Diet and postmenopausal breast cancer - With a focus on low-grade inflammation

Alves Dias, Joana

2017

Document Version:
Publisher's PDF, also known as Version of record

Diet and postmenopausal breast cancer

With a focus on low-grade inflammation

Diet and postmenopausal breast cancer

With a focus on low-grade inflammation

Joana Alves Dias

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended at the CRC Aula, Clinical Research Centre, Entrance 72, Jan Waldenströms gata 35, Skåne University Hospital, Malmö, Friday 10th of February 2017, at 9:00 am.

Faculty opponent
Bethany van Guelpen

Associate Professor at the Department of Radiation Sciences,
Umeå University, Sweden

Organization LUND UNIVERSITY	Document name DOCTORAL DISSERTATION
Department of Clinical Sciences, Malmö	Date of issue
Author: Joana Alves Dias	Sponsoring organization

Title and subtitle: Diet and postmenopausal breast cancer – With a focus on low-grade inflammation

Abstract

Diet-breast cancer studies have shown that “healthy eating patterns” are associated with decreased risk whereas unhealthy patterns (especially those including alcohol) are associated with increased risk, particularly in postmenopausal women. The potential mechanisms behind the observed associations are still under investigation. A great deal of evidence supports the major role of lifelong overexposure to sex hormones in the induction and progression of breast cancer, especially after menopause. However, this alone cannot fully explain the variation of breast cancer incidence across populations, and we hypothesize that an inflammatory environment, promoted by a Western lifestyle, may also play an important role. It is accepted that inflammation is an important feature in cancer development and progression, but also that cancer induces inflammatory processes.

This thesis aimed to investigate the role of diet in the development of postmenopausal breast cancer, with a special interest in low-grade inflammation as a possible pathway. A population-based cohort, the Malmö Diet and Cancer (MDC) Study, consisting of 28,098 participants was used. The baseline examinations, that took place between 1991 and 1996, included blood sampling, anthropometric measurements and the detailed collection of dietary data.

In study I, we inspected the reliability of several biomarkers of inflammation, examining a random sample of 95 people (46 women and 49 men) recruited from the MDC cohort. Six blood samples were taken at different occasions during a 6-week period in 2010-2011 (in fasting and non-fasting states). Intraclass correlation coefficients for the biomarkers were estimated. In study II, the association between diet quality and several inflammatory biomarkers was examined. A group of 667 individuals from the MDC-cardiovascular arm were randomly selected, and baseline data on diet and biomarkers of inflammation were investigated. Studies III and IV used a nested-case control design with 446 breast cancer cases and 910 matched controls. In study III, we analyzed the breast cancer risk associated with specific biomarkers and the possible role of obesity in this association. Finally, the association between dietary patterns derived to explain the variation of certain inflammation markers and breast cancer was explored in study IV.

Our findings indicated a high reliability for the biomarkers of inflammation. Lower concentrations of biomarkers of inflammation were associated with higher diet quality, as assessed by overall adherence to the Swedish nutrition recommendations. We found three inflammation markers (ox-LDL, IL-1β and TNF-α) to be associated with breast cancer independent of obesity, but with diverging directions. We did not find evidence for inflammation-driven dietary patterns to be associated with breast cancer risk.

In conclusion, an overall higher diet quality pattern was associated with lower inflammation. However, inflammation did not seem to explain possible associations between diet and postmenopausal breast cancer, as the dietary patterns identified to explain the variation in biomarkers of inflammation did not associate with breast cancer.

Key words: oxidative stress, low-grade inflammation, biomarkers of inflammation, cytokines, ICC, variation, diet, dietary recommendations, diet quality index, food patterns, reduced rank regression, obesity, CVD, postmenopausal breast cancer, epidemiology,cross sectional, nested case-control

Classification system and/or index terms (if any)

Supplementary bibliographical information | Language: English

ISSN and key title 1652-8220 | ISBN 978-91-7619-406-5

Recipient’s notes | Number of pages 241 | Price | Security classification

Diet and postmenopausal breast cancer

With a focus on low-grade inflammation

Joana Alves Dias

LUND UNIVERSITY

Research group in Nutritional Epidemiology
Department of Clinical Sciences in Malmö
Faculty of Medicine, Lund University, Sweden

Cover photo by leonori/Shutterstock.com.

Faculty of Medicine | Department of Clinical Sciences in Malmö

Lund University, Faculty of Medicine Doctoral Dissertation Series 2017:25
ISBN 978-91-7619-406-5
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2017

To my family

Contents

List of papers

This doctoral thesis is based on the following original papers:

I. **Dias JA**, Hellstrand S, Ericson U, Gullberg B, Nilsson J, Alm R, Persson M, Engström G, Fredrikson GN, Hedblad B, Wirfält E. Plasma variation and reproducibility of oxidized LDL-cholesterol and low-grade inflammation biomarkers among participants of the Malmö Diet and Cancer cohort. ***Biomarkers*** 2016:1-10.

II. **Dias JA**, Wirfält E, Drake I, Gullberg B, Hedblad B, Persson M, Engström G, Nilsson J, Schiopu A, Fredrikson GN, Björkbacka H. A high quality diet is associated with reduced systemic inflammation in middle-aged individuals. ***Atherosclerosis*** 2015;238(1):38-44.

III. **Dias JA**, Fredrikson GN, Ericson U, Gullberg B, Hedblad B, Engström G, Borgquist S, Nilsson J, Wirfält E. Low-grade inflammation, oxidative stress and risk of post-menopausal breast cancer – a nested case-control study from the Malmö Diet and Cancer cohort. ***PLoS One*** 2016;11(7):e0158959.

IV. **Dias JA**, Drake I, Ericson U, Gullberg B, Hedblad B, Engström G, Borgquist S, Nilsson J, Fredrikson GN, Wirfält E. Low-grade inflammation-associated food patterns and risk of postmenopausal breast cancer – a nested case-control study from the Malmö Diet and Cancer cohort. ***Submitted manuscript***.

List of papers not included in this thesis

1. Huseinovic E, Winkvist A, Slimani N, Park MK, Freisling H, Boeing H, Buckland G, Schwingshackl L, Weiderpass E, Rostgaard-Hansen AL, Tjonneland A, Affret A, Boutron-Ruault MC, Fagherazzi G, Katzke V, Kuhn T, Naska A, Orfanos P, Trichopoulou A, Pala V, Palli D, Ricceri F, Santucci de Magistris M, Tumino R, Engeset D, Enget T, Skeie G, Barricarte A, Bonet CB, Chirlaque MD, Amiano P, Quirós JR, Sánchez MJ, **Dias JA**, Drake I, Wennberg M, Boer JMA, Ocké MC, Werschuren WMM, Lassale C, Perez-Cornago A, Riboli E, Ward H and Bertéus Forslund H. Meal patterns across ten European countries – results from the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. Public Health Nutr 2016:1-12.
2. Merritt MA, Tzoulaki I, van den Brandt PA, Schouten LJ, Tsilidis KK, Weiderpass E, Patel CJ, Tjonneland A, Hansen L, Overvad K, His M, Dartois L, Boutron-Ruault MC, Fortner RT, Lagiou P, Bamia C, Palli D, Krogh V, Tumino R, Ricceri F, Mattiello A, Bueno-de-Mesquita HB, Onland-Moret NC, Peeters PH, Skeie G, Jareid M, Quirós JR, Obón-Santacana M, Sánchez MJ, Chamosa S, Huerta JM, Barricarte A, **Dias JA**, Sonestedt E, Idahl A, Lundin E, Wareham EJ, Khaw KT, Travis RC, Ferrari P, Riboli E and Gunter MJ. Nutrient-wide association study of 57 foods/nutrients and epithelial ovarian cancer in the European Prospective Investigation into Cancer and Nutrition study and the Netherlands Cohort Study. Am J Clin Nutr 2016;103(1):161-7.
3. Besevic J, Gunter MJ, Fortner RT, Tsilidis KK, Weiderpass E, Onland-Moret NC, Dossus L, Tjonneland A, Hansen L, Overvad K, Mesrine S, Baglietto L, Clavel-Chapelon F, Kaaks R, Aleksandrova K, Boeing H, Trichopoulou A, Lagiou P, Bamia C, Masala G, Agnoli C, Tumino R, Ricceri F, Panico S, Bueno-de-Mesquita HB, Peeters N, Jareid M, Quirós JR, Duell EJ, Sánchez MJ, Larrañaga N, Chirlaque MD, Barricarte A, **Dias JA**, Sonestedt E, Idahl A, Lundin E, Wareham NJ, Khaw KT, Travis RC, Rinaldi S, Romieu I, Riboli E, Merrit MA. Reproductive factors and epithelial ovarian cancer survival in the EPIC cohort study. Br J Cancer 2015;113(11):1622-31.

4. Lette M, Bemelmans WJ, Breda J, Slobbe LC, **Dias JA**, Boshuizen HC. Health care costs attributable to overweight calculated in a standardized way for three European countries. Eur J Health Econ 2014;17(1):61-9.

Abstract

Diet-breast cancer studies have shown that "healthy eating patterns" are associated with decreased risk whereas unhealthy patterns (especially those including alcohol) are associated with increased risk, particularly in postmenopausal women. The potential mechanisms behind the observed associations are still under investigation. A great deal of evidence supports the major role of lifelong overexposure to sex hormones in the induction and progression of breast cancer, especially after menopause. However, this alone cannot fully explain the variation of breast cancer incidence across populations, and we hypothesize that an inflammatory environment, promoted by a Western lifestyle, may also play an important role. It is accepted that inflammation is an important feature in cancer development and progression, but also that cancer induces inflammatory processes.

This thesis aimed to investigate the role of diet in the development of postmenopausal breast cancer, with a special interest in low-grade inflammation as a possible pathway. A population-based cohort, the Malmö Diet and Cancer (MDC) Study, consisting of 28,098 participants was used. The baseline examinations, that took place between 1991 and 1996, included blood sampling, anthropometric measurements and the detailed collection of dietary data.

In study I, we inspected the reliability of several biomarkers of inflammation, examining a random sample of 95 people (46 women and 49 men) recruited from the MDC cohort. Six blood samples were taken at different occasions during a 6-week period in 2010-2011 (in fasting and non-fasting states). Intraclass correlation coefficients for the biomarkers were estimated. In study II, the association between diet quality and several inflammatory biomarkers was examined. A group of 667 individuals from the MDC-cardiovascular arm were randomly selected, and baseline data on diet and biomarkers of inflammation were investigated. Studies III and IV used a nested-case control design with 446 breast cancer cases and 910 matched controls. In study III, we analyzed the breast cancer risk associated with specific biomarkers and the possible role of obesity in this association. Finally, the association between dietary patterns derived to explain the variation of certain inflammation markers and breast cancer was explored in study IV.

Our findings indicated a high reliability for the biomarkers of inflammation. Lower concentrations of biomarkers of inflammation were associated with higher

diet quality, as assessed by overall adherence to the Swedish nutrition recommendations. We found three inflammation markers (ox-LDL, IL-1β and TNF-α) to be associated with breast cancer independent of obesity, but with diverging directions. We did not find evidence for inflammation-driven dietary patterns to be associated with breast cancer risk.

In conclusion, an overall higher diet quality pattern was associated with lower inflammation. However, inflammation did not seem to explain possible associations between diet and postmenopausal breast cancer, as the dietary patterns identified to explain the variation in biomarkers of inflammation did not associate with breast cancer.

Abbreviations

BIA	Bioelectric impedance analysis
BMI	Body mass index
BMR	Basal metabolic rate
CA	Cluster Analysis
CD14/16	Subtypes of monocytes
cells/µl	Cell count per microliter of whole blood
CI	Confidence interval
COX	Cyclooxygenases
CRP	C-reactive protein
CV	Coefficient of variation
CV_A	Analytical coefficient of variation
CV_B	Between-subject coefficient of variation
CVD	Cardiovascular disease
CV_I	Within-subject biological coefficient of variation
CV_W	Within-subject coefficient of variation
DAG	Directed acyclic graph
DCIS	Ductal carcinoma in situ
dl	Deciliter
DLW	Doubly labeled water
DM2	Type 2 diabetes mellitus
DNA	Deoxyribonucleic acid
DQI-SNR	Diet quality index
EI	Energy intake

EPIC	European Prospective Investigation into Cancer and Nutrition
ER	Estrogen receptor
E%	Energy percentage
FA	Factor Analysis
FFQ	Food frequency questionnaire
FSH	Follicle-stimulating hormone
GLM	General linear model
GSH	Glutathione
HDL	High-density lipoprotein
HEI	Healthy Eating Index
HER2	Human epidermal growth factor receptor 2
HRT	Hormone replacement therapy
IARC	International Agency for Research on Cancer
ICC	Intraclass correlation coefficient
ICD	International Classification of Diseases
IDL	Intermediate-density lipoproteins
IGF-1	Insulin-like growth factor-1
IHC	Immunohistochemistry
IL	Interleukin
IS	Index-based scores
Kcal	Kilocalories
Kg	Kilogram
LCIS	Lobular carcinoma in situ
LDL	Low-density lipoprotein
LH	Luteinizing hormone
LLOD	Lower limit of detection
MAF	Minor allele frequency
MDC	Malmö Diet and Cancer
MDC-CC	Malmö Diet and Cancer – Cardiovascular cohort

mg	Milligram
mg/dl	Milligram per deciliter
MHT	Menopausal hormone therapy
MJ	Megajoule
ml	Milliliter
mmHg	Millimeter of mercury
MUFA	Monounsaturated fatty acids
MRI	Magnetic resonance imaging
NCD	Non-communicable diseases
NO	Nitric oxide
NSAID	Nonsteroidal anti-inflammatory drugs
OC	Oral contraceptives
OR	Odds ratio
Ox-LDL	Oxidized low-density lipoprotein
PA	Physical activity
PAL	Physical activity level
PET	Positron emission tomography
pg/ml	Pictogram per milliliter
PR	Progesterone receptor
PUFA	Polyunsaturated fatty acids
ROS	Reactive oxygen species
RNS	Reactive nitrogen species
RR	Relative risk
SD	Standard deviation
SDG	Swedish dietary guidelines
SE	Standard error
SFA	Saturated fatty acids
SNR	Swedish nutrition recommendations
SOD	Superoxide dismutase

TAG	Triacylglycerol
TCA	Tricarboxylic acid cycle (also known as the Krebs cycle)
TEI	Total energy intake (reported)
TNF	Tumor necrosis factor
VLDL	Very low-density lipoprotein
WBC	White blood cells
WCRF	World Cancer Research Fund
WHO	World Health Organization
WHR	Waist-to-hip ratio
µg/ml	microgram per milliliter
%	Percentage

1. Introduction

Breast cancer is the most common form of cancer affecting women worldwide. The reduction of mortality rates and increase in incidence rates in the past decades translates into large numbers of women being treated for and therefore living with this disease. This represents a heavy burden to societies and is a widespread problem, present both in high- and low-income countries. Focusing on prevention rather than on treatment, might be the answer for this serious health problem.

The etiology of breast cancer is not yet fully understood, but it is accepted that it is a multifactorial disease. In the quest to understand the biological mechanisms of and possible risk factors for the disease, much emphasis has been placed upon the hormonal aspects of breast cancer.

It is accepted that oxidative stress combined with low-grade inflammation can contribute to and participate in several phases of the carcinogenesis. A localized inflammatory environment is characteristic of all tumors and contributes to their progression. However, little is known about the role of oxidative stress and low-grade inflammation in the development of breast cancer.

A few decades ago, diet was considered "the promised land" for researchers. Despite the difficulties of capturing what people eat, major efforts were made to evaluate the influence of diet in the development of non-communicable diseases. Healthy dietary patterns are associated with a lower breast cancer risk. However, it is difficult to pinpoint the roles of specific nutrients.

It is important to understand the whole picture and to unravel all the pieces behind the mechanisms leading to cancer. Investigating environmental factors, which may be modifiable, presents a major opportunity to benefit public health.

This thesis aims to investigate the role of diet in the development of postmenopausal breast cancer. It intends to shed light on factors related to low-grade inflammation and oxidative stress as a possible explanatory pathway through which diet may play a role. This may help to identify areas of research that deserve more attention from the scientific community, and to develop new public health strategies focused on modifiable risk factors that can reduce the burden of the disease.

2. Background

2.1. Breast Cancer

Breast cancer is a heterogeneous disease. Depending on its location in the breast, its stage of development, or whether it is pre- or postmenopausal, the implications are different. Thus, it is important to characterize the disease.

Definition and biology

Biology of the breast and regular functioning

The female breast is composed of several types of tissue: fat tissue, glandular epithelial tissue (comprising ducts and lobules), fiber tissue, blood vessels, nerves, lymph vessels, lymph nodes and skin tissue (**Figure 1**). Suspensory ligaments connect the breast tissue to the *pectoralis major* muscle, overlaying the chest wall. The mammary lobes are composed of small lobules that are similar to little bags. Milk production occurs in the lobules when women are lactating, and milk is distributed via the lactiferous ducts, which converge in the nipple [1].

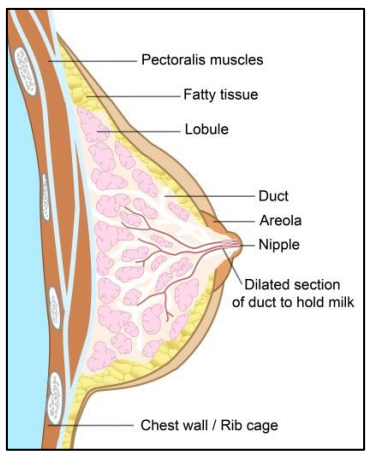

Figure 1. Biology of the female breast

Carcinogenesis and a definition of breast cancer

Normally functioning cells have the potential to become cancer cells. This process is called carcinogenesis and occurs when uncontrolled cell growth and division occur in an indefinite manner, surpassing the tight regulation processes. This may be a long process as it may take several decades from the first genetic mutation until the cancer is diagnosed. In the year 2000, 6 hallmarks of cancer (that is, common characteristics) were proposed to explain the mechanisms through which most cancers develop [2]: a) there is self-sufficiency of growth signals, and these cells are less dependent on external signaling; b) cells become insensitive to anti-growth signals: c) their ability to replicate becomes limitless (i.e., the cells become immortal); d) they are able to evade apoptosis (programmed cell death); e) they can stimulate angiogenesis, enabling access to the nutrients supplied by the new blood vessels; and f) are able to invade tissue and metastasize (i.e., migrate and spread to other locations) (**Figure 2**). This process was later extended using 4 additional traits (two that were considered emerging hallmarks and the other two of which were enabling characteristics) [3]: cellular metabolism is deregulated to the benefit of the tumor proliferation; the cell can avoid destruction by the immune system; genomic instability and mutation enable the tumor; and inflammation occurs that promotes tumor development (**Figure 2**).

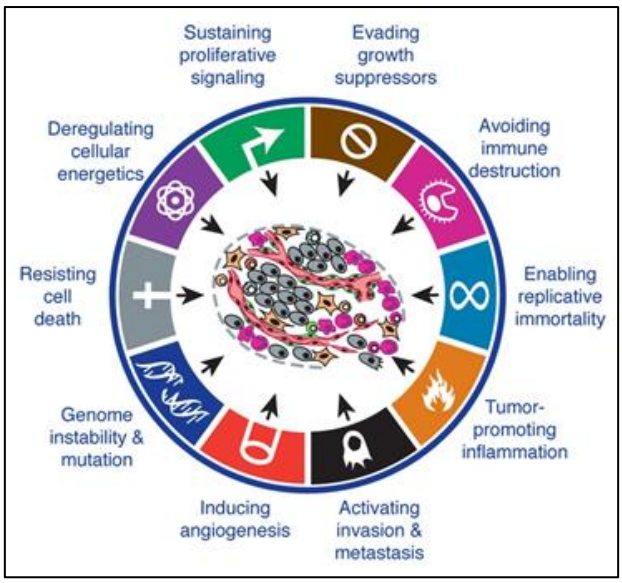

Figure 2. The 10 hallmarks of cancer
Adapted and reprinted from Cell, Vol.144(5), Hanahan D & Weinberg RA, Hallmarks of Cancer: The Next Generation, Pages No. 645-674, Copyright (2011), with permission from Elsevier.

Breast cancer can be classified according to different aspects: invasive versus in situ, histological subtypes, molecular subtypes, grade, stage, etc.

A carcinoma is located in the epithelial tissue (ducts or lobules), whereas a sarcoma is located in the stroma (anything that is not epithelial: fat tissue, fiber tissue, etc.). The majority of breast cancers are carcinomas, and approximately 80-85% occur in the ducts [1]. An in situ cancer is localized and has no ability to spread (also known as "pre-cancer" or ductal carcinoma in situ – DCIS). In contrast, invasive cancers have acquired the ability to invade and infiltrate other tissues surrounding the primary location [1]. According to the receptors the breast cancer cells express at their surface (traditionally identified using immunohistochemistry – IHC), they can be classified into different subgroups. The receptors of interest in this field are estrogen and progesterone receptors (ER, PR) and human epidermal growth factor receptor 2 (HER2) [1].

Diagnosis and treatment

The procedures used to diagnose breast cancer include ultrasonography, mammography, magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. Mammography is a widespread screening tool used in many high-income countries; it has the great benefit of being able to detect abnormal growth, even before any signs or symptoms of problems [4]. If the breast tissue is not dense, mammography can detect a mass before it can be felt in self-exams [4]. After a suspicious mass is identified, other exams and scans are performed to classify the cancer, and enable a more targeted treatment. These tests help determine and classify the tumor depending on how it looks (grade) or how it behaves (stage). For example, grade is defined by observing under a microscope whether the cancer cells are well or poorly differentiated depending on whether they look like the surrounding tissue or are not similar at all [1]. A very common staging system is the TNM system, where T stands for the size of the tumor, N refers to the number of lymph nodes that are affected with breast cancer, and M refers to metastasis [5]. The stage is determined according to the number attributed to each of the components of the TNM.

Stages vary from 0 to 4. Stage 0 indicates a carcinoma in situ; stages 1 to 3 depend on the size of the tumor and the number of lymph nodes affected; and stage 4 indicates that metastasis has occurred. Prognosis and survival rates generally worsen with higher cancer stages. The organs most commonly affected by breast cancer metastasis are the lungs, liver, bones, and brain [1].

Epidemiology

Worldwide

Breast cancer is the most common form of cancer among women worldwide (**Figure 3**) and the second most common overall, after lung cancer. Out of a total estimated 6,657,518 incident cases among women in 2012, 25% were cancers of the breast [6].

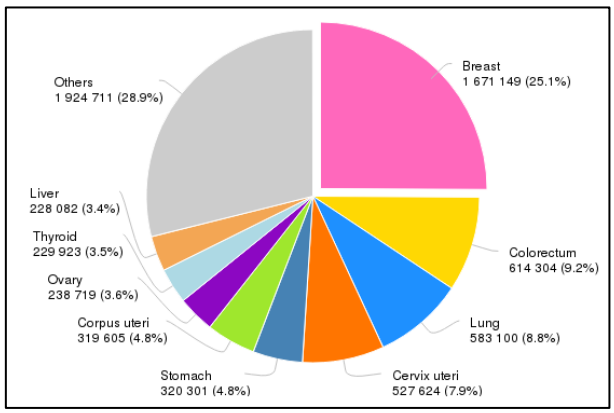

Figure 3. Estimated number of incident cases worldwide (top 10 cancer sites) in 2012 in women
Source: GLOBOCAN 2012, IARC 2016, available from: http://gco.iarc.fr/today, accessed 31/10/2016.

Incidence rates are higher in high-income countries (marked with a darker color in **Figure 4**) but are increasing more rapidly in low-income countries. Across the world, incidence rates can vary 5 to 10-fold [7].

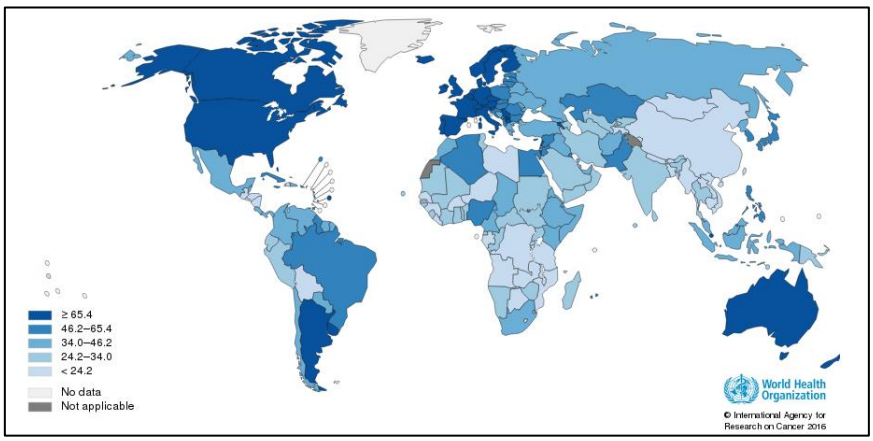

Figure 4. Age-standardized estimated rates of incident breast cancer worldwide in 2012 in women
Source: GLOBOCAN 2012, IARC 2016, available from: http://gco.iarc.fr/today, accessed 31/10/2016.

Differences in incidence rates among countries are thought to be attributable to environmental factors as genetic factors alone cannot explain all the variation. It is hypothesized that a westernized lifestyle, which is more common in high-income countries, and differences in reproductive patterns and hormone use contribute to the current picture [8]. Migrant studies contribute to the view that environmental factors are pivotal in breast cancer trends; populations migrating from countries with lower incidence rates adjust to the host country's rates after a few generations [9]. On the other hand, the more recent "westernization" of low-income countries could be the culprit for the rapid increase in incidence rates in these countries [10]. It is predicted that there will be approximately 2.7 million new cases by the year 2030, and 60% of these will occur in low-income countries, assuming that current trends in incidence rates are held constant [10].

Global differences in the mortality rates of breast cancer are shown in **Figure 5**, with the countries with the highest mortality rates marked with a darker color. This figure differs somewhat from **Figure 4** as the countries with higher incidence rates are not necessarily those with higher mortality rates. This is thought to be because of discrepancies in access to health care and in the success of cancer screening and improvements in diagnosis [7].

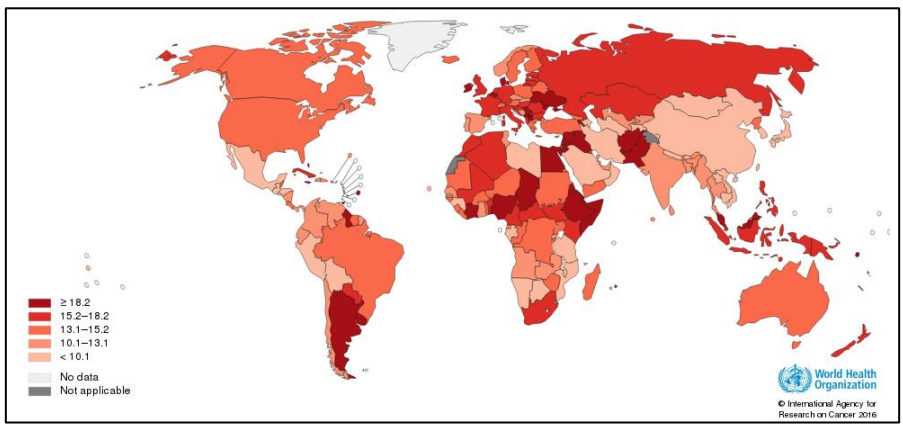

Figure 5. Age-standardized estimated rates of deaths due to breast cancer worldwide in 2012 in women
Source: GLOBOCAN 2012, IARC 2016, available from: http://gco.iarc.fr/today, accessed 31/10/2016.

In Sweden

The scenario is not very different in Sweden as the highest incidence rates of breast cancer are observed in northern and western Europe, among other regions [10]. Breast cancer accounts for one-third of all cancers in Sweden in women (**Figure 6**), and it is the second most common killer (more women die of lung cancer). In 2012, there were 6,625 estimated new cases of breast cancer (**Figure 6**).

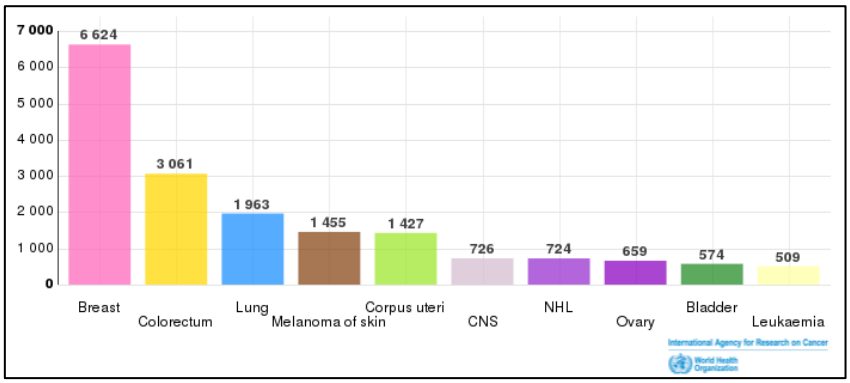

Figure 6. Estimated number of incident cases (top 10 cancer sites) in 2012, among women in Sweden
Source: GLOBOCAN 2012, IARC 2016, available from: http://gco.iarc.fr/today, accessed 31/10/2016.

Figure 7 shows the trends in breast cancer incidence and mortality rates in Sweden and in the southern region (where Skåne and thereby Malmö are included) between 1980 and 2015. While incidence has shown an increasing trend in the past decades, the opposite trend was observed for mortality. The southern region seems to follow the national trends; however in both 1990 and 2010, a higher proportion of women were diagnosed with breast cancer in the southern region than in the nation as a whole. It is predicted that 10% of Swedish women will be diagnosed with breast cancer at some point during their lifetime [11].

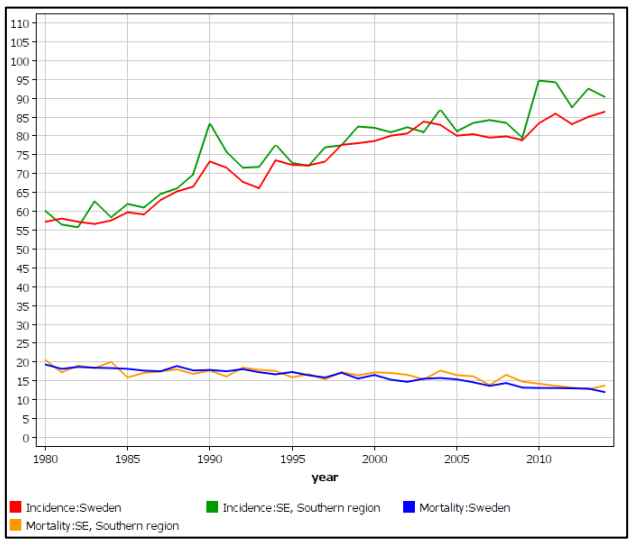

Figure 7. The number of new breast cancer cases and the number of deaths from breast cancer per 100,000 Swedish women in Sweden and in the Southern Region of Sweden between 1980 and 2014
Source: NORDCAN © Association of the Nordic Cancer Registries, IARC 2016, available from: http://www-dep.iarc.fr/NORDCAN, accessed 31/10/2016.

Risk factors for breast cancer

According to the WHO, any characteristic, attribute, or exposure that increases the likelihood of developing a disease can be considered a risk factor [12]. In light of the sufficient-cause model [13], it is likely that the risk of one individual developing a disease is the result of the sum of exposures to several risk factors to varying degrees. Moreover, many common risk factors (e.g., obesity, alcohol consumption, smoking status, etc.) add to the risk of several non-communicable diseases (NCDs).

Risk factors can be classified as non-modifiable or modifiable. The latter are interesting from the public health point of view because promoting changes in exposure to these factors could change incidence rates and consequentially reduce the public health burden.

Age

Age is a strong risk factor for many diseases. The incidence rate of breast cancer increases with age. As we can observe in **Figure 8**, there is a steep increase in the incidence until near the age of 50 years (at which most women have reached menopause) and then there is a certain plateau followed by a slow increase and then a decrease at approximately the age of 70 years. It is hypothesized that different biological mechanisms are responsible for the different pre- and postmenopausal curves [14]. This is further supported by evidence that shows that there is a great divergence in the breast cancer risk after menopause in different countries, suggesting that more external factors could be at play [7].

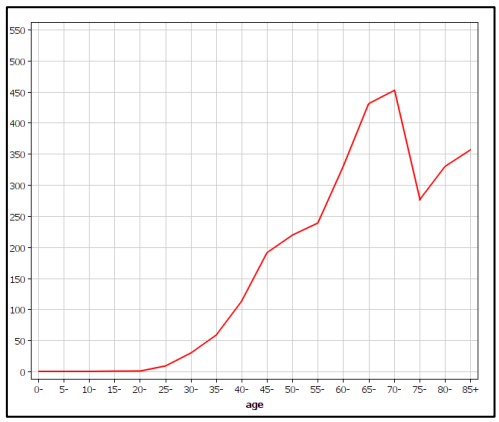

Figure 8. Age incidence curve for breast cancer in Swedish women in 2014
Source: NORDCAN © Association of the Nordic Cancer Registries, IARC 2016, available from: http://www-dep.iarc.fr/NORDCAN, accessed 31/10/2016.

It is also interesting to note the differences in incidence trends across different age categories in the past decades (**Figure 9**). Up until 1990, trends remained fairly stable upwards across all age categories, but there was a change throughout the 1990s with a sharp increase among women aged 60-69 years. In 2014, this was the age category with the highest incidence rates; women older than 80 years came in third [11]. Despite this phenomenon, mortality rates are higher in older age groups.

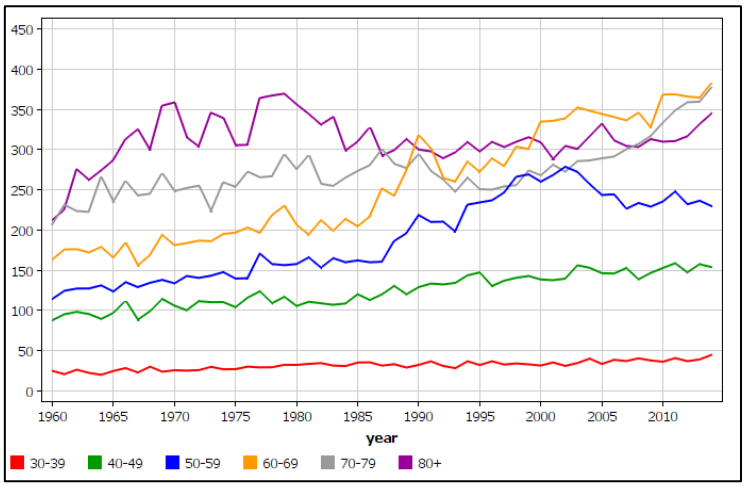

Figure 9. Incidence rate trends: the number of new cases of breast cancer per 100,000 Swedish women across different age categories at 10-year intervals between 1960-2014
Source: NORDCAN © Association of the Nordic Cancer Registries, IARC 2016, available from: http://www-dep.iarc.fr/NORDCAN, accessed 31/10/2016.

Ionizing radiation

Radiation is a carcinogen that interacts with deoxyribonucleic acid (DNA) to produce a range of mutations [15]. Evidence from atomic bomb survivors (in Hiroshima and Nagasaki) has shown the dire effects of whole-body exposure to high-dose radiation; the breast cancer incidence increases sharply, along with the incidence of many other types of cancer [16]. Interestingly, the mechanisms seem to differ when people face a nuclear spill, with exposure to low doses of whole-body radiation. For example, no evidence was found of an increased incidence of cancers among affected residents living near Chernobyl with the exception of increases in thyroid cancer in children [17].

The potential of ionizing radiation as a carcinogenic was documented early on, with early X-ray workers developing skin cancer, and with second cancers developing after subjects were treated with radiation against the first cancer [15]. The risk at lower levels of exposures has not yet been fully characterized and estimated, and a conservative approach is usually preferred. However, it is accepted that the possible negative effect of mammography screening (which

entails exposure to low-dose radiation) is greatly counteracted by its benefits (the ability to detect numerous cancers in early stages, thus improving prognosis) [18, 19].

Genetic factors

A list of all breast cancer susceptibility genes known to date is presented in a publication by Harris [20]. A high individual risk for developing hereditary breast cancer is conferred by germline mutations in high penetrance genes, such as BRCA1, BRCA2, ATM, TP53 and the PTEN [20]. Both non-selective population-based studies and family-based studies have estimated the probability of developing breast cancer if a woman carries a high-risk mutation; it varies from 37% in the first setting [21] to 70% in the second [22]. However, possibly because of their low allele frequency in the population, mutations in these genes account for up to 5-10% of all breast cancers [23].

On the other side of the scale we find low-penetrance susceptibility genes; low-risk genes that are more common in the population (i.e., their minor allele frequency –MAF – is much higher). Acting together with lifestyle risk factors and endogenous factors (hormones), these low-risk genes are more likely to make a greater contribution to breast cancer development [14].

Endogenous hormones

Steroid hormones (such as androgens) are produced in the adrenal gland (generally synthesized from cholesterol) in a process controlled by the gonadotropin-releasing hormone (GnRH) released by the hypothalamus. In short, GnRH stimulates the production of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) at the pituitary gland, and these are released into the general circulation. LH will in turn stimulate the production of androgens and FSH the expression of the aromatase enzyme, which will catalyze the conversion of the androgens into estrogen in the ovaries [24]. After menopause, when the gonads cease function, aromatase is primarily expressed in the adipose tissue [24]. There are three major forms of estrogen that naturally occur in women throughout their lifetime: estradiol (premenopausal), estriol (during pregnancy), and estrone (postmenopausal).

It is believed that lifetime exposure to sex steroid hormones plays a key role in breast cancer development, although the etiology of the disease is not yet fully understood. This role is hypothesized to be due to estrogen's role in stimulating the mitosis of mammary epithelial cells (a mechanism mediated by the estrogen receptor – ER) as estrogen is the major hormone responsible for reproductive system development in females [4]. Epidemiological studies have shown a convincing positive link between circulating concentration of estrogen and breast cancer risk in postmenopausal women [25].

Factors associated with the lifetime exposure of breast tissue to sex hormones are the age at menarche and the age at menopause; early menarche and late menopause are associated with an increased risk of breast cancer [7, 25]. Childbearing is also associated with a lower long-term risk and a higher number of children confer additional increased protection. The age at the first full-term pregnancy seems to play a role, independent of the number of full-term pregnancies: the protective benefit is greater for women whose first pregnancy occurs at a younger age, while a higher age is associated with increased risk [7]. Additionally, longer-term breastfeeding has been reported as a protective factor [26].

Exogenous hormones

After menopause, the use of exogenous hormones is associated with an increased breast cancer risk. Both the duration of exposure and the type of menopausal hormone therapy (MHT, formerly known as hormone replacement therapy (HRT)) are of interest. Current users of MHT are at higher risk of breast cancer compared with never-users, and this risk increases with longer duration of use. However, the risk seems to reduce to the level of the never-users 10 years after cessation. MHT can consist of estrogen alone, estrogen plus progesterone, or the combined use of estrogen and progestin (a synthetic hormone with effects similar to those of progesterone). The latter option as a MHT approach has been associated with a higher risk of breast cancer [27].

There is evidence of an increased risk of breast cancer with the current or recent use of combined oral contraceptives (OC), but this risk drops after the cessation of use. No sufficient evidence has been found for an association between the duration or type of OC use and breast cancer risk [27].

Breast changes

Breast density (i.e., more connective tissue than fat tissue) is associated with a 3 to 5-times increase in breast cancer risk [28, 29]. The mechanisms of this association do not seem to be dependent on hormonal factors. Higher risk might be due to the difficulties associated with distinguishing an abnormal mass from connective breast tissue during mammography screening.

Benign breast tumor (non-cancerous) might be associated with a risk of developing invasive breast cancer. The increased risk is smaller for the non-proliferative type of cancer (which is only significant when combined with a strong family history) but higher for the proliferative types, especially the type with atypia (called atypical hyperplasia), for which the risk can increase to approximately 3 times the average [30].

Both lobular and ductal carcinomas in situ (LCIS and DCIS) represent the non-invasive (pre-cancer) stage (stage 0 in the TNM system). Most DCIS and LCIS cases will not develop into invasive breast cancer; however, these cancers are associated with higher risk of invasive breast cancer, especially for certain types of DCIS and LCIS [31].

Anthropometry

Stature

Adult height has been positively associated with increased breast cancer risk in several epidemiological studies [26]. The mechanism behind this association is not yet understood, but a dose-response relationship is apparent. It is unlikely that tallness itself is the causal factor for breast cancer; rather, it is likely the factors that lead to and promote linear growth in childhood. Factors such as early life nutrition (including during growth spurts), the rate of sexual maturation, and altered hormone profiles are all plausible risk factors for increased risk [32].

Obesity

The effect of body fat on the risk of breast cancer differs according to menopausal status: it increases the risk in postmenopausal women, while it seems to be protective in premenopausal women. The mechanisms behind the decreased risk of breast cancer in premenopausal women are unclear; it is speculated that anovulation and abnormal hormonal profiles, which are more common in obese women, might be behind this protection [26]. However, for postmenopausal women, the evidence of a plausible mechanism is robust, and there is clear dose-response relationship [26]. Adipose tissue is the major producer of endogenous estrogen during menopause, and higher circulating levels of estrogen are observed in obese women. Most common measures of obesity are associated with postmenopausal breast cancer risk: the body mass index (BMI), weight, waist circumference, waist-to-hip ratio (WHR), and weight gain in adult life [33].

Lifestyle factors

A major effort by the World Cancer Research Fund (WCRF) was made in 2007, when the evidence of lifestyle risk factors and several cancers was thoroughly reviewed and summarized in a report [26]. The evidence was deemed convincing, probable or limited depending on the quality of the studies reviewed and how much is understood regarding the possible mechanisms at play. Further evidence regarding breast cancer was presented in 2010 [34], but a new complete and updated report is expected to be released in 2017.

Physical activity

Physical activity can be defined as any bodily movement involving the skeletal muscles. It can be categorized as occupational, household, transport or recreational, depending on whether it is performed at work, at home, when travelling between work and home, or during leisure time. Subjective (questionnaires) or objective measures (such as pedometers or accelerometers) can be used to measure the frequency, intensity and the duration of the activities.

There is probable evidence from prospective studies of a protective association between physical activity (generally leisure time physical activity – PA) and postmenopausal breast cancer. Additionally, there seems to be a dose-response association. A few possible interrelated mechanisms have been implicated: the positive impact of PA in decreasing body fatness, the effects on endogenous hormone metabolism, improved insulin and glucose profiles, and the possible positive impact on the immune system [35]. A modest but significant effect on reducing circulating sex hormones has also been highlighted [36].

Alcohol consumption

The link between alcohol consumption and the risk of breast cancer (both pre- and postmenopausal) is clear and consistent across case-control and cohort studies [26]. This appears to be a linear dose-response relationship, and no threshold has yet been identified. There are several proposed mechanisms through which alcohol increases the breast cancer risk: disturbances in the estrogen pathways that affect hormone levels and the receptors sensitive to these hormones; the promotion of oxidative stress and damage; the induction of mutagenesis by acetaldehyde; and effects on the one-carbon metabolism resulting from effects on the folate pathways [37].

Smoking

It is plausible that exposure to tobacco smoke increases one's risk of several cancers due to the active carcinogenic substances in cigarettes. Evidence linking smoking to breast cancer risk is, however, limited. Only in recent years has a suggestive risk been described [38].

Diet

The human diet has the potential to contain both anti- and pro-carcinogenic chemicals [39]. After Doll and Peto's publication in 1981, which estimated that diet could be involved in 10 to 70% of all cancers in the USA, it became imperative to investigate diet [40]. However, despite many efforts to clarify associations between diet and cancer in recent decades, the findings on the role of diet in the development of breast cancer remain inconsistent and inconclusive. Doll and Peto's estimation was revised fifteen years later by Willett, and the

results are still of great significance: between 20 and 42% of cancer deaths could probably be avoided by dietary changes [41].

Fat intake has been extensively investigated since observational studies indicated that breast cancer rates are higher in countries where fat intake is also high [42]. However, evidence regarding the association between total fat intake and the breast cancer risk has been deemed only limited or suggestive by the WCRF [26]. In fact, a major pooled analysis study did not show any significant associations between fat intake and breast cancer risk [43]. The associations seem to differ according to the type of fat consumed: some studies have reported an increased breast cancer risk with higher consumption of certain unsaturated fatty acids, while omega-3 polyunsaturated fatty acids (PUFAs) seem to be protective [14].

Natural sources of antioxidants and phytoestrogens, such as fruits and vegetables, have also been investigated for their anti-carcinogenic potential, but the results are not consistent [44]. Fiber represents another component of plant foods with the potential to both decrease circulating estrogen and help with weight reduction and thus protect against cancer development. Many reports do not show any significant association between fiber intake and postmenopausal breast cancer, whereas reports from the Malmö Diet and Cancer (MDC) cohort have shown a protective association [45, 46].

Other dietary factors that have been examined with no clear results regarding the development of breast cancer are meat (processed and red meat), fish consumption, milk and dairy products, soy, glycemic index, calcium, selenium, and vitamin D [34]. The inconsistent associations between diet and cancer are thought to be due partly to measurement errors in dietary assessment.

Lately, greater emphasis has been placed on investigating dietary patterns in relation to breast cancer with the aim of moving beyond the reductionist approach of single nutrients' effects on health outcomes. Evidence suggests that some dietary patterns may be associated with breast cancer risk; dietary patterns labeled "prudent/healthy" are associated with a lower risk, while "westernized/unhealthy" patterns are associated with an increased risk [47].

2.2. Low-Grade Inflammation and Oxidative Stress

> *"If genetic damage is the match that lights the fire, inflammation may provide the fuel that feeds the flames."*
>
> *- Fran Balkwill [48]*

In homeostasis, processes such as inflammation and oxidative stress result from regular actions of the metabolism. It is therefore important to understand what happens when an imbalance occurs.

Reactive species of oxygen and metabolism

In oxidation-reduction reactions (commonly known as redox), an exchange of electrons (negatively charged particle) occurs between two atoms: one donates one electron (and thus is oxidized), while one gains an electron (and is then reduced). Atomic molecules with unpaired electrons are (free) radicals, not to be confused with "ions", in which there is an imbalance between negative (electrons) and positive (protons) charges in the same molecule. Radicals can take any type of charge (positive, neutral or negative), although they are most often negatively charged or neutral. Because electrons tend to exist in pairs, coupled in an orbital with opposite directional spins, radicals with an unpaired electron are very unstable molecules. They tend to capture electrons from nearby molecules, thereby destabilizing them (in a redox reaction). This will successively occur in a chain of oxidation reactions, until the free radical encounters a molecule capable of modifying its electron spin or forms a less unstable molecule; that is, the last radical formed will not have sufficient energy to continue the propagation [49].

In biological systems, the most important radicals are the oxygen radicals, commonly known as reactive oxygen species (ROS), which are potent oxidants. Not all radicals are ROS, and vice versa. Examples of ROS are $O_2^{\bullet -}$ (superoxide anion radical), $^{\bullet}OH$ (hydroxyl radical), H_2O_2 (hydrogen peroxide), and diverse other peroxides, in which the symbol $^{\bullet}$ denotes an electron with unpaired spin. However, H_2O_2 is not a radical; along with the oxygen singlet ($^1O_2$), it is considered a non-radical ROS [50]. Additional radicals (or reactive species) that are not ROS include carbonyl species, reactive nitrogen species (RNS, which include nitric oxide – NO), and others that are closely related to the homeostasis of ROS [49].

ROS are produced under normal circumstances as a byproduct of cellular respiration – the process of utilizing oxygen to produce energy specifically from the electron transport chain [51]. ROS also have important roles in cell signaling,

and they can be synthesized by phagocytic cells such as neutrophils and macrophages [52]. The majority of ROS production, however, occurs in the mitochondria [53]. In short, oxygen undergoes a stepwise addition of electrons until it is reduced to water, creating several ROS as intermediates. This process is believed to not be 100% efficient, and some ROS might "leak out"; of these, $^{\bullet}OH$ (resulting from a Fenton reaction with H_2O_2) is the most reactive oxidant form. In the absence of a hydrogen ion and an electron (to form a water molecule), the hydroxyl radical can attack many other biological molecules [54]. Although the hydroxyl radical is an extremely reactive molecule (and thus short-lived), the hydrogen peroxide can travel to the nucleus of a cell, producing greater damage [51]. Another non-radical but highly reactive ROS is the oxygen singlet ($^1O_2$), which is the result of photochemical reactions. The major site of production is the cytoplasm of skin cells via ultraviolet irradiation. The oxygen singlet can also travel to the nucleus and cause damage, unless it meets a scavenger molecule first [51].

Among the most common harmful effects of the ROS are DNA damage; the oxidation of polyunsaturated fatty acids (PUFAs) in lipids (also known as lipid peroxidation); the oxidation of amino acids; and the oxidation and subsequent deactivation of co-factors of specific enzymes [54].

Oxidants and antioxidants

Molecules that can counteract the effect of oxidants such as ROS are called antioxidants. These can be either endogenous or exogenous, depending on their origin. They are usually neutral molecules that can donate an electron without becoming reactive.

Exogenous oxidants

Many oxidants are provided by the environment. Air pollutants are one source of certain oxidants, such as some components of smog or ozone [55]. Tobacco is another major source of biologically active substances that are powerful oxidants, such as the nitrogen dioxide ($NO_2^{\bullet}$) [56]. Another major source of ROS formation is the metabolism of alcohol (ethanol), which is hypothesized to affect mainly the liver and to be associated with alcohol hepatitis [57, 58].

Endogenous antioxidants

Two enzymes are responsible for the chain of reactions that transform oxygen into water: superoxide dismutase (SOD) and catalase. They are located in and around the mitochondria and are important in all cells that are exposed to oxygen. The catalase reaction prevents the Fenton reaction from occurring. At the same time, the glutathione peroxidase enzyme catalyzes a reaction between glutathione (GSH), a main intracellular antioxidant, and the hydroxyl radicals [51].

Exogenous antioxidants

Several vitamins and minerals are considered antioxidants because of their roles in metabolism. Examples are vitamin C (ascorbic acid), vitamin E (tocopherol), carotenoids, and selenium. Some are water soluble (vitamin C), while others are lipid-soluble (vitamin E). The food sources of these antioxidants are diverse: fruits, berries, and vegetables for vitamin C; dietary fats such as margarines and vegetable oils, and meats, fish, eggs, and fruit and vegetables for vitamin E; root vegetables for carotenoids; and grains (varying amounts according to the soil in which they are grown) for selenium.

Oxidative stress

Oxidative stress results from an imbalance between oxidative substances and antioxidants, with negative health consequences. There are several reasons for this imbalance: increased production of ROS; reduced reserves of existing antioxidants; decreased production of antioxidants; or a combination of these factors [49].

All individuals have a stationary level of ROS that oscillates within the normal range (depending on concerted production and elimination actions). Acute oxidation (provoked by any type of agent) can sharply accentuate the production of ROS and can lead to acute oxidative stress, if antioxidant systems are able to return the levels to the stationary levels. However, when this is not possible, higher levels or ROS will be present in an organism destabilizing the homeostasis and inducing many cellular alterations. This is the state of chronic oxidative stress [49].

Several factors are thought to affect oxidative stress; for example, physical activity (PA) is thought to acutely increase the production of ROS. They are important in the redox signaling pathways that result in processes needed for muscle adaptation, as mediators of inflammation after strenuous exercise, and for upregulating the antioxidant system [59, 60]. Obesity is associated with increased levels of oxidative stress, and it is suggested that ROS are the mediators for such deleterious effects as increased inflammation and insulin resistance [61, 62]. Higher energy intake might also be associated with higher ROS production, as suggested by studies showing lower levels of oxidative DNA damage with energy restriction [63]. Finally, chronic inflammatory diseases such as rheumatoid arthritis and cancer are associated with higher oxidative stress [64]. The extent of the damage caused by oxidative stress depends on the ability of the attacked cells to overcome the challenges.

Lipid peroxidation

All cell membranes are composed of phospholipids, along with proteins, cholesterol and vitamin E. Thus, the likely targets of the ˙OH molecule (or any other radical) are the phospholipids; however, their sensitivity varies with the number of bonds in the lipid residue [65]. Polyunsaturated fatty acids (PUFAs) are the most sensitive to a radical attack. In short, when a radical is formed close to the membrane, it attacks the PUFA residues of a phospholipid, forming a lipid radical. This, in turn, will form a peroxyl radical after reacting with oxygen. Finally, the peroxyl radical may react with the side chains to form lipid hydroperoxides and new lipid radicals, thus propagating this reaction further. Lipid hydroperoxides can accumulate in the membrane, causing it to lose function or degrade until it collapses. Lipid peroxidation is therefore the chain of reactions initiated by an attack by hydroxyl radicals [66].

Oxidation of low-density lipoproteins

The particles responsible for the transportation of fat (triacylglycerol – TAG – and cholesterol) in the blood stream are the lipoproteins. The many types of lipoproteins differ from one another in size, density (the fat-to-protein ratio), and in what type of fat they carry: low-density lipoproteins (LDL) carry more cholesterol (more fat to protein), whereas high-density lipoproteins (HDL) carry less fat and more protein. Others, such as chylomicrons, very low-density lipoprotein (VLDL) and IDL (intermediate-density lipoproteins, which are remnants of chylomicrons) are also present and have an important role in fat transport. Lipoproteins can also be characterized by the apolipoproteins (Apo) they express; LDL contains Apo B100, whereas HDL mainly contains Apo A1. In simplified terms, the main function of LDL is to transport cholesterol from the liver to cells, where it is used to produce several important components, such as vitamin D and steroid hormones. Subsequently, HDL will pick up the cholesterol leftovers from the tissues and bring them back to the liver [65].

These lipoproteins are sensitive to oxidation (from oxidative stress, not to be confused with β-oxidation, the process through which acetyl-coA is released from free fatty acids, to enter the Krebs cycle – TCA). LDL is especially susceptible to oxidative changes [67]. The oxidation of LDL molecules is a complex process that includes the oxidation of both protein and lipid parts and the formation of complex products. Very damaged and modified LDL attracts macrophages, which will scavenge and degrade these particles [50]. The formation of foam cells after the oxidation of LDL and its engulfment by macrophages on the inner artery walls is the current explanation for the beginning of the atherosclerotic process [65].

Low-grade inflammation

Inflammation and oxidative stress are closely related, as one can be induced by the other. In many pathologic conditions, both processes occur simultaneously. In fact, it is believed that they are the key mechanisms linking the major non-communicable diseases (NCD).

Chronic inflammation can be described as a prolonged state of inflammation, in which tissue injury and repair attempts coexist in varying combinations. The classic signs of inflammation are heat, pain, redness, and swelling; another feature, loss of function, was added later. Due to its generic nature (i.e., not specific to a pathogen), it is considered a mechanism of innate immunity. A system's acute response to harmful substances (e.g., infection by pathogens) that includes the recruitment of leukocytes (also known as white blood cells – WBC) from the blood to the site of action (e.g., an injured tissue) is an acute inflammatory response. This primary response is fundamental for protection against harmful substances from the environment. However, when this inflammatory process persists, it leads to a shift in the cells present at the site and has consequences for the tissue. This is known as chronic inflammation or low-grade inflammation.

The major players in inflammation

In generic terms, one of the major functions of the innate system is the recruitment of cells to the site of an infection or injury, usually through mediators such as cytokines, with the aim of terminating that menace. It is thus important to understand the cascade of reactions and the key players involved in an inflammatory reaction. Specialized cells, present in all tissues, initiate the process of acute inflammation. They recognize a pathogen by its distinguishing receptors (which differ from those of the host cells) and release inflammatory mediators (these are responsible for the classic signs of inflammation) to stimulate and direct an adaptive response. Among these cells are mast cells, phagocytes (such as macrophages and dendritic cells), basophils and eosinophils, and natural kill cells.

In short, in reaction to an injury (or invasion of a pathogen), vasodilation occurs and is accompanied by vascular permeability, resulting in edema. This is necessary to transport leukocytes to the site via extravasation (passing through the capillary walls). These leukocytes will in turn phagocytose the pathogen and release molecular mediators, such as cytokines, that contribute to the inflammatory response. All of these steps are highly regulated [68].

Cells

All white blood cells (WBCs) are produced in the bone marrow, through a process called hematopoiesis. All cells derived from a multi-potential hematopoietic stem cell (hemocytoblast) can be divided into groups. At first, hemocytoblasts either differentiate into the common lymphoid progenitor cell or the common myeloid progenitor cell. Lymphocytes (which can be further divided into B cells, T cells, and natural killer cells) derive from the lymphoid progenitor, while all the others (i.e., megakaryocytes, erythrocytes, mast cells, and myeloblasts) derive from the myeloid progenitor. Myeloblasts further differentiate into basophils, neutrophils and eosinophils (granular cells) or into monocytes (agranular cells). When monocytes leave the blood stream and enter the tissue, they differentiate into macrophages [68]. The production of a specific cell line is tightly regulated in healthy humans by several stimuli, such as growth factors and cytokines.

All WBC cells have different constitutions, functions and lifetimes. For instance, monocytes (approximately 5% of WBC in adults) and neutrophils (62%) are the only WBCs with phagocytic capacity (along with mast cells), but while neutrophils are specialized in bacteria and fungi, monocytes migrate from the blood stream and differentiate into dendritic cells and macrophages that reside in specific tissues. Both last from a few hours to days. Eosinophils and basophils (approximately 2% and 0.5% of WBCs, respectively) are responsible for modulating allergic inflammatory responses and releasing histamine. They last between 2 weeks (eosinophils) to just a few days or hours (basophils). Finally, the lymphocytes are responsible for the adaptive immune response (and represent approximately 30% of WBCs). Their actions vary between B cells and T cells, and they last for years (memory cells) or weeks (all others).

Cytokines

Injured and affected cells produce eicosanoids and cytokines. The role of eicosanoids is, among other things, to signal immune responses; that is, to mediate local symptoms of inflammation, such as vasodilatation, pain and fever. They derive from fatty acids in the cell membrane, specifically from the oxidation (enzymatic or non-enzymatic) of arachidonic acid (or another PUFA). A well-known family of eicosanoids is the prostaglandins, which are produced with the help of cyclooxygenases (COX-1 and COX-2). The COX-1 enzyme is expressed at a constant level in all cells, whereas COX-2 is absent from most tissues but overexpressed in tumor cells. In addition to their pro-inflammatory role, prostaglandins are known to not only stimulate cell proliferation and induce the mitogenesis of mammary epithelial cells but also to induce the expression of aromatase (the enzyme responsible for estrogen production) [69, 70]. Nonsteroidal anti-inflammatory drugs (NSAIDs) target and inhibit the activity of COX-1 and COX-2, leading to anti-inflammatory, antipyretic and analgesic effects [71, 72];

these drugs have been epidemiologically associated with a reduced risk of breast cancer [73, 74].

Cytokines, on the other hand, are a group of low-molecular-weight proteins with the function of cell signaling: they bind to specific receptors and trigger the signal transduction pathways within. They are also produced by WBCs and other cells. The cytokines include interleukins (IL), which are responsible for the communication between WBCs; chemokines, which promote chemotaxis (the movement of an organism in response to a chemical stimulus); interferons, which have anti-viral effects; and tumor necrosis factors (TNF) [75]. Immune cells are recruited to the site of infection by these cytokines, which also promote the healing of the damaged tissue. Cytokines can also be produced by immune cells to recruit more cells and promote the inflammatory state. Additionally, cytokines can have autocrine, paracrine or endocrine actions (within the cells where they are produced, in a nearby cell, or in a distant cell), and their effects can be several: pleiotropic, redundant, synergistic and even antagonistic [75]. They can also participate in cascade induction. In the four studies presented in this thesis, several interleukins (IL-1β, IL-6, and IL-8) and tumor necrosis factor alpha (TNF-α) were investigated, and their main functions are described in **Table 1**.

Table 1. Cytokines: site of production and main effects

Cytokines	Secreted by	Targets and main effects
IL-1	Monocytes, macrophages, endothelial cells, and epithelial cells	Vasculature (inflammation); hypothalamus (fever); liver (induction of acute phase proteins)
IL-6	Macrophages and endothelial cells	Liver (induction of acute phase proteins); influences adaptive immunity (proliferation and antibody secretion of B cell lineage); has both anti- and pro-inflammatory effects
IL-8	Macrophages, epithelial cells and endothelial cells	Chemotaxis (mainly for neutrophils and other granulocytes); induces phagocytosis; promotes angiogenesis
TNF-α	Macrophages	Vasculature (inflammation); liver (induction of acute phase proteins); causes loss of muscle and body fat (cachexia); induces death in many cell types; activates neutrophils

Table adapted from Kindt, T. J., Kuby Immunology, sixth edition, 2007[68]

Oxidative stress and low-grade inflammation

Chronic low-grade inflammation and oxidative stress are closely related pathophysiological processes that influence each another [76]. Leukocytes fighting an invader increase their oxidative metabolism, thus increasing the production of ROS [76]. In fact, ROS are produced by neutrophils and macrophages as a mechanism to kill tumor cells via ROS-induced apoptosis. ROS play a role in maintaining the homeostatic functions of the macrophages, especially in macrophage polarization. Macrophage polarization refers to the macrophage's activation, i.e., whether it is classically activated (M1) or alternatively activated

(M2). M1-activated macrophages are more pro-inflammatory (secreting high amounts of pro-inflammatory cytokines, such as IL-1β, and TNF-α), whereas M2-activated macrophages are involved in inflammation resolution. An imbalance of the normally balanced M1/M2 ratio is thought to lead to disease [77]. It has also been demonstrated that ox-LDL molecules stimulate the intracellular production of ROS in macrophages [52].

Low-grade inflammation and cancer

The possible link between inflammation and cancer is not new; in fact, it was proposed approximately a century ago by Virchow. Virchow first noted the presence of macrophages at the tumor site [48]. Approximately 20% of all cancers are estimated to be associated with chronic inflammation and infection [78]. However, inflammation is not only the "fertile field" that enables cancer progression; it also predisposes the individual to certain types of cancer [79]. Inflammation at different sites is a strong risk factor for many cancers; examples include the link between bronchitis and lung cancer; gastritis and gastric cancer; pancreatitis and pancreatic cancer, and others [80]. Additionally, chronic inflammatory diseases increase the risk of developing cancer; rheumatoid arthritis, and inflammatory bowel disease [64]. The link between chronic inflammation and several steps of tumorigenesis (such as cell transformation, promotion, proliferation, survival, angiogenesis, invasion and metastasis) is also widely accepted [80, 81]. In fact, tumor-promoting inflammation is one of the hallmarks of cancer (**Figure 2**). Many cytokines released by inflammatory cells have a pro-tumor action. One such example is TNF-α and its role, along with the interleukins (e.g., IL-1), in many steps of the carcinogenesis [80]. These cytokines can be produced by both inflammation cells and tumor cells, further inducing them. Cancer cells produce more cytokines to control their microenvironment [82]. These signals (pro-inflammatory cytokines) released by tumor cells attract macrophages and maintain a positive feedback loop. In other words, cancer causes an increase in inflammation levels, which is accompanied by the increased production of radicals (because tumor cells have higher energy demands than normal cells), which in turn causes further DNA damage. In addition, considerable evidence pointing to the reduced risk of several cancers among long-term NSAID users further highlights inflammation as an important mechanism in the causal pathway of cancer [83].

Using circulating C-reactive protein (CRP) as a marker for low-grade inflammation, a literature review and meta-analysis explored the association between inflammation and postmenopausal breast cancer. An increased breast cancer risk of 7% was found for each doubling in the concentration of CRP [84].

Obesity

Obesity fosters both oxidation and inflammatory states, which are the keys to many diseases associated with abdominal fat. Subclinical inflammation (frequently undetectable), which is often mediated by obesity-induced inflammation, may be very important in cancer risk [81]. The adipose tissue, specifically the adipocytes, is considered a secretory organ where many adipokines (cytokines produced by the adipocytes) are produced. Examples of adipokines are leptin (which is a hormone), adiponectin, IL-6, and TNF-α, among many others [85]. Plasma concentrations of many pro-inflammatory cytokines are also positively associated with BMI [86]. Furthermore, oxidative stress increases with obesity [87].

Biomarkers

Any biological substance, structure or process that can be measured in the human body is a biomarker (i.e., biological marker) that may predict or influence the incidence or outcome of disease. Biomarkers can be classified into markers of exposure, effect and susceptibility [88]. Ideally, biomarkers should be valid and reliable (reproducible).

Validity and reproducibility of measurements

To accurately estimate the association between any biomarker and disease, valid and reliable measures are needed for exposures, covariates and outcomes [89]. The validity (also known as the accuracy) of a measurement tool is related to how well it measures the "reality", or how well it measures what is supposed to measure; that is, the degree to which it comes close to the true value or concentration of a biomarker. In other words, when a measurement is highly accurate there is less chance of systematic error (or bias). In comparison, reliability (also known as precision or reproducibility) is related to the ability to obtain similar answers for repeated measures of same reality. In this domain, random error can be observed, which in theory can be accounted for by increasing the sample size. Because the error is random and not systematic, the higher the number of measurements, the more closely the results will reflect the truth.

Using these two parameters, there are four possible theoretical scenarios for measurements: low validity and low reproducibility; high validity and high reproducibility; low validity and high reproducibility; and high validity and low reproducibility. Thus, a reliable measure might not always be valid and vice versa. It is of great importance to estimate the validity and reproducibility of measures in epidemiology to account for different types of errors.

Variation in biomarkers

There are several sources of biomarker variability: 1) inter-subject; 2) intra-subject (over time); 3) biological sampling; and 4) laboratory variation [89]. Inter-subject variability can result from age, gender, weight, diet or any other characteristics that can influence biomarker concentrations. Variations within the same individual can reflect the individual's health or nutritional status (changes in diet), the biomarker's intrinsic variability (e.g., circadian variations), and variation in exposure to influencing factors (for example, seasonal variations) [13]. Errors related to the collection of the biological specimen, processing and storage are another source of variation [90]. Laboratory variations can fall into two error types: errors within batches and errors between batches (intra-assay and inter-assay variability) [89].

Often, the sensitivity and specificity of biomarkers are measured to establish cut-offs and maximize the predictive value. For this purpose, receiver operating characteristic (ROC) curves are usually used. However, to establish a biomarker as a risk factor for a disease with relevance for a clinical setting, further steps need to be taken; namely, clinical validity and clinical utility must be established [91].

Measuring variation/reliability

In epidemiology, it is not always possible to use the best biomarker available (i.e., the most accurate), often because of budget constraints or because one simply does not exist. It is important to be able to rank individuals according to their risk of biomarker exposure, despite not always being able to measure the biomarkers' "true" value. Epidemiology is usually interested in estimating dose-response associations, given a satisfactory within-individual variation [92, 93].

There are two commonly used approaches to determining the repeated measures reliability of continuous exposures (such as biomarkers): intraclass correlation coefficients (ICCs) and Bland & Altman plots [94, 95]. ICCs were used in **Paper I**, to estimate the variation and reproducibility of several biomarkers of interest for this thesis. The ICC can be defined as the between-person variance divided by the total variance (i.e., the between- and within-subject variability and measurement error); where a coefficient of 1.0 represents exact agreement between 2 or more measures for each subject. The main difference between ICCs and Pearson's correlation coefficients is that instead of measuring total agreement, when the value of the latter is 1.0, it indicates that one measure is a linear combination of the other [89]. The ICC is a recommended measure to predict the reliability and population heterogeneity of biomarkers [96].

There are no established cut-offs defining a good ICC; just as a relative validity coefficient of 0.45 might be considered low/poor in one context and reasonable in another. It is, however, agreed that an ICC<0.40 indicates poor reproducibility, an

ICC between 0.40 and 0.75 indicates fair to good reproducibility, and an ICC≥0.75 indicates excellent reproducibility [97]. ICC estimation can also be a useful tool for predicting the attenuation level of observed relative risks (RRs) and the subsequent correction of RRs.

Biomarkers of oxidative stress

Measuring oxidative stress directly in a living organism is a great challenge. In this thesis, the ox-LDL was measured to approximate the level of oxidative stress in individuals. Ox-LDL levels are associated with oxidative stress [98], a finding that has implications not just for the study of cardiovascular disease (CVD) but also for cancer, specifically breast cancer [99].

Biomarkers of low-grade inflammation

There is no clear definition of low-grade inflammation. To be able to investigate the general systemic levels of inflammation, several cellular and molecular biomarkers rather than one single biomarker were used in this thesis. These biomarkers have clear roles during inflammatory periods and have been implicated in several stages of carcinogenesis [80, 100]. Furthermore, they have been associated with breast cancer [101, 102].

2.3. Diet

"Let food be thy medicine and medicine be thy food."

- Hippocrates

Diet as an exposure

The value of diet in health has been long recognized. Early studies in the 18th century focused on deficiencies that would later translate into disease. Lind conducted the first controlled trial-type investigation in sailors and observed that fresh citrus fruits could prevent scurvy. This effect was later found to be the result of addressing a vitamin C deficiency in their diet [103]. Such issues (i.e., deficiency status) contrast with the issues that concern nutrition epidemiologists today. The focus has shifted towards the major diseases currently affecting Western societies, such as cardiovascular diseases, cancer, diabetes, etc. This focus poses a bigger challenge because most of these diseases have multiple causes and because the potential role of nutrition together with other key players, such as genetics, physical activity, tobacco and alcohol use, is not well understood. Other major difficulties are the unknown period of latency, and the relevant period of exposure for most of these diseases.

For many reasons, diet is a unique exposure. On a macro level, food choices relate to other factors including food availability, cultural practices, personal characteristics, lifestyle, and socioeconomic status. Many dimensions (social, behavioral, personal, and biological) are at play [104].

On an individual basis, diet is a complex exposure. Food is a necessity, and everyone is exposed to it throughout their life to varying degrees in a variable manner; we eat differently during different periods of our lives. People are exposed to food in a repeated manner, at every feeding episode. Diet is rarely a dichotomous variable of "exposed" versus "non-exposed"; rather, it has a continuous nature. Problems with temporality can arise when undiagnosed disease processes affect diet (both consumption and metabolism). Another challenge is that when dietary intake levels are too homogenous within a population (i.e., in cases of low within-population variability), because associations between diet and disease might then be hard to detect. Additionally, different levels of exposure might have different health effects in a population (i.e., genetic variability). Lastly, it may be difficult to disentangle the effects of specific nutrient intakes because a complex set of interrelated exposures, such as nutrient-nutrient interactions, might exist, along with variations in metabolism and effects caused by the food matrix [105].

Sources of variation and measurement error

True dietary intake can be characterized by day-to-day variation superimposed on a consistent underlying pattern at the individual level [103]. Nutritional epidemiology is interested in the long-term dietary intake, and to examine that, it is important to understand the day-to-day variation. The nature of variation in diet can be either systematic or random. Examples of factors that can contribute to systematic variation are the day of the week or the season. Ecological and cultural factors determine the magnitude of these effects. Random variation includes the true variation in food intake. The level of random variation differs across nutrients; it is lower for total energy intake (TEI) and slightly higher for macronutrients, while micronutrients have the greatest day-to-day variation. Both intra-individual (i.e., the variable intake for a specific person) and inter-individual (i.e., different people making different food choices due to lifestyle, cultural and sex-specific factors) variations play a role in the true variation in food intake [103].

Measurement error is also a source of random variation in dietary intake. It includes both systematic and random error and can occur at two levels: between people and within a person. Therefore, four types of measurement error may be encountered: random within-person error, systematic within-person error, random between-people error, and systematic between-people error. Typically, it is not possible to distinguish a random measurement error from the random true day-to-day variation in dietary intake of one individual, but both would benefit from increasing the number of repeated measurements. If we consider the long-term average of dietary intake the true dietary intake for one individual, then true variation over time could be considered "error". However, more measurements or more individuals would not solve a systematic error. Reproducibility studies (i.e., replicated measurements taken in the same sample of subjects) can be used to estimate random within-person error, whereas validation or calibration studies can be used to quantify systematic error [103].

Measuring diet

Because of its intrinsic characteristics, diet is probably one of the most difficult exposures to measure with high precision. This may be the biggest limitation of research in nutritional epidemiology. There are several methods for measuring dietary intake; the most common are food frequency questionnaires (FFQ), food diaries, and 24-hour recalls. All dietary assessment methods are associated with some amount of measurement error. A brief description of the dietary assessment methods most commonly used in nutritional epidemiology is provided in **Table 2**. In general terms, recalls and dietary records are based on the foods actually consumed, whereas FFQs are based on the individual's perceptions of his or her usual intake.

It is interesting to note, however, that regardless of how well dietary intake is measured, there is still a "black box" of unknowns between reported intake and disease. A few questions cannot be answered just by using conventional dietary assessment methods: how much of the estimated intake is the true intake, taking measurement errors into account; how much of that is the real absorbed amount, accounting for interactions with gut microbiota; how much of that amount was needed to shift metabolism, taking genetic variability into account; and how much of that in turn led to an altered risk factor and subsequently to a preclinical disease state.

Table 2. Different dietary assessment methods

	Recalls (24-h recall)	**Records (food diaries)**	**FFQ**
Data collected	Actual intake (over 24 h)	Actual intake (over a certain period)	Usual intake estimate (usually over the past year)
Strengths	Detailed intake data; Small respondent burden	Detailed intake data; No interviewer required; No recall bias	Simple; Time-saving and cost-effective
Limitations	Possible recall bias; Skilled interviewer required; Possible interviewer bias; Time-consuming and expensive; Multiple days required for usual intake estimation; Possible change in diet if repeated measures are used	Large respondent burden; Time-consuming and expensive; Multiple days required for usual intake estimation; Possible change in diet if repeated measures are used	Specific to study groups and research aims; Closed-ended questionnaire; Low accuracy (recall bias); Requires validation of the developed questionnaires

Brief description of most commonly used dietary assessment methods in nutritional epidemiology, adapted from Shim, J.S., Dietary assessment methods in epidemiologic studies, Epidemiology and Health, 2014 [106]. FFQ, Food Frequency Questionnaires.

Validity and reproducibility of dietary assessment methods

In epidemiological investigations, the concept of validity can be divided into internal, relative, and external validity. External validity refers to the ability to generalize the results obtained to the source population, which is related to the representativeness of the study sample. Internal validity refers to the ability to trust the results that a study obtains, i.e., the ability to collect reliable data. To estimate the validity of a dietary assessment method (i.e., whether it is measuring what it was intended to measure), the "truth" would have to be known. This is almost impossible in nutritional epidemiology [107]. Therefore, the validity of one method is usually tested against a superior method (considered a "gold standard", but not without imperfections), and relative validity is estimated. For that purpose, correlations coefficients are typically used to estimate the ability to rank order individuals in terms of dietary exposures [103].

Reproducibility, on the other hand, is the ability to produce the same results using the same method under similar conditions. It refers to the consistency or concordance of the questionnaire measurements, when applied in the same population. However, it can be somewhat more difficult to design reproducibility studies for dietary measurements. A learning effect can occur when two measurements are performed a short time apart. However, if a long period is allowed between the two measurements, dietary habits may have changed. Seasonal variation in diet must also be considered. Thus, reproducibility represents both the questionnaire's performance and true changes in dietary habits. The difficulty of separating these two sources of variation is not worrisome because both will lead to the misclassification of long-term dietary intake [103].

Assessing biomarkers of dietary intake, such as recovery biomarkers, is considered an objective way to calibrate reported intakes. Unlike concentration biomarkers, recovery biomarkers are quantitatively related to dietary intake during a specific period and are therefore useful for estimating absolute intakes. Their major drawback is that only a few recovery biomarkers are available, and their assessment is expensive and cumbersome; therefore, they are not useable in large-scale epidemiological studies. Examples of recovery biomarkers are: doubly labeled water (DLW) for energy intake and 24 h-urinary nitrogen, sodium and potassium for protein, sodium and potassium intake, respectively [103].

Energy adjustment

A very important step when examining reported dietary intakes in relation to disease in epidemiology is adjustment for the reported total energy intake (TEI) [108]. There are several reasons why it is crucial to adjust for TEI.

The first reason is to control for confounding and to examine the effect of a specific dietary factor independent from TEI since total energy might be associated with the disease under study because of body size, PA level, or metabolic efficiency. TEI can also be associated with specific dietary factors; for example, because the absolute intake of nutrients contributes to TEI or because people who eat more will eventually eat more of all nutrients. Another important reason for TEI adjustment is to reduce measurement error and extraneous variation. Errors in the assessment of both nutrients/foods and of total energy are strongly correlated; thus, accounting for variation in TEI helps to reduce errors in the measurement of dietary factors [108]. There are 4 main approaches to addressing the influence of TEI on the analysis of nutrients: the nutrient density method, the standard method, the partition method, and the residual method [108]. Three of these models are interchangeable, and the nutrient density model is analogous, but not mathematically equivalent. Therefore, the choice of which approach to take is based on which coefficient should be used and in what unit the relative risks (RR) should be expressed.

The traditional approach used to account for TEI is the nutrient density model, which involves dividing the nutrient intake by the TEI and can be expressed as the percentage of energy (E %) or as grams (g)/1 Megajoule (MJ) (or g/1000 kilocalories – kcal). The standard model consists of adding TEI and the nutrient (expressed in g) in the same multivariable regression model. In this situation, the beta coefficient (β) of the nutrient could be interpreted as the effect of increasing one unit of the nutrient while holding TEI constant. However, the actual variation independent of TEI can be assessed using the residual method. The residual model uses regression analyses to account for confounding by TEI; nutrient and food intakes are regressed on TEI for each individual. The differences between the actual intake and the intake predicted by the TEI are represented by the residuals resulting from the regression analysis (which are now uncorrelated with TEI). Finally, the partition (decomposition) model separates calories from the exposure of interest and the remaining variables. In addition, when TEI is suspected of being associated with the outcome, it is added to all the models as a confounder [108].

Dietary patterns

Traditionally, diet has been represented in terms of nutrients, foods or food groups. However, in the past decade, the focus has shifted from heavy emphasis on a single-nutrient approach to the association between diet and disease to a more overall diet investigation characterized by dietary patterns (DP). These are considered now complementary approaches: single nutrient/food-focused studies aim to identify biological mechanisms, while DP approach examines the influence of the overall pattern.

Several limitations are associated with examining single nutrients or single foods. It is obvious that people eat a variety of foods in a mixed combination of nutrients. Many of these nutrients can act synergistically or interactively (together with other bioactive compounds in the food matrix), and such information can be lost with a single-nutrient approach. Furthermore, it is very difficult to depict separate effects because many nutrients are highly correlated, as they could be provided by the same food sources [109, 110]. This high correlation also poses statistical problems (known as multicollinearity) because risk estimates can be highly underestimated or overestimated when highly correlated nutrients are entered in the same model [111]. A second statistical problem arising from analyzing many nutrients and/or foods in the same model is the possibility of attaining statistical significance by chance [109]. One single nutrient might have an effect so small it cannot be detected, whereas a DP includes the cumulative effect of several nutrients/foods, which could be sufficiently large to detect an effect. Additionally, the effect of a specific nutrient might be confounded by a particular dietary pattern [109].

Finally, dietary recommendations based on DP are most likely easier to understand (and thus adhere to) than recommendations based on nutrients.

The potential limitations of dietary pattern analysis should also be addressed here. A true health effect that is primarily mediated by a single nutrient might be masked and diluted when using the overall pattern [103]. When associations with specific dietary patterns are found, it might be more difficult to translate a broad message into a simpler recommendation. Dietary patterns based on foods might be limited by the subjectivity of the researchers' decisions regarding combining foods. Finally, the characterization and naming of DPs are also subjective, and might not be comparable from one study to another [103].

Types of dietary pattern analysis

There are two main approaches to dietary patterns (DP); *a priori* (hypothesis-driven) and *a posteriori* (data driven). Among the different data-driven methods, the most commonly used are principal component analysis (PCA), factor analysis (FA), and cluster analysis (CA). These are used when there is no prior hypothesis regarding how diet is associated with a disease. A smaller set of variables results from an aggregated and reduced larger set of variables.

Factor analysis (FA) derives factor scores (scales) based on the correlations (linear combinations) between dietary variables, resulting in factor loadings for each variable. Individuals are then assigned factor scores, which are uncorrelated but not mutually exclusive (individuals are assigned a factor score for each derived factor). CA separates and aggregates individuals into different (mutually exclusive) clusters that are based on intake differences and thus represent behaviors shared by individuals [112, 113]. These exploratory approaches are exclusively based on the population they are derived from, which makes comparisons and reproducibility limited [112]. However, it is common to find a few dietary factors or clusters that are fairly reproducible across populations; "healthy", "alcohol", "traditional" and "sweets" [110]. A major drawback of these methods is the subjective nature of some decisions made by the researchers at many points in the process, especially if these are not clearly reported [112].

Dietary patterns that are constructed using different types of scores and that do not depend on a specific population are hypothesis-driven. This opens the possibility of greater reproducibility across populations. These types of DPs are intuitive and simple to compute. Index-based scores (IS) are the most common score-based *a priori* methods; such scores are assigned to each individual for the total diet, based on food/nutrient recommendations. Other examples are nutrient adequacy or density scores and food group patterning scores [112]. Several dietary indices have been reported: the Mediterranean Diet Score (probably the most widely recognized), the Diet Quality Index, the Healthy Eating Index, and the

Recommended Food Score, among others. Some drawbacks of IS are that summing equally weighted components implies that each component contributes equally to health; furthermore, some scores only dichotomize "consuming" versus "non-consuming" (not taking the full range of variability into account), and the chosen dietary or nutrition guidelines or recommendations may not be disease specific [112].

Reduced rank regression (RRR) has recently been proposed as a hybrid approach between *a priori* and *a posteriori*. It combines the exploratory analysis (letting the data do the talking) with some knowledge of the disease process (when choosing the response variables). RRR can be defined as a regression analysis that derives linear combinations of a set of predictors (such as nutrients or food intakes) that explain the maximum variation in a set of response variables (such as intermediate outcomes). Examples of response variables are nutrient intakes, biomarkers of intake or biomarkers of disease processes. One or more factors are extracted, depending on the number of response variables, and a score is assigned to each subject for each factor. This methodology has the potential to highlight pathways between dietary patterns and the disease outcome using a biomarker intermediate of disease. However, if the postulated intermediates do not represent the reality, a bias might be introduced [114].

Diet and inflammation

Nutrients

Energy restriction leading to weight loss is associated with decreased levels of mediators of inflammation such as CRP, IL-6, and TNF-α [115]. High-glucose and high-fat meals induce postprandial inflammation. This is exaggerated in obese people and in people with type 2 diabetes. Saturated fatty acids (SFA) and trans-monounsaturated fatty acids (MUFA) are considered pro-inflammatory whereas certain types of polyunsaturated fatty acids (namely omega (ω)-3 PUFA) are considered anti-inflammatory. Vitamin C, vitamin E and carotenoids, due to their anti-inflammatory role, decrease the circulating concentrations of biomarkers of inflammation [115].

Whole foods

Components of "healthy diets" are associated with lower inflammation. Despite the difficulties of comparing all studies of whole grains because of divergent classifications, whole grain intakes seem to be negatively associated with low-grade inflammation markers. However, more research is needed to understand what the potential contributors are (vitamins, minerals, dietary fiber, or phytochemicals). Fruits and vegetables are also a source of great interest, because

they are clear sources of many vitamins and minerals with possible anti-inflammatory effects. They are also rich in fiber and other compounds, such as polyphenols. Studies focusing on a single fruit or vegetable have yielded inconsistent results, but there is convincing evidence that higher overall consumption of fruits and vegetables seems to be associated with lower inflammation [115]. Higher frequencies of fish consumption have also been associated with lower levels of inflammation in several studies. The evidence is still controversial or lacking in other investigated components: for example, soya does not seem to be associated with lower inflammation despite the potential of the soybean (probably due to processing). The findings for nuts, tea, coffee and cocoa are also inconclusive [115]. Moderate alcohol consumption (of beer and wine), on the other hand, seems to be inversely associated with low-grade inflammation. Whether the responsible for this association is the alcohol or other compounds, such as phenolic compounds, remains to be determined.

Dietary patterns

Healthy eating patterns are associated with lower circulating concentrations of inflammatory markers. Several studies have investigated the benefits of adhering to a Mediterranean dietary pattern, which is usually rich in vegetables and fruits, legumes, whole grains, fish, olive oil, and low-fat dairy products and includes a moderate consumption of wine. Adherence is measured by attributing a score for all individuals for each of these foods. In several prospective investigations, inflammatory markers were inversely associated with this dietary pattern. The potential role of the Mediterranean diet has also been compared with other diets in a few short-term intervention studies, and inflammatory markers decreased with adherence to the Mediterranean diet, independent of weight loss [115]. Overall, there is a strong suggestion that Mediterranean dietary patterns are useful for decreasing low-grade inflammation markers and therefore have important health effects.

The same trends were observed in an American population using the Healthy Eating Index (HEI). This index was constructed using the Dietary Guidelines for Americans and the Food Guide Pyramid. A negative association was found with CRP for people adhering to the recommendations.

Compared with a healthy non-vegetarian diet, the consumption of a vegetarian diet was also associated with lower levels of low-grade inflammatory markers. However, studies on this subject are usually cross-sectional, and vegetarians could differ from non-vegetarians in other aspects that might not have been considered [115].

The patterns identified via RRR were positively associated with biomarkers of low-grade inflammation, and these patterns were characterized by high meat

consumption and low consumption of fiber-rich foods and were often with alcohol consumption [116].

Diet and breast cancer

A systematic review and meta-analysis of the existing evidence regarding dietary patterns and breast cancer risk investigated the dietary patterns derived from FA or PCA. A decreased risk of breast cancer was observed in those with the highest adherence to healthy/prudent dietary patterns compared with those with the lowest adherence. However, no evidence was found regarding breast cancer and unhealthy/Western dietary patterns. There is evidence of an increased breast cancer risk for those in the highest categories of a "drinker" dietary pattern compared to those in the lowest group [47].

Previous findings from the Malmö Diet and Cancer cohort

Previous findings from the Malmö Diet and Cancer (MDC) cohort have shown that women who eat plant foods and have high fiber and low fat intakes have a reduced risk of breast cancer after menopause. High intakes of ω-6 PUFAs and fat from vegetable oil-based margarines were associated with an increased postmenopausal breast cancer risk [117, 118]. Intakes of energy-adjusted yogurt and regular milk showed protective associations with all invasive breast cancer and protective linear trends for hormone receptor-positive status (ER+PR+) tumors. In contrast, the consumption of vegetable oil-based margarines and dried sauce and soup powders was positively associated with breast tumors with low levels of hormone receptors (ER-PR-) [118, 119]. These findings suggest that non-hormonal mechanisms are at play in the development of breast cancer.

High plasma concentrations of enterolactone and high-fiber diets were associated with reduced breast cancer risk. In contrast, deep fried potatoes were associated with increased risk [46]. Finally, a protective association with diets that provided recommended levels of dietary folate intake was also observed, but this association was dependent on a gene encoding the MTHFR-enzyme, which modified the effect [120].

2.4. Main hypothesis and theoretical framework

Lifelong exposure to sex hormones cannot fully explain previous observations between diet and postmenopausal breast cancer within the MDC. It is plausible that other biological processes are involved in mediating the effect between dietary fat (from vegetable oil-based margarines, fried potatoes, or dried soups and sauces) and breast cancer, especially when associations were observed for tumors that were low in hormone receptors [119]. During the manufacturing process of most foods included in Western diets, many harmful substances can be formed through to exposure to prolonged heat treatment [121]. These substances (i.e., oxysterols) induce cholesterol oxidation and promote oxidative stress but can be counteracted with antioxidants [121]. Furthermore, it is possible that examining dietary patterns instead of single foods/nutrients will improve our understanding of the role of overall diet in the development of breast cancer.

We hypothesize that low-grade inflammation and oxidative stress are a possible pathway that helps explain the inconsistent associations between diet and breast cancer (**Figure 10**).

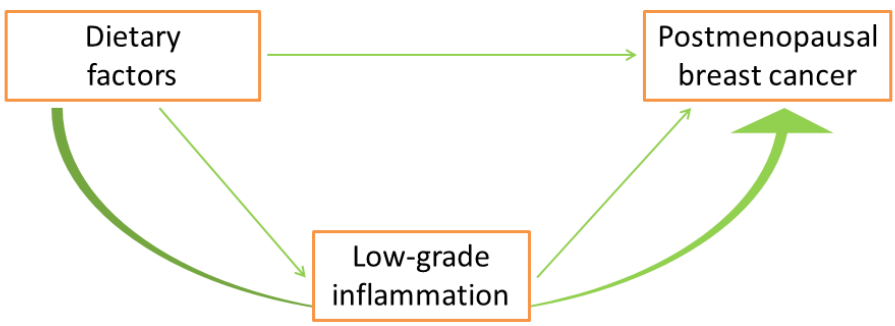

Figure 10. Thesis theoretical framework

The complex picture of the potential mechanisms linking risk factors with breast cancer is depicted in **Figure 11**. The figure provides a quick summary of the associations between diet and breast cancer risk and progression, with a focus on oxidative stress and low-grade inflammation.

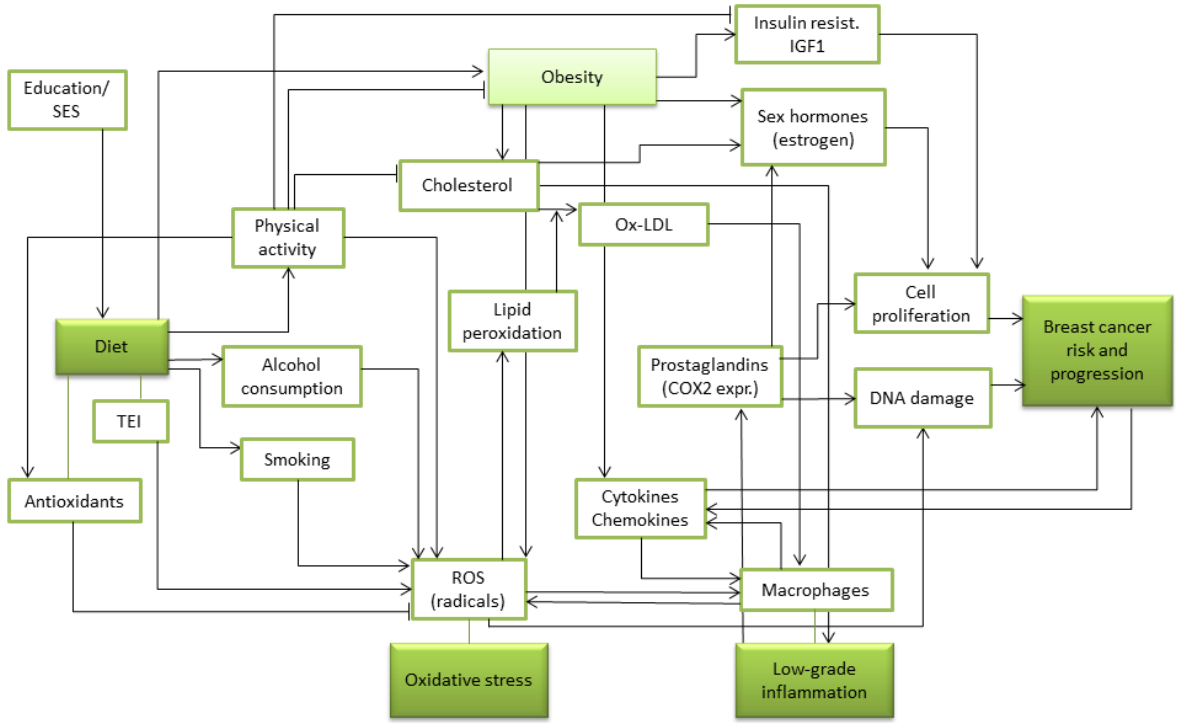

Figure 11. Schematic illustration of possible pathways between diet and breast cancer
Lines with arrows represent positive associations whereas cut lines represent negative associations. Green lines represent specific examples of greater concepts; e.g., TEI and antioxidants are specific parts of the overall diet. ROS, Reactive Oxygen Species; SES, socioeconomic status; TEI, Total Energy Intake; Ox-LDL, oxidized LDL; COX, cyclooxygenases; IGF, Insulin-like grwoth factor.

3. Aims

Overall aim

The aim of this doctoral project is to investigate the association between diet and postmenopausal breast cancer and to explore how this association is affected by low-grade inflammation and oxidative stress.

Specific aims

The specific aims of this doctoral project are as follows:

1. To explore the reliability and reproducibility of selected biomarkers of inflammation (*Paper I*)
2. To investigate how diet quality is associated with biomarkers of systemic inflammation (*Paper II*)
3. To examine the association between selected biomarkers of inflammation and postmenopausal breast cancer (*Paper III*)
4. To investigate the role of obesity in the association between biomarkers of inflammation and postmenopausal breast cancer (*Paper III*)
5. To identify the food patterns associated with three biomarkers of systemic inflammation (*Paper IV*)
6. To determine whether the specific inflammation-related food patterns are associated with postmenopausal breast cancer (*Paper IV*)

4. Materials and Methods

4.1. Subjects

All the studies incorporated into this thesis were conducted within the Malmö Diet and Cancer (MDC) framework. The participants in the variation study described in **Paper I** were recruited from the MDC. The studies described in **Papers II** to **IV** used data collected during the baseline measurements of the MDC and used different sub samples (**Figure 12**).

Source population

Malmö is the third largest city in Sweden, situated in the southern region of Scania ("Skåne"). Now comprising 280,415 inhabitants, it had a population of 223,663 in 1990 [122]. The source population was defined in 1991 as all people residing in Malmö and born between 1926 and 1945 (n=53,325). The population was redefined (in May 1994) to include all men born between 1923 and 1945 and all women born between 1923 and 1950, increasing the source population to a total of 74,138 people (**Figure 12**). The main reason for including younger women was to be able to study premenopausal breast cancer. The recruitment procedures and cohort details have been published previously [123]. In short, personal letters of invitation were sent, and public information campaigns were conducted through local newspapers, in public places or in primary health care centers [124].

The only exclusion criteria were a lack of mental capacity or limited Swedish language skills (n=1,975). An additional 3,241 people died or moved either before getting the invitation letter (n=3,017) or before the completion of the baseline examinations (n=224). Lastly, 17 people could not be identified. A final total of 68,905 people were classified as eligible for participation. Of these, 30,146 people joined the study until the end of the recruitment whereas 21,817 did not reply to the invitation letters, and 16,942 did not want to participate. With a participation rate of approximately 41%, a total of 28,098 participants (of those eligible) completed all baseline examinations, including 17,035 (approximately 61%) women and 11,063 men. Of the total number of participants, 23,016 were recruited by letters of invitation, whereas 5,082 joined spontaneously (following community

advertising). Compensation for participation included small gifts such as T-shirts, pens and plastic bags.

Malmö Diet and Cancer study

The MDC study was jointly planned and initiated by the Faculty of Medicine at Lund University, the Swedish Cancer Society, the Swedish Medical Research Council, and the International Agency for Research on Cancer (IARC) [123]. In 1993, the MDC became an associated member of the European Prospective Investigation into Cancer (EPIC) organized by IARC, Lyon, France [125]. EPIC is a prospective cohort of more than half a million participants from 23 research centers in 10 European countries [126].

The MDC study is a population-based prospective cohort study established with the primary interest of studying the associations between diet and cancer while considering lifestyle factors [123]. The baseline examinations took place in Malmö between March 1991 and October 1996 and consisted of two visits to the study center, separated by nearly two weeks. At the first visit, groups of 6-8 participants were given instructions by trained staff on how to fill out the extensive questionnaire covering lifestyle and socioeconomic factors and on how to record their meals in the 7-day food diary ("Menybok") and the diet history questionnaire ("Kosthistoria"). The participants filled out all questionnaires at home. Additionally, non-fasting blood samples were collected, anthropometric measurements were taken, and blood pressure was measured by trained nurses. The second visit to the study center consisted of individual interviews by trained dietary interviewers to complete the diet history questionnaire and to verify the correctness of the completed questionnaires [123].

Representativeness of the MDC cohort

The MDC participants and non-participants (28,098 versus 40,807) were examined with the objective of studying potential selection bias resulting from the suboptimal participation rate. This examination consisted of the investigation of cancer incidence and mortality at 3 different time periods: prior to, during and after the MDC baseline examinations [127]. The subjective health, sociodemographic factors and lifestyle of the MDC participants were also compared with those of a random sample of participants in corresponding birth-year cohorts using a mailed health survey with a higher participation rate of approximately 75%. Mortality was higher among the non-participants during recruitment and follow-up. This study also suggested that cancer incidence may be lower in non-participants prior to recruitment but higher during the recruitment period. Furthermore, the study indicated that despite the low MDC participation

rate, the sociodemographic structure and the prevalence of smoking and obesity seemed to be similar to those of a study with higher participation rate. In comparison, the MDC had a lower proportion of subjects with foreign backgrounds and a higher proportion of participants that reported good subjective health [127].

Recruitment strategies in the MDC cohort

The impact of the different recruitment strategies used in the MDC study on the interpretation of the findings was investigated [124]. A total of 5,082 MDC participants were recruited through community-directed invitations (thus classified as "passive recruitment"), and 23,016 through personal invitation (considered "active recruitment"). The latter group was further divided into two subgroups: the "early responders" were those who responded to the first letter of invitation, and the "late responders" were those who answered after two or more letters of invitation. The study reported lifestyle and socioeconomic differences between the active and passive responders; for example, the passive responders were more often female, older, and less likely to have a university degree or to be unemployed. Cancer incidence and all-cause mortality during follow-up (until December 31, 1999) was lower among the passively recruited responders [124].

Malmö Diet and Cancer study – Cardiovascular Cohort

During recruitment to the MDC study between October 1991 and February 1994, a random sample of approximately 50% of those who enrolled (n=6,103) was invited to participate in a sub study that focused on the epidemiology of carotid artery disease: the Malmö Diet and Cancer - Cardiovascular Cohort (MDC-CC) [128]. They also underwent an ultrasound examination and a medical history review. A third visit to the study center was scheduled to collect fasting blood samples under standardized conditions. The sub study occurred within a median of seven months after recruitment (8 months on average) and included a total of 5,540 participants.

Paper I

The variation study described in **Paper I** included participants recruited from the MDC who had not been admitted to the hospital in the previous year and were born between 1940 and 1950. Of those who fulfilled the criteria (n=3,586), 1,000 participants were randomly listed. After approximately 600 invitation letters were mailed, recruitment ended when more than 100 people had expressed their interest in participating. Recruitment and data collection took place between November

2010 and June 2011. A total of 113 people expressed their interest in participating in this study, but eleven dropped out before the start, and five terminated the study due to problems with blood sampling. In total, 97 participants completed the blood sample collections. Two individuals with extreme and unexplainable values for two of the biomarkers were excluded from further analysis. The study sample consisted of 95 participants: 49 males and 46 females.

Paper II

The population considered for this study consisted of a random sample selected from the MDC-CC sub-cohort (n=6,103): 700 individuals, aged 63-68 years [129]. Due to a lack of dietary data, 33 people were excluded from the analysis, and the total study sample consisted of 667 people: 276 men and 391 women.

Papers III and IV

The studies described in **Papers III** and **IV** were prospective case-control studies, nested within the MDC cohort. These studies examined the risk of postmenopausal breast cancer associated with diet and chronic inflammation. Women who were free of prevalent cancers (except cervical cancer in situ) and aged ≥55 years at the time of the baseline measurements were eligible (n=8,513). Women in the MDC cohort were excluded from these studies if they had a prevalent cancer (n=1,239), were younger than 55 of years (n=7,431), or both (n=325). Other individuals had no remaining plasma samples (n=141) or could not be identified (n=36). When follow-up ended on December 31, 2010, a total of 459 cases of invasive breast cancer had been diagnosed. Three controls per case (n=1,377) were randomly drawn from the same source of 8,513 women and matched by age at baseline examinations (± 3 months) and date of blood sampling (± 1 month). Additional requirements for the controls were being free of breast cancer, living in Sweden and being alive at the time of the matching. Laboratory analysis of the inflammation biomarkers was successful for 446 cases and 910 controls (approximately 2 controls per case) and thus constituted the study sample for **Paper IV**. The statistical procedure applied in **Paper III** (conditional logistic regression) influenced the study sample, which included only controls matched to the specific cases (446 cases and 885 controls). In this case, seven of the cases had only one control with a successful laboratory analysis, and the remaining eighteen controls were matched with cases that were excluded prior to statistical analysis due to unsuccessful laboratory analysis.

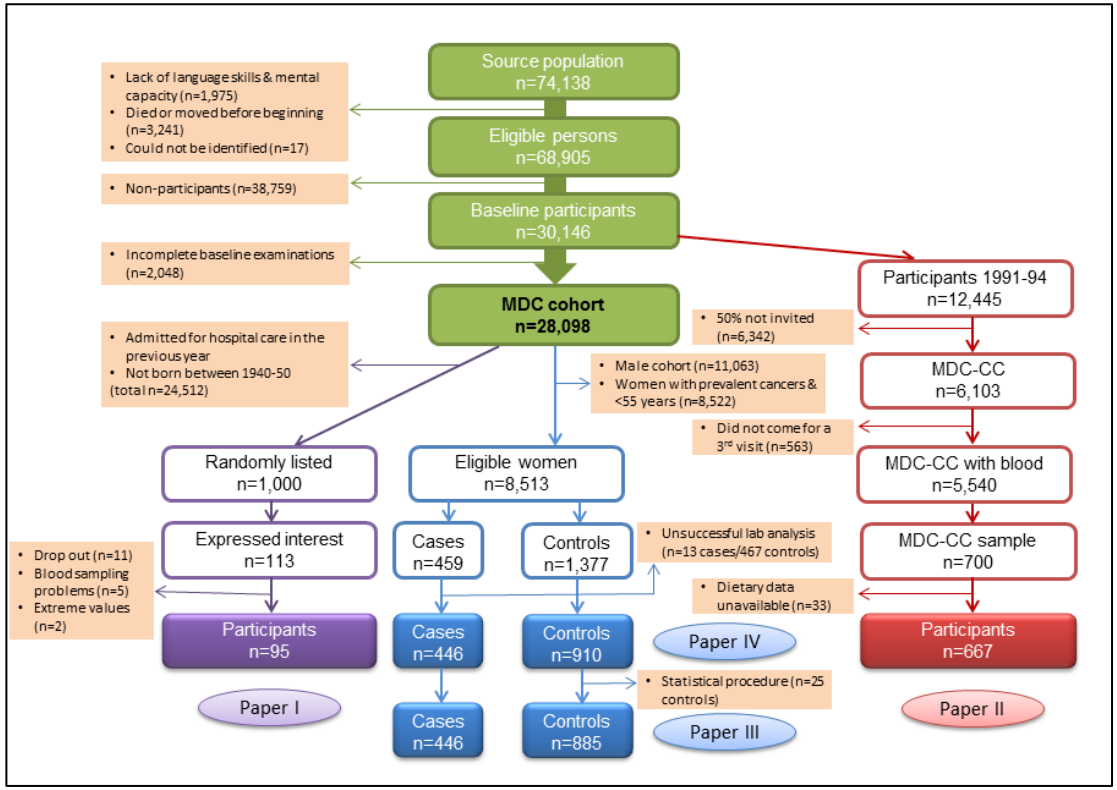

Figure 12. Flowchart of the participants in the Malmö Diet and Cancer cohort and study populations included in Papers I-IV
Sampling process from the source population through the selection of each study sample, including the number of people and exclusion criteria for the different steps.

4.2. Assessment of exposures and covariates

Ethical considerations

The Ethical Committee at Lund University approved the MDC study (LU 51-90). All the participants received detailed information about the study and gave their written informed consent at their first visit to the study center, prior to any examinations. The studies described in **Papers II**, **III** and **IV** were performed under this ethical approval.

The variation study described in **Paper I**, however, required a new ethical approval. The participants received detailed information about the study and were assured that participation was voluntary, and they gave their written informed consent prior to any new data collection (in 2010/11). This study was approved by the Regional Ethical Review Board in Lund, Sweden (Dnr 2010/435).

Anthropometrics

During the baseline examinations of the MDC cohort, trained nurses performed anthropometric measurements at the first visit to the study center. With the participants wearing light indoor clothing and no shoes, height (cm) was measured using a wall-mounted stadiometer, weight (kg) was measured using a calibrated balance-beam scale, and waist and hip circumferences (cm) were measured midway between the lowest rib margin and the iliac crest and horizontally at the level of the greatest lateral extension of the hip, respectively. After a 10-min rest in the supine position, blood pressure was assessed once (mmHg). Body composition (i.e., fat mass and fat-free mass) was estimated with bioelectric impedance analysis (BIA; BIA 103, RJL Systems, Detroit, MI, U.S.A.; single-frequency analyzer) following the manufacturer's instructions, and body fat % (BF%) was calculated using the algorithm provided by the manufacturer.

In the variation study (**Paper I**), the participants' heights and fasting weights were measured while they were wearing light indoor clothing, at the second visit to the study center (in 2010/11). Weight measurements were repeated at the sixth visit.

Timeline

The different timelines for each of the studies included in this thesis are illustrated in **Figure 13**.

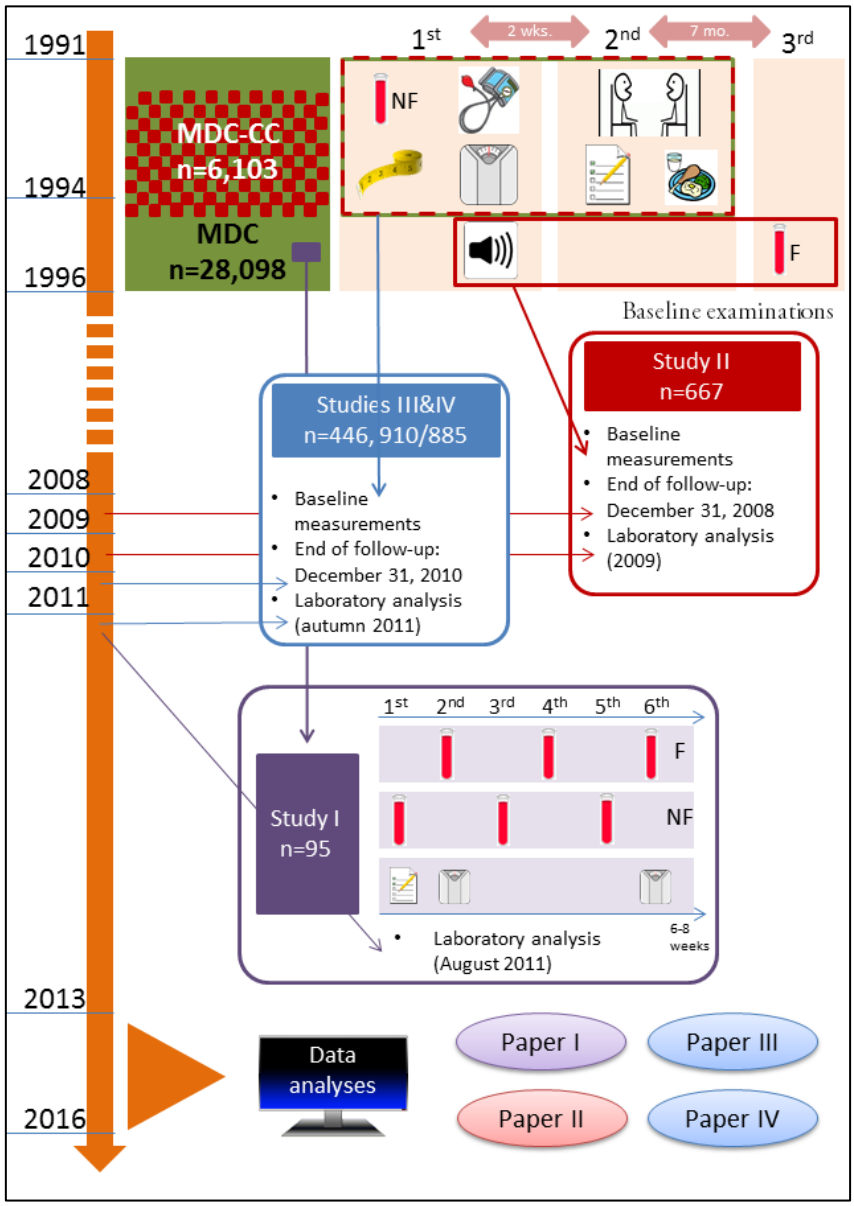

Figure 13. Timeline of events
Depiction of the timeline, including baseline measurements, specific study measurements, sampling, laboratory analyses, and statistical analyses. NF, Non-fasting; F, Fasting.

Questionnaire

A self-administered questionnaire focusing on disease history, lifestyle, socioeconomic factors and hormonal factors was delivered to all participants during the baseline examinations. This extensive questionnaire included questions pertaining to 1) education and work, 2) nationality, 3) social network and support, 4) feelings and thoughts, 5) leisure time physical activity (including sleeping habits), 6) tobacco use, 7) alcohol consumption, 8) health status (including reported health, weight change, past/current diseases, dietary change, use of medications), and 9) family history of diseases, plus an additional section for women addressing 10) reproductive factors (including pill usage, reported ages at menarche and menopause, parity, miscarriages, and breastfeeding). Throughout the different sections of the questionnaire, several questions regarding the psychological wellbeing were posed.

In the variation study (**Paper I**), the participants also completed a brief socioeconomic and lifestyle questionnaire at the time of their first visit (in 2010/11).

Reproducibility of the MDC questionnaire

In 1994, a random sample of participants was invited to participate in a reproducibility study, nearly 3 weeks after their baseline examinations [124]. The questionnaire was then administered a second time. A total of 211 participants responded to the questionnaire twice (a participation rate of approximately 91%); of these, 209 had complete baseline examinations. Kappa statistics were used to determine the concordance between questionnaires (at baseline and three weeks after). Most of the variables showed high agreement/reproducibility (>0.75). For women, the kappa coefficients for selected variables were as follows: 0.84 for education, 0.94 for smoking, 0.77 for alcohol, 0.84 for weight change, and 0.68 for dietary change [124].

Biological material

Non-fasting blood samples collected during the baseline measurements of the MDC were handled by trained staff according to the highest standards and following the guidelines of a strict quality control program [130]. Blood components were separated into fractions within one hour, as described previously by Pero et al. [131]. In short, of the 40 ml of blood donated by each participant, 30 ml was dedicated to blood fractioning, whereas 10 ml was used to prepare a serum sample, which was stored without an anticoagulant. Mononuclear leukocytes were cryopreserved at -140°C, while granulocytes, erythrocytes, plasma and serum were

separately preserved and stored at -80°C [131]. Fasting blood samples were prepared according to the same procedures.

Blood sample collection for the variation study (**Paper I**) occurred on 6 occasions. The participants donated fasting and non-fasting blood alternately (starting with non-fasting blood), so that each participant contributed three fasting and three non-fasting blood samples over a 6 to 8-week period. All plasma samples were stored at -80°C and analyzed at the same time, shortly after the completion of the data collection (August 2011).

Quality control program

The quality control program was developed to ensure validity and control for potential variables in the cell separation and freezing procedures at the MDC baseline examinations [131]. The main objective was to define the baseline variability in the stored samples for future users at the time of sampling and storage. Thus, three main areas for quality control were defined and described: 1) instrument variability, 2) blood cell fraction variability, and 3) variability in the stored blood fractions. The overall conclusion was that the samples obtained at baseline were collected in a reproducible and quality-controlled manner as no differences in yield, purity, or storage were found [130]. The exception was the growth response of mononuclear leukocytes and granulocytes. This led to a change in the procedures in August 1995, and instead of purified mononuclear leukocytes and granulocytes, buffy coats were stored (at -140°C) [130].

Assessment of the biomarkers

In the variation study (**Paper I**), we assessed biomarkers using the plasma collected during the study enrollment (and not the plasma collected at the MDC baseline measurements). The concentration of ox-LDL in the plasma was analyzed with a capture ELISA (Mecordia, Uppsala, Sweden) using the mAb-4E6 antibody against a conformational epitope in oxidized ApoB-100, developed by Holvoet et al. [132]. The analytical coefficient of variation (CV_A), or inter-assay variation, was 5%. The biomarkers of inflammation were analyzed using the Human Pro-Inflammatory 4-plex Ultra-Sensitive Kit (Meso Scale Discovery, Gaithersburg, MD, USA). The CV_A values were 12%, 8%, 12% and 8% for IL-1β, IL-6, IL-8, and TNF-α, respectively. The LLOD for the cytokines ranged between 0.10-0.37 pg/ml, according to the manufacturer's indications. All Meso Scale assay plates were ordered at the same time, and all six samples from each individual were run on the same ELISA plate.

The leukocyte counts described in **Paper II** were assessed at the time of the baseline examinations using fresh heparinized blood [133] and a SYSMEX K1000 automatic counter (Sysmex Europe, Norderstedt, Germany), and were expressed as totals (white blood cells – WBC) and differentials (i.e., neutrophils,

lymphocytes and mixed cells, including monocytes, eosinophils and basophils). Scatter properties and expression of the CD14 and CD16 surface markers were used to identify the different monocyte subsets (i.e., the classical $CD14^{++}CD16^{-}$, non-classical $CD14^{+}CD16^{++}$, and intermediate $CD14^{++}CD16^{+}$ monocytes) [129].

Fasting samples were collected on the 3rd visit only for the MDC-CC participants. This plasma was used to assess the concentration of the inflammatory protein S100A8/A9 using commercially available ELISA kits (BMA Biomedicals, Augst, Switzerland). The C-reactive protein (CRP) concentration was assessed using the high-sensitivity CRP (hs-CRP) test, performed using Tina-quant CRP latex assay (Roche Diagnostics, Basel, Switzerland) on an ADVIA 1650 Chemistry System (Bayer healthcare, NY, USA) [134]. The concentration of cytokines, such as IL-1β, IL-8, and TNF-α, was measured in the plasma using a multiplex immunoassay (Meso Scale Discovery, Gaithersburg, MD, USA) [135].

Information retrieved from the leukocyte count assessments performed at baseline was used in **Paper III**. Other inflammation markers, such as the cytokines used in **Papers III** and **IV**, were assessed during the autumn of 2011, using baseline blood samples. Plasma samples were then thawed, and the concentration of ox-LDL in the plasma was analyzed with an ELISA (Mecordia, Uppsala, Sweden), as described in **Paper I**. The inter-assay variation was 5.6% for the ox-LDL ELISA and all samples were within the range of detection.

The cytokines described in **Papers III** and **IV** were analyzed with the Human Pro-Inflammatory 4-plex Ultra-Sensitive Kit (Meso Scale Discovery, Gaithersburg, MD, USA) [135]. The inter-assay variation was 19%, 17%, 10% and 19% for IL-1β, IL-6, IL-8, and TNF-α, respectively. The IL-1β concentration was below the lower limit of detection (LLOD) in approximately 74% of the samples (LLOD varied from 0.00 and 0.57 pg/ml across 18 plates). The IL-6 concentration was below the LLOD in 1.9% of the samples (between 0.16 and 0.73 pg/ml). Finally, the concentrations of IL-8 and TNF-α were not detectable in three samples (LLODs between 0.04 and 0.15 pg/ml and 0.15 and 0.58 pg/ml, respectively). There were no missing values for ox-LDL or the cytokines. All the samples were analyzed randomly across the 18 plates. In 73% of the samples (n=979), the cases and matched controls were analyzed in the same plate, whereas in 26% of the samples, the cases were analyzed in different plates than the controls.

Dietary assessment

A modified diet history method was specifically developed for use during the baseline examinations of the MDC study, with the aim of capturing total diet and with a special focus on total fat. The chosen methodology took into account the age group of the study population – a middle-aged urban population – as it was

expected that their eating habits would be fairly regular and would include traditionally cooked meals with seldom inclusions of fast food or meals eaten outside the home [136]. The dietary assessment consisted of three parts; a) a 7-day food diary ("Menybok"), b) a 168-item semi-quantitative diet history questionnaire ("Kosthistoria"), and c) a 45 to 60-min diet history interview. The assessment of the total diet is reflected by the combination of "usual diet" (diet history questionnaire) and "current diet" (food diary) methods.

In the 7-day food diary, the participants were asked to record the "cooked/main" meals eaten on a daily basis but with a high day-to-day variation, such as lunch and dinner, over 7 consecutive days. This diary also included cold beverages (i.e., juice, milk, water, soft drinks, and alcoholic beverages) as well as drugs, dietary supplements and natural remedies. The participants recorded a general meal pattern and the frequency and portion size information of the foods they consumed regularly (with a low day-to-day variation) in the diet history questionnaire. Using the preceding year as the reference period, participants reported the frequency of consumption of hot beverages (such as coffee and tea), sandwiches, breakfast cereals, edible fats, milk, yogurt, candies, cakes and snacks. The portion sizes reported in the diet history questionnaire were estimated at home using a booklet with 48 black and white photographs. Each set of photographs included 4 options of different portion sizes of a food/dish.

Both the food diary and the diet history questionnaire were filled out at home. At the time of the diet history interview (performed on the second visit during baseline examinations), the participants estimated the usual portion sizes of foods and dishes reported in the food diary with the help of a more detailed and comprehensive book of photographs. A set of 4 photos showing different portion sizes (A-D) was provided for each of the reported dishes or foods. The participants also described in detail how the dishes and foods recorded in the food diary were prepared (e.g., the specific ingredients in mixed dishes, the type of fats used for cooking, etc.). Dietary interviewers carefully checked for the consistency of the information provided and possible overlapping information from both tools. Over the years during the baseline measurements, a total of 17 trained interviewers conducted the dietary interviews.

The food data collected from the two sources were then entered into the data system using the interactive software KOSTVAR (AIVO AB). The interviewer was aided by a system of "recipe identifiers" provided by the software. Based on the information from all sources, the average daily consumption of food various groups (grams per day) was calculated for each individual and further converted into nutrient and energy intakes using the MDC Food and Nutrient Database, based on the Swedish Food Database PC KOST-93 of the Swedish National Food Administration [137]. A thorough description of the modified diet history

methodology and its utility is available to all interested parties on the corresponding webpage [138].

Validity and reproducibility

To determine the quality of the modified diet history method used in the MDC study, its validity and reproducibility were evaluated [137, 139-141]. The relative validity of the method was investigated in 1984-85 among 206 Malmö residents (105 women and 101 men) aged 50-69 years [140, 141]. The reference method used was an 18-day weighted dietary record kept for 3 consecutive days every other month (6 times in total, over one-year period). Thus, seasonal variation was captured as well as weekdays and weekend days. Two methods were investigated and compared with the reference method: an extensive food frequency questionnaire (FFQ) and a combined food record with a quantitative FFQ (an early version of the modified diet history method). Both methods proved to be reasonably accurate (with high correlation coefficients for most foods and nutrients), but the latter overestimated most food groups [140] and most nutrients [141] to a lesser extent. Energy-adjusted Pearson correlation coefficients for both women and men were between 0.6 and 0.8 for most nutrients and food groups (**Table 3**).

A total of 241 subjects (115 female and 126 male residents of Malmö, aged 50-69 years) participated in the examination of the reproducibility of the two methods described above (111 of these subjects also participated in the validation study) [139]. The methods were applied twice with a one-year interval, and a total of 120 participants used the food record combined with a quantitative FFQ. The energy-adjusted Pearson correlation coefficients for the men and women were between 0.5 and 0.9 for most food groups and nutrients (**Table 3**) [139].

Table 3. Relative validity and reproducibility of the MDC dietary assessment method

Dietary variable (food or nutrient)	Relative validity[1] (men/women)	Reproducibility[2] (men/women)
Energy	0.55/0.55	0.70/0.79
Protein	0.54/0.53	0.63/0.54
Fat	0.64/0.69	0.49/0.52
Carbohydrates	0.66/0.70	0.50/0.49
Fiber	0.74/0.69	0.66/0.70
Alcohol	0.80/0.78	0.70/0.82
Sucrose	0.60/0.74	0.78/0.46
SFA	0.56/0.68	0.64/0.62
MUFA	0.59/0.66	0.46/0.50
PUFA	0.26/0.64	0.68/0.70
Vegetables	0.65/0.53	0.71/0.76

Fruits	0.60/0.77	0.80/0.81
Fish	0.35/0.70	0.78/0.22
Meat	0.84/0.92	0.96/0.57
Eggs	0.57/0.74	0.48/0.56
Milk	0.83/0.84	0.82/0.70
Cheese	0.47/0.59	0.71/0.71
Cream	0.47/0.52	0.48/0.42
Bread	0.50/0.58	0.45/0.65
Cereals	0.74/0.73	0.76/0.61
Potatoes	0.69/0.51	0.82/0.43
Rice and pasta	0.35/0.24	0.41/0.23
Fats and oils	0.54/0.66	0.62/0.46

[1]Energy-adjusted Pearson correlation coefficients for the intakes estimated using the MDC method and the reference method (18 days of weighed records).

[2]Energy-adjusted Pearson correlation coefficients for the intakes estimated using the MDC method on two occasions 1-year apart.

Diet assessment method version

Due to constricted funding, the routines for dietary coding (but not dietary reporting) at the baseline of the MDC were altered in September 1994 [142]. The main change consisted of using standardized recipes for some dishes and standardized portion sizes for some foods instead of individualized ones, thus shortening the interview time. The alteration of coding routines seemed to affect absolute intakes of fat and energy, but it did not appear to significantly alter the ability of ranking individuals [142].

Energy misreporting

Energy misreporting was defined as a ratio of energy intake (EI) to basal metabolic rate (BMR) outside the 95% CI limits of the calculated physical activity level (PAL) [143] following the Goldberg et al. and Black approach [144, 145]. Approximately 12% of the men and 18% of the women were classified as low energy reporters, whereas 3.5% of the men and 2.8% of the women were high energy reporters. Having a larger waist circumference and a high BMI, being a blue-collar worker and having a short education were significantly associated with low energy reporting. Living alone, current smoking and low BMI were significantly associated with high energy reporting [143].

4.3. Classification of outcomes and definition of variables

Ascertainment of breast cancer cases

The Swedish Cancer Registry ("Cancerregistret, Socialstyrelsen") and the Southern Swedish Regional Tumor Registry ("Regionala tumörregistret", Lund) provided data on case definition until the end of follow-up on December 31, 2010. Women with an invasive breast cancer diagnosis (classified according to the International Classification of Diseases system, IDC7=170), defined as all cancers except in situ cancer, were considered cases. Cases were then identified in the MDC cohort by linking them with the registries using the participants' Swedish personal identification number ("personnummer"). These registries are well established and have been estimated to be almost 100% complete [146].

The Swedish Cause of Death Register ("Dödsorsaksregistret", in the National Board of Health and Welfare – "Socialstyrelsen") and the Total Population Register ("Registret över totalbefolkningen", in Statistics Sweden – "SCB") provided information on vital status.

Dietary variables

Diet quality index (Paper II)

In **Paper II**, the following dietary variables were used to construct a dietary index: saturated fatty acids (SFA) (E%), polyunsaturated fatty acids (PUFA) (E%), fish and shellfish (g/week), dietary fiber (g/MJ), fruits and vegetables (g/day) and sucrose (E%). The index was intended to assess adherence to the Swedish Dietary Guidelines (SDG) and the Swedish Nutrition Recommendations (SNR-2005) [147]. During the development of the diet quality index (DQI-SNR) for the MDC, three aspects were taken into account: 1) the components included had to be mutually exclusive (i.e., not highly correlated with each other), 2) information regarding specific nutrients/dietary components had to be available in the MDC database, and 3) to reflect overall diet quality, specific components that have been previously associated with chronic disease were selected [147]. Previous reports indicated the value of the DQI-SNR for ranking individuals on their reported adherence to the SNR, and as a tool to predict overall and CVD-specific mortality [148], and CVD-incidence within the MDC cohort [149].

Participants who adhered to the recommendations for each of the 6 dietary components were given 1 point, while non-adherents were given zero points. The

cut-offs were defined based on the recommended intake levels in the SDG and SNR [147], and 1 point was attributed for each of the components if SFA consumption was ≤14 E% (non-alcohol energy percentage), PUFA consumption was between 5 E% and 10 E%, fish and shellfish consumption was ≥300 g/week, dietary fiber consumption was between 2.4 and 3.6 g/MJ, fruit and vegetable consumption was ≥400 g/day, and sucrose consumption was ≤10 E%. The 6 components were then summed to produce a total score (ranging from 0 to 6), which was further categorized as low (0-1 points), medium (2-3 points), and high (4-6 points).

The cut-offs for three of the components were modified from the original recommendations to meet the specific characteristics of the MDC cohort. Only a small percentage of the MDC participants met the recommendations for SFA (only 4% in **Paper II**); thus, the approach used to calculate the new cut-off was similar to the one presented by Drake et al. [147]. One standard deviation (SD) from the mean intake of the population was added to the SNR-recommended level for a final cut-off of 14 E%. The fiber cut-off used was based on the mean recommended level of intake (between 25 and 35 g/day, which is approximately 3 g/MJ). The values for the cut-offs were determined by subtracting and adding one SD of the population mean. In this study, the cut-off for fruits and vegetables was reduced from ≥500 g/day because fruit juices (included in the original recommendations) were excluded.

Food patterns derived from reduced rank regression (Paper IV)

Taking fiber content, fat content (quality and consistency), and food culture (traditional food use in the region) into consideration, a total of 40 food groups (including beverages) were aggregated in **Paper IV**. These included vegetables, fruits, juices, lean meats, fatty meats, eggs, sausages, lean fish, fatty fish, shellfish, boiled potatoes, fried potatoes, cereals, rice and pasta, fiber-rich bread, white bread, high-fiber crisp bread, low-fiber crisp bread, cakes and buns, cheese, cottage cheese, low-fat milk, whole milk, cream, yogurt, ice cream, butter, oils, solid margarines, soft margarines, (both high in fat content, 61-80%), low/medium-fat margarines (<40%), dressings, marmalade and sugar, fatty snacks, sweets, coffee, tea, soft drinks, ketchup and dried soups/sauces (all expressed as g/day). A thorough description of what was included in each food group is provided in **Table 4**.

Table 4. Description of variables included in the 40 food groups used to derive food patterns from reduced rank regression (RRR) (Paper IV)

Food groups	Variable description
Vegetables	Cooked vegetables, legumes, raw vegetables (carrots, root vegetables, leafy/salad greens, cabbage, tomatoes, other).
Fruits	Non-citrus (including berries) and citrus. Cooked, dried, and fresh.
Juices	Made of vegetables (carrot, mixed vegetable juice) and/or fruits (citrus and non-citrus).
Fatty meats	Beef (HF>10%), lamb (HF>10%), pork (HF>25% and MF=11-24%), and assorted pork spreads/ham
Lean meats	Beef (LF≤10%), lamb (LF≤10%), pork (LF≤25% and MF=11-24%), assorted beef-based spreads, poultry, game, and “clean” viscera (offal: liver, kidney, heart, tongue).
Sausages	Sausages, sausage spreads, and viscera products (liver sausage, blood pudding).
Eggs	Total egg consumption.
Fatty fish	Fish (HF>5%), all types of fatty fish including canned products (includes fish such as salmon, herring, mackerel, eel, octopus, and tuna).
Lean fish	Fish (LF≤5%), all types of lean fish including canned products (includes fish such as codfish, haddock, whitefish, halibut, plaice, flounder), and others (fish fingers and spreads).
Shellfish	Shellfish, seafood, and mollusks.
Fried potatoes	Fried and deep-fried potatoes.
Boiled potatoes	Boiled potatoes.
Rice and pasta	Rice and pasta were included together in one variable at the baseline examinations.
Cereals	Mixed cereals with low-fiber/low-sugar, low-fiber/high-sugar, and high-fiber/high-sugar.
Fiber-rich bread	High-fiber (4.6-5.9% and ≥6% of fiber content) soft “table” bread.
White bread	Low- and medium-fiber (≤3.5% and 3.5-4.5% fiber content) soft “table” bread.
High-fiber crisp bread	High-fiber (>20% of fiber content) hard bread, and high-fiber (>10 g of fiber) wholegrain biscuits.
Low-fiber crisp bread	Low- and medium-fiber (<10% and 10-20% fiber content) hard bread, and low-fiber crackers (cream crackers, wafers).
Cakes and buns	LF (≤15 g of fat) wheat bread, Marie cookies, and HF (>15 g of fat) digestive cookies, gingerbread, wafers with cream filling.
Cheese	MF (11-20% fat) and HF (>20%) cheeses.
Cottage cheese	LF (≤10% fat) cheese.
Whole milk	Between 2.5 and 7% fat.
Low-fat milk	LF (≤0.5% of fat) and MF (0.6-2.4% of fat).
Yogurt	“Fermented milk” (including all fat levels: LF, MF and HF).
Cream	All cream: LF (≤12%), MF (13-30%), and HF (>30%).
Ice cream	Sorbet (water-based ice cream ≤6% fat), and milk-based ice cream (>6% fat).
Butter	Lard and coconut oil, butter, margarine, milk-based (HF, e.g. “Bregott”).
Oils	Olive oil, corn, grapeseed, sunflower seed, and rapeseed oils.
Soft margarines	Most household margarines, MF (61-80% fat) >20% PUFA content.
Solid margarines	Solid margarines, MF (61-80% fat) ≤20% PUFA and ≥20% MUFA.
Low/medium-fat margarines	LF (≤40%) milk-based margarines, LF margarines with high PUFA content (≤40% fat, >10% PUFA, <10% MUFA), and MF margarines (41-60% of fat).

Dressings	Mayonnaise, dressings (all levels of fat).
Marmalade and sugar	Marmalade, jam, honey, puree, and pure sugar.
Sweets	Sugary candy.
Fatty snacks	Nuts, seeds, almonds, chocolate, and snacks/chips.
Coffee	Total coffee consumption.
Tea	Total tea consumption.
Soft drinks	Soft drinks including Coca-Cola, "saft", etc.
Ketchup	Total ketchup consumption.
Dried soups/sauces	Industrial soups, gruel, and nutritional powders.

LF, low fat; MF, medium fat; HF, high fat

Methodological variables

Season

Season was constructed using the date of the measurements taken at the baseline examinations because dietary intakes may vary depending on season. Season was classified as follows: spring (March-May), summer (June-August), autumn (September-November), and winter (December-February). This variable was examined as a confounder in **Paper II**.

Past food habit changes

One question on the baseline questionnaire assessed whether the participants "had substantially changed their eating habits because of disease or other reasons" (yes/no), thus possibly identifying participants with unstable food habits. The food assessments of individuals who report dietary changes might reflect a short period of dietary intake of their lives, thus violating the assumption that the reported usual diet is stable over time. Two previous reports showed an association between report of food habit change and obesity, lifestyle and socioeconomic variables [150, 151]. As part of the sensitivity analyses, we excluded participants who reported past food habit changes, when examining dietary intake and outcomes (**Papers II** and **IV**).

Misreporting of energy and dietary assessment method

The participants were classified as under-, adequate- and over-energy reporters [143]. Individuals who were identified as misreporting their energy intake (i.e., under- and over-reporters) were excluded from the sensitivity analysis in **Paper IV**. In **Paper II**, the variable identifying the dietary assessment method in terms of a slight change made to the interviewing routines in September 1994 [142] (labeled as "old" and "new") was examined (but the results are not presented in the paper).

Biomarkers of inflammation

Several biomarkers of low-grade inflammation and oxidative stress were used across the different papers: ox-LDL (U/l; **Papers I**, **III**, and **IV**); IL-1β (pg/ml; **Papers I-IV**); IL-6 (pg/ml; **Papers I** and **III**); IL-8 (pg/ml; **Papers I-III**); TNF-α (pg/ml; **Papers I-IV**); WBC count and differential (i.e., neutrophils and lymphocytes; **Papers II** and **III**); monocyte subtypes (cells/µl) according to CD14 and CD16 expression (**Paper II**); high-sensitivity CRP (mg/dl; **Paper II**); and S100A8/A9 (µg/ml; **Paper II**).

Methodological variable

To account for a possible batch effect (in **Paper III**), we added the ***plate number*** (from 1 to 18) as a covariate in the fully adjusted models.

Anthropometric, socioeconomic and lifestyle variables

Age and gender

The Swedish personal identification number is a unique combination of the birthdate and 4 additional digits, thus providing each person's birthdate. Age was calculated by subtracting the date of entry into the MDC study at baseline from the birthdate for each participant. In **Paper I**, age was calculated using the date of entry into the specific study. One of the four additional digits identifies gender. Age was investigated as a continuous variable, whereas gender was categorical.

Date at screening (week of blood sampling)

In **Papers III** and **IV**, the controls were matched to the cases based on their age at baseline (± 3 months) and on the date of blood sampling (±1 month). The variable "week of blood sampling" was created by subtracting the last entry date from the first date of the MDC study (for this subsample); it ranged from week 0 (on March 22, 1991) to week 287 (on September 24, 1996) and was investigated as a continuous variable.

Cohabitation

One question on the baseline questionnaire assessed whether the participants lived alone, with a significant other (and with or without children), with parents or with others. In **Paper II**, this was categorized as "living alone" or "cohabiting".

Education level and employment status

On the questionnaire, six alternatives were provided for the participants to report their highest education level achieved. Education level was further categorized

into primary school, elementary/secondary school, upper-secondary/high school and university level. The participants were asked to report their current employment status, and it was defined as white-collar, blue-collar or employer/self-employed (**Paper I**).

Leisure time physical activity

To assess leisure time physical activity (PA), a list of 18 different activities adapted from the Minnesota Leisure Time Physical Activity Questionnaire [152] was included in the baseline questionnaire. A previous report investigated the ability of the PA questionnaire to predict health-related risks and showed moderate correlations with accelerometer measurements [153]. The list included activities such as swimming, cycling, walking, mowing the lawn, playing tennis, golfing, playing football/handball, and a final open option. The participants recorded the minutes per week spent on each activity, which was further multiplied by an activity-specific factor and summed into a score. The score was then divided into tertiles (**Papers II-IV**).

Household and work-related physical activity

The participants reported household physical activity in hours. This was divided into tertiles from light to heavy. One question assessed the degree of physical activity needed to perform job-related activities. It consisted of 5 options ranging from "very light" to "very heavy", and the last two categories were combined into one category ("heavy") because of the very small number of individuals in the "very heavy" category. These variables were investigated in **Paper II**.

Smoking status

The participants reported their smoking status in the baseline questionnaire by answering a question with 4 response options: smoke regularly, smoke sometimes, quit smoking, and never smoked. In **Papers II-IV**, the first two categories were combined into one category: "active smokers". In **Paper I**, snuff users were also included in that category, which was renamed "active users of tobacco".

Alcohol consumption

Information on alcohol consumption was aggregated from 2 sources: a question on the questionnaire and indications in the 7-day food diary the participant had consumed alcohol. The questionnaire assessed whether participants had not drunk any type of alcohol in the past year, if they had drunk in the past year but not in the last 30 days, or if they had consumed some sort of alcohol in the past 30 days. Participants who reported no alcohol consumption on both the 7-day food diary and the questionnaire were classified as "zero consumers". The remaining participants were classified as low, medium and high consumers according to cut-

offs based on a biological risk assumption [154]: 15 and 30 grams/day for women, and 20 and 40 grams/day for men.

Hypertension, prevalent diseases and medication

Blood pressure measurements taken at baseline were used to identify hypertension, which was defined in **Paper II** as blood pressure ≥140/90 mmHg and/or reported current use of medication to lower blood pressure. Diabetes was defined in a similar fashion, based on self-reported diabetes at baseline combined with reporting the use of anti-diabetic drugs. To more accurately determine the history of prevalent cardiovascular (CVD) events (including stroke and coronary events) at baseline, information was retrieved from validated national and regional registries. History of CVD and diabetes at baseline was also examined in sensitivity analysis in **Paper III**.

In **Paper I**, blood pressure was not measured, and hypertension was derived from participants' self-reports on the questionnaire at the time of the study. Other diseases reported included myocardial infarction, stroke, diabetes, rheumatoid arthritis, and angina of the legs, which were included with hypertension in one category of the variable "previously diagnosed diseases". The other category included no diseases or others not described in the list above. A second variable was created, "medication", which included the use of statins, non-steroidal anti-inflammatory drugs (NSAIDs) and medication for hypertension in one category and no medication or other drugs not mentioned above in another.

Body composition

Body mass index (BMI) was calculated by dividing weight by squared height (kg/m^2), measured at baseline. In **Paper II**, it was used as a continuous variable, but in **Papers III** and **IV**, it was further categorized according to the WHO reference cut-offs [155]: underweight (<18.5), normal weight (18.5-25), overweight (25-30), and obese (≥30). Because there were very few people in the underweight category, it was combined with the normal weight category. In **Paper I**, BMI was calculated based on the measurements performed at the 2nd visit (to provide fasting blood samples), and it was used as a continuous variable.

In **Papers II-IV**, the waist-to-hip (WHR) ratio was calculated by dividing the waist circumference by the hip circumference. It was used as a continuous variable in **Paper II**, and it was further divided into tertiles and sextiles in **Papers III** and **IV**. Body fat percentage (BF%), derived from the BIA measurements taken at baseline, was also examined in **Paper II** as a continuous variable.

Reproductive factors

Age at specific events

In the baseline questionnaire, the women reported the years when their menstruation started (menarche) and stopped (menopause), as applicable. They also noted the years when each of their children was born. The age at menarche, age at menopause and age at birth of first child were then calculated by subtracting each of the specific years from each woman's birthdate.

Parity and breastfeeding

The reported number of children was divided into 5 categories: nulliparous (no children) and 1, 2, 3, and 4 or more children. Additionally, each woman reported how many months each of their children was breastfed. All months were then summed into a total for each woman.

Menstrual cycles

To approximate the number of years with menstrual cycles, the time span between the age at menarche and the age at menopause was computed. Interruptions caused by pregnancy and lactation were then subtracted from the time span [156].

Oral contraceptive and menopausal hormone therapy use

A question assessed whether the women had used or were using oral contraceptives (OC). Information on the current use of menopausal hormone therapy (MHT) was retrieved from a questionnaire item assessing the medications used on a regular basis, and from the information reported in the 7-day food diary [157].

4.4. Statistical analyses

All statistical analyses in the four papers were performed using the Statistical Package for the Social Sciences (IBM SPSS Statistics; IBM corporation, Armonk, NY, USA) software for Windows, versions 20.0 (**Papers I** and **II**) and 22.0 (**Papers III** and **IV**). The only exception to this was the derivation of food patterns using RRR in **Paper IV**, which was performed by implementing the PROC PLS procedure in the Statistical Analysis System (SAS) software (version 9.3; SAS Institute, Inc., Cary, NC). Statistical tests were two-sided, and the significance level was set at $p<0.05$ (i.e., $\alpha=0.05$).

Descriptive analyses

Baseline characteristics of the study participants

Papers I-IV include one table each describing the lifestyle, socioeconomic and anthropometric characteristics of the study participants. These were represented either as the mean and SD for continuous variables or as proportions (n and %) for categorical variables. However, due to its skewedness, alcohol consumption was represented in **Paper IV** as the median and IQR. In **Paper I**, the study participants were compared with the MDC cohort from which they were derived, divided by gender. The characteristics of the sample population for **Paper II** were described across categories of the DQI-SNR. In **Paper III**, risk factors and baseline characteristics were described between cases and controls. Additionally, the controls were used to examine characteristics across tertiles of several inflammation markers. The characteristics of the study population in **Paper IV** were investigated across tertiles of the three food patterns.

Chi-square tests were used to test for proportion differences (in categorical variables) in **Papers I-III**, and odds ratios (ORs) were used to describe case/control differences in **Paper IV**, whereas linear models were used to test for level differences (continuous variables). In **Paper II**, univariate general linear models (GLM) were used and in **Paper III**, analysis of variance (ANOVA) was used to describe the differences in controls across tertiles of inflammation markers. ANOVA was also used in **Paper IV** for continuous variables (WHR), but Kruskal-Wallis test was used for the alcohol variable, because transformation could not solve its skewed distribution.

Figures and graphs

Different types of figures and graphics were used to further describe some features in the papers. In **Paper I**, drop-line graphs were used to depict the difference in biomarker concentration levels at two measurement points against the number of days between measurements. We also used a simple scatterplot to illustrate the correlations between biomarker levels measured at different time points (not included in **Paper I**). In **Paper IV**, line charts were used to graphically represent the distribution of biomarker concentrations (and 95% CI) across tertiles of the 3 food patterns. Bar charts were also used in **Paper IV** to illustrate the factor loadings characterizing each of the food patterns.

Variable transformation

Variable distribution

Whenever the variables used in the statistical models did not fulfil the normal distribution criteria, they were transformed using the natural logarithm (ln) prior to analysis. When the variables included zero values in the distribution, a small value was added to enable transformation. The exceptions to this were 3 biomarkers in **Paper II** that required a square root transformation. Furthermore, in **Papers III** and **IV**, the biomarker levels were ranked into tertiles.

The percentage of values for specific biomarkers that was below the lower limit of detection (LLOD) was calculated in **Papers I**, **III** and **IV**. In **Papers III** and **IV**, IL-1β included many zeros (n=340) and values below the LLOD (n=640). This led to the classification of this variable into 4 categories: zeros, values below the LLOD, and all values above the LLOD were split along the median (0.40 pg/ml; n=176/175). The concentration levels for TNF-α were not detected in three samples (below LLOD) and these were included in the first tertile of the distribution. In **Paper I**, the percentage of values below the LLOD varied from 0 to 4.4 across the different biomarkers. Consequently, both a complete case analysis and a sensitivity analysis excluding individuals with values below the LLOD for IL-β (4.4%) were performed, and the results were compared.

Energy adjustment

Different types of energy adjustment (for TEI) were made in the different papers among the four available options (the standard method, the nutrient density method, the partition method, and the residual method) [108].

In **Paper II**, we used the nutrient density method, dividing the nutrient intake by TEI. The results were expressed as the energy percentage (E%) for factors contributing to the total energy or as the intake per Megajoule (g/MJ). Total energy was also entered into the main analysis as a covariate.

In **Paper IV**, we used the residual method. Energy-adjusted food factors were obtained when the intakes of each food factor (dependent variables) were regressed on "dietary" energy intake (i.e., not including energy from alcohol; independent variable) after logarithmic transformations. The energy-adjusted food factors were then used in the RRR analysis.

Main analyses

Reliability analysis (Paper I)

To assess the reliability of the selected biomarkers (ox-LDL, IL-1β, IL-6, IL-8, and TNF-α), we used logarithmically transformed data to estimate the intraclass correlation coefficient (ICC) and 95% CI for each of the scenarios (repeated measures for fasting and for non-fasting blood samples). According to the Shrout and Fleiss convention [95], a two-way mixed model and with absolute agreement was used in this study. The reliability of a single measure was investigated by computing single-measure ICCs, whereas the reproducibility of measurements based on the three measurement points was examined by computing the average-measures ICCs [93].

The same reliability procedure in SPSS yielded the Cronbach's alpha estimates, the within-subject coefficient of variation (CV_W) and the between-subject coefficient of variation (CV_B) using conventional analysis of variance (ANOVA) for repeated measurements, and mixed-effect variance components using the method of moments [158]. Further analyses included the derivation of the within-subject "biological" coefficient of variation (CV_I) by subtracting the square root of the analytical coefficient of variation (CV_A) from the square root of CV_W [159, 160]. The index of individuality, which indicates how one single measurement is different form the others, was computed by dividing CV_I by CV_B. Finally, to investigate the homeostatic set point for each biomarker (i.e., the number of days/measurements needed to determine the mean serum concentration), we used the following formula: $CV_W^2/A^2 \times Z^2$ (A stands for accuracy=20%, Z=1.645) [161].

General linear models (Paper II)

We used general linear models (GLM) to test for the mean differences of continuous variables across categories of the DQI-SNR. This included the main exposures (i.e., the inflammation markers) and the continuous covariates (age, BMI, BF%, waist circumference and WHR). The biomarker concentration levels were assessed across categories of the DQI-SNR with several levels of adjustment. In model I, the adjustments included age, gender and total energy. Smoking status and PA level were added to those covariates in model II, and model III included diseases at baseline (hypertension, diabetes and CVD) as adjustments. We examined the *p*-for-trend for all models by including the continuous DQI-SNR score.

Logistic regression (Papers III and IV)

We used logistic regression models (which are a specific type of GLM) to investigate the association between exposures and outcomes when the outcome was a binary variable. In **Paper III** and **Paper IV**, the main outcome was breast cancer status (yes or no), and the OR and 95% CI were estimated based on the independent variables (inflammation markers in **Paper III** and food patterns in **Paper IV**). The OR represents the odds of being a case among the exposed group divided by the odds of being a case among the non-exposed.

In **Paper III**, the OR and 95% CI were computed as tertiles of the following inflammation markers; ox-LDL, IL-1β, IL-6, IL-8, TNF-α, WBC, lymphocytes and neutrophils. Logistic regression conditioned on the matched case/controls trios was performed with several models of adjustment for confounders. Model I included adjustments for matching variables (age and week of blood sampling), and model II added BMI. Additional adjustments were made in model III by including WHR, parity and MHT. A final model (model IV) included the following additional adjustments: PA level, smoking status, alcohol consumption, and education. In a final step, unconditional logistic regression was performed with the same adjustments, and the results were compared.

In **Paper IV**, we used unconditional logistic regression to estimate the breast cancer risk for each tertile of the three derived food patterns (Factor 1, 2 and 3). The adjustments in model I included the matching variables; in model II, the adjustments included a specific covariate for each food pattern: smoking status for Factor 1; MHT, WHR, PA level and alcohol intake for Factor 2; and alcohol intake for Factor 3. In model III, all the covariates were entered for the three food patterns: matching variables, smoking status, WHR, BMI, MHT, education, alcohol intake, parity and PA level. The final model (IV) excluded BMI and WHR from the previous list of covariates to investigate the possibility of an effect mediated via obesity. In a final step, conditional logistic regression was performed with the same levels of adjustments, and the results were compared.

Reduced rank regression (Paper IV)

RRR is a type of regression method that derives linear combinations of a set of food intakes (i.e., a set of predictors) that explain the maximum variation in a set of response variables that are thought to be intermediate outcomes of disease [114]. In **Paper IV**, we defined the biomarkers of inflammation that were associated with breast cancer in **Paper III** (i.e., ox-LDL, IL-1β and TNF-α) as the set of response variables (in this case, as biomarkers of the disease process). We used the 40 food groups (including beverages) described in **Table 4**, which were intended to represent the overall diet, as the set of predictors for RRR.

Secondary analyses

Additional analyses

A linear regression model with backward exclusion was performed in **Paper II** to determine the independent associations between the biomarkers of inflammation and the DQI-SNR score, adjusting for age and gender. Additionally, season was added to the main models to be examined as a possible confounder.

In **Paper III**, an additional analysis examined whether the season and time of blood collection may have influenced the observed associations in the main analysis. For that purpose, differences among controls were explored, associations with breast cancer risk were assessed, and these variables were added to the models to estimate the possible changes in the associations with inflammation markers. A final step included investigating a possible batch effect by adding the plate number in the main analysis.

Further analysis in **Paper IV** involved investigating the associations between breast cancer and key dietary factors (omega-6 PUFA, fiber and dried soups/sauces) that were previously associated with breast cancer in the MDC cohort. Each variable was added with Factor 3 (the food pattern that showed significant association with breast cancer) in models I and III. Finally, an additional step included a stepwise linear regression analysis to identify specific food groups associated with each inflammation marker, and the results were compared with the food patterns derived from the RRR.

Correlations

To linearly measure the degree of association between one variable and another we used Pearson (**Paper IV**), partial (**Paper II** and **III**), and Spearman correlations (**Paper I**) depending on whether the variables were normally distributed, whether we wanted to keep other variables (possible confounders) constant, or whether the variables were not normally distributed. Spearman correlation coefficients were computed when examining the degree of association between the biomarker levels measured at each visit (computed separately for fasting and non-fasting occasions) and when examining the degree of association between the mean levels of inflammation markers in non-fasting samples and fasting samples (**Paper I**). Partial correlations were computed in **Paper II** to examine the associations between inflammation markers and each component of the DQI-SNR while adjusting for age and gender. In **Paper III**, partial correlation coefficients were used to determine the associations between obesity measures and inflammation markers among controls, while adjusting for matching factors (age and week of blood measurement). Finally, Pearson correlation coefficients were computed in **Paper IV** between the food groups with the highest and lowest factor loadings and each of the response variables, i.e., the inflammation biomarkers.

Interaction analyses

To study possible statistical interactions (i.e., when the effects of one factor vary across levels of another factor), interaction terms between the suspected variables were introduced in the models. In **Paper III,** the potential interactions between the biomarkers and the measures of obesity were examined. For that purpose, the variables categorized into tertiles were considered continuous, and multiplicative interaction terms were created (for biomarker × obesity). The interaction terms were introduced together with the factors as separate variables in basic and fully adjusted models to predict breast cancer risk using unconditional logistic regression.

In **Paper IV**, interaction analyses were performed between each of the three food patterns and BMI, MHT and smoking status. The procedure was similar to that used in **Paper III**: categorical variables were considered continuous, and interaction terms were created using basic models of adjustment. The main analysis was further stratified by the categories of each of these variables.

Sensitivity analyses

We performed additional analyses in each of the four papers to estimate the possible uncertainty of the main observations. In **Paper I**, one individual was found to have large variation between the first and last measurement points of one biomarker, IL-8 (**Figure 14**); when this individual was excluded from the analyses, all ICC estimates improved greatly. Thus, the results presented in **Table 9** for IL-8 did not include this individual.

We excluded participants who reported having changed their food habits from the main analysis and in the fully adjusted models, in **Paper II** and **Paper IV**. Additionally, in **Paper IV**, women who misreported their energy intake were excluded.

In **Paper III**, the exclusions included women who were diagnosed with breast cancer until the 3rd year of follow-up (n=45 for the 1st year, n=64 diagnosed until the 2nd year, and n=83 until the 3rd year). Additionally, women with a history of diabetes (n=52) or CVD (n=27) at baseline were excluded.

Handling of missing values

Individuals who did not have values for all the variables (exposures or confounders) introduced in the fully adjusted models are generally excluded from analyses by default. Doing so can reduce greatly the sample size in multivariable models, which can impact its comparability to the sample used in the basic adjusted models and can create a bias. Participants excluded from analysis might be important, and the information they carry regarding the other variables could be valuable.

To address this issue in **Paper III**, we classified people with missing information in a new category for each of the variables for which most people had missing information. For the MHT variable, 110 women (8.3%) were classified as having missing values, whereas 33 women (2.5%) had missing information for the parity variable. All the other variables were either complete or had very few missing values. Both a complete case analysis (default) and an analysis including missing values in a separate category were performed, and the results were compared. A similar procedure was performed in **Paper IV**.

For the main analyses in both **Papers I** and **II**, complete case analyses were performed due to the low percentage of missing values.

Power calculations

The power calculations conducted during the planning and design of the MDC revealed that when 283 cases were accumulated, the study would have sufficient statistical power (i.e., 80% and α=0.05) to identify a risk gradient of 1 to 1.75 over the quintiles of a nutrient's intake (assuming a true risk gradient of 1 to 3 and a 0.6 correlation coefficient for the relative validity of the dietary variable). This number of cases was reached in 1999, and due to the reasonably high relative validity of most of the dietary variables, the power to examine the hypotheses was reached.

In **Paper I**, power calculations performed before the study was conducted revealed that if the observed ICC was 0.55 with three measurement points and the sample size was n=50, we would be able to exclude (with a power of 92.4%) the possibility that the true ICC was in fact less than 0.35 (Source: computer software StudySize 2.0, CreoStat HB).

5. Results

5.1. Characteristics of the study participants

Longitudinal and cross-sectional studies (Papers I and II)

In a subsample of the MDC (Paper I)

The analytical study population of **Paper I** consisted of 95 people, of whom 49 were men (approximately 52%). This study sample was constituted of slightly younger participants of the MDC cohort. The men had higher BMI ($p<0.05$) and tended to include a greater proportion of smokers, to be less educated and to consume higher levels of alcohol than the women (**Table 5**).

Table 5. Characteristics of the study (I) participants and at MDC baseline divided by gender

	Study sample (2010-11)		Original MDC cohort[a] (1991-96)	
	Men	**Women**	**Men**	**Women**
N (%)	49 (51.6)	46 (48.4)	11,063 (39.4)	17,035 (60.6)
Age (y), mean ± SD at baseline	67.9 ± 1.8	64.8 ± 3.1	75.5 ± 6.7[b] 58.7 ± 7.1	73.4 ± 8.0[b] 56.9 ± 7.9
BMI (kg/m^2), mean ± SD	27.6 ± 3.9	25.5 ± 4.2	26.3 ± 3.5	25.4 ± 4.2
Weight (kg), mean ± SD	85.6 ± 13.5	69.2 ± 13.0	81.7 ± 12.1	68.0 ± 11.7
Total energy intake[c] (kcal/d), mean ± SD	1953 ± 529	1680 ± 399	2534 ± 674	1980 ± 503
Total fat (E %)	35.5 ± 6.5	35.4 ± 6.7	39.8 ± 6.2	38.5 ± 6.0
Fiber (g/1000 kcal)	11.8 ± 4.0	12.8 ± 4.9	8.6 ± 2.5	9.7 ± 2.8
Education, *n* (%)				
Primary school	20 (40.8)	6 (13.1)	5,078 (46.0)	6,699 (39.4)
Elementary school	4 (8.1)	18 (39.1)	2,169 (19.7)	5,163 (30.4)
High school	9 (18.4)	4 (8.7)	13,08 (11.8)	1,183 (7.0)
University	16 (32.7)	18 (39.1)	11,035 (22.5)	16,992 (23.2)
Alcohol consumption, *n* (%)				
Zero consumers	8 (17.0)	9 (20.9)	487 (4.4)	1,297 (7.6)
Low	14 (29.8)	22 (51.2)	7,360 (66.5)	12,951 (76.0)
Medium	17 (36.2)	8 (18.6)	2,409 (21.8)	2,385 (14.0)
High	8 (17.0)	4 (9.3)	807 (7.3)	402 (2.4)
Tobacco use status, *n* (%)				

Active	8 (16.3)	6 (13.0)	3,684 (33.3)	4,872 (28.6)
Former	23 (46.9)	20 (43.5)	4,331 (39.1)	4,691 (27.5)
Never	18 (36.7)	20 (43.5)	3,048 (27.6)	7,472 (43.9)

[a]Baseline information retrieved from the MDC-cohort, between 1991 and 1996.

[b]Age of the 28,098 individuals at baseline, around the time of the beginning of the present study (01/11/2010).

[c]Total energy calculated from macronutrients, and no alcohol. Available dietary data for the study sample; men (n=47) and women (n=43).

In a subsample of the MDC-CC (Paper II)

The analytical study sample of **Paper II** included 667 participants, of whom 276 were men (41%). Lifestyle and socioeconomic characteristics as well as age and anthropometrics are described across the categories of the DQI-SNR in **Table 6** and **Table 7**.

Table 6. Socioeconomic and lifestyle characteristics of the sample population (from study II) across DQI-SNR categories (n=667)

N (%)	DQI-SNR			*p*[a]
	Low (0 or 1 point)	Medium (2 or 3 points)	High (4-6 points)	
No. of participants	77	324	266	
Gender				
Men	31 (40.3)	147 (45.4)	98 (36.8)	
Women	46 (59.7)	177 (54.6)	168 (63.2)	0.109
Educational level[b]				
Primary	44 (57.9)	194 (60.1)	141 (53.0)	
Secondary	19 (25.0)	74 (22.9)	69 (25.9)	
Upper secondary	9 (11.8)	42 (13.0)	35 (13.2)	
University	4 (5.3)	13 (4.0)	21 (7.9)	0.448
Employment status[c]				
Blue-collar workers	41 (55.4)	147 (47.6)	112 (43.8)	
White-collar workers	28 (37.8)	145 (46.9)	130 (50.8)	
Employers/self-employed	5 (6.8)	17 (5.5)	14 (5.4)	0.412
Smoking status[d]				
Never	34 (44.7)	137 (42.4)	128 (48.1)	
Former	12 (15.8)	117 (36.2)	91 (34.2)	
Current	30 (39.5)	69 (21.4)	47 (17.7)	**<0.001**
Alcohol consumption				
Zero	12 (15.6)	32 (9.9)	21 (7.9)	
Low	58 (75.3)	246 (75.9)	216 (81.2)	
Medium/high	7 (3.9)	46 (14.2)	29 (10.9)	0.181
Leisure PA (tertiles)[e]				
1 (lowest)	35 (45.5)	120 (37.5)	65 (24.5)	
2	17 (22.1)	106 (33.1)	98 (37.0)	
3 (highest)	25 (32.5)	94 (29.4)	102 (38.5)	**<0.001**
Household PA (tertiles)[f]				

1 (lowest)	34 (45.3)	121 (38.6)	83 (32.2)	
2	15 (20.0)	90 (28.8)	83 (32.2)	
3 (highest)	26 (34.7)	102 (32.6)	92 (35.6)	0.155
Work PA[g]				
Light	30 (40.5)	129 (41.1)	105 (41.0)	
Medium	32 (43.3)	140 (44.6)	99 (38.7)	
Heavy	12 (16.2)	45 (14.3)	52 (20.3)	0.378
Cohabitation status				
Living alone	18 (23.4)	79 (24.4)	68 (25.6)	
Cohabiting	59 (76.6)	245 (75.6)	198 (74.4)	0.907
Past food habit change				
Yes	14 (18.2)	76 (23.5)	101 (38.0)	
No	63 (81.8)	248 (76.5)	165 (62.0)	**<0.001**
Season				
Winter	34 (44.2)	103 (31.8)	77 (29.0)	
Spring	8 (10.4)	58 (17.9)	61 (22.9)	
Summer	7 (9.0)	39 (12.0)	25 (9.4)	
Fall	28 (36.4)	124 (38.3)	103 (38.7)	0.091
Diseases at baseline				
Diabetes	1 (1.3)	14 (4.3)	18 (6.8)	0.115
Hypertension	60 (77.9)	264 (81.5)	218 (82.0)	0.720
Cardiovascular event	2 (2.6)	6 (1.9)	14 (5.3)	0.065

[a]Chi-square was used to test for differences between the groups.

Number of participants with missing data: b, n=2; c, n=28; d, n=2; e, n=5; f, n=21; g, n=23.

Because a larger percentage of women than men adhered to the recommendations, gender adjustments were included in the appropriate analyses. Anthropometric characteristics did not vary across DQI-SNR categories (**Table 7**).

Table 7. Anthropometric characteristics of the sample population (from study II) across DQI-SNR categories (n=667)

Mean±SD	DQI-SNR			*p*	*p*-trend
	Low (0 or 1 point)	Medium (2 or 3 points)	High (4-6 points)		
Age	65.9±1.0	65.7±1.1	65.5±1.1	**<0.05**	**<0.05**
BMI	25.9±4.1	26.2±3.9	26.5±4.0	0.435	0.202
Body fat %[a]	28.5±7.5	27.9±7.4	28.5±7.1	0.687	0.416
Waist circumference (cm)	84.0±12.0	86.4±12.6	85.0±12.5	0.507	0.575
Waist-to-hip ratio	0.84±0.08	0.86±0.09	0.85±0.10	0.102	0.256

A univariate general linear model (GLM) was used to test for mean differences (*p*) across DQI-SNR for age (adjusting for gender), body mass index (BMI), body fat percentage, waist circumference and waist to hip ratio, adjusting for gender and age.

p-for-trend (*p*-trend) was analyzed using DQI-SNR as a continuous covariate in the GLM model, with the same adjustments.

Number of participants with missing data: a, n=3.

Nested case-control studies (Papers III and IV)

A nested case-control study of the postmenopausal women in the MDC comprised 446 cases and 910 controls with available laboratory data. In **Paper III**, only the controls matched directly with the 446 cases were used due to the choice of statistical method. The cases tended to be more overweight and to have higher WHR than the controls (**Table 8**).

Table 8. Baseline characteristics of the breast cancer cases (n=446) and matched controls (n=885 in study III and n=910 in study IV) aged 55-73 years from the Malmö Diet and Cancer cohort

Mean±SD	**Cases (n=446)**	**Controls[1] (n=885)**	**Controls[2] (n=910)**	**OR (95%CI)**	***p***
Age (y)	62.0±4.8	62.0±4.8	62.0±4.8		(m.v.)
Week of blood sampling	139±77	139±77	139±77		(m.v.)
Height (m)	1.63±0.05	1.63±0.06	1.63±0.06		0.20
Waist-to-hip ratio (WHR)*[§]	0.80±0.07	0.79±0.06	0.79±0.06		**0.04**
Age at menarche (*a*)[§]	13.6±1.1	13.6±1.1	13.7±1.5		0.98
Age at menopause (*b*)	50.2±4.8	50.0±4.6	50.0±4.6		0.27
Age at birth of first child[§]	24.7±1.2	24.4±1.2	24.7±4.4		0.16
Breastfeeding time (mo.)[§]	8.4±2.2	8.1±2.3	8.8±9.2		0.38
Time span *b-a* (y)	36.5±5.0	36.2±4.8	36.2±4.8		0.35
Menstrual cycles (y)	34.4±5.0	34.0±5.0	34.1±5.0		0.16
N (%)					
Education					
Primary school	196 (44.2)	442 (50.1)	454 (49.9)	1 (ref.)	
Elementary school	155 (35.0)	278 (31.4)	287 (31.6)	1.26 (0.97-1.64)	
High school	18 (4.1)	47 (5.3)	47 (5.2)	0.87 (0.49-1.55)	
University	74 (16.7)	117 (13.2)	121 (13.3)	**1.46 (1.03-2.08)**	0.06
Smoking status					
Never smoker	215 (48.2)	458 (51.8)	473 (52.0)	1 (ref.)	
Former smoker	139 (31.2)	240 (27.1)	245 (27.0)	1.26 (0.96-1.65)	
Current smoker	92 (20.6)	186 (21.1)	191 (21.0)	1.05 (0.78-1.41)	0.50
Alcohol consumption					
Zero consumers	30 (6.7)	77 (8.7)	78 (8.6)	1 (ref.)	
Low (<15 g/d)	351 (78.7)	686 (77.5)	707 (77.7)	1.31 (0.83-2.07)	
Medium (15-30 g/d)	49 (11.0)	114 (12.9)	117 (12.9)	1.15 (0.66-2.01)	
High (>30 g/d)	16 (3.6)	8 (0.9)	8 (0.9)	**4.92**	

				(1.91-12.7)	0.06
Leisure time PA					
Tertile 1	158 (35.6)	279 (32.0)	287 (32.0)	1 (ref.)	
Tertile 2	146 (32.9)	292 (33.5)	300 (33.5)	0.88 (0.67-1.17)	
Tertile 3	140 (31.5)	300 (34.5)	309 (34.5)	0.82 (0.62-1.09)	0.18
BMI					
Under/normal weight (<25)	178 (39.9)	406 (45.9)	420 (46.2)	1 (ref.)	
Overweight (25-30)	195 (43.7)	342 (38.7)	349 (38.4)	**1.29 (1.01-1.67)**	
Obese (>30)	73 (16.4)	136 (15.4)	140 (15.4)	1.24 (0.89-1.75)	0.09
Parity					
0	59 (13.6)	99 (11.5)	101 (11.4)	1 (ref.)	
1	91 (21.0)	185 (21.4)	194 (21.8)	0.82 (0.55-1.25)	
2	181 (41.7)	351 (40.6)	361 (40.6)	0.87 (0.60-1.27)	
3	76 (17.5)	142 (16.4)	145 (16.3)	0.91 (0.59-1.40)	
≥ 4	27 (6.2)	87 (10.1)	88 (9.9)	**0.52 (0.30-0.89)**	0.11
Oral contraceptives					
No use	262 (58.9)	562 (63.6)	328 (36.9)	1 (ref.)	
Reported use	183 (41.1)	322 (36.4)	581 (63.9)	1.24 (0.97-1.58)	0.08
MHT					
No use	284 (70.0)	654 (80.2)	165 (19.7)	1 (ref.)	
Current use	122 (30.0)	161 (19.8)	672 (80.3)	**1.80 (1.35-2.42)**	**<0.001**

Controls used in study [1]III and in study [2]IV.

[¥]The ORs and 95% CIs were calculated with conditional logistic regression analysis, in a basic model (unadjusted). The first category was used as the reference. The *p*-values refer to *p*-for-linear-trend as categorical variables were introduced linearly. The *p*-values for continuous variables were derived from a GLM model using paired data. m.v., matching variables. [§]Due to the skewedness of these variables, ln-transformation was used, and means and SD values were back-transformed.* Due to very wide confidence intervals, these variables were introduced as continuous sextiles in the model (this did not affect the p-value greatly).

In this study, higher alcohol consumption, holding a university degree, currently using MHT, and being overweight were associated with increased breast cancer risk. For additional descriptions of how the participants' characteristics varied across levels of exposure (i.e., biomarkers of inflammation and food patterns), refer to **Table 2** in **Paper III** and **Table 3** in **Paper IV**.

Aim 1

We explored the reliability and reproducibility of selected biomarkers of low-grade inflammation both in fasting and in non-fasting blood samples (**Paper I**). Three fasting and three non-fasting samples were assessed for the following biomarkers: ox-LDL, IL-1β, IL-6, IL-8 and TNF-α. The interval range between the first and the last fasting measurements varied between 24 and 49 days for women and 26 to 40 days for men, whereas for the non-fasting measurements, it varied from 26 to 47 days for women and from 27 to 41 days for men.

No clear patterns emerged when we plotted the difference in the biomarker concentrations from the first to the last measurements separately for fasting and non-fasting samples against the range of days between measurements for each individual (**Figures 14** and **15**), indicating that biomarker levels were not dependent on time span.

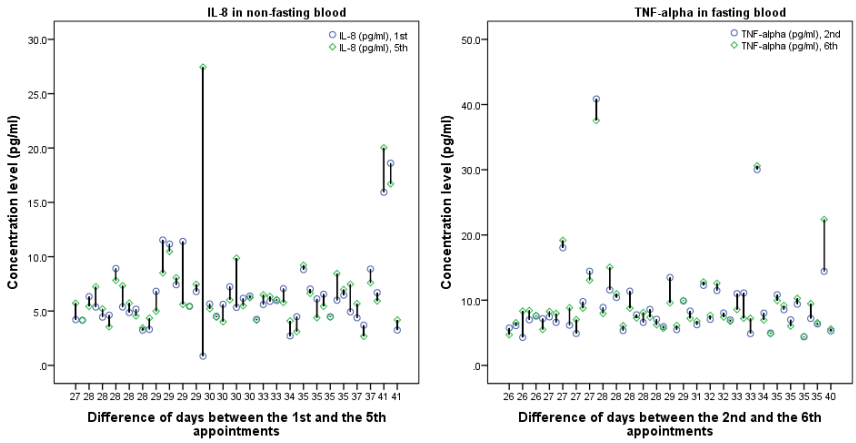

Figure 14. Difference in biomarker concentration levels plotted against the number of days between measurements for men (n=49)
Difference in the biomarker concentration levels between the first and the last measurement (1st and 5th visits for non-fasting samples and 2nd and 6th visits for fasting samples) plotted against the number of days between measurements for each individual. The biomarker with the lowest single-measurements ICC, IL-8 in non-fasting blood, is represented on the left, whereas the biomarker with the highest single-measurements ICC, TNF-α in fasting samples, is represented on the right side of the figure.

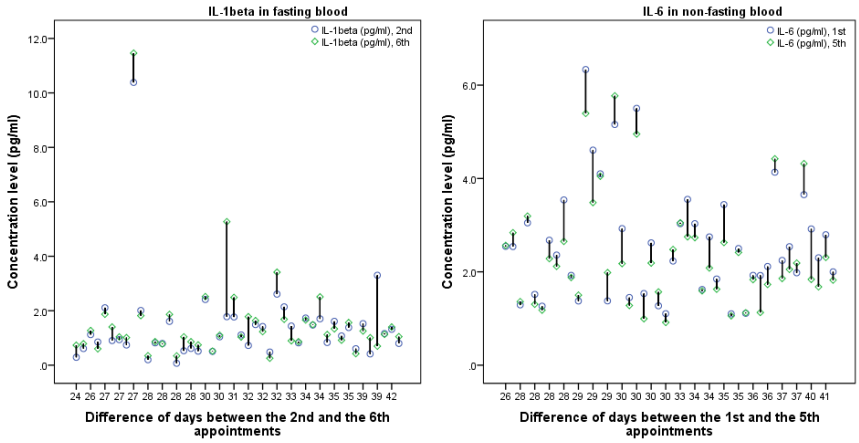

Figure 15. Difference in biomarker concentration levels plotted against the number of days between measurements for women (n=46)
Difference in the biomarker concentration levels between the first and the last measurement (1st and 5th visits for non-fasting samples and 2nd and 6th visits for fasting samples), plotted against the number of days between measurements for each individual. The biomarker with the lowest single-measurements ICC, IL-1β in fasting blood, is represented on the left, whereas the biomarker with the highest single-measurements ICC, IL-6 in non-fasting samples, is represented on the right side of the figure.

A complete description of median levels for the biomarkers on each occasion and Spearman correlations between biomarker measurements on the different occasions can be found in **Table 2** of **Paper I**. The correlations varied between 0.65 and 0.92 in the women and 0.48 and 0.89 in the men (all $p<0.01$). A scatterplot showing the lowest and highest correlation coefficients is presented in **Figure 16** (information not included in **Paper I**).

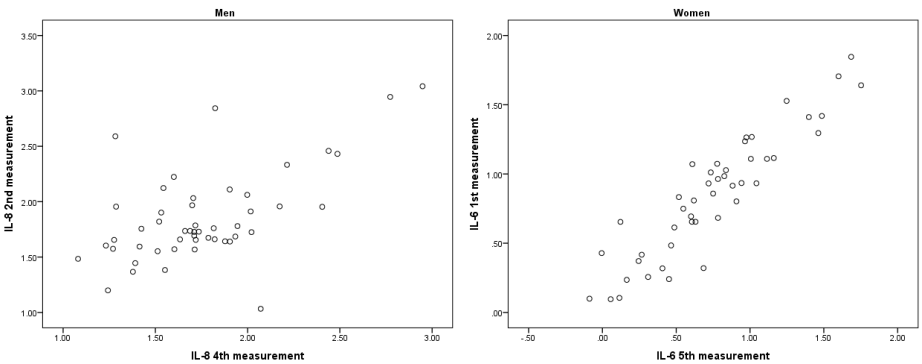

Figure 16. Biomarker concentrations (ln-transformed, pg/ml) for the lowest correlation coefficient (IL-8 measured at the 2nd and 4th visits in the men) and the highest (IL-6 measured at the 1st and 5th visits, in the women)

Table 9. The means and SD for repeated measures of the biomarkers, coefficients of variation (CV), and intraclass correlation coefficients (ICC) of each biomarker, for both fasting (F) and non-fasting (NF) blood samples in men (n=49) and women (n=46)

		Mean± SD	CV_B (%)	CV_W (%)	Index[a]	N[b]	ICC[c] (95% CI)	ICC[d] (95% CI)	α[e]
Men									
NF	Ox-LDL	56.6±13.2	42	10	0.19	1	0.85 (0.78-0.91)	0.95 (0.91-0.97)	0.95
	IL-1β	1.59±1.57	178	44	0.24	13	0.71 (0.58-0.81)	0.88 (0.80-0.93)	0.88
	IL-6	3.19±2.40	126	43	0.34	13	0.61(0.46-0.74)	0.82 (0.71-0.89)	0.82
	IL-8[f]	6.47±2.92	67	18	0.20	2	0.78 (0.67-0.86)	0.91 (0.86-0.95)	0.91
	TNF-α	9.03±5.97	92	31	0.32	7	0.66 (0.52-0.78)	0.85 (0.76-0.91)	0.85
F	Ox-LDL	57.0±12.9	41	9	0.18	1	0.86 (0.78-0.91)	0.95 (0.91-0.97)	0.95
	IL-1β	1.62±1.60	207	40	0.19	11	0.77 (0.67-0.85)	0.91 (0.86-0.95)	0.91
	IL-6	3.66±2.73	127	31	0.24	7	0.76 (0.64-0.84)	0.90 (0.84-0.94)	0.90
	IL-8	6.76±3.12	67	22	0.28	4	0.69 (0.55-0.80)	0.87 (0.79-0.92)	0.87
	TNF-α	9.39±5.94	84	14	0.14	2	0.90 (0.84-0.94)	0.96 (0.94-0.98)	0.96
Women									
NF	Ox-LDL	56.6±10.2	32	12	0.33	1	0.67 (0.53-0.79)	0.86 (0.77-0.92)	0.86
	IL-1β	1.41±1.48	181	34	0.17	8	0.81 (0.71-0.88)	0.93 (0.88-0.96)	0.93
	IL-6	2.45±1.15	87	16	0.16	2	0.87 (0.80-0.92)	0.95 (0.92-0.97)	0.96
	IL-8	5.61±1.89	54	19	0.26	3	0.69 (0.55-0.80)	0.87 (0.79-0.92)	0.87
	TNF-α	7.70±3.51	76	19	0.22	3	0.81 (0.71-0.88)	0.93 (0.88-0.96)	0.93
F	Ox-LDL	57.1±10.7	35	10	0.23	1	0.79 (0.69-0.87)	0.92 (0.87-0.95)	0.92
	IL-1β	1.48±1.55	220	54	0.24	20	0.66 (0.52-0.78)	0.85 (0.76-0.91)	0.86
	IL-6	2.68±1.44	103	20	0.18	3	0.85 (0.77-0.91)	0.95 (0.91-0.97)	0.95
	IL-8	5.95±2.08	59	19	0.28	3	0.71 (0.57-0.81)	0.88 (0.80-0.93)	0.88
	TNF-α	8.10±3.77	75	17	0.19	2	0.84 (0.75-0.90)	0.94 (0.90-0.96)	0.94

[a]Index of individuality; [b]Number of measurements/days needed to estimate the mean serum concentration in individuals within 20% with 90% CI; [c]single-measures ICC; [d]average-measures ICC; [e]Cronbach's alpha; [f]analyses that excluded one individual (n=48).

A great majority of the single-measures ICCs in non-fasting samples were above 0.70, and all were above 0.60 (**Table 9**), independent of the participant's gender. Similar results were observed for the fasting samples. The reliability measures based on the three measurement points (average-measures ICCs) were all above 0.85. Most of the biomarkers would require four measurement points or fewer to estimate the homeostatic set point for the individuals within ±20% and with a 90% CI (represented in the table as "N") as most CV_I values were below 20% (ranging from 7% to 53%). A wide CV_W could translate into a low ICC; however, we observed that when the CV_B was proportionally larger, the average-measures ICC tended to be high (**Table 9**). According to the Cronbach's alpha estimates based on the three measurement points (all above 0.80), internal validity was very high. Finally, the mean levels of each biomarker (measured on 3 occasions) in non-fasting samples were highly correlated with the same biomarkers in fasting samples, with correlation coefficients ranging from 0.81 to 0.98 (all $p<0.001$; **Table 4** in **Paper I**).

Aim 2

In **Paper II**, we examined the associations between diet quality and biomarkers of systemic inflammation. We observed significant inverse associations between high diet quality and several soluble and cellular biomarkers of low-grade systemic inflammation (**Table 10**). Biomarkers such as S100A8/A9, hs-CRP, WBC, TNF-α, neutrophils, lymphocytes, and $CD14^{+}CD16^{++}$ count were significantly and inversely associated with the DQI-SNR score, independent of age, gender, total energy intake, smoking status, PA level, waist circumference and season. In contrast, the percentage of $CD14^{++}CD16^{+}$ was significantly and positively associated with DQI-SNR. All estimates remained virtually the same after past food habit changers were excluded from the sensitivity analysis. Partial correlation coefficients between the biomarkers of inflammation and categories of the different DQI-SNR components (SFAs, PUFAs, fish and shellfish, fiber, fruits and vegetables, and sucrose) can be found in **Table 3** of **Paper II**. Some components seemed to drive associations between inflammatory markers and the DQI-SNR; fiber consumption was associated with the circulating cells (both total and differential), whereas fish and shellfish consumption was associated with TNF-α and the percentage of $CD14^{++}CD16^{+}$. The total index score revealed associations that were not apparent from any of its components, with hs-CRP and S100A8/A9. A backward linear regression model showed that the following biomarkers were independently associated with DQI-SNR: hs-CRP, S100A8/A9, WBC, TNF-α, neutrophils, lymphocytes, mixed cells, $CD14^{++}CD16^{+}$ (%), and $CD14^{+}CD16^{++}$ count (all $p<0.05$).

Table 10. Adjusted means of biomarker concentrations across categories of DQI-SNR

		DQI-SNR			Models		
		Low (0 or 1 point)	Medium (2 or 3 points)	High (4-6 points)	I	II	III
					(*p*-values)		
hs-CRP[a]	mg/dL	0.17	0.18	0.15	**<0.05**	**<0.05**	**<0.05**
S100A8/A9[a]	µg/mL	1.64	1.54	1.46	**<0.05**	**<0.05**	**<0.05**
IL-1β[b]	pg/mL	0.57	0.55	0.55	0.764	0.749	0.756
IL-8[a]	pg/mL	4.94	5.40	5.10	0.784	0.902	0.825
TNF-α[a]	pg/mL	5.08	4.78	4.56	**<0.05**	**<0.05**	**<0.05**
WBC	10^6 cells/mL	6.82	6.06	5.93	**<0.001**	**<0.01**	**<0.01**
Neutrophils[a]	10^6 cells/mL	4.07	3.59	3.48	**<0.01**	**<0.05**	**<0.01**
Lymphocytes	10^6 cells/mL	1.99	1.81	1.79	**<0.01**	**<0.05**	**<0.05**
Mixed cells	10^6 cells/mL	0.51	0.50	0.48	0.076	0.149	0.137
$CD14^{++}$	%	65.8	63.1	65.2	0.544	0.572	0.556
$CD14^{++}CD16^{+a}$	%	3.1	3.6	3.6	**<0.05**	0.060	0.064
$CD14^{+}CD16^{++b}$	%	7.4	7.4	6.8	0.133	**<0.05**	**<0.05**
$CD14^{+}CD16^{+}$ tot	%	11.4	12.0	11.3	0.492	0.240	0.232
$CD14^{++}$	cells/µL	344	313	318	0.186	0.287	0.285
$CD14^{++}CD16^{+a}$	cells/µL	15	18	16	0.571	0.534	0.538
$CD14^{+}CD16^{++b}$	cells/µL	36	35	31	**<0.05**	**<0.01**	**<0.01**
$CD14^{+}CD16^{+}$ tot[a]	cells/µL	51	51	47	0.106	0.071	0.066

Adjusted means (estimated marginal means) were derived from Model I and back-transformed from logarithmic[a] or square root[b] transformations.

Mg, milligrams; dL, deciliter; %, percentage; µg, micrograms; mL, milliliter; pg, pictograms; cells/µL, cell count per microliter of whole blood; 10^6 cells/mL, cell count multiplied by 10^6 per mL whole blood.

GLM was used to determine the *p*-for-trend (*p*-values), using the DQI-SNR as a continuous covariate with the following adjustments; Model I: age, gender, total energy. Model II: I + smoking + PA. Model III: II + waist circumference + season.

Aim 3

The association between selected biomarkers of inflammation and postmenopausal breast cancer was investigated in **Paper III**. The biomarkers investigated in this study were the following: ox-LDL, IL-1β, IL-6, TNF-α, WBC, neutrophils and lymphocytes. We found that ox-LDL, IL-1β, and TNF-α were associated with postmenopausal breast cancer after adjustment for matching factors (model I). The women in the highest tertiles of TNF-α and ox-LDL were at lower risk than those in the lowest tertiles: 0.60 (0.42-0.86) and 0.71 (0.52-0.96), respectively (**Table 11**). The opposite association was observed for IL-1β: the women in the highest category were at increased risk compared to those in the lowest, 1.52 (1.01-2.30; **Table 11**). The observed associations appeared to be independent of age, WHR, BMI, parity, MHT, smoking status, PA level, and education (model IV). In a model that included the three biomarkers that showed significant associations with breast cancer and with basic adjustments (model I), the estimates remained similar, indicating a possible independent effect for each of the biomarkers.

In the sensitivity analyses, when season or plate number were added to the models, the estimates and *p*-values remained the same. When the women diagnosed within the first years of follow-up were excluded, the magnitude of the estimates remained, although most of the associations were attenuated (i.e., they were no longer significant). This happened after the exclusion of cases diagnosed until the 3rd year of follow-up (n=83). Finally, in an *ad hoc* analysis performed when the missing values for parity and MHT were treated as a separate category, most of the associations remained. The exception was the highest category of IL-1β, which was no longer significant in models III and IV.

Table 2 of **Paper III** provides a thorough description of baseline characteristics of the controls (n=885) across levels of inflammation markers, and **Table 3** of the same paper provides a description of the correlation matrix for biomarkers and obesity indicators. The participants in the highest tertile of ox-LDL were older, had higher WHR and higher BMI and were less educated, whereas the highest tertile of TNF-α was associated with younger participants with higher BMI and a higher proportion of MHT use. The women in the highest category of IL-1β had higher WHR and included a higher proportion of MHT users. All biomarkers were positively and significantly associated with WHR; however, WBC, neutrophils and IL-1β were not associated with BMI, unlike the other biomarkers (**Table 3** of **Paper III**). We observed that IL-6 was positively associated with all other biomarkers except ox-LDL. Ox-LDL was associated positively but weakly with IL-8, TNF-α and lymphocyte count, whereas TNF-α was positively and

significantly associated with all biomarkers except circulating cells (both total and differential).

Aim 4

We further examined the role of obesity in the association between low-grade inflammation markers and postmenopausal breast cancer (**Paper III**). For that purpose, we analyzed obesity under 2 conditions: confounding and interaction (or effect modification; illustrated in **Figure 17**). Furthermore, we used two measures of obesity associated with total and abdominal adiposity: BMI and WHR, respectively. These measures were not strongly correlated with each other (r=0.45, p<0.001), indicating that they may capture two different aspects of obesity; therefore, they were included in the multivariable models together.

In this study (**Paper III**), obesity was associated with both exposure (inflammation markers) and the outcome (breast cancer). Therefore, BMI was added to model II in a first step (**Table 11**). When BMI was added to the models, no major changes in the associations with breast cancer were observed for either the biomarkers or for BMI, probably indicating that the associations were independent.

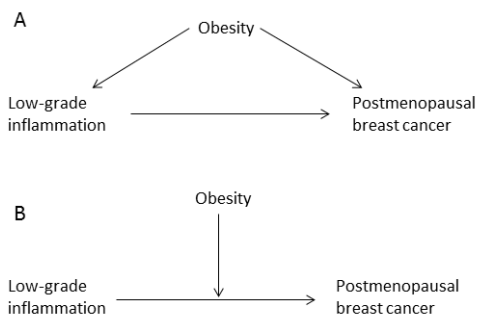

Figure 17. Illustration of possible associations
This figure illustrates the possible associations between exposure (low-grade inflammation) and outcome (breast cancer), considering obesity as either a confounder (A) or an effect modifier (B).

We investigated the possible interactions between the obesity indicators and each of the biomarkers that was significantly associated with postmenopausal breast cancer. No interaction was significant, possibly indicating that associations between biomarkers and breast cancer do not vary across categories of obesity.

Table 11. Associations between inflammation markers and the risk of postmenopausal breast cancer in a nested case-control study of the MDC

			Model I	Model II	Model III	Model IV
	Median	Case/control	OR (95% CI)			
Ox-LDL						
Tertile 1 (n=445)	46.6	168/277	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=442)	62.0	140/302	**0.75** (0.57-1.00)	**0.74** (0.56-0.99)	**0.65** (0.47-0.90)	**0.65** (0.47-0.90)
Tertile 3 (n=444)	85.9	138/306	**0.71** (0.52-0.96)	**0.69** (0.51-0.94)	**0.65** (0.46-0.91)	**0.63** (0.45-0.89)
p-trend			**0.03**	**0.02**	**0.01**	**0.01**
IL-1β						
0 (n=340)	0	105/235	1 (reference)	1 (reference)	1 (reference)	1 (reference)
<LLOD (n=640)	0.10	208/432	1.08 (0.80-1.44)	1.08 (0.80-1.45)	1.08 (0.79-1.45)	1.09 (0.78-1.51)
Cat 1 (n=176)	0.22	65/111	1.45 (0.94-2.24)	1.45 (0.93-2.25)	1.58 (0.99-2.53)	**1.62** (1.00-2.62)
Cat 2 (n=175)	0.76	68/107	**1.52** (1.01-2.30)	1.51 (0.99-2.28)	**1.64** (1.02-2.65)	**1.71** (1.05-2.79)
p-trend			**0.02**	**0.03**	**0.02**	**0.01**
IL-6						
Tertile 1 (n=446)	0.80	159/287	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=442)	1.45	138/304	0.80 (0.59-1.07)	0.76 (0.56-1.03)	0.76 (0.55-1.05)	0.75 (0.54-1.05)
Tertile 3 (n=443)	2.62	149/294	0.87 (0.64-1.18)	0.80 (0.58-1.10)	0.83 (0.58-1.18)	0.80 (0.56-1.15)
p-trend			0.40	0.19	0.32	0.25
IL-8						
Tertile 1 (n=446)	3.10	149/297	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=445)	5.12	153/292	1.06 (0.75-1.50)	1.04 (0.73-1.48)	1.08 (0.74-1.57)	1.13 (0.77-1.65)
Tertile 3 (n=440)	7.61	144/296	0.97 (0.67-1.41)	0.96 (0.66-1.41)	1.05 (0.69-1.59)	1.09 (0.71-1.66)
p-trend			0.80	0.78	0.86	0.76
TNF-α						
Tertile 1 (n=446)	1.40	161/285	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=439)	2.25	156/283	0.86 (0.62-1.19)	0.82 (0.59-1.14)	0.93 (0.65-1.33)	0.91 (0.63-1.32)
Tertile 3 (n=446)	3.28	129/317	**0.60** (0.42-0.86)	**0.56** (0.38-0.81)	**0.65** (0.42-0.98)	**0.65** (0.43-0.99)
p-trend			**0.01**	**0.01**	**0.03**	**0.04**
WBC						

Tertile 1 (n=448)	4.90	145/303	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=447)	6.00	152/295	1.07 (0.81-1.42)	1.07 (0.80-1.42)	1.03 (0.75-1.41)	1.01 (0.73-1.40)
Tertile 3 (n=436)	7.60	149/287	1.08 (0.81-1.44)	1.06 (0.80-1.41)	0.96 (0.70-1.32)	0.93 (0.67-1.30)
p-trend			0.59	0.69	0.81	0.66
Lymphocytes						
Tertile 1 (n=444)	1.50	142/302	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=445)	1.90	148/311	0.83 (0.62-1.10)	0.81 (0.61-1.08)	0.77 (0.56-1.05)	0.79 (0.57-1.09)
Tertile 3 (n=427)	2.50	156/271	0.97 (0.74-1.27)	0.94 (0.72-1.24)	0.97 (0.71-1.30)	0.94 (0.68-1.28)
p-trend			0.81	0.66	0.76	0.62
Neutrophils						
Tertile 1 (n=486)	2.70	171/315	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Tertile 2 (n=392)	3.60	121/271	1.00 (0.75-1.33)	1.00 (0.75-1.33)	1.01 (0.74-1.38)	1.02 (0.74-1.40)
Tertile 3 (n=452)	4.80	154/298	1.21 (0.91-1.62)	1.20 (0.90-1.60)	1.06 (0.77-1.46)	1.04 (0.74-1.46)
p-trend			0.19	0.23	0.72	0.82

Breast cancer risk (OR) was calculated using conditional logistic regression with the lowest category (Tertile 1) as the reference group. The adjustment models included age and week of blood sampling (model I); BMI (model II); WHR, MHT and parity (model III); and smoking, alcohol, PA and education (model IV).

Aim 5

In **Paper IV**, we aimed to identify food patterns associated with three biomarkers of low-grade inflammation (ox-LDL, IL-1β, TNF-α) using the RRR method. Three food patterns were identified that together explained a total of 3.2% of the variation in the response variables, and a total of 8.4% of the variation in the food predictor variables. Food pattern 1 (Factor 1) explained 1.79% of the variation in the inflammation markers, whereas Factor 2 explained an additional 0.95% and Factor 3 a further 0.51%. Each food pattern was characterized by different food groups that loaded positively or negatively (**Table 12**).

Table 12. Top positive and negative loadings characterizing the three food patterns

	Factor 1	Factor 2	Factor 3
Positive loadings	High-fiber crisp bread	Soft drinks	Soft margarines
	Yogurt	Fatty meats	Fatty fish
		Cottage cheese	Cottage cheese
		Lean meats	Lean meats
			Coffee
Negative loadings	Solid margarines	Fruits	Dried soups/sauces
	Whole milk	Lean fish	Fiber-rich bread
	Lean meats	Cottage cheese	Yogurt
	Lean fish	Shellfish	
	Cereals	Yogurt	
	Cream		

Food groups with the highest absolute factor loadings (>0.20).

Table 2 of **Paper IV** provides a thorough description of the associations between food groups and inflammation markers. Few of the food groups that loaded high in the food patterns did not associate with any of the inflammation markers. These were soft drinks, cereals, shellfish, fatty meats, fatty fish, fiber-rich bread and dried soups/sauces.

Food Pattern (or Factor) 1 explained most of the variation in TNF-α, whereas FP 2 added mostly to the explanation of the ox-LDL variation, and FP 3 added to the explanation of the variation in IL-1β. Correlation coefficients are shown in **Table 13** (this information was extracted from **Table 2** of **Paper IV**). **Figures 18**, **19** and **20** show the variation of the biomarker concentration levels across the tertiles of each food pattern in detail.

Two biomarkers (IL-1β and TNF-α) were positively correlated with each other, r=0.16 (p<0.01).

Table 13. Pearson correlations between food patterns (Factors 1, 2 and 3) and inflammation markers (response variables)

	Inflammation markers		
Food patterns	**Ox-LDL**	**IL-1β**	**TNF-α**
Factor 1	-0.09***	0.11***	0.18***
Factor 2	0.15***	0.06*	0.04
Factor 3	-0.02	0.10***	-0.07*

*p<0.05; **p<0.01; ***p<0.01. The inflammation markers were ln-transformed.

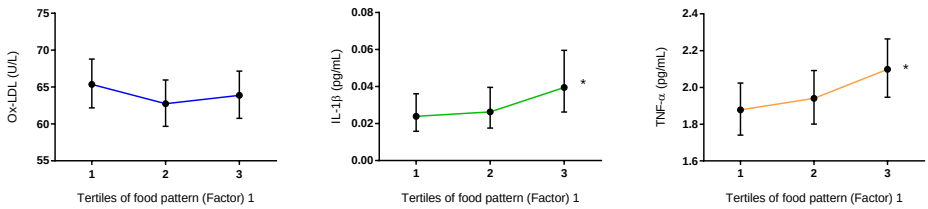

Figure 18. Biomarker concentration levels across the tertiles of Factor 1 (n= 1356)

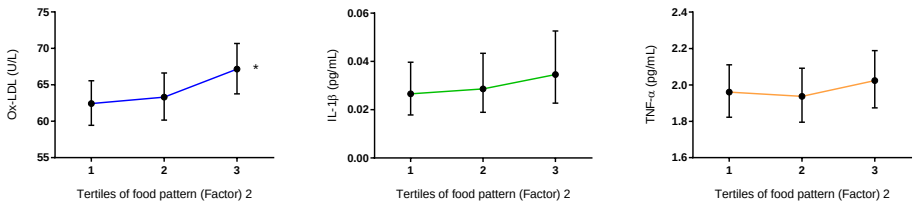

Figure 19. Biomarker concentration levels across the tertiles of Factor 2 (n= 1356)

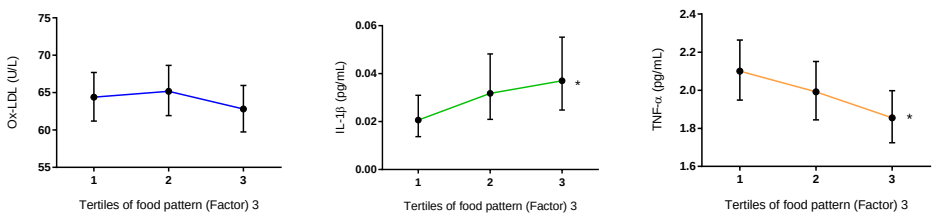

Figure 20. Biomarker concentration levels across the tertiles of Factor 3 (n= 1356)
The geometric mean concentrations and 95% CI of ox-LDL (in blue), IL-1β (in green), and TNF-α (in orange) by tertiles of food pattern scores (Factors 1, 2 and 3), adjusted for age, week of blood sampling, total energy intake, PA level, WHR, BMI, MHT, parity, alcohol consumption, smoking status and education level. Tests for trend (*) were significant for IL-1β and TNF-α in Factors 1 and 3, and for ox-LDL in Factor 2.

Aim 6

Finally, we assessed whether the specific food patterns were associated with postmenopausal breast cancer risk (**Paper IV**). No significant associations were observed between Factor 1 and Factor 2 in model I (with basic adjustments; **Table 14**). In contrast, Factor 3 was associated with increased risk; compared with the women in the lowest tertile, the women in the highest tertile had a 34% (2-77%) risk (**Table 14**). When adjusting for potential confounders other than alcohol intake, the observed associations were no longer significant. Similar estimates were observed in the sensitivity analysis after excluding women who reported having changed their diet and those who misreported their energy intake. When conditional logistic regression was performed, most of the estimates remained similar, but the association for the highest tertile of Factor 3 was not significant in model I.

A description of the participants' characteristics across tertiles of food patterns is available in **Table 3** of **Paper IV**. The highest tertile of Factor 1 was characterized by a lower proportion of never smokers, whereas the women in the highest tertile of Factor 2 had a higher WHR; additionally, a lower proportion were MHT users or had university degrees, and fewer were alcohol consumers. The women in the highest tertile of Factor 3 were heavier alcohol consumers.

Previous results from the MDC in this nested case-control study

In this study, we did not find any significant associations between omega-6 or fiber intakes and breast cancer, unlike previous studies in the MDC cohort [45] (1.21 (0.92-1.60) and 1.08 (0.78-1.48), respectively). In contrast, we found a strong association between the dried soups/sauces variable and an increased risk of breast cancer in the highest tertile compared with the lowest, after basic adjustments were made (1.41 (1.04-1.91)), which was in line with previous results [118, 119]. In model III, both dried soups/sauces and the Factor 3 estimates were strengthened when they were introduced together: 1.53 (1.07-2.16) and 1.41 (1.03-1.94) for the highest tertiles for both variables, respectively. The correlation coefficient for dried soups/sauces and Factor 3 was r=-0.31 (p<0.01).

Table 14. Risk of postmenopausal breast cancer associated with the tertiles of three food patterns (factors)

	OR (95%CI)			
Food patterns	**T1**	**T2**	**T3**	***p*-trend***
Factor 1				
Cases/controls	144/308	156/296	146/306	
Model I	1	1.13 (0.86-1.49)	1.03 (0.77-1.36)	0.86
Model II	1	1.11 (0.84-1.47)	1.02 (0.76-1.35)	0.91
Model III	1	1.23 (0.91-1.66)	0.99 (0.73-1.35)	0.96
Model IV	1	1.22 (0.90-1.64)	0.99 (0.73-1.35)	0.98
Factor 2				
Cases/controls	161/291	134/318	151/301	
Model I	1	0.76 (0.58-1.01)	0.90 (0.69-1.19)	0.47
Model II	1	0.86 (0.64-1.17)	0.90 (0.66-1.22)	0.49
Model III	1	0.87 (0.64-1.18)	0.90 (0.66-1.22)	0.49
Model IV	1	0.89 (0.65-1.20)	0.90 (0.66-1.22)	0.50
Factor 3				
Cases/controls	135/317	147/305	164/288	
Model I	1	1.13 (0.86-1.50)	**1.34 (1.02-1.77)**	**0.04**
Model II	1	1.16 (0.87-1.53)	**1.34 (1.01-1.77)**	**0.04**
Model III	1	1.19 (0.87-1.61)	1.29 (0.95-1.74)	0.10
Model IV	1	1.18 (0.87-1.60)	1.32 (0.97-1.78)	0.07

Unconditional logistic regression with the following adjustments;

Model I: age and week of blood sampling

Model II:

Factor 1: Model I + smoking status

Factor 2: Model I + WHR, MHT, alcohol intake, education, and leisure time physical activity (PA)

Factor 3: Model I + alcohol intake

Model III: Model I + smoking status, BMI, WHR, MHT, alcohol intake, education, parity, and PA

Model IV: Model III, excluding BMI and WHR

*the *p*-trend was obtained when the tertiles of food patterns were introduced as continuous variables in the models.

6. Discussion

The overall purpose of this thesis was to contribute to the understanding of possible mechanisms in the development of postmenopausal breast cancer. The major focus was on oxidative stress and low-grade inflammation pathways as a possible connecting link between diet and breast cancer.

There are two possible explanations for why our main hypothesis was not supported by our results. The simplest explanation is that there is in fact no real association between diet and breast cancer via inflammation. The other possibility is that, if there is an association, it may be too complex to capture in an observational setting.

The latter possibility opens the door to several alternative explanations: 1) perhaps we did not have enough power (especially across strata) to detect associations; 2) the measurement errors associated with assessing the exposures (especially regarding diet) and/or 3) possible unmeasured confounders might have masked the true associations; and 4) we were not looking in the right place (e.g., perhaps we did not use the most appropriate biomarkers, temporality issues, etc.). All alternative explanations warrant further investigation to support or disprove our hypothesis.

The following section aims to describe the strengths and weaknesses of the four studies included in this thesis, and to compare and contrast the results with previous studies in the field. Finally, suggestions for future undertakings within this field are provided.

6.1. Main findings and interpretation

In this thesis, we first examined the reliability of several biomarkers of low-grade inflammation used in our studies. We then inspected how biomarkers of low-grade inflammation were associated with overall diet quality, indicated by a diet quality index. The association between the biomarkers of low-grade inflammation and breast cancer risk was investigated, and the role of obesity was scrutinized. Lastly, food patterns representing diets associated with certain inflammation markers were examined in relation to breast cancer risk.

Reproducibility of biomarkers (Paper I)

Our results in this study indicate a high stability over time for all the included biomarkers (ox-LDL, IL-1β, IL-6, IL-8, and TNF-α) based on the average-measures ICCs with values above 0.80. The reliability was high for some of the biomarkers (ox-LDL and TNF-α for men and IL-6 and TNF-α for women), with ICCs above 0.80, but was reasonable for the other biomarkers based on single-measures ICCs with values above 0.60. A previous study suggested that ICCs above 0.65 should be considered reasonably reliable, as risk estimates would not be attenuated to a major extent [162]. In that case, a "true" RR of 1.5 would be observed as an estimated RR of 1.3 [162].

Our results were in line with previous studies indicating good reproducibility for these biomarkers [163-168]. Hofmann and colleagues found that IL-6 and TNF-α had an excellent reproducibility, and IL-8 a fair to good reproducibility [166]. Similar to our study, Hofmann's study included subjects aged between 55 and 70 years. However, while we examined reproducibility over a 2-month period, Hofmann's observations referred to reproducibility over a 5-year interval. The long-term variability of biomarker levels is of great interest, especially when examining diseases with a long latency period, such as cancer. Because the within-person variation is likely lower over shorter periods, we can expect our ICCs to be slightly overestimated. However, the ICCs for these biomarkers over 2 months and 5 years were very similar. It is therefore possible that these biomarkers are highly reliable over long periods, and thus are useful for investigating biomarker-disease associations. However, we cannot predict the extent to which having a somewhat homogenous sample (i.e., less between-person variation) might have counterbalanced the ICC overestimation.

Several studies have analyzed fasting samples [169, 170], although the reproducibility of such fasting samples remains to be explored. Our results indicate that the ICCs for biomarker measurements in fasting samples were

slightly higher than those for non-fasting samples. These differences between the ICCs for fasting and non-fasting samples could be the result of the composition of the meals consumed prior to blood sampling during non-fasting states. Additionally, the time variation between the meal and sampling could contribute to these differences and to the lower ICCs for non-fasting samples [171]. Such factors affect the within-person variation, potentially lowering the ICC estimates, especially among obese subjects [172]. However, the high correlation coefficients observed in our study between the biomarkers in fasting and non-fasting samples support the usefulness of the non-fasting samples.

A great advantage of this study was the ability to examine the reliability of biomarkers separately by gender. Previous investigations have combined men and women in the analyses, probably to accommodate smaller sample sizes [165, 166, 168], or have investigated just one gender [167, 173]. However, there are some indications of differences according to gender [174] and menopausal status [175]. Despite some variation, our study did not provide strong evidence of major gender differences as most biomarkers were reliable in both groups. This is ultimately of great value for future research examining men and women separately.

A study by Lee and colleagues examined the usefulness of a single biomarker measurement in a cohort of middle-aged Chinese men (n=48) [167]. Our results were consistent with Lee's, which observed high ICCs for IL-1β and IL-6 (0.77 and 0.73) and reasonable ICCs for IL-8 and TNF-α (0.51 and 0.48). Findings from both studies suggest that a single measurement of a biomarker with an increased sample size might be more useful in cohort studies than using repeated measurement points and fewer participants.

Several days/measurements would be needed to be able to estimate the homeostatic set point, within ±20% and with 90% CI, for most of the biomarkers included in this study. The exception was ox-LDL, which would require only one day/measurement. This is also illustrated by the high CV_W (within-person coefficient of variation) of these biomarkers. However, if we relax our assumptions, in line with other studies [176, 177], and estimate the "true" biomarker levels within 50% and with 80% CI, almost all the biomarkers would require only one day/measurement. In any case, as stated in previous sections of this thesis, epidemiology very often has no need to estimate the "true" mean biomarker levels of each individual, unlike in clinical practice. Instead, its main concern is the ability to rank individuals in relation to a certain exposure with sufficient discriminatory power.

Ideally, investigations using biomarkers such as cytokines should explore the reliability of these biomarkers. It is of special importance to perform reliability tests in the same cohort that will be used by future investigations, considering the possibility of different ICC estimations resulting not only from population

characteristics but also differences in laboratory [178] and analytical procedures [179-181]. The value of having repeated measures of these biomarkers during follow-up, albeit seemingly reasonable, remains yet to be explored. Nevertheless, our study supports the feasibility of using of these biomarkers in future studies that aim to relate biomarkers to disease.

Diet quality and low-grade inflammation (Paper II)

Our findings showed an inverse association between a high diet quality, defined according to a DQI-SNR index, and several cellular and soluble biomarkers of systemic inflammation. Many of the dietary components were associated with specific biomarkers. However, the DQI-SNR index was more strongly associated with the inflammation biomarkers than with each specific component separately. For example, IL-8 and TNF-α were significantly and negatively associated with SFA and fish and shellfish consumption, respectively, but only TNF-α was negatively associated with the DQI-SNR. In contrast, hs-CRP was not significantly associated with any of the components, but it was negatively associated with the DQI-SNR. These findings add to the importance of investigating diet as whole, as some health effects of diet can be lost if only a few foods or nutrients are investigated [103].

An important strength of our study was the inclusion of several cellular biomarkers, cytokines and other soluble components with the aim of enabling a better overall picture of the inflammatory state. In fact, most of the investigated biomarkers were significantly associated with diet quality in the same direction (i.e., negatively). The only exception was the percentage of intermediate monocytes ($CD14^{++}CD16^{+}$), which was positively and significantly associated with the DQI-SNR, possibly driven by the association with the fish and shellfish component. Not much is yet known regarding the role of these very short-lived intermediate monocytes [182]; therefore, we were not able to speculate, based on a biological model, why they were positively associated with a higher diet quality in our study. Nevertheless, important and commonly used markers of inflammation, such as hs-CRP (which is a widely used biomarker for CVD risk [183, 184]), were inversely associated with diet quality. Our results were in line with the literature [116]. Additionally, a previous investigation in the same study population found inverse associations between hs-CRP and specific nutrients [185]. In a previous report, diet quality was shown to influence the levels of TNF-α [186]. Our results were in the same direction because among the investigated cytokines, only TNF-α was inversely and significantly associated with overall diet quality. Additionally, WBC count, a classical marker of inflammation [187], and its components (i.e., neutrophils, lymphocytes and mixed cells) were negatively associated with diet quality in our study. Neutrophil activation is hypothesized to be reflected by the

inflammatory protein S100A8/A9 [135]. In our study, both neutrophils and S100A8/A9 protein were retained in a backward linear regression model, suggesting that adherence to a higher-quality diet might contribute both to reduced circulating levels of neutrophils and to a decrease in neutrophil activation. The potential role of both neutrophils and S100A8/A9 protein in the development of carotid artery disease, the risk of coronary events and cardiovascular death has been observed in the same cohort [135]. We can therefore speculate that a high-quality diet (closer to the recommendations) might help to protect individuals from the development of CVD disease via decreased inflammation.

The possible role of dietary intake in modulating inflammation and promoting cellular activation has been reported previously [188]. A "healthy" pattern, derived with principal component analysis, was inversely associated with several biomarkers of inflammation, whereas a "western" pattern was positively associated with the inflammation biomarkers [188]. This finding is consistent with the results of investigations using an *a priori* approach to dietary pattern analysis, that is, index-based scores. The most widely used scores are based on the Mediterranean diet, and these have been associated not only with several health outcomes related to inflammation [189-193] but with reduced inflammation levels [116]. Likewise, the beneficial effects of a healthy Nordic diet have gained attention more recently [194]. Our results contribute to the evidence that diets that follow recommendations based on the Nordic food culture are protective against chronic diseases with an inflammatory component, such as CVD and cancer. It is important to investigate the usefulness of the national recommendations in specific populations because the consumption of a Mediterranean-like diet might be hindered by the lack of availability or the cultural acceptance of certain food items. One study performed in Belgium demonstrated that adhering to the national food-based dietary guidelines was associated with lower levels of biomarkers such as hs-CRP, IL-6, WBC count, and ox-LDL [195]. Overall, we would argue that the common denominators of these dietary patterns may be more important than the differences. In this regard, the consumption of fruits and vegetables, legumes, whole grains, and fatty fish and the limited consumption of red meat, discretionary foods and sugary drinks seem to be of great importance.

Low-grade inflammation and breast cancer (Paper III)

In this study, we found significant associations between several biomarkers of inflammation and postmenopausal breast cancer. We observed a positive association for IL-1β and inverse associations for ox-LDL and TNF-α. Adjusting for obesity did not have a major impact on the results. A major strength of this study was the fact that we investigated several biomarkers of inflammation. Previous research regarding inflammation and breast cancer primarily used CRP as the biomarker of inflammation [84]. A review and meta-analysis of 12 studies reported a 7% increase in breast cancer risk for each doubling of the CRP concentration [84].

In our study, only higher levels of IL-1β were associated with increased breast cancer risk. The role of IL-1β in cell proliferation and differentiation has been described previously, along with its role in apoptosis [196]. Its presence in patients' breast tumor cells has also been reported together with IL-1α [197]. Nevertheless, previous reports have highlighted that the coordinated expression of IL-1β and TNF-α seems to be important in cancer progression [198]. However, in our study, we observed opposite effects for these two biomarkers, despite their positive correlation with each other: higher levels of TNF-α were associated with lower breast cancer risk. It is possible that the coordinated expression of these two cytokines is more important during the cancer progression phase than during the development phase. Several roles in the carcinogenic process have been attributed to TNF-α, e.g., the ability to stimulate fibroblasts and tumor cell growth and the association with increased aromatase activity [199-201]. However, TNF-α is a highly pleiotropic cytokine with reported dual roles in carcinogenesis [78, 101, 202], as it also acts as a cytotoxic factor in cancer cells [203]. The few existing epidemiological investigations add very little to this picture, as no significant associations have been reported [204, 205]. Similarly, a dual role has also been reported for ox-LDL; it is associated with increased cancer risk [99, 206], but it can have a cytotoxic effect on cancer cells by inducing apoptosis and autophagy [207]. Another possible explanation for the increased breast cancer risk associated with decreased ox-LDL levels has been raised and was recently reviewed by two meta-analyses investigating blood lipids and breast cancer [208, 209]. Normal breast cells readily internalize ox-LDL, leading to an increase in proliferative and pro-inflammatory signaling and paving the way for breast cancer development [99]. This suggests that lower circulating cholesterol levels reflect the greater uptake of cholesterol in breast cancer cells when compared to normal cells due to active undiagnosed tumors (i.e., reverse causality).

We found divergent associations between several biomarkers of inflammation and breast cancer and a complex correlation matrix among the biomarkers. This highlights that different biomarkers could be associated with different

inflammatory processes. For example, WBC count, a strong indicator of inflammation that promotes cardiovascular events [210], was not associated with any of the biomarkers that were significantly associated with breast cancer in our study. This is a clear argument for the importance of examining more than one biomarker of inflammation in future studies investigating inflammation and breast cancer.

Furthermore, the question remains regarding whether using these specific biomarkers as biomarkers of inflammation in the development of breast cancer is adequate. Their value in the CVD field is undeniable, and their role in breast cancer progression is also accepted [102, 211]. However, we cannot predict how well these biomarkers perform as markers of the slow and lengthy process of breast cancer development [166]. It is likely that because of a probable shorter time of disease progression between inflammation and CVD and the better-established inflammation markers in this field [183], an association between inflammation and CVD is easier to establish. Access to biomarkers that are measured routinely after baseline examinations could help better establish the potential link between low-grade inflammation and breast cancer development. Nevertheless, not all the hallmarks of cancer, including those related to inflammation, might or must be met by all cancer cells, as some authors argue [212, 213]. Although the biomarkers used in this study were deemed reliable in **Paper I**, a few have limitations during longer periods [166]. Because the inflammation process leading to cancer might play a role during specific key periods [211], it may be impossible to discern the specific biomarkers for each individual at the specific time-points. The specific role of these biomarkers in breast cancer development remains controversial [78], and conflicting findings might be due to underlying undiagnosed tumors.

The release of inflammatory mediators that induce aromatase expression (and the subsequent production of estrogens) occurs in the adipocytes nearby the ducts (in the breast tissue) [214]. This increased local inflammation might not be reflected in circulatory levels, and so systemic inflammation might not be as important in the development of breast cancer. However, the potential role of inflammation in the development of hormone responsive breast tumors (ER+) through the induction of aromatase action has been reported [215]. It is possible that the hormonal factors overshadowed the inflammatory pathways in our study. Inflammation also seems to play a role in tumor types that are non-responsive to hormones. Our study would benefit from investigating the association between inflammation and breast cancer development depending on hormone receptor status. However, a larger sample size would be required to enable investigations across strata, especially because tumors that are non-responsive to hormones are more common in pre-menopausal breast cancer cases [216, 217].

Diet, low-grade inflammation and breast cancer (Paper IV)

In this study, we identified three food patterns (factors) associated with three biomarkers of inflammation previously associated with breast cancer (ox-LDL, IL-1β and TNF-α) using RRR. Factor 1 and Factor 2, which together explained most of the variation in the response variables, were not significantly associated with breast cancer. Factor 3 was positively and significantly associated with breast cancer risk in the minimally adjusted model. However, after adjustment for potential confounders, the associations were attenuated, suggesting that the associations might not be independent from obesity and hormonal factors.

The potential protective role of healthy dietary patterns in breast cancer development has been previously described [47]. Few studies have investigated the link between diet and breast cancer using RRR to derive dietary patterns associated with hypothesized intermediates [218-223]. Higher scores for dietary patterns associated with glycemic index and glycemic load [223] or fatty acid intake (SFA, MUFA and PUFA) [222] were significantly associated with increased breast cancer risk. To the best of our knowledge, our study is the first to examine biomarkers of inflammation as potential intermediates between diet and breast cancer. Our hypothesis was based on a biological model under the assumption that inflammation plays a role in breast cancer development [224]. The selection of intermediates to use in this study was based on the biomarkers of inflammation that were associated with breast cancer, independent of potential confounders, in the previous investigation (**Paper III**). We therefore sought to avoid creating spurious associations by choosing the appropriate intermediates (i.e., those associated with both exposure and disease). Nonetheless, there is a degree of uncertainty regarding what specific processes these biomarkers illustrate, as the associations with breast cancer diverge.

Factor 3 added very little to the explanation of the variation in the inflammation markers (0.51%). It was mostly associated with IL-1β (the only cytokine we found to associate positively with breast cancer risk), and the concentration of the three biomarkers across tertiles of Factor 3 mirrored breast cancer risk. The highest tertile of Factor 3, which was associated with increased breast cancer risk compared with the lowest tertile, was characterized by higher levels of IL-1β and lower levels of ox-LDL and TNF-α. Because both diet and inflammation were measured at baseline, it remains to be discerned whether the associations observed between Factor 3 and breast cancer were driven by the inflammation markers' associations with breast cancer rather than by the specific food pattern (and inherent specific food group choices).

The interpretation of the identified food patterns is not simple. We did not find clear "healthy" or "western" dietary patterns. In fact, no pattern was comprised

solely of unhealthy foods loading low and healthy foods loading high (or vice versa). People who consume both disease risk-enhancing foods and those with protective effects might find that these effects are counterbalanced in the metabolism, and this may have contributed to the overall lack of associations between diet and breast cancer in our study. This adds to the importance of studying dietary intakes as a whole, as the effects of single nutrients or foods might not reflect the real effects in the context of a dietary pattern. However, it is possible that dietary pattern analysis could hide the importance of specific foods that are independent of the overall dietary pattern. This was observed in our results when stronger associations were detected between Factor 3 and breast cancer after adjusting for the variable dried soups/sauces (in models I and II). This variable has previously been described to associate with tumors non-responsive to hormones in the same cohort, suggesting that a non-hormonal mechanism was involved [119]. It was hypothesized that inflammation and oxidative stress resulting from oxidized sterols found in this type of industrialized food product (due to prolonged heat treatment) could contribute to increased breast cancer risk [118]. An alternative explanation is that this variable could be a marker for risky dietary behaviors such as extreme dieting, which could lead to some nutrition deficiency [118]. Although it is challenging to interpret the model when one variable that was used for pattern construction is including as a covariate, we could speculate that the stronger associations observed suggest a somewhat shared pathway effect. However, the question remains whether this pathway is inflammation. The potential subjectivity associated with the creation of food groups at the beginning of the study could also have contributed to the lack of associations observed. Decisions regarding the aggregation of food groups were made based on fiber and fat content and food culture, but it is possible that different decisions could have led to the creation of different food patterns. It is also possible that our study simply did not have enough power to detect associations between diet and breast cancer via the chosen biomarkers.

Many of the dietary patterns previously associated with breast cancer include alcohol as a variable [47]. It is not surprising that "drinker patterns" are associated with increased breast cancer risk given the risk associated with alcohol consumption itself [225]. However, moderate consumption of alcohol is linked to reduced CVD mortality, probably due to decreased inflammation [226]. Furthermore, despite the high relative validity of the alcohol variable in this cohort, women of the MDC cohort tended to underestimate their mean daily intake by about 60% compared with the reference method [140], raising questions of possible misclassification. In an attempt to investigate and possibly identify disease-associated food patterns that would not be driven or confounded by alcohol, we decided to not include this variable in the RRR analysis. Instead, alcohol intake was introduced in the multivariable models as a potential

confounder of the associations between diet, inflammation and breast cancer. The non-inclusion of alcohol might have also contributed to the lack of associations observed between food patterns and breast cancer if this was the major dietary component associated with breast cancer. It is possible that the protective associations observed between "healthy" patterns and breast cancer reflect a healthy lifestyle that does not include alcohol consumption to a major extent. Healthy diets, together with physical activity and maintaining a healthy weight, coexist in an overall healthy lifestyle, which could play a key role in preventing breast cancer. However, it is difficult to correctly disentangle the effects of each contributor, even with multivariable models that adjust for potential confounders [227]. Healthy diets could also play a role in the development of breast cancer through specific dietary factors (e.g., the consumption of fruits and vegetables or red meat), and pathways other than inflammation might mediate the associations between diet and breast cancer.

Other possibilities that warrant further investigation include the insulin/IGF-1 (insulin-like growth factor) and COX-2 pathways. In the first case, circulating insulin and IGF-1 (i.e., growth factors) resulting from hyperinsulinemia (which is closely related to dietary intake and lifestyle factors) promote cell proliferation and cell growth and inhibit apoptosis [228]. Hyperinsulinemia might also be responsible for the increased production of inflammatory cytokines. In the second case, COX-2 overexpression is induced by inflammatory stimuli, such as cytokines and growth factors, and as a result, arachidonic acid is transformed into prostaglandins. The main type of prostaglandins produced via COX-2 is PGE_2, which promotes angiogenesis and induces aromatase expression [71]. It should be worthwhile to expand the narrow picture presented in this thesis to include concurrent pathways that could help explain associations between diet and breast cancer.

The answer to our main question remains open. An overall healthy dietary pattern is reportedly beneficial in relation to breast cancer. However, which specific dietary factors have critical roles in breast cancer development remain to be clarified. Furthermore, whether inflammation is one process through which diet is linked to breast cancer is still unclear.

6.2. Methodological considerations

The main concern in observational studies is their non-randomized nature. Individuals are not assigned a pre-determined level of exposure of a specific factor, while other factors are held constant; instead, the participants report or are subjected to the measurement of exposure levels for several factors. To avoid spurious associations, issues such as random errors (or chance), measurement errors, confounding and other biases (systematic errors) must be accounted for in both the design and the analyses steps of the investigation [13]. However, to affirm causal associations, additional steps need to be taken, such as exploring the consistency of the results with other studies, the strength of associations, the dose-response relationship, the biological plausibility, and the temporality [229].

Study design

The MDC cohort has a prospective design, and thus its main characteristic is that the exposures were measured before the outcomes occurred. This is a major advantage compared with retrospective studies in which exposure information is collected after the outcome has occurred. The prospective design allows people to be followed from their enrollment in the study until one or several health outcomes occur or until death or migration. This benefit is pivotal when investigating the etiology of diseases. It also ensures that the outcome cannot influence the reported exposure (recall bias).

In **Paper I**, a longitudinal design with repeated measures was used to estimate the reliability of several biomarkers. Lifestyle factors were assumed to be stable throughout the study period (approximately 6 to 8 weeks for each subject). Because objective measures were used, we did not expect to encounter a "learning effect" from repeated measures, nor did we expect to see an "order effect" because no treatment effect was being evaluated.

In **Paper II**, a cross-sectional design was used. This means that both exposures (i.e., the dietary factors used to construct the DQI-SNR index) and outcomes (i.e., biomarkers of low-grade inflammation) were measured at approximately the same time, and thus we were not able to separate cause from effect. In this specific case, several biomarkers were measured at the time of the participant's first visit to the study center; information on diet was then retrieved approximately 2 weeks later, and other biomarkers were assessed using fasting samples drawn on a third visit approximately 7 months later. We can argue that diet comes first, for several reasons: 1) the assumption that dietary habits tend to be stable over time; 2) literature referring to the effects of diet on inflammation levels; and 3) biomarkers

measured 7 months later were associated in the same direction (i.e., negatively) as those measured 2 weeks before the dietary assessment. However, we cannot completely exclude the possibility that people with low-quality diets have a higher risk of inflammation due to other factors for which we could not account.

In **Papers III** and **IV**, a nested case-control design was used. Whereas other case-control studies can encounter major problems with recall bias, when a study is nested within a cohort, it ensures that exposures are measured or reported before the outcome. However, the period for the development of breast cancer is not exactly known and is believed to be long. Therefore, in **Paper III**, people diagnosed within the first to third years of follow-up were excluded in the sensitivity analysis. Because inflammation markers were the exposures examined in this paper and because levels of these markers increase with the progression of cancer [230], a temporal bias (also known as reverse causality) could have been introduced. With the successive exclusion of women diagnosed within the first years of the study, the associations lost significance but the magnitude of the estimates remained similar. While we cannot confidently exclude the possibility of reverse causality, the stability of the estimates points instead to loss of power.

Another advantage of conducting a nested case-control study is that biological analyses of the blood samples were not performed for the whole cohort. This was important in terms of both budgetary limitations and conserving the already limited volume of the stored samples. Two matched controls were selected and matched per each breast cancer case, to ensure power for the main analysis. It is generally noted that there is no major gain in efficiency of adding more than 2 controls to the analysis compared with the improved gain in power, and more than 4 would not be necessary [13]. The controls were drawn from the same source population as the breast cancer cases to minimize bias.

Biases and errors

In general terms, there are two main sources of error in epidemiology: random error, also known as "chance", and systematic error, also known as bias. Random error affects the precision of the estimates but can be improved by increasing the sample size. In exposures measured with higher random error, estimated associations with an outcome will move towards the null. In comparison, systematic error affects internal or external validity, and it cannot be improved increasing the sample size. Biased estimates can move in any direction. Many types of biases exist, but they can be divided into 3 categories: 1) selection bias, which relates to how the subjects are selected; 2) information (or measurement) bias, which relates to how relevant information is obtained; and 3) confounding bias, which relates to how information is handled and treated [13].

Random error and precision

Random error affects measurements in a non-systematic manner as the error differs from one measurement to the next, leading to increased variability (or variance). Sources of random error, such as random variation of the process used to select the specific study subjects (i.e., random sampling variation) and random variation in a measurement or estimation process (due to chance) affect the precision of the estimates. These sources of random error can be addressed and statistical power and precision can be improved by increasing the sample size and using repeated measures [13]. The MDC cohort is a large cohort that had previously acquired enough breast cancer cases (n=283, in 1999) to provide enough statistical power to detect associations between diet and breast cancer. Nonetheless, the statistical power was lower for different strata and subgroup analyses, such as across BMI categories.

Significance or hypothesis testing is used to help researchers decide whether results of a study are due to chance and how precise the results are. Researchers use *p*-values or CI for this purpose. P-value is a conditional probability used to help determine whether the null hypothesis (i.e., that there is no association or difference between A and B) should be accepted or rejected. This conditional probability can be defined as the probability of finding an effect when the null hypothesis is in fact true. In other words, it is the probability of a type I error. A significance level is then set at a reasonably low level (of alpha) and used to reject the null hypothesis with some confidence. The most widespread level of significance used in epidemiology is below α=5% (p<0.05). Conversely, a significant finding could be determined by using the CI of an estimate if it does not include the null. However, all statistical hypothesis tests are prone to type I and type II errors. Type I error or alpha (α) error (also known as false positive) occurs when one rejects the null hypothesis when it is in fact true. Type II error or beta (β) error (also known as false negative) happens when one fails to reject a false null hypothesis. The power is thus the probability of correctly rejecting the null hypothesis when the alternative hypothesis is in fact true. Methodological sources of bias may also contribute to not rejecting the null hypothesis when it is false (type II (β) error). Type II error is very common in nutrition epidemiology, and possible reasons for this are the measurement of the incorrect temporal period, low statistical power, and negative confounding.

Using an alpha value of 5% proposes that 1 out of 20 of our estimates will potentially have occurred due to chance. Though we cannot assuredly exclude findings by chance, all the associations examined in the four studies included in this thesis were conducted with an *a priori* hypothesis grounded in proposed biological mechanisms. Typically, results need to be confirmed in other studies (with other designs, population, etc.) to confidently conclude that type I and type II errors did not affect our results.

Selection bias and external validity

In general terms, selection bias occurs when the associations between exposure and disease differ between participants and non-participants. Different types of selection bias include self-selection bias, Berksonian bias, diagnostic bias and differential response rate [13]. The ability to generalize the study results to the source population is not always a fundamental step in exposure-disease studies (etiological risk), in which the internal validity (i.e., exposure and outcome are measured with high validity) is of utmost importance. However, external validity is a necessary feature if inferences from an internally valid study are to be extrapolated to the source population, and possibly to others. To ensure a good external validity, selection bias must be avoided.

The low participation rates in prospective cohort studies can be of concern, especially because participants are more often health-conscious than non-participants. The MDC cohort represents approximately 41% of its source population, and its participants reported better subjective health and had lower mortality during recruitment and follow-up than its non-participants [127]. Differences in exposure and outcome rates might affect the cohort's comparability to the source population. However, socioeconomic and demographic characteristics were similar for both participants and non-participants. The inclusion of more health-conscious participants could lead to a more homogenous population with less between-person variability in exposure levels (e.g., for dietary intakes), which will affect the ability to detect an association between exposure and outcomes. In a broad sense, a wide range of exposures is essential in epidemiology. For example, while constructing the DQI-SNR in **Paper II**, the cut-off for adequate SFA intake was redefined after observing that a very small percentage of the study population adhered to the recommendations. This could be an indication that the ability to detect associations with disease outcomes might be hampered due to a narrow range of SFA intake in the MDC study.

In contrast, if participation is related to both exposure and outcome, we face selection bias. Cancer incidence prior to recruitment was higher among MDC participants [127], which could suggest that enrollment was associated with prevalent cancers. Because prevalent cancers could also be associated with other future cancers and disease outcomes, the presence of prevalent cases was accounted for in **Papers I** and **II**, and was used as an exclusion criterion for sample selection in **Papers III** and **IV**. Self-referral to the study could also be associated with the outcomes and thus incur a self-selection type of bias. Approximately 22% of the MDC participants responded to advertisements for enrollment (classified as "passive recruitment") and were not directly invited [124]. Cancer incidence and all-cause mortality were lower in the passively recruited group and socioeconomic and lifestyle factors differed from those of the other participants. This raises concern of a "healthy cohort" effect. However, it is

hoped that this possibility was counteracted by the fact that the majority of the MDC participants were actively recruited through personal invitations [124].

Overall, the representativeness of the MDC cohort can be considered reasonable. The sampling procedures for all four studies included in this thesis also did not limit the ability to generalize results, taking each specific criterion into account. However, the ability to generalize to other populations is limited to the age group of the MDC cohort and to the specific environmental factors (which influenced lifestyle factors, such as diet) present during the 1990s in Sweden. Nevertheless, high internal validity is arguably very important in etiological research. Biological associations observed in studies with high internal validity should persist regardless of low external validity. In this thesis, we investigated potential etiological associations and thus, concerns regarding internal validity are high.

Information bias (misclassification) and internal validity

Internal validity refers to the degree to which one can trust the results of a study. In other words, it refers to the accuracy of the measurements of exposure and outcome, and it is considered a prerequisite for external validity.

Information bias or misclassification bias are the major sources of bias that threatens internal validity as a result of systematic measurement errors and misclassification of exposures, covariates or outcomes. These biases include systematic errors at the individual level that are observed as random in the population level. Systematic error can be divided into differential and non-differential if it affects groups differently (depending on the actual values of other variables) or all study subjects to the same degree (independent of other variables). While non-differential misclassification most often (but not always [231]) bias estimates towards the null, biased estimates resulting from differential misclassification can be either underestimated or overestimated. Differential misclassification can be the result of recall bias or interviewer bias.

Misclassification of exposures

Biomarkers

The investigation of the reliability, or reproducibility, of several biomarkers used in this thesis was the main focus of the study described in **Paper I**, and the implications are discussed in the previous section (6.1). Nevertheless, questions pertaining to the validity of the biomarkers are also of importance. Misclassification of the various biomarkers' concentrations could arise from different sources: laboratory errors, degradation of the biomarkers during storage time, season and concurrent conditions at the time of blood sampling (such as the flu and circadian variations). Because the biomarkers were measured from blood taken at baseline measurements, we can assume with a certain degree of

confidence that potential misclassifications are independent of disease outcome; thus, we potentially face mostly non-differential misclassification. All the subjects underwent the same measurement procedures at baseline, and all the samples were stored under similar conditions. This was ensured by the quality control program [131]. However, as described previously, the procedures used for handling leukocytes was changed when a deterioration of concentration levels was detected in the stored samples [130].

It is plausible that season could have had an influence on biomarker levels [232]. In fact, in **Paper I** we did observe a tendency (albeit non-significant) for higher levels over time (between November and May) in certain biomarkers. This could have biased our results if people measured over 6 weeks in the spring had had significantly higher levels than people measured over 6 weeks in the winter. Furthermore, this observation could also be of concern when investigating dietary associations (in **Papers II** and **IV**) because misclassification of dietary intakes may have occurred due to seasonal variation. To exclude seasonal effects, we adjusted for season in **Papers II** to **IV**. In **Paper I**, we stratified the analysis by season. The ICCs remained similar across the 4 season categories.

At the individual level, circadian variation could lead to the misclassification of biomarkers. This could be a problem for specific biomarkers, such as IL-6 [169]. To account for this possibility in **Paper I**, the participants were measured at approximately the same time every time they visited the study center, and the number of days between measurements did not differ greatly. We cannot, however, be overconfident regarding the ability to account for circadian variation with this study design. For the biomarkers that measured in samples taken at the baseline examinations (**Papers II-IV**), this possibility was not considered; consequently, those values may suffer from bias to a higher extent.

The degradation of stored biomarkers (i.e., laboratory drift) is another source of bias [89]. This can happen to all subjects at the same rate, but it affects results if blood samples from cases and controls are analyzed at different time points. This is more common in nested case-control studies and can be avoided by performing the laboratory analyses for cases and matched controls at the same time. This was the approach taken in **Papers III** and **IV**. Some of the biomarkers were analyzed during the baseline examinations (this was the case for leukocyte counts) and thus were not affected by storage time. However, it could be that samples from cancer cases, due to their metabolic alterations caused by the disease, could face faster degradation of certain metabolites than samples from controls when stored for the same amount of time. This could lead to temporality problems, as the outcome would be affected by the measured exposure. This was accounted for our in sensitivity analysis by excluding the cases diagnosed within the first years of the study.

A source of error from the laboratory analysis could be the batch effect: errors can occur between or within analytical batches [89]. Variation among batches can sometimes be greater than between-subject variation and thus can introduce biased estimates. Therefore, it is important to identify both types of variation separately. The CV_A (analytical coefficient of variation, or inter-assay variation) for all the biomarkers was within a reasonable range in **Papers I**, **III** and **IV**. Furthermore, a possible batch effect was investigated in **Paper III** by adding the plate number as a confounder to the main analysis, and the estimates did not change.

Diet

The non-differential misclassification of dietary intakes results from random error, which comprises true day-to-day variation in dietary intakes (in relation to a mean daily intake) and random error associated with measuring dietary intakes [103]. As a result, individuals' intakes can be misclassified (thus increasing variance and decreasing statistical power), but this is not expected to affect group means. However, increased variance could lead to the attenuation of correlation coefficients (for example, in the validation studies), regression coefficients and risk estimates of diet-disease associations. Specific statistical methods can be used to de-attenuate diet-disease associations that are impacted by random measurement errors [233]. The attenuation coefficients can be estimated using the intra- to inter-person ratio of specific dietary intakes, which are measured with a superior method used in validation studies. In other words, the higher the intra- or within-person variation (i.e., higher misclassification of the true intake), the higher the attenuation factor. Subsequently, correlation coefficients and risk estimates can be corrected using the attenuation factor. Although this thesis could benefit from this procedure, as could other studies from the MDC cohort investigating dietary exposures in relation to a disease outcome, such information was not available. In any case, we expect a low rate of non-differential misclassification in our studies due to the previously observed medium to high relative validity of most food groups and nutrient intakes [139-141]. If anything, one could presume the true risk estimates to be higher than those observed, provided the measurements are free of systematic errors.

Systematic differential measurement errors are more worrisome because biased estimates could move either towards or away from the null, and there is no statistical method that can correct for bias. This type of bias could be the result of a dietary assessment method that does not cover specific food groups that are frequently consumed by specific groups of people. It could also derive from people or specific groups of people under- or over-reporting their consumption of specific foods. The classic example is when obese people tend to under-report perceived "bad" food choices and over-report the consumption of fruits and vegetables. The common practice in nutritional epidemiology is to identify groups

that potentially misreport their diet in a systematic fashion and run sensitivity analyses that exclude them. If the estimates remain similar after such exclusions, one can be more confident about trusting the results.

All dietary assessment methods are prone to errors, and the modified diet history method used in the MDC is no exception. It has been argued that the measurement errors associated with the most commonly used dietary assessment methods might be the major source of a confusing message regarding the association between fat and breast cancer [234]. It is possible that the use of a combination of two types of dietary data in the MDC, current diet (food diary) and usual diet (diet history questionnaire), helped capture a more comprehensive picture of the individuals' usual intake by including different dimensions. This possibility is supported by the relatively high correlation coefficients for most foods and nutrients in the validation studies [140, 141], and when compared with other methods with lower validity used in other cohorts [103]. Nevertheless, one should keep in mind that even the reference methods used in validation studies include measurement error and that these errors correlate among the different assessment methods, leading to an overestimation of the correlation coefficients [235]. In addition, the use of two sources of dietary information could bring the specific errors associated with each method into play, making it difficult to determine which errors affect which food groups. For example, the consumption of berries was mostly reported in the FFQ, which is affected by the memory of the respondents (and ultimately could lead to recall bias), whereas coffee consumption (reported in the same instrument) is likely to be easier to remember and report. In comparison, on the 7-day food records, the consumption of certain foods may have been affected by the subjects' efforts to make reporting easier and less cumbersome [106] or simply may have not captured the consumption of foods that were not consumed during that week. An example of this is fish consumption, which was mainly captured by the 7-day food record. If consumption was infrequent, even regular fish eaters would be wrongly classified as zero consumers if they did not report eating fish during those 7 days. This effect is further shown by the low relative validity of this variable, especially among the men of the MDC [140]. A final thought regarding the MDC methodology is the difficulties pertaining to portion size estimation. Estimating the portion sizes requires a great deal of motivation, good cognitive function and again, good memory. It is, however, possible that the use of comprehensive booklets with pictures aided a more precise estimation, at least while keeping the 7-day food record.

The use of dietary patterns rather than the reported food groups to investigate diet-disease associations could add another layer of misclassification, especially when the DQI-SNR was used in **Paper II**. The DQI-SNR was created with the aim of reflecting overall diet quality, taking dietary and nutrition recommendations into consideration. However, this could be impaired by the reduction of dietary habits

and their complexities into an index that is the result of the sum of 6 dichotomous components representing adherence or non-adherence to certain recommendations. Nevertheless, the ability to adequately categorize individuals' diets as low, medium and high quality has been previously acknowledged. A high diet quality score, as identified by the DQI-SNR, reflects adherence to many SDGs and SNRs [147]. Furthermore, the DQI-SNR index has been shown to be more predictive of disease risk (especially CVD incidence and overall and CVD-specific mortality) than the reported consumption of specific foods or nutrients [148, 149]. However, it is important to note that some of the components used in the DQI-SNR did not show high validity, such as fish and PUFA consumption for men, or high reproducibility, such as sucrose consumption for women [139-141]. Another important detail is that the cut-offs for two index components (fish and shellfish, and fruits and vegetables) were based on the absolute intakes (g/day) instead of relative intakes. This might lead to some misclassification because the diet history method was not appropriate for estimating absolute intakes and because the consumption of fruits and vegetables tends to be over reported [140]. Overall, dietary indices are useful for comparing individuals in the high and low categories, but strong inferences regarding people in the middle categories should be avoided as this group could have very different diets [236].

Measurements of the diet at one point in time may not represent the total diet usually consumed. Despite all the pitfalls associated with dietary assessment, diet is still a useful exposure to measure, mostly because researchers expect a somewhat stable diet pattern in certain periods of life. This is potentially the case for this cohort, which, because of their age and the food availability on Swedish shelves in the 1990s, we assume that they have dietary habits that are more established than those of younger cohorts. It could have been useful, however, to have access to dietary changes that might have occurred during follow-up. However, the added value of additional measurements could be expected to be small [237]. Nonetheless, the time point at which dietary exposure is crucial for the development of cancer remains to be clarified. Due to long latency and the unknown periods of cancer development [26], it is not clear to us how important "current" dietary habits are compared with dietary habits from a decade or two ago or with an individual's overall dietary history.

In this context, and despite its self-reported nature, the variable identifying people who have substantially changed their dietary habits proved to be useful for detecting people with unstable dietary habits [150]. Estimates obtained in the main analyses after exclusion of the people who reported having changed their diet in the sensitivity analyses (in **Papers II** and **IV**) remained stable. This gives us some confidence that our results are free of possible bias due to misclassification caused by unstable dietary habits. Furthermore, concerns regarding correlations with other confounders in the MDC have been raised in the past [151]. The participants with

unstable dietary habits were more likely to be obese, to be highly educated and to be non-consumers of alcohol and non-smokers [151]. Obesity is a major risk factor for many non-communicable diseases. In fact, the reasons for dietary changes among obese people might be related to diseases such as metabolic syndrome [238]. This is especially important if people changed their diets due to diseases associated with breast cancer. Additionally, recent dietary changes might not eliminate the risk for many conditions acquired throughout the years.

Obesity is also related to the underreporting of energy intake, especially when self-reported data is used [239] as was the case in the MDC cohort [143]. This could be the result of failure to report consuming specific food items because of social desirability or the result of dieting (perceived or real energy restriction), which is somewhat captured by the "diet change" variable. To further investigate misreporting, the EI:BMR ratio was used to calculate the plausibility of reported energy intakes. The inclusion of groups who misreported their dietary intake could lead to spurious associations because misreporting could be associated with the outcome of interest [238]. Again, in the sensitivity analysis (in **Paper IV**), we excluded people who were classified as misreporting their energy intake, and the estimates remained similar.

Another possible source of systematic error in dietary assessment could be seasonal variation in dietary intakes [240, 241]. However, this possibility was accounted for in the analyses in **Paper II** by including the variable season as a covariate in the fully adjusted models, and no influences were observed. In **Paper IV**, the possible influence of season was addressed in the design by matching controls who underwent baseline measurements at approximately the same time as the cases. Matching by the date of baseline examinations may have solved additional methodological issues that could be sources of concern: the diet assessment method version and interviewer variability. Finally, energy adjustment was performed in both the papers that examined diet (**Papers II** and **IV**) with the aim of reducing the influence of measurement errors related to reported intakes [242].

Misclassification of covariates

Overall, a multivariable model including covariates that were weakly measured could produce attenuated estimates between exposure and outcome. Self-reported data are typically more prone to error than direct measurements. Weight, height, waist circumference and hip circumference were directly measured using standardized procedures during the baseline of the MDC. Other variables, such as physical activity (PA), smoking, education, parity, alcohol consumption, age at menopause, age at menarche, and medication were self-reported by the participants and thus were less precise.

Notably, the degree of error varies depending on the variable. For example, it is recognized that participants in the MDC had problems reporting PA because of the use of an extensive reporting method. Associations between leisure time PA and anthropometric measurements and metabolic components, previously reported, were stronger in men [153]. We suspect that the finding of no significant associations between PA and breast cancer in our studies might be due to some misclassification associated with this variable, especially in women.

Information regarding MHT, retrieved both from the lifestyle questionnaire and the 7-day food diary, only referred to current use. The high agreement between information obtained from the two sources denotes the high validity of this variable [157]. However, information regarding past use and duration, and thus increased risk, might be lost. Still, in the MDC, current use was a strong predictor of breast cancer, which is consistent with the literature [27].

Misclassification of the outcome (breast cancer status)

The Swedish cancer registry is nearly 100% complete, and therefore, the potential misclassification of breast cancer status is very low. However, one cannot completely exclude potential detection bias as mammography screening attendance can differ across different socioeconomic groups. High socioeconomic groups are expected to undergo screening more than other groups [243]; thus, the detection of breast cancer might be higher in this group. Another factor to consider is the loss to follow-up (i.e., no information available on whether they became cases), but the percentage of individuals who emigrated from Sweden during follow-up is very small (about 0.5%).

There is some uncertainty surrounding the in situ breast cancers. It is not known the proportion of these cancers that progress to invasive. Therefore, only women diagnosed with invasive breast cancer were included in the analysis. The inclusion of in situ breast cancers could have obscured a true association between diet and breast cancer.

Confounding

Confounding is a major threat to causal inference and internal validity in observational studies. When exposed and non-exposed groups differ in terms of characteristics other than the exposure of interest and these characteristics are associated with the outcome, associations between exposure and outcome may be biased because of confounding. Depending on the direction of the associations between the confounding factor and exposure and outcome, measures of association can be either overestimated or underestimated. Confounding factors should not be part of the causal pathway between exposure and disease (i.e., mediating factors), and they might only occur in the context of a specific study (i.e., specific methodological issues) [13].

The ability to adjust for potential confounders is limited to what was measured or reported during baseline examinations. The participants in the MDC cohort underwent extensive examinations; therefore, in our studies, we were able to account for many potential confounders, especially those related to hormonal exposure, which is of great importance when examining breast cancer risk. We could also account for various lifestyle and socioeconomic factors known to be associated with diet, such as PA, smoking status, alcohol consumption, and obesity. These factors are also associated with breast cancer. Our choices regarding which confounders to include in the multivariable models were guided mainly by the information available in the literature that reflected a biological mechanism. Additionally, several "methodological" variables were also considered. These were inherent to the MDC cohort (such as method version or the week of blood sampling) and could have contributed to some selection bias. A final layer of decision-making when building the multivariable models focused on avoiding over-adjustment and keeping the models as parsimonious as possible. In this regard, in **Paper III**, for example, we did not include variables related to hormonal factors (such as age at menarche, age at menopause, age at first child, and breastfeeding) in the multivariable model, even though the literature refers to these as risk factors for breast cancer. Other studies within the MDC cohort with a prospective design have included these factors [156, 244]. However, in our studies, we did not see an association between these factors and the outcome of interest, nor did we observe a difference in the estimates when they were included in the multivariable model.

Residual and unmeasured confounding

Residual confounding may still exist despite our efforts in adjusting for potential risk factors and confounders. This could be due to poor measurement of these factors, as most were self-reported (e.g., PA, smoking, alcohol consumption). Using large categories might also lead to residual confounding by not adjusting

adequately for the confounder; for example, tertiles of PA might be a very crude form of categorization that could miss nuances. In **Paper I**, attempts to account for possible confounders were made by stratifying important lifestyle factors, but no major patterns were observed for the different ICCs. That is, most ICCs remained stable for all the biomarkers of inflammation across the strata. However, one cannot completely exclude the possibility of residual confounding because of the loss of power within the strata.

Finally, we cannot exclude that other potential confounders that were not measured or are not known could also be a potential source of bias. For example, information regarding family history of breast cancer, specifically, was not available in the MDC. However, this risk factor is recognized to be more important in younger women [245], and therefore, its influence as a confounding factor in the age group investigated is potentially low.

Interaction

An interaction exists when the degree of association between exposure and outcome varies across the levels of another variable. This is synonymous with effect modification if no bias is present. It is important to note that statistical interaction might not correspond to biological interaction. Biological interaction is usually tested in stable factors, such as genetic factors, as it presupposes a causal relationship. Interaction is scale dependent as it can be evaluated using an additive scale (risk differences) or a multiplicative scale (risk ratios). In general terms, a lack of interaction in one of the scales assumes the presence of interaction in the other. It is worth noting that biological interaction is mostly tested using the additive scale, and this scale is more relevant from a public health perspective [13]. Unlike confounding where stratification or adjustment for the confounder will result in less biased estimates, in the case of an interaction, it should be reported as such and stratification by the interacting factor is mandatory if the interaction effect is to be removed.

The investigation of potential interactions was not the main focus of this thesis. We did, however, have some concerns regarding possible interactions that could affect our estimates and thus conducted statistical analyses to inspect these concerns. Our main concern was with obesity in **Paper III**. With a diet-inflammation-breast cancer framework in mind, obesity could play several roles: confounder, mediator, or effect modifier. This would imply different approaches and interpretations regarding the inclusion of BMI or WHR in the multivariable models. For instance, if we wished to examine the "total effect" of inflammation on breast cancer and we considered that obesity might be part of the pathway (or vice versa, that obesity leads to increased inflammation which in turn increases the

breast cancer risk), we should not adjust for obesity. If the adjustment had been performed, we would be capturing only the "direct effect" of inflammation on breast cancer risk. If we considered obesity as just a confounder (i.e., obesity is associated with exposure and outcome but is not on the pathway; which was the approach we took), it would be wise to adjust for it. We also considered the possibility of obesity being an effect modifier (e.g., only obese subjects would present an effect between inflammatory levels and breast cancer, while normal-weight people might not be impacted by the same levels of inflammation). For this purpose, we tested for interaction on a multiplicative scale (OR) and did not observe significant effects. Thus, we performed analyses assuming that obesity is just a confounder of the association between inflammation levels and breast cancer. It is important to underline that not observing an interaction does not mean it does not exist, especially in studies with smaller sample sizes (**Papers III** and **IV**). The same line of thought was applied in **Paper IV**, where we tested for interactions for BMI, MHT and smoking status.

7. Conclusions

Overall, this thesis investigated the associations between diet and postmenopausal breast cancer and explored how low-grade inflammation and oxidative stress could affect this association. Taken together, the four papers included in this thesis suggest the following:

1. A single non-fasting blood sample measured at baseline could, with reasonable reliability, be used to rank study participants according to the concentrations of most of the examined biomarkers of low-grade inflammation and oxidative stress (i.e., ox-LDL, IL-1β, IL-6, IL-8, and TNF-α).
2. A high-quality diet characterized by high adherence to the Swedish Nutrition Recommendations and the Swedish Dietary Guidelines was associated with lower concentrations of several soluble and cellular biomarkers of low-grade inflammation (i.e., hs-CRP, TNF-α, S100A8/9, WBC, neutrophils, lymphocytes, and $CD14^{+}CD16^{++}$) in middle-aged individuals.
3. Three inflammation markers were associated with breast cancer risk; IL-1β positively, and ox-LDL and TNF-α negatively. However, no overall consistent association between inflammation and postmenopausal breast cancer risk could be concluded.
4. The observed positive associations between the obesity indicators and postmenopausal breast cancer appear to be independent from associations between inflammatory markers and postmenopausal breast cancer.
5. Three food patterns were associated with and explained the variation of three biomarkers of systemic inflammation (i.e., ox-LDL, IL-1β, and TNF-α).
6. Although one food pattern (Factor 3) was associated with increased breast cancer risk, the significance level was lost in the multivariable analysis. The other two food patterns were not significantly associated with postmenopausal breast cancer risk.

In conclusion, our results suggest that diet is associated with low-grade inflammation. However, the role of diet in the development of postmenopausal

breast cancer, whether via inflammation or another pathway, remains to be clarified.

8. Future challenges and public health perspective

Epidemiological studies, despite all the pitfalls, remain an important source for understanding the association between diet and disease. Randomized controlled trials are the highest standard in medical research, however little can be achieved by using this type of design when investigating the impact of diet on NCDs. It is, thus, important that researchers are aware of all types of biases and errors that can influence study results, and take action to prevent them or to control for them. Major challenges that researchers are confronted with in the field of cancer epidemiology include the difficulty to prove causality, and the fact that cancer has a multifaceted nature. Lifestyle factors (including diet, and physical activity), genetic make-up and other environmental exposures together play a role in the development of many NCDs. To pinpoint the causes or risk factors for multifactorial diseases remains a challenge and this thesis is just one humble step further. Each of the four projects included in this thesis faced their own challenges and shortcomings (discussed in section 6). General and paper-specific lessons were learned, and future directions in this field may include:

Within the MDC

- More people should be included in the studies (following on the work in Papers III and IV), in order to be able to investigate interactions among exposures, and estimate RR across strata (for example in different categories of hormone receptor status).
- Other biomarkers could be investigated in relation to breast cancer, in order to increase understanding of the complex picture of the mechanisms.
- Investigate other possible mechanisms where inflammation may play a co-adjuvant role (for example the COX-2 and the insulin/IGF-1 pathways).
- Examine other potentially diet-induced mechanisms: the role of modern food habits and other single compounds introduced with manufactured and highly processed foods (e.g., AGE – advanced glycation end-products, ALE – advanced lipoxidation end-products).

- Other possibilities include the study of the role of gut microbiota in modulating diet-induced inflammation, and the examination of the role of genetic variation in relation to inflammation and breast cancer.

In other settings

- Epidemiological studies should rely on reasonably reliable biomarkers, and should therefore include reliability investigations of the biomarkers under consideration.
- Include repeated measures of exposures and covariates during follow-up (e.g., every 5 years) and investigate possible changes.
- Choose appropriate tools that enable repeated measurements of high quality but at a low cost.
- Studies investigating dietary factors should rely on high quality dietary data, and therefore should seek to examine the validity and reproducibility of the dietary factors.
- More emphasis should be put into creating better dietary assessment methods. We might never overcome some biases inherent to dietary reporting (which should be measured against a golden standard such as DLW), but respondent and researcher burden (as well as costs) could be reduced with introducing the help of new technologies.
- Based on the information from the reliability studies, some effort should be made to investigate and use the de-attenuation factors for risk estimates.
- Investigate our main research question in cohorts that were set up later, and thus were potentially less influenced by external hormonal factors such as the use of MHT. It is possible that inflammatory factors play a bigger role when not overshadowed by that of the hormone therapy.
- With the ever increasing range of dietary exposures presented to us in supermarket shelves, it would be valuable to investigate which dietary patterns are nowadays associated with inflammation.

Gaining a comprehensive picture of the potential risk factors for breast cancer, and to what extent they influence the disease, might never be possible due to the complexities of the disease. Nevertheless, research within public health should never lose its focus of helping people making decisions that could lead to improved health. It is of extreme importance to focus research on modifiable risk factors with the aim of reducing the burden of disease in health care facilities in the future.

It is clear to us that healthy dietary patterns are of great health benefit, although the possible mechanisms behind this might never be fully understood. Even though we did not find evidence for inflammation-associated food patterns to be associated with postmenopausal breast cancer, we did find a clear benefit of adhering to the Swedish nutrition recommendations and dietary guidelines, as it was associated with lower levels of systemic inflammation. This should not be ignored, as it reinforces a clear public health message on some of the established consensus in nutrition and health. The role of nutritionists is of great importance when translating science to lay people, in an era of an ever growing number of "fad diets" available, and of increased skepticism towards the research community.

Popular Summary (in English)

Breast cancer is the most common form of cancer affecting women worldwide. In the past decades there has been a lowering in the proportion of women dying from breast cancer, due to advancements in treatment. However, the proportion of women being diagnosed with breast cancer has increased. This health problem affects high and low income countries alike, and has increased more in low income countries in recent years. The difference is thought to be due to the so-called "westernization" of the way of living. In a "Western" lifestyle, people are more inactive, they eat less healthy, and they are more exposed to external hormones. These hormones could be in the form of oral contraceptives, and/or menopausal hormone therapy (MHT). The lifelong exposure to sex hormones such as estrogen is considered to be the main risk factor for breast cancer after menopause. Estrogen stimulates cell division and multiplication, which could lead to cancer in the long run, when it happens in an uncontrolled manner. Besides the use of external hormones, other factors could contribute to an increased exposure to estrogen: early menarche, late menopause, having no children, or having the first child at later age. Breastfeeding is also thought to decrease the risk due to delaying the menstrual cycle.

Obesity, which results from an imbalance of energy intake and energy expenditure, is a major risk factor for many diseases such as cardiovascular disease, diabetes and many types of cancer. Women with higher body mass index (BMI) and higher waist-to-hip ratio are more likely to be diagnosed with breast cancer after menopause than normal-weight women. This is likely to be due to the fact that the adipose tissue is metabolically active: it produces several hormones and inflammation markers (of which adipokines are an example). These metabolites will affect the metabolism both within the adipose tissue and outside (when transported in the blood stream). After menopause, the circulating levels of estrogen decrease because the ovaries stop the production of these sex hormones. However, this is not the case in obese women, as estrogens are still produced, mostly in the adipose tissue. Obesity is thus associated with higher levels of circulating estrogens after menopause, but also with more systemic inflammation.

Higher levels of systemic inflammation could be the result of several factors other than obesity: inflammatory diseases (such as rheumatoid arthritis), unhealthy diets, inactivity, smoking, and alcohol consumption. Inflammation is also associated

with cancer, as it intervenes in several steps of cancer development and progression. We hypothesized that an inflammatory environment, promoted by a Western lifestyle, is associated with postmenopausal breast cancer risk.

In this doctoral thesis, which is based on four studies, we investigated the role of diet in the development of postmenopausal breast cancer, highlighting systemic inflammation as a possible link. We used the Malmö Diet and Cancer (MDC) cohort, where 28,098 people were investigated between 1991 and 1996. Participants underwent body measurements and blood sampling, answered a lifestyle questionnaire, and reported dietary habits. Information regarding the diagnosis of breast cancer and other diseases was retrieved from several national Swedish registries, until December 2010.

Our results showed (study I) that some biomarkers of inflammation (ox-LDL, IL-1β, IL-6, IL-8, and TNF-α) were reasonably reliable. This is of importance when researchers want to investigate disease processes and use biomarkers as indicators of these processes. If biomarkers are not stable over time, the conclusions drawn will not be reliable. In study II, we observed that lower levels of several inflammation markers were linked to a healthy diet, following the Swedish Nutrition Recommendations and Dietary Guidelines. We used a diet quality index (DQI-SNR) to quantify diet quality based on adherence to the recommendations of 6 components: saturated fatty acids, polyunsaturated fatty acids, fish and shellfish, dietary fiber, fruits and vegetables, and sucrose.

Three biomarkers of inflammation (ox-LDL, IL-1β, and TNF-α) were associated with breast cancer, in study III. However, the associations diverged: as we expected and hypothesized, IL-1β was associated with increased risk, but the other two were associated with decreased risk. The associations remained similar even after taking obesity into account. In the final study (study IV), we observed that dietary patterns associated with the three inflammation markers were not linked to postmenopausal breast cancer, after controlling for other important factors.

Taken together, our results suggest that although overall diet quality is an important ally in reduced systemic inflammation, its role in postmenopausal breast cancer development/prevention is yet unclear.

Populärvetenskaplig sammanfattning (in Swedish)

Bröstcancer är den vanligaste formen av cancer som drabbar kvinnor över hela världen. Under de senaste årtiondena har andelen kvinnor som dör av bröstcancer minskat, främst på grund av framsteg inom behandling, men andelen som får diagnosen bröstcancer har ökat. Sjukdomen är numera ett hälsoproblem i såväl hög- som låginkomstländer och har framför allt ökat i låginkomstländer under de senaste åren. Anledningen tros vara det ”västerländska” sättet att leva. Med en ”västerländsk” livsstil, är människor mer inaktiva, äter mindre hälsosamt, och använder oftare hormon-preparat, i form av p-piller, och/eller hormonläkemedel efter klimakteriet. Livslång överexponering för könshormoner såsom östrogen anses vara den viktigaste riskfaktorn för bröstcancer efter klimakteriet. Östrogen stimulerar celldelning och cellförökning, vilket kan leda till cancer på lång sikt, om det ske på ett okontrollerat sätt. Förutom hormon användning, bidrar även andra faktorer till en ökad exponering för östrogen: tidig pubertet, sent klimakterium, barnlöshet och att föda sitt första barn vid högs ålder. Amning anses också minska östrogen exponeringen, eftersom lång amningsperiod medför färre menstruationscykler.

Fetma är resultatet av en obalans mellan energiintag och energiförbrukning, och är en viktig riskfaktor för många sjukdomar såsom hjärt-kärlsjukdomar, diabetes och många typer av cancer. Kvinnor med fetma (högre BMI) och högre midja-till-höft-kvot är mer benägna att få diagnosen bröstcancer efter klimakteriet än normalviktiga kvinnor. Detta beror sannolikt på att den fettväven är metabolt aktiv: den producerar flera hormoner och ämnen kopplade till inflammation (varav adipocytokiner är ett exempel). Dessa ämnen påverkar metabolismen både inom och utanför fettvävnaden (då de transporteras i blodet). Efter klimakteriet, minskar östrogennivåerna i blodet eftersom äggstockarna slutar producera könshormoner. Så är dock inte fallet hos överviktiga kvinnor, eftersom östrogen fortsätter att produceras i fettväven. Fetma är alltså förknippad med högre nivåer av cirkulerande östrogen efter klimakteriet, men också med inflammation som kan påverka hela kroppen.

En högre nivå av systemisk inflammation kan bero på flera andra faktorer än fetma: inflammatoriska sjukdomar (såsom reumatoid artrit), ohälsosam kost,

inaktivitet, rökning och alkoholkonsumtion. Inflammation anses vara kopplad till flera steg av cancerutveckling. Därför är vårt antagande att en inflammatorisk miljö, som stöds av en västerländsk livsstil, är förknippad med postmenopausal bröstcancerrisk.

I denna avhandling, som är baserad på fyra studier, undersökte vi kostens roll vid utveckling av postmenopausal bröstcancer, och speciellt om systemisk inflammation är en möjlig länk. Vi använde Malmö Kost Cancer (MKC) studien, där 28,098 personer undersöktes mellan 1991 och 1996. Deltagarnas kroppssammansättning undersöktes (inklusive vikt och längd), de lämnade blodprover och gav detaljerad information om sina kostvanor och livsstil. Information om bröstcancer diagnos och om andra sjukdomar hämtades från flera nationella svenska register, fram till december 2010.

Våra resultat visade (studie I) att vissa biomarkörer för inflammation (blodnivåer av ox-LDL, IL-1β, IL-6, IL-8 och TNF-α) är någorlunda pålitliga. Detta är av betydelse för forskare som önskar undersöka sjukdomsprocesser och då använda biomarkörer som indikatorer på dessa processer. Om biomarkörerna inte är stabila över tiden, kommer slutsatserna inte vara tillförlitliga. I studie II såg vi att lägre nivåer av flera inflammationsmarkörer var kopplade till en hälsosam kost (enligt de svenska näringsrekommendationerna och kostråden). Vi använde ett kostkvalitetsindex (DQI-SNR) baserat på 6 komponenter: mättade fettsyror, fleromättade fettsyror, fisk och skaldjur, kostfiber, frukt och grönsaker, och sackaros, för att kvantifiera kostkvalité och följsamhet till rekommendationerna.

I studie III, visade tre biomarkörer för inflammation (ox-LDL, IL-1β, and TNF-α) samband med bröstcancer, men sambanden hade olika riktning. Som förväntat var IL-1β kopplat till ökad risk, medan de andra två var kopplade till minskad risk. Sambanden förblev desamma även när vi tog hänsyn till graden av fetma. I den fjärde studien (studie IV), kunde vi konstatera att kostvanor som visade samband med de tre inflammationsmarkörerna inte var kopplade till postmenopausal bröstcancer, när analysmodellen kontrollerade för andra viktiga faktorer.

Sammantaget tyder våra resultat på att, även om en hälsosam kost är en viktig faktor som bidrar till minskad grad av systemisk inflammation, är det ännu oklart hur kostens kvalité bidrar till utveckling, eller förebyggande, av postmenopausal bröstcancer.

Resumo para a comunidade não científica (in Portuguese)

O cancro da mama é o tipo de cancro que mais afecta mulheres em todo o mundo. Nas últimas décadas, os avanços no tratamento levaram a que a proporção de mulheres que morre devido a cancro da mama tenha diminuído. Ainda assim, a proporção de mulheres diagnosticadas com esta doença tem aumentado. Este problema de saúde afecta igualmente países desenvolvidos e em desenvolvimento, mas, recentemente, o número de casos tem aumentado mais rapidamente nos países em desenvolvimento. Pensa-se que esta diferença se deverá à ocidentalização do estilo de vida. Num estilo de vida moderno (também conhecido como ocidental), as pessoas são mais sedentárias, alimentam-se de forma menos saudável, e estão mais expostas a hormonas externas. Tais hormonas podem vir na forma de contraceptivos orais e/ou de terapia hormonal na menopausa (THM). A exposição ao longo da vida a hormonas sexuais, como os estrogénios, é considerado o maior factor de risco para o cancro da mama após a menopausa. Os estrogénios são responsáveis, entre outras coisas, por estimular a divisão e multiplicação celular, o que se acontecer de forma descontrolada a longo prazo, poderá levar ao desenvolvimento de cancro. Para além do uso de hormonas externas, outros factores podem contribuir para um aumento de exposição aos estrogénios: idade da menarca (primeira menstruação) precoce, idade da menopausa atrasada, não ter filhos, ou ter o primeiro filho mais tarde. Pensa-se que a amamentação está também associada a uma diminuição do risco, por atrasar o ciclo menstrual.

A obesidade, que resulta de um desiquilíbrio entre energia ingerida e energia gasta, é um dos maiores factores de risco associados a doenças como as doenças cardiovasculares, diabetes, e vários tipos de cancro. Mulheres com maior índice de massa corporal (IMC) e maior rácio de perímetros de cintura/anca, são mais propensas a ser diagnosticadas com cancro da mama pós-menopausa do que mulheres com peso normal. É provável que tal se deva ao facto de o tecido adiposo ser um tecido metabolicamente activo, que produz várias hormonas e biomarcadores de inflamação (dos quais as adipocitocinas são um exemplo). Estes metabolitos podem afectar o metabolismo localmente (no tecido adiposo) e distalmente (quando transportados pela corrente sanguínea). Após a menopausa, os

níveis de estrogénios em circulação diminuem, dado que os ovários param a sua produção. Contudo, este não é o caso em mulheres obesas, pois a produção de estrogénios será continuada pelo tecido adiposo. A obesidade está portanto associada a maiores níveis de estrogénios em circulação após a menopausa, mas também a maior inflamação sistémica.

Elevados níves de inflamação sistémica podem resultar de outros factores que não a obesidade: doenças inflamatórias (como a artrite reumatóide), alimentação pouco saudável, inactividade física, uso de tabaco, e consumo de álcool. A inflamação está também associada ao cancro, pois é fulcral a sua participação em vários passos no processo de desenvolvimento e progressão do cancro. A nossa hipótese é que um ambiente inflamatório, promovido por um modo de vida ocidental, esteja associado ao aumento do risco de cancro da mama pós-menopausa.

Nesta tese de doutoramento, que é baseada em 4 artigos científicos, investigamos o papel da alimentação no desenvolvimento do cancro da mama pós-menopausa, evidenciando a inflamação sistémica como um possível caminho. Para isso, foi usada a coorte "Malmö Diet and Cancer" (MDC), que engloba 28098 pessoas investigadas entre 1991 e 1996. Os intervenientes no estudo foram submetidos a medições corporais, forneceram amostras de sangue, responderam a um extenso questionário, e reportaram os seus hábitos alimentares. A informação relacionada com o diagnóstico do cancro da mama, assim como outras doenças, foi retirada de vários registos nacionais (suecos), até Dezembro de 2010.

Os nossos resultados mostraram (no estudo I) que vários biomarcadores de inflamação (ox-LDL, IL-1β, IL-6, IL-8 e TNF-α) são razoavelmente fiáveis. Isto é importante quando investigadores pretendem averiguar processos de doença e usam biomarcadores como indicadores de tais processos. Se os biomarcadores não forem estáveis ao longo do tempo, as conclusões retiradas dos estudos não serão confiáveis. No estudo II, observamos que valores mais baixos de biomarcadores de inflamação estão associados a uma alimentação saudável, que segue as recomendações nutricionais e as guias alimentares suecas. Usámos um índice de qualidade alimentar (DQI-SNR) para quantificar a qualidade da alimentação baseada na aderência às recomendações a 6 componentes: ácidos gordos saturados, ácidos gordos poliinsaturados, peixe e marisco, fibra alimentar, frutas e vegetais, e sacarose.

Os nossos resultados (no estudo III) mostraram também que três biomarcadores de inflamação (ox-LDL, IL-1β e TNF-α) estão associados ao cancro da mama. Contudo, as associações divergiram: IL-1β está associado a um aumento de risco (como esperado), mas os outros dois biomarcadores estão associados a uma diminuição do risco. As associações mantiveram-se idênticas após o controlo de outros factores (confundidores), como a obsidade. No último estudo (estudo IV), obeservamos que os padrões alimentares associados aos três biomarcadores de

inflamação não estão associados ao cancro da mama pós-menopausa, após o controlo para outros factores importantes.

Em conjunto, os nossos resultados sugerem que apesar da importância duma alimentação de alta qualidade na redução da inflamação sistémica, o seu papel no desenvolvimento/prevenção do cancro da mama pós-menopausa ainda não é claro.

Acknowledgments

"No man is an island"

- John Donne

This thesis is a product of many contributors, and I am extremely thankful to everyone who touched me and my work along the way. I see this thesis as a product of history.

Part I

I am who I am because of, firstly, my grandparents. I would like to start by thanking my grandfather **Manuel Dias** ("*avô*") and my grandmother **Marcelina Grossinho** ("*avó*"), for being who they are. Thank you for deciding to move to another city in the 60ies and forward; searching for a better life and for better health care for my father. Back in those days it represents a much more courageous act than me moving to another country nowadays. *Avô*, thank you for working endless hours for a low pay; for cycling to work in another city (around 1h each way); and for hanging on through those difficult times always with a smile. I know no one ever thanked you before, and if they would try you would promptly just answer "I just did what I had to do". *Avó*, thank you for being the backbone of the family. I know that being a housewife is underrated, but you were always and still are more important than that; behind every great man's achievement is an even greater woman's support. Thank you both for all you have also done during my existence, I believe that you contributed to a great childhood for me and my brother and gave us many tools to develop as persons and fully functional adults. I would also like to appreciate my grandmother from the mother side, **Maria da Conceição Serôdio**, for contributing with a great childhood place to play.

The second generation after that is also of great importance: I would like to express my gratitude to my parents, my father **Fernando Dias**, and my mother, **Maria de Fátima Alves**. I am a combination of you and the environment. Thank you *daddy* for always bringing the best side of me: fun, happy, friendly, and

relaxed towards life. Thank you also for the great values you taught me and that I will carry with me in politics, morals, and life in general. Thank you also for the countless great philosophical conversations we would have late at night, and for holding my hand through all major thunderstorms (both real and metaphorical). To my *mom*, I would like to thank her for giving me more of the "dark side", making me a control freak, giving me all the necessary work ethics, lecturing me about Portuguese grammar (more than 6 years abroad take its toll even on a mother tongue), and for pushing me forward all the way to and through the PhD. All of those characteristics can be tough in a personal context, but they got me to where I am today, and I am very thankful for that. If women's rights are something that still needs to be advocated for nowadays even in free societies, it for sure starts at home, with striving mothers that believe that their daughters can achieve more than they were given the opportunity during their time. I also learned by watching your example, taking your bachelor's and master's degrees during my existence, in order to improve and advance in your career. Thank you both! You are both my Yin and my Yang.

Then it is my generation. I cannot express in words how much I owe to my brother, **João Dias**. You have been the best example of how a big brother makes an impression on his little sister. Children like to copy, so in early years I would find myself wanting the same things as you: doing swimming competitions, being sporty, listening to rock music, playing with cars; and climbing trees, to name a few. You are a true inspiration and you managed to fight your way until where you are now: you finished your master's degree; you are an IronMan (and you and I know that an athlete is 60% of his own motivation); and, at the same time, you have raised a wonderful family. You proved to everyone that you were capable of becoming a normal fully functional adult. Furthermore, you love what you do! That is a great achievement: to be able to excel at work because your motivation is endless. A big thank you to **Sandra Leitão** for not only taking good care of my brother, but for being also a great source of inspiration!

The next generation deserves my appreciation as well. **Francisco**, my little nephew, came to the world when I was in a tough spot, and reminded me that even in times of sadness, a beautiful thing can happen. Thank you for all the love you give me unconditionally and constantly. After came little **Mafalda** along, making the Swedish winter feel cozy and comfy. I love you both until the end of times.

To all my cousins, aunts and uncles, and the rest of my family, a big thank you!

Part II

The second and important part of my acknowledgments goes to the true inspirational figures I had along the way throughout my young career. Some people, even without knowing, contributed to my current "self". The first person I would like to thank is Professor **Pedro Moreira**. You were the most inspirational lecturer I had during my bachelor's in Nutrition and Food Sciences. Every time, you managed to spike our curiosity about varied subjects regarding food and health and by the end of each lecture I would find myself asking the same question: "how can I get into research?" That was the steering wheel that led me to Karolinska Institutet (KI) for my bachelor's thesis. The same thesis I had to defend to you as my opponent. From all the constructive criticism you gave me, I understood I would have to dig deeper into research to better understand it. And now, here I am. Thank you!

So many people I could be thankful for my time at KI, and those acknowledgments are already part of my master's thesis. I could briefly mention **Eric Poortvliet** for inspiring me to learn more and more about statistics, and also Professor **Agneta Yngve** for being supportive during the time I was there. While writing my master's thesis I had the opportunity to become an intern at the World Health Organization, under Dr **João Breda** supervision. It was a great chance to work in such high level institution in public health, and there was where I decided I really wanted to pursue a PhD degree. During those 2 years at KI, I was inspired even more to go into the research field, as I had the opportunity to work closely with some research projects from different angles. I appreciate all the coffee breaks (aka "*fika*"), all the conference meetings, all the courses, and everyone that little by little became part of my routines and my early days in Sweden.

Part III

The third and last section of the acknowledgments includes the most recent part of history, and probably the most important since it directly influenced its result. I have to give the most important and special acknowledgment to my supervisor, Professor Emerita **Elisabet Wirfält**. There are so many things to thank you for; thank you for taking the wild decision of having me coming for a 6-month research experience all the way from across Europe. I don't know what you saw during our Skype interview, but I am extremely thankful for the chance you've given me. Some years ago I would complain that I needed a real mentor, someone who would work together with me in my development, someone who would care. You revealed to be much more than that from the beginning. We had a good chemistry since day 1 and I can only wish to continue working with people like you throughout my life. So again, thank you Elisabet (or "Elis" as I would fondly

nick name you in my agenda) for all the opportunities you have given me to develop as a student, as a researcher, and as a person. You were a true mentor!

Another grateful thank you goes to my main supervisor for half of my PhD, Professor **Gunilla N. Fredrikson**. Thank you for giving me the opportunity to complete my PhD. Thank you for the constructive criticism in all my papers, for believing in me without even knowing me well, and for always being there when I needed.

A warm thank you goes to my other co-supervisors that made it possible for me to complete this thesis; Professor **Jan Nilsson**, **Ulrika Ericson,** and **Isabel Drake**. Thank you for all the advising and inspiration talks. A special thank you goes also to **Isabel Drake** for all the support, especially in the last paper. That one set the tone for this thesis. Thank you all for the constructive criticism on each respective areas of expertise: inflammation, nutrition and statistics.

I would also like to thank all my co-authors for their great contributions: **Gunnar Engström**, **Bo Gullberg**, **Bo Hedblad**, **Margaretha Persson**, **Signe Borgquist**, **Alexandru Schiopu**, **Ragnar Alm**, and **Harry Björkbacka**. I have learned from each one of you and advanced a step further for each piece of knowledge you transmitted to me. I would also like to thank Professor **Marju Orho-Melander** for letting me feel part of a big research group, and for all the good feedback.

Also colleagues, friends, and other researchers that had an impact on me and my performance though these years deserve my appreciation: a big extended thank you! That list includes **Sophie Hellstrand**, **Gull Rukh**, **Louise Brunkwall**, **George Hindy**, **Emily Sonestedt**, **Peter Wallström**, **Tanja Stocks** (thank you for making me feel more "normal" during lunch!), **Yan Borné**, **Eliana Garcia Vaz**, **Alexander Perfilyev**, among others. I might have forgotten a name (it always happens on these occasions) so feel free to add your name to the list: _____. A warm thank you goes to **Gunilla Hughes Wulkan** for helping me with so many details towards the end of my PhD; life certainly became easier with you around. A big "thank you" to **Andrew Clelland** for taking the photo in the back cover of this thesis and to **Pedro Carvalho**, **Patricia Herbert** and **Alexandra Mendes** for the help with translating the popular science summary into Portuguese (yes, I needed help with my mother tongue!).

Big thanks to **Sanjib Saha**, **Ali Kiadaliri**, **Jens Hellstrand** and **Gawain Heckley** (and respective wives: **Labonya Saha**, **Naghmeh Yazdani**, **Sophie Hellstrand** and **Lotta Heckley**) for being such great company! For all the mocking, all the inspirational talks about the unknown future, and all the competitions (being that in swimming, go-karting, beach volley or publication fikas!). Thank you for keeping me sharp and giving me the opportunity to get to know other cultures (especially through feeding!).

I have to thank my dear friend **Sophie Hellstrand**. There are several reasons why I should thank you, but I start by acknowledging you for giving me the chance to fight for me PhD here! If it had not been for you, I don't really know what would be of me today. These small gestures make a big difference. Thank you for the best friendship and travel companionship during all of these years, even when we were under stress. I am happy I could share so many special moments in our lives with you.

I would also like to thank my dear friend **Gull Rukh**. It was a great inspiration to see you doing so well during the final stage of your PhD. Sharing the office with you became more than just 2 people sitting in the same room. Thank you for all the wise advices. Thank you **Yan Borné** for coming in to our office every now and then to complain about something (it felt good to know we were not alone!), and for all the talks and laughter we shared the three of us.

A warm thank you goes to all my friends and adopted families in Portugal, UK, Sweden, and all around the world. Friends that deserve a spot in this section and that are always in my heart are **Patrícia Esbjörnsson Dias**, **Raquel Lemos**, **Patricia Herbert**, **Ana Paula**, (and respective husbands/partners **Philip Esbjörnsson Dias**, **Johan Björserud**, **Richard Herbert**, and **Fernando Paes de Vasconcellos**), **Agota Puskas**, **Alexandra Mendes**, **Inês Monteiro**, **Cláudio Baptista**, and many others I cannot fit in here. Thank you all for all the Skype talks, the trips together, the sessions of game night and late deep talks, the (very) loud laughs we shared, and the unlimited friendship. If I was a ship, you were my anchor and safe harbor. I would also like to thank **Rui Martins** for all the positive input during the last semester, while writing my thesis, which had a great impact in the outcome. And lastly, I am very thankful to all the friends I made along the way: the "Portuguese crew", the "Internations crew", the "international PhD students group", the "gym group", and all the others in my salsa, bachata, kizomba and Swedish courses.

A final word of appreciation goes to all the **participants of the Malmö Diet and Cancer study** which made hundreds of theses (not just mine) and thousands of scientific articles possible. You probably did not realize this at the time, how great your contribution to science would be. Thank you! And extended thanks to everyone who worked to make this cohort possible during the designing, planning, and data collection phases.

References

1. Schatten, H., *Cell and Molecular Biology of Breast Cancer*. Humana Press ed. 2013: Springer New York Heidelberg Dordrecht London.
2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
3. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
4. Veronesi, U., et al., *Breast cancer.* Lancet, 2005. **365**(9472): p. 1727-41.
5. Sobin, L.H. and C. Wittekind, *International Union Against Cancer TNM Classification of Malignant Tumours.* Sixth ed. 2002, New York, NY: John Wiley & Sons.
6. Ervik, M., et al. *Cancer Today*. 2016 31/10/2016]; Available from: http://gco.iarc.fr/today.
7. Key, T.J., P.K. Verkasalo, and E. Banks, *Epidemiology of breast cancer.* Lancet Oncol, 2001. **2**(3): p. 133-40.
8. Bray, F., P. McCarron, and D.M. Parkin, *The changing global patterns of female breast cancer incidence and mortality.* Breast Cancer Res, 2004. **6**(6): p. 229-39.
9. Ziegler, R.G., et al., *Migration patterns and breast cancer risk in Asian-American women.* J Natl Cancer Inst, 1993. **85**(22): p. 1819-27.
10. Li, C.I., *Breast Cancer Epidemiology*. 2010: Spring New York Dordrecht Heidelberg London.
11. NORDCAN. *The NORCAN project - Cancer Fact Sheets and Statistics*. 08/07/2016 31/10/2016]; Available from: http://www-dep.iarc.fr/NORDCAN.
12. WHO. *World Health Organization. Health topics - Risk factors*. 31/10/2016]; Available from: http://www.who.int/topics/risk_factors/en/.
13. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern Epidemiology*. Third ed. 2008, Philadelphia, PA, USA: Lippincott Williams & Wilkins.
14. Dumitrescu, R.G. and I. Cotarla, *Understanding breast cancer risk -- where do we stand in 2005?* J Cell Mol Med, 2005. **9**(1): p. 208-21.
15. Gilbert, E.S., *Ionising radiation and cancer risks: what have we learned from epidemiology?* Int J Radiat Biol, 2009. **85**(6): p. 467-82.
16. Preston, D.L., et al., *Solid cancer incidence in atomic bomb survivors: 1958-1998.* Radiat Res, 2007. **168**(1): p. 1-64.

17. Williams, D., *Radiation carcinogenesis: lessons from Chernobyl.* Oncogene, 2008. **27 Suppl 2**: p. S9-18.
18. Biglia, N., et al., *Management of risk of breast carcinoma in postmenopausal women.* Endocr Relat Cancer, 2004. **11**(1): p. 69-83.
19. Evans, J.S., J.E. Wennberg, and B.J. McNeil, *The influence of diagnostic radiography on the incidence of breast cancer and leukemia.* N Engl J Med, 1986. **315**(13): p. 810-5.
20. Harris, T.J. and F. McCormick, *The molecular pathology of cancer.* Nat Rev Clin Oncol, 2010. **7**(5): p. 251-65.
21. Ford, D., et al., *Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium.* Am J Hum Genet, 1998. **62**(3): p. 676-89.
22. Thorlacius, S., et al., *Population-based study of risk of breast cancer in carriers of BRCA2 mutation.* Lancet, 1998. **352**(9137): p. 1337-9.
23. Easton, D., D. Ford, and J. Peto, *Inherited susceptibility to breast cancer.* Cancer Surv, 1993. **18**: p. 95-113.
24. Nelson, L.R. and S.E. Bulun, *Estrogen production and action.* J Am Acad Dermatol, 2001. **45**(3 Suppl): p. S116-24.
25. Key, T.J. and P.K. Verkasalo, *Endogenous hormones and the aetiology of breast cancer.* Breast Cancer Res, 1999. **1**(1): p. 18-21.
26. WCRF, *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Prespective*. 2007, Washington, DC, USA: American Institute for Cancer Research.
27. Ross, R.K., et al., *Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin.* J Natl Cancer Inst, 2000. **92**(4): p. 328-32.
28. McCormack, V.A. and I. dos Santos Silva, *Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(6): p. 1159-69.
29. Pettersson, A., et al., *Mammographic density phenotypes and risk of breast cancer: a meta-analysis.* J Natl Cancer Inst, 2014. **106**(5).
30. Simpson, P.T., et al., *The diagnosis and management of pre-invasive breast disease: pathology of atypical lobular hyperplasia and lobular carcinoma in situ.* Breast Cancer Res, 2003. **5**(5): p. 258-62.
31. Sanders, M.E., et al., *The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up.* Cancer, 2005. **103**(12): p. 2481-4.
32. Baer, H.J., et al., *Adult height, age at attained height, and incidence of breast cancer in premenopausal women.* Int J Cancer, 2006. **119**(9): p. 2231-5.
33. Friedenreich, C.M., *Review of anthropometric factors and breast cancer risk.* Eur J Cancer Prev, 2001. **10**(1): p. 15-32.
34. WCRF/AICR, *Continuous Update Project. Food, Nutrition, Physical Activity, and the Prevention of Breast Cancer*. 2010.

35. McTiernan, A., *Mechanisms linking physical activity with cancer.* Nat Rev Cancer, 2008. **8**(3): p. 205-11.
36. Ennour-Idrissi, K., E. Maunsell, and C. Diorio, *Effect of physical activity on sex hormones in women: a systematic review and meta-analysis of randomized controlled trials.* Breast Cancer Res, 2015. **17**(1): p. 139.
37. Dumitrescu, R.G. and P.G. Shields, *The etiology of alcohol-induced breast cancer.* Alcohol, 2005. **35**(3): p. 213-25.
38. Macacu, A., et al., *Active and passive smoking and risk of breast cancer: a meta-analysis.* Breast Cancer Res Treat, 2015. **154**(2): p. 213-24.
39. Sugimura, T., *Nutrition and dietary carcinogens.* Carcinogenesis, 2000. **21**(3): p. 387-95.
40. Doll, R. and R. Peto, *The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today.* J Natl Cancer Inst, 1981. **66**(6): p. 1191-308.
41. Willett, W.C., *Diet, nutrition, and avoidable cancer.* Environ Health Perspect, 1995. **103 Suppl 8**: p. 165-70.
42. Armstrong, B. and R. Doll, *Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices.* Int J Cancer, 1975. **15**(4): p. 617-31.
43. Hunter, D.J., et al., *Cohort studies of fat intake and the risk of breast cancer--a pooled analysis.* N Engl J Med, 1996. **334**(6): p. 356-61.
44. Key, T.J., et al., *The effect of diet on risk of cancer.* Lancet, 2002. **360**(9336): p. 861-8.
45. Mattisson, I., et al., *Intakes of plant foods, fibre and fat and risk of breast cancer--a prospective study in the Malmo Diet and Cancer cohort.* Br J Cancer, 2004. **90**(1): p. 122-7.
46. Sonestedt, E., et al., *Plant foods and oestrogen receptor alpha- and beta-defined breast cancer: observations from the Malmo Diet and Cancer cohort.* Carcinogenesis, 2008. **29**(11): p. 2203-9.
47. Brennan, S.F., et al., *Dietary patterns and breast cancer risk: a systematic review and meta-analysis.* Am J Clin Nutr, 2010. **91**(5): p. 1294-302.
48. Balkwill, F. and A. Mantovani, *Inflammation and cancer: back to Virchow?* Lancet, 2001. **357**(9255): p. 539-45.
49. Lushchak, V.I., *Free radicals, reactive oxygen species, oxidative stress and its classification.* Chem Biol Interact, 2014. **224**: p. 164-75.
50. Liou, G.Y. and P. Storz, *Reactive oxygen species in cancer.* Free Radic Res, 2010. **44**(5): p. 479-96.
51. Gutteridge, J.M., *Biological origin of free radicals, and mechanisms of antioxidant protection.* Chem Biol Interact, 1994. **91**(2-3): p. 133-40.
52. Bae, Y.S., et al., *Macrophages generate reactive oxygen species in response to minimally oxidized low-density lipoprotein: toll-like receptor 4- and spleen tyrosine kinase-dependent activation of NADPH oxidase 2.* Circ Res, 2009. **104**(2): p. 210-8, 21p following 218.

53. Skulachev, V.P., *Mitochondria-targeted antioxidants as promising drugs for treatment of age-related brain diseases.* J Alzheimers Dis, 2012. **28**(2): p. 283-9.
54. Basaga, H.S., *Biochemical aspects of free radicals.* Biochem Cell Biol, 1990. **68**(7-8): p. 989-98.
55. Altemose, B., et al., *Aldehydes in Relation to Air Pollution Sources: A Case Study around the Beijing Olympics.* Atmos Environ (1994), 2015. **109**: p. 61-69.
56. Eiserich, J.P., et al., *Dietary antioxidants and cigarette smoke-induced biomolecular damage: a complex interaction.* Am J Clin Nutr, 1995. **62**(6 Suppl): p. 1490S-1500S.
57. Bailey, S.M. and C.C. Cunningham, *Contribution of mitochondria to oxidative stress associated with alcoholic liver disease.* Free Radic Biol Med, 2002. **32**(1): p. 11-6.
58. Zima, T., et al., *Oxidative stress, metabolism of ethanol and alcohol-related diseases.* J Biomed Sci, 2001. **8**(1): p. 59-70.
59. Ji, L.L., *Redox signaling in skeletal muscle: role of aging and exercise.* Adv Physiol Educ, 2015. **39**(4): p. 352-9.
60. Sacheck, J.M. and J.B. Blumberg, *Role of vitamin E and oxidative stress in exercise.* Nutrition, 2001. **17**(10): p. 809-14.
61. Keaney, J.F., Jr., et al., *Obesity and systemic oxidative stress: clinical correlates of oxidative stress in the Framingham Study.* Arterioscler Thromb Vasc Biol, 2003. **23**(3): p. 434-9.
62. Matsuda, M. and I. Shimomura, *Increased oxidative stress in obesity: implications for metabolic syndrome, diabetes, hypertension, dyslipidemia, atherosclerosis, and cancer.* Obes Res Clin Pract, 2013. **7**(5): p. e330-41.
63. Rogers, A.E., S.H. Zeisel, and J. Groopman, *Diet and carcinogenesis.* Carcinogenesis, 1993. **14**(11): p. 2205-17.
64. Shacter, E. and S.A. Weitzman, *Chronic inflammation and cancer.* Oncology (Williston Park), 2002. **16**(2): p. 217-26, 229; discussion 230-2.
65. Steinberg, D., *Low density lipoprotein oxidation and its pathobiological significance.* J Biol Chem, 1997. **272**(34): p. 20963-6.
66. Tappel, A.L., *Vitamin E and free radical peroxidation of lipids.* Ann N Y Acad Sci, 1972. **203**: p. 12-28.
67. Parthasarathy, S., et al., *Oxidized low-density lipoprotein.* Methods Mol Biol, 2010. **610**: p. 403-17.
68. Kindt, T.J., *Kuby Immonology*. sixth ed. 2007, New York: W. H. Freeman.
69. Harris, R.E., et al., *Genetic induction and upregulation of cyclooxygenase (COX) and aromatase (CYP19): an extension of the dietary fat hypothesis of breast cancer.* Med Hypotheses, 1999. **52**(4): p. 291-2.
70. Singh-Ranger, G. and K. Mokbel, *The role of cyclooxygenase-2 (COX-2) in breast cancer, and implications of COX-2 inhibition.* Eur J Surg Oncol, 2002. **28**(7): p. 729-37.

71. Davies, G., et al., *Cyclooxygenase-2 (COX-2), aromatase and breast cancer: a possible role for COX-2 inhibitors in breast cancer chemoprevention.* Ann Oncol, 2002. **13**(5): p. 669-78.
72. Piazza, G.A., et al., *NSAIDs: Old Drugs Reveal New Anticancer Targets.* Pharmaceuticals (Basel), 2010. **3**(5): p. 1652-1667.
73. Gonzalez-Perez, A., L.A. Garcia Rodriguez, and R. Lopez-Ridaura, *Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis.* BMC Cancer, 2003. **3**: p. 28.
74. Mazhar, D., R. Ang, and J. Waxman, *COX inhibitors and breast cancer.* Br J Cancer, 2006. **94**(3): p. 346-50.
75. Khan, M.M., *Immunopharmacology*. 2008: Springer Science+Business Media LLC.
76. Biswas, S.K., *Does the Interdependence between Oxidative Stress and Inflammation Explain the Antioxidant Paradox?* Oxid Med Cell Longev, 2016. **2016**: p. 5698931.
77. Tan, H.Y., et al., *The Reactive Oxygen Species in Macrophage Polarization: Reflecting Its Dual Role in Progression and Treatment of Human Diseases.* Oxid Med Cell Longev, 2016. **2016**: p. 2795090.
78. Porta, C., et al., *Cellular and molecular pathways linking inflammation and cancer.* Immunobiology, 2009. **214**(9-10): p. 761-77.
79. Allavena, P., et al., *Pathways connecting inflammation and cancer.* Curr Opin Genet Dev, 2008. **18**(1): p. 3-10.
80. Aggarwal, B.B., et al., *Inflammation and cancer: how hot is the link?* Biochem Pharmacol, 2006. **72**(11): p. 1605-21.
81. Grivennikov, S.I., F.R. Greten, and M. Karin, *Immunity, inflammation, and cancer.* Cell, 2010. **140**(6): p. 883-99.
82. Allavena, P., et al., *The inflammatory micro-environment in tumor progression: the role of tumor-associated macrophages.* Crit Rev Oncol Hematol, 2008. **66**(1): p. 1-9.
83. Baron, J.A., *Epidemiology of non-steroidal anti-inflammatory drugs and cancer.* Prog Exp Tumor Res, 2003. **37**: p. 1-24.
84. Chan, D.S., et al., *Circulating C-Reactive Protein and Breast Cancer Risk-Systematic Literature Review and Meta-analysis of Prospective Cohort Studies.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(10): p. 1439-49.
85. Rajala, M.W. and P.E. Scherer, *Minireview: The adipocyte--at the crossroads of energy homeostasis, inflammation, and atherosclerosis.* Endocrinology, 2003. **144**(9): p. 3765-73.
86. Vona-Davis, L. and D.P. Rose, *Adipokines as endocrine, paracrine, and autocrine factors in breast cancer risk and progression.* Endocr Relat Cancer, 2007. **14**(2): p. 189-206.
87. Furukawa, S., et al., *Increased oxidative stress in obesity and its impact on metabolic syndrome.* J Clin Invest, 2004. **114**(12): p. 1752-61.

88. Toniolo, P., *Application of biomarkers in cancer epidemiology*. IARC Scientific publications,. 1997, Lyon; Carey, NC: International Agency for Research on Cancer ; Oxford University Press distributor. xvii, 318 p.
89. Vineis, P., *Sources of variation in biomarkers.* IARC Sci Publ, 1997(142): p. 59-71.
90. Tworoger, S.S. and S.E. Hankinson, *Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis.* Cancer Causes Control, 2006. **17**(7): p. 889-99.
91. Strimbu, K. and J.A. Tavel, *What are biomarkers?* Curr Opin HIV AIDS, 2010. **5**(6): p. 463-6.
92. Gisev, N., J.S. Bell, and T.F. Chen, *Interrater agreement and interrater reliability: key concepts, approaches, and applications.* Res Social Adm Pharm, 2013. **9**(3): p. 330-8.
93. McGraw, K.O. and S.P. Wong, *Forming inferences about some intraclass correlation coefficients.* Psychological Methods, 1996. **1**(1): p. 30-46.
94. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.
95. Shrout, P.E. and J.L. Fleiss, *Intraclass correlations: uses in assessing rater reliability.* Psychol Bull, 1979. **86**(2): p. 420-8.
96. Rosner, B., D. Spiegelman, and W.C. Willett, *Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error.* Am J Epidemiol, 1992. **136**(11): p. 1400-13.
97. Hallgren, K.A., *Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial.* Tutor Quant Methods Psychol, 2012. **8**(1): p. 23-34.
98. Frijhoff, J., et al., *Clinical Relevance of Biomarkers of Oxidative Stress.* Antioxid Redox Signal, 2015. **23**(14): p. 1144-70.
99. Delimaris, I., et al., *Oxidized LDL, serum oxidizability and serum lipid levels in patients with breast or ovarian cancer.* Clin Biochem, 2007. **40**(15): p. 1129-34.
100. Germano, G., P. Allavena, and A. Mantovani, *Cytokines as a key component of cancer-related inflammation.* Cytokine, 2008. **43**(3): p. 374-9.
101. Esquivel-Velazquez, M., et al., *The role of cytokines in breast cancer development and progression.* J Interferon Cytokine Res, 2015. **35**(1): p. 1-16.
102. Nicolini, A., A. Carpi, and G. Rossi, *Cytokines in breast cancer.* Cytokine Growth Factor Rev, 2006. **17**(5): p. 325-37.
103. Willett, W., *Nutritional Epidemiology*. third ed. Vol. 40. 2013, New York, USA: Oxford University Press.
104. Gross, R., et al., *The four dimensions of food and nutrition security: definitions and concepts*. 2000.

105. Willett, W., *Nutritional epidemiology: issues and challenges.* Int J Epidemiol, 1987. **16**(2): p. 312-7.
106. Shim, J.S., K. Oh, and H.C. Kim, *Dietary assessment methods in epidemiologic studies.* Epidemiol Health, 2014. **36**: p. e2014009.
107. Block, G., *A review of validations of dietary assessment methods.* Am J Epidemiol, 1982. **115**(4): p. 492-505.
108. Willett, W.C., G.R. Howe, and L.H. Kushi, *Adjustment for total energy intake in epidemiologic studies.* Am J Clin Nutr, 1997. **65**(4 Suppl): p. 1220S-1228S; discussion 1229S-1231S.
109. Hu, F.B., *Dietary pattern analysis: a new direction in nutritional epidemiology.* Curr Opin Lipidol, 2002. **13**(1): p. 3-9.
110. Jacobs, D.R., Jr. and L.C. Tapsell, *Food, not nutrients, is the fundamental unit in nutrition.* Nutr Rev, 2007. **65**(10): p. 439-50.
111. Elmstahl, S. and B. Gullberg, *Bias in diet assessment methods--consequences of collinearity and measurement errors on power and observed relative risks.* Int J Epidemiol, 1997. **26**(5): p. 1071-9.
112. Moeller, S.M., et al., *Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006.* J Am Diet Assoc, 2007. **107**(7): p. 1233-9.
113. Reedy, J., et al., *Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study.* Am J Epidemiol, 2010. **171**(4): p. 479-87.
114. Hoffmann, K., et al., *Application of a new statistical method to derive dietary patterns in nutritional epidemiology.* Am J Epidemiol, 2004. **159**(10): p. 935-44.
115. Calder, P.C., et al., *Dietary factors and low-grade inflammation in relation to overweight and obesity.* Br J Nutr, 2011. **106 Suppl 3**: p. S5-78.
116. Barbaresko, J., et al., *Dietary pattern analysis and biomarkers of low-grade inflammation: a systematic literature review.* Nutr Rev, 2013. **71**(8): p. 511-27.
117. Wirfalt, E., et al., *Postmenopausal breast cancer is associated with high intakes of omega6 fatty acids (Sweden).* Cancer Causes Control, 2002. **13**(10): p. 883-93.
118. Wirfalt, E., et al., *Fat from different foods show diverging relations with breast cancer risk in postmenopausal women.* Nutr Cancer, 2005. **53**(2): p. 135-43.
119. Wirfalt, E., et al., *Food sources of fat and sex hormone receptor status of invasive breast tumors in women of the Malmo Diet and Cancer cohort.* Nutr Cancer, 2011. **63**(5): p. 722-33.
120. Ericson, U., et al., *Folate intake, methylenetetrahydrofolate reductase polymorphisms, and breast cancer risk in women from the Malmo Diet and Cancer cohort.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(4): p. 1101-10.

121. Valenzuela, A., J. Sanhueza, and S. Nieto, *Cholesterol oxidation: health hazard and the role of antioxidants in prevention.* Biol Res, 2003. **36**(3-4): p. 291-302.
122. SCB. *Population by region and every fith year*. 2016 17/10/2016]; Available from: http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101A/FolkmangdTatort/table/tableViewLayout1/?rxid=4f261aa4-3a43-485b-8ee7-47ff5a8134ff.
123. Berglund, G., et al., *The Malmo Diet and Cancer Study. Design and feasibility.* J Intern Med, 1993. **233**(1): p. 45-51.
124. Manjer, J., et al., *Invitation to a population-based cohort study: differences between subjects recruited using various strategies.* Scand J Public Health, 2002. **30**(2): p. 103-12.
125. Riboli, E. and R. Kaaks, *The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S6-14.
126. Riboli, E., *Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC).* Ann Oncol, 1992. **3**(10): p. 783-91.
127. Manjer, J., et al., *The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants.* Eur J Cancer Prev, 2001. **10**(6): p. 489-99.
128. Hedblad, B., et al., *Relation between insulin resistance and carotid intima-media thickness and stenosis in non-diabetic subjects. Results from a cross-sectional study in Malmo, Sweden.* Diabet Med, 2000. **17**(4): p. 299-307.
129. Berg, K.E., et al., *Elevated CD14++CD16- monocytes predict cardiovascular events.* Circ Cardiovasc Genet, 2012. **5**(1): p. 122-31.
130. Pero, R.W., et al., *Quality control program for storage of biologically banked blood specimens in the Malmo Diet and Cancer Study.* Cancer Epidemiol Biomarkers Prev, 1998. **7**(9): p. 803-8.
131. Pero, R.W., et al., *The Malmo biological bank.* J Intern Med, 1993. **233**(1): p. 63-7.
132. Holvoet, P., et al., *Oxidized LDL and malondialdehyde-modified LDL in patients with acute coronary syndromes and stable coronary artery disease.* Circulation, 1998. **98**(15): p. 1487-94.
133. Adamsson Eryd, S., et al., *Incidence of coronary events and case fatality rate in relation to blood lymphocyte and neutrophil counts.* Arterioscler Thromb Vasc Biol, 2012. **32**(2): p. 533-9.
134. Persson, M., et al., *The epidemiology of Lp-PLA(2): distribution and correlation with cardiovascular risk factors in a population-based cohort.* Atherosclerosis, 2007. **190**(2): p. 388-96.
135. Cotoi, O.S., et al., *Plasma S100A8/A9 correlates with blood neutrophil counts, traditional risk factors, and cardiovascular disease in middle-aged*

healthy individuals. Arterioscler Thromb Vasc Biol, 2014. **34**(1): p. 202-10.
136. Wirfalt, E. and E. Sonestedt, *The modified diet history methodology of the Malmö Diet Cancer cohort*. 2016.
137. Callmer, E., et al., *Dietary assessment methods evaluated in the Malmo food study*. J Intern Med, 1993. **233**(1): p. 53-7.
138. Hwasser, A. *MKC kostmetod*. 2016 17/10/2016]; Available from: http://www.med.lu.se/klinvetmalmo/befolkningsstudier/malmoe_kost_cancer_och_malmoe_foerebyggande_medicin/malmoe_kost_cancer/mkc_kostmetod.
139. Elmstahl, S., et al., *The Malmo Food Study: the reproducibility of a novel diet history method and an extensive food frequency questionnaire.* Eur J Clin Nutr, 1996. **50**(3): p. 134-42.
140. Elmstahl, S., et al., *The Malmo Food Study: the relative validity of a modified diet history method and an extensive food frequency questionnaire for measuring food intake.* Eur J Clin Nutr, 1996. **50**(3): p. 143-51.
141. Riboli, E., et al., *The Malmo Food Study: validity of two dietary assessment methods for measuring nutrient intake.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S161-73.
142. Wirfalt, E., et al., *A methodological report from the Malmo Diet and Cancer study: development and evaluation of altered routines in dietary data processing.* Nutr J, 2002. **1**: p. 3.
143. Mattisson, I., et al., *Misreporting of energy: prevalence, characteristics of misreporters and influence on observed risk estimates in the Malmo Diet and Cancer cohort.* Br J Nutr, 2005. **94**(5): p. 832-42.
144. Black, A.E., *Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations.* Int J Obes Relat Metab Disord, 2000. **24**(9): p. 1119-30.
145. Goldberg, G.R., et al., *Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording.* Eur J Clin Nutr, 1991. **45**(12): p. 569-81.
146. Barlow, L., et al., *The completeness of the Swedish Cancer Register: a sample survey for year 1998.* Acta Oncol, 2009. **48**(1): p. 27-33.
147. Drake, I., et al., *Development of a diet quality index assessing adherence to the Swedish nutrition recommendations and dietary guidelines in the Malmo Diet and Cancer cohort.* Public Health Nutr, 2011. **14**(5): p. 835-45.
148. Drake, I., et al., *Scoring models of a diet quality index and the predictive capability of mortality in a population-based cohort of Swedish men and women.* Public Health Nutr, 2013. **16**(3): p. 468-78.

149. Hlebowicz, J., et al., *A high diet quality is associated with lower incidence of cardiovascular events in the Malmo diet and cancer cohort.* PLoS One, 2013. **8**(8): p. e71095.
150. Sonestedt, E., B. Gullberg, and E. Wirfalt, *Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer.* Public Health Nutr, 2007. **10**(8): p. 769-79.
151. Sonestedt, E., et al., *Past food habit change is related to obesity, lifestyle and socio-economic factors in the Malmo Diet and Cancer Cohort.* Public Health Nutr, 2005. **8**(7): p. 876-85.
152. Richardson, M.T., et al., *Comprehensive evaluation of the Minnesota Leisure Time Physical Activity Questionnaire.* J Clin Epidemiol, 1994. **47**(3): p. 271-81.
153. Li, C., et al., *Ability of physical activity measurements to assess health-related risks.* Eur J Clin Nutr, 2009. **63**(12): p. 1448-51.
154. Psychiatrists, R.C.o., *Alcohol: Our Favorite Drug*. 1986, London: Tavistock.
155. WHO. *BMI classification - Global Database on Body Mass Index*. 17/10/2016]; Available from: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html.
156. Ericson, U., et al., *High folate intake is associated with lower breast cancer incidence in postmenopausal women in the Malmo Diet and Cancer cohort.* Am J Clin Nutr, 2007. **86**(2): p. 434-43.
157. Merlo, J., et al., *Self-administered questionnaire compared with a personal diary for assessment of current use of hormone therapy: an analysis of 16,060 women.* Am J Epidemiol, 2000. **152**(8): p. 788-92.
158. Montgomery, D.C., *Design and Analysis of Experiments*. 8th ed. 2012: John Wiley Sons.
159. Wasenius, A., et al., *Diurnal and monthly intra-individual variability of the concentration of lipids, lipoproteins and apoproteins.* Scand J Clin Lab Invest, 1990. **50**(6): p. 635-42.
160. Widjaja, A., et al., *Within- and between-subject variation in commonly measured anthropometric and biochemical variables.* Clin Chem, 1999. **45**(4): p. 561-6.
161. Basiotis, P.P., et al., *Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence.* J Nutr, 1987. **117**(9): p. 1638-41.
162. Hankinson, S.E., et al., *Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period.* Cancer Epidemiol Biomarkers Prev, 1995. **4**(6): p. 649-54.
163. Clendenen, T.V., et al., *Temporal reliability of cytokines and growth factors in EDTA plasma.* BMC Res Notes, 2010. **3**: p. 302.
164. Gu, Y., et al., *Reproducibility of serum cytokines and growth factors.* Cytokine, 2009. **45**(1): p. 44-9.

165. Ho, G.Y., et al., *Variability of serum levels of tumor necrosis factor-alpha, interleukin 6, and soluble interleukin 6 receptor over 2 years in young women.* Cytokine, 2005. **30**(1): p. 1-6.
166. Hofmann, J.N., et al., *Intra-individual variability over time in serum cytokine levels among participants in the prostate, lung, colorectal, and ovarian cancer screening Trial.* Cytokine, 2011. **56**(2): p. 145-8.
167. Lee, S.A., et al., *Intra-individual variation of plasma adipokine levels and utility of single measurement of these biomarkers in population-based studies.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(11): p. 2464-70.
168. Navarro, S.L., et al., *Reliability of serum biomarkers of inflammation from repeated measures in healthy individuals.* Cancer Epidemiol Biomarkers Prev, 2012. **21**(7): p. 1167-70.
169. Cava, F., et al., *Biological variation of interleukin 6 (IL-6) and soluble interleukin 2 receptor (sIL2R) in serum of healthy individuals.* Cytokine, 2000. **12**(9): p. 1423-5.
170. Gonzalez, C., et al., *Biological variation of interleukin-1beta, interleukin-8 and tumor necrosis factor-alpha in serum of healthy individuals.* Clin Chem Lab Med, 2001. **39**(9): p. 836-41.
171. Nestel, P.J., et al., *Circulating inflammatory and atherogenic biomarkers are not increased following single meals of dairy foods.* Eur J Clin Nutr, 2012. **66**(1): p. 25-31.
172. Blackburn, P., et al., *Postprandial variations of plasma inflammatory markers in abdominally obese men.* Obesity (Silver Spring), 2006. **14**(10): p. 1747-54.
173. Kaplan, R.C., et al., *Within-individual stability of obesity-related biomarkers among women.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(6): p. 1291-3.
174. Fernandez-Real, J.M., et al., *Circulating interleukin 6 levels, blood pressure, and insulin sensitivity in apparently healthy men and women.* J Clin Endocrinol Metab, 2001. **86**(3): p. 1154-9.
175. Sites, C.K., et al., *Menopause-related differences in inflammation markers and their relationship to body fat distribution and insulin-stimulated glucose disposal.* Fertil Steril, 2002. **77**(1): p. 128-35.
176. Sonestedt, E., et al., *Variation in fasting and non-fasting serum enterolactone concentrations in women of the Malmo Diet and Cancer cohort.* Eur J Clin Nutr, 2008. **62**(8): p. 1005-9.
177. Stumpf, K. and H. Adlercreutz, *Short-term variations in enterolactone in serum, 24-hour urine, and spot urine and relationship with enterolactone concentrations.* Clin Chem, 2003. **49**(1): p. 178-81.
178. Skogstrand, K., et al., *Effects of blood sample handling procedures on measurable inflammatory markers in plasma, serum and dried blood spot samples.* J Immunol Methods, 2008. **336**(1): p. 78-84.
179. Chaturvedi, A.K., et al., *Evaluation of multiplexed cytokine and inflammation marker measurements: a methodologic study.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(9): p. 1902-11.

180. Lash, G.E., et al., *Comparison of three multiplex cytokine analysis systems: Luminex, SearchLight and FAST Quant.* J Immunol Methods, 2006. **309**(1-2): p. 205-8.
181. Ledur, A., et al., *Variable estimates of cytokine levels produced by commercial ELISA kits: results using international cytokine standards.* J Immunol Methods, 1995. **186**(2): p. 171-9.
182. Ziegler-Heitbrock, L. and T.P. Hofer, *Toward a refined definition of monocyte subsets.* Front Immunol, 2013. **4**: p. 23.
183. Ridker, P.M., *Clinical application of C-reactive protein for cardiovascular disease detection and prevention.* Circulation, 2003. **107**(3): p. 363-9.
184. Ridker, P.M., et al., *C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women.* N Engl J Med, 2000. **342**(12): p. 836-43.
185. Fredrikson, G.N., et al., *Association between diet, lifestyle, metabolic cardiovascular risk factors, and plasma C-reactive protein levels.* Metabolism, 2004. **53**(11): p. 1436-42.
186. Serrano-Martinez, M., et al., *A Mediterranean dietary style influences TNF-alpha and VCAM-1 coronary blood levels in unstable angina patients.* Eur J Nutr, 2005. **44**(6): p. 348-54.
187. Farhangi, M.A., et al., *White blood cell count in women: relation to inflammatory biomarkers, haematological profiles, visceral adiposity, and other cardiovascular risk factors.* J Health Popul Nutr, 2013. **31**(1): p. 58-64.
188. Nettleton, J.A., et al., *Associations between dietary patterns and flow cytometry-measured biomarkers of inflammation and cellular activation in the Atherosclerosis Risk in Communities (ARIC) Carotid Artery MRI Study.* Atherosclerosis, 2010. **212**(1): p. 260-7.
189. Esposito, K., et al., *Mediterranean diet and metabolic syndrome: An updated systematic review.* Rev Endocr Metab Disord, 2013. **14**(3): p. 255-63.
190. Guallar-Castillon, P., et al., *Major dietary patterns and risk of coronary heart disease in middle-aged persons from a Mediterranean country: the EPIC-Spain cohort study.* Nutr Metab Cardiovasc Dis, 2012. **22**(3): p. 192-9.
191. Kontou, N., et al., *The mediterranean diet in cancer prevention: a review.* J Med Food, 2011. **14**(10): p. 1065-78.
192. Salas-Salvado, J., et al., *The role of diet in the prevention of type 2 diabetes.* Nutr Metab Cardiovasc Dis, 2011. **21 Suppl 2**: p. B32-48.
193. Sofi, F., et al., *Adherence to Mediterranean diet and health status: meta-analysis.* BMJ, 2008. **337**: p. a1344.
194. Akesson, A., et al., *Health effects associated with foods characteristic of the Nordic diet: a systematic literature review.* Food Nutr Res, 2013. **57**.

195. Hoebeeck, L.I., et al., *The relationship between diet and subclinical atherosclerosis: results from the Asklepios Study.* Eur J Clin Nutr, 2011. **65**(5): p. 606-13.
196. Zhao, R., H. Zhou, and S.B. Su, *A critical role for interleukin-1beta in the progression of autoimmune diseases.* Int Immunopharmacol, 2013. **17**(3): p. 658-69.
197. Yudkin, J.S., et al., *Inflammation, obesity, stress and coronary heart disease: is interleukin-6 the link?* Atherosclerosis, 2000. **148**(2): p. 209-14.
198. Soria, G., et al., *Inflammatory mediators in breast cancer: coordinated expression of TNFalpha & IL-1beta with CCL2 & CCL5 and effects on epithelial-to-mesenchymal transition.* BMC Cancer, 2011. **11**: p. 130.
199. Gaiotti, D., et al., *Tumor necrosis factor-alpha promotes human papillomavirus (HPV) E6/E7 RNA expression and cyclin-dependent kinase activity in HPV-immortalized keratinocytes by a ras-dependent pathway.* Mol Carcinog, 2000. **27**(2): p. 97-109.
200. Montesano, R., et al., *Tumour necrosis factor alpha confers an invasive, transformed phenotype on mammary epithelial cells.* J Cell Sci, 2005. **118**(Pt 15): p. 3487-500.
201. Purohit, A., S.P. Newman, and M.J. Reed, *The role of cytokines in regulating estrogen synthesis: implications for the etiology of breast cancer.* Breast Cancer Res, 2002. **4**(2): p. 65-9.
202. Vendramini-Costa, D.B. and J.E. Carvalho, *Molecular link mechanisms between inflammation and cancer.* Curr Pharm Des, 2012. **18**(26): p. 3831-52.
203. Cao, W., et al., *TNF-alpha promotes Doxorubicin-induced cell apoptosis and anti-cancer effect through downregulation of p21 in p53-deficient tumor cells.* Biochem Biophys Res Commun, 2005. **330**(4): p. 1034-40.
204. Izano, M., et al., *Chronic inflammation and risk of colorectal and other obesity-related cancers: The health, aging and body composition study.* Int J Cancer, 2016. **138**(5): p. 1118-28.
205. Krajcik, R.A., S. Massardo, and N. Orentreich, *No association between serum levels of tumor necrosis factor-alpha (TNF-alpha) or the soluble receptors sTNFR1 and sTNFR2 and breast cancer risk.* Cancer Epidemiol Biomarkers Prev, 2003. **12**(9): p. 945-6.
206. Suzuki, K., et al., *Serum oxidized low-density lipoprotein levels and risk of colorectal cancer: a case-control study nested in the Japan Collaborative Cohort Study.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(11 Pt 1): p. 1781-7.
207. Zabirnyk, O., et al., *Oxidized low-density lipoproteins upregulate proline oxidase to initiate ROS-dependent autophagy.* Carcinogenesis, 2010. **31**(3): p. 446-54.
208. Ni, H., H. Liu, and R. Gao, *Serum Lipids and Breast Cancer Risk: A Meta-Analysis of Prospective Cohort Studies.* PLoS One, 2015. **10**(11): p. e0142669.

209. Touvier, M., et al., *Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies.* Br J Nutr, 2015. **114**(3): p. 347-57.
210. Madjid, M., et al., *Leukocyte count and coronary heart disease: implications for risk assessment.* J Am Coll Cardiol, 2004. **44**(10): p. 1945-56.
211. Coussens, L.M. and Z. Werb, *Inflammation and cancer.* Nature, 2002. **420**(6917): p. 860-7.
212. Floor, S.L., et al., *Hallmarks of cancer: of all cancer cells, all the time?* Trends Mol Med, 2012. **18**(9): p. 509-15.
213. Sonnenschein, C. and A.M. Soto, *The aging of the 2000 and 2011 Hallmarks of Cancer reviews: a critique.* J Biosci, 2013. **38**(3): p. 651-63.
214. Simpson, E.R. and K.A. Brown, *Obesity and breast cancer: role of inflammation and aromatase.* J Mol Endocrinol, 2013. **51**(3): p. T51-9.
215. Mittal, R., et al., *Mechanistic Insight of Drug Resistance with Special Focus on Iron in Estrogen Receptor Positive Breast Cancer.* Curr Pharm Biotechnol, 2014.
216. Colditz, G.A., et al., *Risk factors for breast cancer according to estrogen and progesterone receptor status.* J Natl Cancer Inst, 2004. **96**(3): p. 218-28.
217. Vona-Davis, L. and D.P. Rose, *The obesity-inflammation-eicosanoid axis in breast cancer.* J Mammary Gland Biol Neoplasia, 2013. **18**(3-4): p. 291-307.
218. Fung, T.T., et al., *A dietary pattern derived to correlate with estrogens and risk of postmenopausal breast cancer.* Breast Cancer Res Treat, 2012. **132**(3): p. 1157-62.
219. Harris, H.R., L. Bergkvist, and A. Wolk, *An estrogen-associated dietary pattern and breast cancer risk in the Swedish Mammography Cohort.* Int J Cancer, 2015. **137**(9): p. 2149-54.
220. McCann, S.E., et al., *Dietary patterns related to glycemic index and load and risk of premenopausal and postmenopausal breast cancer in the Western New York Exposure and Breast Cancer Study.* Am J Clin Nutr, 2007. **86**(2): p. 465-71.
221. Pot, G.K., et al., *Dietary patterns derived with multiple methods from food diaries and breast cancer risk in the UK Dietary Cohort Consortium.* Eur J Clin Nutr, 2014. **68**(12): p. 1353-8.
222. Schulz, M., et al., *Identification of a dietary pattern characterized by high-fat food choices associated with increased risk of breast cancer: the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study.* Br J Nutr, 2008. **100**(5): p. 942-6.
223. Woo, H.D., et al., *Glycemic index and glycemic load dietary patterns and the associated risk of breast cancer: a case-control study.* Asian Pac J Cancer Prev, 2013. **14**(9): p. 5193-8.
224. Hecht, F., et al., *The role of oxidative stress on breast cancer development and therapy.* Tumour Biol, 2016.

225. Chen, W.Y., et al., *Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk.* JAMA, 2011. **306**(17): p. 1884-90.
226. Koene, R.J., et al., *Shared Risk Factors in Cardiovascular Disease and Cancer.* Circulation, 2016. **133**(11): p. 1104-14.
227. Christenfeld, N.J., et al., *Risk factors, confounding, and the illusion of statistical control.* Psychosom Med, 2004. **66**(6): p. 868-75.
228. Voudouri, K., et al., *Insulin-like growth factor and epidermal growth factor signaling in breast cancer cell growth: focus on endocrine resistant disease.* Anal Cell Pathol (Amst), 2015. **2015**: p. 975495.
229. Potischman, N. and D.L. Weed, *Causal criteria in nutritional epidemiology.* Am J Clin Nutr, 1999. **69**(6): p. 1309S-1314S.
230. Grivennikov, S.I. and M. Karin, *Inflammation and oncogenesis: a vicious connection.* Curr Opin Genet Dev, 2010. **20**(1): p. 65-71.
231. Fosgate, G.T., *Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null.* Emerg Themes Epidemiol, 2006. **3**: p. 7.
232. Myrianthefs, P., et al., *Seasonal variation in whole blood cytokine production after LPS stimulation in normal individuals.* Cytokine, 2003. **24**(6): p. 286-92.
233. Keogh, R.H. and I.R. White, *A toolkit for measurement error correction, with a focus on nutritional epidemiology.* Stat Med, 2014. **33**(12): p. 2137-55.
234. Bingham, S.A., et al., *Are imprecise methods obscuring a relation between fat and breast cancer?* Lancet, 2003. **362**(9379): p. 212-4.
235. Day, N., et al., *Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium.* Int J Epidemiol, 2001. **30**(2): p. 309-17.
236. Waijers, P.M., E.J. Feskens, and M.C. Ocke, *A critical review of predefined diet quality scores.* Br J Nutr, 2007. **97**(2): p. 219-31.
237. Goldbohm, R.A., 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**(6): p. 420-9.
238. Rosell, M.S., et al., *Associations between diet and the metabolic syndrome vary with the validity of dietary intake data.* Am J Clin Nutr, 2003. **78**(1): p. 84-90.
239. Lissner, L., *Measuring food intake in studies of obesity.* Public Health Nutr, 2002. **5**(6A): p. 889-92.
240. Hartman, A.M., et al., *Variability in nutrient and food intakes among older middle-aged men. Implications for design of epidemiologic and validation studies using food recording.* Am J Epidemiol, 1990. **132**(5): p. 999-1012.
241. Subar, A.F., et al., *Differences in reported food frequency by season of questionnaire administration: the 1987 National Health Interview Survey.* Epidemiology, 1994. **5**(2): p. 226-33.

242. Kipnis, V. and L.S. Freedman, *Impact of exposure measurement error in nutritional epidemiology.* J Natl Cancer Inst, 2008. **100**(23): p. 1658-9.
243. Lundqvist, A., et al., *Socioeconomic inequalities in breast cancer incidence and mortality in Europe-a systematic review and meta-analysis.* Eur J Public Health, 2016. **26**(5): p. 804-813.
244. Ericson, U.C., et al., *Increased breast cancer risk at high plasma folate concentrations among women with the MTHFR 677T allele.* Am J Clin Nutr, 2009. **90**(5): p. 1380-9.
245. Pharoah, P.D., et al., *Family history and the risk of breast cancer: a systematic review and meta-analysis.* Int J Cancer, 1997. **71**(5): p. 800-9.