent questionnaires were excluded. After additionally excluding individuals who were missing variables that were adjusted for in the multivariate analysis, 3981 sub-cohort members and 2971 CRC cases (Men: 484 proximal colon, 584 distal colon, 189 rectosigmoid, 436 rectum, 48 not classified; Women: 514 proximal colon, 372 distal colon, 95 rectosigmoid, 225 rectum, 30 not classified) were ultimately included in the analysis.

3

Assessment of Energy Restriction

Individual food intake data of the cohort members was not available for the three periods of energy restriction. Therefore, proxy variables were used to describe exposure to energy restriction. With respect to the Economic Depression years, the best proxy available was the employment status of the cohort member's father during this time frame. On the baseline questionnaire, individuals were asked 'was your father employed during the crisis of 1932-1940?' Cohort members could either answer yes, along with the years that he was employed, or no. For analyses purposes, answers were categorised into 'father was employed during the economic depression' or 'father was not employed during the economic depression'.

For the years spanning both WW2 and the Hunger Winter, place of residence was used to approximate the exposure to energy restriction. Place of residence in 1942 (midpoint of the war years) was used as a proxy for energy restriction during WW2. On the baseline questionnaire, cohort members were asked to list the last four residences they lived in, the province where these residences were located, and the years that they lived in these

locations. The location of residence during the mid-point of WW2 was dichotomised to ‘lived in a city in 1942’, or ‘lived in a rural area in 1942’. Additionally, cohort members were specifically asked where they lived during the winter of 1944-45. Residing in a western city during the Hunger Winter was considered an indicator for severe energy restriction. Eleven cities in the west of the Netherlands were considered famine cities based on the definition by Stein *et al.*²⁰: Amsterdam, Rotterdam, The Hague, Utrecht, Zaandam, Hilversum, Amersfoort, Dordrecht, Vlaardingens/Schiedam, Delft, and Leiden. Answers to this question were classified into three categories: lived in a western city, lived in a western rural area, and lived in a non-western area of the Netherlands during the winter of 1944-1945.

Data Analyses

Data were analysed using Stata (version 9.2, Statacorp, College Station, TX, USA). Because the aetiology of CRC may differ according to sex and sub-site of the cancer, men and women were analysed separately for overall colorectal (including rectosigmoid), colon, proximal colon, distal colon, rectosigmoid, and rectal cancers. Cox proportional hazard analysis using the case-cohort approach was used to obtain hazard ratios (HR) and 95% confidence intervals (CI) for the association between the three proxy measures of energy restriction and the incidence of CRC overall, and for each sub-site. The proportional hazards assumption was tested using the scaled Schoenfeld residuals and by visual inspection of the hazard curves. When the proportional hazards test appeared to be violated, analyses were done for short and long follow-up by splitting up the follow-up time at 8 years. To account for the additional variance introduced by sampling the sub-cohort from the entire cohort, standard errors were estimated using the robust option.

HRs were estimated using age-adjusted and multivariate adjusted models. Potential confounders that introduced a more than 10% change in HRs or contributed significantly to the prediction were included as covariates in the multivariate model. These ultimately included age, family history of CRC, height, body mass index, baseline alcohol and energy intake, recreational physical activity, level of education, and smoking history.⁷ HRs were expressed using ‘father was employed during the economic depression’, ‘lived in a rural area in 1942’, and ‘lived in a non-western area of the Netherlands during the Hunger Winter’ as the reference categories for the

different proxies of energy restriction.

Previously in the NLCS, sub-group analysis was conducted after 7.3 years of follow-up to assess if timing of exposure modified the effect of energy restriction on colon carcinoma cancer risk.⁷ In that study there were too few cases in each sub-group to reach definitive conclusions; therefore, we repeated this sub-group analysis in terms of total CRC to include the maximum number of cases. For each of the three exposure periods, individuals were classified into a sub-group depending on if they were exposed to energy restriction before, during, or after the adolescent growth spurt. Men and women were analysed separately. For men, the growth spurt was defined as occurring between the ages of 12 and 15. For women, the growth spurt was defined as 2 years before the reported age of menarche to 1 year after the reported year of menarche (mean age Economic Depression: 13.1 years, WW2: 14.3 years, Hunger Winter: 14.8 years). The long exposure periods of the Economic Depression and WW2 years were restricted to those years when food shortages were reported to be the most severe: 1933-1934 and 1942-1943 respectively. Therefore, only individuals who met the definition of going through their growth spurt (i.e. males who were 12-15 years of age) during 1933-34 or 1942-43 were included in the sub-group. While many more CRC cases were available compared to the analysis at 7.3 years follow-up, the female sub-group 'exposed before growth spurt to the Hunger Winter' still had too few cases for the model to converge. Due to an extremely low number of cases, associations for women exposed to the Hunger Winter before their adolescent growth spurt are not reported.

Finally, tests for heterogeneity were done to evaluate differences between subtypes of tumours (e.g., colon vs. rectum) using the competing risks procedure in STATA. However, the SE for the difference of the log-hazard ratios from this procedure assumes independence of both estimated hazard ratios, which would overestimate that SE and thus overestimate the *P* values for their difference. Therefore, these *P* values and the associated confidence intervals were estimated based on a bootstrapping method that was developed for the case-cohort design.²¹ Each bootstrap analysis was based on 1,000 replications.

RESULTS

Baseline characteristics of the sub-cohort are presented in **Table 1** and **Table 2** for men and women, respectively. Men who lived in a western rural area during the Hunger Winter had a higher level of occupational physical activity and a higher level of education compared to men who lived in a non-western area or a western city. With respect to the war years, men who lived in a western city in 1942 had a lower BMI and a lower daily intake of calories at baseline compared to men who did not. Additionally, these men had a higher level of education and a lower level of occupational physical activity. Men who had an employed father during the Economic Depression had a higher level of education and had a lower level of occupational physical activity than men whose father was not employed. Women who lived in a western city during the Hunger Winter had a lower energy intake and a higher intake of alcohol at baseline compared to women who lived in a non-western or western rural area, and a higher level of education. Women who lived in a western rural area had the lowest proportion of CRC family history. Women who lived in an urban area in 1942 were shorter, had a lower energy intake and a higher alcohol intake at baseline, had a higher level of education, exercised more, and smoked less than women who lived in a rural area. Women whose father was employed during the Economic Depression were younger, taller, had a lower BMI, a higher intake of alcohol and energy at baseline, were better educated, and smoked less than women whose father did not have a job.

Table 1: Distribution ^a of potential confounders according to time period of exposure to energy restriction in the male sub-cohort of the NLCS (1986) after 16.3 years of follow-up

	Hunger Winter ^b (1944-45)				War Years (1940-44)				Economic Depression (1932-1940)	
	Lived in non-western area	Lived in western rural	Lived in western city	Lived in rural area in 1942	Lived in urban area in 1942	Father had a job	Father did not have a job	1705	206	
Number of individuals ^b	980	249	415	729	753	1705	206			
Baseline and population characteristics										
Age	61.1 (4.2)	60.4 (4.2)	61.0 (4.4)	61.2 (4.2)	61.5 (4.2)	61.1 (4.2)	61.6 (4.5)			
Age range during period of exposure ^d	12-28 (19)	12-27 (18)	12-28 (19)	12-26 (17)	12-26 (17)	1-17 (8)	1-16 (9)			
Height (m)	176.2 (6.4)	178.2 (6.8)	176.8 (6.4)	176.5 (6.6)	176.1 (6.5)	176.6 (6.6)	176.1 (6.6)			
Weight (kg)	78.2 (9.2)	77.5 (9.4)	77.3 (9.3)	78.7 (9.6)	76.6 (9.2) ^e	77.8 (9.3)	77.7 (9.6)			
BMI (kg/m ²)	25.1 (2.5)	24.4 (2.6)	24.7 (2.5)	25.2 (2.6)	24.7 (2.5) ^e	24.9 (2.6)	25.0 (2.6)			
Baseline alcohol intake (g/d)	14.8 (16.5)	14.7 (16.7)	14.9 (17.6)	13.9 (15.2)	14.2 (16.6)	14.9 (16.8)	14.1 (15.6)			
Baseline energy intake (kcal)	2194 (520)	2197 (504)	2107 (471)	2213 (524)	2112 (473) ^e	2180 (510)	2133 (506)			
Family history of CRC (%)										
No	95.3	96.0	94.7	94.9	95.2	94.9	94.2			
Yes	4.7	4.0	5.3	5.1	4.8	5.1	5.8			
Level of education (%)										
Low	49.0	45.4	33.7 ^e	58.0	39.9 ^e	41.9	59.7 ^e			
Medium	33.1	34.9	40.2	30.9	41.8	37.4	27.2			
High	17.8	19.7	26.0	11.1	18.3	20.8	13.1			
Recreational physical activity (%)										
<30 min/d	16.9	20.5	17.8	16.7	17.3 ^e	16.8	21.4			
30-60 min/d	32.2	33.5	27.0	33.2	28.2	31.6	28.6			
60-90 min/d	18.7	21.7	22.4	18.0	19.7	20.1	15.1			
>90 min/d	32.1	24.5	32.8	32.1	34.9	31.6	35.0			
Smoking status (%)										
Current smoker	35.6	35.3	32.8	36.5	35.5	36.1	32.0			
Ex-smoker	50.8	49.8	54.2	51.7	51.5	51.1	56.8			
Never smoked	13.6	14.9	13.0	11.8	12.0	12.8	11.2			

^a presented as mean (standard deviation) or percentage (%) when indicated; ^b Hunger Winter: it was assumed that individuals residing in a western area of the Netherlands, especially a western urban area during the Hunger Winter would have experienced the greatest degree of energy restriction; War Years: it was assumed that individuals who resided in an urban area during the mid-point of the War would have experienced the greatest degree of energy restriction; Economic Depression: it was assumed that individuals who had an unemployed father during the Economic Depression would have experienced a higher degree of energy restriction compared to those whose father was employed; ^c subcohort members who were not missing data for the given exposure and any of the considered baseline characteristics; ^d age range (mean age); ^e analysis of variance, t-test or chi-square test: p<0.05

Table 2: Distribution^a of potential confounders according to time period of exposure to energy restriction in the female sub-cohort of the NLCS (1986) after 16.3 years of follow-up

	Hunger Winter (1944-45)			War Years (1940-44)		Economic Depression (1932-1940)	
	Lived in non-western area	Lived in western rural	Lived in western city	Lived in rural area in 1942	Lived in urban area in 1942	Father had a job	Father did not have a job
Number of individuals ^b	1077	286	554	700	824	1728	221
Baseline and population characteristics							
Age	61.2 (4.3)	61.1 (4.1)	61.4 (4.3)	61.2 (4.1)	61.4 (4.4)	61.2 (4.3)	61.7 (4.2) ^c
Age range during period of exposure ^d	12-28 (19)	12-28 (19)	12-28 (19)	10-26 (17)	10-26 (17)	1-17 (8)	1-17 (9)
Age at menarche	13.7 (1.8)	13.6 (1.7)	13.7 (1.8)	13.7 (1.7)	13.7 (1.8)	13.7 (1.7)	13.7 (2.1)
Height (m)	165.4 (6.3)	165.6 (5.8)	165.4 (5.9)	165.4 (6.0)	165.1 (6.1) ^c	165.5 (6.0)	163.6 (6.8) ^e
Weight (kg)	68.4 (10.0)	69.2 (10.3)	68.2 (9.9)	68.7 (9.8)	68.7 (10.2)	68.8 (10.0)	68.8 (10.6)
BMI (kg/m ²)	25.0 (3.5)	25.2 (3.5)	24.9 (3.5) ^c	25.1 (3.4)	25.2 (3.5) ^c	24.9 (3.4)	25.7 (3.9) ^c
Baseline alcohol intake (g/d)	5.5 (9.2)	5.4 (8.2)	6.8 (10.6) ^c	4.7 (7.8)	6.3 (10.0) ^c	6.1 (9.2)	4.5 (8.0) ^c
Baseline energy intake (kcal)	1709 (410)	1701 (402)	1661 (365) ^c				
Family history of CRC (%)							
No	93.2	96.3	94.7 ^c	94.8	94.9	94.2	94.5
Yes	6.8	3.7	5.3	5.2	5.1	5.8	5.5
Level of education (%)							
Low	59.6	59.1	44.2 ^c	70.1	46.3 ^c	52.8	70.7 ^c
Medium	32.4	30.1	44.5	23.5	44.6	37.3	24.3
High	8.0	10.8	11.2	6.5	9.1	10.0	5.1
Recreational physical activity (%)							
<30 min/d	23.9	17.8	21.7 ^c	31.4	23.8 ^c	21.9	28.9
30-60 min/d	32.1	39.0	27.2	36.3	27.3	32.1	26.6
60-90 min/d	22.1	23.8	25.6	22.0	24.5	23.6	22.5
>90 min/d	22.0	19.3	25.4	20.4	24.4	22.4	22.0
Smoking status (%)							
Current smoker	20.3	15.2	24.9	23.9	17.1 ^c	19.9	31.2 ^c
Ex-smoker	18.8	21.2	23.3	23.9	14.3	21.2	18.4
Never smoked	60.9	63.6	51.8	52.2	68.0	58.9	50.5

^a Presented as mean (standard deviation) or percentage where indicated.

^b Sub-cohort members who were not missing data for the given exposure and any of the considered baseline characteristics

^c Analysis of variance, *t*-test or chi-square test: *P*<0.05

^d age range (mean age)

Associations between the three proxy measures of energy restriction and CRC risk are presented in **Table 3** for men and **Table 4** for women. Men who reported living in a western city during the winter of 1944-1945 had a lower risk of CRC compared to men who lived in a non-western area (HR 0.81, 95% CI: 0.68-0.98). With respect to tumour sub-site, living in a western city during the Hunger Winter was protective against risk of tumours of the total colon (HR: 0.76, 95% CI: 0.60-0.96), proximal colon (HR 0.72, 95% CI: 0.54-0.96), and rectum (HR: 0.71, 95% CI: 0.53-0.96). Living in a city in 1942 and having a father without a job during the economic depression were not associated with CRC, or the specific CRC sub-sites. Tests for heterogeneity suggest that there is no difference in risk between the tumour sub-sites.

In women, living in a western city during the Hunger Winter was not associated with total CRC risk (HR: 0.99, 95% CI: 0.83-1.18). Analyses according to tumour sub-site were heterogeneous. Living in a western city was inversely associated with tumours of the distal colon, rectosigmoid, and rectum, while an increased risk was observed for proximal colon tumours (HR: 1.23, 95% CI: 0.97-1.56). Residing in a city in 1942 was inversely associated with risk of rectal cancer (HR 0.66, 95%CI: 0.47-0.94), but was not associated with tumours at other sub-sites. Having a father without a job during the economic depression was not associated with CRC risk or the risk of cancer at each sub-site. Tests for heterogeneity indicate that there is no difference in risk between the tumour sub-sites.

Table 3: Age-adjusted and multivariate-adjusted hazard ratios (HR) and 95% Confidence Intervals (CI) for colorectal cancers according to three time periods of energy restriction in the male population of the NLCS, after 16.3 years of follow-up

	Hunger Winter (1944-45)			War Years (1940-44)		Economic Depression (1932-1940)	
	Lived in non-western area ^a	Lived in western rural area	Lived in western city	Lived in rural area in 1942 ^a	Lived in urban area in 1942	Father had a job ^a	Father did not have a job
PY at risk	13, 798	3646	5774	10,366	10,455	24,098	2850
Total CRC							
N cases	905	218	314	666	654	1484	184
HR ^b (95% CI)	1 (referent)	0.95 (0.77-1.16)	0.83 (0.69-0.99)	1 (referent)	0.96 (0.82-1.12)	1 (referent)	1.03 (0.83-1.28)
HR ^c (95% CI)	1 (referent)	0.96 (0.78-1.19)	0.81 (0.68-0.98)	1 (referent)	0.95 (0.81-1.12)	1 (referent)	0.92 (0.72-1.18)
Proximal colon							
N cases	255	67	78	210	169	422	45
HR ^b (95% CI)	1 (referent)	1.04 (0.77-1.41)	0.73 (0.55-0.97)	1 (referent)	0.82 (0.65-1.03)	1 (referent)	0.88 (0.62-1.24)
HR ^c (95% CI)	1 (referent)	1.03 (0.76-1.41)	0.72 (0.54-0.96)	1 (referent)	0.78 (0.61-1.00)	1 (referent)	0.79 (0.53-1.18)
Distal colon							
N cases	297	68	118	195	226	502	63
HR ^b (95% CI)	1 (referent)	0.90 (0.66-1.22)	0.95 (0.74-1.21)	1 (referent)	1.13 (0.91-1.41)	1 (referent)	1.04 (0.77-1.41)
HR ^c (95% CI)	1 (referent)	0.92 (0.67-1.26)	0.89 (0.69-1.15)	1 (referent)	1.07 (0.84-1.36)	1 (referent)	0.95 (0.66-1.37)
Rectosigmoid							
N cases	87	30	31	76	63	153	21
HR ^b (95% CI)	1 (referent)	1.35 (0.87-2.10)	0.85 (0.55-1.30)	1 (referent)	0.81 (0.57-1.15)	1 (referent)	1.14 (0.71-1.84)
HR ^c (95% CI)	1 (referent)	1.46 (0.92-2.32)	0.85 (0.55-1.31)	1 (referent)	0.82 (0.58-1.17)	1 (referent)	1.12 (0.65-1.95)
Rectum							
N cases	247	50	71	186	169	369	49
HR ^b (95% CI)	1 (referent)	0.78 (0.56-1.10)	0.69 (0.51-0.92)	1 (referent)	0.89 (0.71-1.13)	1 (referent)	1.11 (0.80-1.55)
HR ^c (95% CI)	1 (referent)	0.81 (0.57-1.15)	0.71 (0.53-0.96)	1 (referent)	0.95 (0.74-1.23)	1 (referent)	0.94 (0.64-1.38)

^a Reference category

^b Model adjusted for age

^c Model adjusted for age, family history of colorectal cancer, BMI, energy intake at baseline, alcohol intake at baseline, recreational physical activity, level of education, and smoking status

Table 4: Age-adjusted and multivariate-adjusted hazard ratios (HR) and 95% Confidence Intervals (CI) for colorectal cancers according to three time periods of energy restriction in the female population of the NLCS, after 16.3 years of follow-up

	Hunger Winter (1944-45)		War Years (1940-44)		Economic Depression (1932-1940)		
	Lived in non-western area ^a	Lived in western rural area	Lived in western city	Lived in rural area in 1942 ^a	Lived in urban area in 1942	Father had a job ^a	Father did not have a job
PY at risk	16,070	4217	8117	10,425	3234	25,576	3234
Total CRC							
N cases	658	178	327	451	497	1040	124
HR ^b (95% CI)	1 (referent)	1.05 (0.84-1.30)	0.98 (0.82-1.16)	1 (referent)	0.94 (0.80-1.11)	1 (referent)	0.90 (0.71-1.14)
HR ^c (95% CI)	1 (referent)	1.08 (0.87-1.35)	0.99 (0.83-1.18)	1 (referent)	0.94 (0.78-1.13)	1 (referent)	0.91 (0.71-1.16)
Proximal colon							
N cases	254	75	157	176	237	442	54
HR ^b (95% CI)	1 (referent)	1.15 (0.86-1.55)	1.21 (0.97-1.52)	1 (referent)	1.14 (0.91-1.43)	1 (referent)	0.94 (0.68-1.29)
HR ^c (95% CI)	1 (referent)	1.20 (0.89-1.63)	1.23 (0.97-1.56)	1 (referent)	1.15 (0.90-1.47)	1 (referent)	0.90 (0.65-1.26)
Distal colon							
N cases	209	55	90	140	140	319	37
HR ^b (95% CI)	1 (referent)	1.02 (0.73-1.41)	0.85 (0.65-1.11)	1 (referent)	0.86 (0.66-1.11)	1 (referent)	0.90 (0.63-1.31)
HR ^c (95% CI)	1 (referent)	1.05 (0.75-1.48)	0.87 (0.66-1.15)	1 (referent)	0.87 (0.66-1.15)	1 (referent)	0.92 (0.63-1.34)
Rectosigmoid							
N cases	54	12	20	37	31	85	7
HR ^b (95% CI)	1 (referent)	0.86 (0.45-1.62)	0.73 (0.43-1.22)	1 (referent)	0.71 (0.44-1.16)	1 (referent)	0.63 (0.29-1.39)
HR ^c (95% CI)	1 (referent)	0.86 (0.44-1.67)	0.67 (0.39-1.17)	1 (referent)	0.71 (0.40-1.24)	1 (referent)	0.63 (0.28-1.41)
Rectum							
N cases	121	31	53	92	74	194	22
HR ^b (95% CI)	1 (referent)	0.96 (0.64-1.46)	0.84 (0.60-1.18)	1 (referent)	0.69 (0.50-0.96)	1 (referent)	0.88 (0.55-1.41)
HR ^c (95% CI)	1 (referent)	0.99 (0.65-1.54)	0.85 (0.61-1.20)	1 (referent)	0.66 (0.47-0.94)	1 (referent)	0.90 (0.56-1.45)

^a Reference category
^b Model adjusted for age
^c Model adjusted for age, family history of colorectal cancer, BMI, energy intake at baseline, alcohol intake at baseline, recreational physical activity, level of education, and smoking status

Results from the sub-group analyses can be seen in **Table 5**, for both men and women. Our cohort has no men who were exposed to the Hunger Winter before their growth spurt. In men, exposure to energy restriction during or after the growth spurt was protective against CRC for those who resided in a western city during the Hunger Winter compared to those who resided in non-western area (HR: 0.56, 95% CI: 0.31-1.03 and HR: 0.76, 95% CI: 0.61-0.94, respectively). Residing in a western rural area during the Hunger Winter also seemed to be protective of CRC in men as opposed to residing in a non-western area at that time (HR: 0.52, 95% CI: 0.26-1.02). In women, too few cases were exposed to the Hunger Winter before the growth spurt to reach definitive conclusions. There were no clear associations or differences between sub-groups during the War Years or Economic Depression for both men and women.

Table 5. Hazard Ratios (HR) and 95% Confidence Intervals (CI) of CRC for men and women in the NLCs (1986-2002) exposed to energy restriction before, during, and after the adolescent growth spurt^a, according to three different time periods of energy restriction.

Exposure	Before Growth Spurt			During Growth Spurt			After Growth Spurt		
	N cases	PY at risk	HR ^b (95% CI)	N cases	PY at risk	HR ^b (95% CI)	N cases	PY at risk	HR ^b (95% CI)
MEN									
Hunger Winter									
Lived in a non-west area ^c				119	2734	1 (referent)	767	11063	1 (referent)
Lived in western rural area	No males exposed before growth spurt			23	896	0.56 (0.31-1.03)	192	2749	1.06 (0.84-1.33)
Lived in western city				33	1119	0.52 (0.26-1.02)	265	4655	0.76 (0.61-0.94)
War Years									
Lived in a rural area in 1942 ^c	33	901	1 (referent)	126	2596	1 (referent)	499	6839	1 (referent)
Lived in a city in 1942	39	722	1.32 (0.59-2.93)	118	2680	0.93 (0.64-1.34)	470	7043	0.87 (0.71-1.05)
Economic Depression									
Father had a job ^c	1024	18,420	1 (referent)	279	3985	1 (referent)	143	1684	1 (referent)
Father had no job	128	2029	1.09 (0.33-1.42)	31	518	0.87 (0.51-1.40)	19	303	0.63 (0.29-1.36)
WOMEN									
Hunger Winter									
Lived in non-west area	15	390	--- ^d	88	3511	1 (referent)	538	12169	1 (referent)
Lived in western rural area	1	81	---	21	1017	0.87 (0.50-1.53)	151	3119	1.14 (0.89-1.46)
Lived in a western city	2	232	---	41	1840	0.89 (0.57-1.39)	277	6045	1.08 (0.89-1.32)
War Years									
Lived in a rural area in 1942	28	1093	1 (referent)	81	2701	1 (referent)	336	6657	1 (referent)
Lived in city in 1942	32	1398	0.68 (0.30-1.53)	97	951	1.01 (0.66-1.54)	353	7560	0.94 (0.76-1.18)
Economic Depression									
Father had a job	644	18290	1 (referent)	296	5156	1 (referent)	100	2131	1 (referent)
Father had no job	73	2189	0.90 (0.66-1.23)	30	659	0.77 (0.46-1.29)	17	385	0.92 (0.43-1.97)

^a Age of growth spurt defined as: Men (12-15 years old), Women (2 years before reported age of menarche – 1 year after reported age of menarche)

^b HR adjusted for age, family history of CRC, BMI, baseline alcohol intake, baseline energy intake, baseline recreational physical activity, level of education, and smoking status; ^c reference category; ^d too few cases for model to converge

DISCUSSION

In this large prospective cohort study, we examined the association between energy restriction in childhood and adolescence and CRC risk at a number of tumour sub-sites. After 16.3 years of follow-up in the NLCS, almost 3000 CRC cases were available for analysis. In men, we observed that living in a city classified as having experienced the greatest degree of caloric rationing during the Hunger Winter²⁰ was inversely associated with the risk of developing CRC. While inverse non-statistically significant associations were also observed in women, it appears that risk may differ according to sub-site.

Prior to discussing these results further, it is important to address an important limitation in our study. Because information on individual food for the cohort was not available for the exposure periods, three proxy measures of energy restriction were used to describe energy restriction during the Economic Depression, WW2, and the Hunger Winter. It is possible that using the father's employment status, place of residence during WW2, and place of residence during the Hunger Winter has resulted in some misclassification of exposure; however, in a follow-up study in the NLCS, female sub-cohort members were asked if they really had experienced hunger during the winter of 1944-1945. Of the women who reported severe hunger, 80% lived in a western city during this winter.²² These results indicate that place of residence during the Hunger Winter is a reasonably adequate proxy of exposure, although unavoidable misclassification may have caused attenuation of the association. While it has been proposed that individual famine data may be more accurate than area-exposure data,²³ other studies in the Netherlands have used place of residence during the Hunger Winter as a measure of energy restriction,^{24,25} and their results indicate that it is an adequate proxy.

The strong inverse associations we observed during the period of the Hunger Winter, especially for those who lived in a western city, suggest that in men, it may be caloric restriction as opposed to specific nutrient restriction that modulates the risk of CRC. This observation is in line with evidence from animal studies²⁶⁻²⁸ and could explain why we observed weaker, non-statistically significant associations during the other two periods of exposure when caloric restriction was assumed to be less severe. Alternatively, new insights reported by Trienekens *et al.*²⁹ suggest that while there was heavy

rationing in the years prior to the Hunger Winter, food distribution was generally satisfactory in the Netherlands, regardless of geographical area. However, this was only true up until the period of the Hunger Winter (1944-1945). It could therefore be argued that differences are not to be expected when using 'place of residence during the war years (1940-44)' as a proxy for exposure to energy restriction as the contrast in exposures is too small.

In the female population, our observations were more heterogeneous. It was interesting to observe that energy restriction during the Hunger Winter was inversely associated with tumours of the distal colon, rectosigmoid, and rectum, but a slightly elevated risk was observed with tumours of the proximal colon. Also, energy restriction during WW2 was significantly protective against rectal tumours, but again, positively associated with tumours of the proximal colon, although not statistically significant. Further research is needed to elucidate the association between energy restriction and risk of CRC at different sub-sites. We can only speculate about why we observed differences between men and women in terms of CRC risk. The majority of female cases were exposed to the Hunger Winter at least one year following their first menstrual period. For women in this sub-group, circulating estrogen levels would have been much higher compared to the women exposed before and during the adolescent growth spurt. It has been suggested that the beneficial effect of caloric restriction on CRC risk is mediated through its influence on IGF (especially IGF-1).³⁰ However, IGF is also influenced by estrogen.³¹ Therefore, some interaction between these two hormones may have offset a positive influence of caloric restriction, and could explain why we did not observe a reduced HR for CRC in women exposed to the Hunger Winter.

Our findings are generally consistent with other epidemiological studies. The association between energy restriction in childhood and colon carcinoma risk was previously investigated in the NLCS after 7.3 years of follow-up. While limited in terms of size, a weak, inverse association was reported.⁷ After 16.3 years, it appears that the protective association previously observed for women residing in a Western city during the Hunger Winter and colon carcinoma risk has been attenuated. Stratifying our data by early and late follow-up time (not shown) suggests that at both early and late follow-up there was an increased risk of proximal colon cancer, and a decreased risk of distal colon cancer. Furthermore, the difference in risk

between sub-sites becomes greater with the later follow-up time. This may help explain why different HR for total colon cancer were observed in women at 7.3 vs. 16.3 years of follow-up, and reiterates the importance of assessing risk by colorectal sub-site. As of yet, there is no clear biological explanation for this observation, but as none of the HR calculated at 7.3 years of follow-up or in the present study were statistically significant, it is possible that these the associations observed when only considering later follow-up are due to chance alone.

3 Information pertaining to exposure to energy restriction around the time period of WW2 has also been collected in several Nordic countries. In Norway, a country that experienced a 20% reduction in energy intake during WW2, it has been observed that individuals born during or shortly after the War have a decreased risk of CRC compared to individuals who were not born around this period.⁹ It is interesting to observe that this trend extends to Estonia, but not to Finland; two countries that both experienced severe restriction.¹⁰ It has been proposed that the effects of caloric restriction may have been counteracted in the Finnish population because approximately 70,000-80,000 Finnish children were sent to Sweden during WW2, a country that experienced less severe rationing.¹⁰ Furthermore, observations from the Boyd Orr cohort study in the United Kingdom suggest that there is a positive association between childhood energy intake and adult mortality from cancer, including gastrointestinal cancers.⁸

It is important to determine the mechanisms through which energy restriction in childhood is associated with a decreased risk of CRC and how this might differ by sub-site. Adult attained height is marker of an aggregated fetal and childhood experience, and can be considered a proxy measure for important nutritional exposures, which impact on several hormonal and metabolic axes.¹ In the present study, height was not acting as an intermediate variable in the risk between exposure to energy restriction and CRC. This was not surprising, as adult height may be most influenced early in childhood, and most individuals in our study were exposed to the Hunger Winter in the later years of childhood and puberty. Long term effects on insulin, insulin like growth factors (especially IGF-1), and adipokines such as leptin have also been proposed as possible mechanisms,^{4, 9, 10, 13, 28} but how exposure during childhood and adolescence translates into protection during adulthood remains elusive. The exposure time frames that we considered,

especially the Hunger Winter, were short-term, transient periods of energy restriction.²⁹ While not directly related to the strong one-time population starvation studies considered here, it has been argued that acute periods of energy restriction are dangerous for humans because it can lead to the phenomenon of weight cycling, that is, repetitive periods of weight loss and weight gain that may result in a modest acceleration of the carcinogenic response.^{32, 33} However, research also suggests that subtle nutritional stimuli at specific critical periods of development can induce persistent changes in gene expression and metabolism through modulation of genetic and epigenetic mechanisms.^{34, 35} This idea is supported by the observation that being overweight during adolescence may be a more powerful predictor of CRC risk than being overweight during adulthood.³⁶

We have observed that exposure to the Hunger Winter during adolescence is associated with a lower risk of developing a tumour characterized by the CpG island methylator phenotype (CIMP) after 7.3 years of follow-up.³⁷ CIMP is a distinct characteristic of epigenetic instability in CRC development,³⁸ and our observations support the hypothesis that adolescence may be a critical period of development for epigenetic changes that influence CRC risk. Evidence suggests that CIMP is highly prevalent in the proximal colon, and in females.³⁸ While the present study suggests a positive association between exposure to the Hunger Winter and risk of tumours at the proximal colon in women, this may be a finding of chance. Although we did not observe statistically significant differences with respect to risk at tumour sub-sites, focusing on associations between energy restriction and molecular characteristics of CRC may shed light on potential differences in risk

In conclusion, data from this unique period of Dutch history has allowed us to obtain rare observational evidence suggesting that severe energy restriction during adolescence may lower CRC risk at a number of tumour sub-sites, especially in men. These results give further insight into the role of energy intake during this period of life in CRC development.

REFERENCES

1. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AIRC, 2007.
2. Albanes D, Salbe AD, Levander OA, *et al.* The effect of early caloric restriction on colonic cellular growth in rats. *Nutr Cancer*. 1990; 13: 73-80.
3. Kritchevsky D. Colorectal cancer: the role of dietary fat and caloric restriction. *Mutat Res*. 1993; 290: 63-70.
4. Kritchevsky D, Klurfeld DM. Influence of caloric intake on experimental carcinogenesis: a review. *Adv Exp Med Biol*. 1986; 206: 55-68.
5. Kritchevsky D, Klurfeld DM. Interaction of fiber and energy registration in experimental colon carcinogens. *Cancer Lett*. 1997; 114: 51-2.
6. Ross MH, Bras G. Lasting influence of early caloric restriction on prevalence of neoplasms in the rat. *J Natl Cancer Inst*. 1971; 47: 1095-113.
7. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Energy restriction early in life and colon carcinoma risk: results of The Netherlands Cohort Study after 7.3 years of follow-up. *Cancer*. 2003; 97: 46-55.
8. Frankel S, Gunnell DJ, Peters TJ, *et al.* Childhood energy intake and adult mortality from cancer: the Boyd Orr Cohort Study. *BMJ*. 1998; 316: 499-504.
9. Svensson E, Grotmol T, Hoff G, *et al.* Trends in colorectal cancer incidence in Norway by gender and anatomic site: an age-period-cohort analysis. *Eur J Cancer Prev*. 2002; 11: 489-95.
10. Svensson E, Moller B, Tretli S, *et al.* Early life events and later risk of colorectal cancer: age-period-cohort modelling in the Nordic countries and Estonia. *Cancer Causes Control*. 2005; 16: 215-23.
11. den Hartog A. Werklozen en hun voeding in de jaren 1930-39 (Unemployed people and their nutrition during the years 1930-39 [in Dutch]). *Voeding*. 1983; 44: 92-8.
12. Dols M, van Arcken A. De voedselvoorziening in Nederlands tijdens en onmiddellijk na den tweeden wereldoorlog 1940-1945 (Food

- supply during and after the second World War 1940-1945 in the Netherlands [in Dutch]). *Voeding*. 1946; 6: 193-207.
13. Elias SG, Keinan-Boker L, Peeters PH, *et al.* Long term consequences of the 1944-1945 Dutch famine on the insulin-like growth factor axis. *Int J Cancer*. 2004; 108: 628-30.
 14. Burger G, Stanstead H, Drummond J. *Malnutrition and starvation in Western Netherlands, September 1944-45. Part I and II*. The Hague: General State Printing Office; 1948.
 15. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr*. 1994; 48: 253-65.
 16. Goldbohm RA, van 't Veer P, van den Brandt PA, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr*. 1995; 49: 420-9.
 17. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol*. 1990; 43: 285-95.
 18. Van den Brandt PA, Schouten LJ, Goldbohm RA, *et al.* Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol*. 1990; 19: 553-8.
 19. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of municipalities by cancer registries and PALGA using hospital discharge data. *Tijdschr Soc Gezondheidsz*. 1994; 72: 80-4.
 20. Stein Z, Susser M, Saenger G, Marolla F. *Famine and human development, the Dutch Hunger winter of 1944-1945*. New York: Oxford University Press; 1975.
 21. de Vogel S, Bongaerts BW, Wouters KA, *et al.* Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. *Carcinogenesis*. 2008; 29: 1765-73.
 22. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Diet in adolescence and the risk of breast cancer: results of the Netherlands Cohort Study. *Cancer Causes Control* 1999; 10: 189-99.

23. Elias SG, Peeters PH, Grobbee DE, *et al.* Breast cancer risk after caloric restriction during the 1944-1945 Dutch famine. *J Natl Cancer Inst.* 2004; 96: 539-46.
24. Lumey LH, Stein AD. Offspring birth weights after maternal intrauterine undernutrition: a comparison within sibships. *Am J Epidemiol.* 1997; 146: 810-9.
25. van Noord PA, Kaaks R. The effect of wartime conditions and the 1944-45 'Dutch famine' on recalled menarcheal age in participants of the DOM breast cancer screening project. *Ann Hum Biol.* 1991; 18: 57-70.
26. Albanes D. Total calories, body weight, and tumour incidence in mice. *Cancer Res.* 1987; 47: 1987-92.
27. Shimokawa I, Yu BP, Masoro EJ. Influence of diet on fatal neoplastic disease in male Fischer 344 rats. *J Gerontol.* 1991; 46: B228-32.
28. Weindruch R, Albanes D, Kritchevsky D. The role of calories and caloric restriction in carcinogenesis. *Hematol Oncol Clin North Am.* 1991; 5: 79-89.
29. Trienekens G. The food supply in the Netherlands during the Second World War. In: Smith DF, Phillips J (eds), *Food, Science, Policy and Regulation in the Twentieth Century: International and Comparative Perspectives.* Routledge: London, 2000: 117-130.
30. Okasha M, Gunnell D, Holly J, *et al.* Childhood growth and adult cancer. *Best Pract Res Clin Endocrinol Metab.* 2002; 16: 225-41.
31. Yee D, Lee AV. Crosstalk between the insulin-like growth factors and estrogens in breast cancer. *J Mammary Gland Biol Neoplasia.* 2000; 5: 107-15.
32. Elias SG, Peeters PH, Grobbee DE, van Noord PA. Transient caloric restriction and cancer risk (The Netherlands). *Cancer Causes Control.* 2007; 18: 1-5.
33. Thompson HJ, Zhu Z, Jiang W. Protection against cancer by energy restriction: all experimental approaches are not equal. *J Nutr.* 2002; 132: 1047-9.
34. Waterland RA. Epigenetic mechanisms and gastrointestinal development. *J Pediatr.* 2006; 149: S137-42.

35. Heijmans BT, Tobi EW, Stein AD, *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008; 105: 17046-9.
36. Must A, Jacques PF, Dallal GE, Bajema CJ, Dietz WH. Long-term morbidity and mortality of overweight adolescents. A follow-up of the Harvard Growth Study of 1922 to 1935. *N Engl J Med.* 1992; 327: 1350-5.
37. Hughes LA, van den Brandt PA, de Bruine AP, *et al.* Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. *PloS One.* 2009; 4: e7951.
38. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006; 38: 787-93.

3

4

Early life exposure to famine and colorectal cancer risk:

A role for epigenetic mechanisms?

Laura AE Hughes
Piet A van den Brandt
Adriaan P de Bruïne
Kim AD Wouters
Sarah Hulsmans
Angela Spiertz
R Alexandra Goldbohm
Anton FPM de Goeij
James G Herman
Matty P Weijnenberg
Manon van Engeland

ABSTRACT

Background: Exposure to energy restriction during childhood and adolescence is associated with a lower risk of developing colorectal cancer (CRC). Epigenetic dysregulation during this critical period of growth and development may be a mechanism to explain such observations. Within the Netherlands Cohort Study on diet and cancer, we investigated the association between early life energy restriction and risk of subsequent CRC characterized by the (promoter) CpG island methylation phenotype (CIMP).

Methods: Information on diet and risk factors was collected by baseline questionnaire (N=120,856). Three indicators of exposure were assessed: place of residence during the Hunger Winter (1944-45) and World War 2 years (1940-44), and father's employment status during the Economic Depression (1932-40). Methylation specific PCR (MSP) on DNA from paraffin embedded tumour tissue was performed to determine CIMP status according to the Weisenberger markers. After 7.3 years of follow-up, 603 cases and 4631 sub-cohort members were available for analysis. Cox regression was used to calculate hazard ratios (HR) and 95% confidence intervals for CIMP+ (27.7%) and CIMP- (72.3%) tumours according to the three time periods of energy restriction, adjusted for age and gender.

Results: Individuals exposed to severe famine during the Hunger Winter had a decreased risk of developing a tumour characterized by CIMP compared to those not exposed (HR: 0.65, 95%CI: 0.45-0.92). Further categorizing individuals by an index of '0-1' '2-3' or '4-7' genes methylated in the promoter region suggested that exposure to the Hunger Winter was associated with the degree of promoter hypermethylation ('0-1 genes methylated' HR: 1.01, 95%CI: 0.74-1.37; '2-3 genes methylated' HR: 0.83, 95% CI: 0.61-1.15; '4-7 genes methylated' HR: 0.72, 95 % CI: 0.49-1.04). No associations were observed with respect to the Economic Depression and WW2 years.

Conclusions: This is the first study indicating that exposure to a severe, transient environmental condition during adolescence and young adulthood may result in persistent epigenetic changes that later influence CRC development.

INTRODUCTION

It has long been hypothesized that the risk of adult disease is associated with dietary patterns experienced early in development.¹⁻⁵ Increasing evidence supports a role for epigenetic mechanisms in explaining how such early life exposures influence disease risk.⁶⁻¹⁰

There are indications that severe, short-term energy restriction can induce persistent changes in epigenetic information. It was recently reported that individuals who were prenatally exposed to famine during the ‘Dutch Hunger Winter’ of 1944-45, had less DNA methylation of the imprinted IGF-2 gene 60 years later compared to their unexposed, same-sex sibling.⁶ While this study did not look at specific adult disease as an endpoint, it showed for the first time that transient dietary patterns early in human gestation can induce permanent epigenetic changes. It also strengthened the hypothesis that likely, there are periods of temporal susceptibility to epigenetic modulations. Evidence suggests that windows of susceptibility also exist later in development,¹¹⁻¹³ but this has never been investigated with respect to energy restriction.

In a large Dutch population, we have observed that individuals who experienced severe caloric restriction as adolescents through exposure to the Hunger Winter have a lower risk of developing colorectal cancer (CRC) later in life.^{14, 15} CRC is one of the best described cancers in terms of genetic and epigenetic events involved.¹⁶⁻¹⁹ A distinct characteristic of epigenetic instability in CRC is the CpG island methylator phenotype (CIMP), characterized by numerous promoter CpG island hypermethylated tumour suppressor- and DNA repair genes.²⁰⁻²⁴ This in turn is associated with transcriptional silencing of gene expression.²⁵ Furthermore, many CIMP tumours also exhibit microsatellite instability (MSI),²⁶ a form of genetic instability associated with defective mismatch repair that is commonly achieved by promoter methylation of the mismatch repair gene *hMLH1*.¹⁶

The Netherlands Cohort Study on diet and cancer (NLCS) offers a rare opportunity to study the association between energy restriction during childhood and adolescence and epigenetic changes associated with CRC. This large and well defined cohort is comprised of individuals who grew up during the Hunger Winter, as well as two preceding periods of more moderate energy restriction: World War 2 (1940-44), and the Economic Depression (1932-40). Although individual food intake data is not available

for these three time periods, a number of proxy measures have been collected from cohort members that reflect their exposure to energy restriction. Additionally, colorectal tumour tissue samples have been collected from cases and subsequently analyzed for a wide range of molecular characteristics. Using this population, the present study investigated the association between energy restriction in childhood and adolescence and the risk of having a CIMP tumour later in life. Because CIMP tumours are associated with MSI, the relationship between energy restriction during childhood and adolescence and having a tumour characterized by MSI was also investigated.

METHODS

Ethics Statement

The study protocol of the NLCS was approved by the Medical Ethics Committees of the University Hospital Maastricht and TNO Nutrition in February 1985 and July 1986, respectively. At the start of the study in 1986, participants were informed in writing of the details of the study and its objectives. In accordance with the regulations of the time, written informed consent was assumed when participants returned the completed baseline questionnaire. Paraffin embedded tumour material of the CRC patients was collected after approval by the ethical review boards of Maastricht University, the NCR (national cancer registry), and a national pathology database (PALGA) in 1997.

Study population and design

The NLCS is a prospective cohort study that was initiated in 1986 with the purpose of investigating the association between diet and the development of cancer. The cohort included 58,279 men and 62,573 women between the ages of 55-69 years at baseline who completed a self-administered questionnaire on dietary habits, and lifestyle, health, and demographic information. Municipal registries from throughout the Netherlands were used to constitute an efficient sampling frame.²⁷⁻²⁹ The NLCS uses a case-cohort approach for data processing and analysis; a subcohort of 5000 individuals was randomly selected from the larger cohort upon recruitment into the study. These individuals have been followed-up biennially from baseline in 1986 for migration and vital status to estimate

person time at risk for the entire cohort. All subcohort members who reported prevalent cancer (excluding skin cancer) at baseline were excluded from analyses (leaving N=4650). Further details of the design of the NLCS have been described in previous publications.²⁷⁻²⁹

Incident CRC cases were identified by annual record linkage to nine regional cancer registries and PALGA. This method of record linkage has been described elsewhere.³⁰ The completeness of cancer follow-up has been estimated to be almost 100%.³¹ CRCs were classified as proximal colon cancer (International Classification of Diseases for Oncology, first edition codes 153.0, 153.1, 153.4, 153.5 and 153.6), distal colon cancer (codes 153.2, 153.3 and 153.7), rectosigmoid (code 154.0) and rectal cancer (code 154.1). All prevalent cancer cases at baseline other than non-melanoma skin cancer and subjects with incomplete or inconsistent questionnaires were excluded.

Colorectal tumour samples were retrieved from August 1999 to December 2001, as described previously.³² In total, 734 incident CRC patients were identified from a follow-up period of 7.3 years after baseline, of whom a PALGA report of the lesion as well as sufficient DNA was available.³²

4

Assessment of Energy Restriction

Individual food intake data of the cohort members was not available for the three periods of energy restriction. Therefore, proxy variables were used to describe exposure to energy restriction.

With respect to the Economic Depression years, the best proxy available was the employment status of the cohort member's father during this time frame. It was documented that families with an unemployed father consumed fewer calories and had less variation in their food patterns than families with an employed father.^{14, 33} On the baseline questionnaire, individuals were asked 'was your father employed during the crisis of 1932-1940?' Cohort members could either answer yes, along with the years that he was employed, or no. For analyses purposes, answers were categorized into 'father was employed during the economic depression' or 'father was not employed during the economic depression'.

For the years spanning both WW2 and the Hunger Winter, place of residence was used to approximate the exposure to energy restriction. During WW2, food rationing was introduced and because of poor food availability in

large cities, nutritional differences emerged between people living in urban versus rural areas. It was documented that individuals living in a city with more than 40,000 inhabitants consumed fewer calories daily, and ate diets consisting of a higher percentage of carbohydrates (70% vs 65%) and a lower percentage of fat (10% vs 15%).^{14, 34} Therefore, place of residence in 1942 (midpoint of the war years) was used as a proxy for energy restriction during WW2. On the baseline questionnaire, cohort members were asked to list the last four residences they lived in, the province where these residences were located, and the years that they lived in these locations. The location of residence during the mid-point of WW2 was dichotomized to ‘lived in a city in 1942’, or ‘lived in a rural area in 1942’.

4 Additionally, cohort members were specifically asked where they lived during the winter of 1944-45. Residing in a Western city during the Hunger Winter was considered an indicator for severe energy restriction. While the diet remained nutritionally balanced, individuals living in western urban areas experienced rationing of less than 700 kcal per day.^{35, 36} Eleven Western cities were considered famine cities based on the definition by Stein *et al.*³⁷: Amsterdam, Rotterdam, The Hague, Utrecht, Zaandam, Hilversum, Amersfoort, Dordrecht, Vlaardingen/Schiedam, Delft, and Leiden. Answers to this question were classified into three categories: lived in a Western city, lived in a Western rural area, and lived in a non-Western area of the Netherlands during the winter of 1944-1945.

Promoter methylation analyses

CIMP in tumour tissue of CRC cases was defined by CpG island promoter hypermethylation of at least 3 out of 5 methylation markers (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1*), as proposed by Weisenberger *et al.*²³ Methylation status of these markers, as well as that of the *MLH1* and *APC* gene; used to create a methylation index with the CIMP markers, were determined by bisulfite modification of 500 ng genomic DNA using a commercially available kit (Zymo Research), and subsequent methylation specific PCR (MSP).^{38, 39} We chose to use MSP as a method because it is effective, specific and does not require specific equipment. It has been shown that results from MSP are in accordance with other technologies, such as MethyLight.²⁰ To facilitate MSP analysis on DNA retrieved from formalin-fixed, paraffin-embedded tissue, DNA was first amplified with

flanking PCR primers that amplify bisulfite-modified DNA but do not preferentially amplify methylated or unmethylated DNA. The resulting fragment was used as a template for the MSP reaction. All PCRs were carried out with controls for unmethylated alleles (DNA from normal lymphocytes), methylated alleles [normal lymphocyte DNA treated *in vitro* with *SssI* methyltransferase (New England Biolabs, Ipswich, MA)] and a control without DNA. Ten microliters of each MSP reaction was directly loaded on to nondenaturing 6% polyacrylamide gels, stained with ethidium bromide and visualised under UV illumination. The MSP analyses were successful for 81%, 79%, 79%, 90%, 83%, 93%, and 93% out of the 734 cases for *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1*, *MLH1*, and *APC* respectively.

Microsatellite instability analyses

MSI was determined by a pentaplex PCR using the MSI markers BAT-26, BAT-25, NR-21, NR-22 and NR-24, as described in detail by Suraweera *et al.*⁴⁰ MSI analyses were successful on 662 (90%) of the 734 available DNA samples.

Data analyses

Data were analyzed with Stata (version 9.2, Statacorp, College Station, TX, USA). Cox proportional hazards analysis using the case-cohort approach was used to obtain hazard ratios (HR) and 95% confidence intervals (CI) for the association between the three proxy measures of energy restriction and CRC characterized by CIMP and MSI status. The proportional hazards assumption was tested using the scaled Schoenfeld residuals and visual inspection of the hazard curves. To account for the additional variance introduced by sampling the subcohort from the entire cohort, standard errors were estimated using the robust option.

Potential confounding variables considered for multivariate analysis were age, sex, BMI (kg/m²), height (cm), family history of CRC (yes/no), smoking status (never smoker, ex-smoker, current smoker), level of education (primary school, junior high school, senior high school, higher vocational school, or university), total energy intake (kcal/day), alcohol intake (g/day), recreational physical activity (<30 minutes/day, 30-60 minutes/day, 61-90 minutes/day, >90 minutes/day), and consumption of red

meat, fruit, vegetables, and grains (g/day). For women, hormonal factors such as contraceptive use (yes/no), age at menarche, age at menopause, and number of children were also considered. Those variables that changed the risk estimates for the association of energy restriction and CRC by more than 10% were ultimately included as confounders. After adjustment for age and sex, effect estimates remained extremely stable. HRs were therefore estimated using a model that only included age and sex, and expressed using ‘father was employed during the economic depression’, ‘lived in a rural area in 1942’, and ‘lived in a non-western area of the Netherlands during the Hunger Winter’ as the reference categories for the different proxies of energy restriction.

4 To assess how exposure to energy restriction was associated with the extent of promoter methylation in the CRC tumours, a methylation index was created that consisted of the five CIMP markers (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, and *SOCS1*), and the *MLH1* and *APC* genes. Cases were categorized into one of three groups: ‘0-1 genes methylated’, ‘2-3 genes methylated’, or ‘4-7 genes methylated’. Of the 734 cases, 556 had sufficient information to be classified into one of the three categories.

Tests for heterogeneity were done to evaluate differences between subtypes of tumours (e.g., *CIMP* vs. *non-CIMP*) using the competing risks procedure in STATA. However, the SE for the difference of the log-hazard ratios from this procedure assumes independence of both estimated hazard ratios, which would overestimate that SE and thus overestimate the *P* values for their difference. Therefore, these *P* values and the associated confidence intervals were estimated based on a bootstrapping method that was developed for the case-cohort design.⁴¹ For each bootstrap sample, *X* subcohort members were randomly drawn from the subcohort of *X* subjects and *Y* cases from the total of *Y* cases outside the subcohort, both with replacement, out of the data set of *X* + *Y* observations. The log-hazard ratios were obtained from this sample using STATA's competing risks procedure and recalculated for each bootstrap replication. The confidence interval and *P* value of the differences in hazard ratio of the subtypes were then calculated from the replicated statistics using the accelerated bias corrected method in STATA. Each bootstrap analysis was based on 1,000 replications.⁴²

RESULTS

The distribution of exposure variables for the study population is presented in **Table 1**.

For all of the case groups, there were proportionally more individuals residing in a non-western area during the Hunger Winter compared to subcohort members. Proportionally, there were fewer non-CIMP and non-MSI cases residing in an urban area during WW2 than rural area than CIMP+ cases, MSI+ cases, and sub-cohort members. With respect to the Economic Depression, more MSI+ cases had an employed father compared to the other case groups, and sub-cohort members.

Table 1. Distributions of energy restriction exposure variables for CRC cases characterized by CIMP^a, MSI, and sub-cohort members in the Netherlands Cohort Study on diet and cancer after 7.3 years of follow-up

	sub-cohort (N=4650) ^b	CRC cases characterized by:			
		CIMP+ (N=167) ^b	Non-CIMP (N=436) ^b	MSI+ (N=84) ^b	MSI- (N=578) ^b
Exposure variables					
Hunger Winter 1944-45^e					
Non-Western	2369 (58.3) ^c	100 (68.5) ^d	224 (60.9) ^d	46 (62.2) ^d	304 (62.5) ^d
Western Rural	626 (15.4)	13 (8.9)	49 (13.3)	9 (12.2)	64 (13.7)
Western City	1067 (26.3)	33 (22.6)	95 (25.8)	19 (25.7)	118 (24.3)
War Years 1940-44^e					
Rural Area in 1942	1648 (48.8)	58 (47.9)	179 (52.5)	32 (49.2)	233 (52.7)
Urban area in 1942	1732 (51.2)	63 (52.1)	162 (47.5)	33 (50.8)	209 (47.3)
Economic Depression 1932-40^e					
Father had a job	3864 (88.3)	142 (88.1)	371 (90.3)	75 (92.6)	485 (89.3)
Father did not have a job	508 (11.6)	19 (11.8)	40 (9.7)	6 (7.4)	58 (10.7)

^a CIMP status defined as $\geq 3/5$ genes methylated according to the Weisenberger markers (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, and *SOCS1*)

^b due to missing data for exposure variables, numbers may not add up to 4650, 167, 84 and 578, respectively

^c presented as number of sub-cohort members (% of total number of sub-cohort members in the exposure category)

^d presented as number of cases (% of total number of cases in the exposure category)

^e *Hunger Winter*: it was assumed that individuals residing in a western area of the Netherlands, especially a western urban area during the Hunger Winter would have experienced the greatest degree of energy restriction; *War Years*: it was assumed that individuals who resided in an urban area during the mid-point of the War would have experienced the greatest degree of energy restriction; *Economic Depression*: it was assumed that individuals who had an unemployed father during the Economic Depression would have experienced a higher degree of energy restriction compared to those whose father was employed.

Associations between the three proxy measures of energy restriction and risk of tumours with specific characteristics are presented in **Table 2**. Individuals who reported living in a western area of the Netherlands during the Hunger Winter had a statistically significant lower risk of developing a CIMP tumour compared to those who did not (HR: 0.65, 95% CI: 0.45-0.92). Inverse associations remained when considering 'western area' as either western rural area (HR: 0.51, 95% CI: 0.28-0.91), or western city (HR: 0.72, 95%CI: 0.48-1.08). There was no association between exposure to energy restriction during WW2 and the Economic Depression and having a CIMP tumour. Tests for heterogeneity showed no significant difference between the HRs observed for tumours with and without CIMP. Inverse associations were observed with respect to the three time periods of energy restriction and risk of a tumour with MSI, but these were far from statistically significant. Tests for heterogeneity showed no significant difference between the HRs observed for tumours with and without MSI.

HRs for the three proxies of energy restriction according to a methylation index of 7 genes are presented in **Table 3**. Of the 556 cases with sufficient information to be categorized, 37% were identified as having 0-1 genes methylated, 37% as having 2-3 genes methylated, and 26% as having > 4-7 genes methylated. Categorizing individuals by '0-1 genes methylated', '2-3 genes methylated' and '4-7 genes methylated', suggested that residing in a western area during the Hunger Winter was associated with the degree of promoter hypermethylation ('0-1 genes methylated' HR: 1.01, 95%CI: 0.74-1.37; '2-3 genes methylated' HR: 0.83, 95% CI: 0.61-1.15; '4-7 genes methylated' HR: 0.72, 95 % CI: 0.49-1.04) compared to individuals who did not reside in a western area.

Table 2. Age and sex adjusted and hazard rate ratios (HR) for CRC characterized by CIMP and MSI, according to three time periods of energy restriction in the Netherlands Cohort Study on diet and cancer (7.3 years of follow-up)

	Hunger Winter (1944-45) ^c			War Years (1940-44) ^c		Economic Depression(1932-40)		
	Non-western ^a	Western rural	Western city	Western areas combined ^b	Rural area in 1942 ^a	Urban area in 1942	Father had a job ^a	Father did not have a job ^a
Person years at risk	16,894	4483	7630	12,133	11,737	12,391	27,575	3608
CRC endpoints characterized by:								
CIMP								
N cases	100	13	33	46	58	63	142	19
HR (95% CI)	1 (referent)	0.51 (0.28-0.91)	0.72 (0.48-1.08)	0.65 (0.45-0.92)	1 (referent)	0.99 (0.69-1.43)	1 (referent)	0.98 (0.61-1.60)
Non-CIMP								
N cases	224	49	95	114	179	162	371	40
HR (95% CI)	1 (referent)	0.85 (0.62-1.18)	0.94 (0.73-1.21)	0.91 (0.73-1.23)	1 (referent)	0.84 (0.67-1.05)	1 (referent)	0.81 (0.57-1.13)
MSI								
N cases	46	9	19	28	32	33	75	6
HR (95% CI)	1 (referent)	0.77 (0.37-1.59)	0.90 (0.52-1.54)	0.85 (0.53-1.37)	1 (referent)	0.94 (0.57-1.54)	1 (referent)	0.58 (0.25-1.34)
Non-MSI								
N cases	304	64	118	182	233	209	485	58
HR (95% CI)	1 (referent)	0.82 (0.62-1.09)	0.86 (0.68-1.08)	0.84 (0.69-1.03)	1 (referent)	0.83 (0.68-1.01)	1 (referent)	0.89 (0.66-1.19)

^a Reference category

^b Category created from combining individuals who lived in a western rural area and a western urban area

Table 3. Age and sex adjusted hazard rate ratios (HR) for a methylation index according to three time periods of energy restriction in the Netherlands Cohort Study on diet and cancer (7.5 years of follow-up)

	Hunger Winter (1944-45)				War Years (1940-44)			Economic Depression(1932-40)	
	Non-western ^a	Western rural	Western city	Western areas combined ^b	Rural area in 1942 ^a	Urban area in 1942	Father had a job ^a	Father did not have a job	
PY at risk	16,894	4483	7630	12,133	12,391	11,737	27,575	3608	
Number of genes methylated:									
0-1									
N cases	102	24	49	73	82	85	171	20	
HR (95% CI)	1 (referent)	0.92 (0.58-1.46)	1.06 (0.74-1.50)	1.01 (0.74-1.37)	1 (referent)	0.95 (0.69-1.30)	1 (referent)	0.86 (0.54-1.39)	
2-3									
N cases	107	24	39	63	90	66	180	18	
HR (95% CI)	1 (referent)	0.87 (0.55-1.37)	0.82 (0.56-1.19)	0.84 (0.61-1.15)	1 (referent)	0.68 (0.49-0.94)	1 (referent)	0.75 (0.46-1.24)	
4-7									
N cases	84	11	32	43	49	55	122	15	
HR (95% CI)	1 (referent)	0.51 (0.27-1.30)	0.83 (0.55-1.26)	0.72 (0.49-1.04)	1 (referent)	1.02 (0.68-1.52)	1 (referent)	0.90 (0.52-1.56)	

^a Reference category

^b Category created from adding together individuals who lived in a western rural area and a western urban area

DISCUSSION

To our knowledge, this study is the first to show that energy restriction during adolescence and early adulthood is associated with the CIMP phenotype in CRC, suggesting that exposure to a transient environmental condition during this period of life may result in persistent epigenetic changes that later influence CRC development. We observed that individuals exposed to the Hunger Winter of 1944-45, a period of severe short-term energy restriction, had a decreased risk of developing a CIMP tumour later in life compared to individuals who were not exposed. Additionally, considering a methylation index suggested that the degree of hypermethylation was inversely associated with exposure to energy restriction. These observations, along with the fact that there was no association between exposure and risk of MSI, suggest that energy restriction during childhood and adolescence is associated with CIMP risk independently of MSI. Furthermore, the strong associations we observed during the period of the Hunger Winter suggest that it may be severe caloric restriction as opposed to specific nutrient restriction that modulates the risk of colorectal tumours with the CIMP phenotype, as we observed no associations during the other two periods of exposure when caloric restriction was assumed to be less severe.

Before discussing our results further, it is important to address some important limitations in our study. Although the number of CRC cases after 7.3 years of follow up in the NLCS was substantial, the number of these cases with the CIMP and or MSI phenotype was small. Because information on individual food for the cohort was not available for the exposure periods, three proxy measures of energy restriction were used to describe energy restriction during the Economic Depression, WW2, and the Hunger Winter. It is possible that using the father's employment status, place of residence during WW2, and place of residence during the Hunger Winter has resulted in some misclassification of exposure; however, in a follow-up study in the NLCS, female sub-cohort members were asked if they really had experienced hunger during the winter of 1944-1945. Of the women who reported severe hunger, 80% lived in a western city during this winter.⁴³ These results indicate that place of residence during the Hunger Winter is a reasonably adequate proxy of exposure, although unavoidable misclassification may have caused attenuation of the association. While it has been proposed that

individual famine data may be more accurate than area-exposure data,⁴⁴ other studies in the Netherlands have used place of residence during the Hunger Winter as a measure of energy restriction^{45, 46} and their results indicate that it is an adequate proxy.

Although there is thought that environmental influences on epigenetics are likely to be most important during prenatal and early postnatal development, when epigenetic mechanisms undergo establishment and maturation,¹⁰ our findings strengthen the hypothesis that childhood and adolescence may also be periods of temporal susceptibility to epigenetic modulations.^{12, 13} Perhaps, this is especially true for tissue specific changes, such as in the colorectal tract where cells are constantly being regenerated. The epigenetic process involved in normal gastrointestinal development, including the expression of gut enzymes, hormones, transporters, and the maturation of gastrointestinal immunoregulation, is an understudied area of research.⁸ Therefore, we can only speculate how an environmental exposure, such as energy restriction, during childhood and adolescence might induce persistent epigenetic changes that later influence methylation patterns in CRC.

Waterland *et al.* propose that environmental influences on the developmental establishment of DNA methylation can occur through two mechanisms: by affecting the supply of dietary methyl donors and/or activity of DNA methyltransferases to induce either hyper- or hypomethylation at metastable epialleles, or by altering transcriptional activity of specific genes during ontogenic periods when DNA methylation is being established.¹⁰ The pre-pubertal and pubertal years are a period of rapid growth and hormonal change, and there could be a link between the influx of growth hormones and regulation of epigenetic mechanisms controlling cancer causing genes. Growth hormone affects cellular growth through the actions of insulin growth factors, and it has been suggested that the beneficial effect of caloric restriction on CRC risk is mediated through decreasing IGF-1.⁴⁷ During puberty, IGF-1 can be as high as four times the normal adult serum concentration.⁴⁷ It could be possible that energy restriction during this critical period of growth may permanently influence the growth hormone-IGF axis, subsequently influencing methylation patterns later in life. Alternatively, imprinted genes, which are vital for human development, may be involved in modulating future risk. Loss of imprinting of the IGF-2 gene is an example of

epigenetic dysregulation in gastrointestinal carcinogenesis.⁸ While it has been reported that adult diet does not correlate with IGF-2 loss of imprinting, a study in mice has shown that genomic imprinting may be modifiable through altering the diet in the postnatal period.⁴⁸ It would be interesting to determine if genomic imprinting is also modifiable during childhood and adolescence.

Our observations provide insight into how energy restriction early in life may influence the development of colorectal cancer,^{15, 49-52} and also generate hypotheses for future research. In contrast to the nutritional circumstances of the Netherlands during the mid-20th century, most Western countries now find themselves in a situation of having too much, rather than too little. It has been observed that being overweight during adolescence may be a more powerful predictor of CRC risk than being overweight during adulthood.⁵³ Determining if energy excess during childhood and adolescence influences epigenetic mechanisms in a way that increases future disease risk is important to elucidate; both in developed countries where the prevalence of childhood obesity is reaching epidemic proportions, and in developing countries where transitioning nutritional habits and exposures may have implications on future CRC trends.

In conclusion, while a dark moment in Dutch history, unique evidence from the Hunger Winter shows that exposure to a period of severe transient energy restriction during adolescence is inversely associated with the risk of having a CIMP tumour later in life. These findings suggest that adolescence may be a critical period of development for epigenetic changes that influence CRC risk.

REFERENCES

1. Barker DJ. The fetal and infant origins of adult disease. *Bmj*. 1990; 301: 1111.
2. Barker DJ. Fetal growth and adult disease. *Br J Obstet Gynaecol*. 1992; 99: 275-6.
3. Barker DJ. Intrauterine programming of adult disease. *Mol Med Today*. 1995; 1: 418-23.
4. Rose G. Familial Patterns in Ischaemic Heart Disease. *Br J Prev Soc Med*. 1964; 18: 75-80.
5. Trichopoulos D. Hypothesis: does breast cancer originate in utero? *Lancet*. 1990; 335: 939-40.
6. Heijmans BT, Tobi EW, Stein AD, *et al*. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*. 2008; 105: 17046-9.
7. Morgan HD, Sutherland HG, Martin DI, *et al*. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet*. 1999; 23: 314-8.
8. Waterland RA. Epigenetic mechanisms and gastrointestinal development. *J Pediatr*. 2006; 149: S137-42.
9. Waterland RA, Jirtle RL. Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. *Nutrition*. 2004; 20: 63-8.
10. Waterland RA, Michels KB. Epigenetic epidemiology of the developmental origins hypothesis. *Annu Rev Nutr*. 2007; 27: 363-88.
11. Bjornsson HT, Sigurdsson MI, Fallin MD, *et al*. Intra-individual change over time in DNA methylation with familial clustering. *Jama*. 2008; 299: 2877-83.
12. Fraga MF, Ballestar E, Paz MF, *et al*. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 2005; 102: 10604-9.
13. Weaver IC, Cervoni N, Champagne FA, *et al*. Epigenetic programming by maternal behavior. *Nat Neurosci*. 2004; 7: 847-54.
14. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al*. Energy restriction early in life and colon carcinoma risk: results of The Netherlands Cohort Study after 7.3 years of follow-up. *Cancer*. 2003; 97: 46-55.
15. Hughes LA, van den Brandt PA, Goldbohm RA, *et al*. Childhood and adolescent energy restriction and subsequent colorectal cancer risk:

- results from the Netherlands Cohort Study. *Int J Epidemiol.* 2010; 39: 1333-44.
16. Derks S, Postma C, Carvalho B, *et al.* Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis.* 2008; 29: 434-9.
 17. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer.* 2004;4:988-93.
 18. Sieber OM, Heinimann K, Tomlinson IP. Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer.* 2003; 3: 701-8.
 19. Wong JJ, Hawkins NJ, Ward RL. Colorectal cancer: a model for epigenetic tumorigenesis. *Gut.* 2007; 56: 140-8.
 20. Barault L, Charon-Barra C, Jooste V, *et al.* Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res.* 2008; 68: 8541-6.
 21. Samowitz WS, Albertsen H, Herrick J, *et al.* Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology.* 2005; 129: 837-45.
 22. Toyota M, Ahuja N, Ohe-Toyota M, *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A.* 1999; 96: 8681-6.
 23. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006; 38: 787-93.
 24. Shen L, Toyota M, Kondo Y, *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A.* 2007; 104: 18654-9.
 25. Ferguson LR, Karunasinghe N, Philpott M. Epigenetic events and protection from colon cancer in New Zealand. *Environ Mol Mutagen.* 2004; 44: 36-43.
 26. Slattery ML, Wolff RK, Curtin K, *et al.* Colon tumour mutations and epigenetic changes associated with genetic polymorphism: insight into disease pathways. *Mutat Res.* 2009; 660: 12-21.

27. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr.* 1994; 48: 253-65.
28. Goldbohm RA, van 't Veer P, van den Brandt PA, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr.* 1995; 49: 420-9.
29. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol.* 1990; 43: 285-95.
30. Van den Brandt PA, Schouten LJ, Goldbohm RA, *et al.* Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol.* 1990; 19: 553-8.
31. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of municipalities by cancer registries and PALGA using hospital discharge data. *Tijdschr Soc Gezondheidsz.* 1994; 72: 80-4.
32. Brink M, de Goeij AF, Weijnenberg MP, *et al.* K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis.* 2003; 24: 703-10.
33. den Hartog A. Werklozen en hun voeding in de jaren 1930-39 (Unemployed people and their nutrition during the years 1930-39 [in Dutch]). *Voeding.* 1983; 44: 92-8.
34. Dols M, van Arcken A. De voedselvoorziening in Nederlands tijdens en onmiddellijk na den tweeden wereldoorlog 1940-1945 (Food supply during and after the second World War 1940-1945 in the Netherlands [in Dutch]). *Voeding.* 1946; 6: 193-207.
35. Burger G, Stanstead H, Drummond J. Malnutrition and starvation in Western Netherlands, September 1944-45. Part I and II. The Hague: General State Printing Office 1948.
36. Elias SG, Keinan-Boker L, Peeters PH, *et al.* Long term consequences of the 1944-1945 Dutch famine on the insulin-like growth factor axis. *Int J Cancer.* 2004; 108: 628-30.
37. Stein Z, Susser M, Saenger G, *et al.* *Famine and human development, the Dutch Hunger winter of 1944-1945.* New York: Oxford University Press 1975.

38. Derks S, Lentjes MH, Hellebrekers DM, *et al.* Methylation-specific PCR unraveled. *Cell Oncol.* 2004; 26: 291-9.
39. Herman JG, Graff JR, Myohanen S, *et al.* Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A.* 1996; 93: 9821-6.
40. Suraweera N, Duval A, Reperant M, *et al.* Evaluation of tumour microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology.* 2002; 123: 1804-11.
41. Wacholder S, Gail MH, Pee D, *et al.* Alternative variance and efficiency calculations for the case-cohort design. *Biometrika.* 1989; 76: 117-23.
42. de Vogel S, Bongaerts BW, Wouters KA, *et al.* Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. *Carcinogenesis.* 2008; 29: 1765-73.
43. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Diet in adolescence and the risk of breast cancer: results of the Netherlands Cohort Study. *Cancer Causes Control.* 1999; 10: 189-99.
44. Elias SG, Peeters PH, Grobbee DE, *et al.* Breast cancer risk after caloric restriction during the 1944-1945 Dutch famine. *J Natl Cancer Inst.* 2004; 96: 539-46.
45. Lumey LH, Stein AD. Offspring birth weights after maternal intrauterine undernutrition: a comparison within sibships. *Am J Epidemiol.* 1997; 146: 810-9.
46. van Noord PA, Kaaks R. The effect of wartime conditions and the 1944-45 'Dutch famine' on recalled menarcheal age in participants of the DOM breast cancer screening project. *Ann Hum Biol.* 1991; 18: 57-70.
47. Okasha M, Gunnell D, Holly J, *et al.* Childhood growth and adult cancer. *Best Pract Res Clin Endocrinol Metab.* 2002; 16: 225-41.
48. Waterland RA, Lin JR, Smith CA, *et al.* Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. *Hum Mol Genet.* 2006; 15: 705-16.

49. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Energy restriction in childhood and adolescence and risk of prostate cancer: results from the Netherlands Cohort Study. *Am J Epidemiol.* 2001; 154: 530-7.
50. Frankel S, Gunnell DJ, Peters TJ, *et al.* Childhood energy intake and adult mortality from cancer: the Boyd Orr Cohort Study. *Bmj.* 1998; 316: 499-504.
51. Svensson E, Grotmol T, Hoff G, *et al.* Trends in colorectal cancer incidence in Norway by gender and anatomic site: an age-period-cohort analysis. *Eur J Cancer Prev.* 2002; 11: 489-95.
52. Svensson E, Moller B, Tretli S, *et al.* Early life events and later risk of colorectal cancer: age-period-cohort modelling in the Nordic countries and Estonia. *Cancer Causes Control.* 2005; 16: 215-23.
53. Must A, Jacques PF, Dallal GE, *et al.* Long-term morbidity and mortality of overweight adolescents. A follow-up of the Harvard Growth Study of 1922 to 1935. *N Engl J Med.* 1992; 327: 1350-5.

5

Body size and colorectal cancer risk after 16.3 years follow-up:

An analysis from the Netherlands Cohort Study

Laura AE Hughes
Colinda CJM Simons
Piet A van den Brandt
R Alexandra Goldbohm
Manon van Engeland
Matty P Weijenberg

Submitted for publication

ABSTRACT

Background: Body size is a well documented risk factor for colorectal cancer (CRC), but how risk differs according to tumour sub-site in men and women remains unclear.

Methods: The Netherlands Cohort Study includes 120,852 men and women who were 55-69 years and completed a self-administered questionnaire at baseline (1986). Using the case-cohort approach, sex-specific hazard ratios (HR) and 95% confidence intervals (CI) were calculated for total colorectal, proximal colon, distal colon, rectosigmoid, and rectal cancers, according to BMI, BMI at age 20, BMI change, trouser/skirt size (as a proxy for waist circumference), and height. After 16.3 years of follow-up, 2316 cases were available for multivariate analysis.

Results: In men, BMI (per 5kg/m² increase) and trouser size (per 2 unit increase) were both associated with total CRC (HR: 1.24, 95%CI: 1.05-1.46 and HR: 1.12, 95% CI: 1.05-1.19; *P*-trend across quintiles= 0.02, respectively). Specifically, BMI was associated with tumours of the distal colon (per 5kg/m² HR: 1.42, 95% CI: 1.13-1.79; *P*-trend across quintiles = 0.05) and rectosigmoid (HR: 1.49, 95%CI: 1.05-2.11; *P*-trend across quintiles = 0.05), and trouser size was associated with the distal colon (HR: 1.16, 95% CI: 1.04-1.29; *P*-trend across quintiles <0.01). Body fat was not associated with CRC risk in women unless considered simultaneously with physical activity; a high trouser/skirt size and low level of physical activity increased risk of all tumour sub-types. Height was associated with CRC in women (per 5 cm increase, HR: 1.09, 95%CI: 1.01-1.17; *P*-trend across quintiles =0.05), but not men.

Conclusion: Body fat is a risk factor for CRC in men, especially for tumours of the distal colon. Height is a risk factor for CRC in women, but associations with body fat may depend on level of physical activity.

INTRODUCTION

There is convincing evidence that a large body size increases the risk of colorectal cancer (CRC).¹ Adiposity, especially when centrally distributed, is known to raise inflammatory responses, increase circulating estrogens, and decrease insulin sensitivity.¹⁻³ Adult attained height is genetically determined,^{4, 5} but is also a marker of an aggregated foetal and childhood experience, and can be considered a proxy for important nutritional exposures that impact on hormonal and metabolic axes later in life.^{1, 6} Meta analyses of prospective cohort data suggest a clear dose-response association between total body fatness (using body mass index (BMI) as an indicator), and abdominal fatness (using waist circumference as an indicator) and CRC.^{1, 7-9} Such an association has also been observed with respect to adult attained height.¹ Data is more consistent and suggests a higher risk for colon cancer than rectal cancer, and recent reviews indicate that associations between body size and CRC are stronger for men than for women.¹⁰⁻¹²

Although the relationship between body size and CRC risk has been well investigated, there are issues requiring further elucidation. It may be more rational to consider risk according to the proximal colon, distal colon, and rectum rather than the traditional division of colon and rectum, because several anatomic, embryologic, and physiologic differences exist between sub-anatomic locations of the colorectal tract, and therefore, arising tumours may have different etiologic pathways.¹³⁻¹⁶ Case-control and prospective data suggests that BMI is more strongly associated with tumours of the distal colon than the proximal colon.¹⁷⁻²⁴ Furthermore, although the incidence of CRC is higher in Westernized countries than other parts of the world,²⁵ as is the prevalence of obesity,²⁶ limited attention has focused on how a 'Western lifestyle' (i.e. large body size, and low levels of physical activity) influence CRC risk simultaneously. To our knowledge, the interaction between body size and physical activity has been examined in one case-control study, in which it was observed that those with a high BMI and a low level of physical activity had a greater risk of colon and rectal cancer compared to those with a low BMI and a high level of physical activity.^{20, 27}

Using data from the Netherlands Cohort Study (NLCS), a large prospective cohort with 16.3 years of follow-up, the aim of the present study was three-fold: 1) to investigate the association between measures of anthropometry and sex-specific risk of CRC; 2) to investigate the association

between measures of anthropometry and sex-specific risk of CRC at specific tumour sub-sites, including the proximal colon, distal colon, rectosigmoid, and rectum; and 3) to assess how body fat influences risk of CRC when considered simultaneously with physical activity.

METHODS

Study populations and design

The NLCS was initiated in 1986 to investigate the association between diet and the development of cancer. It includes 58,279 men and 62,573 women between the ages of 55-69 years at baseline who completed a self-administered questionnaire involving 150 food items as well as questions on dietary habits, lifestyle, health, and demographics. Municipal registries from throughout the Netherlands were used to constitute an efficient sampling frame.²⁸ The NLCS uses a case – cohort approach for data processing and analysis; case subjects were derived from the entire cohort, and the number of person-years at risk for the entire cohort was estimated from a sub-cohort of 5000 men and women who were randomly sampled from the full cohort at baseline. All sub-cohort members who reported prevalent cancer (excluding skin cancer) at baseline were excluded from analyses, leaving 4654. Further details of the NLCS design have been described.²⁸⁻³⁰

Incident CRC cases were identified by annual record linkage to nine regional cancer registries and a national pathology database (PALGA).³¹ The completeness of cancer follow-up is almost 100%.³² CRCs were classified as proximal colon cancer (international Classification of Diseases for Oncology, first edition codes 153.0, 153.1, 153.4, 153.5, and 153.6), distal colon cancer (codes 153.2, 153.3, and 153.7), rectosigmoid (code 154.0), and rectal cancer (code 154.1). All prevalent cancer cases at baseline other than non-melanoma skin cancer and subjects with incomplete or inconsistent baseline questionnaires were excluded. After additionally excluding individuals who were missing variables adjusted for in the multivariate analysis, 3197 sub-cohort members and 2316 CRC cases (men: 327 proximal colon, 427 distal colon, 33 unspecified colon, 125 rectosigmoid, 299 rectum; women: 459 proximal colon, 327 distal colon, 27 unspecified colon, 87 rectosigmoid, 205 rectum) were ultimately included in the analysis.

The study protocol was approved by the Medical Ethics Committees of the University Hospital Maastricht and TNO Nutrition. On recruitment, participants were informed in writing of the details of the study and its objectives. In accordance with the regulations at that time, informed consent was assumed when participants returned the completed baseline questionnaire.

Assessment of body size, diet, and lifestyle factors

Height (cm), body weight (kg), and body weight at age 20 (kg) were self-reported on the baseline questionnaire. From these variables, BMI and BMI at age 20, and BMI change were subsequently calculated. At baseline, individuals were also asked to report their lower body (trouser or skirt) size from their clothing label (Dutch sizes). Trouser/skirt size has been shown to be an adequate proxy measure for waist circumference when predicting cancer risk in the NLCS, and details of how trouser/skirt size corresponds to waist measurements in men and women in this Dutch population has been published.³³

In the present study, both occupational physical activity of the longest/last held job and baseline non-occupational physical activity were used to assess risk, depending on sex (Simons *et al.* 2010, submitted). For analyses with men, occupational physical activity was used. Participants were asked to report job title(s) and job duration(s) on the baseline questionnaire. Assessment of physical activity at work was based on the job held for the longest amount of time. Total energy expenditure was based on a rating system developed by Hettinger *et al.*³⁴ Participants were classified into three energy expenditure groups: <8 kJ/minute, 8– 12 kJ/minute, and ≥ 12 kJ/minute. Non-occupational physical activity was used to assess risk in women, because most women of this generation had not held a job or had worked for only a short period of time, mostly in the distant past. Non-occupational physical activity covers both recreational physical activity and the physical activity involved in getting to and from work (e.g., walking, cycling). It was calculated by adding up the number of minutes spent per day on cycling/walking to work, going to the shops, and walking the dog and the number of hours spent per week on gardening/odd jobs, recreational cycling/walking, and sports/exercise, as reported on the baseline

questionnaire. For purposes of this analysis, this was classified as low (<30 minutes/day), intermediate (30-90 minutes/day), and high (>90 minutes/day).

Statistical analyses

Data were analyzed with Stata (version 10, Statacorp, College Station, TX, USA), separately for men and women. Cox proportional hazards analysis using the case-cohort approach was used to estimate hazard ratios (HR) and 95% confidence intervals (CI) for the association between total CRC risk and BMI (per 5 kg/m² increase, and according to sex specific quintiles), BMI at age 20 (per 5 kg/m² increase, and according to approximate sex specific quintiles), BMI change (<0 kg/m², 0-4 kg/m², 4.0-8 kg/m², and 8.0 kg/m²),³⁵ trouser/skirt size (per 2 unit size increase, and according to sex specific quintiles), and height (per 5 cm increase, and according to sex specific quintiles). To test for linear trend across categories, the median anthropometric variable within quintiles as a continuous variable was used.

Multivariate models were adjusted for age, family history of CRC (yes/no), smoking status (never smoker, ex-smoker, current smoker), socioeconomic status (level of education: primary school, junior high school, senior high school, higher vocational school, or university), total energy intake (kcal/day), alcohol intake (0, 0.1-4, 5-14, 15-29, ≥30g/day), recreational physical activity (women only) (<30 minutes/day, 30-60 minutes/day, 61-90 minutes/day, >90 minutes/day), and occupation physical activity (men only) (<8 kj/h, 8-12 kj/h, ≥12 kj/h at longest held job). Physical activity was adjusted differently for men and women, because we have observed in our cohort that occupational physical activity predicts risk better for males, and recreational physical activity predicts risk better for females (Simons *et al.* 2010 submitted). BMI change was additionally adjusted for BMI at age 20, and height was additionally adjusted for body weight.³⁵ Although previous studies have additionally adjusted waist circumference for height, we chose not to do this due to the strong correlation between trouser/skirt size and height. We also considered an extra model for women that was additionally adjusted for oral contraceptive use (ever, never), duration of oral contraceptive use (years), use of hormone replacement therapy (ever, never), and duration of hormone replacement therapy (years).

This extra adjustment did not influence effect estimates and therefore, is not presented in the results.

Furthermore, we created a two-way interaction between trouser/skirt size and physical activity. Categories of physical activity were based on self-reported recreational physical activity (women) and occupational physical activity (men), and trouser/skirt size was dichotomized based on the median size for men and women. We created this interaction with trouser/skirt size because evidence suggests that fat distribution is more important than body weight or BMI for CRC risk, especially in women.³⁶ We also considered this interaction for colon and rectum tumours separately. Although a previous case-control study considered a three-way interaction that accounted for body fat, physical activity and total energy intake,²⁰ a major limitation of food frequency questionnaires is the under reporting of food intake.³⁷ This is especially a problem with overweight and obese individuals.³⁸ Therefore, we decided not to consider total energy intake as an independent variable or an interaction term in the present study.

For all analyses, the proportional hazards assumption was tested using the scaled Schoenfeld residuals and visual inspection of the hazard curves. Additionally, analyses were repeated according to short and long follow-up by splitting up the follow-up time at 8 years. To account for the additional variance introduced by sampling the sub-cohort from the entire cohort, standard errors were estimated using the robust option. Statistical significance was tested at the 0.05 level.

Tests for heterogeneity were done to evaluate differences between sub-localizations of tumours (e.g. proximal colon vs. distal colon) using the competing risks procedure in STATA. However, the SE for the difference of the log-hazard ratios from this procedure assumes independence of both estimated hazard ratios, which would overestimate that SE and thus overestimate the *P* values for their difference. Therefore, these *P* values and the associated confidence intervals were estimated based on a bootstrapping method that was developed for the case-cohort design, as described previously.³⁹ Each bootstrap analysis was based on 1000 replications

RESULTS

Baseline characteristics of the sub-cohort are presented in **Tables 1** and **2** for men and women, respectively.

Table 1: Baseline (1986) characteristics of male sub-cohort members of the NLCS according to quintiles of BMI^a

	Quintile of body mass index (BMI) (kg/m ²)				
	1 (lowest)	2	3	4	5 (highest)
Age at baseline ^b	61.0 (4.3)	61.0 (4.1)	60.9 (4.2)	61.0 (4.2)	61.2 (4.2)
Total energy intake (kcal)	2115 (445)	2194 (485)	2202 (480)	2147 (505)	2172 (523)
Weight (kg)	67.9 (6.1)	74.2 (5.6)	77.4 (5.4)	81.4 (6.3)	88.9 (8.9)
Height (cm)	177.5 (6.7)	177.1 (6.6)	176.7 (6.0)	176.7 (6.5)	176.0 (6.5)
BMI ^c	22	23.7	24.9	26.1	28.4
BMI at age 20 (kg/m ²) ^c	18.9	20.5	21.7	22.9	24.7
Trouser size	50 (2)	50 (2)	52 (2)	52 (2)	54 (2)
Occupational physical activity (%)					
< 8 kJ/min	63	61	62	60	60
8-12 kJ/min	26	25	24	26	26
>12 kJ/min	11	13	14	14	14
Level of education (%)					
Primary school	18	19	23	27	29
Junior /senior high school	18	19	21	22	21
Higher vocational school	39	40	36	33	35
University	24	23	20	18	15
Smoking status (%)					
Never smoker	12	11	11	12	14
Ex-smoker	47	54	57	59	57
Current smoker	41	35	32	30	29
Alcohol intake (%)					
0 g/day	20	11	13	11	10
0.1-4 g/day	22	17	19	24	21
5-14 g/day	24	30	29	29	25
15-29 g/day	21	26	23	22	28
≥ 30 g/day	14	16	17	15	16
Family history of CRC (%)					
No	96	95	93	96	95
Yes	4	5	7	4	5

^a Sub-cohort members who were not missing data for the given exposure and any of the considered baseline characteristics.

^b Presented as mean (SD) or % where indicated

^c Median BMI

Table 2: Baseline (1986) characteristics of female sub-cohort members of the NLCS according to quintiles of BMI ^a

	Quintile of body mass index (BMI) (kg/m ²)				
	1 (lowest)	2	3	4	5 (highest)
Age at baseline ^b	61.2 (4.5)	61.0 (4.3)	61.1 (4.2)	61.6 (4.2)	61.4 (4.1)
Total energy intake (kcal)	1719 (412)	1710 (401)	1725 (420)	1655 (350)	1630 (372)
Weight (kg)	57.4 (5.6)	63.6 (4.6)	67.8 (4.9)	71.5 (5.6)	81.7 (8.5)
Height (cm)	166.2 (6.1)	165.6 (5.9)	165.9 (5.9)	164.4 (6.2)	163.9 (5.9)
Baseline BMI ^c	21.3	23.4	24.7	26.3	29.4
BMI at age 20 (kg/m ²)	18	20	21.3	22.6	24.6
Trouser size	40 (2)	42 (2)	43 (2)	44 (2)	46 (2)
Recreational physical activity %					
<30 mins/day	20	19	23	25	25
30-60 mins/day	30	32	32	33	33
60-90 mins/day	25	26	23	21	21
>90 mins/day	25	23	23	22	22
Level of education %					
Primary school	26	27	31	38	39
Junior /senior high school	17	24	26	25	28
Higher vocational school	43	38	36	29	29
University	13	12	7	8	4
Smoking status %					
Never smoker	49	58	54	65	63
Ex-smoker	24	18	25	21	19
Current smoker	27	24	21	14	18
Alcohol intake %					
0 g/day	28	31	32	33	37
0.1-4 g/day	33	35	37	39	38
5-14 g/day	22	20	18	18	16
15-29 g/day	12	11	9	8	8
≥ 30 g/day	6	3	4	3	3
Family history of CRC %					
No	94	94	93	95	94
Yes	6	6	7	5	6

^a Sub-cohort members who were not missing data for the given exposure and any of the considered baseline characteristics.

^b Presented as mean (SD) or % where indicated

^c Median value

In both men and women, a higher trouser/skirt size corresponded with increasing BMI, and total energy intake was highest in the third quintile of BMI. Furthermore, the greatest proportion of individuals with a University level education was observed in the lowest quintile of BMI, as was the greatest proportion of current smokers. In men, levels of occupational physical activity were relatively similar across quintiles of BMI, whereas in women, higher levels of recreational physical activity was reported by those in the lowest quintile of BMI

Age-adjusted results are not presented as these were comparable with multivariate results. Associations between body size and risk of CRC in men are presented in **Table 3**. BMI modelled per 5kg/m² increase was associated with total CRC (HR: 1.24, 95%CI: 1.05-1.46), and tumours of the distal colon (HR: 1.42, 95%CI: 1.13-1.79) and rectosigmoid (HR: 1.49, 95%CI: 1.05-2.11). When BMI was modelled in quintiles, associations did not reach statistical significance, however, there was a significant linear trend for tumours of the distal colon ($P=0.05$) and rectosigmoid ($P=0.05$). BMI at age 20 was associated only with tumours of the distal colon (per 5kg/m² increase HR: 1.09, 95%CI: 1.00-1.19; highest versus lowest quintile HR: 1.47, 95%CI: 1.03-2.08; P -trend=0.05). With respect to a positive BMI change, there was a suggested dose-response association for total CRC and all tumour sub-types, but this only reached statistical significance for tumours of the distal colon (P -trend=0.05). Trouser size, modelled per 2 unit size increase was associated with total CRC (HR: 1.12, 95%CI: 1.05-1.19) and tumours of the distal colon (HR: 1.16, 95%CI: 1.04-1.29). When trouser size was modelled in quintiles, there was a strong dose response association for total CRC (highest versus lowest quintile HR: 1.63, 95% CI: 1.17-2.29; P -trend=0.02) and tumours of the distal colon (highest versus lowest quintile HR: 2.56, 95%CI: 1.55-4.24; P -trend=<0.01). For other sub-types there was no association. With respect to height, an inverse association was observed for tumours of the rectosigmoid (per 5cm increase HR: 0.77, 95%CI: 0.65-0.93; highest versus lowest quintile HR: 0.38, 95%CI: 0.17-0.83; P -trend: 0.01). The strength of these associations did not differ when data was split according to long and short follow-up time. Tests for heterogeneity between tumour sub-types were not statistically significant, for all risk factors considered.

Associations between measures of anthropometry and risk of CRC in women are presented in **Table 4**. There were no statistically significant associations between BMI, BMI at age 20, BMI change and CRC risk, both overall, and for tumour sub-sites. With respect to trouser/skirt size, females in the highest quintile had a borderline statistically significant risk of proximal colon tumours, compared to those in the lowest quintile (HR: 1.46, 95%CI: 0.98-2.18), but there was no clear trend (P -trend=0.46). Height modelled per 5 cm increase was associated with a higher risk of total CRC (HR: 1.09, 95% CI: 1.01-1.17), and also when the highest quintile was compared to the lowest quintile (HR: 1.32, 95%CI: 1.03-1.71; P -trend=0.05). There was a borderline statistically significant association between height and distal colon tumours (HR: 1.11, 95%CI: 0.99-1.24) and rectal tumours (HR: 1.14, 95%CI: 0.98-1.32) when modelled per 5 cm increase. Females in the highest quintile also had a statistically significant risk of distal tumours when compared to those in the lowest quintile (HR: 1.53, 95%CI: 1.03-2.27; P -trend=0.05). This association was borderline statistically significant for rectal tumours (HR: 1.49, 95%CI: 0.98-2.27; P -trend=0.06). When data was split at 8 years of follow-up, it appeared that associations for all anthropometric variables attenuated over time (data not shown). Tests for heterogeneity between tumour sub-types were not statistically significant, for all risk factors considered.

Table 3: Multivariate adjusted hazard ratio (HR)^a and 95% confidence intervals (CI) for the association between anthropometric variables and colorectal endpoints in males of the NLCS after 16.3 years of follow-up

Exposure	PY at risk	Total Colorectal Cancer		Proximal Colon		Distal colon		Rectosigmoid		Rectum	
		N cases	HR (95% CI)	N cases	HR (95% CI)	N cases	HR (95%CI)	N cases	HR (95%CI)	N cases	HR (95%CI)
BMI (kg/m²)											
per 5 kg/m ²	18,595	1211	1.24 (1.05-1.46)	327	1.19 (0.92-1.54)	427	1.42 (1.13-1.79)	125	1.49 (1.05-2.11)	299	1.02 (0.79-1.32)
1	3773	232	1 (referent)	61	1 (referent)	78	1 (referent)	16	1 (referent)	60	1 (referent)
2	3931	238	0.95 (0.74-1.24)	75	1.15 (0.78-1.70)	78	0.92 (0.64-1.34)	24	1.34 (0.67-2.61)	59	0.92 (0.61-1.39)
3	3889	240	0.99 (0.77-1.28)	58	0.89 (0.59-1.33)	89	1.07 (0.75-1.53)	24	1.34 (0.69-2.60)	59	0.90 (0.60-1.35)
4	3688	247	1.05 (0.81-1.36)	59	0.93 (0.62-1.40)	89	1.10 (0.77-1.58)	34	2.04 (1.08-3.82)	60	0.93 (0.62-1.41)
5	3315	254	1.25 (0.96-1.62)	74	1.35 (0.90-1.98)	93	1.38 (0.95-1.98)	27	1.62 (0.85-3.10)	61	1.01 (0.67-1.51)
<i>P-trend</i>			0.08		0.43		0.05		0.05		0.96
BMI at age 20											
per 5 kg/m ²	18,595	1211	1.15 (0.97-1.37)	327	0.98 (0.75-1.27)	427	1.09 (1.00-1.19)	125	1.40 (0.98-2.01)	299	1.12 (0.86-1.46)
1	3771	238	1 (referent)	69	1 (referent)	84	1 (referent)	21	1 (referent)	29	1 (referent)
2	4090	240	0.92 (0.72-1.19)	69	0.98 (0.66-1.44)	82	0.96 (0.67-1.37)	19	0.85 (0.45-1.63)	163	1.18 (0.77-1.80)
3	3723	229	0.99 (0.77-1.28)	62	0.93 (0.63-1.38)	72	0.84 (0.58-1.22)	26	1.24 (0.68-2.26)	91	1.16 (0.75-1.79)
4	3622	254	1.09 (0.84-1.41)	60	0.86 (0.58-1.29)	89	1.03 (0.72-1.47)	31	1.43 (0.79-2.60)	16	1.34 (0.88-2.03)
5	3389	250	1.21 (0.93-1.56)	67	1.05 (0.72-1.56)	100	1.47 (1.03-2.08)	28	1.37 (0.75-2.51)	111	1.11 (0.73-1.71)
<i>P-trend</i>			0.07		0.97		0.05		0.1		0.46

Continues on next page

Continuing from previous page											
BMI change ^b											
<0	1646	103	0.87 (0.63-1.19)	27	0.91 (0.56-1.47)	36	0.87 (0.56-1.36)	7	0.41 (0.17-1.02)	29	0.99 (0.61-1.61)
0-3.9	10,298	657	1 (referent)	178	1 (referent)	224	1 (referent)	74	1 (referent)	163	1 (referent)
4.0-8.0	5701	375	1.07 (0.88-1.30)	96	0.97 (0.71-1.32)	140	1.23 (0.93-1.62)	37	1.01 (0.63-1.61)	91	1.02 (0.75-1.40)
≥8.0	950	76	1.38 (0.93-2.05)	26	1.59 (0.93-2.72)	27	1.56 (0.91-2.65)	7	1.29 (0.51-3.26)	16	1.16 (0.63-2.15)
<i>P-trend</i>			0.09		0.33		0.05		0.19		0.73
Height (cm) ^c											
per 5 cm	18,595	1211	0.96 (0.89-1.04)	327	1.01 (0.90-1.13)	427	0.99 (0.89-1.11)	125	0.77 (0.65-0.93)	299	0.95 (0.85-1.07)
1	5047	337	1 (referent)	94	1 (referent)	96	1 (referent)	32	1 (referent)	72	1 (referent)
2	2781	167	0.85 (0.65-1.12)	39	0.73 (0.48-1.12)	66	1.13 (0.77-1.65)	26	0.73 (0.41-1.29)	62	0.84 (0.56-1.26)
3	3819	262	0.94 (0.73-1.20)	68	0.91 (0.62-1.31)	91	1.03 (0.73-1.47)	34	0.90 (0.51-1.57)	63	0.91 (0.60-1.37)
4	3480	224	0.84 (0.64-1.09)	59	0.82 (0.55-1.22)	84	0.96 (0.65-1.41)	19	0.54 (0.27-1.06)	58	0.92 (0.60-1.42)
5	3469	221	0.80 (0.60-1.08)	67	0.90 (0.59-1.37)	90	0.97 (0.63-1.48)	14	0.38 (0.17-0.83)	44	0.71 (0.43-1.17)
<i>P-trend</i>			0.16		0.66		0.70		0.01		0.35
Trouser size											
per 2 sizes	18,595	1211	1.12 (1.05-1.19)	327	1.09 (0.99-1.20)	427	1.16 (1.04-1.29)	125	1.07 (0.93-1.22)	299	1.06 (0.96-1.17)
1	2760	128	1 (referent)	43	1 (referent)	32	1 (referent)	15	1 (referent)	35	1 (referent)
2	4563	280	1.31 (0.98-1.74)	70	0.96 (0.62-1.48)	95	1.78 (1.14-2.78)	25	1.04 (0.53-2.04)	83	1.42 (0.90-2.23)
3	6185	421	1.46 (1.11-1.91)	116	1.21 (0.81-1.80)	156	2.14 (1.40-3.27)	48	1.51 (0.82-2.79)	87	1.09 (0.70-1.63)
4	3269	239	1.56 (1.15-2.10)	59	1.16 (0.74-1.82)	87	2.23 (1.40-3.54)	26	1.51 (0.76-3.01)	63	1.50 (0.94-2.39)
5	1819	143	1.63 (1.17-2.29)	39	1.32 (0.81-2.15)	57	2.56 (1.55-4.24)	11	1.10 (0.49-2.49)	31	1.33 (0.77-1.29)
<i>P-trend</i>			0.02		0.14		<0.01		0.28		0.34

a Models adjusted for: age, total energy intake, recreational physical activity, level of education, family history of CRC, alcohol, smoking

b Model additionally adjusted for BMI at age 20

c Model additionally adjusted for body weight (kg)



Table 4: Multivariate adjusted hazard ratio (HR)a and 95% confidence intervals (CI) for the association between anthropometric variables and colorectal endpoints in females of the NLCS after 16.3 years of follow-up

Exposure	Total Colorectal Cancer			Proximal Colon			Distal colon			Rectosigmoid			Rectum			
	PY at risk	N cases	HR (95% CI)	N cases	HR (95% CI)	CI	N cases	HR (95% CI)	CI	N cases	HR (95% CI)	CI	N cases	HR (95% CI)	CI	
BMI (kg/m²)																
per 5 kg/m ²	27,19	1106	0.98 (0.88-1.10)	459	1.02 (0.87-1.18)		327	0.95 (0.79-1.14)		87	1.01 (0.74-1.38)		205	1.05 (0.85-1.31)		
	5															
1	5239	228	1 (referent)	94	1 (referent)		70	1 (referent)		16	1 (referent)		42	1 (referent)		
2	5623	211	0.88 (0.69-1.13)	87	0.88 (0.63-1.23)		62	0.85 (0.58-1.24)		17	1.08 (0.53-2.21)		41	1.00 (0.63-1.58)		
3	5499	223	0.94 (0.73-1.20)	94	0.99 (0.71-1.40)		58	0.81 (0.55-1.19)		26	1.86 (0.94-3.69)		38	0.87 (0.54-1.40)		
4	5606	222	0.91 (0.71-1.16)	95	0.98 (0.70-1.37)		64	0.85 (0.58-1.23)		13	0.85 (0.40-1.84)		40	0.93 (0.58-1.49)		
5	5229	222	0.97 (0.76-1.24)	89	0.91 (0.65-1.28)		73	1.04 (0.72-1.50)		15	1.06 (0.50-2.26)		44	1.07 (0.67-1.69)		
<i>P-trend</i>			0.90		0.84			0.84			0.93			0.90		
BMI at age 20																
per 5 kg/m ²	27,19	1106	1.04 (0.90-1.19)	459	1.03 (0.84-1.25)		327	1.05 (0.85-1.31)		87	0.91 (0.61-1.35)		205	1.06 (0.82-1.36)		
	5															
1	5628	213	1 (referent)	90	1 (referent)		60	1 (referent)		19	1 (referent)		38	1 (referent)		
2	5247	227	1.21 (0.95-1.56)	100	1.28 (0.90-1.76)		67	1.22 (0.83-1.79)		18	1.09 (0.56-1.13)		43	1.26 (0.78-2.01)		
3	5212	236	1.25 (0.98-1.61)	96	1.15 (0.82-1.60)		74	1.36 (0.95-1.99)		18	0.97 (0.50-1.89)		46	1.34 (0.84-2.13)		
4	5576	212	1.08 (0.84-1.38)	87	0.98 (0.70-1.38)		65	1.21 (0.82-1.79)		15	0.85 (0.42-1.72)		35	0.98 (0.60-1.61)		
5	5532	218	1.12 (0.87-1.43)	86	1.14 (0.81-1.62)		61	1.09 (0.74-1.62)		17	1.07 (0.54-2.12)		43	1.29 (0.80-2.07)		
<i>P-trend</i>			0.68		0.97			0.68			0.89			0.62		

Continues on next page

Continuing from previous page											
BMI change^b											
<0	3268	134	0.94 (0.73-1.22)	51	0.86 (0.60-1.23)	44	1.08 (0.72-1.60)	12	1.26 (0.63-2.50)	19	0.67 (0.39-1.14)
0-3.9	12,177	507	1 (referent)	212	1 (referent)	147	1 (referent)	37	1 (referent)	97	1 (referent)
4.0-8.0	8787	343	0.89 (0.74-1.06)	144	0.86 (1.67-1.10)	98	0.89 (0.67-1.20)	26	0.97 (0.56-1.68)	69	0.96 (0.68-1.35)
≥8.0	2963	122	0.95 (0.73-1.24)	52	0.98 (0.68-1.40)	38	1.00 (0.67-1.50)	12	1.40 (0.69-2.82)	20	0.83 (0.49-1.39)
<i>P-trend</i>			0.55		0.86		0.58		0.83		0.78
Height (cm)^c											
per 5 cm	27,195	1106	1.09 (1.01-1.17)	459	1.04 (0.95-1.15)	327	1.11 (0.99-1.24)	87	1.02 (0.82-1.25)	205	1.14 (0.98-1.32)
1	6353	250	1 (referent)	104	1 (referent)	68	1 (referent)	24	1 (referent)	49	1 (referent)
2	4932	186	0.96 (0.75-1.23)	89	1.11 (0.80-1.54)	62	1.17 (0.80-1.72)	10	0.53 (0.25-1.12)	20	0.52 (0.30-0.89)
3	5599	209	0.97 (0.77-1.24)	103	1.15 (0.83-1.58)	49	0.84 (0.56-1.25)	13	0.61 (0.30-1.24)	40	0.93 (0.60-1.45)
4	5708	231	1.03 (0.81-1.31)	78	0.83 (0.59-1.18)	76	1.28 (0.88-1.86)	23	1.02 (0.55-1.86)	45	1.04 (0.68-1.59)
5	4603	230	1.32 (1.03-1.71)	85	1.19 (0.84-1.70)	72	1.53 (1.03-2.27)	17	0.90 (0.44-1.84)	51	1.49 (0.98-2.27)
<i>P-trend</i>			0.05		0.67		0.05		0.78		0.06
Trouser/skirt size											
per 2 sizes	27,195	1106	1.03 (0.98-1.09)	459	1.06 (0.98-1.14)	327	1.02 (0.94-1.11)	87	1.03 (0.90-1.18)	205	1.04 (0.94-1.16)
1	5317	210	1 (referent)	80	1 (referent)	66	1 (referent)	15	1 (referent)	36	1 (referent)
2	6861	284	1.02 (0.80-1.29)	126	1.20 (0.86-1.67)	80	0.89 (0.63-1.28)	20	1.07 (0.54-2.13)	52	1.08 (0.68-1.71)
3	7586	283	0.90 (0.71-1.14)	114	0.95 (0.68-1.33)	83	0.83 (0.58-1.19)	29	1.38 (0.73-2.60)	52	0.96 (0.61-1.53)
4	4761	197	0.98 (0.76-1.27)	76	0.97 (0.67-1.39)	57	0.92 (0.62-1.36)	19	1.39 (0.68-2.87)	44	1.28 (0.79-2.08)
5	2670	132	1.17 (0.87-1.58)	63	1.46 (0.98-2.18)	41	1.15 (0.74-1.80)	4	0.58 (0.19-1.78)	21	1.07 (0.59-1.93)
<i>P-trend</i>			0.62		0.46		0.72		0.90		0.56

^a Models adjusted for: age, total energy intake, recreational physical activity, level of education, family history of CRC, alcohol, smoking

^b additionally adjusted for BMI at age 20

^c additionally adjusted for body weight (kg)

Table 5 shows HRs for the interaction between trouser size and occupational physical activity in men. Compared to the reference category of men with a small trouser size and a high level of physical activity, there were no statistically significant associations for total CRC, but men with a small trouser size and a low level of physical activity were at increased risk of distal colon tumours (HR: 1.63, 95%CI: 1.03-2.56). A borderline statistically significant association was also observed for men with a large trouser/skirt size and all levels of physical activity with respect to tumours of the distal colon.

Table 6 shows CRC risk according to the interaction between trouser/skirt size and baseline recreational physical activity in women. It appears that physical activity is important for modulating CRC risk in women. Compared to the reference category of women with a small skirt/trouser size and a high level of physical activity, a number of statistically significant associations were observed. However, the greatest risk for each sub-type considered appeared to be for those with the highest trouser/skirt size and the lowest level of physical activity: total CRC HR: 1.83, 95%CI: 1.28-2.63; proximal colon HR: 1.70, 95%CI: 1.08-2.67; distal colon HR: 1.95, 95%CI: 1.21-3.17; rectum HR: 2.56, 95%CI: 1.36-4.79. The interaction between trouser/skirt size and physical activity was only statistically significant for the proximal colon ($P=0.05$).

Table 5: Multivariate Adjusted Hazard ratios (HR)^a and 95% Confidence Intervals (CI) for the Interaction Between Occupational Physical Activity and Trouser Size, and Risk of Colorectal Cancer in Men of the NLCS

Trouser size	Physical Activity					
	Low		Intermediate		High	
	N/PY at risk ^b	HR (95%CI)	N/PY at risk ^b	HR (95%CI)	N/PY at risk ^b	HR (95%CI)
Total CRC						
≤52	632/ 10,139	1.20 (0.90-1.60)	249/ 4392	1.03 (0.76- 1.39)	123/ 2293	1 (referent)
>52	263/ 3821	1.15 (0.83- 1.61)	123/ 1490	1.38 (0.95- 2.02)	75/ 940	1.34 (0.87-2.06)
Proximal Colon						
≤52	169/ 10,139	1.03 (0.67- 1.58)	83/ 4392	1.10 (0.70- 1.72)	37/ 2293	1 (referent)
>52	81/ 3821	1.14 (0.70- 1.84)	29/ 1490	1.07 (0.61- 1.89)	15/ 940	0.84 (0.42-1.67)
Distal Colon						
≤52	233/ 10,139	1.63 (1.03-2.56)	66/ 4392	1.10 (0.70- 1.72)	30/ 2293	1 (referent)
>52	103/ 3821	1.54 (0.95- 2.54)	41/ 1490	1.72 (0.98-3.03)	26/ 940	1.76 (0.94-3.30)
Rectum						
≤52	140/ 10,139	0.94 (0.61-1.45)	68/ 4392	0.92 (0.59-1.45)	40/ 2293	1 (referent)
>52	51/ 3821	0.89 (0.52- 1.53)	37/ 1490	1.51 (0.86- 2.66)	22/ 940	1.41 (0.75-2.67)

^a Adjusted for trouser size, physical activity, total energy intake, family history of CRC, socioeconomic status, alcohol consumption, and smoking status

^b Number of cases/person years at risk

Table 6: Multivariate Adjusted Hazard Ratios (HR)^a and 95% Confidence Intervals (CI) for the Interaction Between Non-Occupational Physical Activity and Trouser/skirt size, and risk of Colorectal Cancer in Women of the NLCS

Trouser size	Physical Activity					
	Low		Intermediate		High	
	N/PY at risk ^b	HR (95%CI)	N/PY at risk ^b	HR (95%CI)	N/PY at risk ^b	HR (95%CI)
Total CRC						
≤44	159/ 4471	1.45 (1.08-1.94)	357/ 12,026	1.25 (0.97- 1.60)	122/ 5054	1 (referent)
>44	97/ 2283	1.83 (1.28-2.63)	116/ 4447	1.14 (0.82- 1.59)	51/ 1617	1.44 (0.95- 2.18)
Proximal Colon[*]						
≤44	89/ 4471	1.60 (1.10-2.31)	203/ 12,026	1.40 (1.02- 1.93)	63/ 5054	1 (referent)
>44	52/ 2283	1.70 (1.08- 2.67)	66/ 4447	1.16 (0.76- 1.77)	33/ 1617	1.69 (1.03- 2.79)
Distal Colon						
≤44	63/ 4471	1.28 (0.86- 1.91)	141/ 12,026	1.08 (0.77- 1.52)	54/ 5054	1 (referent)
>44	42/ 2283	1.95 (1.21- 3.17)	50/ 4447	1.18 (0.74- 1.84)	17/ 1617	1.12 (0.61- 2.03)
Rectum						
≤44	38/ 4471	1.67 (0.98-2.85)	86/ 12,026	1.43 (0.89- 2.29)	25/ 5054	1 (referent)
>44	27/ 2283	2.56 (1.36- 4.79)	35/ 4447	1.70 (0.93- 3.10)	12/ 1617	1.66 (0.77- 3.54)

^{*} Statistically significant interaction ($P < 0.05$)

^a Adjusted for trouser/skirt size, physical activity, total energy intake, family history of CRC, socioeconomic status, alcohol consumption, and smoking status

^b Number of cases/person years at risk

DISCUSSION

The results of the present study are in agreement with the WCRF consensus that body fat, especially central body fat, is associated with increased risk of CRC in males.¹ Furthermore, when we investigated associations according to anatomic sub-site of the colon and rectum, this risk appeared especially strong for tumours of the distal colon. Although we observed only a weak association between waist measurement and CRC risk in women, we have shown that physical activity may influence this association. Height was associated with an increased risk of incident CRC in women, especially tumours of the distal colon and rectum, whereas we observed no association between height and CRC in men.

Strengths of this study include the prospective cohort design, and a large population in combination with sufficient follow-up time to allow precise analysis of CRC risk at different sub-sites. The NLCS has almost complete ascertainment of CRC cases, and follow-up of the sub-cohort is almost 100%. Although measures of body size in this study were obtained by self-report, there are many examples in the literature showing that this method is a reasonably valid and reliable tool for assessing body weight and height in cohort studies.⁴⁰⁻⁴³

Several case-control studies have considered the association between BMI and site specific tumours of the colon in men, and generally report a stronger risk for tumours of the distal colon than proximal colon.¹⁷⁻²¹ This is supported by results of a recent prospective cohort study by Laake *et al.*²² In our study we considered several indicators of body fat in addition to BMI, including BMI at age 20, BMI change, and trouser size as a proxy for waist circumference, and consistently observed an elevated risk of distal colon tumours compared to other sub-sites. Although we did not observe statistically significant heterogeneity between proximal and distal colon tumours as was observed by Laake *et al.*,²² the bootstrapping method we used is quite conservative. In the present study we also considered associations for the rectum and rectosigmoid. The weaker associations that we observed with respect to associations between body fat and the rectum align with previous research,¹ but we also observed a positive association between BMI and tumours of the rectosigmoid. To our knowledge, no other prospective study has considered the association between anthropometric measures and the rectosigmoid. This association is plausible, as the rectosigmoid and distal

colon are anatomically linked. Our findings should be replicated in other prospective studies, but they support the hypothesis that aetiology of tumours differs between anatomical sub-sites of the colorectal tract in men.

We did not observe clear associations between BMI, BMI at age 20, or BMI change and CRC risk in the women of our population, but our results suggest that women with high skirt or trouser size are at increased risk of a tumour in the proximal colon, although this association did not reach statistical significance. In contrast, positive associations between BMI and distal colon tumours have been reported in prospective cohorts of Swedish²³ and Norwegian women.²² In an American cohort, both BMI and waist circumference were associated with proximal and distal tumours.²⁴ The proportion of obese women in the NLCS is small, and perhaps this prevented us from detecting associations of similar strength. It has also been suggested that the impact of a given risk factor along the large bowel may differ according to the prevalence of other environmental factors, and thus, according to sex and country.⁴⁴

Alternatively, it is plausible that a metabolic profile reflecting a combination of risk factors has a greater influence on cancer risk than do the individual effects of body size and obesity.^{45, 46} Perhaps this is especially true for women; it has been observed that breast cancer risk may vary according to various combinations of the components of energy balance.⁴⁶ Energy imbalance results when there is a change in body energy stores due to an excess or deficit of energy intake in comparison to energy expenditure.⁴⁷ While increasing body fat is the outcome of such imbalance, energy intake and physical activity are also important aspects to this equation.^{46, 47} Each of these components has been investigated separately with respect to CRC risk,¹ however, to our knowledge, only one case-control study has considered risk of CRC according to a model of energy balance.²⁰

Slattery *et al.*²⁰ observed that when physical activity was high, individuals with a high energy intake and a large BMI had a non-significant increased risk of colon cancer, and this was slightly more pronounced for men than women. We were unable to consider a model that included energy intake, but we did investigate the interaction between body fat and physical activity. Because evidence suggests that fat distribution is more important than body weight or BMI for CRC risk, especially in women,³⁶ we investigated this interaction using trouser/skirt size instead of BMI. Our

observations for women were intriguing. As previously noted, we did not observe statistically significant associations with respect to trouser/skirt size and CRC risk in our general analysis. However, when we considered the interaction between trouser/skirt size and physical activity, it appears that women with a large skirt/trouser size and the lowest level of physical activity are at the greatest risk of CRC compared to women with a small skirt/trouser size and a high level of physical activity. This risk was highest for rectal tumours; however, the interaction was only statistically significant for tumours of the proximal colon. These findings require confirmation in other prospective studies, but support case-control data.^{20, 27} They also highlight the importance of considering energy balance as a risk factor for CRC, and may help explain previous contradictory findings with respect to the association between body fat and CRC risk in women.

Height is a reflection of earlier life exposures, such as childhood energy intake.⁶ In our study, height was associated with increased risk of CRC in women, which is in accordance with previous research.^{1, 24, 36, 48, 49} A recent report from the Iowa Women's Health Study suggested that risk is greater for tumours of the proximal colon,²⁴ whereas our data suggests that risk is greater for tumours of the distal colon. However, the former study did not report tests for heterogeneity, and in our study, such tests did not reach statistical significance. We did not observe any association with respect to height and CRC risk in men, and actually observed an inverse association for tumours of the rectosigmoid. We can not explain this observation, and suggest that it may be a chance finding. We have shown previously in the NLCS population that energy restriction during childhood is inversely related to later CRC risk,⁵⁰ and therefore, while we have not observed strong associations between height and CRC in the present study, it is plausible that early life nutritional factors may influence different carcinogenic pathways than later life exposure to increased body fat.²⁴

We can only speculate as to why body size and body fat in particular might differentially influence aetiology of colorectal tumour sub-sites. Obesity is associated with a chronic state of low-grade inflammation and thus increased circulatory levels of inflammatory markers like C - reactive protein, interleukin-6 and tumour necrosis factor- α .⁵¹ These in turn may induce insulin resistance and hyperinsulinemia. It is hypothesized that such conditions influence CRC risk,^{10, 52, 53} however, it remains unclear how they

directly or indirectly influence specific tumour sub-sites. There are some distinct differences between the proximal and distal colon, both anatomically and genetically.¹⁴⁻¹⁶ The distal colon is associated more with physical/chemical stimuli and associated less with water absorption and electrolyte transport than the proximal colon.¹⁴ Genetically speaking, CRC is one of the best described cancers in terms of pathways and events involved. 85% of CRCs are thought to arise via the chromosomal instability (CIN) pathway, which is characterized by mutations in the *KRAS* oncogene and the *APC* tumour suppressor gene.⁵⁴ The remaining 15% are thought to arise via promoter methylation and/or microsatellite instability (MSI).⁵⁴ It has been shown that CIN is more associated with tumours of the distal colon, whereas promoter methylation and MSI is more associated with tumours of the proximal colon.⁵⁵ Interestingly, recent studies suggest that overweight and obesity do not appear to differentially exert risk on CRC via epigenetic mechanisms and MSI.⁵⁶⁻⁵⁸ Based on the observation that body fat appears to be associated with a higher risk of distal colon tumours than proximal colon tumours, a plausible hypothesis is that obesity and its associated process may influence risk via the CIN pathway. However, to our knowledge, this has not been investigated in any population based studies.

5

To conclude, in this Dutch population we observed that overweight is a strong risk factor for CRC in men, especially for tumours of the distal colon, and trouser size predicts this risk independently of BMI. In women, the risk of body fat on CRC risk may differ according to level of physical activity. Together, these findings reiterate the importance of maintaining an ideal energy balance in CRC prevention.

REFERENCES

1. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AIRC, 2007.
2. Dagenais GR, Yi Q, Mann JF, *et al.* Prognostic impact of body weight and abdominal obesity in women and men with cardiovascular disease. *Am Heart J.* 2005; 149: 54-60.
3. Bigaard J, Tjonneland A, Thomsen BL, *et al.* Waist circumference, BMI, smoking, and mortality in middle-aged men and women. *Obes Res.* 2003; 11: 895-903.
4. Soranzo N, Rivadeneira F, Chinappan-Horsley U, *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet.* 2009; 5: e1000445.
5. Hirschhorn JN, Lettre G. Progress in genome-wide association studies of human height. *Horm Res.* 2009; 71 Suppl 2: 5-13.
6. Okasha M, Gunnell D, Holly J, *et al.* Childhood growth and adult cancer. *Best Pract Res Clin Endocrinol Metab.* 2002; 16: 225-41.
7. Dai Z, Xu YC, Niu L. Obesity and colorectal cancer risk: a meta-analysis of cohort studies. *World J Gastroenterol.* 2007; 13: 4199-206.
8. Ning Y, Wang L, Giovannucci EL. A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies. *Obes Rev.* 2010; 11: 19-30.
9. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr.* 2007; 86: 556-65.
10. Giovannucci E, Michaud D. The role of obesity and related metabolic disturbances in cancers of the colon, prostate, and pancreas. *Gastroenterology.* 2007; 132: 2208-25.
11. Jacobs ET, Thompson PA, Martinez ME. Diet, gender, and colorectal neoplasia. *J Clin Gastroenterol.* 2007; 41: 731-46.
12. Pischon T, Notlings U, Boeing H. Obesity and cancer. *Proc Nutr Soc.* 2008; 67: 128-45.

13. Morois S, Mesrine S, Josset M, *et al.* Anthropometric Factors in Adulthood and Risk of Colorectal Adenomas: The French E3N-EPIC Prospective Cohort. *Am J Epidemiol.* 2010.
14. Li FY, Lai MD. Colorectal cancer, one entity or three. *J Zhejiang Univ Sci B.* 2009; 10: 219-29.
15. Breivik J, Lothe RA, Meling GI, *et al.* Different genetic pathways to proximal and distal colorectal cancer influenced by sex-related factors. *Int J Cancer.* 1997; 74: 664-9.
16. Lindblom A. Different mechanisms in the tumourigenesis of proximal and distal colon cancers. *Curr Opin Oncol.* 2001; 13: 63-9.
17. Caan BJ, Coates AO, Slattery ML, *et al.* Body size and the risk of colon cancer in a large case-control study. *Int J Obes Relat Metab Disord.* 1998; 22: 178-84.
18. Gerhardsson de Verdier M, Hagman U, Steineck G, *et al.* Diet, body mass and colorectal cancer: a case-referent study in Stockholm. *Int J Cancer.* 1990; 46: 832-8.
19. Russo A, Franceschi S, La Vecchia C, *et al.* Body size and colorectal-cancer risk. *Int J Cancer.* 1998; 78: 161-5.
20. Slattery ML, Potter J, Caan B, *et al.* Energy balance and colon cancer--beyond physical activity. *Cancer Res.* 1997; 57: 75-80.
21. West DW, Slattery ML, Robison LM, *et al.* Dietary intake and colon cancer: sex- and anatomic site-specific associations. *Am J Epidemiol.* 1989; 130: 883-94.
22. Laake I, Thune I, Selmer R, *et al.* A prospective study of body mass index, weight change, and risk of cancer in the proximal and distal colon. *Cancer Epidemiol Biomarkers Prev.* 2010; 19: 1511-22.
23. Terry P, Giovannucci E, Bergkvist L, *et al.* Body weight and colorectal cancer risk in a cohort of Swedish women: relation varies by age and cancer site. *Br J Cancer.* 2001; 85: 346-9.
24. Oxentenko A, Bardia A, Vierkant RA, *et al.* Body Size and Incident Colorectal Cancer: A Prospective Study of Older Women. *Cancer Prev Res.* 2010; doi: 10.1158/1940-6207.CAPR-10-0116.
25. International Agency for Research on Cancer (IARC). 2008. GLOBOCAN 2008: Colorectal Cancer Incidence and Mortality Worldwide in 2008.
<http://globocan.iarc.fr/factsheets/cancers/colorectal.asp>

26. World Health Organization. 2010. Information sheet on obesity and overweight factsheet.
http://www.who.int/dietphysicalactivity/media/en/gsf_s_obesity.pdf
27. Slattery ML, Caan BJ, Benson J, *et al.* Energy balance and rectal cancer: an evaluation of energy intake, energy expenditure, and body mass index. *Nutr Cancer*. 2003; 46: 166-71.
28. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol*. 1990; 43: 285-95.
29. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr*. 1994; 48: 253-65.
30. Goldbohm RA, van 't Veer P, van den Brandt PA, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr*. 1995; 49: 420-9.
31. Van den Brandt PA, Schouten LJ, Goldbohm RA, *et al.* Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol*. 1990; 19: 553-8.
32. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of municipalities by cancer registries and PALGA using hospital discharge data. *Tijdschr Soc Gezondheidsz* 1994; 72: 80-4.
33. Hughes LA, Schouten LJ, Goldbohm RA, *et al.* Self-reported clothing size as a proxy measure for body size. *Epidemiology*. 2009; 20: 673-6.
34. Hettinger TH, Mueller BH, Gebhard H. *Ermittlung des Arbeitsenergieumsatzes bei Dynamisch Muskulaerer Arbeit*. Dortmund 1989.
35. Pylypchuk RD, Schouten LJ, Goldbohm RA, *et al.* Body mass index, height, and risk of lymphatic malignancies: a prospective cohort study. *Am J Epidemiol*. 2009; 170: 297-307.
36. Pischon T, Lahmann PH, Boeing H, *et al.* Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst*. 2006; 98: 920-31.
37. Johnson RK. Dietary intake--how do we measure what people are really eating? *Obes Res*. 2002; 10 Suppl 1: 63S-8S.

38. Briefel RR, Sempos CT, McDowell MA, *et al.* Dietary methods research in the third National Health and Nutrition Examination Survey: underreporting of energy intake. *Am J Clin Nutr.* 1997; 65: 1203S-9S.
39. de Vogel S, Bongaerts BW, Wouters KA, *et al.* Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. *Carcinogenesis.* 2008; 29: 1765-73.
40. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, *et al.* Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health.* 2000; 54: 143-8.
41. Nyholm M, Gullberg B, Merlo J, *et al.* The validity of obesity based on self-reported weight and height: Implications for population studies. *Obesity (Silver Spring).* 2007; 15: 197-208.
42. Spencer EA, Appleby PN, Davey GK, *et al.* Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr.* 2002; 5: 561-5.
43. Wada K, Tamakoshi K, Tsunekawa T, *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Int J Obes.* 2005; 29: 1093-9.
44. Morois S, Mesrine S, Josset M, *et al.* Anthropometric Factors in Adulthood and Risk of Colorectal Adenomas: The French E3N-EPIC Prospective Cohort. *Am J Epidemiol.*
45. McKeown-Eyssen G. Epidemiology of colorectal cancer revisited: are serum triglycerides and/or plasma glucose associated with risk? *Cancer Epidemiol Biomarkers Prev.* 1994; 3: 687-95.
46. Silvera SA, Jain M, Howe GR, *et al.* Energy balance and breast cancer risk: a prospective cohort study. *Breast Cancer Res Treat.* 2006; 97: 97-106.
47. Willett WC, Stampfer MJ. Implications of total energy intake and epidemiological analyses. In: Willett WC (ed.) *Nutritional Epidemiology.* Oxford University Press, New York (NY) 1998, pp 273-301.

48. Engeland A, Tretli S, Austad G, *et al.* Height and body mass index in relation to colorectal and gallbladder cancer in two million Norwegian men and women. *Cancer Causes Control.* 2005; 16: 987-96.
49. Wei EK, Giovannucci E, Wu K, *et al.* Comparison of risk factors for colon and rectal cancer. *Int J Cancer.* 2004; 108: 433-42.
50. Hughes LA, van den Brandt PA, Goldbohm RA, *et al.* Childhood and adolescent energy restriction and subsequent colorectal cancer risk: results from the Netherlands Cohort Study. *Int J Epidemiol.* 2010; 39: 1333-44.
51. Monteiro R, Azevedo I. Chronic inflammation in obesity and the metabolic syndrome. *Mediators Inflamm.* 2010; doi: 10.1155/2010/289645
52. Elwing JE, Gao F, Davidson NO, *et al.* Type 2 diabetes mellitus: the impact on colorectal adenoma risk in women. *Am J Gastroenterol.* 2006; 101: 1866-71.
53. Hu FB, Manson JE, Liu S, *et al.* Prospective study of adult onset diabetes mellitus (type 2) and risk of colorectal cancer in women. *J Natl Cancer Inst.* 1999; 91: 542-7.
54. Derks S, Postma C, Carvalho B, *et al.* Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis.* 2008; 29: 434-9.
55. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology.* 2007; 50: 113-30.
56. Campbell PT, Jacobs ET, Ulrich CM, *et al.* Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumour microsatellite instability status. *J Natl Cancer Inst.* 2010; 102: 391-400.
57. Slattery ML, Curtin K, Sweeney C, *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer.* 2007; 120: 656-63.

58. Slattery ML, Wolff RK, Curtin K, *et al.* Colon tumour mutations and epigenetic changes associated with genetic polymorphism: insight into disease pathways. *Mutat Res.* 2009; 660: 12-21.

6

Physical activity, sitting time and colorectal cancer risk in the Netherlands Cohort Study:

Regular long-term physical activity may matter

Colinda CJM Simons
Laura AE Hughes
Manon van Engeland
R Alexandra Goldbohm
Piet A van den Brandt
Matty P Weijenberg

Submitted for publication

ABSTRACT

Background: Physical activity has been associated with a lower colon cancer risk, but not so much rectal cancer risk. We analyzed cumulative occupational energy expenditure and sitting hours (in men only), baseline non-occupational physical activity and former sports participation in relation to colorectal cancer endpoints.

Methods: The Netherlands Cohort Study (NLCS) includes 120,852 participants aged 55-69 years in 1986 when completing a self-administered questionnaire. After 16.3 years, 1,819 male and 1,366 female colorectal cancer cases were available for case-cohort analyses.

Results: Higher levels of cumulative occupational energy expenditure were not significantly associated with decreased hazard ratios for distal colon cancer (T3 versus T1: HR: 0.79, 95% CI: 0.60-1.01, P -trend=0.06), and significantly associated with increased hazard ratios for rectal cancer (T3 versus T1: HR: 1.73, 95% CI: 1.30-2.29, P -trend<0.001). Less cumulative occupational sitting hours were significantly associated with decreased hazard ratios for distal colon cancer (T3 versus T1: HR: 0.66, 95% CI: 0.51-0.85, P -trend=0.001). The median duration of work history in male sub-cohort members was 38 years. Baseline non-occupational physical activity was not consistently associated with colorectal cancer endpoints in men. In women, hazard ratios were significantly decreased for >90 minutes of daily baseline non-occupational physical activity versus \leq 30 minutes for distal colon cancer (HR: 0.69, 95% CI: 0.50-0.96, P -trend=0.06) and rectal cancer (HR: 0.59, 95% CI: 0.39-0.90, P -trend=0.02).

Conclusions: Former sports participation might have a small added protective effect in active participants. In conclusion, our data suggest a potential protective effect of regular long-term physical activity, which was especially evident for distal colon cancer.

INTRODUCTION

Physical activity has been associated with a decreased colon cancer risk in numerous studies, with indications of linear dose-response patterns. The evidence for rectal cancer risk is less clear.¹⁻¹⁵ Health recommendations include moderate physical activity for at least 30 minutes every day and limiting sedentary habits.⁵

Several biological mechanisms may underlie the inverse association between physical activity and colorectal cancer risk. Physical activity may improve immune function, reduce free radicals, increase prostaglandins and increase gut motility. Physical activity may also improve insulin sensitivity, reduce the level of insulin and insulin-like growth factor 1, and increase the level of insulin-like growth factor binding protein 3. Although physical activity may protect against colorectal cancer through protecting against weight gain, overweight and obesity, studies have suggested an independent effect of physical activity.^{2, 4, 15, 16}

There is no clear evidence indicating that some types of physical activity are better predictors of a decreased colorectal cancer risk than others. In terms of translating study findings to health recommendations, it is important to investigate different types (*e.g.*, occupational, recreational) and aspects of physical activity (*e.g.*, duration, intensity and frequency) as well as sedentary behavior. It is also interesting whether physical activity during specific age periods may have a particular protective effect as colorectal cancer development takes several decades to complete, and exposures early in life could influence future cancer risk.¹⁷ An effect of physical activity through insulin-related growth factors, for example, may especially occur during puberty when most bodily growth occurs. Intriguing in this respect is the finding that increased height has been associated with an increased colorectal cancer risk independent of weight.⁵

To help clarify the unresolved issues, we analyzed cumulative occupational energy expenditure, cumulative occupational sitting hours, baseline non-occupational physical activity and former sports participation in relation to colorectal cancer risk within the large Netherlands Cohort Study (NCLS). The NCLS made possible an analysis of colorectal cancer risk overall, at sub- sites, and for men and women separately.

METHODS

Study population and design

The NLCS includes 120,852 men and women aged 55-69 years at baseline in 1986. Using a case-cohort approach, cases are enumerated for the entire cohort and a randomly selected sub-cohort ($n = 5000$) is followed up for vital status to calculate the person-time at risk. Incident cancer cases are ascertained through linkage with the population-based cancer registry and the national pathology database (PALGA).¹⁸⁻¹⁹ The estimated completeness of cancer follow-up is $>96\%$.²⁰ Follow-up of the sub-cohort is $\sim 100\%$. More details of the NLCS can be found elsewhere.²¹

After 16.3 years, 1,819 male and 1,366 female colorectal cancer cases and 2,169 male and 2,247 female sub-cohort members were available for analysis. These numbers excluded participants with a history of cancer other than non-melanoma skin cancer and/or incomplete or inconsistent questionnaires. Among the 1,819 male colorectal cancer cases, there were 511 male proximal colon cancer cases (International Classification of Diseases for Oncology, first edition codes 153.0, 153.1, 153.4, 153.5 and 153.6), 608 male distal colon cancer cases (codes 153.2, 153.3 and 153.7) and 464 male rectal cancer cases (codes 154.1). Among the 1,366 female colorectal cancer cases, these numbers were 560, 417 and 252, respectively. The total number of male and female colon cancer cases was 1,165 and 1,009, respectively, and includes colon cancer cases that could not be classified as either proximal or distal colon cancer cases (codes 153.8 and 153.9). The remaining cases were rectosigmoid cancer cases (code 154.0) and were only included when analyzing the risk of overall colorectal cancer.

The NLCS has been approved by the institutional review boards of the TNO Nutrition and Food Research Institute (Zeist) and Maastricht University (Maastricht).

Questionnaire

All participants completed a mailed self-administered questionnaire on diet and other risk factors for cancer at baseline. The questionnaire included a semi-quantitative food-frequency questionnaire (FFQ) assessing regular consumption of 150 food items in the preceding year. Nutrient intakes were calculated using the Dutch food composition table 1986-1987.²² The

FFQ was found a valid instrument when compared with a 9-day diet record²³ and was shown to be reproducible over a period of at least 5 years.²⁴

Occupational physical activity

Occupational energy expenditure and sitting time were derived from participants' reported occupational history using a rating system developed by Hettinger et al.²⁵ A maximum of five jobs could be mentioned.

Occupational energy expenditure distinguishes between work of low activity (*e.g.*, desk jobs, driving), work of moderate activity (*e.g.*, jobs that involve walking ≤ 3 km per hour, jobs heavy on one arm, jobs moderately heavy on both arms) and work of high activity (*e.g.*, jobs that involve walking > 3 km per hour, jobs heavy on both arms, physical labor such as digging, bending and heavy lifting). These three types of work correspond to an occupational energy expenditure of < 8 , 8-12 and > 12 kJ/min. Occupational sitting time distinguishes between individuals sitting > 6 -8, 2-6 and < 2 working hours per day. Cumulative measures were calculated by taking into account all reported jobs and the duration of each reported job.

Because a substantial number of women could not be classified, the job type could not be coded, they never had paid employment or information on occupational history was missing, we did not analyze occupational energy expenditure and sitting time in women. We considered classifying women who did not report any paid employment but reported to be homemakers at baseline in the middle category of occupational energy expenditure and sitting time, however, this type of imputation would have introduced misclassification, and 27.6% of female sub-cohort members could still not be classified (in contrast with 10.1% of male sub-cohort members).

Baseline non-occupational physical activity

Baseline non-occupational physical activity was calculated based on two questions. The first (open-ended) question was 'How many minutes do you spend on average per day walking or cycling to your work, to go shopping or to walk your dog?' The reported time spent per day was categorized into ≤ 10 , > 10 -30, > 30 -60 and > 60 minutes per day. The second question was 'How many hours of your leisure time do you spend on average per week on 1) recreational cycling/walking, 2) gardening/doing odd jobs and 3) sports/gymnastics?' Answering possibilities were never, < 1 , 1-2 and > 2

hours per week. The time spent on these activities and the minutes spent per day on walking or cycling to work, to go shopping or to walk the dog were summed to obtain an overall measure of baseline non-occupational physical activity with categories ≤ 30 , $>30-60$, $>60-90$ and >90 minutes per day.

Former sports participation

Former sports participation was assessed by means of the question ‘In case you formerly played sports, please indicate what sport(s), was it in a competition system, how many hours per week did you spend on each sport (exercise included) and in what period?’ A maximum of three sports could be reported. Participants were categorized as ever having played sports if any information was given on the type(s) of sport(s) played, the number of hours spent on a sport per week and the period in time in which participants’ played a particular sport.

The total duration of former sports participation was calculated accounting for overlap. The type of sport was used to assign a metabolic equivalents score (MET-score) for intensity based on the compendium of Ainsworth et al.²⁶ The number of hours per week spent on a particular sport was derived immediately from the questionnaire. With respect to intensity and the number of hours per week of sports participation, we investigated the longest played sport.

Information on age and the period of former sports participation was used to differentiate between participants who played sports after puberty and those who started playing sports before puberty. Puberty was defined as ages 12-15 years in boys and the period between two years before menarche to one year after menarche in girls. The mean age at menarche was 13.7 years (range: 9-22 years).

Physical activity patterns

For purposes of sensitivity analyses, we combined the cumulative measures of occupational energy expenditure and sitting time with the measure of baseline non-occupational physical activity in men because men low in occupational energy expenditure or high in occupational sitting time may be active outside their job and vice versa. In these measures of physical activity in daily life, we distinguished between four groups of individuals based on low versus high occupational energy expenditure or sitting time and

low versus high non-occupational physical activity. We also combined the measure of occupational energy expenditure in men and the measure of baseline non-occupational physical activity in women with the measure of former sports participation and the measure of sports participation around puberty because sports activity is likely only a relatively small part of total physical activity.

Statistical analyses

Survival analysis was done using Cox proportional hazards model using the statistical package Stata (Stata Corp., College Station, TX) to estimate hazard ratios (HR) and corresponding 95% confidence intervals (CI). Hazard ratios were estimated for men and women separately to enhance comparability with previous studies. To account for the additional variance introduced by sampling the sub-cohort from the entire cohort, standard errors were estimated using the robust Huber-White sandwich estimator.²⁷

The proportional hazards assumption was tested using the scaled Schoenfeld residuals.²⁸ Because significant results emerge easily from this test when dealing with a large number of cases and long follow-up (as in our cohort), it is recommended to use more than one method to assess the proportional hazards assumption.²⁹ We also visually inspected the log-log transformed hazard curves and performed analyses for short and long follow-up by splitting the duration of follow-up at 8 years.

Hazard ratios and 95% confidence intervals were estimated with age-adjusted and multivariate adjusted models and models adjusted for other types of physical activity. In the multivariate model we adjusted for potential confounders selected a priori from literature and those that introduced a more than 10% change in hazard ratios. Selected a priori were family history of colorectal cancer, smoking status, body mass index (BMI, kg/m²), total energy intake, and intake of alcohol, meat and processed meat. Other covariates considered were socioeconomic status, self-reported diabetes, and intake of folate, β -carotene, vitamin B6, vitamin C, vitamin E, fruits, vegetables, water through foods and fluids, calcium, total fat (energy-adjusted), magnesium and haem iron. None introduced a more than 10% change. Models for cumulative occupational energy expenditure and cumulative sitting time in men were furthermore adjusted for baseline non-occupational physical activity and vice versa. Models for cumulative

occupational energy expenditure and cumulative sitting time were also adjusted for each other. Models for former sports participation were adjusted for cumulative occupational energy expenditure in men and baseline non-occupational physical activity in women.

In an extra analysis, we checked whether BMI may be an intermediate factor in the studied associations by comparing results derived from models adjusted for BMI and those not adjusted for BMI. Analyses were repeated substituting BMI for trouser/skirt size as a proxy measure for waist circumference³⁰ because physical activity has been associated with both BMI and waist circumference.³¹ To check for the influence of preclinical disease, we excluded the first two years of follow-up in multivariate analyses for our main exposure variables. Statistical significance was indicated by a *P*-value smaller than 0.05 for two-sided testing.

RESULTS

Baseline characteristics are presented in **Table 1**. In both men and women, colorectal cancer cases as compared to sub-cohort members were not evidently less physically active. The median duration of employment in male sub-cohort members and colorectal cancer cases was 38 and 39 years, respectively. Fifty percent of sub-cohort members and colorectal cancer cases held their last held job until 1983 or later. With respect to potential confounders, the percentage of colorectal cancer cases with a family history of colorectal cancer was clearly higher than among sub-cohort members. **Table 2** furthermore shows that cumulative occupational energy expenditure and cumulative sitting time were significantly different measures.

Table 1. Baseline characteristics sub-cohort members and colorectal cancer cases in the Netherlands Cohort Study.

	Men				Women			
	Sub-cohort		CRC cases		Sub-cohort		CRC cases	
	(N ^a = 2070)		(N ^a = 1739)		(N ^a = 2078)		(N ^a = 1260)	
	N ^b	%	N ^b	%	N ^b	%	N ^b	%
Occupational PA								
Cumulative energy expenditure (kJ/min)								
T1	632		477		-	-	-	-
T2	620		607		-	-	-	-
T3	613		517		-	-	-	-
Cumulative sitting time (hours)								
T1	615		475		-	-	-	-
T2	636		499		-	-	-	-
T3	614		627		-	-	-	-
	Median		Median		Median		Median	
Cumulative duration work history (years)	38		39		-		-	
Stop year last held job	1983		1983		-		-	
	N ^b	%	N ^b	%	N ^b	%	N ^b	%
Non-occupational PA								
Baseline non-occupational PA								
≤30 min/day	364	17.7	269	15.6	480	23.4	346	27.8
>30-60	638	31.1	548	31.7	648	31.6	374	30.0
>60-90	390	19.0	348	20.2	474	23.1	292	23.4
>90	660	32.2	562	32.5	451	22.0	234	18.8
Daily walking/biking								
≤10 min/day	687	33.5	579	33.5	597	29.1	393	31.5
>10-30	456	22.2	367	21.3	607	29.6	372	29.9
>30-60	477	23.3	428	24.8	587	28.6	349	28
>60	432	21.1	353	20.4	262	12.8	132	10.6
Recreational walking/biking								
Never	155	7.6	110	6.4	174	8.5	116	9.3
<1 hrs/week	417	20.3	352	20.4	497	24.2	287	23
1-2	644	31.4	503	29.1	663	32.3	426	34.2
>2	836	40.7	762	44.1	719	35.0	417	33.5
Gardening/doing odd jobs								
Never	307	15.0	218	12.6	778	37.9	466	37.4
<1 hrs/week	286	13.9	211	12.2	386	18.8	231	18.5
1-2	444	21.6	403	23.3	425	20.7	284	22.8
>2	1,015	49.5	895	51.8	464	22.6	265	21.3
Sports/gymnastics								
Never	1,484	72.3	1,183	68.5	1,344	65.5	842	67.6
<1 hrs/week	226	11	226	13.1	277	13.5	135	10.8
1-2	176	8.6	164	9.5	298	14.5	195	15.7
>2	166	8.1	154	8.9	134	6.5	74	5.9

Continues on next page



Continuing from previous page

Sports history

Former sports participation

Never	936	45.2	785	45.1	1,107	53.3	698	55.4
Ever	1,134	54.8	954	54.9	971	46.7	562	44.6

Total duration former sports

Never played	936	46.5	785	46.3	1,107	55.7	698	58.2
<11 (yrs)	477	23.7	389	22.9	476	23.9	255	21.3
≥11	598	29.7	523	30.8	405	20.4	246	20.5

Intensity longest played sport

Never played	936	46.7	785	46.3	1,107	55.9	698	58.3
<7 (METs)	273	13.6	243	14.3	467	23.6	260	21.7
≥7	796	39.7	668	39.4	406	20.5	239	20.0

Hrs/week longest played sport

Never played	936	47.3	785	47	1,107	56.5	698	59.3
<3	427	21.6	334	20	552	28.2	309	26.3
≥3	614	31.1	552	33	301	15.4	170	14.4

Sports around puberty

Never played	936	46.5	785	46.3	1,107	55.9	698	58.4
After puberty	497	24.7	421	24.8	423	21.4	239	20.0
Started before puberty	578	28.7	491	28.9	450	22.7	259	21.7

Potential
confounders

	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%
Age (years)	61.3 (4.2)		61.9 (4.1)		61.4 (4.3)		62.3 (4.0)	
Family history of CRC (yes)		5.3		9.1		6.0		9.7
Smoking status								
Never smoker		12.8		11.7		57.6		58.4
Ex		51.8		58.0		20.8		21.6
Current		35.4		30.3		21.6		20.0
Alcohol intake								
0 g/day		14.6		12.0		32.3		32.5
0.1-29		70.7		70.4		64.1		62.5
≥30		14.7		17.6		3.6		5.1
BMI (kg/m ²)	24.9 (2.6)		25.2 (2.6)		25.0 (3.5)		25.0 (3.5)	
Meat intake (g/day)	106 (42.9)		105 (41.1)		93.1 (39.5)		92.5 (40.1)	
Processed meat intake (g/day)	17.2 (17.6)		17.7 (17.6)		11.2 (12.4)		11.2 (11.7)	
Total energy intake (kcal/day)	2,172 (510)		2,148 (489)		1,688 (397)		1,679 (304)	

Abbreviations: N, number of; PA, physical activity

^a Excluding sub-cohort members and colorectal cancer cases with missing data on potential confounders.^b Numbers do not add up to the total number of sub-cohort members or cases because of missing data on exposure variables.

Table 2. Cross tabulation of cumulative occupational energy expenditure (kJ/min) and cumulative occupational sitting hours in male sub-cohort members, excluding those with missing values on potential confounders.

Cumulative occupational energy expenditure (kJ/min) ^b	Men		
	Cumulative number of occupational sitting time (hours) ^a		
	T1 (low) N (%)	T2 N (%)	T3 (high) N (%)
T1 (low)	169 (27.5)	274 (43.1)	189 (30.8)
T2	88 (14.3)	201 (31.6)	331 (53.9)
T3 (high)	358 (58.2)	161 (25.3)	94 (15.3)
Total	615	636	614

Note: the P -value for the χ^2 test in male sub-cohort members was <0.001.

^a The range of the cumulative number of occupational sitting hours in tertiles 1, 2 and 3 was 365-33,580, 3,3945-59,860 and 60,225-155,125, respectively.

^b The range of the cumulative amount of occupational energy expenditure in tertiles 1, 2 and 3 was 7-264, 265-370 and 370.5-1,317.5 kJ/min, respectively.

Tables 3 and 4 show the multivariate results for men and women, respectively. The proportional hazards assumption was violated on some occasions, yet log-log transformed hazard curves seemed proportional upon visual inspection. Age-adjusted results are not presented as these were comparable with multivariate results.

Table 3. Hazard ratios (HR) and 95% confidence intervals (CI) for colorectal cancer subsites in relation to physical activity in men.

	Colon				Proximal colon				Distal colon				Rectum			
	PY ^a at risk	N	HR ^b	95% CI	N	HR ^b	95% CI	N	HR ^b	95% CI	N	HR ^b	95% CI	N	HR ^b	95% CI
Occupational PA																
Cumulative energy expenditure (kJ/min) ^c																
T1 (low)	8701	332	1	(referent)	136	1	(referent)	181	1	(referent)	98	1	(referent)	98	1	(referent)
T2	3437	403	1.14	(0.94-1.39)	179	1.22	(0.94-1.59)	211	1.12	(0.88-1.42)	139	1.40	(1.05-1.87)	139	1.40	(1.05-1.87)
T3 (high)	8336	299	0.85	(0.69-1.05)	139	0.94	(0.71-1.25)	146	0.79	(0.60-1.01)	163	1.73	(1.30-2.29)	163	1.73	(1.30-2.29)
<i>P-trend</i>				0.14			0.66			0.06						<0.001
Cumulative sitting time (hrs) ^d																
T1 (low)	8176	289	0.73	(0.60-0.89)	139	0.83	(0.64-1.08)	138	0.66	(0.51-0.85)	139	1.02	(0.78-1.33)	139	1.02	(0.78-1.33)
T2	8947	326	0.76	(0.62-0.92)	139	0.78	(0.60-1.00)	176	0.76	(0.60-0.96)	117	0.76	(0.58-1.00)	117	0.76	(0.58-1.00)
T3 (high)	8353	419	1	(referent)	176	1	(referent)	224	1	(referent)	144	1	(referent)	144	1	(referent)
<i>P-trend</i>				0.002			0.15			0.001						0.92
Non-occupational PA																
Baseline non-occupational PA																
≤30 min/day	4546	166	1	(referent)	78	1	(referent)	82	1	(referent)	67	1	(referent)	67	1	(referent)
>30-60	8996	356	1.04	(0.83-1.32)	166	1.02	(0.75-1.38)	168	1.01	(0.74-1.36)	134	1.00	(0.72-1.39)	134	1.00	(0.72-1.39)
>60-90	5343	225	1.14	(0.88-1.47)	85	0.92	(0.65-1.30)	135	1.37	(0.99-1.89)	89	1.15	(0.80-1.64)	89	1.15	(0.80-1.64)
>90	9008	360	1.06	(0.84-1.33)	151	0.92	(0.68-1.25)	196	1.18	(0.88-1.60)	148	1.10	(0.80-1.52)	148	1.10	(0.80-1.52)
<i>P-trend</i>				0.61			0.42			0.09						0.40

Continues on next page

Continuing from previous page													
Daily walking/biking													
≤10 min/day	9086	369	1	(referent)	175	1	(referent)	181	1	(referent)	139	1	(referent)
>10-30	6447	226	0.88	(0.71-1.09)	93	0.77	(0.58-1.03)	119	0.94	(0.71-1.23)	103	1.04	(0.78-1.40)
>30-60	6469	288	1.11	(0.90-1.36)	117	0.96	(0.73-1.26)	157	1.22	(0.94-1.57)	103	1.06	(0.79-1.41)
>60	5891	224	0.94	(0.76-1.17)	95	0.83	(0.62-1.10)	124	1.09	(0.83-1.42)	93	1.03	(0.77-1.38)
<i>P-trend</i>				0.89			0.34			0.23			0.79
Recreational walking/biking													
Never	1867	68	1	(referent)	30	1	(referent)	34	1	(referent)	28	1	(referent)
<1 hrs/week	5570	229	1.03	(0.73-1.45)	98	1.00	(0.63-1.58)	121	1.08	(0.69-1.69)	88	0.99	(0.61-1.60)
1-2	8709	314	0.85	(0.61-1.18)	135	0.83	(0.54-1.30)	164	0.88	(0.57-1.35)	129	0.91	(0.57-1.44)
>2	11,747	496	1	(0.73-1.38)	217	0.99	(0.64-1.51)	262	1.06	(0.70-1.60)	193	1.02	(0.65-1.59)
<i>P-trend</i>				0.99			0.99			0.86			0.80
Gardening/doing odd jobs													
Never	3959	140	1	(referent)	69	1	(referent)	59	1	(referent)	55	1	(referent)
<1 hrs/week	3995	138	1.00	(0.74-1.35)	49	0.73	(0.48-1.11)	84	1.42	(0.96-2.09)	46	0.83	(0.54-1.29)
1-2	6061	257	1.25	(0.96-1.63)	113	1.12	(0.79-1.59)	133	1.51	(1.06-2.15)	106	1.30	(0.90-1.87)
>2	13,878	572	1.16	(0.92-1.48)	249	1.03	(0.76-1.40)	305	1.48	(1.08-2.03)	231	1.20	(0.86-1.67)
<i>P-trend</i>				0.13			0.35			0.04			0.09
Sports/													
gymnastics													
Never	19,878	732	1	(referent)	320	1	(referent)	380	1	(referent)	310	1	(referent)
<1 hrs/week	3214	151	1.18	(0.93-1.49)	70	1.28	(0.94-1.74)	76	1.12	(0.84-1.51)	56	1.05	(0.76-1.46)
1-2	2461	124	1.39	(1.07-1.81)	59	1.51	(1.08-2.11)	59	1.29	(0.93-1.79)	30	0.78	(0.51-1.17)
>2	2341	100	1.20	(0.91-1.59)	31	0.87	(0.57-1.30)	66	1.52	(1.11-2.10)	42	1.15	(0.79-1.66)
<i>P-trend</i>				0.02			0.30			0.01			0.94
Continues on next page													



Continuing from previous page

Sports history

Former sports participation													
Never	12,736	499	1	(referent)	213	1	(referent)	264	1	(referent)	208	1	(referent)
Ever	15,404	616	1	(0.85-1.16)	269	0.96	(0.78-1.18)	323	1	(0.83-1.22)	234	1.10	(0.89-1.36)
Total duration former sports													
Never played	12,736	499	1	(referent)	213	1	(referent)	264	1	(referent)	208	1	(referent)
≤11 (yrs)	6503	250	1	(0.83-1.22)	102	0.97	(0.74-1.27)	138	1.04	(0.82-1.33)	87	0.82	(0.62-1.08)
≥11	8110	338	1.03	(0.85-1.23)	158	1.14	(0.89-1.45)	167	0.95	(0.76-1.20)	135	0.98	(0.76-1.25)
<i>P-trend</i>				0.79			0.32			0.72			0.76
Intensity longest played sport													
Never played	12,736	499	1	(referent)	213	1	(referent)	264	1	(referent)	208	1	(referent)
<7 (METs)	3841	155	1.01	(0.80-1.28)	70	1.08	(0.79-1.47)	82	1.01	(0.75-1.35)	60	0.95	(0.69-1.32)
≥7	10,676	432	1.03	(0.87-1.21)	189	1.06	(0.85-1.34)	223	1	(0.81-1.23)	162	0.90	(0.72-1.14)
<i>P-trend</i>				0.77			0.59			0.97			0.39
Hrs/week longest played sport													
Never played	12,736	499	1	(referent)	213	1	(referent)	264	1	(referent)	208	1	(referent)
<3	6026	218	0.91	(0.74-1.11)	100	0.99	(0.75-1.30)	113	0.88	(0.68-1.14)	82	0.80	(0.60-1.07)
≥3	8102	350	1.1	(0.92-1.31)	150	1.11	(0.87-1.42)	183	1.08	(0.86-1.35)	137	1.02	(0.80-1.30)
<i>P-trend</i>				0.38			0.41			0.58			0.99
Sports around puberty													
Never played	12,736	499	1	(referent)	213	1	(referent)	264	1	(referent)	208	1	(referent)
After puberty	6853	278	1.08	(0.89-1.30)	128	1.18	(0.91-1.52)	141	1.02	(0.80-1.31)	98	0.87	(0.67-1.14)
Started before puberty	7760	310	0.97	(0.81-1.17)	132	0.98	(0.76-1.25)	164	0.97	(0.77-1.22)	124	0.94	(0.73-1.21)
<i>P-trend</i>				0.82			0.98			0.80			0.56

Abbreviations: PA, physical activity; PY, person years; T, tertile

^aNumbers do not add up to the same total because of missing values on exposure variables.^bAdjusted for age, family history of colorectal cancer, smoking status, alcohol intake, BMI, meat intake, processed meat intake and total energy intake.^cThe range of the cumulative amount of occupational energy expenditure in tertiles 1, 2 and 3 was 7-264, 265-370 and 370.5-1,317.5 kJ/min, respectively.^dThe range of the cumulative number of occupational sitting time in tertiles 1, 2 and 3 was 365-33,580, 3,3945-59,860 and 60,225-155,125 hours, respectively.

Table 4. Hazard ratios (HR) and 95% confidence intervals (CI) for colorectal cancer subsites in relation to physical activity in women.

	Colon		Proximal colon		Distal colon		Rectum	
	PY ^a at risk	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI
Non-occupational PA								
Baseline non-occupational PA								
≤30 min/day	6924	1 (referent)	144	1 (referent)	107	1 (referent)	67	1 (referent)
>30-60	9669	0.75 (0.60-0.93)	154	0.78 (0.60-1.02)	111	0.74 (0.55-1.00)	70	0.77 (0.53-1.10)
>60-90	6995	0.84 (0.67-1.06)	123	0.87 (0.66-1.16)	87	0.82 (0.60-1.12)	52	0.78 (0.53-1.15)
>90	6792	0.70 (0.55-0.88)	96	0.71 (0.53-0.96)	72	0.69 (0.50-0.96)	38	0.59 (0.39-0.90)
<i>P-trend</i>		0.01		0.07		0.06		0.02
Daily walking/biking								
≤10 min/day	8724	1 (referent)	160	1 (referent)	125	1 (referent)	75	1 (referent)
>10-30	9015	0.88 (0.71-1.08)	158	0.93 (0.72-1.20)	107	0.82 (0.62-1.10)	69	0.90 (0.63-1.27)
>30-60	8711	0.87 (0.70-1.07)	138	0.88 (0.67-1.14)	102	0.84 (0.63-1.12)	67	0.90 (0.64-1.28)
>60	3930	0.79 (0.60-1.04)	61	0.85 (0.60-1.19)	43	0.75 (0.52-1.10)	16	0.47 (0.27-0.83)
<i>P-trend</i>		0.07		0.25		0.12		0.03
Recreational walking/biking								
Never	2439	1 (referent)	47	1 (referent)	41	1 (referent)	21	1 (referent)
<1 hrs/week	7225	0.77 (0.56-1.05)	108	0.78 (0.53-1.16)	86	0.70 (0.46-1.06)	60	1.00 (0.59-1.70)
1-2	9913	0.88 (0.65-1.18)	191	1.02 (0.70-1.48)	122	0.71 (0.48-1.05)	77	0.93 (0.55-1.55)
>2	10,803	0.81 (0.60-1.08)	171	0.88 (0.61-1.29)	128	0.70 (0.47-1.04)	69	0.78 (0.46-1.32)
<i>P-trend</i>		0.51		0.87		0.26		0.18
Gardening/doing odd jobs								
Never	11,381	1 (referent)	185	1 (referent)	160	1 (referent)	75	1 (referent)
<1 hrs/week	5794	1.01 (0.80-1.26)	112	1.24 (0.95-1.63)	58	0.73 (0.53-1.01)	42	1.15 (0.77-1.72)
1-2	6276	1.1 (0.89-1.36)	120	1.24 (0.95-1.62)	83	0.97 (0.72-1.29)	52	1.31 (0.90-1.90)
>2	6928	0.88 (0.71-1.09)	100	0.93 (0.70-1.23)	76	0.80 (0.59-1.07)	58	1.32 (0.91-1.90)
<i>P-trend</i>		0.48		0.94		0.25		0.09
Sports/gymnastics								
Never	19,693	1 (referent)	351	1 (referent)	254	1 (referent)	144	1 (referent)
<1 hrs/week	4118	0.72 (0.56-0.93)	56	0.75 (0.55-1.03)	39	0.73 (0.51-1.05)	31	1.03 (0.68-1.57)
1-2	4499	0.99 (0.79-1.24)	84	1.07 (0.81-1.42)	56	0.97 (0.71-1.34)	39	1.24 (0.84-1.83)
>2	2070	0.94 (0.67-1.31)	26	0.75 (0.48-1.18)	28	1.09 (0.70-1.69)	13	0.89 (0.49-1.62)
<i>P-trend</i>		0.48		0.44		0.91		0.66

Continues on next page

Continuing from previous page													
Sports history													
Former sports participation													
Never	16,439	525	1	(referent)	289	1	(referent)	219	1	(referent)	123	1	(referent)
Ever	14,323	411	1.10	(0.94-1.29)	232	1.08	(0.88-1.32)	166	1.13	(0.90-1.41)	105	0.99	(0.75-1.31)
Total duration former sports													
Never played	16,439	525	1	(referent)	289	1	(referent)	219	1	(referent)	123	1	(referent)
sports	7,027	183	0.85	(0.69-1.04)	103	0.87	(0.68-1.13)	75	0.83	(0.62-1.10)	48	0.94	(0.66-1.34)
<11 (yrs)	5,956	187	0.97	(0.78-1.20)	105	0.98	(0.75-1.27)	76	0.96	(0.72-1.29)	44	1	(0.69-1.45)
													0.95
<i>P for trend</i>													
Intensity longest played sport													0.58
Never played	16,439	525	1	(referent)	289	1	(referent)	219	1	(referent)	123	1	(referent)
sports	6,877	190	0.91	(0.74-1.11)	111	0.97	(0.76-1.25)	75	0.85	(0.64-1.14)	50	1.01	(0.71-1.43)
<7 (METs)	5,983	178	0.90	(0.73-1.12)	97	0.88	(0.67-1.16)	75	0.94	(0.70-1.26)	42	0.94	(0.65-1.38)
													0.80
<i>P for trend</i>													
Hrs/week longest played sport													0.51
Never played	16,439	525	1	(referent)	289	1	(referent)	219	1	(referent)	123	1	(referent)
sports	8,147	224	0.89	(0.74-1.08)	128	0.93	(0.73-1.19)	90	0.86	(0.65-1.12)	58	0.99	(0.71-1.37)
<3	4,451	127	0.89	(0.70-1.13)	68	0.85	(0.63-1.15)	55	0.95	(0.68-1.32)	31	0.94	(0.62-1.42)
													0.77
<i>P for trend</i>													
Sports around puberty													0.50
Never played	16,439	525	1	(referent)	289	1	(referent)	219	1	(referent)	123	1	(referent)
sports	6,230	173	0.87	(0.71-1.08)	97	0.89	(0.68-1.16)	72	0.88	(0.65-1.18)	40	0.89	(0.61-1.29)
Started before	6,636	195	0.93	(0.76-1.14)	110	0.96	(0.74-1.23)	78	0.90	(0.68-1.19)	51	1.04	(0.73-1.47)
puberty													0.93
													0.38
<i>P for trend</i>													
													0.61

Abbreviations: PA, physical activity; PY, person years

^a Numbers do not add up to the same total because of missing values on exposure variables.^b Adjusted for age, family history of colorectal cancer, smoking status, alcohol intake, BMI, meat intake, processed meat intake and total energy intake.

Less cumulative occupational sitting time was significantly associated with lower hazard ratios for colon cancer in men. Hazard ratios were significantly decreased and lowest for distal colon cancer in men in the lowest tertile versus men in the highest tertile (HR: 0.66, 95% CI: 0.51-0.85, P -trend=0.001). Higher levels of cumulative occupational physical activity appeared associated with decreased hazard ratios for distal colon cancer, however, estimates were not statistically significant. In contrast, there was a significant positive dose-response relationship with rectal cancer risk in men for cumulative occupational energy expenditure (highest tertile versus lowest: HR: 1.73, 95% CI: 1.30-2.29, P -trend<0.001).

Baseline non-occupational physical activity was significantly inversely associated with colon cancer and rectal cancer risk in women (P -trend=0.01 and 0.02, respectively), but not in men. Hazard ratios were significantly decreased and lowest for rectal cancer in women who had >90 minutes of baseline non-occupational physical activity per day versus \leq 30 minutes per day (HR: 0.59, 95% CI: 0.39-0.90). Decreased hazard ratios were also observed for proximal colon cancer (HR: 0.71, 95% CI: 0.53-0.96) and distal colon cancer (HR: 0.69, 95% CI: 0.50-0.96).

When investigating components of baseline non-occupational physical activity, hazard ratios for rectal cancer were observed to be particularly decreased in women who walked/cycled >60 minutes per day versus \leq 10 minutes per day (HR: 0.47, 95% CI: 0.27-0.83, P -trend=0.03). No associations were observed with recreational walking/cycling, gardening/doing odd jobs and sports/gymnastics in women. In men, there were significant positive dose-response relationships with distal colon cancer risk for gardening/doing odd jobs (>2 hours per week versus never: HR: 1.48, 95% CI: 1.08-2.03, P -trend=0.04) and sports/gymnastics (>2 hours per week versus never: HR: 1.52, 95% CI: 1.11-2.10, P -trend=0.01), but there were no associations with other components of baseline non-occupational physical activity.

Further analyses demonstrated that associations with cumulative occupational energy expenditure and cumulative sitting time were unchanged after adjustment for baseline non-occupational physical activity in men. Associations for cumulative occupational energy expenditure were more pronounced for short follow-up, whereas this was the case for cumulative occupational sitting time for long follow-up (data not shown). When

adjusting cumulative occupational energy expenditure and cumulative sitting time for each other, hazard ratios were essentially the same (data not shown). Associations with baseline non-occupational physical activity were essentially the same after adjustment for cumulative occupational energy expenditure in men, and were evident in analyses for short and long follow-up in men and women (data not shown). Former sports participation by itself was not associated with any of the colorectal cancer endpoints in men and women (Tables 3 and 4, respectively), nor after adjustment for cumulative occupational energy expenditure in men and baseline non-occupational physical activity in women (data not shown).

The results of sensitivity analyses for physical activity patterns are presented in **Table 5** for men and **Table 6** for women. Patterns indicating higher physical activity levels in the daily lives of men were significantly inversely associated with the risk of distal colon cancer, although associations were less clear than when considering cumulative occupational energy expenditure, cumulative occupational sitting time and baseline non-occupational physical activity separately. A significant positive trend for rectal cancer risk in men was observed when investigating cumulative occupational energy expenditure in combination with baseline non-occupational physical activity.

Table 5. Hazard ratios (HR) and 95% confidence intervals (CI) for colorectal cancer subsites in relation to physical activity patterns in men.

	Colon			Proximal colon			Distal colon			Rectum		
	PY ^a at risk	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	
Cumulative occupational energy expenditure ^c and non-occupational PA ^d												
Low, low	2662	112	1 (referent)	56	1 (referent)	53	1 (referent)	37	1 (referent)	198	1.01 (0.69-1.49)	
Low, active	14,369	620	1.03 (0.79-1.33)	259	0.85 (0.62-1.18)	336	1.18 (0.84-1.65)	198	1.18 (0.84-1.65)	22	1.28 (0.71-2.31)	
Active, low	1350	43	0.79 (0.51-1.23)	20	0.73 (0.41-1.29)	21	0.82 (0.46-1.45)	22	0.82 (0.46-1.45)	140	1.48 (0.98-2.22)	
Active, active	6895	253	0.80 (0.60-1.07)	117	0.72 (0.50-1.05)	124	0.85 (0.58-1.24)	140	0.85 (0.58-1.24)		0.003	
<i>P-trend</i>			0.02		0.08		0.03		0.03			
Cumulative occupational sitting time ^e and non-occupational PA ^d												
Low, low	1167	56	1 (referent)	27	1 (referent)	27	1 (referent)	23	1 (referent)			
Low, active	7116	361	1.00 (0.69-1.45)	149	0.85 (0.53-1.36)	195	1.13 (0.70-1.81)	119	0.83 (0.49-1.39)			
Active, low	2855	99	0.74 (0.48-1.12)	49	0.76 (0.44-1.28)	47	0.72 (0.42-1.24)	36	0.64 (0.36-1.16)			
Active, active	14,148	512	0.75 (0.52-1.06)	227	0.68 (0.43-1.07)	265	0.80 (0.51-1.28)	219	0.78 (0.48-1.28)		0.44	
<i>P-trend</i>			0.001		0.03		0.01		0.01			
Occupational energy expenditure ^c and former sports												
Low, no sports	7120	298	1 (referent)	126	1 (referent)	160	1 (referent)	104	1 (referent)			
Low, sports	10,019	437	1.02 (0.84-1.24)	189	1.06 (0.81-1.37)	232	1 (0.78-1.27)	133	0.88 (0.66-1.18)			
Active, no sports	4075	152	0.84 (0.65-1.08)	71	0.92 (0.66-1.29)	72	0.73 (0.53-1.02)	82	1.38 (0.99-1.92)			
Active, sports	4262	147	0.77 (0.59-0.99)	68	0.83 (0.59-1.16)	74	0.73 (0.53-1.01)	81	1.28 (0.92-1.79)		0.03	
<i>P-trend</i>			0.02		0.21		0.02		0.02			
Occupational energy expenditure ^c and sports around puberty												
Low, no sports	7120	298	1 (referent)	126	1 (referent)	160	1 (referent)	104	1 (referent)			
Low, after puberty	4436	199	1.16 (0.91-1.49)	87	1.23 (0.89-1.69)	107	1.15 (0.85-1.56)	56	0.88 (0.61-1.27)			
Low, started before puberty	515	220	0.93 (0.73-1.17)	95	0.96 (0.70-1.30)	114	0.89 (0.66-1.18)	70	0.88 (0.62-1.24)			
Active, no sports	4075	152	0.83 (0.64-1.08)	71	0.92 (0.65-1.28)	72	0.73 (0.53-1.01)	82	1.39 (1.00-1.94)			
Active, after puberty	1978	68	0.76 (0.54-1.06)	38	0.98 (0.64-1.50)	29	0.61 (0.39-0.95)	37	1.26 (0.82-1.92)			
Active, started before puberty	2075	70	0.74 (0.53-1.04)	28	0.69 (0.43-1.10)	39	0.78 (0.51-1.18)	41	1.35 (0.88-2.06)			
<i>P-trend</i>			0.01		0.12		0.01		0.01			

Abbreviations: PA, physical activity; PY, person years

^aNumbers do not add up to the same total because of missing values on exposure variables.^bAdjusted for age, family history of colorectal cancer, smoking status, alcohol intake, BMI, meat intake, processed meat intake and total energy intake.^cLow and active correspond to tertiles 1 and 2 of cumulative occupational energy expenditure (kJ/min) and tertile 3 of cumulative occupational energy expenditure (kJ/min), respectively^dAs measured at baseline; low and active correspond to ≤ 30 min/day and >30 min/day, respectively.^eLow and active correspond to tertile 1 of cumulative occupational sitting time (hours) and tertiles 2 and 3 of cumulative occupational sitting time (hours), respectively.

Table 6. Hazard ratios (HR) and 95% confidence intervals (CI) for colorectal cancer subsites in relation to physical activity patterns in women.

	Colon			Proximal colon			Distal colon			Rectum	
	PY ^a at risk	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI	N cases ^a	HR ^b 95% CI
Non-occupational PA^c and former sports											
Low, no sports	4132	166	1 (referent)	87	1 (referent)	70	1 (referent)	40	1 (referent)		
Low, sports	2792	96	0.87 (0.63-1.20)	57	0.99 (0.67-1.47)	37	0.79 (0.51-1.24)	27	1.03 (0.61-1.73)		
Active, no sports	12,070	352	0.75 (0.59-0.95)	200	0.82 (0.61-1.10)	144	0.71 (0.51-0.98)	82	0.72 (0.48-1.08)		
Active, sports	11,386	310	0.69 (0.55-0.88)	173	0.75 (0.55-1.01)	126	0.66 (0.48-0.92)	78	0.74 (0.49-1.12)		
<i>P-trend</i>			0.002		0.03		0.02		0.08		
Non-occupational PA^c and sports around puberty											
Low, no sports	4132	166	1 (referent)	87	1 (referent)	70	1 (referent)	40	1 (referent)		
Low, after puberty	1348	33	0.59 (0.37-0.93)	17	0.58 (0.32-1.04)	16	0.68 (0.37-1.25)	8	0.63 (0.28-1.39)		
Low, started before puberty	1049	52	1.33 (0.86-2.00)	34	1.68 (1.03-2.74)	16	0.93 (0.51-1.72)	13	1.28 (0.65-2.54)		
Active, no sports	12,070	352	0.75 (0.59-0.95)	200	0.82 (0.61-1.10)	144	0.71 (0.51-0.98)	82	0.71 (0.48-1.07)		
Active, after puberty	4834	138	0.73 (0.55-0.97)	79	0.8 (0.56-1.15)	55	0.68 (0.46-1.01)	32	0.71 (0.43-1.17)		
Active, started before puberty	5557	142	0.65 (0.49-0.86)	76	0.67 (0.47-0.95)	61	0.65 (0.45-0.95)	38	0.72 (0.45-1.17)		
<i>P-trend</i>			0.003		0.03		0.03		0.15		

Abbreviations: PA, physical activity; PY, person years

^aNumbers do not add up to the same total because of missing values on exposure variables.^bAdjusted for age, family history of colorectal cancer, smoking status, alcohol intake, BMI, meat intake, processed meat intake and total energy intake.^cAs measured at baseline; low and active correspond to ≤ 30 min/day and >30 min/day, respectively.

When investigating cumulative occupational energy expenditure in men and baseline non-occupational physical activity in women in combination with former sports participation, we observed significant inverse dose-response relationships with colon cancer risk, but also a significant positive trend for rectal cancer risk in men (Tables 5 and 6). Compared with women low in physical activity who did not play sports, active women who played sports were observed to have a significantly decreased distal colon cancer risk (HR: 0.66, 95% CI: 0.48-0.92), more so than active women who did not play sports (HR: 0.71, 95% CI: 0.51-0.98). However, confidence intervals included both hazard ratios, warranting caution when interpreting this result.

When including the time period of sports participation, hazard ratios were lowest and significantly decreased in active men who started playing sports after puberty and active women who started playing sports before puberty (men: HR for distal colon cancer: 0.61, 95% CI: 0.39-0.95; women: HR for proximal colon cancer: 0.67, 95% CI: 0.47-0.95 and HR for distal colon cancer: 0.65, 95% CI: 0.45-0.95). Conversely, there was a significantly increased hazard ratio for proximal colon cancer in women with a low level of physical activity who started playing sports before puberty, although the confidence interval was wide (HR: 1.68, 95% CI: 1.03-2.74), and a non-significant increased hazard ratio for rectal cancer. In men, non-significant increased hazard ratios for rectal cancer were evident as well.

Additional analyses provided no indication that BMI or waist circumference may be an intermediate factor (data not shown). Excluding the first two years of follow-up gave comparable estimates (data not shown).

DISCUSSION

The inverse associations observed with colon cancer risk in men for cumulative occupational energy expenditure and cumulative sitting time and in women for baseline non-occupational physical activity are consistent with the findings in numerous previous observational studies, suggesting a protective effect of physical activity.^{1-15, 32} Experimental evidence includes a randomized controlled trial on the effect of a 12-month exercise intervention in previously sedentary men and women. In this trial, changes were found in some biomarkers for colorectal cancer (colon crypt cell proliferation,

expression of Bcl-2 and Bax),^{33, 34} but not others (concentrations of colon mucosal prostaglandins and C-reactive proteins levels).^{35, 36}

Decreases in risk in our study ranged between 22-53% for the highest physical activity levels as compared with the lowest. It has been estimated by Friedenreich et al.¹⁵ that approximately 8% of male colon cancer cases and 17% of female colon cancer cases in Europe could be prevented if physical activity levels were increased above the sedentary level (defined as <600 MET-minutes over 7 days). These estimates rose to approximately 17% and 19%, respectively, when a more stringent cut-off point was used. Other studies have reported similar estimates.^{37, 38}

Our study adds to the current body of observational evidence that the potential protective effect of physical activity may be particularly evident for distal colon cancer risk. With respect to a possible mechanism involved, it has been shown previously that chromosomal instability is more associated with tumours of the distal colon, whereas promoter hypermethylation is more associated with proximal tumours,³⁹ and environmental factors such as physical activity may act through one but not another. The potential protective effect might also extend to rectal cancer risk as we observed a strongly decreased rectal cancer risk for more than 60 minutes of daily walking/cycling versus 10 minutes or less in women. An association between physical activity and a reduced rectal cancer risk was also reported by Howard et al.⁷ and Slattery et al.⁴⁰ However, we cannot explain the positive dose-response relationship that we observed between rectal cancer risk and cumulative occupational energy expenditure in men, and chance findings cannot be excluded.

Special emphasis may be placed on our results for cumulative occupational sitting time as sedentary behavior is becoming more common and may differ from low physical activity as shown in our cohort in Table 2. Illustrative in this respect may be a Danish cohort study that found colon cancer risk not associated with occupational and leisure time physical activity, yet observed a borderline significant inverse association with the number of activities engaged in, with low numbers thought indicative of inactivity or a sedentary lifestyle.⁴¹

The best predictor of a decreased distal colon cancer risk in men in our study was cumulative occupational sitting time. The best predictor in women was baseline non-occupational physical activity. Together, these

findings suggest that regular, if not daily, physical activity may protect against distal colon cancer. In addition, being physically active over a long period of time seems important as the median duration of the longest held job in men in our sub-cohort (which is considered representative for the whole cohort) was 38 years. In women, it may be assumed that baseline non-occupational physical activity was representative for some time before baseline or even a considerable part of women's lives as 50.9% of female sub-cohort members reported to be exclusively homemakers at baseline.

That consistently higher physical activity levels may confer the greatest decrease in colon cancer risk was previously suggested by Lee et al.⁴² in their prospective study of Harvard alumni. In this study, physical activity was assessed twice, 11-15 years apart. Activity levels at either assessment alone were not associated with colon cancer risk, yet alumni who reported to be moderately or highly physically active at both assessments had a substantially decreased risk. More recently, data from the Cancer Prevention II Study have shown that physical activity consistently at or above the sample median over 15 years was associated with a significantly 55% lower risk of colon cancer mortality as compared with physical activity below the median over 15 years. Change in physical activity was not associated with colon cancer mortality, and neither measure was associated with colon cancer incidence.⁴³

The observed increased distal colon cancer risks for gardening/doing odd jobs and sports/gymnastics in men in our study are surprising. Because overall baseline non-occupational physical activity was inversely associated with colon cancer risk in women and potential protective effects were observed in men for cumulative occupational energy expenditure and cumulative occupational sitting hours, we suspect that compensation effects may be in play. It is conceivable that men who were inactive in the job compensated for the lack of physical activity in their spare time. Men who were retired at the time of filling in the baseline questionnaire (43.7% of male sub-cohort members) may have also reported high non-occupational physical activity levels not representative of most part of their lives. In comparison, a recent case-control study of Canadian men who worked in the period between 1920-1980 has found high occupational physical activity inversely associated with colon cancer risk, but not recreational physical activity.¹⁴

When investigating former sports participation, we observed inverse associations with distal colon cancer risk when analyzing the combined effect of daily physical activity and former sports participation in sensitivity analyses. These analyses should be interpreted cautiously because measures of physical activity are involved that reflect different time periods. Although the decreased hazard ratios may be mostly reflecting the effect of daily physical activity, there was slight suggestion of a small added benefit of former sports participation. In contrast, two studies on cancer risk in Finnish athletes⁴⁴ and Norwegian athletes⁴⁵ did not find a reduced colon cancer risk as compared with the general population or with other sub-groups of athletes, except for Norwegian ball game athletes. However, in this group there were only two cases.⁴⁵

Our observation that sports participation started before puberty in active participants showed significantly decreased colon cancer risks, yet in combination with low daily physical activity showed no or a significantly increased risk, is puzzling. Previously, one case-control study found physical activity during youth inversely associated with colon cancer risk,⁴⁶ while another did not.⁴⁷ Howard et al⁷ have suggested that recent physical activity may be more strongly associated with a reduction in colorectal cancer risk than physical activity in the distant past. In our cohort, it has been shown that individuals exposed to famine in childhood and adolescence during the Hunger Winter were at a significantly lower risk of colorectal cancer,⁴⁸ especially risk of a tumour characterized by the promoter hypermethylation phenotype.⁴⁹ It could be that in our specific study population physical activity as an early life exposure was not extreme enough to influence future cancer risk regardless of physical activity levels in adulthood. On the other hand, chance findings cannot be ruled out.

Strengths of our study include the prospective design and high completeness of follow-up, making selection and information bias unlikely. The NLCS also enabled us to adjust for important risk factors. However, as physically active individuals may be individuals with a generally healthy lifestyle, residual confounding cannot be excluded.

A limitation of this study may be the single baseline measurement of physical activity and dietary factors, and that our measures of physical activity were not tested for validity and reproducibility. Another limitation may be that physical activity was self-reported. However, our measure of

(cumulative) occupational energy expenditure and sitting time were based on occupational history using an expert assessment scheme, which likely provides a more accurate assessment than self-reports. It has been shown that individuals are able to report accurately on their occupational history.⁵⁰ In addition, our physical activity measures cover a large part of total lifetime physical activity, which may be difficult to assess otherwise and made possible the exploration of physical activity patterns by combining different measures. Future studies may be encouraged to also examine lifetime physical activity patterns.

In conclusion, our findings are generally consistent with the generally inverse association found of physical activity with colon cancer risk. Decreased risks were particularly evident for distal colon cancer. Results for rectal cancer were mixed. Measures of regular long-term physical activity seemed to predict risk best.

REFERENCES

1. Colditz GA, Cannuscio CC, Frazier AL. Physical activity and reduced risk of colon cancer: implications for prevention. *Cancer Causes Control* 1997;8:649-67.
2. Slattery ML. Physical activity and colorectal cancer. *Sports Med* 2004;34:239-52.
3. Samad AK, Taylor RS, Marshall T, Chapman MA. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* 2005;7:204-13.
4. Harriss DJ, Cable NT, George K, Reilly T, Renehan AG, Haboubi N. Physical activity before and after diagnosis of colorectal cancer: disease risk, clinical outcomes, response pathways and biomarkers. *Sports Med* 2007;37:947-60.
5. World Cancer Research Fund / American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AICR, 2007.
6. Wolin KY, Lee IM, Colditz GA, Glynn RJ, Fuchs C, Giovannucci E. Leisure-time physical activity patterns and risk of colon cancer in women. *Int J Cancer* 2007;121:2776-81.
7. Howard RA, Freedman DM, Park Y, Hollenbeck A, Schatzkin A, Leitzmann MF. Physical activity, sedentary behavior, and the risk of colon and rectal cancer in the NIH-AARP Diet and Health Study. *Cancer Causes Control* 2008;19:939-53.
8. Moradi T, Gridley G, Bjork J, Dosemeci M, Ji BT, Berkel HJ, et al. Occupational physical activity and risk for cancer of the colon and rectum in Sweden among men and women by anatomic subsite. *Eur J Cancer Prev* 2008;17:201-8.
9. Nilsen TI, Romundstad PR, Petersen H, Gunnell D, Vatten LJ. Recreational physical activity and cancer risk in subsites of the colon (the Nord-Trondelag Health Study). *Cancer Epidemiol Biomarkers Prev* 2008;17:183-8.
10. Harriss DJ, Atkinson G, Batterham A, George K, Cable NT, Reilly T, et al. Lifestyle factors and colorectal cancer risk (2): a systematic review and meta-analysis of associations with leisure-time physical activity. *Colorectal Dis* 2009;11:689-701.

11. Pan SY, DesMeules M. Energy intake, physical activity, energy balance, and cancer: epidemiologic evidence. *Methods Mol Biol* 2009;472:191-215 [Abstract only].
12. Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer* 2009;100:611-6.
13. Spence RR, Heesch KC, Brown WJ. A systematic review of the association between physical activity and colorectal cancer risk. *Scand J Med Sci Sports* 2009;19:764-81.
14. Parent ME, Rousseau MC, El-Zein M, Latreille B, Desy M, Siemiatycki J. Occupational and recreational physical activity during adult life and the risk of cancer among men. *Cancer Epidemiol* 2010.
15. Friedenreich CM, Neilson HK, Lynch BM. State of the epidemiological evidence on physical activity and cancer prevention. *Eur J Cancer* 2010;46:2593-604.
16. McTiernan A. Mechanisms linking physical activity with cancer. *Nat Rev Cancer* 2008;8:205-11.
17. Fuemmeler BF, Pendzich MK, Tercyak KP. Weight, dietary behavior, and physical activity in childhood and adolescence: implications for adult cancer risk. *Obes Facts* 2009;2:179-86.
18. Casparie M, Tiebosch ATMG, Burger G, Blauwgeers H, Van de Pol A, Van Krieken JHJM, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 2007;29:19-24.
19. van den Brandt PA, Schouten LJ, Goldbohm RA, Dorant E, Hunen PMH. Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol* 1990;19:553-8.
20. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of municipalities by cancer registries and PALGA using hospital discharge data. *Tijdschr Soc Gezondheidsz* 1994;72:80-4.
21. van den Brandt PA, Goldbohm RA, van 't Veer P, Volvovics A, Hermus RJJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol* 1990;43:285-95.

22. Nevo Table: Dutch food composition table 1986-1987. The Hague, Netherlands: Voorlichtingsbureau voor de Voeding, 1986.
23. Goldbohm RA, van den Brandt PA, Brants HAM, van 't Veer P, Al M, Sturmans F, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr* 1994;48:253-65.
24. Goldbohm RA, van 't Veer P, van den Brandt PA, van 't Hof MA, Brants HAM, Sturmans F, et al. Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr* 1995;49:420-9.
25. Hettinger TH, Mueller BH, Gebhard H. Ermittlung des Arbeitsenergieumsatzes bei Dynamisch Muskulaerer Arbeit. Schriftenreihe der Bundesarbeit fuer Arbeitsschutz. Dortmund, 1989.
26. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Jr., Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc* 1993;25:71-80.
27. Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics* 1994;50:1064-72.
28. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982;69:239-41.
29. Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. New York: Wiley, 1997.
30. Hughes LA, Schouten LJ, Goldbohm RA, van den Brandt PA, Weijnenberg MP. Self-reported clothing size as a proxy measure for body size. *Epidemiology* 2009;20:673-6.
31. May AM, Bueno-de-Mesquita HB, Boshuizen H, Spijkerman AM, Peeters PH, Verschuren WM. Effect of change in physical activity on body fatness over a 10-y period in the Doetinchem Cohort Study. *Am J Clin Nutr* 2010;92:491-9.
32. Lynch BM. Sedentary behavior and cancer: a systematic review of the literature and proposed biological mechanisms. *Cancer Epidemiol Biomarkers Prev* 2010;19:2691-709.
33. McTiernan A, Yasui Y, Sorensen B, Irwin ML, Morgan A, Rudolph RE, et al. Effect of a 12-month exercise intervention on patterns of

- cellular proliferation in colonic crypts: a randomized controlled trial. *Cancer Epidemiol Biomarkers Prev* 2006;15:1588-97.
34. Campbell KL, McTiernan A, Li SS, Sorensen BE, Yasui Y, Lampe JW, et al. Effect of a 12-month exercise intervention on the apoptotic regulating proteins Bax and Bcl-2 in colon crypts: a randomized controlled trial. *Cancer Epidemiol Biomarkers Prev* 2007;16:1767-74.
 35. Abrahamson PE, King IB, Ulrich CM, Rudolph RE, Irwin ML, Yasui Y, et al. No effect of exercise on colon mucosal prostaglandin concentrations: a 12-month randomized controlled trial. *Cancer Epidemiol Biomarkers Prev* 2007;16:2351-6.
 36. Campbell KL, Campbell PT, Ulrich CM, Wener M, Alfano CM, Foster-Schubert K, et al. No reduction in C-reactive protein following a 12-month randomized controlled trial of exercise in men and women. *Cancer Epidemiol Biomarkers Prev* 2008;17:1714-8.
 37. de Vries E, Soerjomataram I, Lemmens VE, Coebergh JW, Barendregt JJ, Oenema A, et al. Lifestyle changes and reduction of colon cancer incidence in Europe: A scenario study of physical activity promotion and weight reduction. *Eur J Cancer* 2010;46:2605-16.
 38. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjønneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. *Bmj* 2010;341:c5504.
 39. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* 2007;50:113-30.
 40. Slattery ML, Curtin K, Wolff RK, Herrick JS, Caan BJ, Samowitz W. Diet, physical activity, and body size associations with rectal tumour mutations and epigenetic changes. *Cancer Causes Control* 2010;21:1237-45.
 41. Johnsen NF, Christensen J, Thomsen BL, Olsen A, Loft S, Overvad K, et al. Physical activity and risk of colon cancer in a cohort of Danish middle-aged men and women. *Eur J Epidemiol* 2006;21:877-84.

42. Lee IM, Paffenbarger RS, Jr., Hsieh C. Physical activity and risk of developing colorectal cancer among college alumni. *J Natl Cancer Inst* 1991;83:1324-9 [Abstract only].
43. Wolin KY, Patel AV, Campbell PT, Jacobs EJ, McCullough ML, Colditz GA, et al. Change in physical activity and colon cancer incidence and mortality. *Cancer Epidemiol Biomarkers Prev* 2010.
44. Pukkala E, Kaprio J, Koskenvuo M, Kujala U, Sarna S. Cancer incidence among Finnish world class male athletes. *Int J Sports Med* 2000;21:216-20.
45. Robsahm TE, Hestvik UE, Veierod MB, Fagerlie A, Nystad W, Engebretsen L, et al. Cancer risk in Norwegian world class athletes. *Cancer Causes Control* 2010;21:1711-9.
46. Levi F, Pasche C, Lucchini F, Tavani A, La Vecchia C. Occupational and leisure-time physical activity and the risk of colorectal cancer. *Eur J Cancer Prev* 1999;8:487-93.
47. Marcus PM, Newcomb PA, Storer BE. Early adulthood physical activity and colon cancer risk among Wisconsin women. *Cancer Epidemiol Biomarkers Prev* 1994;3:641-4.
48. Hughes LA, van den Brandt PA, Goldbohm RA, de Goeij AF, de Bruine AP, van Engeland M, et al. Childhood and adolescent energy restriction and subsequent colorectal cancer risk: results from the Netherlands Cohort Study. *Int J Epidemiol* 2010;39:1333-44.
49. Hughes LA, van den Brandt PA, de Bruine AP, Wouters KA, Hulsmans S, Spiertz A, et al. Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. *PLoS One* 2009;4:e7951.
50. Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiologic purposes. *Am J Epidemiol* 1983;118:583-91.

7

Body size, physical activity and risk of colorectal cancer with or without the CpG island methylator phenotype (CIMP):

Laura AE Hughes
Colinda CJM Simons
Piet A van den Brandt
R Alexandra Goldbohm
Anton FPM de Goeij
Adriaan P de Bruïne
Manon van Engeland
Matty P Weijenberg

PLoS ONE 2011, in press

ABSTRACT

Background: Colorectal cancer (CRC) is a disease associated with epigenetic instability. We investigated how measures of anthropometry and physical activity influence the risk of the CpG island methylator phenotype (CIMP) in CRC.

Methods: In the Netherlands Cohort Study (N=120,852), risk factors were self-reported at baseline in 1986. After 7.3 years of follow-up, 603 cases and 4631 sub-cohort members were available for case-cohort analysis. Methylation specific PCR on DNA from paraffin embedded tumour tissue was performed to determine CIMP status using the Weisenberger markers. Cox regression was used to calculate hazard rate ratios (HR) and 95% confidence intervals for CIMP (27.7%) and non-CIMP (72.3%) tumours according to BMI, BMI at age 20, BMI change, trouser/skirt size, height, and baseline physical activity. All models were adjusted for age and sex.

Results: BMI modeled per 5kg/m² increase was associated with both CIMP and non-CIMP tumours, however, HRs were attenuated when additionally adjusted for trouser/skirt size. Trouser/skirt size, per 2 size increase, was associated with both tumour sub-types, even after adjustment for BMI (CIMP HR: 1.20, 95%CI: 1.01-1.43; non-CIMP HR: 1.14, 95%CI: 1.04-1.28). Associations per 5 cm increase in height (CIMP HR: 1.14, 95%CI: 1.00-1.30; non-CIMP HR: 1.11, 95%CI: 1.01-1.22) were attenuated when mutually adjusting for body weight. BMI at age 20 was positively associated with increased risk of CIMP tumours and the association was significantly less pronounced for non-CIMP tumours (*P*-heterogeneity=0.01). Physical activity was inversely associated with both sub-types, but a dose-response association was observed only for non-CIMP tumours (*P*-trend=0.02).

Conclusion: Our findings suggest that measures of anthropometry reflecting a large body size increase the risk of both CIMP and non-CIMP tumours, and that body fat at young age may differentially influence risk. Physical activity appears to decrease the risk of CRC regardless of these molecular sub-types.

INTRODUCTION

It is well accepted that indicators of energy balance influence the risk of colorectal cancer (CRC). A high body mass index (BMI), waist circumference, and adult attained height are clear risk factors for CRC, while physical activity has been shown to be protective.¹ Although CRC is one of the best described cancers in terms of genetic and epigenetic events involved,²⁻⁵ little is known about how measures of anthropometry and physical activity are associated with different molecular sub-sets of this disease. Elucidating potential differences in risk between molecular sub-types of CRC may lead to a better understanding of CRC, treatment, and prevention. This is especially important as the global prevalence of overweight and obesity continues to rise.

A distinct characteristic of epigenetic instability in CRC is the CpG island methylator phenotype (CIMP), characterized by numerous promoter CpG island hypermethylated tumour suppressor- and DNA repair genes.⁶⁻⁹ This in turn is associated with transcriptional silencing of gene expression.¹⁰ Few studies have investigated associations between indicators of energy balance and CIMP status, and those that have, only considered BMI as an indicator of body size. In a case-control setting, Slattery *et al.* reported an association between a high BMI and CIMP low but not CIMP high colon tumours,¹¹ and no association between BMI and CIMP status in rectal tumours.¹² Vigorous physical activity was associated with both CIMP high and CIMP low colon tumours, but not rectal tumours.^{11, 12} It has been hypothesized that DNA methylation is a consequence of inflammation.^{13, 14} Central adiposity is also associated with chronic inflammation.¹⁵ Therefore, considering waist circumference as a risk factor for CIMP in addition to BMI is important.

Using the Netherlands Cohort Study on diet and cancer (NLCS), we investigated the association between BMI, clothing size (as a proxy for waist circumference) and physical activity and risk of developing a tumour characterized by CIMP. Furthermore, in attempt to elucidate if timing of exposure is important for modulating this risk, we also investigated associations according to BMI at age 20, BMI change, and adult-attained height, which is an indicator of *in utero* and early life exposures.¹⁶

METHODS

Study populations and design

The NLCS is a prospective cohort study that was initiated in 1986 to investigate the association between diet and the development of cancer. It includes 58,279 men and 62,573 women between the ages of 55-69 years at baseline who completed a self-administered food frequency questionnaire involving 150 food items as well as questions on dietary habits, lifestyle, health, and demographics.^{17, 18} Municipal registries from throughout the Netherlands were used to constitute an efficient sampling frame.¹⁷⁻¹⁹ The NLCS uses a case – cohort approach for data processing and analysis; case subjects were derived from the entire cohort, and the number of person-years at risk for the entire cohort was estimated from a sub-cohort of 5000 men and women who were randomly sampled from the full cohort at baseline. All sub-cohort members who reported prevalent cancer (excluding skin cancer) at baseline were excluded from analyses, leaving 4654. Further details of the NLCS design have been described.¹⁷⁻¹⁹

Incident CRC cases were identified by annual record linkage to nine regional cancer registries and a national pathology database (PALGA).²⁰ The completeness of cancer follow-up is almost 100%.²¹ Paraffin embedded tumour material from CRC patients was retrieved, as described previously.²² In total, 734 incident CRC patients were identified from a follow-up period of 7.3 years after baseline, excluding the first 2 years of follow-up, of whom a PALGA report of the lesion as well as sufficient DNA was available.²²

The study protocol was approved by the Medical Ethics Committees of the University Hospital Maastricht and TNO Nutrition. On recruitment, participants were informed in writing of the details of the study and its objectives. In accordance with the regulations at that time, written informed consent was obtained when participants returned the completed baseline questionnaire. Tumour material was collected after approval by the ethical review boards of Maastricht University, the National Cancer Registry, and PALGA.

Ascertainment of risk factors

Anthropometric variables

Height (cm), body weight (kg), and body weight at age 20 (kg) were self-reported on the baseline questionnaire. From these variables, BMI and

BMI at age 20, and BMI change were subsequently calculated. At baseline, individuals were also asked to report their lower body (trouser or skirt) clothing size from their clothing label (Dutch sizes). Trouser/skirt size has been shown to be an adequate proxy measure for waist circumference when predicting cancer risk in the NLCS, and details of how clothing size corresponds to waist measurements in men and women in this Dutch population has been published.²³ BMI, BMI at age 20, skirt/trouser size, and height were categorized into approximate sex-specific quartiles. As in previous NLCS analyses, BMI change was categorized as: $<0\text{kg/m}^2$, $0\text{-}4\text{ kg/m}^2$, $4\text{-}8\text{ kg/m}^2$, and $>8\text{kg/m}^2$.²⁴

Baseline physical activity

With respect to physical activity and CRC risk in the NLCS population, occupational physical activity appears to be more important for men and recreational physical activity for women (Simons *et al.*, submitted). Therefore, we used these two variables to create sex-specific categories of 'low,' 'intermediate' and 'high' physical activity.

Occupational physical activity was derived from data on participants' occupational history. Using information on the type of job and the duration, energy expenditure and sitting time was calculated for the longest and last held job. Energy expenditure was based on a rating system developed by Hettinger *et al.*²⁵ and distinguishes between work of low, moderate and high activity which corresponds to an energy expenditure of <8 , $8\text{-}12$ and >12 kJ/min. Men were categorized into the 'low' category if their occupational physical activity was <8 kJ/minute, 'intermediate' if they fell into the $8\text{-}12$ kJ/minute category, and high if their occupational physical activity was >12 kJ/minute.

Baseline non-occupational physical activity was calculated based on two questions. The first (open-ended) question was 'How many minutes do you spend on average per day walking or cycling to your work, to go shopping or to take out your dog?' The reported time spent per day was categorized into ≤ 10 , $>10\text{-}30$, $>30\text{-}60$ and >60 minutes per day. The second question was 'How many hours of your leisure time do you spend on average per week on 1) recreational cycling, walking, 2) gardening/doing odd jobs and 3) sports, gymnastics?' Answering possibilities were never, <1 hour per week, $1\text{-}2$ hours per week and >2 hours per week. The time spent on these

activities and the minutes spent per day on walking or cycling to work, to go shopping or to take out the dog were summed to obtain an overall measure of baseline non-occupational physical activity, with categories <30, >30-60, >60-90 and >90 minutes per day. Low physical activity was defined as <30 minutes/day, intermediate as 30-90 minutes/day, and high as >90 minutes/day.

Promoter Methylation Analyses

CIMP in tumour tissue of CRC cases was defined by CpG island promoter hypermethylation of at least 3 out of 5 methylation markers (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1*), as proposed by Weisenberger *et al.*⁹ were determined by bisulfite modification of 500 ng genomic DNA using a commercially available kit (Zymo Research), and subsequent methylation specific PCR (MSP).^{26, 27} We chose to use MSP as a method because it is effective, specific and does not require specific equipment. It has been shown that results from MSP are in accordance with other technologies, such as MethyLight.²⁸ Additionally, the methylation status of two other genes, *APC* and *MLH1*, were determined and we added them to the CIMP markers to create a methylation index of seven genes.

To facilitate MSP analysis on DNA retrieved from formalin-fixed, paraffin-embedded tissue, DNA was first amplified with flanking PCR primers that amplify bisulfite-modified DNA but do not preferentially amplify methylated or unmethylated DNA. The resulting fragment was used as a template for the MSP reaction. All PCRs were carried out with controls for unmethylated alleles (DNA from normal lymphocytes), methylated alleles [normal lymphocyte DNA treated *in vitro* with *SssI* methyltransferase (New England Biolabs, Ipswich, MA)] and a control without DNA. Ten microliters of each MSP reaction was directly loaded on to nondenaturing 6% polyacrylamide gels stained with ethidium bromide and visualised under UV illumination. The MSP analyses were successful for 81%, 79%, 79%, 90%, 83%, 93%, and 93% out of the 734 cases for *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1*, *MLH1*, and *APC* respectively.

Statistical analyses

Data were analyzed with Stata (version 10, Statacorp, College Station, TX, USA). Cox proportional hazards analysis using the case-cohort

approach was used to obtain hazard ratios (HR) and 95% confidence intervals (CI) for the association between measures of anthropometry and physical activity and CRC characterized by CIMP status. To improve statistical power, we considered men and women together. Tests for effect modification by sex were not statistically significant. The proportional hazards assumption was tested using the scaled Schoenfeld residuals and visual inspection of the hazard curves. To account for the additional variance introduced by sampling the sub-cohort from the entire cohort, standard errors were estimated using the robust option. Statistical significance was tested at the 0.05 level.

For all anthropometric variables in question, three models were considered. The first was adjusted only for age and sex. The second was additionally adjusted for variables identified as being associated with both CRC and energy balance from previous literature. These included family history of CRC (yes/no), smoking status (never smoker, ex-smoker, current smoker), socioeconomic status (level of education: primary school, junior high school, senior high school, higher vocational school, or university), total energy intake (kcal/day), alcohol intake (0, 0.1-4, 5-14, 15-29, ≥ 30 g/day), physical activity (low, medium, high as previously described), and consumption of red meat, fruit, vegetables, and grains (g/day). Finally, models were mutually adjusted for other anthropometric variables. Models including BMI and BMI at age 20 were mutually adjusted for skirt/trouser size, BMI change was adjusted for BMI at age 20, trouser/skirt size was adjusted for BMI, and height was adjusted for body weight.

We modeled physical activity adjusted for age and sex, and additionally adjusted for trouser/skirt size, family history of CRC (yes/no), smoking status (never smoker, ex-smoker, current smoker), socioeconomic status (level of education: primary school, junior high school, senior high school, higher vocational school, or university), total energy intake (kcal/day), alcohol intake (0, 0.1-4, 5-14, 15-29, ≥ 30 g/day), clothing size, and consumption of red meat, fruit, vegetables, and fiber (g/day).

To assess how measures of anthropometry and physical activity were associated with the extent of promoter methylation in the CRC tumours, we used the aforementioned methylation index to categorize cases into one of three groups: ‘0-1 genes methylated’, ‘2-3 genes methylated’, or ‘4-7 genes methylated’. Of the 734 cases, 556 had sufficient information to be classified into one of the three categories. Models including anthropometric variables

were adjusted for age, sex, and mutually adjusted for other anthropometric variables as previously described, and the model for physical activity was adjusted for age and sex.

Tests for heterogeneity were done to evaluate differences between sub-types of tumours (e.g., *CIMP* vs. non-*CIMP*) using the competing risks procedure in STATA. However, the SE for the difference of the log-hazard ratios from this procedure assumes independence of both estimated hazard ratios, which would overestimate that SE and thus overestimate the *P* values for their difference. Therefore, these *P* values and the associated confidence intervals were estimated based on a bootstrapping method that was developed for the case-cohort design, as described previously.²⁹ Each bootstrap analysis was based on 1000 replications.

RESULTS

Baseline and molecular characteristics are described in **Table 1**. Proportionally, *CIMP* CRC cases were more likely to be female, have a tumour in the proximal colon, and be older than non-*CIMP* cases.

Table 1: Characteristics of NLCS study participants according to colorectal cancer status after 7.3 years of follow-up

	MEN			WOMEN		
	Sub-cohort ^a	CIMP+	Non-CIMP	Sub-cohort	CIMP+	Non-CIMP
Baseline characteristics						
Total (% of cases)	2219	83 (25)	245 (75)	2390	81 (30)	190 (70)
Age (years) ^b	61.3 (4.2)	63.0 (4.1)	62.8 (4.1)	61.5 (4.3)	63.3 (4.0)	62.8 (4.0)
Height (cm)	176.4 (6.7)	176.5 (6.7)	176.9 (6.6)	165.1 (6.2)	166.5 (6.2)	166.2 (6.3)
BMI (kg/m ²)	25.0 (2.6)	25.8 (2.6)	25.3 (2.9)	25.1 (3.6)	25.4 (2.9)	25.6 (3.5)
BMI (kg/m ²) at age 20	21.8 (2.4)	22.4 (2.5)	21.8 (2.2)	21.4 (2.8)	21.5 (2.2)	21.6 (2.6)
BMI change						
Clothing size	51 (4)	52 (6)	52 (3)	44 (3)	44(3)	44(3)
Recreational physical activity (%)						
<30 min/day	19	21	13	27	33	31
30-60 min/day	31	30	32	31	23	32
60-90 min/day	18	20	18	22	18	22
>90 min/day	31	29	37	21	28	15
Occupational physical activity (%)						
<8 kJ/min	57	72	63	58	52	51
8-12 kJ/min	27	18	22	36	40	40
>12 kJ/min	16	11	15	6	8	10
Sports participation during youth (%)						
Never	50	52	49	62	73	62
Only after puberty	23	24	20	20	15	22
Before, during, and after puberty	27	24	31	19	12	16
Molecular characteristics						
Site (%)						
Proximal colon	-	54	20	-	67	25
Distal colon	-	24	39	-	18	37
Rectosigmoid	-	5	11	-	4	17
Rectum	-	17	30	-	11	22
MSI status (%)						
Stable	-	72	95	-	60	97
Unstable	-	28	5	-	40	3
BRAF V600E mutation status						
Wildtype	-	64	94	-	54	93
Mutated	-	36	6	-	46	7

^a From sub-cohort of 5000 individuals selected randomly at baseline

^b Mean (SD) or percentages where indicated

Associations between measures of anthropometry, physical activity and CRC risk according to CIMP status are shown in **Table 2**. After adjustment for age and sex, BMI modeled per 5 kg/m² increase was statistically significantly associated with CIMP tumours (HR: 1.29, 95% CI: 1.01-1.66), and a statistically significant dose-response relationship was observed when modeling quartiles of BMI (P -trend=0.02). However, when models were mutually adjusted for trouser/skirt size, these associations were greatly attenuated. BMI was also positively, though not statistically significantly, associated with non-CIMP tumours. These were also attenuated when mutually adjusted for trouser/skirt size. The HRs for CIMP and non-CIMP tumours did not differ significantly from one another (P -heterogeneity= 0.78).

BMI at age 20 modeled per 5kg/m² increase was associated with increased risk of both sub-types, but HRs did not reach statistical significance. When BMI at age 20 was modeled in quartiles, a statistically significant trend was observed for CIMP tumours (P =0.03). This trend became borderline significant when the model was adjusted for trouser/skirt size (P -trend =0.07). The test for heterogeneity between tumour sub-types was statistically significant with respect to BMI at age 20, even after adjustment for trouser/skirt size (P -heterogeneity=0.01). No statistically significant observations were observed with respect to BMI change.

Trouser/skirt size, modeled per 2 unit size increase, was associated with both CIMP (HR: 1.20, 95%CI: 1.05-1.37) and non-CIMP tumours (HR: 1.14, 95% CI: 1.05-1.24) after multivariate adjustment, and these associations remained even when models were mutually adjusted for BMI. When trouser/skirt size was considered in approximate quartiles, the test for trend was significant for CIMP (P =0.02) and borderline significant for non-CIMP (P =0.06) tumours, although these were attenuated when models were mutually adjusted for BMI. Associations observed for CIMP and non-CIMP tumours were not statistically significantly different from each other (P -heterogeneity= 0.61).

Height, modeled per 5cm increase was associated with both CIMP (HR: 1.16, 95% CI: 1.01-1.33) and non-CIMP (HR: 1.10, 95% CI: 1.00-1.21) tumours after multivariate adjustment. A borderline statistically significant trend was observed for risk of non-CIMP tumours when height was considered in quartiles (highest versus lowest quartile HR: 1.45, 95% CI:

1.03-2.06; $P=0.06$) (P -heterogeneity=0.98). These associations were attenuated when the model was mutually adjusted for body weight.

With low physical activity as the reference category, there was no inverse dose-response association between physical activity and CIMP tumours, although the HR for intermediate physical activity was statistically significant (HR: 0.50, 95%CI: 0.30-0.82). An inverse association was observed in a dose-response fashion for non-CIMP tumours (intermediate physical activity HR: 0.80, 95%CI: 0.61-1.04; high physical activity HR: 0.69, 95%CI: 0.47-0.96; P -trend= 0.02), however, associations with CIMP and non-CIMP tumours did not differ significantly from each other (P -heterogeneity= 0.33).

Table 2: Hazard Ratios (HR) and 95% Confidence Intervals (CI) for colorectal cancer characterized by CIMP, according to measures of anthropometry in the NLCS after 7.3 years of follow-up

	PY at risk	N cases	CIMP+			Non-CIMP			<i>P</i> -heterogeneity ^d
			HR (95%CI) ^a	HR (95%CI) ^b	HR (95%CI) ^c	N	HR (95%CI) ^a	HR (95%CI) ^b	
BMI at baseline									
Per 5 kg/m ²	24,272	117	1.29 (1.01-1.66)	1.25 (0.96-1.64)	1.04 (0.75-1.49)	318	1.19 (0.98-1.14)	1.19 (0.98-1.14)	1.08 (0.85-1.38)
Quartiles									
1	8080	29	1 (referent)	1 (referent)	1 (referent)	92	1 (referent)	1 (referent)	1 (referent)
2	4491	20	1.28 (0.71-2.30)	1.22 (0.67-2.23)	1.11 (0.62-2.00)	57	1.01 (0.71-1.45)	1.00 (0.69-1.43)	0.97 (0.67-1.40)
3	7901	44	1.52 (0.94-2.44)	1.45 (0.90-2.35)	1.22 (0.73-2.04)	109	1.19 (0.89-1.59)	1.18 (0.88-1.59)	1.09 (0.79-1.50)
4	3799	24	1.85 (1.07-3.21)	1.77 (1.01-3.11)	1.45 (0.90-2.35)	60	1.22 (0.86-1.75)	1.22 (0.84-1.75)	1.03 (0.69-1.54)
<i>P</i> -trend			0.02	0.03	0.32		0.15	0.18	0.71
BMI @ 20									
Per 5 kg/m ²	24,272	117	1.28 (0.96-1.72)	1.28 (0.96-1.72)	1.22 (0.90-1.65)	318	1.14 (0.92-1.39)	1.10 (0.90-1.34)	1.17 (0.94-1.44)
Quartiles									
1	6112	21	1 (referent)	1 (referent)	1 (referent)	76	1 (referent)	1 (referent)	1 (referent)
2	6139	24	1.17 (0.64-2.11)	1.23 (0.67-2.26)	1.16 (0.64-2.11)	91	1.20 (0.87-1.66)	1.21 (0.87-1.66)	1.24 (0.89-1.72)
3	6070	42	2.01 (1.18-3.43)	2.09 (1.22-3.59)	1.94 (1.14-3.32)	74	0.99 (0.71-1.38)	0.98 (0.70-1.37)	1.02 (0.72-1.43)
4	5951	30	1.53 (0.87-2.70)	1.57 (0.88-2.82)	1.43 (0.81-2.55)	77	1.08 (0.78-1.51)	1.04 (0.75-1.45)	1.12 (0.80-1.57)
<i>P</i> -trend			0.03	0.03	0.07		0.95	0.81	0.80
BMI change									
<0	2896	10	0.61 (0.31-1.21)	0.63 (0.32-1.24)	0.52 (0.25-1.07)	28	0.72 (0.47-1.10)	0.74 (0.48-1.13)	0.65 (0.42-1.02)
0-3.9	11,594	62	1 (referent)	1 (referent)	1 (referent)	158	1 (referent)	1 (referent)	1 (referent)
4.0-7.9	7709	34	0.77 (0.51-1.18)	0.74 (0.45-1.13)	0.89 (0.55-1.28)	102	0.94 (0.72-1.22)	0.94 (0.72-1.23)	1.00 (0.76-1.31)
≥8.0	2073	11	0.93 (0.48-1.79)	0.86 (0.44-1.68)	1.06 (0.55-2.05)	30	1.11 (0.73-1.68)	1.08 (0.70-1.66)	1.23 (0.79-1.89)
<i>P</i> -trend			0.89	0.86	0.37		0.28	0.37	0.09

Continues on next page

Continuing from previous page									
Trousers/skirt size at baseline									
Per 2 sizes	24,272	117	1.20 (1.06-1.35)	1.20 (1.05-1.37)	1.20 (1.01-1.43)	318	1.14 (1.05-1.24)	1.14 (1.05-1.24)	1.15 (1.04-1.28)
Quartiles									
1	9561	31	1 (referent)	1 (referent)	1 (referent)	104	1 (referent)	1 (referent)	1 (referent)
2	7805	43	1.65 (1.03-2.64)	1.66 (1.05-2.69)	1.62 (0.99-2.65)	109	1.16 (0.84-1.61)	1.17 (0.88-1.56)	1.14 (0.85-1.54)
3	4440	26	1.66 (0.98-2.83)	1.66 (0.97-2.85)	1.61 (0.87-2.97)	60	1.19 (0.86-1.65)	1.12 (0.80-1.58)	1.07 (0.74-1.54)
4	2465	17	2.01 (1.10-3.67)	2.01 (1.07-3.77)	1.90 (0.86-4.17)	45	1.46 (1.04-2.04)	1.53 (1.04-2.24)	1.39 (0.87-2.23)
<i>P-trend</i>			0.01	0.02	0.1		0.002	0.06	0.28
									0.61
Height at baseline									
Per 5 cm	24,272	117	1.14 (1.00-1.30)	1.16 (1.01-1.33)	1.05 (0.91-1.22)	318	1.11 (1.01-1.21)	1.10 (1.00-1.21)	1.05 (0.95-1.17)
Quartiles									
1	6589	28	1 (referent)	1 (referent)	1 (referent)	75	1 (referent)	1 (referent)	1 (referent)
2	6294	31	1.17 (0.70-1.98)	1.19 (0.71-2.01)	1.06 (0.63-1.80)	81	1.16 (0.84-1.61)	1.16 (0.83-1.62)	1.10 (0.79-1.53)
3	6612	32	1.18 (0.70-1.38)	1.19 (0.71-2.01)	0.99 (0.58-1.68)	85	1.19 (0.86-1.65)	1.13 (0.80-1.57)	1.08 (0.77-1.51)
4	4777	26	1.38 (0.79-2.38)	1.47 (0.82-2.61)	1.05 (0.58-1.89)	77	1.46 (1.04-2.04)	1.45 (1.03-2.06)	1.24 (0.86-1.79)
<i>P-trend</i>			0.28	0.23	0.96		0.04	0.06	0.31
									0.98
Physical activity^e									
Low	9468	58	1 (referent)	1 (referent)	-	162	1 (referent)	1 (referent)	-
Intermediate	9312	29	0.50 (0.30-0.82)	0.50 (0.30-0.81)	-	115	0.80 (0.61-1.04)	0.81 (0.61-1.07)	-
High	4252	21	0.82 (0.49-1.38)	0.81 (0.48-1.36)	-	44	0.67 (0.47-0.96)	0.69 (0.47-1.01)	-
<i>P-trend</i>			0.22	0.22	-		0.02	0.04	-

^a Models adjusted for age and sex
^b Models adjusted for age, sex, and additionally for family history of CRC, smoking status, socioeconomic status, total energy intake, alcohol intake, and consumption of red meat, fruit, vegetables, and grains. Models with anthropometric variables were additionally adjusted for baseline physical activity (based on recreational physical activity for women (low= <30 minutes/day, intermediate= 30-90 minutes/day, high= >90 minutes/day) and occupational physical activity at the longest held job for men (low= <8 kJ/minute, intermediate = 8-12 kJ/minute, and high= >12 kJ/minute). The model for physical activity was additionally adjusted for trousers/skirt size.
^c Models adjusted for age and sex, and mutually adjusted for other anthropometric variables. Models with BMI and BMI at age 20 were additionally adjusted for trousers/skirt size; the model for BMI change was additionally adjusted for BMI at age 20; the model with trousers/skirt size was additionally adjusted for BMI; the model with height was additionally adjusted for body weight.
^d *P* value for test that HR for two tumor subtypes are equal (based on model 3 for anthropometric variables, and model 1 for physical activity).
^e Physical activity variable is based on baseline non-occupational physical activity for females, and occupational physical activity in males, as described in the methods.



Associations for a tumour methylation index in relation to anthropometric risk factors and physical activity are shown in **Table 3**. There was no clear pattern with respect to the degree of methylation, however, when modeled per 2 unit size increase, trouser/skirt size was associated in a dose-response manner with tumours displaying the highest level of methylation (4-7 genes methylated HR: 1.29, 95%CI: 1.06-1.58; p-trend=0.08).

Table 3: Hazard ratios (HR)^a and 95 % confidence intervals (CI) for a methylation index^b according to anthropometric risk factors and physical activity in the NLCS 7.3 years of follow-up

	PY at risk	0-1 genes methylated		2-3 genes methylated		4-7 genes methylated	
		N	HR (95% CI)	N	HR (95% CI)	N	HR (95% CI)
BMI at baseline							
Per 5 kg/m ²	24,272	116	1.07 (0.74-1.54)	154	1.05 (0.75-1.48)	100	1.02 (0.67-1.56)
Quartiles							
1	8080	44	1 (referent)	42	1 (referent)	24	1 (referent)
2	4491	12	0.48 (0.24-0.97)	31	1.15 (0.70-1.90)	20	1.33 (0.71-2.50)
3	7901	40	0.85 (0.52-1.39)	50	1.20 (0.75-1.91)	37	1.15 (0.63-2.10)
4	3799	20	0.71 (0.38-1.33)	31	1.37 (0.79-2.39)	19	1.19 (0.52-2.70)
<i>P-trend</i>			0.43		0.28		0.75
BMI @ 20							
Per 5 kg/m ²	24,272	116	1.28 (0.95-1.73)	154	1.23 (0.93-1.63)	100	1.06 (0.75-1.49)
Quartiles							
1	6112	23	1 (referent)	37	1 (referent)	20	1 (referent)
2	6139	40	1.81 (1.07-3.06)	34	0.91 (0.56-1.46)	20	1.02 (0.54-1.92)
3	6070	24	1.07 (0.59-1.91)	39	1.06 (0.67-1.67)	38	1.84 (1.06-3.19)
4	5951	29	1.31 (0.75-2.29)	44	1.24 (0.80-1.94)	22	1.09 (0.59-2.03)
<i>P-trend</i>			0.82		0.28		0.31
BMI change							
<0 kg/m ²	6115	9	0.54 (0.25-1.15)	15	0.70 (0.39-1.28)	8	0.58 (0.26-1.28)
0-3.9	5973	59	1 (referent)	88	1 (referent)	52	1 (referent)
4.0-7.9	6100	39	1.05 (0.68-1.63)	39	0.70 (0.47-1.03)	28	0.81 (0.51-1.29)
≥8.0	6085	9	0.99 (0.47-2.08)	12	0.91 (0.49-1.72)	12	1.34 (0.69-2.57)
<i>P-trend</i>			0.28		0.63		0.35

Continues on next page

Continuing from previous page

Trouser size at baseline

Per 2 sizes	24,272	116	1.13 (0.96-1.34)	154	1.12 (0.97-1.29)	100	1.29 (1.06-1.58)
Quartiles							
1	9561	48	1 (referent)	55	1 (referent)	28	1 (referent)
2	7805	26	0.68 (0.40-1.14)	56	1.25 (0.84-1.87)	33	1.52 (0.86-2.68)
3	4440	25	1.06 (0.62-1.81)	24	0.90 (0.52-1.56)	24	1.79 (0.88-3.64)
4	2465	17	1.24 (0.59-2.58)	19	1.22 (0.63-2.38)	15	2.00 (0.83-4.80)
<i>P-trend</i>			0.56		0.87		0.08

Height at baseline

Per 5 cm	24,272	116	1.04 (0.89-1.22)	154	1.11 (0.96-1.28)	100	1.07 (0.92-1.24)
Quartiles							
1	6589	30	1 (referent)	36	1 (referent)	19	1 (referent)
2	6294	26	0.89 (0.52-1.51)	38	1.09 (0.68-1.74)	32	1.61 (0.91-2.88)
3	6612	33	1.06 (0.63-1.80)	41	1.11 (0.70-1.78)	27	1.23 (0.67-2.26)
4	4777	27	1.17 (0.65-2.11)	39	1.34 (0.80-2.25)	22	1.31 (0.67-2.55)
<i>P-trend</i>			0.52		0.30		0.68

Physical activity^c

Low	9468	56	1 (referent)	81	1 (referent)	47	1 (referent)
Intermediate	9312	48	0.85 (0.55-1.31)	56	0.82 (0.57-1.18)	28	0.60 (0.37-1.02)
High	4252	18	0.72 (0.41-1.28)	17	0.54 (0.31-0.94)	18	0.88 (0.50-1.54)
<i>P-trend</i>			0.25		0.02		0.42

^a All models were adjusted for age and sex. Models with BMI and BMI at age 20 were additionally adjusted for trouser/skirt size, models with BMI change were additionally adjusted for BMI at age 20, models with trouser/skirt size were additionally adjusted for BMI, and models with height were additionally adjusted for body weight

^b Methylation index including the five CIMP markers (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, and *SOCS1*), *MLH1*, and the *APC* gene

^c Based on baseline non-occupational physical activity for women (low= <30 minutes/day, intermediate= 30-90 minutes/day, high= >90 minutes/day) and occupational physical activity at the longest held job for men (low= <8 kJ/minute, intermediate = 8-12 kJ/minute, and high= >12 kJ/minute).

DISCUSSION

These data suggest that adult body fatness and height may increase the risk of CRC, but are not differentially associated with CIMP and non-CIMP tumours. Contrarily, BMI at age 20 may be a stronger risk factor for CIMP+ tumours. Baseline physical activity appears to decrease the risk of CRC regardless of CIMP status.

A major strength of this study is that we investigated associations in a prospective cohort setting. The NLCS has almost complete ascertainment of colorectal cancer and very little loss to follow-up. Although the number of total CRC after 7.3 years of follow-up in the NLCS was substantial, the number of cases with the CIMP phenotype was small. With limited power to detect associations, it is possible that some findings arose by chance. Another potential limitation of this study is that anthropometric variables were obtained by self-report. However, there are many examples in the literature showing that this method is a valid and reliable tool for assessing body weight and height in cohort studies.³⁰⁻³³

To our knowledge, associations between indicators of energy balance and CIMP status of colorectal tumours have been reported only by Slattery *et al.*, in a case-control setting.^{11, 12} In addition to the study design, there are differences between our studies which should be taken into account when comparing results.

The NLCS utilized the Weisenberger panel of genes to define CIMP (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1*), whereas the study of Slattery *et al.*, utilized the so-called Classic panel (*MINT1*, *MINT2*, *MINT31*, *p16* and *hMLH1*), as well as different cut-offs for CIMP tumours.^{34, 35} The 'right' definition of CIMP is still a highly debated topic, as is the ideal gene panel and the appropriate method of methylation detection.^{12, 36} While the Weisenberger panel has been validated, different markers of methylation may be more or less informative with respect to studying different exposures. The prevalence of CIMP in the NLCS population is higher than in the study by Slattery *et al.* (27% vs. 11%).¹² However, a difference in primer designs and PCR conditions may substantially change sensitivity and specificity of a particular marker for the detection of CIMP in CRC.³⁷ Therefore, it is likely that differences in prevalence are not due to the different methods per se, but rather a difference in choice of primers. The MSP analyses that have been conducted in the NLCS have a high detection signal, and subsequently, a

higher prevalence of CIMP has been observed. In the present study, we attempted to clarify our observations by constructing a methylation index with different cut-off points that included seven genes commonly methylated in CRC.

In the present study we considered colon and rectal tumours together to increase statistical power. A sensitivity analysis revealed that this did not bias our findings (data not shown). Furthermore, we suggest that idea of combining sub-localizations of tumours may be acceptable when studying molecular endpoints, because this may help explain differences in etiology according to sub-localization.

We observed that BMI was associated with both tumour CIMP and non-CIMP tumours; however, after adjustment for clothing size, these associations disappeared. This is in contrast to case-cohort data suggesting that BMI is associated only with CIMP negative colon tumours and not with rectal tumours.^{11, 12} In our study, trouser/skirt size appears to be a strong, independent predictor of both tumour sub-types, even after adjustment for BMI. This is logical, because waist circumference, an indicator of central adiposity, is a stronger predictor of CRC than BMI.^{1, 38} When we considered associations according to a methylation index, we did not observe clear associations with respect to BMI and degree of promoter methylation, however, we did observe that trouser/skirt size was associated with the highest level of methylation. That we observed associations between trouser/skirt size and both CIMP and non-CIMP tumours suggests that central adiposity may influence CRC risk through both a methylation and a non-methylation pathway.

Although very few studies have considered associations between BMI and CIMP, a number have considered endpoints in the same pathway as CIMP. Colorectal cancer has distinct molecular sub-sets, which evolve through different pathways.³⁹ The path to a serrated adenocarcinoma appears to take one of two major routes: the traditional serrated pathway or the sessile serrated pathway.⁴⁰ The sessile serrated pathway is characterized by a high degree of CIMP, BRAF V600E mutations, and ultimately develops into microsatellite instability (MSI).⁴⁰ MSI may serve as a marker for CIMP or other molecular events in CRC,⁴¹ therefore, it may be informative to consider the findings of the present study in the context of that research. Two case-control studies have reported that BMI appears associated⁴² with

microsatellite stable (MSS) tumours, and less with MSI tumours.⁴³ Neither study reported associations according to waist circumference. In a recent pooled analysis of NLCS and data from the Melbourne Collaborative Cohort Study, we observed similar associations, although the test for heterogeneity between the tumour sub-types was not statistically significant (Hughes *et al.*, submitted).

There is evidence to suggest that early life exposures influence epigenetic mechanisms associated with adult disease risk.^{44, 45} Therefore, we also investigated associations between BMI at age 20, BMI change, height and CIMP status. Height is a marker of an aggregated fetal and childhood experience, and can be considered a proxy measure for important nutritional exposures, which affect several hormonal and metabolic axes.¹ In the NLCS, we have observed that childhood and adolescent energy restriction is associated with a decreased risk of CRC later in life,^{46, 47} which is supported by other population based studies.⁴⁸⁻⁵⁰ We also recently reported that exposure to severe energy restriction during childhood and adolescence was associated with a low risk of developing a CIMP tumour.⁴⁵ Furthermore, pooled data suggest that taller individuals are at greater risk of developing a MSI tumour (Hughes *et al.* submitted). According to the present study, height is not differentially associated with the risk of tumours, however, we did observe significant heterogeneity with respect to BMI at age 20 and tumour sub-types. Taken together, our findings suggest that body size may differentially influence CIMP status during different periods of life, potentially affecting later CRC risk. The association between BMI at age 20 and CIMP tumours was stronger than with non-CIMP tumours, which is in line with previous findings for severe energy restriction during childhood and adolescence. Although our bootstrapping method is quite conservative, we did not observe a clear association with respect to BMI at age 20 and the methylation index and therefore we can not rule out that the differential association with CIMP status is a chance finding. The hypothesis that timing of exposure may influence epigenetic mechanisms requires further investigation.

That we did not observe any clear associations between BMI change and risk of tumours was surprising. This may indicate that metabolic changes in fat may be more important for modulating risk over time, rather than BMI. Alternatively, considering men and women together may have attenuated

these observations. Campbell *et al.* report that adult weight gain was associated with CRC in men, but not in women, and only with respect to individuals who gained more than 21 kg since age 20.⁴² Finally, only considering two time points may not be indicative of true BMI change.

Our findings with respect to physical activity support those of Slattery *et al.*,¹¹ and suggest that high levels of daily exercise are associated with a decreased risk of both CIMP and non-CIMP tumours. Observations with respect to our methylation index suggest that physical activity may be more protective of tumours with increasing methylation, but more research is required before firm conclusions can be drawn.

Preliminary evidence suggests that molecular markers can be used to classify colorectal cancers into distinct sub-types, which have implications for both etiology and prevention.⁴¹ Tumours arising from the sessile serrated pathway make up only 12% of all colorectal cancers.⁴⁰ As overweight and obesity are such strong risk factors for CRC, there is a possibility that these conditions may differentially influence risk through pathways and molecular mechanisms other than what we investigated here. More research is needed to clarify the association between indicators of energy balance and epigenetic mechanisms leading to CRC; preferably in a prospective cohort setting, with many cases.⁵¹ Furthermore, as the field of molecular pathological epidemiology⁴¹ continues to evolve, standardizing methods and definitions of molecular endpoints should be addressed. This will become especially critical as more opportunities for pooling data arise.

In conclusion, our findings suggest that measures of anthropometry reflecting a large body size increase the risk of both CIMP and non-CIMP tumours, and that body fat at young age may differentially influence risk. Physical activity appears to decrease the risk of CRC regardless of these molecular sub-types. Our observations reiterate the importance of a healthy body weight with respect to general CRC prevention.

REFERENCES

1. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AIRC, 2007.
2. Derks S, Postma C, Carvalho B, *et al.* Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis*. 2008; 29: 434-9.
3. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer*. 2004; 4: 988-93.
4. Sieber OM, Heinimann K, Tomlinson IP. Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer*. 2003; 3: 701-8.
5. Wong JJ, Hawkins NJ, Ward RL. Colorectal cancer: a model for epigenetic tumorigenesis. *Gut*. 2007; 56: 140-8.
6. Samowitz WS, Albertsen H, Herrick J, *et al.* Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology*. 2005; 129: 837-45.
7. Shen L, Toyota M, Kondo Y, *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A*. 2007; 104: 18654-9.
8. Toyota M, Ahuja N, Ohe-Toyota M, *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A*. 1999; 96: 8681-6.
9. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*. 2006; 38: 787-93.
10. Ferguson LR, Karunasinghe N, Philpott M. Epigenetic events and protection from colon cancer in New Zealand. *Environ Mol Mutagen*. 2004; 44: 36-43.
11. Slattery ML, Curtin K, Sweeney C, *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer*. 2007; 120: 656-63.

12. Slattery ML, Curtin K, Wolff RK, *et al.* Diet, physical activity, and body size associations with rectal tumour mutations and epigenetic changes. *Cancer Causes Control.* 2010; 21: 1237-45.
13. Hodge DR, Peng B, Cherry JC, *et al.* Interleukin 6 supports the maintenance of p53 tumour suppressor gene promoter methylation. *Cancer Res.* 2005; 65: 4673-82.
14. Kang GH, Lee HJ, Hwang KS, *et al.* Aberrant CpG island hypermethylation of chronic gastritis, in relation to aging, gender, intestinal metaplasia, and chronic inflammation. *The American journal of pathology.* 2003; 163: 1551-6.
15. Monteiro R, Azevedo I. Chronic inflammation in obesity and the metabolic syndrome. *Mediators of inflammation.* 2010; doi: [10.1155/2010/289645](https://doi.org/10.1155/2010/289645)
16. Okasha M, Gunnell D, Holly J, *et al.* Childhood growth and adult cancer. *Best Pract Res Clin Endocrinol Metab.* 2002; 16: 225-41.
17. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr.* 1994; 48: 253-65.
18. Goldbohm RA, van 't Veer P, van den Brandt PA, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr.* 1995; 49: 420-9.
19. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol.* 1990; 43: 285-95.
20. Van den Brandt PA, Schouten LJ, Goldbohm RA, *et al.* Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol.* 1990; 19: 553-8.
21. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of municipalities by cancer registries and PALGA using hospital discharge data. *Tijdschr Soc Gezondheidsz* 1994; 72: 80-4.
22. Brink M, de Goeij AF, Weijenberg MP, *et al.* K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis.* 2003; 24: 703-10.

23. Hughes LA, Schouten LJ, Goldbohm RA, *et al.* Self-reported clothing size as a proxy measure for body size. *Epidemiology*. 2009; 20: 673-6.
24. Pylypchuk RD, Schouten LJ, Goldbohm RA, *et al.* Body mass index, height, and risk of lymphatic malignancies: a prospective cohort study. *Am J Epidemiol*. 2009; 170: 297-307.
25. Hettinger TH, Mueller BH, Gebhard H. *Ermittlung des Arbeitsenergieumsatzes bei Dynamisch Muskulaerer Arbeit*. Dortmund 1989.
26. Derks S, Lentjes MH, Hellebrekers DM, *et al.* Methylation-specific PCR unraveled. *Cell Oncol*. 2004; 26: 291-9.
27. Herman JG, Graff JR, Myohanen S, *et al.* Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A*. 1996; 93: 9821-6.
28. Barault L, Charon-Barra C, Jooste V, *et al.* Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res*. 2008; 68: 8541-6.
29. de Vogel S, Bongaerts BW, Wouters KA, *et al.* Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. *Carcinogenesis*. 2008; 29: 1765-73.
30. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, *et al.* Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *Journal of epidemiology and community health*. 2000; 54: 143-8.
31. Nyholm M, Gullberg B, Merlo J, *et al.* The validity of obesity based on self-reported weight and height: Implications for population studies. *Obesity*. 2007; 15: 197-208.
32. Spencer EA, Appleby PN, Davey GK, *et al.* Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public health nutrition*. 2002; 5: 561-5.
33. Wada K, Tamakoshi K, Tsunekawa T, *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Int J Obes*. 2005; 29:1093-9.

34. Frazier ML, Xi L, Zong J, *et al.* Association of the CpG island methylator phenotype with family history of cancer in patients with colorectal cancer. *Cancer Res.* 2003; 63: 4805-8.
35. Park SJ, Rashid A, Lee JH, *et al.* Frequent CpG island methylation in serrated adenomas of the colorectum. *Am J Pathol.* 2003; 162: 815-22.
36. Samowitz WS. The CpG island methylator phenotype in colorectal cancer. *J Mol Diagn.* 2007; 9: 281-3.
37. Ogino S, Kawasaki T, Kirkner GJ, *et al.* Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn.* 2007; 9: 305-14.
38. Pischon T, Lahmann PH, Boeing H, *et al.* Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst.* 2006; 98: 920-31.
39. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology.* 2007; 50: 113-30.
40. Noffsinger AE. Serrated polyps and colorectal cancer: new pathway to malignancy. *Annual review of pathology.* 2009; 4: 343-64.
41. Ogino S, Stampfer M. Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. *J Natl Cancer Inst.* 2010; 102: 365-7.
42. Campbell PT, Jacobs ET, Ulrich CM, *et al.* Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumour microsatellite instability status. *J Natl Cancer Inst.* 2010; 102: 391-400.
43. Slattery ML, Curtin K, Anderson K, *et al.* Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumours. *J Natl Cancer Inst.* 2000; 92: 1831-6.
44. Heijmans BT, Tobi EW, Stein AD, *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008; 105: 17046-9.
45. Hughes LA, van den Brandt PA, de Bruine AP, *et al.* Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. *PloS one.* 2009; 4:e7951.

46. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Energy restriction early in life and colon carcinoma risk: results of The Netherlands Cohort Study after 7.3 years of follow-up. *Cancer*. 2003; 97: 46-55.
47. Hughes LA, van den Brandt PA, Goldbohm RA, *et al.* Childhood and adolescent energy restriction and subsequent colorectal cancer risk: results from the Netherlands Cohort Study. *Int J Epidemiol*. 2010; 39: 1333-44.
48. Frankel S, Gunnell DJ, Peters TJ, *et al.* Childhood energy intake and adult mortality from cancer: the Boyd Orr Cohort Study. *Bmj*. 1998; 316: 499-504.
49. Svensson E, Grotmol T, Hoff G, *et al.* Trends in colorectal cancer incidence in Norway by gender and anatomic site: an age-period-cohort analysis. *Eur J Cancer Prev*. 2002; 11: 489-95.
50. Svensson E, Moller B, Tretli S, *et al.* Early life events and later risk of colorectal cancer: age-period-cohort modelling in the Nordic countries and Estonia. *Cancer Causes Control*. 2005; 16: 215-23.
51. Ogino S, Cantor M, Kawasaki T, *et al.* CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut*. 2006; 55: 1000-6.

8

Body size and risk for colorectal cancers showing BRAF mutations or microsatellite instability:

A pooled analysis of the Netherlands Cohort Study and Melbourne Collaborative Cohort Study

Laura AE Hughes
Elizabeth J Williamson
Manon van England
Mark A Jenkins
Graham Giles
John Hopper
Melissa Southey
Joanne Young
Daniel Buchanan
Michael Walsh
Piet A van den Brandt
R Alexandra Goldbohm
Matty P Weijenberg
Dallas R English

Submitted for publication

ABSTRACT

Background: How body size influences risk of molecular sub-types of colorectal cancer (CRC) is unclear. We investigated whether measures of anthropometry differentially influence risk of tumours according to *BRAF* mutation and microsatellite instability (MSI) status.

Methods: Data from the Netherlands Cohort Study (N=120,852) and Melbourne Collaborative Cohort Study (N=40,514) were pooled, and included 734 and 717 colorectal cancer cases from each study, respectively. Hazard ratios (HR) and 95% confidence intervals for body mass index (BMI), waist measurement, and height were calculated and compared for sub-types defined by *BRAF* mutation and MSI status, measured from archival tissue.

Results: Results were consistent between studies. When pooled, BMI modelled in 5kg/m² increments was positively associated with *BRAF* wild type (HR: 1.16, 95%CI: 1.08-1.26) and MS-stable tumours (HR: 1.15, 95%CI: 1.06-1.24). Waist measurement was also associated with *BRAF* wild type (highest vs. lowest quartile, HR: 1.59, 95%CI: 1.33-1.90) and MS-stable tumours (highest vs. lowest quartile HR: 1.68, 95%CI: 1.31-2.15). The HRs for *BRAF* mutation tumours and MSI tumours were smaller and non-significant, but differences between the HRs by tumour sub-types were not significant. Height, modelled per 5 cm increase, was positively associated with *BRAF* wild type and *BRAF* mutation tumours, but the HR was greater for tumours with a *BRAF* mutation than *BRAF* wild type (HR: 1.23, 95%CI: 1.11-1.37, *P-heterogeneity*=0.03). Similar associations were observed with respect to height and MSI tumours (HR 1.26 (95%CI: 1.13-1.40, *P-heterogeneity*=0.02).

Conclusions: Generally, overweight increases the risk of CRC. Taller individuals have an increased risk of developing a tumour with a *BRAF* mutation or MSI

INTRODUCTION

Although it is well documented that body size influences colorectal cancer (CRC) risk,¹ how this risk differs according to molecular phenotype of tumours is less clear. Elucidating such differences may lead to a better understanding of CRC etiology and prevention.

One well defined sub-group of CRC arises through the (sessile) serrated pathway, via serrated epithelium or hyperplastic polyps.² A number of genetic and epigenetic abnormalities contribute to the resulting adenoma/carcinoma. The somatic V600E mutation of the *BRAF* proto-oncogene is thought to occur early in disease progression.³ Microsatellite instability (MSI), which occurs when short, repetitive DNA sequences undergo an increase or decrease in repeat length⁴ is thought to occur later in the pathway.² In sporadic CRC, MSI almost always arises through loss of function in the mismatch repair gene *MLH1*, resulting from promoter methylation.² An underlying abnormality of tumours harbouring a *BRAF* mutation and/or MSI is the CpG island methylator phenotype (CIMP),² which is characterized by numerous promoter CpG island hypermethylated tumour suppressor and DNA repair genes.⁵⁻⁹ This in turn is associated with transcriptional silencing of gene expression.¹⁰

Only a handful studies have investigated the association between body size and this pathway, and those were all case-control in design, and only considered BMI as a risk factor.^{4, 11-13} Central adiposity is thought to be a better predictor of CRC than BMI,¹⁴ and height reflects early-life environmental and hormonal exposures,¹ therefore, investigating these risk factors in addition to BMI is warranted.

The Netherlands Cohort Study (NLCS) and the Melbourne Collaborative Cohort Study (MCCS) are prospective cohorts in which colorectal tumour tissue samples have been collected and subsequently analyzed for molecular characteristics. We pooled data from individual participants from the two studies to produce a more accurate estimate of relative risk.¹⁵⁻¹⁶ The purpose of the present study was to investigate if body size (BMI, indicators of central adiposity, and height) differentially influence risk of tumours according to *BRAF* mutation and MSI status.

METHODS

Study populations and design

NLCS

The NLCS was initiated in 1986 to investigate the association between diet and the development of cancer. It includes 58,279 men and 62,573 women between the ages of 55-69 years at baseline who completed a self-administered food frequency questionnaire involving 150 food items as well as questions on dietary habits, lifestyle, health, and demographics. Municipal registries from throughout the Netherlands were used to constitute an efficient sampling frame. The NLCS uses a case – cohort approach for data processing and analysis; case subjects were derived from the entire cohort, and the number of person-years at risk for the entire cohort was estimated from a sub-cohort of 5000 men and women who were randomly sampled from the full cohort at baseline. All sub-cohort members who reported prevalent cancer (excluding skin cancer) at baseline were excluded from analyses, leaving 4654. Further details of the NLCS design have been described.¹⁷⁻¹⁹

Incident CRC cases were identified by annual record linkage to nine regional cancer registries and a national pathology database (PALGA).²⁰ The completeness of cancer follow-up is almost 100%.²¹ Paraffin embedded tumour material from CRC patients was retrieved, as described previously.²² In total, 734 incident CRC patients were identified from a follow-up period of 7.3 years after baseline, excluding the first 2 years of follow-up, of whom a PALGA report of the lesion as well as sufficient DNA was available.²²

The study protocol was approved by the Medical Ethics Committees of the University Hospital Maastricht and TNO Nutrition. Tumour material was collected after approval by the ethical review boards of Maastricht University, the National Cancer Registry, and PALGA.

MCCS

The MCCS is a prospective cohort study of 17,045 men and 24,469 women, who were between the age of 27 and 75 years at recruitment from 1990 to 1994 (almost all were aged 40-69 years), and includes 5425 migrants from Italy and 4535 from Greece. For this analysis, 202 participants who had a colorectal cancer diagnosed before baseline were excluded, leaving a total of 41,312 individuals. Subjects were recruited via the electoral rolls (registration to vote is compulsory for adults in Australia), advertisements,

and community announcements in local media (e.g., television, radio, and newspapers). Comprehensive lists of Italian and Greek surnames were also used to target southern European migrants listed in the phone books and on electoral rolls. A structured interview schedule was used to obtain information on potential risk factors. Information on current diet was obtained from a dietary questionnaire that contained a 121-item food frequency questionnaire that was developed for the MCCS.²³

Cases were participants who had a first diagnosis of invasive cancer of the colon or rectum during follow-up to 31 December 2004 identified by linkage to population-based cancer registries in all Australian states. Addresses and vital status of the subjects were determined by record linkage to electoral rolls, the National Death Index, Victorian death records, from electronic phone books, and from responses to mailed questionnaires and newsletters. Archival tumour tissue was sought for all primary, histopathologically confirmed adenocarcinomas diagnosed in Victoria.

The study protocol was approved by the Cancer Council Victoria's Human Research Ethics Committee. Participants gave written consent for participation and for the investigators to obtain their medical records.

Anthropometry

NLCS

Height (cm) and body weight (kg) were self-reported on the baseline questionnaire. BMI was subsequently calculated. At baseline, individuals were also asked to report their lower body (trouser or skirt) clothing size from their clothing label (Dutch sizes). Clothing size appears to predict cancer risk independently of BMI, and corresponds to waist measurements.²⁴

MCCS

Height (cm), weight (kg), and waist circumference (cm) were measured at baseline attendance for each participant according to written protocols that were based on standard procedures.²⁵ Weight was measured to 100 g using digital electronic scales, height to 1 mm using a stadiometer, and waist circumferences was measured to 1 mm using a 2-m metal anthropometric tape. BMI was calculated from weight and height.

Molecular Pathology

Although CIMP has been measured in both studies, it was not considered here because different methods of measurement were used, leading to different proportions of colorectal cancers showing the phenotype.

NLCS

BRAF V600E mutation analysis was done by a semi-nested PCR and subsequent RFLP analyses as previously described.²⁶ MSI was determined by a pentaplex PCR using the MSI markers BAT-26, BAT-25, NR-21, NR-22 and NR-24, as described in detail by Suraweera *et al.*²⁷ Tumours were classified as MSI-high if two or more markers showed instability and as MSI-low/MS-stable (MSS) if one or none of the markers examined showed instability.

MCCS

BRAF V600E mutation analysis was done by a real-time PCR-based allelic discrimination method.²⁸ Microsatellite instability (MSI) was examined using 10 microsatellite markers.²⁹ Tumours were classified as MSI-high, MSI-intermediate, MSI-low, and MS-stable, but for comparability with the NLCS, MSI-low and MS-stable were combined into one category.

Statistical Analyses

Data were analyzed with Stata (version 10, Statacorp, College Station, TX, USA). Hazard ratios (HR), 95% confidence intervals (95% CI), and P values were calculated using Cox regression with age as the time scale,³⁰ To account for its case-cohort design, survival times in the NLCS were adapted using Prentice weights.³¹

To estimate HRs separately for different molecular sub-types and to test their difference (i.e. *BRAF* mutation vs. *BRAF* wild type), Cox models based on competing risks were fitted using a data duplication method.³² Person-years of follow-up were calculated from baseline until the date of colorectal cancer diagnosis, death, or end of follow-up.

Study specific analyses

BMI (per 5 kg/m²) and height (per 5cm) were fitted as continuous covariates to estimate linear trends on the log hazard scale and as categorical variables. These variables were additionally considered as sex and study specific quartiles. With respect to waist measurements, trouser/skirt size and waist circumference were fitted categorically for NLCS and MCCS data,

respectively, according to approximate sex specific quartiles. All models were stratified by sex, and the MCCS was additionally stratified by ethnicity (based on country of birth and classified into two groups: Southern European (born in Italy or Greece) and Anglo-Celtic (born in Australia, New Zealand, United Kingdom, or Ireland)).³³

Several potential confounding variables were considered for multivariate analysis. For the NLCS these included family history of CRC (yes/no), smoking status (never smoker, ex-smoker, current smoker), education (primary school, junior high school, senior high school, higher vocational school, or university), total energy intake (kcal/day), alcohol intake (0, 0.1-4, 5-14, 15-29, ≥ 30 g/day), recreational physical activity (<30 minutes/day, 30-60 minutes/day, 61-90 minutes/day, >90 minutes/day), and consumption of red meat, fruit, vegetables, fiber, and grains (g/day). For women, hormonal factors such as contraceptive use (yes/no), HRT use, age at menarche, age at menopause, and number of children were also tested. For the MCCS, the variables considered were smoking status (current smoker, former smoker, never smoker), education (primary school, some high/technical school, completed high school, and completed tertiary degree/diploma), total dietary energy intake (kcal/day), current alcohol consumption (males 0, 1-39, ≥ 40 grams/day, females 0, 1-19, ≥ 20 grams/day), current level of physical activity (none, low, moderate, high), meat, fruit, and vegetable (servings/day) intake, fiber from cereal products, multivitamin, and fiber supplements (yes/no). For women, parity, total months of lactation for all live births, age at first live birth, age at menarche, HRT use (never, former, current), oral contraceptive pill usage (never, ever), menopausal status, and age at menopause were also considered. In the respective studies, no variables changed the estimated HRs by more than 10% and therefore none were included in the final models.

Pooled Analyses

Models were stratified by study, ethnicity, and sex. HRs, 95% CIs and p values were estimated for a 5 kg/m² increase in BMI and a 5cm increase in height. Waist measurement was not considered continuously for the pooled data, due to differences in measurement. BMI, waist measurement and height were additionally considered according to sex and study specific quartiles. The NLCS was coded to reflect a non-Southern European ethnicity.

Because of the increase in sample size, sex specific analyses were also conducted.

Tests based on Schoenfeld residuals and graphical methods using Kaplan-Meier curves showed no evidence that proportional hazard assumptions were violated for any analyses.

RESULTS

Generally, men and women in the NLCS were older, taller, and had a lower BMI than those in the MCCS. There was a higher proportion of current and ex-smokers in the NLCS than the MCCS (**Tables 1 and 2**).

Table 1: Baseline demographic, anthropometric, dietary and lifestyle characteristics of the Netherlands Cohort Study (NLCS)

	MEN	WOMEN
Total in cohort	58,279	62,573
Total in sub-cohort ^a	2232	2399
Age (years) ^b	61.3 (4.2)	61.5 (4.3)
Weight (kg)	77.8 (9.5)	68.5 (10.3)
Height (cm)	176.4 (6.7)	165.1 (6.3)
BMI (kg/m ²)	25.0 (2.6)	25.1 (3.6)
Trouser size (Dutch size)	51 (4)	44 (3)
Smoking (%)		
Former	51	20
Current	36	21
Never	13	59
Alcohol intake (%)		
0 g/day	14	33
0.1-4	21	36
5-14	28	19
15-29	23	9
≥30	15	3
Education (%)		
Primary school	27	35
Some high school	21	23
Completed high school	34	33
Completed tertiary school	18	8
Family History of CRC (%)		
No	95	94
Yes	5	6
Recreational Physical Activity (%)		
<30 mins/day	19	27
30-60	32	31
60-90	18	22
≥90	31	21
Total energy (kcal/day)	2148 (526)	1655 (417)
Red meat (g/day)	104.8 (44.2)	92.3 (41.2)
Fibre (g/day)	28.4 (9.0)	24.9 (7.3)
Folate (µg/day)	208.5 (76.3)	195.2 (70.9)
Hormone replacement therapy (%)		
Ever	-	12
Never	-	85
Unknown	-	3

^a The NLCS uses a case-cohort approach for data processing and estimating person time at risk for the entire cohort. This sub-cohort consists of 5000 individuals, of which, 4631 were available for the present analysis.

^b Mean (SD) or percentages where indicated

Table 2: Baseline demographic, anthropometric, dietary and lifestyle characteristics of the Melbourne Collaborative Cohort Study (MCCS)

	MEN	WOMEN
Total in cohort	16,942	24,370
Age (years) ^a	55.8 (8.8)	55.0 (8.6)
Weight (kg)	80.8 (11.8)	68.2 (12.4)
Height (cm)	172.4 (7.4)	159.8 (6.7)
BMI (kg/m ²)	27.2 (3.6)	26.7 (4.9)
Waist circumference (cm)	93.5 (10.0)	80.0 (11.8)
Smoking (%)		
Former	45	22
Current	15	9
Never	41	69
Alcohol intake (%) ^b		
Lifetime abstainer	14	39
Ex-drinker	5	3
Low intake	27	19
Medium intake	29	21
High intake	25	17
Education (%)		
Primary school	19	20
Some high school	31	43
Completed high school	25	18
Completed tertiary school	25	19
Recreational Physical Activity (%)		
None	23	22
Low	18	21
Moderate	34	36
High	25	20
Total energy (kcal/day)	2468 (777)	2023 (666)
Red meat (times/week)		
0-4	18	28
5-6	21	25
7-9	27	27
≥10	34	21
Fibre (g/day)	32.2 (12.7)	30.0 (11.2)
Folate (µg/day)	330.2 (149.0)	321.3 (143.9)
Hormone replacement therapy (%)		
Never	-	75
Former	-	9
Current	-	17

^a Mean (SD) or percentages where indicated^b Sex specific categorization

Of 734 incident tumours in the NLCS, 697 were successfully analyzed for *BRAF* V600E mutation status and 658 for MSI (**Table 3**). Of 717 tumours identified in the MCCS, *BRAF* mutation status was determined for 582 and MSI status for 585. Further details of the cancers in each cohort are in previous publications.(33-36) For both studies, 16% of cancers evaluated had *BRAF* mutations. In the NLCS, 14% of tumours from males and 19% from females had *BRAF* mutations, while in the MCCS, the proportions were 11% and 19%, respectively. Cases with *BRAF* mutation tended to be older and taller than non-*BRAF* cases. With respect to MSI status, 13% of the cancers evaluated in the NLCS and 15% in the MCCS were MSI-high. In the NLCS, 12% of tumours from males and 14% from females were MSI, while in the MCCS, the proportions were 12% and 18%, respectively. MSI-high cases tended to be taller than MSS cases.

Table 3: Characteristics of participants in the NLCS and MCCS by colorectal cancer status

	Colorectal cancer diagnosed during follow-up				
	No cancer	V600E BRAF mutation status		MSI status	
		Mutation	Wild type	MSI-high	MSI-stable
NLCS					
Total (%)	4631 ^a	112 (16) ^b	585 (84)	84 (13)	574 (87)
Sex (%)					
Male	48	47	57	50	55
Female	52	53	43	50	45
Age (years) ^c	61.4 (4.2)	62.8 (4.1)	63.0 (4.1)	62.9 (4.0)	63.5 (4.5)
Height (cm)					
Male	176.4 (6.7)	177.8 (7.9)	176.6 (6.6)	177.0 (7.3)	176.9 (6.8)
Female	165.1 (6.2)	166.5 (5.3)	166.0 (6.6)	167.2 (5.3)	166.1 (6.6)
BMI (kg/m ²)	25.1 (3.1)	25.4 (2.9)	25.6 (3.2)	25.6 (3.3)	25.5 (3.1)
Clothing size					
Male	51 (4)	52 (4)	52 (4)	52 (2)	52 (3)
Female	44 (3)	44 (3)	44(3)	44 (3)	44 (3)
MCCS					
Total (%)	40,595	95 (16)	487 (84)	90 (15)	495 (85)
Sex (%)					
Male	41	35	53	40	52
Female	59	65	47	60	48
Age (years)	55.2 (8.7)	62.4 (6.4)	60.4 (7.4)	60.4 (7.9)	60.7 (7.3)
Height (cm)					
Male	172.4 (7.4)	173.8 (6.5)	171.8 (7.0)	176.1 (5.9)	171.4 (6.9)
Female	159.8 (6.7)	161.4 (6.6)	159.4 (6.7)	159.7 (6.0)	159.9 (7.0)
BMI (kg/m ²)	26.9 (4.4)	26.8 (4.3)	27.5 (4.1)	26.9 (4.1)	27.5 (4.2)
Waist circumference (cm)					
Male	93.4 (10.0)	95.6 (9.5)	96.2 (9.4)	94.1 (8.9)	96.5 (9.4)
Female	79.9 (11.8)	82.7 (11.8)	82.6 (12.3)	83.1 (11.0)	82.6 (12.5)
Pooled Data					
Total (%)	45,226	207 (16)	1072 (84)	174 (14)	1069 (86)
Sex (%)					
Male	42	42	55	45	54
Female	58	58	45	55	46
Age	55.8 (8.5)	62.6 (5.2)	61.8 (6.0)	61.9 (6.7)	61.9 (5.9)
Height					
Male	172.9 (7.4)	176.3 (7.6)	174.5 (7.2)	176.6 (6.7)	174.4 (7.3)
Female	160.3 (6.8)	163.8 (6.5)	162.8 (7.4)	162.9 (6.8)	163.0 (7.4)
BMI	26.7 (4.4)	26.0 (3.7)	26.5 (3.8)	26.3 (3.7)	26.4 (3.8)
Site (%)					
Proximal colon	-	62	27	74	27
Distal colon	-	17	31	13	30
Rectosigmoid/Rectum	-	18	39	9	40
Unknown	-	2	3	4	3
MSI status (%)					
MSI-stable	-	57	92	-	-
MSI-high	-	43	8	-	-

^a Sub-cohort derived from total cohort to estimate person-time at risk^b N (%)^c Mean (SD), or percentages where indicated

Associations between body size and CRC risk according to *BRAF* status are shown in **Table 4**. For both studies, BMI modelled per 5 kg/m² increase was associated with *BRAF* wild type tumours (NLCS HR: 1.28 per 5 kg/m², 95%CI: 1.12-1.45; MCCS HR: 1.10, 95%CI: 0.99-1.21). This association remained when the data were pooled (HR: 1.16, 95%CI: 1.08-1.26). After pooling and considering BMI in sex and study specific quartiles, there was a positive dose-response association only for *BRAF* wild type tumours, and no association with respect to tumours with a *BRAF* mutation. Tests for heterogeneity between *BRAF* sub-types were not statistically significant.

With respect to waist measurements, in the MCCS there were similar dose-response associations for both *BRAF* mutated and *BRAF* wild type tumours when considering quartiles of waist circumference, however, these only reached statistical significance for the *BRAF* wild type group. This was also observed in the NLCS for clothing size, although there was less evidence of a dose response. When the data were pooled, there was a statistically significant positive dose-response association for *BRAF* wild type tumours, (highest vs. lowest quartile HR: 1.59, 95CI: 1.33-1.90, *P*-trend<0.001). A statistically significant trend was also observed for *BRAF* mutated tumours (*P*=0.03).

When height was modelled per 5 cm increment, there was an increased risk for both tumours with and without *BRAF* mutations, although the HR was greater for *BRAF* mutated tumours. For the pooled data, the HR for tumours with *BRAF* mutations was 1.23 per 5cm increment (95%CI: 1.11-1.37) compared with 1.08 (95%CI: 1.03-1.13) for tumours without *BRAF* mutations (*P*-heterogeneity=0.03). Height was also associated with both tumour sub-types when modelled according to sex and study specific quartiles, showing a positive dose-response association that was stronger for mutated than wild type tumours.

Table 4: Hazard Ratios (HR) and 95% Confidence Intervals (CI) for the association between body mass index, waist measurements, height and BRAF mutation status of colorectal tumours in the NLCS, MCCS, and a pooled analysis of the two studies.

	NLCS				MCCS				Pooled Analysis					
	BRAF mutation		BRAF wild type		BRAF mutation		BRAF wild type		BRAF mutation		BRAF wild type			
	PY	N	HR (95% CI)	N	HR (95% CI)	PY	N	HR (95% CI)	N	HR (95% CI)	N	HR (95% CI)		
Continuous														
BMI														
Per 5 kg/m ²	35,129	109	1.11 (0.85-1.45)	550	1.28(1.12-1.45)	466,772	95	1.01 (0.80-1.27)	487	1.10(0.99-1.21)	204	1.05 (0.88-1.25)	1037	1.16 (1.08-1.26)
<i>P</i> heterogeneity ^c			0.35					0.50						0.29
Height (cm)														
Per 5 cm	34,351	110	1.19 (1.03-1.37)	561	1.08(1.01-1.15)	466,804	95	1.29 (1.10-1.52)	487	1.07 (1.00-1.15)	204	1.23 (1.11-1.37)	1044	1.08 (1.03-1.13)
<i>P</i> heterogeneity			0.23					0.04						0.03
Quartiles^d BMI (kg/m²)														
1	11,133	28	1 (referent)	140	1 (referent)	114,954	25	1 (referent)	92	1 (referent)	53	1 (referent)	232	1 (referent)
2	6632	26	1.57 (0.92-2.67)	89	0.99 (0.76-1.29)	116,243	22	0.81 (0.46-1.44)	123	1.22 (0.93-1.60)	48	1.14 (0.77-1.70)	212	1.11 (0.92-1.34)
3	11,445	37	1.24 (0.76-2.02)	204	1.39 (1.12-1.72)	117,680	27	0.99 (0.58-1.71)	128	1.20 (0.91-1.58)	64	1.12 (0.78-1.61)	332	1.31 (1.11-1.55)
4	5917	18	1.22 (0.67-2.21)	117	1.40 (1.09-1.80)	117,895	21	0.87 (0.49-1.55)	144	1.35 (1.03-1.78)	39	1.04 (0.69-1.58)	261	1.38 (1.15-1.66)
<i>P</i> -trend			0.56		<0.001			0.83		0.04		0.77		<0.001

Continues on next page



Continuing from previous page

Waist measurement														
1	13,725	32	1 (referent)	176	1 (referent)	118,689	17	1 (referent)	80	1 (referent)	49	1 (referent)	256	1 (referent)
2	10,595	36	1.37 (0.85-2.21)	176	1.16 (0.94-1.43)	111,966	21	1.14 (0.61-2.13)	101	1.22 (0.91-1.64)	57	1.28 (0.87-1.87)	277	1.18 (0.99-1.40)
3	6285	24	1.46 (0.86-2.48)	105	1.15 (0.91-1.46)	121,233	28	1.38 (0.77-2.51)	136	1.45 (1.10-1.92)	52	1.42 (0.95-2.11)	241	1.28 (1.07-1.53)
4	3747	11	1.15 (0.58-2.27)	84	1.59 (1.23-2.06)	114,810	29	1.52 (0.84-2.72)	163	1.66 (1.27-2.18)	40	1.40 (0.92-2.13)	274	1.59 (1.33-1.90)
<i>P-trend</i>			0.21		0.009			0.07		<0.001		0.03		<0.001
Height (cm)														
1	9954	26	1 (referent)	146	1 (referent)	120,026	18	1 (referent)	135	1 (referent)	44	1 (referent)	281	1 (referent)
2	9049	26	1.12 (0.65-1.92)	141	1.11 (0.88-1.40)	122,910	23	1.13 (0.61-2.09)	141	1.13 (0.88-1.43)	49	1.11 (0.74-1.67)	282	1.12 (0.95-1.32)
3	9150	27	1.17 (0.68-2.01)	143	1.15 (0.92-1.45)	111,235	25	1.45 (0.79-2.66)	98	0.95 (0.72-1.25)	52	1.28 (0.86-1.92)	241	1.06 (0.89-1.26)
4	7195	30	1.73 (1.02-2.92)	127	1.29 (1.03-1.64)	112,634	29	2.08 (1.13-3.81)	113	1.31 (1.00-1.73)	59	1.87 (1.26-2.77)	240	1.31 (1.09-1.56)
<i>P-trend</i>			0.03		0.04			0.05		0.17		0.004		0.01

^a Analysis stratified by sex and ethnicity

^b Analysis stratified by sex, ethnicity, and study

^c P value for test that IIR for two tumour subtypes are equal

^d Approximate quartiles, sex and study specific

BMI: NLCs: men: <23.4 kg/m², 23.4-24.8 kg/m², 24.9-26.5 kg/m², ≥26.6 kg/m² and women: <23.8 kg/m², 23.8-24.6 kg/m², 24.7-29.9 kg/m², ≥29.9 kg/m²; MCCS: men <24.8 kg/m², 24.8-26.8 kg/m², 26.9-29.1 kg/m², ≥29.2 kg/m² and women: <23.2 kg/m², 23.2-25.9 kg/m², 25.9-29.4 kg/m², ≥29.4 kg/m²

Clothing size: NLCs: men <50, 52, 54 ≥ 56; women <42, 44, 46 ≥48

Waist circumference: MCCS: men <87.0cm, 87.0-93.0cm, 93.1-99.5 ≥99.5cm; women <71.2cm, 71.3-78.3cm, 78.3-87.0 ≥87.1cm

Height: NLCs: men <173cm, 173-176cm, 177-182, ≥183cm, and women <162 cm, 162-165 cm, 166-171 ≥172 cm; MCCS men <167.6cm, 167.6-172.5 cm, 172.6-177.5cm, ≥177.6 cm, and women <153.3 cm, 153.3-160.0 cm, 160.0-164.3, ≥164.4 cm



Table 5: Hazard Ratios (HR) and 95% Confidence Intervals (CI) for the association between body mass index, waist measurements, height and MSI status of colorectal tumours in the NLCS, MCCS, and a pooled analysis of the two studies.

	NLCS				MCCS				Pooled Analysis			
	MSI-high		MSI-stable		MSI-high		MSI-stable		MSI-high		MSI-stable	
	N	HR (95% CI)	N	HR (95% CI)	N	HR (95% CI)	N	HR (95% CI)	N	HR (95% CI)	N	HR (95% CI)
Continuous												
BMI												
Per 5 kg/m ²	35,129	81 1.27 (0.92-1.77)	544	1.22 (1.08-1.39)	466,772	90 0.98 (0.78-1.24)	495	1.10 (1.00-1.22)	171	1.07 (0.89-1.30)	1,039	1.15 (1.06-1.24)
<i>P</i> heterogeneity ^c			0.81			0.36			0.53			
Height (cm)												
Per 5 cm	34,351	83 1.20 (1.03-1.41)	548	1.10 (1.03-1.17)	466,804	90 1.32 (1.13-1.54)	495	1.07 (0.99-1.15)	171	1.26 (1.13-1.40)	1,043	1.08 (1.03-1.14)
<i>P</i> heterogeneity			0.29			0.02			0.02			
Quarti-												
BMI (kg/m²)												
1	11,133	23 1 (referent)	140	1 (referent)	114,954	19 1 (referent)	98	1 (referent)	42	1 (referent)	238	1 (referent)
2	6632	14 1.02 (0.53-2.00)	97	1.08 (0.83-1.41)	116,243	27 1.30 (0.72-1.34)	118	1.10 (0.84-1.43)	41	1.19 (0.77-1.84)	215	1.10 (0.91-1.32)
3	11,445	29 1.20 (0.69-2.07)	198	1.35 (1.09-1.67)	117,680	23 1.07 (0.57-2.00)	133	1.18 (0.90-1.54)	52	1.14 (0.76-1.72)	331	1.27 (1.08-1.51)
4	5917	15 1.21 (0.62-2.37)	109	1.33 (1.03-1.71)	117,895	21 1.04 (0.55-1.96)	146	1.31 (1.01-1.71)	36	1.11 (0.70-1.76)	255	1.33 (1.11-1.60)
<i>P</i> trend	0.46		0.04		0.90		0.04		0.67		<0.001	

Continues on next page



Continuing from previous page

Waist measurement

1	13,725	31	1 (referent)	160	1 (referent)	118,689	17	1 (referent)	80	1 (referent)	39	1 (referent)	255	1 (referent)
2	10,595	33	1.19 (0.69-2.06)	166	1.20 (0.97-1.48)	111,966	21	1.61 (0.81-3.20)	101	1.14 (0.86-1.53)	48	1.33 (1.87-2.03)	276	1.18 (0.99-1.39)
3	6285	23	0.88 (0.44-1.79)	100	1.26 (0.99-1.59)	121,233	28	1.97 (1.01-3.82)	136	1.37 (1.04-1.80)	42	1.37 (0.88-2.13)	247	1.30 (1.09-1.56)
4	3747	10	1.47 (0.72-3.00)	78	1.51 (1.16-1.97)	114,810	29	1.59 (0.80-3.15)	163	1.66 (1.28-2.16)	35	1.40 (0.87-2.24)	248	1.60 (1.33-1.91)
<i>P-trend</i>			0.74		0.04			0.43		<0.001		0.43		<0.001

Height (cm)

1	9954	15	1 (referent)	147	1 (referent)	120,026	18	1 (referent)	136	1 (referent)	33	1 (referent)	283	1 (referent)
2	9049	26	2.00 (1.06-3.76)	126	0.98 (0.77-1.24)	122,910	18	1.02 (0.53-1.95)	147	1.14 (0.90-1.45)	44	1.44 (0.93-2.25)	273	1.06 (0.90-1.26)
3	9150	22	1.73 (0.90-3.33)	137	1.08 (0.86-1.37)	111,235	27	1.88 (1.02-3.46)	97	0.91 (0.69-1.19)	49	1.79 (1.15-2.80)	234	1.01 (0.84-1.20)
4	7195	20	2.07 (1.06-4.06)	138	1.38 (1.09-1.73)	112,634	27	2.29 (1.21-4.35)	115	1.30 (0.99-1.70)	47	2.18 (1.38-3.44)	253	1.35 (1.13-1.60)
<i>P-trend</i>			0.20		0.04			0.003		0.38		0.003		0.006

^a Analysis stratified by sex and ethnicity

^b Analysis stratified by sex, ethnicity, and study

^c *P* value for test that HR for two tumour subtypes are equal

^d Approximate quartiles, sex and study specific

BMI: NLCS: men: <23.4 kg/m², 23.4-24.8 kg/m², 24.9-26.5 kg/m², ≥26.6 kg/m² and women: <23.8 kg/m², 23.8-24.6 kg/m², 24.7-29.9 kg/m², ≥29.9 kg/m²; MCCS: men: <24.8 kg/m², 24.8-26.8 kg/m², 26.9-29.1 kg/m², ≥29.2 kg/m² and women: <23.2 kg/m², 23.2-25.9 kg/m², 25.9-29.4 kg/m², ≥29.4 kg/m²

Clothing size: NLCS: men <50, 52, 54 ≥ 56; women <42, 44, 46 ≥48

Waist circumference: MCCS: men <87.0cm, 87.0-93.0cm, 93.1-99.5 ≥99.5cm; women <71.2cm, 71.3-78.3cm, 78.3-87.0 ≥87.1cm

Height: NLCS: men <173cm, 173-176cm, 177-182, ≥183cm, and women <162 cm, 162-165 cm, 166-171 ≥172 cm; MCCS men <167.6cm, 167.6-172.5 cm, 172.6-177.5cm, ≥177.6 cm, and women <155.3 cm, 155.3-160.0 cm, 160.0-164.3, ≥164.4 cm



Table 5 shows associations according to MSI status. HRs for pooled BMI and waist measurements were similar to those observed for *BRAF*; a statistically significant association was observed only for MSS tumours. Height was also associated with both tumour sub-types. For the pooled data, the HR for MSI-high tumours was 1.26 per 5cm increase (95%CI: 1.13-1.40) compared with 1.08 (95%CI: 1.03-1.14) for MSS tumours (*P*-heterogeneity=0.02).

Stratifying the analyses according to sex showed no differences between men and women.

DISCUSSION

We present prospective cohort data on associations between body size and molecular sub-sets of CRC, specifically *BRAF* mutation and MSI status. We pooled data from the NLCS and MCCS and observed that there was little heterogeneity between tumour sub-types when considering associations according to BMI and waist measurement, however, BMI and waist measurements were strong risk factors for tumours without *BRAF* mutations or MSI. Interestingly, associations with height showed more heterogeneity; our findings suggest that height is a stronger risk factor for tumours with *BRAF* mutations or MSI.

Both the NLCS and MCCS have almost complete ascertainment of colorectal cancer and little loss to follow-up. In the MCCS, body size at baseline was measured according to standard protocols. Measures of body size in the NLCS were obtained by self-report, however, there are many examples in the literature showing that this method is a valid and reliable tool for assessing body weight and height in cohort studies.³⁷⁻⁴⁰ As *BRAF* mutated tumours and MSI are uncommon, pooling the NLCS and MCCS allows for greater precision than in the individual studies.

To our knowledge only one study has investigated the association between BMI and *BRAF* status in CRC tumours. Using a case-control study, Slattery *et al.*¹² reported that obesity (BMI ≥ 30 kg/m²) was not associated with *BRAF* mutated tumours (associations with *BRAF* wild type tumours were not reported). Central adiposity has been identified as a stronger predictor of CRC than BMI,^{1,14} and based on our observations, we propose that waist measurements may be a better predictor of CRC than BMI when analyzing data according to molecular sub-type, especially when there is

limited power to detect associations. Furthermore, it may reduce heterogeneity between studies. When HRs were estimated according to waist measurements, the associations observed in the study specific analyses were more comparable with each other than those observed for BMI, and the pooled analysis suggested a dose-response association with respect to both *BRAF* wild type and MSS tumours. Future studies should consider this variable in addition to BMI.

Two case-control studies have previously considered the association between BMI and MSI.^{4, 41} While the study by Campbell *et al.*⁴ distinguished between MSS and MSI-low tumours, we combined these two groups because our MSI-low group was very small. The presence and role of MSI-low remains controversial.⁴² Mutations often observed in MSI-high tumours appear to be absent in MSI-low tumours and no large differences in clinical or molecular characteristics have been observed between MSS and MSI-low patients.⁴³ Campbell *et al.* reported that BMI was associated only with MSS and MSI-low tumours, but not MSI-high tumours. We also observed a positive association between BMI and MSS tumours, however, tests for heterogeneity between tumour sub-types did not reach statistical significance in our study. Slattery *et al.*⁴¹ reported associations according to MSI positive (equivalent to MSI-high) or MSI negative (equivalent to MSS/MSI-low) status, with similar findings to Campbell *et al.*, and additionally reported a positive association between BMI and MSI-positive tumours in women. When we performed sex specific analyses on the pooled data, we observed no differences between men and women, and no associations with MSI tumours. Neither case-control study reported associations according to waist circumference.

Our observations from the pooled data are intriguing. For all molecular sub-types considered, we observed positive associations, which is consistent with substantial evidence that adult attained height is a strong risk factor for CRC in general.¹ However, HRs were significantly higher for tumours with *BRAF* mutations and MSI than for *BRAF* wild type or MSS tumours, respectively. Height is a marker of an aggregated fetal and childhood experience, and can be considered a proxy measure for important nutritional exposures, which affect several hormonal and metabolic axes.¹ Several studies show that childhood energy restriction is associated with a decreased risk of CRC later in life.⁴⁴⁻⁴⁸ Furthermore, we recently reported

from the NLCS that exposure to severe energy restriction during childhood and adolescence was associated with a low risk of developing a CIMP positive tumour.⁴⁹ Although little research has been done in this area, our current findings contribute to the hypothesis that genetic and epigenetic events in CRC development may be influenced by early life environmental exposures. This hypothesis may explain why there was little heterogeneity between tumour sub-types when considering associations according to BMI and waist circumferences; these were measured at baseline, in adulthood. Analyzing additional cohorts to draw firmer conclusions is necessary.

In conclusion, our findings reiterate the importance of a healthy body weight in CRC prevention. Furthermore, positive associations between height, *BRAF* mutation and MSI provide more evidence for the hypothesis that early life events may influence genetic and epigenetic mechanisms.

REFERENCES

1. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AIRC, 2007
2. Noffsinger AE. Serrated polyps and colorectal cancer: new pathway to malignancy. *Annu Rev Pathol.* 2009; 4: 343-64.
3. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* 2007; 50: 113-30.
4. Campbell PT, Jacobs ET, Ulrich CM, *et al.* Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumour microsatellite instability status. *J Natl Cancer Inst.* 2010; 102: 391-400.
5. Samowitz WS, Albertsen H, Herrick J, *et al.* Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* 2005; 129: 837-45.
6. Shen L, Toyota M, Kondo Y, *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A.* 2007; 104: 18654-59.
7. Toyota M, Ahuja N, Ohe-Toyota M, *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A.* 1999; 96: 8681-86.
8. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006; 38: 787-93.
9. Barault L, Charon-Barra C, Jooste V, *et al.* Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res.* 2008; 68: 8541-46.
10. Ferguson LR, Karunasinghe N, Philpott M. Epigenetic events and protection from colon cancer in New Zealand. *Environ Mol Mutagen.* 2004; 44: 36-43.
11. Caan BJ, Coates AO, Slattery ML, *et al.* Body size and the risk of colon cancer in a large case-control study. *Int J Obes Relat Metab Disord.* 1998; 22: 178-84.

12. Slattery ML, Curtin K, Sweeney C, *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer.* 2007; 120: 656-63.
13. Slattery ML, Curtin K, Wolff RK, *et al.* Diet, physical activity, and body size associations with rectal tumour mutations and epigenetic changes. *Cancer Causes Control.* 2010; 21: 1237-45.
14. Pischon T, Lahmann PH, Boeing H, *et al.* Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst.* 2006; 98: 920-31.
15. Taioli E, Bonassi S. Pooled analysis of epidemiological studies involving biological markers. *Int J Hyg Environ Health.* 2003; 206: 109-15.
16. Smith-Warner SA, Spiegelman D, Ritz J, *et al.* Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. *Am J Epidemiol.* 2006; 163: 1053-64.
17. Goldbohm RA, van den Brandt PA, Brants HA, *et al.* Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr.* 1994; 48: 253-65.
18. Goldbohm RA, van 't Veer P, van den Brandt PA, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr* 1995; 49: 420-29.
19. van den Brandt PA, Goldbohm RA, van 't Veer P, *et al.* A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol.* 1990; 43: 285-95.
20. Van den Brandt PA, Schouten LJ, Goldbohm RA, *et al.* Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. *Int J Epidemiol.* 1990; 19: 553-8.
21. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of municipalities by cancer registries and PALGA using hospital discharge data. *Tijdschr Soc Gezondheidsz.* 1994; 72: 80-84.
22. Brink M, de Goeij AF, Weijenberg MP, *et al.* K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis.* 2003; 24: 703-10.
23. Ireland P, Jolley D, Giles GG, *et al.* Development of the Melbourne FFQ: a food frequency questionnaire for use in an Australian

- prospective study involving an ethnically diverse cohort. *Asia Pac J Clin Nutr* 1994; 3: 19-31.
24. Hughes LA, Schouten LJ, Goldbohm RA, *et al.* Self-reported clothing size as a proxy measure for body size. *Epidemiology*. 2009; 20: 673-6.
 25. Lohman TG, Roche AF, Martorell R. Anthropometric standardization reference manual. Kinetics Books, 1988.
 26. Sieben NL, Roemen GM, Oosting J, *et al.* Clonal analysis favours a monoclonal origin for serous borderline tumours with peritoneal implants. *J Pathol*. 2006; 210: 405-11.
 27. Suraweera N, Duval A, Reperant M, *et al.* Evaluation of tumour microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology*. 2002; 123: 1804-11.
 28. Young J, Barker MA, Simms LA, *et al.* Evidence for BRAF mutation and variable levels of microsatellite instability in a syndrome of familial colorectal cancer. *Clin Gastroenterol Hepatol*. 2005; 3: 254-63.
 29. Lindor NM, Burgart LJ, Leontovich O, *et al.* Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumours. *J Clin Oncol*. 2002; 20: 1043-8.
 30. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am J Epidemiol*. 1997; 145: 72-80.
 31. Barlow WE, Ichikawa L, Rosner D, *et al.* Analysis of case-cohort designs. *J Clin Epidemiol*. 1999; 52: 1165-72.
 32. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics*. 1995; 51: 524-32.
 33. English DR, Young JP, Simpson JA, *et al.* Ethnicity and risk for colorectal cancers showing somatic BRAF V600E mutation or CpG island methylator phenotype. *Cancer Epidemiol Biomarkers Prev*. 2008; 17: 1774-80.
 34. de Vogel S, Bongaerts BW, Wouters KA, *et al.* Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. *Carcinogenesis*. 2008; 29: 1765-73.

35. de Vogel S, Weijenberg MP, Herman JG, *et al.* MGMT and MLH1 promoter methylation versus APC, KRAS and BRAF gene mutations in colorectal cancer: indications for distinct pathways and sequence of events. *Ann Oncol.* 2009; 20: 1216-22.
36. de Vogel S, Wouters KA, Gottschalk RW, *et al.* Genetic variants of methyl metabolizing enzymes and epigenetic regulators: associations with promoter CpG island hypermethylation in colorectal cancer. *Cancer Epidemiol Biomarkers Prev.* 2009; 18: 3086-96.
37. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, *et al.* Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health.* 2000; 54: 143-8.
38. Nyholm M, Gullberg B, Merlo J, *et al.* The validity of obesity based on self-reported weight and height: Implications for population studies. *Obesity.* 2007; 15: 197-208.
39. Spencer EA, Appleby PN, Davey GK, *et al.* Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr.* 2002; 5: 561-5.
40. Wada K, Tamakoshi K, Tsunekawa T, *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Int J Obes.* 2005; 29: 1093-9.
41. Slattery ML, Curtin K, Anderson K, *et al.* Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumours. *J Natl Cancer Inst.* 2000; 92: 1831-6.
42. Smits KM, Cleven AH, Weijenberg MP, *et al.* Pharmacoeogenomics in colorectal cancer: a step forward in predicting prognosis and treatment response. *Pharmacogenomics.* 2008; 9: 1903-16.
43. Imai K, Yamamoto H. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. *Carcinogenesis.* 2008; 29: 673-80.
44. Frankel S, Gunnell DJ, Peters TJ, *et al.* Childhood energy intake and adult mortality from cancer: the Boyd Orr Cohort Study. *Bmj.* 1998; 316: 499-504.
45. Dirx MJ, van den Brandt PA, Goldbohm RA, *et al.* Energy restriction early in life and colon carcinoma risk: results of The Netherlands Cohort Study after 7.3 years of follow-up. *Cancer.* 2003; 97: 46-55.

46. Svensson E, Moller B, Tretli S, *et al.* Early life events and later risk of colorectal cancer: age-period-cohort modeling in the Nordic countries and Estonia. *Cancer Causes Control.* 2005; 16: 215-23.
47. Svensson E, Grotmol T, Hoff G, *et al.* Trends in colorectal cancer incidence in Norway by gender and anatomic site: an age-period-cohort analysis. *Eur J Cancer Prev.* 2002; 11: 489-95.
48. Hughes LA, van den Brandt PA, Goldbohm RA, *et al.* Energy restriction during childhood is associated with a lower risk of colorectal cancer: an analysis from the Netherlands Cohort Study. *Int J Epidemiol.* 2010; 39: 1333-44
49. Hughes LA, van den Brandt PA, de Bruine AP, *et al.* Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. *PLoS One.* 2009; 4:e7951.

9

General Discussion

The aim of this thesis was to elucidate how indicators of energy balance such as energy restriction, body size, and physical activity influence risk of CRC according to a) anatomical sub-sites of the large intestine, and b) molecular phenotypes reflective of epigenetic instability in these tumours. This chapter provides a summary of the main findings, as well as a discussion about the limitations and insights offered by molecular pathological epidemiology in the context of the study aims.

SUMMARY OF MAIN FINDINGS

First, it was shown that clothing size, specifically trouser and skirt size, predicts cancer risk independently of BMI (chapter 2). Therefore, this variable was used throughout the thesis as a proxy measure for central adiposity, as waist circumference was not available for the entire NLCS population. The main findings of the studies described in chapters 3-8 of this thesis are summarized in **Table 1** (associations according to anatomical sub-sites of the large intestine) and **Table 2** (associations according to molecular phenotypes reflective of epigenetic instability in tumours).

Table 1: Summary of the main findings of this thesis, with respect to the studies that considered anatomical sub-sites of CRC as the endpoint ^a

	Proximal Colon		Distal Colon		Rectum	
	MEN	WOMEN	MEN	WOMEN	MEN	WOMEN
Later Life Indicators of Exposure						
<i>BMI</i>						
Per 5 kg/m ²	1.19 (0.92-1.54)	1.02 (0.87-1.18)	1.42 (1.13-1.79)	0.95 (0.79-1.14)	1.02 (0.79-1.32)	1.05 (0.85-1.31)
Highest vs. lowest quintile	1.35 (0.90-1.98)	0.91 (0.65-1.28)	1.38 (0.96-1.98) ^b	1.04 (0.72-1.50)	1.01 (0.67-1.51)	1.07 (0.67-1.69)
<i>Trouser/skirt size</i>						
Per 2 unit increase	1.09 (0.99-1.20)	1.06 (0.98-1.14)	1.16 (1.04-1.29)	1.02 (0.94-1.11)	1.06 (0.96-1.17)	1.04 (0.94-1.16)
Highest vs. lowest quintile	1.32 (0.81-2.15)	1.46 (0.98-2.18)	2.56 (1.55-4.24) ^b	1.15 (0.74-1.80)	1.33 (0.77-1.29)	1.07 (0.59-1.93)
<i>Occupational physical activity</i>						
< 8kJ/day	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)
8-12	0.98 (0.76-1.26)	0.92 (0.72-1.18)	0.74 (0.57-0.95)	0.96 (0.72-1.28)	1.25 (0.96-1.79)	0.72 (0.49-1.05)
>12	0.78 (0.56-1.09)	0.79 (0.46-1.36)	0.71 (0.52-0.97) ^b	0.92 (0.50-1.69)	1.31 (0.96-1.26) ^b	0.98 (0.47-2.05)
<i>Recreational physical activity</i>						
≤30 min/day	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)
30-60	1.02 (0.75-1.38)	0.78 (0.60-1.02)	1.01 (0.74-1.36)	0.74 (0.55-1.00)	1.00 (0.72-1.39)	0.77 (0.53-1.10)
60-90	0.92 (0.65-1.30)	0.87 (0.66-1.16)	1.37 (0.99-1.89)	0.82 (0.60-1.12)	1.15 (0.80-1.64)	0.78 (0.53-1.15)
>90	0.92 (0.68-1.25)	0.71 (0.53-0.96)	1.18 (0.88-1.60)	0.69 (0.50-0.96)	1.10 (0.80-1.52)	0.59 (0.39-0.90) ^b

Continues on next page

Continuing from previous page

Early Life Indicators of Exposure

<i>BMI at age 20</i>						
Per 5 kg/m ²	0.98 (0.75-1.27)	1.03 (0.84-1.25)	1.09 (1.00-1.19)	1.05 (0.85-1.31)	1.12 (0.86-1.46)	1.06 (0.82-1.36)
Highest vs. lowest quintile	1.05 (0.72-1.56)	1.14 (0.81-1.62)	1.47 (1.03-2.08) ^b	1.09 (0.74-1.62)	1.11 (0.73-1.71)	1.29 (0.80-2.07)
<i>Height</i>						
Per 5 cm increase	1.01 (0.90-1.13)	1.04 (0.95-1.15)	0.99 (0.89-1.11)	1.11 (0.99-1.24)	0.95 (0.85-1.07)	1.14 (0.98-1.32)
Highest vs. lowest quintile	0.90 (0.59-1.37)	1.19 (0.84-1.70)	0.97 (0.63-1.48)	1.53 (1.03-2.27) ^b	0.71 (0.43-1.17)	1.49 (0.98-2.27)
<i>Place of Residence in Hunger Winter (1944-45)^c</i>						
Non-western area (ref.)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)
Western rural area	1.03 (0.76-1.41)	1.20 (0.89-1.63)	0.92 (0.67-1.26)	1.05 (0.75-1.48)	0.81 (0.57-1.15)	0.99 (0.65-1.54)
Western urban area	0.72 (0.54-0.96)	1.23 (0.97-1.56)	0.89 (0.69-1.15)	0.87 (0.66-1.15)	0.71 (0.53-0.96)	0.85 (0.61-1.20)

^a Multivariate adjusted HRs presented, as described in chapters 3, 5 and 6.

^b p-trend for categories of exposure <0.05

^c It was assumed that individuals residing in a western area of the Netherlands, especially a western urban area during the Hunger Winter would have experienced the greatest degree of energy restriction, as explained in chapters 3 & 4.

Table 2: Summary of the main findings of this thesis, with respect to the studies that considered molecular phenotypes of CRC as the endpoint.^a

	CIMP status		MSI status		BRAF mutation status	
	CIMP	Non-CIMP	MS instable	MS stable	BRAF mutated	BRAF wild type
Later Life Indicators of Exposure						
<i>BMI (NLCS only)</i>						
Per 5 kg/m ²	1.29 (1.01-1.66)	1.19 (0.98-1.44)	1.27 (0.92-1.77)	1.22 (1.08-1.39)	1.11 (0.85-1.45)	1.28 (1.12-1.45)
Highest vs. lowest quartile	1.85 (1.07-3.21)	1.22 (0.86-1.75)	1.21 (0.62-2.37)	1.33 (1.03-1.71) ^b	1.22 (0.67-2.21)	1.40 (1.09-1.80) ^b
<i>BMI (pooled with MCCS)</i>						
Per 5 kg/m ²	- ^c	-	1.07 (0.89-1.30)	1.15 (1.06-1.24)	1.05 (0.88-1.25)	1.16 (1.08-1.26)
Highest vs. lowest quartile	-	-	1.11 (0.70-1.76)	1.33 (1.11-1.60) ^b	1.04 (0.69-1.58)	1.38 (1.15-1.66) ^b
<i>Trouser-skirt size (NLCS only)</i>						
Highest vs. lowest quartile	2.01 (1.10-3.67) ^b	1.46 (1.04-2.04) ^b	1.47 (0.72-3.00)	1.51 (1.16-1.97) ^b	1.15 (0.58-2.27)	1.59 (1.23-2.06) ^b
<i>Trouser-skirt size (pooled with MCCS)</i>						
Highest vs. lowest quartile	-	-	1.40 (0.87-2.24)	1.60 (1.33-1.91) ^b	1.40 (0.92-2.13) ^b	1.59 (1.33-1.90) ^b
<i>Physical activity^f</i>						
Low (ref.)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)
Intermediate	0.50 (0.30-0.82)	0.80 (1.61-1.04)	0.64 (0.32-1.27)	0.74 (0.58-0.95)	0.69 (0.40-1.17)	0.76 (0.59-0.98)
High	0.82 (0.49-1.38)	0.67 (0.47-0.96) ^b	0.89 (0.43-1.87)	0.68 (0.49-0.93) ^b	0.77 (0.40-1.48)	0.70 (0.51-0.96) ^b

Continues on next page

Continuing from previous page

Early Life Indicators of Exposure

<i>BMI at age 20</i>						
Per 5 kg/m ²	1.28 (0.96-1.72)	1.10 (0.90-1.34)	1.02 (0.62-1.67)	1.12 (0.93-1.34)	1.08 (0.70-1.67)	1.11 (0.93-1.32)
Highest vs. lowest quartile	1.57 (0.88-2.82) ^b	1.04 (0.75-1.45)	0.96 (0.46-2.01)	1.13 (0.84-1.53)	1.19 (0.63-2.23)	1.13 (0.84-1.52)
<i>Height (NLCS only)</i>						
Per 5 cm increase	1.14 (1.00-1.30)	1.11 (1.01-1.21)	1.20 (1.03-1.41)	1.10 (1.03-1.17)	1.19 (1.03-1.37)	1.08 (1.01-1.15)
Highest vs. lowest quartile	1.38 (0.79-2.38)	1.46 (1.04-2.04) ^b	2.07 (1.06-4.06)	1.38 (1.09-1.73) ^b	1.73 (1.02-2.92) ^b	1.29 (1.03-1.64) ^b
<i>Height (pooled with MICS)</i>						
Per 5 cm increase	-	-	1.26 (1.13-1.40) ^d	1.08 (1.03-1.14)	1.23 (1.11-1.37) ^d	1.08 (1.03-1.13)
Highest vs. lowest quartile	-	-	2.18 (1.38-3.44) ^b	1.35 (1.13-1.60) ^b	1.87 (1.26-2.77) ^b	1.31 (1.09-1.56) ^b
<i>Place of Residence in Hunger Winter (1944-45)^e</i>						
Non-western area (ref)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)	1 (referent)
Western area	0.65 (0.45-0.92)	0.91 (0.73-1.23)	0.85 (0.53-1.37)	0.84 (0.69-1.03)	1.00 (0.66-1.52)	0.80 (0.66-0.97)

^a For purposes of comparison, HRs are adjusted for age and sex. Multivariate adjusted HRs were presented for molecular endpoints in chapter 7, however, even after such adjustments, HRs remained stable.

^b *P-trend* for categories of exposure <0.05

^c pooling was not possible

^d *P-heterogeneity* <0.05 (test that HR for two tumor subtypes, i.e. CIMP and non-CIMP are equal)

^e It was assumed that individuals residing in a western area of the Netherlands, especially a western urban area during the Hunger Winter would have experienced the greatest degree of energy restriction, as explained in chapters 3 & 4.

^f Based on baseline non-occupational physical activity for women (low= <30 minutes/day, intermediate= 30-90 minutes/day, high= >90 minutes/day) and occupational physical activity at the longest held job for men (low= <8 kJ/minute, intermediate = 8-12 kJ/minute, and high= >12 kJ/minute).

We observed that indicators of energy balance appear to be associated with risk of CRC differently according to sex and sub-localization of the large intestine. With the exception of physical activity, stronger associations were generally observed for men. Energy restriction during childhood and adolescence was associated with an inverse risk of CRC in men, especially for tumours of the proximal colon and rectum. Associations for women were less clear.

Body fat, as reflected by BMI, BMI at age 20 and trouser size, was associated with an increased risk of CRC in men, especially tumours of the distal colon. Such associations were weaker in women. For both men and women, trouser/skirt size appears to predict risk independently of BMI. With respect to height, no clear associations were observed for men. In contrast, height was associated with increased risk of CRC in women, especially tumours of the distal colon and rectum.

Sex specific differences were also observed with respect to physical activity; occupational physical activity appeared to be more protective of CRC in men, especially tumours of the distal colon, and non-occupational physical activity appeared to be more protective in women, especially for tumours of the distal colon and rectum.

With respect to molecular endpoints, it appears that timing of exposure may be important for influencing CRC risk. We observed that individuals exposed to severe energy restriction during childhood or adolescence had a decreased risk of developing a tumour characterized by CIMP compared to those not exposed. Although other indicators of energy balance reflective of earlier life exposures, such as BMI at age 20 and height, increased the risk of all tumour subtypes, they did so differentially; those with a higher BMI at age 20 had a higher risk of developing a CIMP tumour, and taller individuals were at higher risk of developing a tumour characterized by *BRAF* mutations or MSI. In contrast, aspects of energy balance reflective of exposures during adulthood, such as baseline BMI, trouser/skirt size, and physical activity, did not significantly differentially influence risk of molecular phenotypes.

INTERPRETATION OF OUR RESULTS

CIMP, MSI and *BRAF* mutations characterise the serrated neoplasia pathway to CRC, and appear to be highly correlated.^{7, 12} One may then question why we observed increased risk of some molecular endpoints with regard to a given exposure, but not all. In **Figure 1**, it can be seen that overlap of CIMP, MSI, and *BRAF* mutations in the NLCS population is incomplete.

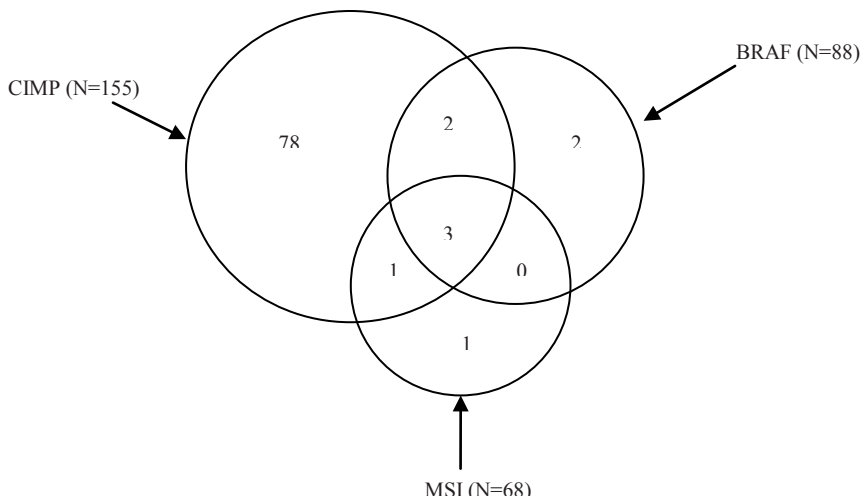


Figure 1:

Overlap between CIMP, BRAF mutations and MSI in CRC tumours of the NLCS. Numbers are based on cases with complete analyses of all three molecular aberrations. The sizes of the different areas in this figure do not exactly reflect the numbers of the applicable subsets.

CRC is a complex disease; heterogeneity and associations with a given risk factor are far from black and white. Although CIMP, MSI, and *BRAF* mutations strongly correlate, the question remains whether CRC with these characteristics develops through one distinct molecular pathway. There is also a possibility that the observed differences and incomplete overlap are attributable to limitations inherent to molecular pathological epidemiology.

Issues of Methodology

Measurement error and misclassification

In traditional epidemiology, measurement error and misclassification of exposure variables can be problematic. For this thesis in particular, it is plausible that a metabolic profile reflecting a combination of risk factors has a greater influence on CRC risk than do the individual effects of body size, physical activity, and energy intake/restriction.^{62, 63} It would have been interesting to consider a model of ‘energy balance’, however, as discussed in chapter 5, measurement error of exposure variables limited this possibility. Technological advances with respect to recording/tracking energy intake and physical activity in large populations may enable future studies to assess the simultaneous influence of these variables on CRC risk. In molecular pathological epidemiology, there is the additional problem of possible measurement error and misclassification of molecular endpoints. This combination of error (exposure plus molecular endpoints) is a serious concern. Nevertheless, errors can be minimized when molecular and histochemical assays are thoroughly validated and monitored for precision and accuracy,¹³ and if cases with poor quality tumour material are interpreted with caution, or excluded from the analysis.

Sample Size

One of the most important limitations faced in this field of research is sample size. The NLCS is one of the largest studies with information collected on molecular endpoints of CRC; there are 120,852 individuals in the cohort. Despite this size, any molecular pathological epidemiology study conducted within the larger cohort will undergo multiple exclusions based on availability of tumour material and valid assay results.¹³ Therefore, the sample size for a study with molecular endpoints will always be smaller than the parent study. To analyse molecular data for associations, a subset analysis for different outcomes is performed (i.e. CIMP vs. non-CIMP; MSI vs. MSS; *BRAF* mutated vs. *BRAF* wild type tumours). The sample size for a subset, especially the rarer event (e.g. CIMP; MSI; *BRAF* mutations) may be too small to provide adequate statistical power, or limit the number of possible sub-groups to be distinguished. This may have especially influenced our findings in chapters 4 and 7. One solution for such a problem is to pool data from independent studies.

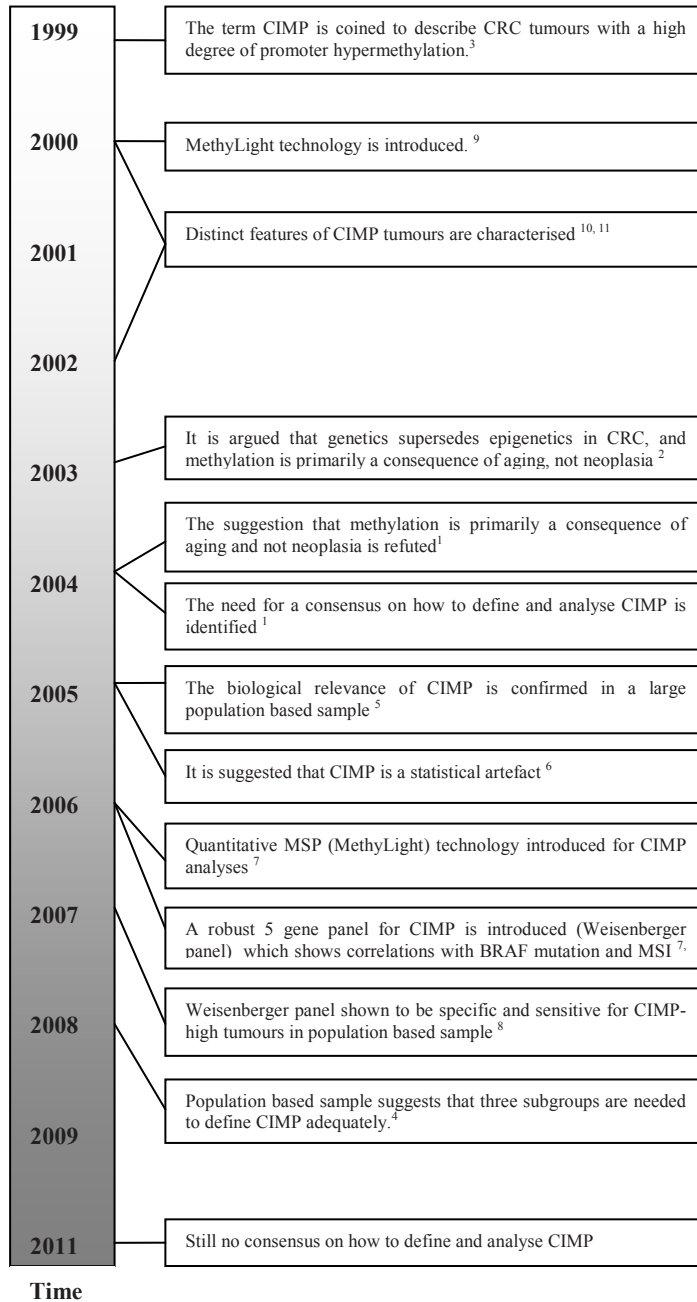
In chapter 8 of this thesis, we pooled data from the NLCS and the Melbourne Collaborative Cohort Study (MCCS), a large, prospective cohort study in Australia. The key motivation for doing so was to overcome small sample sizes regarding molecular data in each cohort, and to improve the precision of risk estimates for the association between body size and risk of tumours displaying MSI or *BRAF* mutations. The prevalence of *BRAF* mutation and MSI was similar between the two studies (NLCS: 16% *BRAF* mutation, 13% MSI; MCCS: 16% *BRAF* mutation, 15% MSI). To successfully pool data with minimum error, it is important that study designs, data collection and analytical methods are sufficiently comparable.

In **Table 2**, one can see that pooling was possible with respect to *BRAF* mutations and MSI. This was because both the NLCS and MCCS used similar definitions and laboratory analyses to define these phenotypes. However, pooling was not possible for CIMP. This highlights unique challenges of pooling molecular data, as well as the fact that the use of CIMP as an indicator of epigenetic instability in CRC research has not been without controversy.

The CpG island methylator phenotype (CIMP)

Progress and problems

It has been more than a decade since the findings that led to the definition of CIMP in CRC.³ Although it is now generally well accepted that these cancers represent a clinically and etiologically distinct group of tumours, one can see in **Figure 2** that the path to acceptance was not without challenges. In fact, a number of unresolved issues surrounding the phenotype still remain.

**Figure 2:**

A timeline of the major advances and setbacks surrounding CIMP over the past decade

In recent years, attention has primarily focused on potential biological and environmental causes of CIMP,^{1, 14} as well as the feasibility of epigenetic interventions in treatment and prevention.¹⁵ However, as molecular pathological epidemiology continues to offer possibilities for answers regarding environmental risk factors and disease trends, more practical issues associated with CIMP also require attention.

The most significant bottleneck facing this field of research is that there is no gold standard or precise definition of the phenotype. The need for a standard definition was clearly stated by Issa in 2004,¹ but to date, there is still no consensus. As a consequence, it is difficult to establish a true prevalence of CIMP because studies continue to utilize different combinations of gene panels, marker cut-offs, and PCR techniques in their respective research programs. A number of large, unselected series of cases are now available for analysis with respect to molecular endpoints. To our knowledge, the NLCS and MCCS are the first to pool such data, and comparing more populations and pooling data remains difficult. Using CIMP as a prime example, it can be seen in **Table 3** that there is substantial heterogeneity between the definition and analysis of CIMP in each study.

Table 3: Comparison of population based samples that have analyzed CRC tumours for CIMP

Study	Study Design	Gene panel	Cut-off points used to define phenotype	Laboratory analysis used to detect tumour methylation	Prevalence of CIMP (CIMP cases/Total cases)
Netherlands Cohort Study (NLCS) ¹⁶	Prospective Cohort (N=120,852)	Weisenberger panel ^a	CIMP: 3 – 5/5 genes methylated non-CIMP: 0- 2 genes/5 methylated	Methylation specific PCR (MSP)	164/ 607 (27%)
Melbourne Collaborative Cohort Study (MCCS) ¹⁷	Prospective Cohort (N=41,328)	Weisenberger panel	CIMP: 3 – 5/5 genes methylated non-CIMP: 0 - 2 genes/5 methylated	MethylLight (quantitative real time PCR)	73/506 (14%)
Nurses Health Study & Health Professionals Follow-up Study ^{8,18}	Prospective Cohort(s) (N= 121,700 & (N=51,500)	Weisenberger panel + Weisenberger panel	CIMP: 3 – 5/5 genes methylated non-CIMP: 0- 2 genes/5 methylated ¹⁸	MethylLight (quantitative real time PCR)	133/904 (15%)
Northern Sweden Health and Disease Study (NSHDS) & The Colorectal Cancer in Umeå Study (CRUMS) ¹⁹	Prospective Cohort(s) (N= 166,000 & (N=414)	Weisenberger panel + <i>CDKN2A (p16), CRABP1, hMLH1</i>	CIMP high: 5 – 8/8 genes methylated CIMP low: 1-4/ 8 genes methylated Non-CIMP: 0/8 genes methylated ⁸	MethylLight (quantitative real time PCR)	CIMP low: 363/920 (38%) CIMP high: 136/920 (15%)
		Weisenberger panel + <i>CDKN2A (p16), CRABP1, hMLH1</i>	CIMP: 6 – 8/8 genes methylated CIMP-low: 1 – 5/8 genes methylated non-CIMP: 0/8 genes methylated	MethylLight (quantitative real time PCR)	CIMP low: 215/590 (36%) CIMP high: 74/590 (13%)

Continues on next page

Continuing from previous page

Kaiser Permanente Medical Care Program of Northern California/Utah and Minnesota population ²⁰	Case-Control (N= 3564)	Classic panel ^b	CIMP: 2 – 5/5 genes methylated non-CIMP: 0 – 1/5 genes methylated	Methylation specific PCR (MSP)	276/1100 (25%)
	Population based case series (N=864)	Classic panel	CIMP: 2 – 5/5 genes methylated non-CIMP: 0 or 1/5 genes methylated	Methylation specific PCR (MSP)	256/864 (29%)
Barnault <i>et al.</i> , ⁴	Population based case series (N=582)	Classic panel	CIMP low: 1 – 3/5 genes methylated	Methylation specific PCR (MSP)	CIMP low: 199/582 (34.2%)
			CIMP high: 4 – 5/5 genes methylated non-CIMP: 0/5 genes methylated		CIMP high: 97/ 582 (16.7%)

^a genes in panel: *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1*^b genes in panel: *MINT1*, *MINT2*, *MINT31*, *CDKN2A (p16)*, *hMLH1*

Gene panels and threshold values

Which gene panel and cut-off points are ‘best’ for identifying CIMP in CRC tumours? **Table 3** shows that studies have adopted one of two panels. The so-called ‘classic panel’, which includes *MINT1*, *MINT2*, *MINT31*, *CDKN2A (p16)* and *MLH1*,¹ or the robust five gene panel more recently introduced by Weisenberger *et al.*,⁷ which includes the genes *CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1*. Previous research shows that *MINT1*, *MINT2*, and *MINT31* are not specific for *BRAF*-mutated CIMP tumours.⁷ However, in studies validating the Weisenberger markers^{8,21} it is emphasized that such findings do not indicate that these MINT markers or other CpG islands are inappropriate for assessment of CIMP in CRC, because it remains a possibility that a difference in primer designs and PCR conditions may substantially change sensitivity and specificity of a particular marker for the detection of CIMP.⁸ Ogino *et al.* propose that a panel of (at least) four markers including *RUNX3*, *CACNA1G*, *IGF2*, and *MLH1* should constitute a sensitive and specific CIMP panel for the purpose of research and clinical use,⁸ but it is unknown how many studies have adapted this advice.

Furthermore, there is debate whether CIMP should be distinguished as two categories: ‘CIMP’ and ‘non-CIMP’, or as three categories: ‘CIMP-high’, ‘CIMP low’ and ‘non-CIMP’. Two prospective cohort studies report similar prevalence using an 8 gene panel to distinguish these three categories (the five genes in the Weisenberger panel plus *CDKN2A (p16)*, *CRABP*, and *MLH1*).^{8,19} However, it was also reported that differences between CIMP-low (1/8 to 5/8 methylated promoters) and non-CIMP (0/8 methylated promoters) were not large.⁸ Additional studies are necessary to assess whether CIMP-low represents a distinct phenotype in CRC, and, the debate surrounding CIMP will likely continue until a biological cause for CIMP has been determined.

Analytical methods

There is also debate surrounding which method is best to detect promoter hypermethylation in tumours. Although the NLCS and MCCS both utilized the Weisenberger panel to define CIMP, there is an obvious difference in CIMP prevalence between the two studies. Could this be because of the difference in methods? It has been suggested that a quantitative analysis, for example MethyLight (realtime PCR), is needed for

studying methylation.^{1, 18} However, qualitative MSP has been shown to be effective and specific and does not require specific equipment.^{4, 22} Even though MethyLight is quantitative, there is a chance that data can differ from study to study, depending on what value is set as the ‘percentage of methylated reference’ (PMR) and the percentage of tumour cells present in the sample. The PMR is the value at which a given loci is declared methylated; some studies report using a PMR of >10 to declare methylation,^{7, 17} whereas another study has reported using a lower PMR of > 4.²³ However, it is also important to note that not all studies specifically report this value. A higher PMR results in a stricter definition of methylation, and consequently, a stricter definition of CIMP.

We suggest that the different prevalence observed between the NLCS and MCCS is not because of the methods per se, but rather a difference in choice of primers and/or location of methylation in the markers. The MSP analyses that have been conducted in the NLCS have a high detection signal, and subsequently, a higher prevalence of CIMP has been observed. Also, the primer/probe location of analysed CpG nucleotides may have differed between studies, and although most studies analyse methylation “around the transcription start site”, no standard protocol for where to look for methylation exists. Promoter CpG islands of genes have often been reported as ‘unmethylated’ or ‘hypermethylated’, based on data of only a small number of CpG dinucleotides independent of location or the assays which have been used. It is now known that the location of core regions and the density of methylation required for gene silencing can vary per gene, therefore, a broader view than just the classical dogma of promoter CpG island methylation and gene silencing is needed to interpret data on DNA methylation, gene expression and clinico-pathological associations.²⁴ In the future, this may be accomplished by novel technologies that enable (semi)epigenome wide analyses of methylation profiles for specific genes.

That we were unable to pool CIMP data between NLCS and MCCS because of such differences reiterates that a consensus regarding the concept of CIMP is critical. To reach such a consensus, a study in which methods and panels are compared within the same population is required. And although this discussion has been limited to CIMP, it is important to note that such issues likely exist with respect to other molecular endpoints as well, and these may influence observations between studies.

Potential role of exposure variables in the aetiology of CRC

Molecular pathological epidemiology offers an opportunity to gain insight into potential mechanisms of disease aetiology. Two remaining methodological limitations are involved in this discussion, and must first be addressed.

By definition, molecular pathological epidemiology involves subset analyses on tumour sub-types, which increases the chance of false positive findings.^{13, 25} This phenomenon is exacerbated when one crosses a number of exposures to a number of molecular endpoints. In the literature, questions have been raised whether molecular pathological epidemiology studies should be hypothesis driven, or exploratory like genome wide association studies (GWAS).¹³ Although some multiple testing was unavoidable during the course of this thesis, we had clear hypotheses to drive our research questions, which were based on the biological plausibility of indicators of energy balance being associated with CRC risk (outlined in **Table 4**), as well as previous epidemiological observations. Such an approach increases the likelihood that reported significant findings did not arise by chance alone. However, regardless if findings are the result of an exploratory analysis or are hypothesis driven, they should be interpreted with caution, and validated in other studies.²⁵

Closely related to the issue of multiple testing, is the concept of causation. Although molecular pathological epidemiology provides a means to pinpoint carcinogenic mechanisms, such as the role of energy balance in CRC development, it is important to realize that experimental data is needed before a causal relationship can be confirmed.¹³

Table 4: Summary of hypothesised biological mechanism(s) of how indicators of energy balance influence CRC risk

Risk factor	Chapter	Physiological relevance	Hypothesised mechanism(s)
Energy restriction during childhood or adolescence	3, 4	- indicator of a severe, transient environmental exposure earlier in life	- inhibits cell proliferation - accelerates apoptosis - reduces IGF-1 - enhances DNA repair - lower sex steroid hormones ²⁶⁻³²
Body Mass Index	5,7,8	- indicator of peripheral body fat	- not clear, but as BMI is an indicator of body fat, mechanisms are likely similar to what is known for central adiposity
Body Mass Index at age 20	5, 7	-indicator of peripheral body fat in the period after puberty -can be used to calculate adult BMI change (adult weight gain is more dependent on fat accumulation rather than lean tissue) ³³	- not clear, but as BMI is an indicator of body fat, mechanisms are likely similar to what is known for central adiposity
Trouser/skirt size	5,7,8	- proxy for central adiposity (visceral fat), which is more metabolically active than peripheral fat - visceral fat actively secretes IGF-1 and cytokines such as TNF- α , interleukin-6, and adiponectin ^{34, 35}	- Alterations in insulin/IGF axes leading to insulin resistance and hyperinsulinemia: - Insulin is an important growth factor for colonic mucosal cells and colonic cancer cells <i>in vitro</i> . ^{34, 36, 37} - IGF-1 inhibits apoptosis and promotes cell-cycle progression ^{34, 38, 39} - Increases in cytokine levels leading to tissue inflammation ³⁴
Adult-attained height	5,7,8	- related to rate of growth, cell divisions, and age of sexual maturity, which in turn influence hormonal environment ⁴⁰ - proxy for important nutritional exposures in-utero and childhood ^{33, 40}	- taller individuals have a longer colorectal tract - taller individuals undergo more cell divisions stimulated by IGF-1 and growth hormones, creating more potential for error during DNA replication. ³³ - Epigenetic modifications can be inherited, or acquired in-utero ^{41, 42}
Physical activity	6	-indicator of energy expenditure	- helps maintain a healthy body weight - improves immune function - reduces free radicals - increases gut motility - improves insulin sensitivity, reduces the level of insulin and insulin-like growth factor 1, and increases the level of insulin-like growth factor binding protein 3 ⁴³⁻⁴⁶

The Serrated Neoplasia Pathway to CRC: Is timing of exposure important?

- *Later life indicators of energy balance*

Throughout this thesis we presented HRs for BMI that were both adjusted and unadjusted for trouser/skirt size (and vice-versa). It may be argued that mutual adjustment of these variables is in fact an over adjustment, but our results support that central adiposity is a consistent, independent risk factor for CRC.^{33, 47} This was also true with respect to associations with specific molecular endpoints. Previous studies considering body fat and risk of CIMP, *BRAF* mutations, and/or MSI, have only considered BMI as an independent risk variable.^{20, 48, 49} Therefore, this finding suggests that the metabolic processes involved with abdominal body fat may directly or indirectly influence CRC development (**Table 4**).

It has been hypothesized that DNA methylation is a consequence of inflammation.^{50, 51} Central adiposity is also associated with chronic inflammation.⁵² We therefore hypothesized that individuals with a high trouser/skirt size may be at higher risk for a tumour characterized by CIMP. Because we observed that a higher trouser/skirt size was associated with a higher risk of all tumours, our results do not support a greater association with CIMP tumours, or that body fat is associated with a higher risk of *BRAF* mutated or MSI tumours more than *BRAF* wild type or MSS tumours. Similarly, physical activity was inversely associated with CRC risk (**Table 4**), in accordance with literature,³³ but not differentially with respect to the molecular endpoints considered. Could adult body fat and adult physical activity instead be influencing risk via the chromosomal instability (CIN) pathway?

CIN is characterised by losses and gains in chromosomes (aneuploidy), and mutations in the *KRAS* and *APC* genes. It is also associated with male sex, and tumours of the distal colon.^{53,54} In this thesis, we observed that later life indicators of energy balance (BMI, trouser/skirt size, and physical activity) were especially associated with tumours of the distal colon and especially in men. To our knowledge, no large, population based studies have published on CIN, however, there is biological plausibility that body fat is associated with chromosome stability. Telomeres are the physical ends of chromosomes. Recently, an inverse association between telomere length and body weight has been reported.^{55, 56} This is interesting because it has also

been observed that shorter telomere length is associated with CIN and MS-stable colorectal tumours, but not to MSI colorectal tumours.⁵⁷ Taken together, this may suggest that a higher BMI or waist circumference may be associated with CIN in CRC.

- *Earlier life indicators of energy balance*

Although adult exposures did not appear to be differentially associated with epigenetic instability in CRC, the opposite was true for variables reflective of exposures earlier in life, namely, adult-attained height, severe energy restriction, and BMI at age 20.

We observed that taller individuals were at higher risk of developing a tumour characterized by *BRAF* mutations or MSI, but not CIMP. We also reported that exposure to severe energy restriction was associated with a decreased risk of CIMP, but not MSI or *BRAF* mutations. GWAS are beginning to unravel the complex genetics involved in determining the heritability of human height.^{58, 59} Influences on genetic processes very early in development may in turn influence (epi)genetic mechanisms later in life. Alternatively, (epi)genetic instability associated with adult-attained height may be a reflection of events transmitted from previous generations. It has been shown that alterations in DNA methylation can be transmitted via eggs or sperm, and measured in blood,^{41, 42} however, it is important to note that we measured methylation at the moment of diagnosis, and in tumour tissue, so we can not draw conclusions about heritability. In contrast to height, environmental exposures during periods of growth such as puberty, a time period with an influx of growth and sex steroid hormones, may be more important for exerting risk on (epi)genetic mechanisms that are not inherited.

Generally, age 20 is considered past the age of puberty and growth spurt. However, we can not rule out that this is reflective of BMI during or before puberty; other studies are required to tease this out. We did observe that BMI at age 20 was associated with a higher risk of CIMP tumour than a non-CIMP tumour later in life. This builds on the hypothesis that methylation is an early event in CRC progression,¹² and as was observed with exposure to energy restriction during adolescence, suggests that exposures long before a given CRC event may already have implications for disease risk later in life.

Whether body size, physical activity or energy restriction can directly or indirectly cause a molecular change, or if they instead create an environment which can provide a selective advantage for clonal expansion of a cell with a specific molecular change,¹³ can not be determined from the present research. Therefore, the hypotheses we lay forth in this discussion are speculative.

IMPLICATIONS FOR PUBLIC HEALTH & FUTURE RESEARCH

The findings reported in this thesis require confirmation in other studies, but with respect to indicators of energy balance and general CRC risk, our observations support the growing body of evidence that a healthy body weight and adequate levels of long-term physical activity protects against CRC. Educational campaigns by organizations such as the World Cancer Research Fund (WCRF), World Health Organization (WHO), and local governments to bring attention to the dangers of overweight and the importance of being active are warranted in an effort to curb rising CRC rates.

Our findings with respect to energy restriction early in life are more difficult to translate. We do not advocate severe energy restriction as was experienced during the Dutch Hunger Winter of 1944-45, but our observations do underscore the risk associated with overeating. We propose that much can be learned from this unique and sobering ‘experiment of history.’ In this thesis, we have shown that early life exposures may be associated with decreased risk of CRC later in life, perhaps through associations with (epi)genetic mechanisms. This raises intriguing questions about possible future public health trends, which in turn, may generate hypotheses for future research.

As indicated in chapter 1 of this thesis, the WHO has identified childhood obesity as a major health concern.⁶⁰ For instance, it is estimated that 17% of American children and adolescents aged 2-19 are now overweight, and 32% are at risk of becoming overweight.⁶¹ This trend extends to Europe, and the International Obesity Task Force (IOTF) has identified that childhood and adolescent obesity is also increasing in developing nations, even in those populations more historically associated with malnutrition. The long-term consequences of being obese during this critical period of growth and development remain largely undetermined. How

will the future prevalence of CRC be influenced by the increasing prevalence of childhood obesity and positive energy balance? Furthermore, what are the implications, both with respect to CRC risk and epigenetic programming, for children and adolescents in developing countries, where nutritional habits and exposures are in transition? Although ethical issues arise when investigating energy restriction in human populations, animal studies may be a means to investigate energy restriction and cancer risk.

We acknowledge that CRCs displaying CIMP, *BRAF* mutations and MSI only account for ~15% of tumours. The other ~85% is thought to arise via the chromosomal instability pathway. As described earlier in this chapter, it may be plausible to hypothesize that adult body fat and adult physical activity influence CRC risk via the chromosomal instability pathway. In order to fully appreciate and elucidate how indicators of energy balance influence the aetiology of CRC, future molecular pathological epidemiological studies should make an effort to determine chromosomal instability status in their populations, and study associations between indicators of energy balance and this molecular endpoint.

REFERENCES

1. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer*. 2004; 4: 988-93.
2. Yamashita K, Dai T, Dai Y, *et al*. Genetics supersedes epigenetics in colon cancer phenotype. *Cancer cell*. 2003; 4: 121-31.
3. Toyota M, Ahuja N, Ohe-Toyota M, *et al*. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A*. 1999; 96: 8681-6.
4. Barault L, Charon-Barra C, Jooste V, *et al*. Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res*. 2008; 68: 8541-6.
5. Samowitz WS, Albertsen H, Herrick J, *et al*. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology*. 2005; 129: 837-45.
6. Anacleto C, Leopoldino AM, Rossi B, *et al*. Colorectal cancer "methylator phenotype": fact or artifact? *Neoplasia*. 2005; 7: 331-5.
7. Weisenberger DJ, Siegmund KD, Campan M, *et al*. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*. 2006; 38: 787-93.
8. Ogino S, Kawasaki T, Kirkner GJ, *et al*. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn*. 2007; 9: 305-14.
9. Eads CA, Danenberg KD, Kawakami K, *et al*. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic acids research*. 2000; 28: E32.
10. Toyota M, Ohe-Toyota M, Ahuja N, *et al*. Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. *Proc Natl Acad Sci U S A*. 2000; 97: 710-5.
11. van Rijnsoever M, Grieu F, Elsaleh H, *et al*. Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. *Gut*. 2002; 51: 797-802.
12. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*. 2007; 50: 113-30.

13. Ogino S, Chan AT, Fuchs CS, *et al.* Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut*. 2010;doi:10.1136/gut.2010.217182.
14. Teodoridis JM, Hardie C, Brown R. CpG island methylator phenotype (CIMP) in cancer: causes and implications. *Cancer Lett*. 2008; 268: 177-86.
15. Smits KM, Cleven AH, Weijnenberg MP, *et al.* Pharmacoepigenomics in colorectal cancer: a step forward in predicting prognosis and treatment response. *Pharmacogenomics*. 2008 ;9: 1903-16.
16. Hughes LA, van den Brandt PA, de Bruine AP, *et al.* Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms. *PloS one*. 2009; 4:e7951.
17. English DR, Young JP, Simpson JA, *et al.* Ethnicity and risk for colorectal cancers showing somatic BRAF V600E mutation or CpG island methylator phenotype. *Cancer Epidemiol Biomarkers Prev*. 2008; 17:1774-80.
18. Ogino S, Cantor M, Kawasaki T, *et al.* CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut*. 2006; 55: 1000-6.
19. Dahlin AM, Palmqvist R, Henriksson ML, *et al.* The role of the CpG island methylator phenotype in colorectal cancer prognosis depends on microsatellite instability screening status. *Clin Cancer Res*. 2010; 16: 1845-55.
20. Slattery ML, Curtin K, Sweeney C, *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer*. 2007; 120: 656-63.
21. Nosho K, Irahara N, Shima K, *et al.* Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PloS one*. 2008; 3: e3698.
22. Goel A, Nagasaka T, Arnold CN, *et al.* The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. *Gastroenterology*. 2007; 132: 127-38.
23. Ogino S, Kawasaki T, Brahmandam M, *et al.* Precision and performance characteristics of bisulfite conversion and real-time PCR

- (MethyLight) for quantitative DNA methylation analysis. *J Mol Diagn.* 2006; 8: 209-17.
24. van Vlodrop IJH, Niessen HEC, Derks S, *et al.* Analysis of promoter CpG island methylation in cancer: location, location, location! *Submitted.* 2010.
 25. Ogino S, Stampfer M. Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. *J Natl Cancer Inst.* 2010; 102: 365-7.
 26. Albanes D. Total calories, body weight, and tumor incidence in mice. *Cancer Res.* 1987; 47: 1987-92.
 27. Albanes D, Salbe AD, Levander OA, *et al.* The effect of early caloric restriction on colonic cellular growth in rats. *Nutr Cancer.* 1990; 13: 73-80.
 28. Kritchevsky D. Colorectal cancer: the role of dietary fat and caloric restriction. *Mutat Res.* 1993; 290: 63-70.
 29. Rous P. The Influence of Diet on Transplanted and Spontaneous Mouse Tumors. *J Exp Med.* 1914; 20: 433-51.
 30. Silverman J, Powers J, Stromberg P, *et al.* Effects on C3H mouse mammary cancer of changing from a high fat to a low fat diet before, at, or after puberty. *Cancer Res.* 1989; 49: 3857-60.
 31. Tannenbaum A. The dependence of tumor formation on the composition of the calorie-restricted diet as well as on the degree of restriction. 1945. *Nutrition.* 1996; 12: 653-4.
 32. Weindruch R, Albanes D, Kritchevsky D. The role of calories and caloric restriction in carcinogenesis. *Hematol Oncol Clin North Am.* 1991; 5: 79-89.
 33. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. Washington DC: AIRC, 2007.
 34. Kang HW, Kim D, Kim HJ, *et al.* Visceral obesity and insulin resistance as risk factors for colorectal adenoma: a cross-sectional, case-control study. *Am J Gastroenterol.* 2010; 105: 178-87.
 35. Kissebah AH, Krakower GR. Regional adiposity and morbidity. *Physiol Re.* 1994; 74: 761-811.
 36. Bjork J, Nilsson J, Hultcrantz R, *et al.* Growth-regulatory effects of sensory neuropeptides, epidermal growth factor, insulin, and

- somatostatin on the non-transformed intestinal epithelial cell line IEC-6 and the colon cancer cell line HT 29. *Scand J Gastroenterol.* 1993; 28: 879-84.
37. Watkins LF, Lewis LR, Levine AE. Characterization of the synergistic effect of insulin and transferrin and the regulation of their receptors on a human colon carcinoma cell line. *Int J Cancer.* 1990; 45: 372-5.
 38. Aaronson SA. Growth factors and cancer. *Science.* 1991; 254: 1146-53.
 39. Kaaks R, Lukanova A. Energy balance and cancer: the role of insulin and insulin-like growth factor-I. *Proc Nutr Soc.* 2001; 60: 91-106.
 40. Okasha M, Gunnell D, Holly J, *et al.* Childhood growth and adult cancer. *Best Pract Res Clin Endocrinol Metab.* 2002; 16: 225-41.
 41. Heijmans BT, Tobi EW, Stein AD, *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008; 105: 17046-9.
 42. Pembrey ME, Bygren LO, Kaati G, *et al.* Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet.* 2006; 14: 159-66.
 43. Friedenreich CM, Neilson HK, Lynch BM. State of the epidemiological evidence on physical activity and cancer prevention. *Eur J Cancer.* 2010; 46: 2593-604.
 44. Harriss DJ, Cable NT, George K, *et al.* Physical activity before and after diagnosis of colorectal cancer: disease risk, clinical outcomes, response pathways and biomarkers. *Sports Med.* 2007; 37: 947-60.
 45. McTiernan A. Mechanisms linking physical activity with cancer. *Nat Rev Cancer.* 2008; 8: 205-11.
 46. Slattery ML. Physical activity and colorectal cancer. *Sports Med.* 2004; 34: 239-52.
 47. Pischon T, Lahmann PH, Boeing H, *et al.* Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst.* 2006; 98: 920-31.
 48. Campbell PT, Jacobs ET, Ulrich CM, *et al.* Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. *J Natl Cancer Inst.* 2010; 102: 391-400.

49. Slattery ML, Curtin K, Wolff RK, *et al.* Diet, physical activity, and body size associations with rectal tumor mutations and epigenetic changes. *Cancer Causes Control.* 2010; 21: 1237-45.
50. Hodge DR, Peng B, Cherry JC, *et al.* Interleukin 6 supports the maintenance of p53 tumor suppressor gene promoter methylation. *Cancer Res.* 2005; 65: 4673-82.
51. Kang GH, Lee HJ, Hwang KS, *et al.* Aberrant CpG island hypermethylation of chronic gastritis, in relation to aging, gender, intestinal metaplasia, and chronic inflammation. *Am J Pathol.* 2003; 163: 1551-6.
52. Monteiro R, Azevedo I. Chronic inflammation in obesity and the metabolic syndrome. *Mediators of inflammation.* 2010; doi: 10.1155/2010/289645.
53. Derks S, Postma C, Carvalho B, *et al.* Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis.* 2008; 29: 434-9.
54. Li FY, Lai MD. Colorectal cancer, one entity or three. *Journal of Zhejiang University Science.* 2009; 10: 219-29.
55. Lee M, Martin H, Firpo MA, *et al.* Inverse association between adiposity and telomere length: The fels longitudinal study. *Am J Hum Biol.* 2010; doi: 10.1002/ajhb.21109.
56. Kim S, Parks CG, DeRoo LA, *et al.* Obesity and weight gain in adulthood and telomere length. *Cancer Epidemiol Biomarkers Prev.* 2009; 18 :816-20.
57. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature.* 1997; 386: 623-7.
58. Hirschhorn JN, Lettre G. Progress in genome-wide association studies of human height. *Horm Res.* 2009; 71 Suppl 2: 5-13.
59. Soranzo N, Rivadeneira F, Chinappan-Horsley U, *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS genetics.* 2009; 5: e1000445.

60. World Health Organization. 2010. Information sheet on obesity and overweight factsheet. http://www.who.int/dietphysicalactivity/media/en/gsfes_obesity.pdf.
61. Ogden CL, Carroll MD, Curtin LR, *et al.* Prevalence of high body mass index in US children and adolescents, 2007-2008. *Jama*. 2010; 303: 242-9.
62. McKeown-Eyssen G. Epidemiology of colorectal cancer revisited: are serum triglycerides and/or plasma glucose associated with risk? *Cancer Epidemiol Biomarkers Prev*. 1994; 3: 687-95.
63. Silvera SA, Jain M, Howe GR, *et al.* Energy balance and breast cancer risk: a prospective cohort study. *Breast Cancer Res Treat*. 2006; 97: 97-106.

English summary

Overweight is a reflection of energy imbalance, and occurs when the calories consumed (from foods, drinks and alcohol) outnumber the calories used by the body (for normal body functions and basal metabolic rate, daily activities, and exercise). It is well accepted that a large body size (both body fat and adult attained height) increases the risk of colorectal cancer (CRC), while high levels of physical activity and caloric restriction have been shown to decrease the risk of CRC. How this risk differs according to men and women, and according to different anatomical sub-sites of the large intestine is unclear. Furthermore, CRC is not a single disease. Rather, CRC encompasses a heterogeneous complex of diseases characterized by numerous molecular abnormalities. By distinguishing underlying pathways responsible for these differences, the exposure factors that are thought to cause CRC and differential risk between individuals and populations may be better understood.

Of interest to this thesis is the serrated neoplasia pathway to CRC. Tumours arising from this pathway are characterised in part by a high degree of epigenetic instability, namely, widespread DNA methylation in the promoter region of several genes resulting in transcriptional silencing of the gene in question. This phenomenon has been labeled the CpG island methylator phenotype (CIMP). Tumours arising from this pathway also display a high degree of microsatellite instability (MSI), a form of genetic instability, and mutations in the *BRAF* proto-oncogene. This pathway was of particular interest because it is thought that timing of exposure to a given risk factor may be important for establishing epigenetic mechanisms such as methylation.

To date, epidemiologic data for associations between indicators/aspects of energy balance and risk of molecular abnormalities associated with epigenetic instability (CIMP, MSI, *BRAF* mutations) is rare. The aim of this thesis was to elucidate potential differences in risk factors for the sub-sites of the large intestine and sex, and for molecular subtypes of these tumours. This may lead to a better understanding of CRC aetiology and prevention.

All studies in this thesis were completed within the Netherlands Cohort Study on diet and cancer (NLCS). The NLCS is a prospective cohort study that includes 120,852 men and women who were between the ages of 55 and 69 at baseline (1986), and completed a self-administered

questionnaire on dietary habits, lifestyle, and other risk factors for cancer. The cohort was followed-up for cancer incidence through linkage to the Netherlands Cancer Registry, and PALGA (the nationwide registry for histopathology and cytopathology in the Netherlands). Further, a sub-cohort of 5000 individuals was randomly selected after baseline exposure measurement, to estimate accumulation of person time at risk in the cohort through biennial follow-up of vital status. Overall CRC could be investigated among 3360 cases from a follow-up period of 16.3 years after baseline. Tumour material was collected from CRC patients identified within the first 7.3 years of follow-up, excluding the first 2.3 years, and of these, 734 had sufficient DNA for molecular analyses. The research described in Chapters 3, 5 and 6 was based on data from 16.3 years of follow-up of the cohort. The research described in Chapters 4, 7 and 8 was based on data from 7.3 years of follow-up.

We begin in **Chapter 2** by describing a study in which we investigated the utility of self-reported trouser/skirt size as a predictor of cancer risk. Although epidemiological studies frequently utilize waist circumference as an indicator of intra-abdominal fat, and subsequently cancer risk, this variable was collected only from sub-cohort members of the NLCS. In weight-stable sub-cohort members, we observed that self-reported trouser/skirt size correlated well with self-reported waist circumference. Furthermore, self-reported clothing size appeared to predict risk of cancer in men and women independently of BMI. These findings indicate that self-reported trouser/skirt size can be used as a surrogate marker for waist circumference in the NLCS, and therefore, it was considered as an independent risk factor in the analyses presented in Chapters 3-8 of this thesis.

In **Chapter 3**, we investigated the association between exposure to energy restriction during childhood and adolescence and risk of CRC later in life. Energy restriction appears to be one of the most convincing interventions for lowering cancer risk in laboratory animals. When investigating risk in humans, epidemiological researchers often depend on historical data, for example exposure to a period of known energy restriction, such as a famine or war, as a proxy measure for energy restriction earlier in life. The NLCS is comprised of individuals who grew up during the years of the Economic Depression (1932-40), World War 2 (1940-44) and the Hunger Winter of

1945. A number of proxy measures were collected from cohort members that reflect their exposure to energy restriction during these three time periods. We observed no associations between exposure to the Economic Depression, WW2 years and CRC risk, and hypothesized that energy restriction was not severe enough during these time periods to influence risk. We did however observe that men who resided in a city exposed to the Hunger Winter had a lower risk than those who did not. This was true for total CRC, and tumours of the proximal colon and rectum. Inverse associations were also observed for women who were exposed to the Hunger Winter, but these did not reach statistical significance.

We then investigated the association between these exposures and CRC risk according to molecular endpoints, specifically CIMP and MSI, in **Chapter 4**. Adolescence is a period of rapid growth and change, and we hypothesized that exposure to a severe environmental exposure during this period of life, such as what would have been experienced during the Hunger Winter, may influence the development of epigenetic mechanisms. We observed that men and women exposed to the Hunger Winter had a lower risk of developing a tumour characterized by CIMP compared to those not exposed. We did not observe such an association with respect to MSI.

In **Chapter 5**, we considered adult aspects of body size and risk of CRC, specifically BMI at age 20, BMI at baseline, BMI change, trouser/skirt size and height. We observed that risk appears to differ according to sex. In men, risk of CRC increased with increasing BMI, BMI at age 20, BMI change and trouser size. This was especially true for tumours of the distal colon. In women, body fat was not associated with CRC risk in women unless considered simultaneously with physical activity; a high trouser/skirt size and low level of physical activity increased risk of all tumor subtypes. Height was associated with increased risk of CRC in women, but not men.

We also observed sex specific differences with respect to physical activity and CRC risk, as described in **Chapter 6**. Occupational physical activity appeared to be more protective of CRC in men, especially tumours of the distal colon, and non-occupational physical activity appeared to be more protective in women, especially for tumours of the distal colon and rectum.

Next, we wanted to investigate whether BMI, BMI at age 20, trouser/skirt size, height and physical activity differentially influenced risk of molecular subtypes of CRC tumours. In **Chapter 7**, we considered this with

respect to CIMP status. We observed that individuals with a high BMI and/or trouser/skirt size were at greater risk of developing CRC regardless of CIMP status, and physical activity was protective regardless of CIMP status. However, individuals who reported a higher BMI at age 20 had a greater risk of developing a CIMP tumour than a non-CIMP tumour. This observation builds on the hypothesis that methylation is an early event in CRC progression. However, we did not observe a difference in risk with respect to height, which is a proxy for nutritional exposures early in life.

For the study presented in **Chapter 8**, we used data from both the NLCS and the Melbourne Collaborative Cohort Study (MCCS) to assess the association between body size and risk of CRC characterized by *BRAF* mutations and MSI. We pooled data from the studies together to improve the precision of risk estimates, and observed that BMI and waist measurements were associated with all tumour subtypes (*BRAF* mutated and *BRAF* non-mutated; MSI and MS-stable). However, taller individuals were at greater risk of developing a tumour characterized by either a *BRAF* mutation or MSI. These findings provide more evidence that early life exposures may influence (epi) genetic mechanisms later in life. We hypothesize that there was not enough statistical power to observe a differential association in Chapter 7 with respect to height and CIMP.

This thesis concludes with a summary of the main findings, a discussion about the limitations and insights offered by molecular pathological epidemiology, a proposed hypothesis to explain our observations, and implications for future research (**Chapter 9**). We first acknowledge that CRC is a complex disease and that heterogeneity and associations with a given risk factor are far from black and white. Another issue is that in any molecular epidemiological study, sample size will be a problem. One way to overcome this issue is to pool data from similar studies, as we did in Chapter 8 of this thesis. We also address the issue that CIMP as an indicator of epigenetic instability in CRC research has not been without controversy. In order to overcome this controversy, a consensus about gene panels, threshold values and laboratory methods used to define CIMP must be reached.

Based on our observations, we suggest that timing of exposure is important for risk of CRCs characterized by epigenetic instability. This hypothesis may influence future research questions, as the incidence of

childhood overweight and obesity continues to increase and the long-term consequences of being obese during this critical period of growth and development remain largely undetermined.

Nederlandse samenvatting

Overgewicht is een reflectie van een verstoorde energie balans en vindt plaats wanneer de geconsumeerde calorieën (uit voedsel) talrijker zijn dan de calorieën gebruikt door het lichaam (normale lichaamsfuncties en basaalmetabolisme, dagelijkse activiteiten en lichaamsbeweging). Over het algemeen wordt aangenomen dat een groot lichaamsvolume (zowel lichaamsvet en -lengte van een volwassen persoon) het risico op dikkedarmkanker¹ (DDK) verhoogt, terwijl een hoge fysieke activiteit en calorie beperking het risico op DDK verlaagt. Hoe dit risico op kanker verschilt tussen mannen en vrouwen en in verschillende anatomische delen van de dikke darm is onduidelijk. Bovendien staat DDK niet op zichzelf. Beter gezegd is DDK een heterogene ziekte gekarakteriseerd door talrijke moleculaire afwijkingen. Door de onderliggende metabole routen te karakteriseren voor deze verschillen zouden de bekende oorzaken van DDK en de verschillende risico's tussen individuen en populaties beter begrepen kunnen worden.

Van belang in het kader van dit proefschrift is de ontwikkeling van colorectaal carcinoom via de geserreerde route. Tumoren die voortkomen uit dit proces worden gedeeltelijk gekarakteriseerd door een hoge mate van epigenetische instabiliteit. Zo kan genexpressie geblokkeerd worden door DNA methylering in de promotorregio van het gen in kwestie. Dit verschijnsel wordt aangeduid als *CpG island methylator phenotype* (CIMP). Tumoren die via deze route ontstaan correleren met een hoge mate van microsatelliet instabiliteit (MSI), een vorm van genetische instabiliteit, en mutaties in het proto-oncogen *BRAF*. Ongeveer 15% van alle DDKs ontstaan via dit proces en deze tumoren zijn interessant omdat men denkt dat de periode van blootstelling aan de gegeven risicofactor belangrijk kan zijn voor het vaststellen van epigenetische mechanismen als methylering.

Vandaag de dag zijn epidemiologische data over associaties tussen indicatoren van de energie balans en het risico op DDK gekarakteriseerd door moleculaire afwijkingen geassocieerd met epigenetische instabiliteit (CIMP, MSI, *BRAF* mutaties) zeldzaam. Het doel van dit proefschrift was het onderzoeken van potentiële verschillen in risicofactoren voor kanker in verschillende delen van de dikke darm, naar geslacht, en voor moleculaire subtypes van deze tumoren. Dit zou kunnen leiden tot een beter inzicht in de etiologie en preventie van DDK.

¹Onder DDK wordt verstaan kanker van de dikke (colon) en endel (rectum) –darm

Alle studies in dit proefschrift zijn uitgevoerd binnen de Nederlandse Cohort Studie naar voeding en kanker (NLCS). De NLCS bestaat uit 120.852 mannen en vrouwen die bij aanvang van de studie in 1986 een leeftijd tussen de 55 en 69 hadden en een vragenlijst over eetgewoonte, levensstijl en andere risicofactoren voor kanker, ingevuld hebben. Het cohort werd gevolgd voor het optreden van incidentie kankergevallen door een koppeling met de Nederlandse Kankerregistratie en PALGA (Pathologisch-Anatomisch Landelijk Geregistreerd Archief). Bij aanvang van de studie, na de beginmeting, is een subcohort van 5000 personen willekeurig geselecteerd voor het schatten van de opgebouwde persoonstijd door middel van de follow-up van de vitale status. Na een follow-up periode van 16,3 jaar, werd DDK vastgesteld bij 3360 personen uit het totale cohort. Tumormateriaal werd verzameld van DDK patiënten geregistreerd in de eerste 7,3 jaar van de follow-up, met uitzondering van de eerste 2,3 jaar. Van 734 patiënten was voldoende DNA materiaal aanwezig voor moleculaire analyses. Het onderzoek dat beschreven wordt in de hoofdstukken 3, 5 en 6 was gebaseerd op data verkregen gedurende 16,3 jaar na aanvang van het cohort. Het onderzoek beschreven in de hoofdstukken 4, 7 en 8 was gebaseerd op data verkregen gedurende 7.3 jaar na aanvang van de studie.

Als eerste, wordt in **hoofdstuk 2** een studie beschreven over de bruikbaarheid van zelfgerapporteerde broek/rok maat als voorspeller van het risico op kanker. Ondanks het feit dat epidemiologische studies vaak taille omtrek gebruiken als een indicator voor intra-abdominaal vet en vervolgens het risico op kanker, werd deze variabele slechts verzameld voor een deel, de subcohortleden, van de NLCS. We observeerden in subcohortleden met een stabiel gewicht dat zelfgerapporteerde broek/rok maat goed correleert met zelfgerapporteerde taille omtrek. Daarbij bleek dat de zelfgerapporteerde kledingmaat het risico op kanker in mannen en vrouwen voorspelt onafhankelijk van BMI. Deze bevindingen geven aan dat zelfgerapporteerde broek/rok maat als een surrogaat marker voor taille omtrek gebruikt kan worden in de NLCS en werd daardoor beschouwd als een onafhankelijke risicofactor in de analyses beschreven in hoofdstukken 3 tot en met 8 van dit proefschrift.

In **hoofdstuk 3** onderzochten we de associatie tussen energie beperking gedurende de kinderjaren en adolescentie en het risico op DDK op latere leeftijd. Energie beperking blijkt een van de meest overtuigende

interventies te zijn voor het verlagen van het risico op kanker in proefdieren. Voor het onderzoeken van deze verbanden bij mensen zijn epidemiologische onderzoekers voornamelijk afhankelijk van historische data. De blootstelling aan een periode van energie beperking, zoals een hongersnood of oorlog, kan daarbij dienen als een indirecte maat voor energie beperking op jongere leeftijd. NLCS deelnemers groeiden op ten tijde van de Grote Depressie (1932-1940), de Tweede Wereldoorlog (1940-1944) en de Hongerwinter van 1945. Deelnemers waren gevraagd te rapporteren of hun vader werk had tijdens de Depressie en wat het woonadres was tijdens de Oorlog en de Hongerwinter. Deze informatie werd beschouwd als proxy maten voor de mate van energie beperking gedurende de jeugd. We zagen geen associaties tussen de blootstelling aan de Grote Depressie en de Tweede Wereldoorlog en het risico op DDK en veronderstelden dat de energie beperking niet hevig genoeg was gedurende deze periodes om van invloed te zijn op het risico. We observeerden echter wel dat mannen die in een stad woonden die getroffen was door de Hongerwinter een lager risico op DDK hadden dan degene die daar niet woonden. Dit was in het bijzonder het geval voor totale DDK en tumoren in het proximale deel van het colon en het rectum. Inverse associaties werden eveneens geobserveerd voor vrouwen die de Hongerwinter ondervonden, deze waren echter niet statistisch significant.

Vervolgens onderzochten we in **hoofdstuk 4** de associatie tussen deze periodes en het risico op DDK gekarakteriseerd door moleculaire veranderingen, in het bijzonder CIMP en MSI. Adolescentie is een periode van snelle groei en verandering en we stelde de hypothese op dat de gevolgen van blootstelling aan heftige omstandigheden gedurende deze periode van het leven, zoals dat wat men ervaart gedurende de Hongerwinter, de ontwikkeling van epigenetische mechanismen beïnvloedt. We observeerden dat mannen en vrouwen die blootgesteld waren aan de Hongerwinter een lager risico hadden om tumoren te ontwikkelen die door CIMP gekarakteriseerd werden in vergelijking met diegene die niet blootgesteld waren aan de Hongerwinter. We vonden geen associatie met MSI.

In **hoofdstuk 5** beschouwden we de volwassen aspecten van lichaamsgrootte en het risico op DDK, in het bijzonder BMI op 20-jarige leeftijd, BMI bij aanvang van de studie, BMI verandering, broek/rok maat en lengte. We observeerden dat het risico afhankelijk lijkt te zijn van geslacht. In mannen leidt een hogere BMI, BMI op 20-jarige leeftijd, BMI verandering

en broekmaat, tot een hoger het risico op DDK. Dit was voornamelijk het geval bij tumoren in de distale deel van de dikke darm. Lichaamsvet was bij vrouwen niet geassocieerd met het risico op DDK, tenzij het samen met lichamelijke activiteit werd beschouwd. Een grote broek/rok maat en een lage maat van fysieke activiteit verhoogden het risico op alle tumor types. Lichaamslengte was geassocieerd met een verhoogd risico op DDK in vrouwen, echter niet in mannen.

We namen eveneens geslachtspecifieke verschillen waar met betrekking tot lichamelijke activiteit en het risico op DDK, zoals beschreven in **hoofdstuk 6**. Beroepsmatige lichamelijke activiteit bleek beter beschermend tegen DDK in mannen, vooral tegen tumoren in het distale deel van de dikke darm, en niet-beroepsmatige fysieke activiteit bleek beter beschermend in vrouwen, vooral tegen tumoren in het distale deel van de dikke darm en het rectum.

Vervolgens wilden we onderzoeken in welke mate BMI, BMI op 20-jarige leeftijd, broek/rok maat, lengte en fysieke activiteit van verschillende invloed zijn op het risico op moleculaire subtypes van DDK tumoren. In **hoofdstuk 7** beschouwden we dit met betrekking tot CIMP status. We observeerden dat individuen met een hoge BMI en/of een grote broek/rok maat een hoger risico liepen om DDK te ontwikkelen ongeacht de CIMP status. Lichamelijke activiteit acteerde als een beschermende factor ongeacht de CIMP status. Individuen die een hoger BMI op 20-jarige leeftijd rapporteerden liepen echter een groter risico op het ontwikkelen van een CIMP tumor dan een non-CIMP tumor. Deze observatie past bij de hypothese dat methylering optreedt in een vroeg stadium in de ontwikkeling van DDK. We namen echter geen verschil in risico waar met betrekking tot lengte, welke als een indirecte maat voor de voedingsstatus in de jonge levensjaren gezien kan worden.

Voor de beschreven studie in **hoofdstuk 8** gebruikten we data van zowel de NLCS als de *Melbourne Collaborative Cohort Study* (MCCS) om de associatie tussen lichaamsgrootte en het risico op DDK gekarakteriseerd door *BRAF* mutaties en MSI te bepalen. De data van beide studies zijn samengevoegd om de nauwkeurigheid van de risicoschattingen te vergroten en we zagen dat BMI en omtrekmetingen geassocieerd waren met alle tumor types (*BRAF* gemuteerd en niet gemuteerd, MSI en MS-stabiel). Langere mensen liepen echter een groter risico op het ontwikkelen van een tumor die

gekaracteriseerd werd door of een *BRAF* mutatie of MSI. Deze bevindingen verschaffen extra bewijs dat blootstellingen gedurende de jonge jaren (epi) genetische mechanismen kunnen beïnvloeden. We stellen dat er niet genoeg statistische power was om verschillende associaties waar te nemen in hoofdstuk 7 met betrekking tot lengte en CIMP.

Dit proefschrift sluit af met een samenvatting van de belangrijkste bevindingen, een discussie over de beperkingen en inzichten verschaft door de moleculaire pathologische epidemiologie, een voorgestelde hypothese om onze waarnemingen te verklaren en suggesties voor toekomstig onderzoek (**hoofdstuk 9**). Als eerste erkennen we dat DDK een complexe ziekte is en dat de heterogeniteit en associaties met een gegeven risicofactor complex zijn. Een andere kwestie is dat in iedere moleculaire epidemiologische studie de grootte van de steekproef een probleem zal zijn. Een manier om deze kwestie te overwinnen is door de data van verschillende studies te poolen, zoals we deden in hoofdstuk 8 van dit proefschrift. We brengen eveneens de kwestie naar voren dat CIMP als een indicator voor epigenetische instabiliteit niet oncontroversieel is in DDK onderzoek. In een poging deze controverse te ondervangen, moet overeenstemming bereikt worden de gebruikte markers, afkappunten en laboratorium methodes die gebruikt worden om CIMP te definiëren.

Gebaseerd op onze waarnemingen suggereren we dat de timing van blootstelling belangrijk is voor het risico op DDK gekarakteriseerd door epigenetische instabiliteit. Deze hypothese kan toekomstige onderzoeksvragen beïnvloeden, vooral zolang de incidentie van overgewicht gedurende de kinderjaren blijft stijgen en de lange termijn gevolgen van obesitas tijdens deze kritieke periode van groei en ontwikkeling grotendeels onduidelijk blijven.

Acknowledgements

Surviving the past few years abroad and completing this PhD project would have been impossible without the help of many people.

To my Promotion team...

- Matty, I can't help but feel like I hit the 'supervisor jackpot' with you! I couldn't have asked for better guidance over the course of my PhD. Your approachability, friendliness and knowledge made working on this project a pleasure, and I am thankful that we are continuing our collaboration.
- Manon, your enthusiasm for this research is contagious! Thank you letting this epidemiologist loose in the Pathology lab and for all of your support over the course of my project, and now during my post-doc.
- Piet, thanks for all of your critiques and insights with respect to my articles, and also for switching the NLCS meetings to English. Needless to say, that's made life much easier ☺.

In 2009, I was lucky enough to have the opportunity to visit the group of Professor Dallas English at the University of Melbourne, and live in Australia for four months...

- Dallas, you made so much time for me during my visit; I felt incredibly welcome in Melbourne, and what I learned while working with the MCCS data has been invaluable to my development as an epidemiologist. Thank you also to Mark Jenkins and Fizz Williamson for their support on my project.
- Thanks also to Jen & Ian, Ben & Kim, Dave & Michelle and Shaz & Craig for making my Australian social life so wonderful!

To my colleagues...

- Thanks to everyone in the NLCS and the departments of Epidemiology and Pathology for creating a great working atmosphere. I must especially thank Sandra Bausch, Adriaan

de Bruïne, Ton de Goeij and Leo Schouten for your collaboration and helpful feedback on various articles; Bas, for the great drawing of the CRC tract for my thesis introduction; Harry and Jos for all the computer help; Nathalie, Yvonne, Connie, Jolanda and Petra for answering questions; and of course, my fellow PhD students, old and new!

To special friends...

- Anke & Colinda, thanks for agreeing to be my paranymphs (no pressure, haha), and for not caring about my messy desk. I couldn't have made it through the last months without your camaraderie, stimulating discussions about epidemiology and frequent tea/coffee breaks.
- Ruth, my 'soul mate' from Deb1, it feels like we've known each other forever. Thanks (no, really!) for moving to London, and for the hospitality you and Paul always show me when I visit. Hannah: It's been a long time since we shared an office, but you made my early days in Maastricht so much fun! I'll always remember it. Stephanie, thanks for the cocktail breaks, movie nights and for acting so Canadian!
- Kim & Marelle (plus Douwe!), Robbert-Jan & Felicia (plus Quinn!) and Lex & Jessica (plus Renske!): thanks for all the gezellig dinners, laughs and shots of Sambuca. You guys are wonderful friends.
- My favourite ex-pats: Audrey, Danni, Katarina, Patrick and Sofia, you guys are awesome (Audrey, I'm still in awe of your love of Holland and Dutch customs. And of course, your infinite supply of maple syrup has been a lifesaver).
- To the friends at home who have kept me close despite the distance, I appreciate the visits, emails, letters, Skype chats, care packages, wedding involvement and epic cycling adventures more than you'll ever know (Krista, it's safe to say life would not be complete without that little jaunt to

Laos). So, Ashleigh, Erica, Jodey, Kat, Krista, Laura, Meghan B., Meghan S. and Robyn- thank you!

To family...

- In Goirle: Ad, Corrie en Max, gullie het d'r vûr gezurgd de'k mezelf jullie kiet as twidde thûs zâág, en het waar verrekkes leuk deh gullie altij benuut waart naor we'k ammel deej! Ik beloof oe de'k oht Tilburgs zal verstaon en misschien ok wel praoten, ge wit oit noit...☺
- In Canada and England: Grandma, thanks for always being interested in what I am doing. Now, I just have to make it to age 90 too! Nancy and Brian, thanks for all of the exotic 'study breaks' and for your support in general! Steve and Wendy, thanks for taking the time to visit us in Maastricht when you were passing through; it's great you can come back for my defense! Maureen, JJ and Steph, thanks for providing me with a second home in Oxfordshire the past few years- the red wine and board games are always much appreciated.
- Katy and Julia, I guess absence really does make the heart grow fonder; I am glad you are my sisters. Mum, your positivity and dignity through the ups and downs of life, especially recently, is humbling. We love you very much!
- I never imagined that my father would miss this day. If it wasn't for his globe-trotting spirit, I doubt that my life would have taken this track. So, I dedicate this book to him. Dad, I miss you.

To Richard:

- Last but not least, Richard: I'm not sure you realized what you were getting into all those years ago at the *C'est la Vie*, but I have no regrets. Simply, bedankt! Xo

Curriculum vitae

Laura Hughes was born on June 3rd, 1982 in Vancouver, Canada. After completing secondary school (Cardinal Newman Catholic Secondary School, Toronto) in 2001, she started undergraduate studies at the University of Guelph. Four years later, including six months abroad at the University of Aberdeen (Scotland), Laura graduated with a BSc. Honours degree in Nutrition and Nutraceutical Science. Soon after, she was accepted into graduate studies in Human Health and Nutrition at the University of Guelph, and received her MSc. in the summer of 2006 after fulfilling an internship of the topic of lipid induced mitochondrial dysfunction in type 2 diabetes at Maastricht University, the Netherlands. Intrigued by life in Europe, Laura decided to pursue a second Masters degree, this time in Epidemiology, at Maastricht University. For her thesis, she co-authored an article entitled, 'Higher dietary flavone, flavonol, and catechin intakes are associated with less of an increase in BMI over time in women: a longitudinal analysis from the Netherlands Cohort Study', and graduated cum laude in 2007. From 2007 to 2010, she was employed as a PhD student in the department of Epidemiology at Maastricht University, where she worked on the project described in this thesis, and completed a three month research visit to the University of Melbourne, Australia. As of February 2011, Laura works as a post-doctoral fellow in the department of Epidemiology at Maastricht University.

List of publications

Published in International Journal with Impact Factor

1. **LAE Hughes**, CCJM Simons, PA van den Brandt, RA Goldbohm, AFPM de Goeij, AP de Bruïne, M van Engeland, MP Weijnenberg. Body size, physical activity and risk of colorectal cancer with or without the CpG island methylator phenotype (CIMP). *PLoS ONE* 2011 (in press).
2. **LAE Hughes**, PA van den Brandt, RA Goldbohm, AFPM de Goeij, AP de Bruine, M van Engeland, MP Weijnenberg. Childhood and adolescent energy restriction and subsequent colorectal cancer risk: results from the Netherlands Cohort Study. *Int J Epid.* 2010; 39 (5): 1333-44.
3. **LAE Hughes**, PA van den Brandt, AP de Bruine, KAD Wouters, S Hulsmans, A Spiertz, RA Goldbohm, AFPM de Goeij, JG Herman, MP Weijnenberg, M van Engeland. Early life exposure to famine and colorectal cancer risk: a role for epigenetic mechanisms? *PLoS ONE.* 2009; 4 (11): e7951. doi:10.1371/journal.pone.0007951
4. CCJM Simons, **LAE Hughes**, ICW Arts, RA Goldbohm, PA van den Brandt, MP Weijnenberg. Dietary flavone, flavonol and catechin intake and risk of colorectal cancer in the Netherlands Cohort Study. *Int J Cancer.* 2009; 125 (12): 2945-52.
5. **LAE Hughes**, LJ Schouten, RA Goldbohm, PA van den Brandt, MP Weijnenberg. Self-reported clothing size as a proxy measure for body size. *Epidemiology.* 2009; 20 (5): 673-6.
6. KM Smits, AHG Cleven, MP Weijnenberg, **LAE Hughes**, JG Herman, AP de Bruïne, M van Engeland. Pharmacogenomics in colorectal cancer; a step forward in predicting prognosis and treatment response. *Pharmacogenomics.* 2008; 9 (12): 1903-16.

7. **LAE Hughes**, ICW Arts, T Ambergen, HA Brants, PC Dagnelie, RA Goldbohm, PA van den Brandt, MP Weijenberg. Higher dietary flavone, flavonol, and catechin intakes are associated with less of an increase in BMI over time in women: An analysis from the Netherlands Cohort Study. *Am J Clin Nutr.* 2008; 88 (5): 1341-52.

Letters to the Editor:

1. **LAE Hughes**, LJ Schouten, RA Goldbohm, PA van den Brandt, MP Weijenberg. Am I Shrinking? On Clothing Size and Body Size Response. *Epidemiology* 2010; 21 (1): 160.

Other:

1. **LAE Hughes**. Flavonoïden helpen lichaamsgewicht handhaven. *Voeding NU* 2009; 11 (3): 34.

Submitted:

1. **LAE Hughes**, EJ Williamson, M van England, MA Jenkins, GG Giles, J Hopper, M Southey, J Young, D Buchanan, M Walsh, PA van den Brandt, R Alexandra Goldbohm, MP Weijenberg, and DR English. Anthropometric variables and risk for colorectal cancers showing BRAF V600E mutations or microsatellite instability.
2. **LAE Hughes**, CCJM Simons, PA van den Brandt, RA Goldbohm, M van Engeland, MP Weijenberg. Body size and risk of colorectal cancer after 16.3 years of follow-up.
3. CCJM Simons, **LAE Hughes**, PA van den Brandt, RA Goldbohm, M van Engeland, MP Weijenberg. Physical activity and risk of colorectal cancer after 16.3 years of follow-up.

4. AMJ Gilsing, MP Weijenberg, **LAE Hughes**, T Ambergen, PC Dagnelie, RA Goldbohm, PA van den Brandt, LJ Schouten. Meat consumption and changes in BMI over time: longitudinal analyses from the Netherlands Cohort Study.