From THE INSTITUTE OF ENVIRONMENTAL MEDICINE
Karolinska Institutet, Stockholm, Sweden

Diet and other environmental factors and the risk of rheumatoid arthritis

in a population-based prospective cohort study

Daniela Di Giuseppe

Karolinska Institutet

Stockholm 2014

Published by Karolinska Institutet. Printed by Åtta.45 Tryckeri AB, 2014

ISBN 978-91-7549-510-1

Diet and other environmental factors and the risk of rheumatoid arthritis in a population-based prospective cohort study

THESIS FOR DOCTORAL DEGREE (Ph.D.)

by

Daniela Di Giuseppe

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Rockefeller, Nobels väg 11

Torsdagen den 24 april, 2014, kl 13.00

Principal Supervisor:
Professor Alicja Wolk
Karolinska Institutet
Institute of Environmental Medicine
Unit of Nutritional Epidemiology

Co-supervisor(s):
Professor Johan Askling
Karolinska Institutet
Department of Medicine, Solna
Clinical Epidemiology Unit

Professor Matteo Bottai
Karolinska Institutet
Institute of Environmental Medicine
Unit of Biostatistics

Opponent:
Professor Kay-Tee Khaw
University of Cambridge
Department of Public Health and Primary Care
Division of Clinical Gerontology

Examination Board:
Professor Paul Dickman
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Professor Thomas Skogh
Linköpings Universitet
Department of Clinical and Experimental Medicine
Division of Rheumatology

Professor Pär Sparén
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

ABSTRACT

Rheumatoid arthritis (RA) is an autoimmune inflammatory disease that affects primarily the joints. RA leads to cartilage destruction and bone erosion, with substantial loss of quality of life. RA is associated with an increased risk of cardiovascular disease, osteoporosis, gastrointestinal disorders, thus increasing disability and mortality. RA affects 0.5-1% of the adult population, and is three times more common among women than among men. Twin studies have shown that the relative contribution of genetic factors to RA is about 50%, leaving the remaining part to environmental factors. Few epidemiological studies have examined risk factors for RA. Even though cigarette smoking is an established risk factor for RA, the role of its characteristics in the development of the disease is less clear. In addition, analyses of other risk factors have led to inconclusive and often conflicting results.

Aims of this thesis were: 1) to analyze the association between characteristics of cigarette smoking (intensity, duration and cessation) and RA risk in a population-based prospective cohort study and by summarizing published evidence; 2) to evaluate the association of alcohol consumption and risk of RA; 3) to estimate the dose-response relationship between long-chain n-3 polyunsaturated fatty acids (PUFAs) and risk of RA; 4) to evaluate long-term intake of alcohol and long-chain n-3 PUFAs, as well as the long-term consumption of fish in relation to RA; 5) to prospectively evaluate the association between physical activity and RA.

The data used to assess the association between selected exposures and the development of RA were obtained by means of questionnaires administered in 1987 and 1997 to the Swedish Mammography Cohort. Among the 35 187 women that did not have RA or non-RA joint conditions before the start of follow-up in 2003, 224 developed RA before 2010. Results showed a twofold increased risk among current smokers compared with never smokers, even when their exposure to smoking was low (<7 cigarettes per day). The risk of RA decreased over time following smoking cessation, but remained elevated after more than 15 years since smoking cessation compared with never smokers. Moderate alcohol consumption (a median of 6 glasses of alcohol per week) was associated with a 37% decreased risk of RA. In addition, long-chain n-3 PUFA dietary intake was inversely associated with RA risk, and women with an intake of more than 0.21 grams per day of long-chain n-3 PUFAs had a 35% decreased risk compared with women with a lower intake (≤0.21 grams per day). A consistent moderate long-term intake of both alcohol and long-chain n-3 PUFAs was associated with a halved risk of RA. Long-term consumption of fish was inversely associated with RA, but after adjustment for their content of long-chain n-3 PUFAs the association disappeared. Leisure-time activity (combination of walking and exercising) was associated with a decreased risk of RA.

Results from this thesis showed that modifiable lifestyle factors, including smoking, alcohol consumption, long-chain n-3 PUFAs intake and physical activity, are associated with RA development.

To Giorgio

LIST OF PUBLICATIONS

I. **Di Giuseppe D**, Orsini N, Alfredsson L, Askling J, Wolk A.
Cigarette smoking and smoking cessation in relation to risk of rheumatoid arthritis in women. *Arthritis Res Ther*. 2013 Apr 22;15(2):R56.

II. **Di Giuseppe D**, Discacciati A, Orsini N, Wolk A.
Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. *Arthritis Res Ther*. 2014 Mar 5;16(2):R61

III. **Di Giuseppe D**, Alfredsson L, Bottai M, Askling J, Wolk A.
Long term alcohol intake and risk of rheumatoid arthritis in women: a population-based cohort study. *BMJ*, 2012 Jul 10;345:e4230

IV. **Di Giuseppe D**, Wallin A, Bottai M, Askling J, Wolk A.
Long-term intake of dietary long-chain n-3 polyunsaturated fatty acids and risk of rheumatoid arthritis: a prospective cohort study of women. *Ann Rheum Dis*. 2013 Aug 12.

V. **Di Giuseppe D**, Bottai M, Askling J, Wolk A.
Physical activity and risk of rheumatoid arthritis in women: a population-based prospective study. *Submitted*

RELATED PUBLICATIONS

Wallin A, **Di Giuseppe D**, Burgaz A, Håkansson N, Wolk A.
Validity of food frequency questionnaire-based estimates of long-term long-chain n-3 polyunsaturated fatty acid intake. *Eur J Nutr.* 2013 Jul 26

Wallin A, **Di Giuseppe D**, Orsini N, Patel PS, Forouhi NG, Wolk A.
Fish consumption, dietary long-chain n-3 fatty acids, and risk of type 2 diabetes: systematic review and meta-analysis of prospective studies. *Diabetes Care*, 2012 Apr;35(4):918-29.

Di Giuseppe D, Crippa A, Orsini N, Wolk A.
Fish consumption and risk of rheumatoid arthritis: a dose-response meta-analysis. *Submitted*

Di Giuseppe D, Wolk A.
Diet and rheumatoid arthritis development: what does the evidence say? *International Journal of Clinical Rheumatology,* accepted for publication

CONTENTS

LIST OF ABBREVIATIONS

ACPA	Anti-citrullinated peptide antibody
ACR	American college of rheumatology
BMI	Body mass index
CI	Confidence interval
CRP	C-reactive protein
DHA	Docosahexaenoic acid
DMARDs	Disease-modifying anti-rheumatic drugs
DPA	Docosapentaenoic acid
EIRA	Epidemiological Investigation of Rheumatoid Arthritis
EPA	Eicosapentaenoic acid
ESR	Erythrocyte sedimentation rate
EULAR	European league against rheumatism
Fc	Fragment crystallizable
FFQ	Food frequency questionnaire
HLA	Human leukocyte antigen
HR	Hazard ratio
IPR	Inpatient Register
MHC	Major histocompatibility complex
MMPs	Matrix metallaproteinases
NBHW	National board of health and welfare
NOQAS	Newcastle–Ottawa Quality Assessment Scale
NSAIDs	Non-steroid anti-inflammatory drugs
OPR	Outpatient Register
PAR	Population attributable risk
PIN	Personal identification number
PPF	Population prevented fraction
PPV	Positive predictive value
PUFA	Polyunsaturated fatty acids
RA	Rheumatoid arthritis
RANKL	Receptor activator of nuclear factor kappa-B ligand
RF	Rheumatoid factor
RR	Relative risk
SMC	Swedish Mammography Cohort
SRR	Swedish Rheumatology Register
TNF	Tumor necrosis factor
YLD	Years lived with disability

1 INTRODUCTION

Rheumatoid arthritis is an autoimmune inflammatory disease that affects the joints.[1] Rheumatoid arthritis affects 0.5-1% of adults in developed countries[2 3] and has tremendous influences on quality of life and on costs for both individuals and society.[4]

Twin studies have shown that the contribution of genetic factors to rheumatoid arthritis is about 50%, leaving the remaining part to environmental factors.[5] Cigarette smoking is one of the few environmental factors that has been linked to the development of rheumatoid arthritis. Other lifestyle and environmental factors, such as diet and physical activity, have been examined in very few studies with mixed results.

The main aim of this thesis was to examine the role of cigarette smoking, alcohol consumption, diet (i.e. long-chain n-3 polyunsaturated fatty acids), and physical activity in the etiology of rheumatoid arthritis.

2 BACKGROUND

2.1 RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a chronic autoimmune inflammatory disease that principally attacks the joints. RA is classified as an autoimmune disease because the immune system attacks the individual's own cells and tissues. RA can be characterized by the production of two known antibodies, rheumatoid factor (RF) and anti-citrullinated peptide antibody (ACPA), against common autoantigens that are widely expressed outside the joints. RA is a chronic inflammatory disease, because of the imbalance of inflammatory cytokines, which causes joint destruction.

RA is a complex genetic disease, with several genes, environmental factors, and stochastic factors acting in concert to cause pathological events (**Figure 1**).[1] Twin studies have shown that the relative contribution of genetic factors to RA is about 50%, leaving the remaining part to environmental factors.[5] An issue in RA prevention is the timing of exposure to environmental factors, since some studies have suggested that the influence of environmental risk factors on RA could begin even before birth.[6 7]

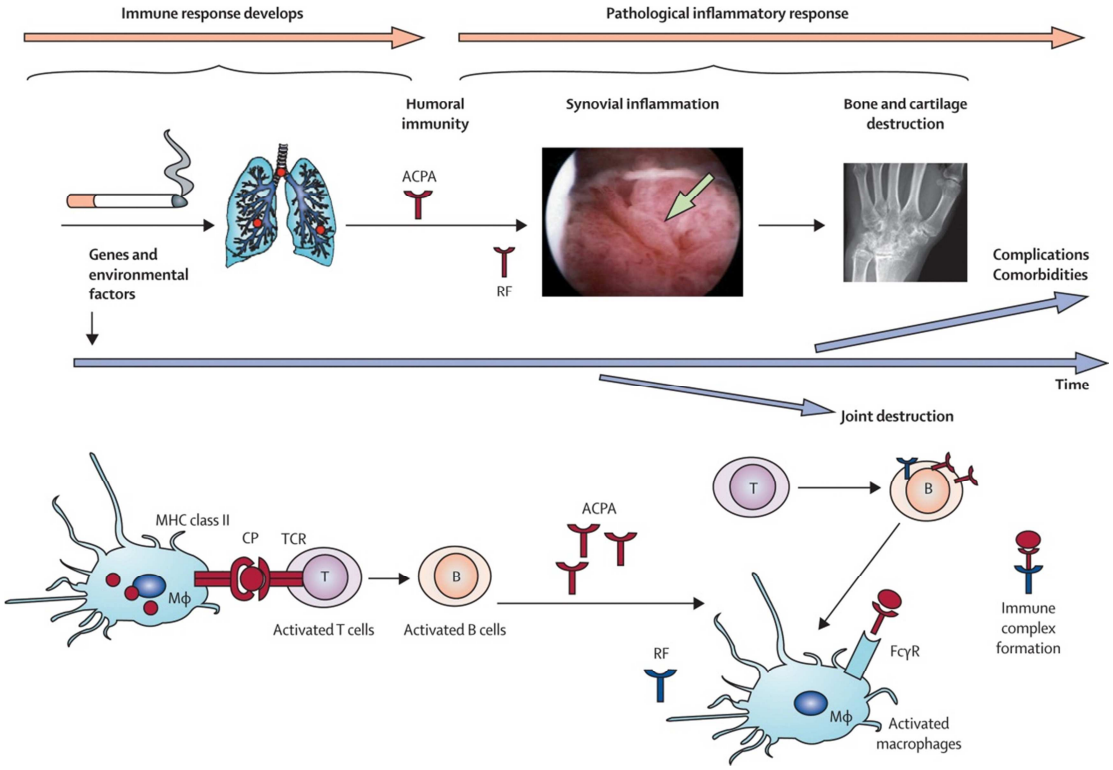

Figure 1. Hypothetical model for molecular pathogenesis of ACPA-positive rheumatoid arthritis. *Reprinted with permission from Lancet*[1]

ACPA: anti-citrullinated peptide antibody, RF: rheumatoid factor, CP: citrullinated proteins and peptide, MHC: major histocompatibility complex, TCR: T cell receptor, FcγR: fragment crystallizable gamma receptor.

2.1.1.1 Pathogenesis

RA can be considered a clinical syndrome that includes several disease subsets.[8] These different subsets involve several inflammatory cascades, which all lead to persistent synovial inflammation and damage to articular cartilage and bone,[9] by actions that include the innate as well as the adaptive immune system and imbalances in regulation of cytokines and other inflammatory mediators.

One such key inflammatory cascade in RA pathogenesis includes overproduction and overexpression of tumor necrosis factor (TNF).[10] TNF overproduction has several causes and leads to overproduction of multiple cytokines including interleukin 6, which also drives persistent inflammation and joint destruction.[11]

Synovial inflammation is characterized by the presence of many interacting immune cells.[1] Antigen-presenting cells, such as B cells, communicate with T cells through the Major Histocompatibility Complex (MHC). Macrophages activated by signals from T cells and by immune complexes produce many proinflammatory cytokines, including TNF, interleukin 1 and interleukin 6. These molecules enhance cytokine release, production of cartilage-destructive enzymes and expression of bone destruction-related molecules, such as RANKL (receptor activator of nuclear factor kappa-B ligand).[12]

Cartilage destruction is caused by matrix metalloproteinases (MMPs), molecules that degrade the structural proteins of the extracellular matrix of cartilage and that are released by proinflammatory cytokines. Bone erosion is caused by osteoclast activated from macrophage-like precursors after stimulation by RANKL,[13] by T cells that act directly on osteoclasts and by fibroblast-like synoviocytes active in pannus tissue.[14]
Fibroblast-like synoviocytes show abnormal behavior in RA that also leads to fibroblast invading cartilage which correlates with joint destruction.[15] However, it is still not clear if RA starts in the joints and then spreads out into the bones or the other way around (**Figure 2**).[16]

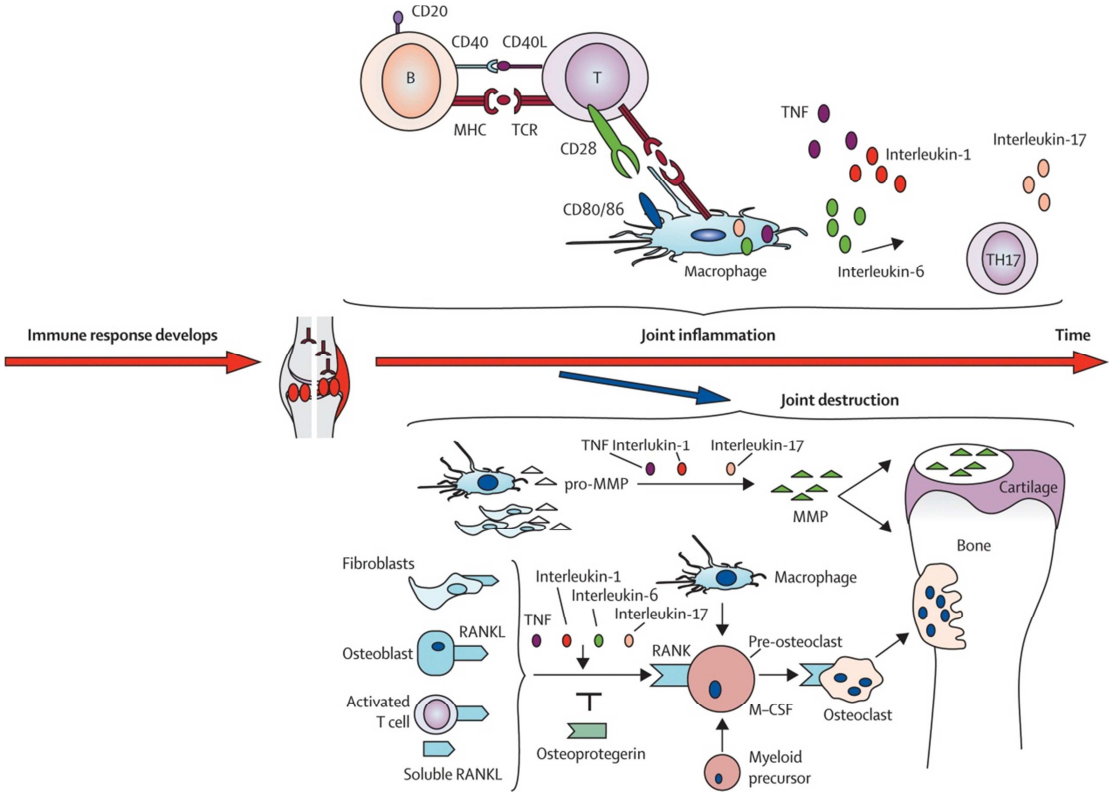

Figure 2. Immunological pathways in the arthritic joint (upper part shows joint inflammation, lower part joint destruction). *Reprinted with permission from Lancet*[1]

CD: cluster of differentiation, MHC: major histocompatibility complex, TCR: T cell receptor, TNF: tumor necrosis factor, TH17: T helper 17 cells, RANKL: receptor activator of nuclear factor kappa-B ligand, MMP: matrix metalloproteinase, M-CSF: macrophage colony-stimulating factor

2.1.1.2 Genes

The reported association between certain HLA-D/DR alleles and risk for RA suggests that MHC class II-dependent T-cell and B-cell activation are major drivers of the disease.[17] Most of HLA-DR alleles involved in RA have a common aminoacid motif, named the shared epitope, in the β-chain of the HLA-DR molecule.[18] A second identified gene involved in the development of RA is PTPN22, a gene that codes for a tyrosine phosphatase which has a role in T-cell and B-cell signaling.[19] HLA DRB1 shared epitope and PTPN22 risk alleles are associated only with ACPA or RF positive RA, thus indicating that subsets of RA should be analyzed as separate entities. One of the HLA alleles thought to be involved in ACPA negative RA is HLA DRB1*03, however this association needs to be confirmed.[20]

2.1.1.3 Treatment strategies

Treatment strategies of RA include early dynamic and tightly controlled treatment and targeted approaches.

Pharmaceutical treatments for RA include cortisone, non-steroid anti-inflammatory drugs (NSAIDs) and disease-modifying anti-rheumatic drugs (DMARDs, such as methotrexate, sulfadalazine, hydroxychloroquine, leflunomide, and glucocorticoids).

In addition, biological treatments include TNF-alpha (infliximab, etanercept, adalimumab) and interleukin 1 (anakinra) blockers and inhibitors, and agents targeting T and B lymphocytes (abatacept and rituximab). Treatment with inflammatory inhibitors should be administered effectively and as early as possible in the course of the disease to reduce future joint damage and functional disorders.[1]

Surgical treatments, such as arthroplasty, arthrodesis, synovectomy of joints and tendons, nerve decompression and reconstructive tendon surgery, have decreased in the past years in favor of biological treatments.

2.1.1.4 Co-morbidities

Co-morbid conditions are common in patients with RA.[21] Co-morbidity means the existence of two or more diseases in the same person. Some co-morbidities, as cardiovascular disease, are associated with RA, and their frequencies and impact are increased in RA patients. Co-morbidities increase disability and shorten life expectancy, thus increasing impact and mortality of RA.[22 23]

Cardiovascular diseases are the most important co-morbidities in those with RA. Patients with RA are at increased risk of ischemic heart disease[24] and heart failure[25] compared to the general population. The increased cardiovascular risk in RA patients can be explained by a higher prevalence of traditional cardiovascular risk factors, such as smoking, or by effects of treatments, such as NSAIDs, corticosteroid and DMARDs (methotrexate).[21] Treatments for RA are also related to gastrointestinal disorders,[26] as well as malignancies[27] and infections.[28]

Osteoporosis is another co-morbidity observed in RA patients, due to a shared mechanism via cytokine-induced osteoclast activation. RA patients also have a higher frequency and severity of periodontal disease, which shares common RA risk factors, including smoking and HLA-DR B1 04 alleles, as well as pathological processes.[29]

2.1.1.5 Incidence and prevalence

Rheumatoid arthritis affects 0.5-1% of adults in developed countries.[2 3] The disease is three times more frequent in women than in men. Onset of RA is usually in middle-age, but it may occur at any ages. The prevalence increases with age.[30] Incidence ranges from 5 to 50 per 100 000 adults in developed countries and increases with age as well.[31] The reported disease prevalence is higher in northern Europe and North America compared with developing countries, which may reflect differences in risk factors as well as case ascertainment and survival.[32]

A recent study estimated the RA incidence in Sweden.[33] Using the Swedish National Patient Register, 8 826 incident cases were identified during the period 2006-2008 in a population of 7 331 508 people aged ≥18 years. The overall incidence was 41 per 100 000 (56 for women, 25 for men). The incidence increased with age and peaked in the 70-79 year age group for both women and men (**Figure 3**).

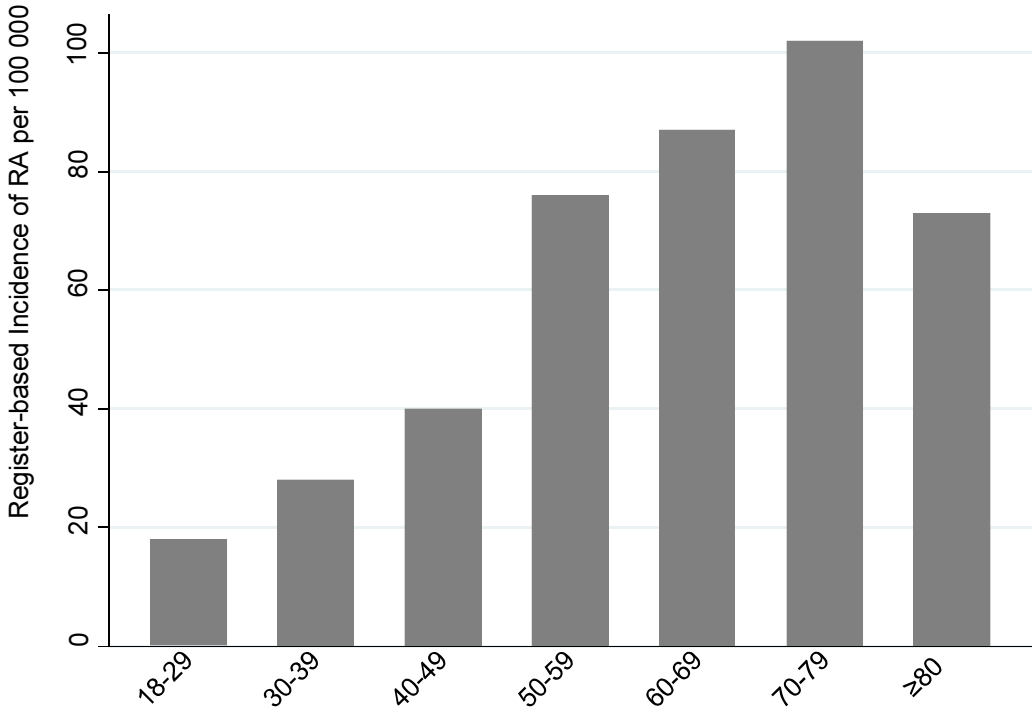

Figure 3. Mean annual incidence of adult rheumatoid arthritis per 100 000 women in Sweden, 2006-2008[33]

2.1.1.6 Costs and quality of life

RA is associated with substantial loss of quality of life and elevated costs for both patients and society.

Costs related to RA include direct medical and non-medical cost and indirect costs (i.e. productivity loss, increased co-morbidities, burden for caregivers, and premature mortality).[34 35] Introduction of biological drugs increased considerably the costs related to drug use, as they are 30-40 times more costly than traditional DMARDs.[36] RA-attributable direct health care costs have been estimated at €14 billion per year in Europe.[4] Costs of RA management increase with increasing disease severity, in particular with functional disability. In Sweden, the costs of RA increased by one third between 1990 and 2010, and were approximately €600 million in 2010.[37] Of the total costs, drug related costs increased from 3% to 33% between 1990 and 2010, while indirect costs (including sick leave and disability pension) decreased.

The decrease in quality of life is considerable in patients with RA that regularly score among the groups with lowest utility values.[4] Utility may be defined as "a cardinal measure of the preference for, or desirability of, a specific level of health status or specific health outcome" and range between 0 and 1.[38] Mean utilities in population samples of RA have been estimated to be between 0.45 and 0.55. As a comparison, only multiple sclerosis appears to have a similar effect on quality of life. Moreover, RA accounts for 0.8% of total global Years Lived with Disability (YLD), around the same percentage as obsessive-compulsive disorders and meningitis.[39]

2.1.1.7 Diagnosis

A diagnosis of RA prior to 2010 was given according to the classification criteria defined by the American College of Rheumatology (ACR) in the mid-1980s.[40] According to these criteria, a patient was defined as having RA if he or she met at least four of the following seven criteria:

- Morning stiffness lasting at least 1 hour, present for at least 6 weeks
- At least three joint areas simultaneously with soft-tissue swelling or fluid, for at least 6 weeks
- At least one area swollen in a wrist, metacarpaophalangeal, or proximal interphalangeal joint, for at least 6 weeks
- Simultaneous involvement of the same joint areas on both sides of the body, for at least 6 weeks
- Subcutaneous nodules seen by a doctor
- Positive rheumatoid factor (RF)
- Radiographic changes on hand and wrist radiographs (erosions or unequivocal bony decalcification).

However, these criteria have been criticized due to the low sensitivity in detecting early disease. In particular, two of the seven criteria (presence of nodules and erosions) are generally not present in early stages of the disease. Therefore, there was a need for new criteria that could take into account the disease pathogenesis that can be used for early diagnosis and treatment decisions. In 2010, the ACR in collaboration with the European League Against Rheumatism (EULAR) produced new criteria.[41] The new criteria were applied to patients with at least 1 joint with definite clinical synovitis (swelling) and with the synovitis not better explained by another disease (such as systemic lupus erythematosus, psoriatic arthritis, and gout). The 2010 criteria was based on a score of four categories and 6 out of 10 is needed for classification of a patient as having definite RA:

A. Joint involvement
    - Score 0 if 1 large joint
    - Score 1 if 2-10 large joints
    - Score 2 if 1-3 small joints (with or without involvement of large joints)
    - Score 3 if 4-10 small joints (with or without involvement of large joints)
    - Score 5 if >10 joints (at least 1 small joint)
B. Serology (at least 1 test result is needed for classification)
    - Score 0 if negative RF and negative ACPA
    - Score 2 if low-positive RF or low-positive ACPA
    - Score 3 if high-positive RF or high-positive ACPA
C. Acute-phase reactants (at least 1 test result is needed for classification)
    - Score 0 if normal C-reactive protein (CRP) and normal erythrocyte sedimentation rate (ESR)
    - Score 1 if abnormal CRP or abnormal ESR
D. Duration of symptoms
    - Score 0 if <6 weeks
    - Score 1 if ≥6 weeks

Similar to the 1987 criteria, the 2010 criteria utilize the presence or absence of RF, a high-affinity autoantibody directed against the fragment crystallizable (Fc) portion of immunoglobulin, as one of the domain. In addition, the 2010 criteria utilize the presence or absence of ACPA.

2.2 ENVIRONMENTAL RISK FACTORS

2.2.1 Cigarette smoking

One established environmental risk factor for RA is cigarette smoking.[1] A large number of case-control studies[42-56] and fewer cohort studies[57-62] have shown that cigarette smoking is directly associated with the risk of developing RA. Previous analyses examined primarily smoking status (current, former and never smokers)[42 45 48 50 51 55-61] and lifelong exposure to smoking analyzed as pack-years of smoking.[43 44 46 47 52-54 59 60]

Less attention has been given to other important aspects of cigarette smoking, such as the number of years a person has smoked (duration),[47 59 62] the mean number of cigarettes smoked (intensity)[47 49 57 59 61 62] and smoking cessation.[47 59 60] Only one prospective cohort study simultaneously addressed all aspects of duration, intensity and lifetime smoking, as well as smoking cessation.[59] Results concerning duration and intensity of smoking have shown that the risk of RA increases in a dose-response manner. However, it is not clear if light smoking is also associated with an increase in RA risk.

A meta-analysis published in 2010 showed that in men the risk of RA was doubled among current smokers compared to never smokers, while it was 30% higher among women who were current smokers.[63] The meta-analysis also showed that cigarette smoking increases the risk of RA significantly especially among heavy smokers (more than 20 pack-years of cigarette smoking) and the risk of RF-positive RA.

Experimental data suggest that smoking is involved in RA development through a triggering mechanism, thus implying that a small amount of cigarette smoking theoretically may be enough to induce RA.[64 65] This causal model has been presented for ACPA positive RA cases. In detail, when the lung encounters smoke, macrophages are activated and some cells go into apoptosis, necrosis, or both. This process could lead to increased citrullination in certain proteins in the lungs, a process that changes the aminoacid arginine to citrulline. Therefore proteins result with a different charge, leading to a different folding and an additional sensitivity to degradation. Some of these modified proteins bind specifically to MHC class II molecules on antigen-presenting cells, such as dendritic cells or macrophages that contain the shared epitope peptide-binding motif. Moreover, smoking might further contribute to T-cell and B-cell activation by triggering antigen presenting cells in the lung, thus enhancing cell-cell interactions, like T-cell receptor with MHC class II or CD40 ligand with CD40, that finally results in a high quantity of ACPA antibodies.

2.2.2 Alcohol

Among lifestyle factors, alcohol has been the most studied in association with RA. In fact, long-term consumption of alcohol in moderate amounts may affect immune function and could down regulate production of pro-inflammatory molecules involved in the development of RA.[66-68]

The first epidemiological study to analyze this association was a hospital based case-control study among women in the Netherlands,[49] that observed a reduced risk of RA associated with alcohol consumption. However, subsequent studies (2 case-control and 2 prospective cohort studies) did not observe any association between alcohol and RA.[54 69-71] In 2008 the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) group examined the association of alcohol consumption with RA in both their case-control study and in the Danish Case-Control Study on Rheumatoid Arthritis (CACORA), observing an inverse association with RA.[72] After the EIRA study, two other case-control studies reported an inverse association as well.[73 74]

The accumulated evidence on the association between alcohol consumption and RA risk has been quantitatively summarized in two recent meta-analyses and clearly indicates a protective role of moderate consumption of alcohol (<15 grams per day) in the development of RA.[75 76] Results from paper III of this thesis were included in these meta-analyses.

2.2.3 Physical activity

Physical activity has an important role in the prevention, management, and rehabilitation of a variety of diseases. Research has gradually provided data regarding the amount of physical activity, and particularly the energy expenditure caused by it, that is necessary to prevent the development of various diseases,[77 78] including cardiovascular disease (CVD).[79] Only one prospective cohort study has analyzed the potential role of physical activity in preventing RA, but found no association between exercise and risk of RA.[70]

In patients with RA, physical activity appears to be beneficial for maintaining joint flexibility.[80 81] Moreover, physical activity improves aerobic capacity, muscle function, bone density, daily activity performance and quality of life, exactly as in healthy persons.[82-84] Some studies have also shown that moderate-intensity exercise is not associated with progression of joint destruction.[82 85] Physical activity and exercise are only used as secondary prevention in patients with RA.

2.2.4 Diet

Diet has been evaluated in several studies for its role in the management of established RA,[86] while fewer studies have examined diet in relation to the development of RA.[87]

2.2.4.1 Fish

Fish consumption is considered protective against several chronic diseases, including cancer[88-90] and cardiovascular diseases.[91 92] Few observational studies (4 case-control and 3 prospective cohort studies) have analyzed the association between fish consumption and RA, and results are mixed.[69 93-98] Two hospital-based case-control studies conducted in Greece have reported a lack of association between fish consumption and RA.[95 96] In contrast, a population-based case-control study observed a statistically significant reduced risk of RA among women who consumed 2 or more servings of broiled or baked fish per week.[98] The EIRA study, a large population-based case-control study conducted in Sweden,[97] and the Diet, Cancer and Health (DCH)[69] cohort study found a modest decrease in risk of total RA with consumption of oily fish, while the Nurses' Health Study did not show an association between total fish and RA.[93]

2.2.4.2 Meat

Meat consumption is an important dietary source of protein and essential nutrients including iron, zinc and vitamin B12. However, there is accumulating evidence that red meat consumption increases the risk of cardiovascular diseases[99 100] and colon cancer.[101 102] Epidemiological studies (1 case-control and 2 prospective cohort studies) have reported a lack of association with RA.[69 93 98]

2.2.4.3 Dairy products

Dairy products is a broad term used to indicate milk and products derived from milk, including yogurt, cheese, cream, and butter. Only two studies have examined dairy product consumption in relation to RA risk. A case-control study conducted in Washington reported no association between dairy products and milk beverages and risk of RA.[98] In contrast, the Iowa Women's Health Study, a prospective cohort study, reported an inverse association between total dairy products and risk of RA.[103]

2.2.4.4 Fruits and vegetables

Fruits and vegetables play an important role in diet due to their protective action against several chronic diseases.[104 105] Fruits and vegetables could play a role in reducing the risk of RA, especially thanks to their high content of antioxidant nutrients.

Studies on the association between fruits and vegetables and RA are limited, and results are not consistent. A case-control study and the Diet, Cancer, and Health cohort study found no association between fruit and vegetable consumption and RA risk.[69 98] A more recent case-control study in Greece found an inverse association between cooked vegetables and RA, but no association with raw vegetables.[95] An inverse association, although not statistically significant, was observed in the Iowa Women's Health Studies between fruit and vegetable consumption and RA.[106] Among fruits, oranges and

grapefruit juice consumption showed the lowest relative risks, while among vegetables, cruciferous vegetable consumption was associated with the lowest risk.

2.2.4.5 Coffee and tea

Coffee and tea are two of the most consumed beverages in the world.

An increase in RA risk associated with high coffee intake was observed in a prospective cohort study in Finland[71] and in a matched case-control study conducted in Denmark.[46] However, other cohort studies failed to replicate these results. The Black Women's Health Study (BWHS),[107] the Iowa Women's Health Study,[108] the Nurses' Health Study,[109] and the Danish Diet, Cancer and Health cohort[69] reported no association with RA for total or caffeinated coffee intake.

Studies that have also examined decaffeinated coffee intake, have found an increase in RA risk associated with this beverage.[107 108] The reason could lie in the use of solvents in the decaffeination process of coffee beans that may play a role in the development of RA.[110]

It is hypothesized that tea has both anti-oxidative and anti-inflammatory properties,[111] but results from observational studies on RA are mixed. The Iowa Women's Health Study observed a decreased risk of RA with high consumption of tea,[108] while the Nurses' Health Study found no association.[109] In contrast, the Black Women's Health Study found a positive association between tea consumption and RA.[107]

2.2.4.6 Nutrients

Foods provide the human body with essential nutrients, that are utilized to survive and grow. Few studies have examined the role of different nutrients in the prevention of RA.

Long-chain n-3 polyunsaturated fatty acids (PUFAs), commonly known as long-chain omega-3, are mainly found in fish. Indeed, the observed inverse association between fish consumption and RA risk has been attributed to their content of long-chain n-3 PUFAs. The n-3 PUFA eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) are metabolized to competitive inhibitors of n-6 PUFAs (prostaglandins and leukotrienes) and suppress the production of the inflammatory cytokines, such as TNFα and interleukin 1β,[112] involved in RA development. However, only two studies have directly examined these nutrients and found no association.[69 98]

Dietary vitamin D is also found mainly in fish. Some studies have indicated that vitamin D may reduce the development of autoimmune diseases.[113 114] The prospective Iowa Women's Health Study reported an inverse association between both dietary and supplemental vitamin D intake and RA.[103] However, other studies were not able to confirm these findings.[69 115 116]

Adequate levels of selenium are important for immunity, and selenium is also involved in regulating excessive immune responses and chronic inflammation.[117] Two nested case-control studies conducted in Finland analyzed the serum concentrations of selenium in men and women with and without RA.[118 119] An elevated risk of RA was observed for low levels of selenium, but the association was not statistically significant.

Fruits and vegetables are rich in antioxidants that may protect against oxidative stress. Products of free radical oxidation are present in the synovial fluid of patients with RA, indicating a role of free radicals and oxidative stress in the RA inflammation process.[120 121] Among the four studies examining associations between antioxidants and RA,[69 98 106 122] only one prospective study observed an inverse association.[106] The association between antioxidants and RA was also examined using serum antioxidant concentrations in three studies. Two nested case-control studies conducted in Finland [118 119] observed an elevated risk of RA for low levels of serum α-tocopherol, and β-carotene, but none of the associations were statistically significant. A case-control study in Washington County, Maryland, analyzed the difference in serum concentration of α-tocopherol and β-carotene between RA cases and controls, finding a statistically significant decrease only for β-carotene.[123] A randomized, double-blind, placebo-controlled trial conducted in US, The Women's Health Study, also evaluated vitamin E supplementation and found no association with RA.[124]

2.2.5 Other factors

Occupational silica dust has been shown to be associated with RA in the EIRA study.[125] Other epidemiological studies have linked the association of occupations such as drilling, mining, and sand blasting with increased RA risk due to exposure to silica through the respiratory tract.[125-128] In addition, substantial exposure to inhaled organic solvents in occupations such as upholstering, hair-dressing, and concrete work has been associated with risk of RA.[129]

Hormones are related to both incidence and clinical expression of RA. Women are two to four times more likely than men to develop RA.[130 131] Use of oral contraceptives has been inversely associated with RA in most studies, but not all.[132] The role of breast feeding is not clear: the Nurses' Health Study showed an inverse association with duration of breast feeding,[133] and two other studies confirmed the inverse association,[134 135] however one study showed a positive association.[136]

3 AIMS

The overall aim of this thesis was to evaluate the association between lifestyle factors and diet and risk of rheumatoid arthritis.

The specific aims were:

- To evaluate the association between cigarette smoking and risk of developing rheumatoid arthritis in a large population of Swedish women, with attention to specific characteristics of smoking including duration, intensity and cessation (**Paper I**), and by summarizing the published evidence in a dose-response meta-analysis (**Paper II**).

- To prospectively estimate the association between alcohol consumption and rheumatoid arthritis in the Swedish Mammography Cohort (**Paper III**).

- To analyze the long-term consumption of alcohol in association with risk of rheumatoid arthritis (**Paper III**).

- To evaluate the association between dietary intake of long-chain n-3 polyunsaturated fatty acids and risk of rheumatoid arthritis in the Swedish Mammography Cohort (**Paper IV**).

- To analyze the long-term consumption of long-chain n-3 polyunsaturated fatty acids (**Paper IV**), as well as the long-term consumption of fish (**Paper IV**) in association with rheumatoid arthritis.

- To analyze the association between physical activity and risk of rheumatoid arthritis in the Swedish Mammography Cohort (**Paper V**).

4 SUBJECTS AND METHODS

4.1 STUDY POPULATION

This thesis is based on data from the Swedish Mammography Cohort (SMC), a cohort study of Swedish women established in 1987.[137]

Between March 1987 and December 1990, all women living in Uppsala (n=48 517) and Västmanland County (n=41 786) and born between 1914 and 1948 were invited to participate in a population-based mammography screening program. The invitation included a six-page questionnaire with question regarding alcohol intake and diet (67-item food frequency questionnaire (FFQ)), parity, weight, height and educational status. The response rate was 74% (n= 61 433) (**Figure 4**).

In the fall of 1997 a second questionnaire was sent to women who were still alive (n=56 030). The 1997 questionnaire collected information regarding alcohol and diet (96-item FFQ), and additional information regarding dietary supplements, physical activity, cigarette smoking and anthropometric measures including body weight across the life course. The response rate was 70% and after exclusion of women with incorrect or missing personal identification number the final 1997 cohort consisted of 38 984. Women who answered the 1997 questionnaire were the study base of **Paper I, III, IV** and **V**.

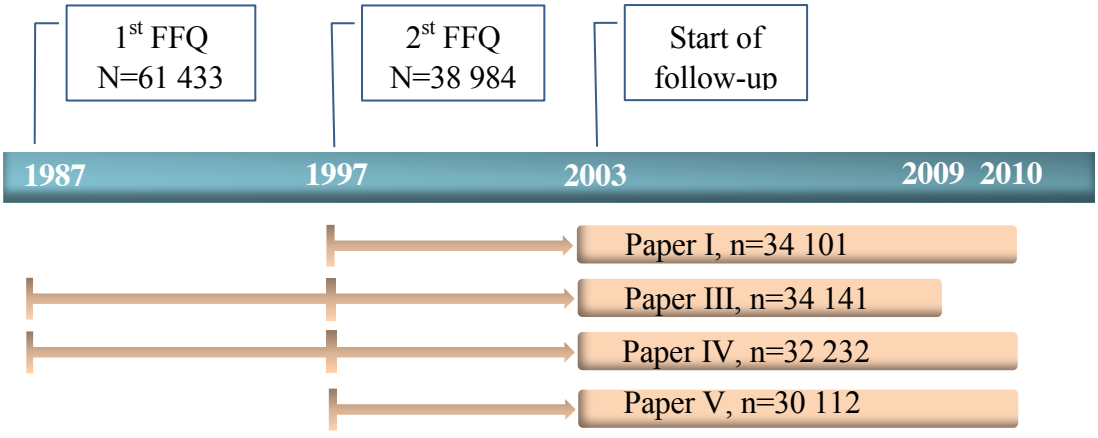

Figure 4. The Swedish Mammography Cohort study populations for **Paper I, III, IV** and **V**. The length of each bar represent the follow-up time for each study.

4.1.1 Exclusions

From the 37 239 women still alive at start of follow-up (1st January 2003), all women diagnosed with a non-RA joint condition (ICD-10 codes M07-M12, M14, M45, M46, M30-M36) were excluded (1899 women in **Paper III** and 2052 women in **Paper I, IV** and **V**). The diagnoses considered for exclusion were psoriatic and enteropathic arthropathies, juvenile arthritis, gout, other arthropathies, and systemic connective tissue disorders, such as systemic lupus erythematosus. The reason for this exclusion is because of the difficulty in the diagnosis of RA, a disease that can be misdiagnosed as other joint conditions in early stages. Moreover, in **Paper I, III** and **V**, women who did not answer the questions regarding cigarette smoking status (n=797), drinking status (n=440) and leisure-time activity (n=4705) were excluded from the analysis. Women with extreme energy intake (i.e. 3 standard errors from the mean value on the log-transformed scale) were also excluded (n=502 in **Paper III** and n=496 in **Paper IV**). Moreover, women who consumed fish oil supplements (n=2167) were excluded from the main analysis in **Paper IV**, but included in sensitivity analysis.

4.2 EXPOSURES ASSESSMENT

Smoking
In the 1997 questionnaire a section regarding cigarette smoking, that was not included in the 1987 questionnaire, was introduced. The questions regarding smoking included current smoking status at the time of filling in the questionnaire, the year they started smoking, and the year of smoking cessation. Moreover, a question regarding the number of cigarettes smoked per day was included, with the option to report current intensity of smoking as well as intensity at different ages.

In **Paper I** different variables were considered in order to cover all aspects of cigarette smoking. Smoking status was categorized as never, former and current smoker. Intensity of cigarette smoking was calculated as the lifetime average number of cigarettes per day, while duration was calculated as the number of years a woman had smoked during her life, for both former and current smokers. Intensity and duration of smoking were categorized into tertiles and relative risks were calculated using never smokers as the reference group. Smoking cessation was evaluated using the number of years since quitting smoking and the age at smoking cessation. Moreover, a variable that took into account the lifetime exposure to cigarettes smoking (pack-years) was calculated. Pack-years were calculated by multiplying the average number of cigarettes smoked per day by the number of years the person had smoked, divided by 20 (number of cigarettes in one pack). Pack-years of smoking was categorized into quartiles.

Alcohol
The average number of glasses of alcoholic beverages per week, defined as 15 grams of ethanol, corresponding to approximately 500mL of beer, 150 mL of wine, or 50 mL of liquor was calculated using five questions from the 1987 FFQ (**Paper III)**. Women were asked to indicate their average consumption over the previous six months of beer (0.5%, 2.8% and 4.5% alcohol by volume), wine (12.5% to 14.5% alcohol), and liquor (40% alcohol).

The section of the 1997 FFQ with information on alcohol consumption included questions on status of alcohol drinking (from which the variable alcohol status was derived), year since quitting drinking, frequency of beer (2.8% and 4.5% alcohol), wine (less and more than 18% alcohol (fortified wine)), and liquor (40% alcohol) drinking. Eight predefined response categories were provided ranging from "never" to "three or more times per day". Moreover, a question regarding the amount of drinking per occasion was included (**Figure 5**). Also, an open ended question collected information regarding the daily or weekly consumption of light beer (0.5% alcohol).

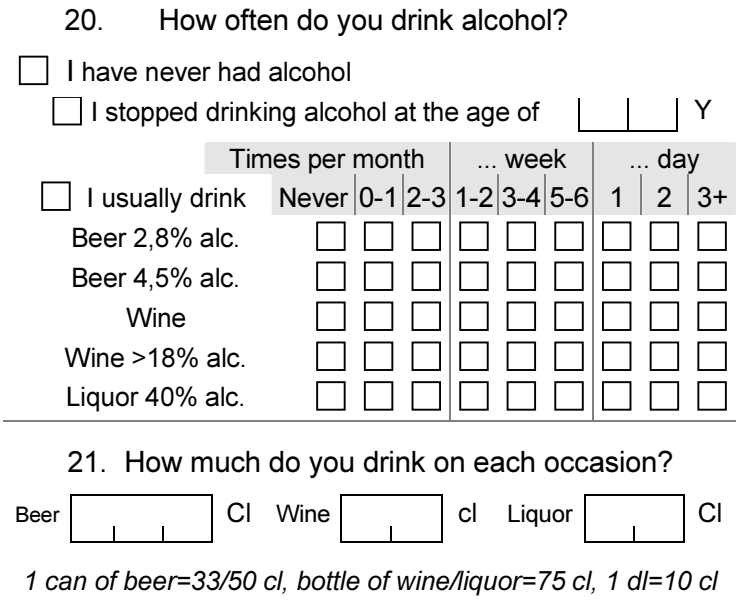

Figure 5. Questions used in the 1997 food-frequency questionnaire to assess alcohol consumption in the Swedish Mammography Cohort

In **Paper III,** to evaluate the association between alcohol consumption and risk of RA, women were categorized as never drinkers, former drinkers, occasional drinkers (≤2 glasses of alcohol per week), and regular drinkers (>2 glasses of alcohol per week) according to 1997 consumption. Moreover, the number of glasses per week was analyzed as a four level variable, from less than 1 or never to more than 4 glasses per week. Relative risks for consumption of beer, wine and liquor were reported separately and mutually adjusted. Moreover, an analysis regarding the long-term consumption of alcohol was performed by combining information about number of glasses of alcohol per week from the 1987 FFQ and the 1997 FFQ.

Diet

In 1987 the FFQ included questions regarding 67 food items. The eight predefined responses ranged from "never or seldom" to "four or more times per day". Total fish consumption in 1987 was calculated using two questions, one regarding the consumption of fatty fish (salmon, mackerel, herring) and the second on consumption of other types of fish (**Paper IV**).

The 1997 FFQ was more comprehensive and included 96 food items. In this questionnaire participants reported their average frequency of consumption of each food item during the previous year. Eight predefined response categories were provided ranging from "never" to "three or more times per day". There were also additional open questions regarding dairy foods, coffee, tea, light beer, soft drinks, sugar/honey, and bread. The 1997 questionnaire, moreover, collected information on the use of dietary supplements, such as multivitamins and fish oil supplements. Total fish consumption in 1997 was calculated using three questions based on the consumption of herring/mackerel, salmon/whitefish, and cod/saithe/fish fingers.

Intake of individual nutrients was calculated by multiplying the average frequency of consumption of each food by the nutrient content of age-specific portion sizes. Values for each nutrient amount in foods was obtained from the Swedish National Food Administration Database.[138] The Swedish National Food Agency database is based on analyses of representative foods on the Swedish market. For prepared foods and dishes cooking losses are taken into consideration. The database is considered to be virtually complete with regard to the Swedish food supply (2071 foods and dishes are included – of which 196 are fish/seafood items and 407 are meat/poultry items). Nutrient intakes were adjusted for total energy intake through the use of the residual method.[139]

In **Paper IV** long-chain n-3 polyunsaturated fatty acids (PUFAs) intake was calculated by summing the intake of eicosapentaenoic acid (EPA C20:5), docosapentaenoic acid (DPA C22:5), and docosahexaenoic acid (DHA C22:6). Dietary intake of long-chain n-3 PUFAs in the study population derived mainly from fish consumption: 74.16% of total EPA dietary intake, 42.46% of DPA, and 67.08% of DHA was contained in fish. To calculate the long-chain n-3 PUFA intake, the portion-size of salmon was considered, according to mean values from a total of 5922 days of weighted food records kept by 213 randomly selected women of the SMC, as 72 g among women aged ≤63 years, 66g among women aged 64-71 and 63g among women aged ≥72, while the portion size of codfish was 94 g among women aged ≤63 years, 116g among women aged 64-71 and 116g among women aged ≥72. This variable was categorized in two different ways: in a multi-category model it was categorized in quintiles, while in a threshold model two categories were used, ≤0.21 grams per day and more than 0.21 grams per day (0.21g/day was the first quintile of the distribution on long-chain n-3 PUFAs). A long-term analysis was performed combining information regarding long-chain n-3 PUFAs and total fish consumption from the 1987 and 1997 FFQ. In this analysis, long-term n-3 PUFAs was categorized as in the threshold model, while fish consumption was categorized as less than one serving per week and one or more servings per week.

Physical activity

Information regarding physical activity was collected on the 1997 questionnaire. There were five questions regarding different daily activities and women reported their current level of activity as well as their physical activity when they were 15, 30 and 50 years old. The first question inquired about the number of hours per day dedicated to home/household work. The six predefined answers ranged from less than one hour per day to more than eight. The second question asked about physical activity in regards to type of work/occupation, and the six predefined answers ranged from work requiring

mostly sitting down to heavy manual labor. Two questions covered leisure time activities such as walking/cycling (six answers ranging between hardly ever to more than 1.5 hours per day) and exercise (five answers from less than one hour per week to more than five hours per week). Another question covered leisure-time inactivity such as watching TV/reading (five answers from less than one hour per day to more than six hours per day). A final open-ended question was about the number of sleeping hours per day.

In **Paper V** the variables regarding physical activity were analyzed separately. Moreover, a combined variable for leisure-time activity was constructed as the combination of minutes per day of walking/cycling and weekly hours of exercise. The combined variable had four levels according to the combination of less or more than 20 minutes per day (min/day) of walking/cycling and less or more than 1 hour per week of exercise. A second combined variable was calculated as total energy expenditure. The 24-hour energy expenditure score was calculated by adding the products of duration and intensity, expressed as metabolic equivalent (MET, kcal/kg per hour), for each type of physical activity and inactivity, including sleep.

4.2.1 Validation

The reproducibility and validity of the FFQs have been assessed for foods, nutrients, dietary supplements by comparison with diet records and biological markers.[140-142] The validity of the FFQ was also assessed among men using multiple 24-hour recall interviews.[143]

A validation of the 1987 FFQ was performed by comparing the four one-week weighted dietary records (three to four months apart) in a random sample of 184 women from the SMC cohort.[140] The validity of alcohol intake, as measured by correlation coefficient, was 0.9 (**Paper III**). In the same subgroup, the correlation between the FFQ-based estimation of EPA and DHA (**Paper IV**) and their relative content in adipose tissue was 0.46 and 0.45 respectively.[141]

The validity of the 1997 FFQ has been evaluated in 248 middle-aged and elderly men (40-74 year old) from the Cohort of Swedish Men (COSM) who received the same questionnaire in 1997 that was sent to the women in the SMC. The estimates of alcohol intake (**Paper III**) based on the 1997 FFQ had good validity compared with 14 interviews that measured 24-hour recall of intake (correlation coefficient of 0.81).[143]
The validity of EPA and DHA (**Paper IV**) estimates were 0.64 and 0.60 respectively.[143] The validity of the estimate of long-chain n-3 PUFA intake from the 1997 FFQ was further examined in a sub-cohort of the SMC. Adipose tissue was obtained in 2003-2004 from 239 randomly selected women, aged 55-75, and the validity was 0.41.[142] The validity of leisure-time activity and inactivity estimates (**Paper V**) was assessed comparing the questionnaire with 7-day activity records.[144] The correlations were 0.42 and 0.52 respectively.

4.2.2 Missing data

The use of self-administered questionnaires is a way to reduce costs related to interviewer-related instruments. However, they are also less easily monitored and more susceptible to misunderstanding of questions and non-response, leading to missing data.

In this thesis missing data regarding the exposure was handled in two different ways. In **Paper I, III** and **V**, women who did not answer questions regarding cigarette smoking status, drinking status and leisure-time activity were excluded from the analysis. We therefore focused only on the women who reported information on the exposure of interest.

In **Paper III** and **IV**, a second approach was also considered, treating non-response as "zero consumption". This approach is based on the work of Hansson et al,[145] who analyzed missing data from a 56-item FFQ sent to participants of a case-control study in Sweden. 58% of cases and 50% of controls were contacted for a telephone interview, during which they were asked to provide all the omitted information. The proportion of actual non-consumption among missing reports was 74.1% for alcoholic beverages and 82% for total fish. This approach was used in **Paper III** when analyzing frequency of beer, wine and liquor consumption, and in **Paper IV** to calculate total fish consumption and intake of long-chain n-3 PUFAs. In particular, if a more conservative approach were used, the long-chain n-3 PUFAs nutrient would have been treated as missing whenever the information concerning one of its component or source items was missing, leading to an unacceptable loss of information.

For example, focusing only on 65 predefined questions regarding average consumption of foods in the 1997 FFQ, only 6.4% of the women in the SMC cohort who were still alive at the start of follow-up (1 January 2003) completed this section without reporting any missing value. In this section the median number of missing values was 4, 25% of the women did not answer 12 questions out of 65, and only 164 women did not report any answer (**Figure 6**).

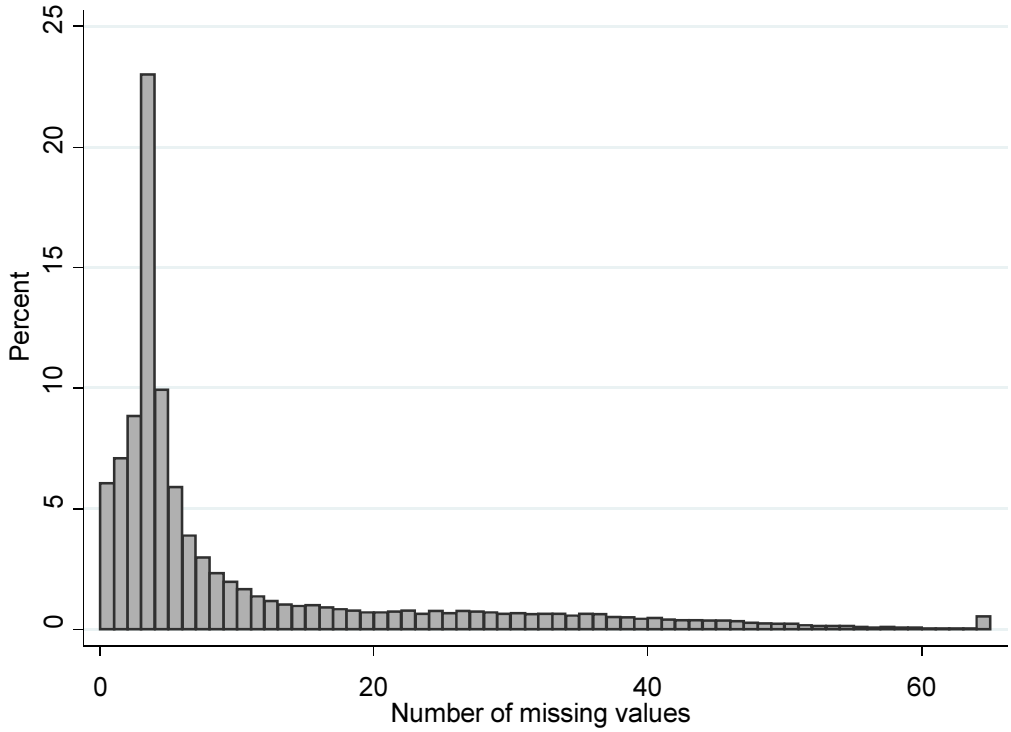

Figure 6. Number of missing values reported in one section of the food-frequency questionnaire, consisting of 65 questions with 8 predefined answers.

4.3 CASE DEFINITION

4.3.1 The Patient Register

The Swedish National Inpatient Register (IPR) of the National Board of Health and Welfare (NBHW) is a collection of information regarding inpatients at public hospitals in Sweden. The IPR started in the 1960's, but only since 1987 includes all inpatient care in Sweden.

From 2001, the IPR was joined by the National Outpatient Register (OPR) that contains data on outpatient specialist visits, including day-surgery and psychiatric visits from both public and private caregivers. Primary care from general practitioners is not yet covered in the Swedish registers.

Coverage of the OPR was assessed at approximately 80% in 2007. In 2011 an external review and validation of the IPR was conducted.[146] The study reported that 99% of all somatic and psychiatric hospital discharges were registered in the IPR and a primary diagnosis was listed for 99% of all discharges. The review of the validity found positive predictive values (PPVs) of 85-95% for most diagnoses given in the IPR, with a PPV for RA of 95.9%. In a study based on the hospital discharge register, almost one thousands medical records were validated, and a validity (ICD-codes for RA vs. the American College of Rheumatology (ACR) criteria) of approximately 90% was found.[147] Similarly, in ongoing and past studies based on the OPR, approximately 200

medical records were scrutinized. A high validity of the register-based RA diagnosis, close to 90%, was found.[148]

4.3.2 The Swedish Rheumatology Register

The Swedish Rheumatology Register (SRR) was initiated in 1995 and collects information from rheumatologists in Sweden on newly diagnosed patients with RA and patients on biologic treatments. In recent years, patients with established RA, patients with long symptom duration at diagnosis and other rheumatologic diseases, regardless of treatment, have been included. The SRR has been estimated to include approximately 50% of all newly diagnosed patients with RA in Sweden.[149]

4.3.3 Identification of incident cases

Linkage to the registers was made using the Swedish Personal Identity Number (PIN).[150] The Swedish personal identity number (date of birth combined with a unique four digit number, e.g. 19470102-0259) serves as a unique identifier in Swedish health care, and in many other areas of the Swedish society (e.g. taxes). This linkage method allows for a virtually 100% coverage of the Swedish health care system. Although typing mistakes are possible when recording patients information, data from the SMC were checked for consistency in the PIN number with the information from the register (date of birth and gender of the person and an inbuilt control internal number to check for correctness) and the error rate was less than 1 per 5000 persons.

RAis is a disease for which the onset and the diagnosis could differ by several months. A study has estimated that the median lag time between onset of RA symptoms and the first rheumatologist's encounter is 17 months.[151] Therefore, early identification of incident RA cases is difficult. Incident cases of RA were identified as the first diagnosis code for RA (ICD-10 codes M05 and M06).

The main source used to identify newly diagnosed RA cases within the SMC was the OPR, since the SRR had a coverage of only 50% and RA is not a disease that usually leads to hospitalization in its first stages. For this reason, the start of follow-up had to be delayed in comparison to the data collection in 1997. The start of follow-up, however, could not correspond to the start of the OPR, in 2001. An additional two-years delay in the start of follow-up was in fact necessary to remove all prevalent cases that were registered for the first time when the OPR started. **Figure 7** shows the pick of first diagnosis identified in 2002 in the OPR for this reason. Longer delays were considered (start of follow-up in 2004 and 2006) in sensitivity analyses.

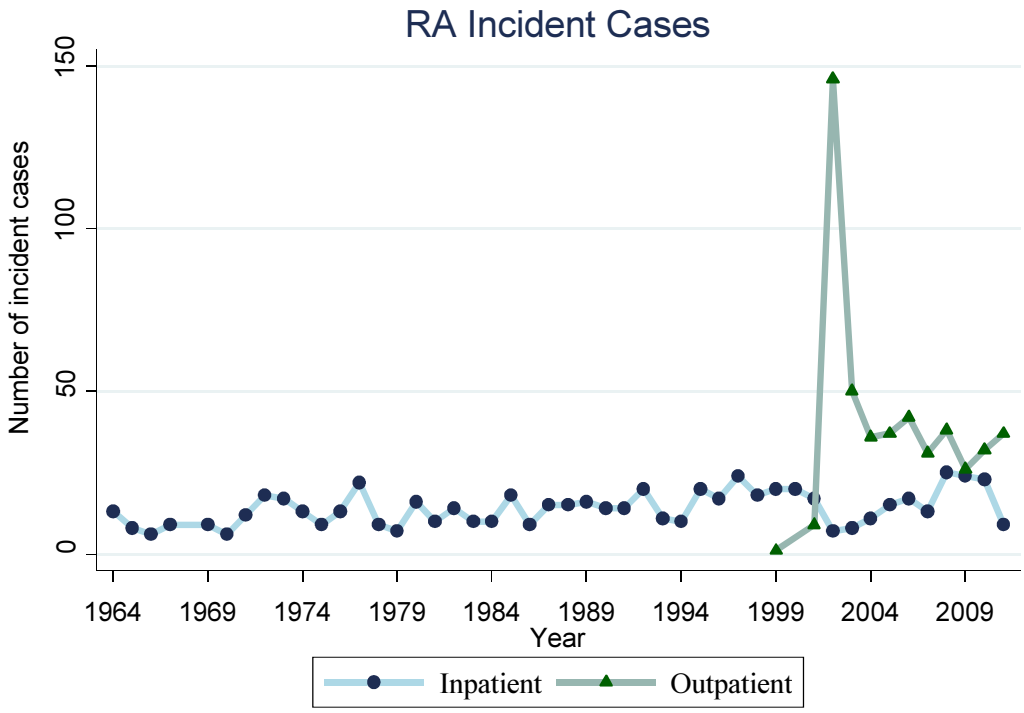

Figure 7. Number of incident cases identified in the Inpatient and Outpatient Registers.

For cases with a first diagnosis in both OPR and SRR, information was merged and the earlier date of diagnosis was used. Moreover, using information from the SRR, cases were excluded if date of first symptoms and date of diagnosis were more than 365 days apart.

The IPR, that includes data on hospitalization, was used only to identify prevalent cases of RA, therefore for exclusion of prevalent RA cases from the cohort at the start of follow-up. Newly diagnosed RA cases identified through the IPR during the follow-up period were considered as prevalent cases and therefore excluded from the cohort. However, the inclusion of these cases was considered in sensitivity analyses.

Deaths occurring in the cohort were identified through the Swedish Death Register.

4.4 STATISTICAL METHODS

4.4.1 Survival analysis

Survival analysis is the analysis of the time to the occurrence of an event, which in this thesis was the diagnosis of RA. Rather than analyzing the probability density function of the time to event T, f(t), or its cumulative distribution F(t), survival analysis usually considers the survival function, S(t), and the hazard function, h(t).

The survival function is the complement of the cumulative distribution function:

$$S(t) = 1 - F(t) = \Pr(T > t)$$

The survival function is defined as the probability of surviving beyond a given time t, i.e. the probability that there is no event before time t. The function is equal to 1 at time 0 and decreases towards 0 as t goes to infinity.

The hazard function h(t) is:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

The hazard function can be interpreted as the instantaneous event rate at time t, conditional on survival up to time t. The hazard is not a probability and can vary between 0 and infinity. Over time the hazard can increase, decrease, fluctuate, or remain constant.

Sometimes it is not possible to determine the time of the failure event, at least not for every subject. For example, during the study period, subjects are followed up and data are collected. If during the follow-up period a subject does not experience the event or is lost to follow-up, then that observation is called "censored".

4.4.1.1 Cox Model

To analyze censored survival data the Cox proportional hazards regression model was used in this thesis.[152] The model define the hazard for the j-th subject as

$$h(t|\boldsymbol{x_j}) = h_0(t)\exp(\boldsymbol{x_j}\boldsymbol{\beta}_x)$$

where $h_0(t)$ denotes an unspecified baseline hazard, **x** is the vector of the covariates, and **β** is the vector of the parameters of the model. Results from a Cox model are given in terms of hazard ratio (HR, also called relative risk in this thesis), comparing one subject to another:

$$HR = \frac{h(t|\boldsymbol{x_j})}{h(t|\boldsymbol{x_m})} = \frac{\exp(\boldsymbol{x_j}\boldsymbol{\beta}_x)}{\exp(\boldsymbol{x_m}\boldsymbol{\beta}_x)}$$

The Cox model is a linear model for the log of the hazard ratio, a fact that simplifies the estimation and the interpretation of the **β** parameters.

The Cox model is based on the assumption that the hazards are proportional. Therefore, the ratio of the two hazards does not depend on time, t. To test this important assumption of the model the Schoenfeld's residuals method was used. This method tests the hypothesis of a zero slope for a model of the Schoenfeld's residuals (difference between the observed value and the estimated value from the model) as a function of time.[153] Other ways to test the proportionality assumption are graphical methods or the inclusion of an interaction term between a covariate and the survival time.

4.4.1.2 Age as time-scale

Typically in cohort studies, the time-scale used in Cox regression model is time-on-study, i.e. the follow-up time or time since baseline. It is however possible to use age as time-scale, where subjects enter the analysis at their baseline age and exit at their event or censoring age. The use of age as time-scale is an efficient way to adjust the model for it. By using age as the time-scale, it is not necessary to assume a parametric model for the relationship between age and the outcome. For example, whether age is included in the model as a continuous variable or as a categorical variable could affect the results. Instead, the use of age as time-scale is a non-parametric way to adjust for it.

4.4.1.3 Restricted cubic splines

To model the dose-response relationship between the exposures and RA, restricted cubic splines were used in **Paper I** and **IV**. Restricted cubic splines consist of three or more polynomial segments with boundaries called knots. Between consecutive knots the curve is a cubic polynomial, and a straight line before the first knot and after the last knot.

A restricted cubic spline Cox proportional hazards model with q knots can be written as

$$h(t|\boldsymbol{x}) = h_0(t)\exp(\boldsymbol{\beta_1 f_1}(\boldsymbol{x}) + \boldsymbol{\beta_2 f_2}(\boldsymbol{x}) + \cdots + \boldsymbol{\beta_{q-1} f_{q-1}}(\boldsymbol{x}))$$

where the spline covariates f_1, f_2, f_{q-1} are transformation of **x** that depend on the original variables **x**, the knots, and the distances between knots.

4.4.1.4 Probabilistic sensitivity analysis

In the appendix of **Paper I, III, IV** and **V** a probabilistic sensitivity analysis was performed to evaluate possible changes in the estimates due to bias in the definition of incident cases of RA. The possible misclassification of prevalent cases as incident cases has been considered and the uncertainty about the amount of misclassification has been modeled in a priori distribution, in the form of a uniform distribution with values that range from 0 to 20 (a maximum of 20% of misclassified cases was considered appropriate). Simulations were performed with 200 random draws from the a priori distribution: each draw corresponded to the percentage of prevalent cases that should be randomly excluded. For each draw the corresponding relative risk was calculated, and from the distribution of relative risks obtained the median relative risk and the standard deviation were calculated.

The probabilistic sensitivity analysis was performed in each study considering different pattern of exclusions. For example, in **Paper IV**, we considered three patterns of exclusions:

- Cases excluded randomly from all categories, as if the prevalent cases did not change their eating habits due to the disease
- Cases excluded only from the lower category, as if the prevalent cases were all decreasing the consumption of foods rich in long-chain n-3 PUFAs

- Cases excluded only from the upper category, as if the prevalent cases were all increasing the consumption of foods rich in long-chain n-3 PUFAs.

Similar patterns were assumed in **Paper I, III** and **V** for smoking, alcohol intake and physical activity.

4.4.1.5 Statistical Software

All the survival analysis using Cox models and the probabilistic sensitivity analyses were performed in SAS, version 9.2. The dose-response analyses presented in Paper I and IV were conducted in Stata, version 11.1.

4.4.2 Population Attributable Risk

The population attributable risk (or population attributable fraction) indicates the proportion of cases that would not occur in a population if the risk factor were eliminated. It can be calculated as:

$$PAR = \frac{P_e\,(RR_e - 1)}{[1 + P_e\,(RR_e - 1)]}$$

where P_e is the prevalence of the exposure and RR_e is the relative risk of the disease due to the exposure.

The equivalent of the population attributable risk when analyzing a protective factor is the population prevented fraction, calculated as

$$PPF = P_e\,(1 - RR_e)\,.$$

4.4.3 Meta-analysis

In the last few years the number of published articles has grown exponentially (see **Figure 8**). According to PubMed, the number of medical articles published between 1960 and 1970 was around 1 million, while in the last decade the number of articles published was more than 8 million.

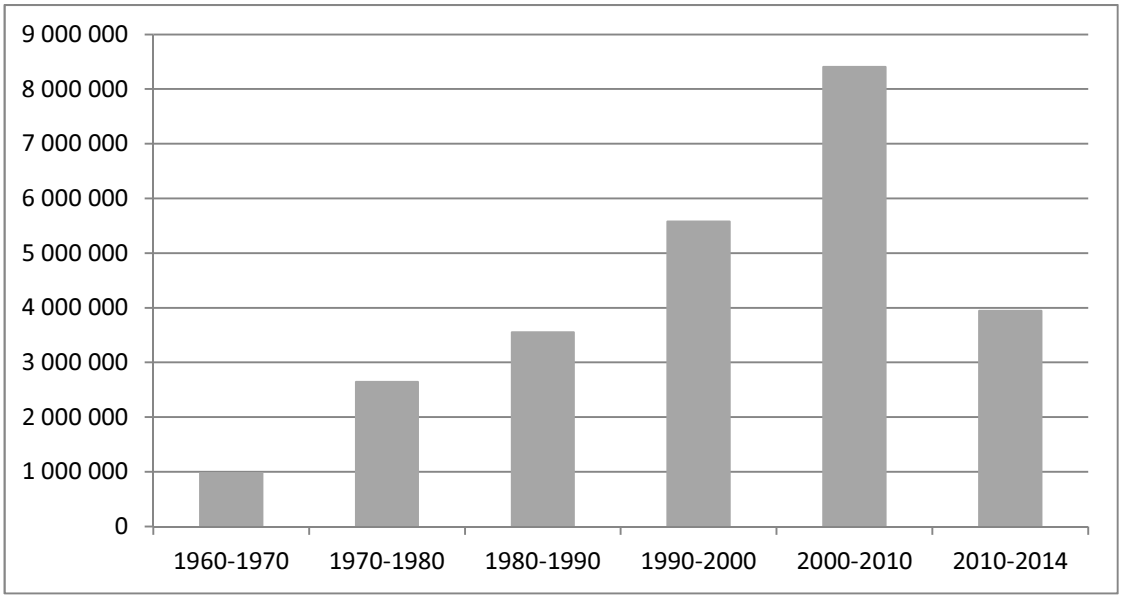

Figure 8. Number of medical articles published divided in decades, according to PubMed

The scientific information has become difficult to handle, and researchers may have problems drawing conclusions based on such a large amount of results, often based on different methodologies. Meta-analysis is a tool to combine and summarize results from studies with disparate research methods and findings.

This method was invented in 1904 by the mathematician Karl Pearson to summarize results on the effectiveness of inoculation against typhoid fever.[154] However, the term meta-analysis was first used in 1976, when Glass presented the methodology to the American Educational Research Association annual meeting.[155] In the last two decades the number of meta-analyses has grown, and extensive research regarding the underlying methods has been conducted.

The main aim of a meta-analysis is to combine the results of different studies into one single estimate. When performing meta-analysis it is possible to choose between two methods: a fixed-effect model, in which it is assumed that there is a true effect size common to all studies, and a random-effects model, in which it is allowed to the true effect size to differ between studies. In the random-effects model the effect estimates in the studies are considered a random sample from a particular distribution of the true effect size and therefore the combined estimate is the mean of this distribution.

To combine the results from different studies, it is first necessary to assign a weight to each study. In the majority of the meta-analyses the weights are assigned based on the inverse of the overall study error variance (1/variance). Therefore, a high weight is assigned to studies with a precise estimate of the population relative risk (low variance), while a low weight is given to studies with a less precise estimate of the population relative risk (high variance). Under the fixed-effect model, the weight of the study *i* is the inverse of the within-study error variance (V_i):

$$W_i = \frac{1}{V_i}$$

Under the random-effects model, there are two sources of variance that need to be taken into account: the within-study variance V_i specific for each study and the between-study variance τ^2 that is common to all the studies. Therefore, the weight is defined as:

$$W_i = \frac{1}{V_i + \tau^2}$$

Under the random-effects model the confidence interval of the combined effect will always be wider and the weights will always be more similar to each other than under the fixed-effect model due to the inclusion of the between-study variance.

The combined log-relative risk is given by the weighted mean across all studies:

$$\log(RR_{overall}) = \frac{\sum_{i=1}^{k} W_i \log(RR_i)}{\sum_{i=1}^{k} W_i}$$

where k is the number of studies included in the meta-analysis.

The variance is calculated as:

$$V_{\log(RR_{overall})} = \frac{1}{\sum_{i=1}^{k} W_i}$$

and the standard error (SE) is given by:

$$SE_{\log(RR_{overall})} = \sqrt{V_{\log(RR_{overall})}}$$

In **Paper II** this approach was used to summarize the relative risk estimates in the highest category of pack-years of smoking versus never smokers. However, the relationship between a continuous exposure and an outcome may take on different shapes (e.g. linear, U-shaped, J-shaped). Epidemiological studies therefore often group the exposure variable into more than two categories. A tool to summarize dose-

response results has been described in a paper by Greenland and Longnecker in 1992[156] and is called a dose-response meta-analysis.

The first stage of a two-stage dose-response meta-analysis is to estimate the dose-response trend within each study using a log-linear model where the dependent variable is the logarithm of the relative risk as function of the exposure dose:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

where **y** is the n x 1 vector of adjusted relative risks and **X** is the n x p matrix containing the values of the dose (excluding the reference level) and/or some transformation of it (e.g. splines, polynomials). The relative risks presented in each study are not independent, since they share the same reference exposure level. Therefore, the covariance matrix $\boldsymbol{\Sigma}$ of the error terms is comprised of the variance of each log-relative risk on the main diagonal and the non-zero covariance between log-relative risks estimates in the off-diagonal elements.

The aim of the second stage is to combine the study-specific trend estimates using established methods for multivariate meta-analysis.[157-159] A dose-response meta-analysis is presented in **Paper II**.

4.4.3.1 Heterogeneity

The studies included in a meta-analysis can differ on many levels, from the study design to the study population, from the exposure assessment to the number of confounding variables adjusted for. Such diversity is usually referred to as heterogeneity, and could cause the observed discrepancies in the results of the studies.

The classical measure of heterogeneity is the Cochran's Q statistic

$$Q = (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

which is distributed as a χ^2 distribution with m-1 degrees of freedom, where m is the number of study included in the meta-analysis. However, if the p-value obtained from this statistic is large (e.g. >0.05) we can conclude only that the test did not detect a problem with the model, not that the problem does not exist. Moreover, the Q test has low power (it is more likely that the test will be statistically non-significant), especially when the number of studies is small, i.e. most meta-analyses.[160]

Another statistic that is widely reported to quantify heterogeneity is I^2, which is the percentage of total variation attributable to heterogeneity. It is calculated as:

$$I^2 = 100\% \, \frac{(Q - df)}{Q}$$

where df is the number of degree of freedom of the Q statistic. A value of 0% indicates no observed heterogeneity, and larger values indicate increasing heterogeneity.

4.4.3.2 Publication bias

Meta-analyses are based on a systematic review of the relevant literature. In **Paper II,** MEDLINE (PubMed) and EMBASE databases were searched and relevant papers were identified to be included in the meta-analysis. The studies were selected based on well specified criteria, inherent to the relevance of each study for the meta-analysis and the accuracy of exposure and outcome definition, as well as the possibility of extracting the information needed for the analyses (i.e. relative risks and their 95% confidence intervals).

A problem of meta-analyses is the so called publication bias. Publication bias is the association of publication probability with the statistical significance of the study results, i.e. studies with statistically significant results are more likely to be published as compared to papers with negative findings.

One of the techniques used to assess publication bias is the funnel plot. The funnel plot is a graphical representation of the precision in the estimation of the treatment effect, increasing as the sample size of component studies increases. Results from small studies will scatter at the bottom of the graph. In the absence of publication bias, the plot will resemble a symmetrical, inverted funnel. Instead, the more asymmetrical is the graph, the more likely it is that the publication bias is substantial. An example of funnel plot is reported in **Figure 9**, where we can observe symmetry, indicating no evidence of publication bias.

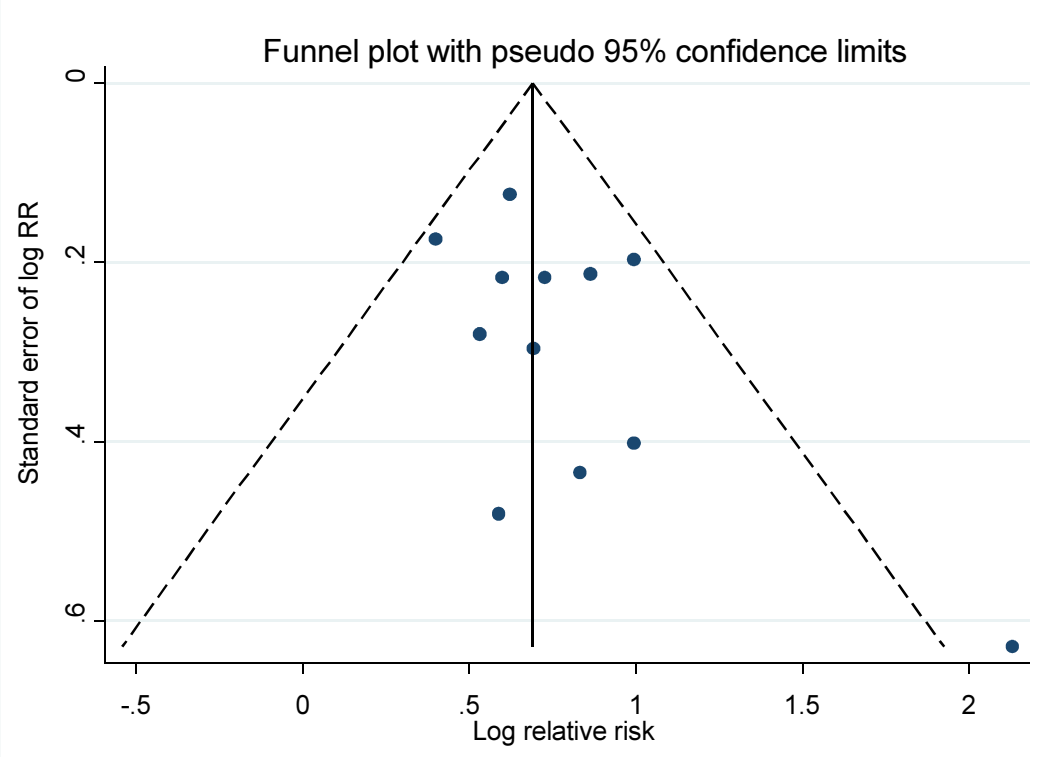

Figure 9. Funnel plot of the studies included in the meta-analysis on the association between pack-years of smoking and rheumatoid arthritis (**Paper II**)

A way to statistically evaluate the symmetry of the funnel plot is the Egger's test.[161] The Egger test is based on the estimate of a linear regression of the log relative risks on their standard errors, weighting by 1/(variance of the log relative risks). Under the null hypothesis of no publication bias, the regression line would be vertical in the funnel plot. In **Paper II** there was no evidence of asymmetry, i.e. no evidence of publication bias.

4.4.3.3 Statistical Software

The meta-analyses reported in **Paper II** were performed with Stata, version 12.1, using the command *metan* to summarize the relative risks estimates of women and men in the higher category of pack-years of smoking vs. never smokers from the included studies. To perform the dose-response meta-analysis the command *glst*[162] and *mvmeta*[163] were used.

5 RESULTS

The Swedish Mammography Cohort with baseline in 1997 includes data on 38 984 women. Due to the limitations that the data sources imposed (described in section 4.3.3), the follow-up in **Paper I, III, IV** and **V** started 1 January 2003 and therefore the study population was limited to the 35 187 women still alive and that did not developed rheumatoid arthritis (RA) or non-RA joint conditions before 2003. In this cohort 224 women developed RA before 31 December 2010 (end of follow-up in **Paper I, IV** and **V**).

5.1 CHARACTERISTICS OF THE STUDY POPULATION

The median age of the study participants in 1997 was 60 years (range 48-83). Regarding cigarette smoking, 23% of women were current smokers at the time of the questionnaire, and 23% were former smokers, with a median smoking intensity of 10 cigarettes per day. Swedish women drank a median of 4.35 grams of alcohol per day, with only 10% of the population drinking on average more than 1 glass per day of alcohol (15 grams of alcohol). Moreover, women in the SMC cohort were moderately active, with 70.5% of women walking or bicycling more than 20 minutes per day and 56.9% of women exercising one hour or more per week.

Comparing the 1987 FFQ and the 1997 FFQ, 64% of the women remained in the same category of fish consumption, therefore maintaining stable fish consumption over the 10 year period. Regarding alcohol consumption, 57% of the women kept their alcohol intake stable.

5.2 MAIN RESULTS

5.2.1 Cigarette smoking and rheumatoid arthritis (Paper I)

During the follow-up period, between 1 January 2003 and 31 December 2010 (254 996 person-years), 219 RA cases were identified among 34 101 women included in the study cohort. Among cases, 37% of women were current smokers, while in the whole cohort the prevalence of current smokers was 23%.

The risk of RA was doubled for women who were current smokers at the time of the questionnaire (RR=2.20, 95% CI: 1.58-3.04) and it was 68% higher for former smokers (RR=1.68, 95% CI: 1.19-2.38) compared to never smokers. All characteristics of smoking, such as intensity, duration and pack-years, were associated with an increased risk of RA, even for the lowest levels of the exposure (**Figure 10**).

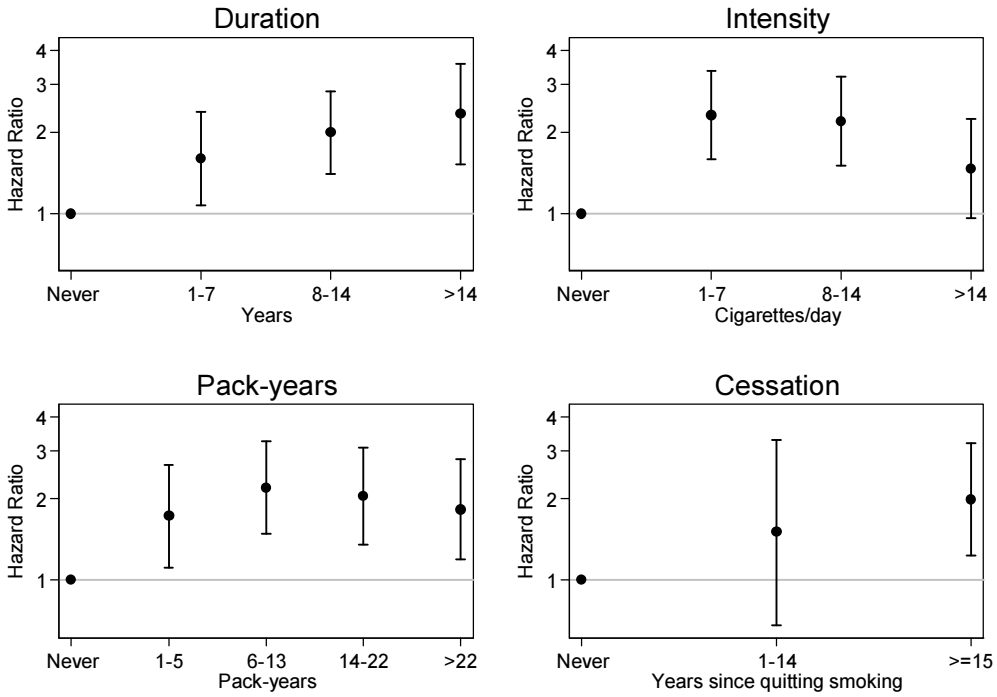

Figure 10. Relative risk of rheumatoid arthritis by smoking status, intensity, duration, and pack-years of cigarette smoking among ever smokers and years since quitting smoking among former smokers in the Swedish Mammography Cohort, 2003-2010.
Adjusted for age (continuous), menopause status (pre-/post-menopause), parity (0, 1, 2, ≥3 children), alcohol use (none, former, current), educational level (less than high school, high school, university), and BMI (quartiles). Intensity, duration and pack-years were not mutually adjusted[164].

Compared to never smokers, women who stopped smoking at least 15 years before the start of the follow-up still had a two-times higher risk of developing RA. Results regarding smoking cessation that were adjusted for smoking duration were still statistically significant (**Table 1**).

Table 1. Multivariable adjusted relative risk* and 95 percent confidence interval of rheumatoid arthritis by duration and years since quitting smoking in the Swedish Mammography Cohort, 2003-2008

			Years since quitting smoking			
	Cases	Never smokers	Cases	1-14 Years	Cases	>14 Years
Duration						
Never smokers	79	1.00 (ref)		-		-
1-20 years		-	1	0.78 (0.11-5.65)	29	1.81 (1.15-2.82)
>20 years		-	19	1.55 (0.93-2.59)	10	2.08 (1.06-4.05)

Adjusted for age (continuous), menopause status (pre-/post-menopause), parity (0, 1, 2, ≥3 children), alcohol use (none, former, current), educational level (less than high school, high school, university), and BMI (quartiles).

5.2.2 Alcohol and rheumatoid arthritis (Paper III)

Among the 34 141 women included in this study, 197 developed RA during the follow-up period (from 1 January 2003 to 31 December 2009; 226 032 person-years). The percentage of women who developed RA that were occasional drinkers (≤2 drinks per week) in 1997 was 53%, and 33% were regular drinkers (>2 drinks per week), compared to 49% and 34.5% of the whole cohort.

Occasional drinkers were used as the reference group in the analysis of the risk of RA related to alcohol status since they were the largest group. No difference in the risk of RA was observed among never drinkers when compared to occasional drinkers, while regular drinkers had a 19% lower risk of RA (RR=0.81, 95% CI: 0.59-1.11), although non statistically significant. The analysis regarding the number of glasses of alcohol per week showed that women who consumed more than 4 glasses per week, compared to the reference group of women who drank 1 or less glass per week, had a 37% lower risk of RA (RR=0.63, 95% CI: 0.42-0.96) (**Figure 11**). The analysis were adjusted only for smoking, since the inclusion in the Cox model of additional covariates such as body mass index (BMI), educational level, parity, menopausal status, and consumption of meat and dairy products did not change the estimate (RR=0.65, 95% CI: 0.43-0.99), but slightly reduced the statistical power of the analysis (wider confidence intervals).

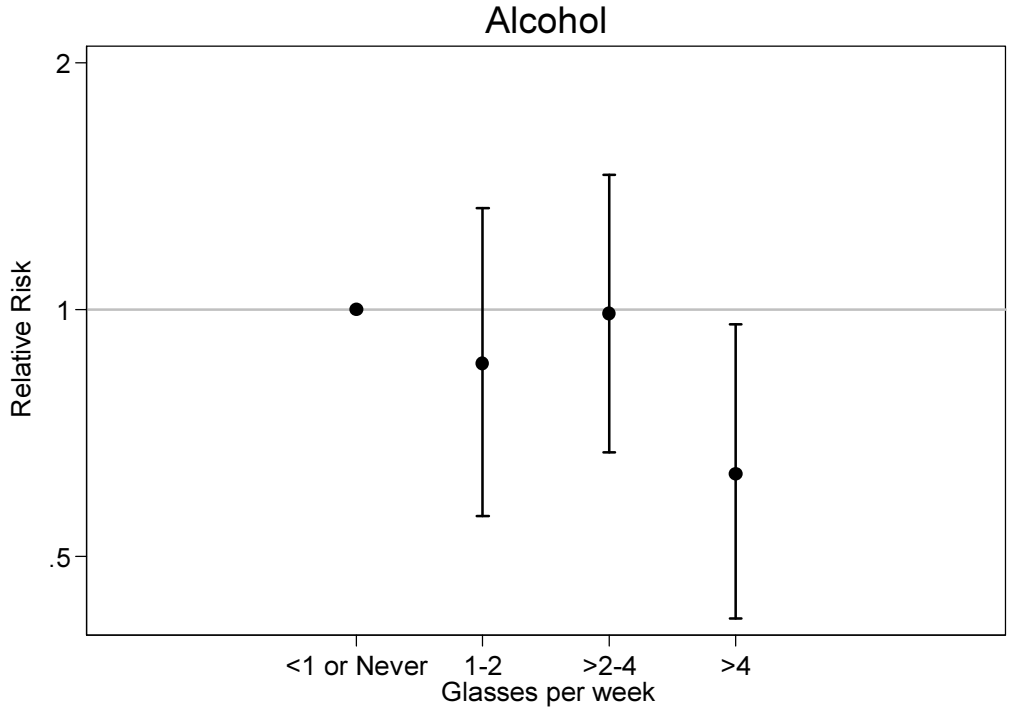

Figure 11. Relative risk and 95% confidence interval of rheumatoid arthritis by number of glasses of alcohol per week in 1997 in the Swedish Mammography Cohort, follow-up 2003-2009.

Adjusted for age (continuous), and smoking status (categorized as never, former, current ≤10 cigarettes/day or >10 cigarettes/day). Additional adjustment for BMI, educational level, parity, menopausal status, meat and dairy products consumption did not change the estimates. Median number of glasses per week in the highest category = 6

The estimated relative risk for women who drank more than 4 glasses of alcohol per week compared to women who consumed one or less glass per week was different among current smokers and never smokers. The relative risk was 0.77 (95% CI: 0.41-1.47) among current smokers, while it was 0.38 (95% CI: 0.15-0.97) among never smokers. However, there was no statistical interaction between smoking and alcohol.

Results regarding beer, wine and liquor consumption showed a similar level of reduction in RA risk for all three types of alcoholic beverages, supporting the hypothesis of a possible protective effect of alcohol against RA.

5.2.3 Long-chain n-3 PUFA and rheumatoid arthritis (Paper IV)

During the follow-up period (from 1 January 2003 to 31 December 2010; 241 120 person-years) 205 newly diagnosed cases of RA were identified among the 32 232 women included in the final study cohort. The variable regarding the long-chain n-3 PUFAs' was divided according to the quintiles of its distribution. The first quintile of the distribution was 0.21 grams per day (corresponding to more than 70 grams per week of salmon or 500 grams per week of cod) and women with a lower level of intake were considered as the reference group. The reference group included 27% of cases and 20% of the entire cohort.

The multivariable adjusted relative risks of RA were between 0.62 and 0.70 for quintiles 2 to 5 compared to the reference group. A threshold model in which the first quintile was used as threshold indicated a 35% decreased risk (95% CI: 0.48-0.90) comparing women with an intake of more than 0.21 grams per day of long-chain n-3 PUFAs to women with a lower intake. The risk among women who reported in 1987 an intake of long-chain n-3 fatty acids higher than 0.21 grams per day was 33% lower (RR=0.67, 95% CI 0.51-0.88) compared to women with a lower intake. Levels of EPA intake higher than 0.1 grams per day also reduced the risk of RA, but the estimates were not statistically significant (**Table 2**). DHA was not associated with risk of RA.

Table 2. Relative risk of rheumatoid arthritis in relation to dietary eicosapentaenoic (EPA) and docosahexaenoic acid (DHA) intake among 32 232 women in the Swedish Mammography Cohort, follow-up 2003-2010

	No. of cases	Multivariable-adjusted RR† ‡ (95% CI)
EPA[§]		
<0.06	53	1.00
0.06-0.085	38	0.75(0.41-1.37)
0.085-0.108	40	0.82(0.37-1.82)
0.108-0.152	34	0.47(0.18-1.23)
≥ 0.152	40	0.66(0.21-2.05)
DHA[§]		
<0.123	53	1.00
0.123-0.171	39	0.89(0.49-1.62)
0.171-0.216	32	0.74(0.33-1.66)
0.216-0.290	43	1.41(0.55-3.62)
≥ 0.290	38	1.13(0.36-3.56)

† RR= relative risk, CI= confidence interval
‡ Adjusted for age (continuous), cigarette smoking (never, former, current ≤10 cigarettes/day or >10 cigarettes/day), alcohol intake (never, former, current <2 drinks/week, ≥2 drinks/week), use of aspirin (yes, no), and intake energy (quintiles).
§ Estimates of EPA and DHA were based on data from the food-frequency questionnaire in 1997

There was a slight decrease in RA risk associated with fish consumption. The risk was lower for consumption of more than one serving per week of lean fish (**Table 3**), although none of the estimates were statistically significant. After adjustment for long-chain n-3 PUFAs the inverse association disappeared for both total and fatty fish, while it was attenuated for lean fish, and remained non statistically significant.

Table 3. Relative risk of rheumatoid arthritis by fish consumption in the Swedish Mammography Cohort, 2003-2009

	No. of women in the cohort	No. of cases	Multivariable-adjusted RR † (95%CI)	Multivariable-adjusted RR ‡ (95%CI)
Fish, servings per week				
<1	11927	78	1.00	1.00
1-2	9569	63	0.93 (0.66-1.30)	1.14 (0.77-1.68)
>2	10736	64	0.88 (0.63-1.24)	1.14 (0.70-1.84)
Fatty Fish, servings per week				
<3 times/month	13461	89	1.00	1.00
0.5-1 times/week	12084	81	0.95 (0.70-1.30)	1.22 (0.80-1.86)
>1	6687	35	0.83 (0.55-1.24)	0.95 (0.50-1.79)
Lean Fish, servings per week				
Never	3219	24	1.00	1.00
≤1	15170	99	0.81 (0.52-1.28)	0.86 (0.55-1.35)
>1	13843	82	0.78 (0.49-1.24)	0.86 (0.53-1.40)

† Adjusted for age (continuous), energy (quartiles), alcohol status (categorized as never, former, sporadic (2 or less drinks per week), and regular drinkers (more than 2 drinks per week), use of aspirin (yes, no), and smoking status (categorized as never, former, current ≤10 cigarettes/day or >10 cigarettes/day).
‡ Additionally adjusted for long-chain n-3 polyunsaturated fatty acids.

5.2.4 Physical activity and rheumatoid arthritis (Paper V)

Among the 30 112 women included in the study cohort, 201 had a first diagnosis of RA during the study period (from 1 January 2003 to 31 December 2010; 226 477 person-years). Among the cases 14.2% were in the lowest category of leisure time-activity, while 9.7% of the women in the whole cohort were in this group. Only 7% of women in the cohort walked more than 1.5 hours per day and only 11.3% exercised more than 5 hours per week.

Women who combined 20 or more minutes per day of walking/bicycling with 1 or more hour per week of exercise had a 35% decreased risk of developing RA (RR=0.65, 95% CI: 0.43-0.96) (**Figure 12**). None of the single physical activities analyzed separately was statistically significantly associated with risk of RA. However, there was a decrease in risk for hours of home/household work (32% decrease for more than 6 hours per day), hours of exercise per week (20% decrease for 2 or more hours per week), walking/standing at work (15% decrease compared to sitting), and duration of walking/bicycling per day (9% for 20 or more minutes per day). Moreover, there was an increased RA risk among women in the high category of inactive leisure-time (27% increase for more than 2 hour watching TV/sitting per day). Women with a daily energy expenditure of less than 35.3 MET-hours/day (5^{th} percentile of the distribution) had a 75% increased risk of RA (RR= 1.75, 95% CI: 0.97-3.17).

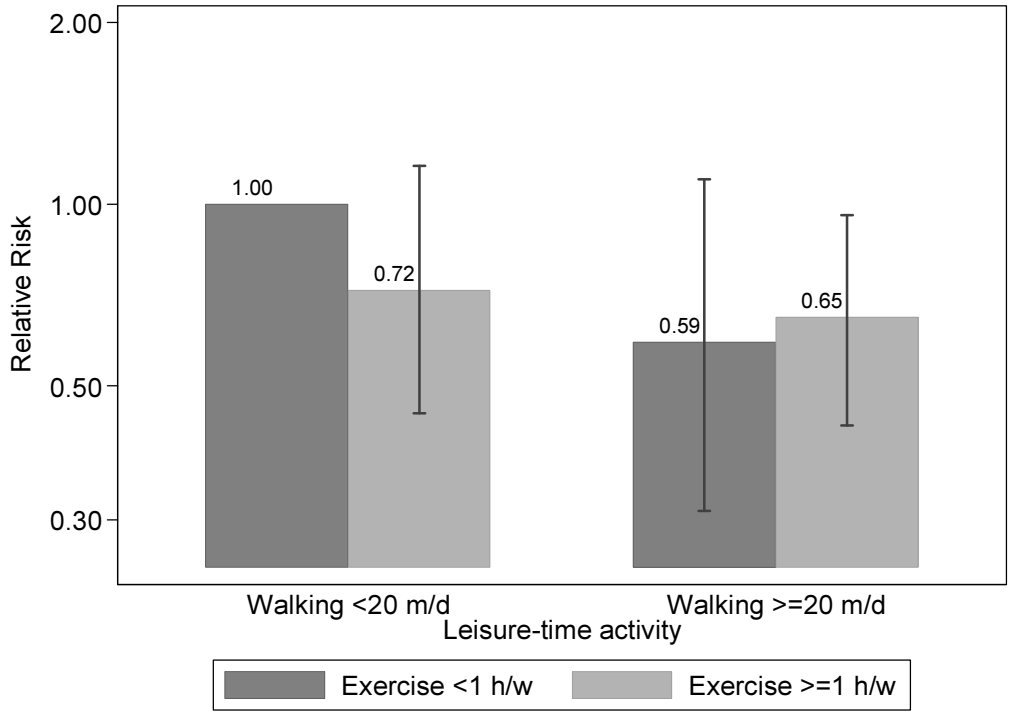

Figure 12. Multivariable-adjusted relative risk and 95% confidence interval of rheumatoid arthritis by leisure-time activity in the Swedish Mammography Cohort, follow-up 2003-2010.

Adjusted for age (continuous), smoking status (categorized as never, former, current ≤10 cigarettes/day or >10 cigarettes/day), alcohol intake (never, former, current <2 drinks per week, ≥2 drinks per week), body mass index (quartiles), and educational level (<10, 10-12, >12 years, other).

5.3 ANALYSES OF LONG-TERM EXPOSURES

In **Paper III** and **Paper IV** analyses of long-term exposures were performed by combining information from the 1987 and 1997 questionnaires.

In **Paper II**, women were divided according to their alcohol consumption (never drinkers, 3 or fewer glasses per week of alcohol, and more than three glasses) in both 1987 and 1997. The resulting variable had 9 levels, of which two did not include any participants (never drinkers in 1997 and ≤3 glasses per week or >3 in 1987), since women who reported being never drinkers in 1997 were also never drinkers in 1987. Never drinkers in both 1987 and 1997 were used as the reference group. The risk of RA was 52% lower (RR=0.48, 95% CI: 0.24-0.98) among women who drank more than 3 glasses per week in both 1987 and 1997 (**Figure 13**).

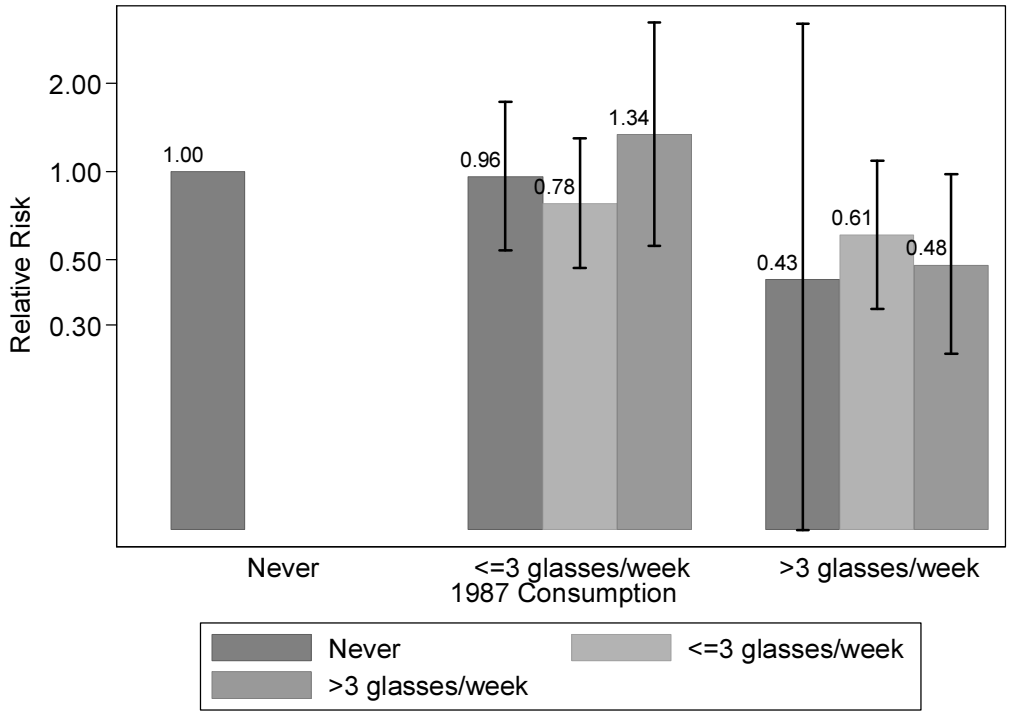

Figure 13. Multivariable adjusted relative risks of rheumatoid arthritis by weekly alcohol intake in 1987 and 1997 in the Swedish Mammography Cohort, follow-up 2003-2009

In **Paper IV**, both long-chain n-3 PUFA dietary intake and fish consumption were analyzed taking into account information from the 1987 and 1997 questionnaires. Long-term intake of more than 0.21 grams per day of long-chain n-3 PUFAs was associated with a 52% decrease in RA risk (RR=0.48, 95% CI: 0.33-0.71). Also the total consumption of one or more serving of fish per week in both 1987 and 1997 was associated with a 29% decrease in the risk of RA, but the estimate was not statistically significant (95% CI: 0.48-1.04). Moreover, after adjusting for long-chain n-3 PUFA intake, the inverse association disappeared (RR=1.14, 95% CI: 0.70-1.87). This result supports the hypothesis of previous studies that a possible protective role of fish against RA is mainly attributable to its content in long-chain n-3 PUFAs.

5.4 SENSITIVITY ANALYSES

5.4.1 Alternative definitions of start of follow-up

In **Paper I, III, IV** and **V** sensitivity analyses were performed to test the assumption regarding the identification of incident RA cases. In the main analyses, newly diagnosed cases of RA were identified from the Outpatient Register (OPR) and the Swedish Rheumatology Register (SRR) and the follow-up period started in 2003. However, these decisions could be questionable and could have led to misclassification of the outcome, with inclusion of prevalent cases as incident cases. To evaluate if these assumptions have in some way influenced the results, other assumptions were taken into consideration.

First, the start of follow-up was delayed from 2003 to 2004 and subsequently to 2006. The gap between 2001, start of the OPR, and 2003, start of follow-up, could have not been enough to avoid the inclusion of prevalent cases that were not yet included in a national register. Therefore, wash-out periods of 3 and 5 years were considered and the results did not show differences when compared to the main analysis (**Figure 14**).

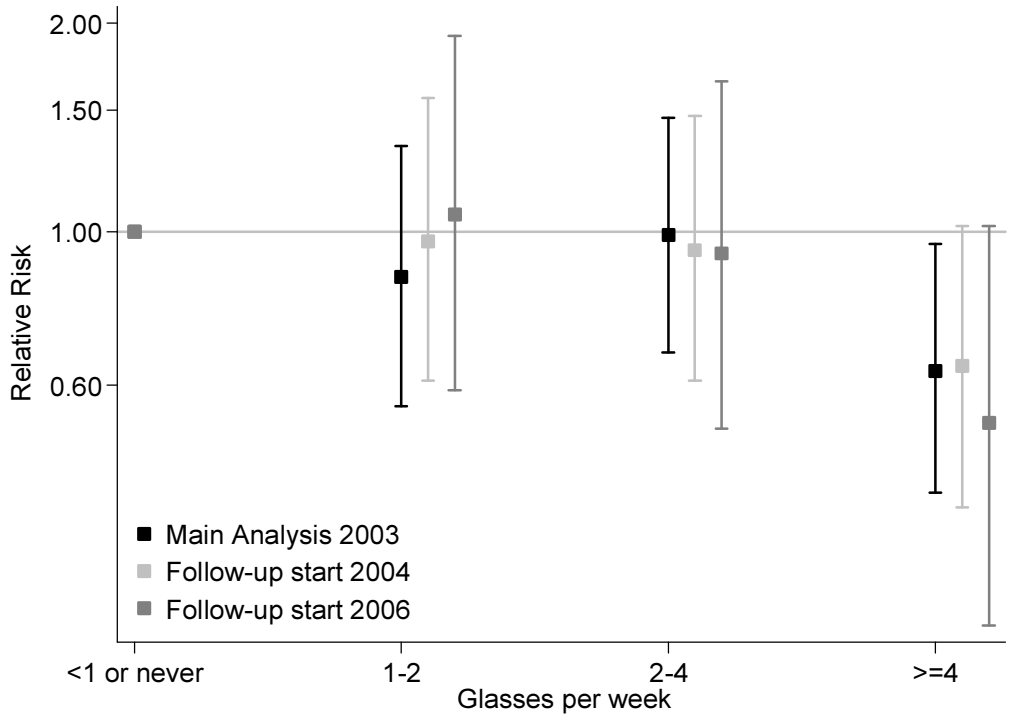

Figure 14. Multivariable adjusted relative risks and 95% confidence intervals for analyses performed with three different starts of follow-up period by number of glasses of alcohol per week in the Swedish Mammography Cohort

Newly diagnosed cases identified during the follow-up period through the Inpatient Register (IPR) were excluded from the analyses and considered as prevalent cases, since RA is a disease that in first stages does not usually lead to hospitalization. It is possible that this assumption could have been too strict, and therefore two additional models were considered, with the inclusion of the cases identified using the IPR. Two different start of follow-up dates were considered (2003 and 2004) and again results did not show remarkable differences.

5.4.2 Presence of prevalent cases

To evaluate the direction of the bias due to the inclusion of prevalent cases as incident cases, a probabilistic sensitivity analysis was conducted. Results from these sensitivity analyses for **Paper I, III, IV** and **V,** reported in their respective appendix, did not show significative changes in the results and therefore supported the conclusions of each paper.

The maximum amount of prevalent cases included was set to 20%. However, this cut-off could have been too low. Results from an additional sensitivity analysis with a maximum inclusion of 40% of prevalent cases are presented in **Table 4** for long-chain n-3 PUFAs. Estimates were similar under assumption 1 (prevalent cases did not change eating habits) and 3 (prevalent cases increased their consumption of foods rich in long-chain n-3 polyunsaturated fatty acids), while they were slightly increased under assumption 2 (prevalent cases decreased their consumption of foods rich in long-chain n-3 polyunsaturated fatty acids). However, it is more likely that prevalent cases did not change or increase consumption of fish rather than decreased it.

Table 4. Median values‡ and standard deviation of the distributions of RA relative risks by dietary long-chain n-3 polyunsaturated fatty acids intake according to three alternative assumptions about the behavior of prevalent cases among incident cases

Long-chain n-3 polyunsaturated fatty acids‡	Assumption 1	Assumption 2	Assumption 3
≤0.21 g/d	1.00	1.00	1.00
0.22-0.29 g/d	0.62±0.07	0.76±0.05	0.62±0.002
0.30-0.37 g/d	0.63±0.07	0.76±0.05	0.63±0.004
0.38-0.49 g/d	0.69±0.07	0.86±0.05	0.71±0.004
>0.49 g/d	0.66±0.07	0.82±0.05	0.55±0.04

1. Prevalent cases did not change eating habits; 2. Prevalent cases decreased their consumption of foods rich in long-chain n-3 polyunsaturated fatty acids; 3. Prevalent cases increased their consumption of foods rich in long-chain n-3 polyunsaturated fatty acids.

5.4.3 Missing values

In this thesis, partial non-response for foods items was considered as non-consumption and therefore considered as zero.[145] This hypothesis was tested in each paper through sensitivity analyses. In **Paper IV**, the long-term fish consumption was analyzed excluding women who did not respond to questions regarding fish consumption: results did not change, but wider confidence intervals were observed (**Table 5**).

Table 5. Multivariable adjusted relative risk and 95% confidence intervals of rheumatoid arthritis by long-term consumption of fish during 1987 and 1997, after exclusions of women with missing values for fish consumption

		Cases	Multivariable-adjusted RR (95%CI)	Cases	Multivariable-adjusted RR (95%CI)
			Intake 1987		
Fish§	Intake 1997		<1 serving/week		≥1 serving/week
	<1 serving/week	22	1.00	14	0.72 (0.37-1.42)
	≥1 serving/week	44	1.04 (0.62-1.75)	83	0.71 (0.44-1.15)

§ Adjusted for age (continuous), cigarette smoking (never, former, current ≤10 cigarettes/day or >10 cigarettes/day), alcohol intake (never, former, current <2 drinks/week, ≥2 drinks/week), use of aspirin (yes, no), red meat consumption (quartiles), dairy food consumption (quartiles), and energy intake (quartiles).

5.4.4 Fish-oil supplements (Paper IV)

In **Paper IV** women who used fish oil supplements were excluded from the main analysis, since the main focus of the paper was on dietary intake of long-chain n-3 fatty acids. A sensitivity analysis was performed including those women in the cohort who consumed fish oil supplements. Among a total of 34 399 women, 6.3% used fish oil supplements, while the prevalence of fish oil supplements consumption was 7.7% among women with RA (n= 222). In this sensitivity analysis the relative risk estimates for long-chain n-3 PUFAs were similar (RR=0.62, 95% CI: 0.41-0.93). The use of fish oil supplements was not associated with the development of RA. Due to the lack of information from the questionnaire, however, it was not possible to evaluate the effect of dose, duration, and composition of the supplements on RA.

5.5 POPULATION ATTRIBUTABLE RISK

The population attributable risk for current smokers was 0.22, therefore 22% of the cases would not have occurred in the population if all women were non-smokers (or former smokers) (**Paper I**).

The population prevented fraction in **Paper IV** was 0.28 and corresponded to the proportion of the hypothetical total load of the disease that has been prevented by exposure to more than 0.21 grams per day of long-chain n-3 PUFAs. The population prevented fraction was 0.25 for exposure to long-term intake of more than 0.21 grams per day of long-chain n-3 PUFAs and was 0.13 for long-term consumption of one or more servings per week of fish. The population prevented fraction calculated for leisure-time activity (**Paper V**) was 0.22 and for consumption of more than 4 glasses per week of alcohol (**Paper III**) was 0.08.

5.6 META-ANALYSIS (PAPER II)

The meta-analysis presented in **Paper II** was based on an extensive search of PubMed and EMBASE databases. Twenty-nine studies reported an estimate of the association between cigarette smoking and RA. Of these, only 3 cohort studies[165-167] and 7 case-control studies[46 52 54 168-171] reported an estimate of the association between lifelong exposure to cigarette smoking, expressed as pack-years, and risk of RA. Results from **Paper I** of this thesis were included in the meta-analyses presented in **Paper II**. All the studies included were evaluated in terms of quality using the Newcastle–Ottawa Quality Assessment Scale (NOQAS) for cohort and case-control studies, and each study was judged based on the selection of the study groups, the comparability of the groups, and the ascertainment of the exposure and the outcome. The score ranged between 0 (poor) and 9 (excellent), and all studies included in **Paper II** scored between 5 and 8.

The dose-response relationship between pack-years of cigarette smoking and risk of RA was modeled using a restricted cubic spline model. The non-linear dose-response trend (**Figure 15**) showed a statistically significant increased risk of developing RA with an increasing number of pack-years smoked up to 20 pack-years, and then the relative risk stabilized approximately around the value of two.

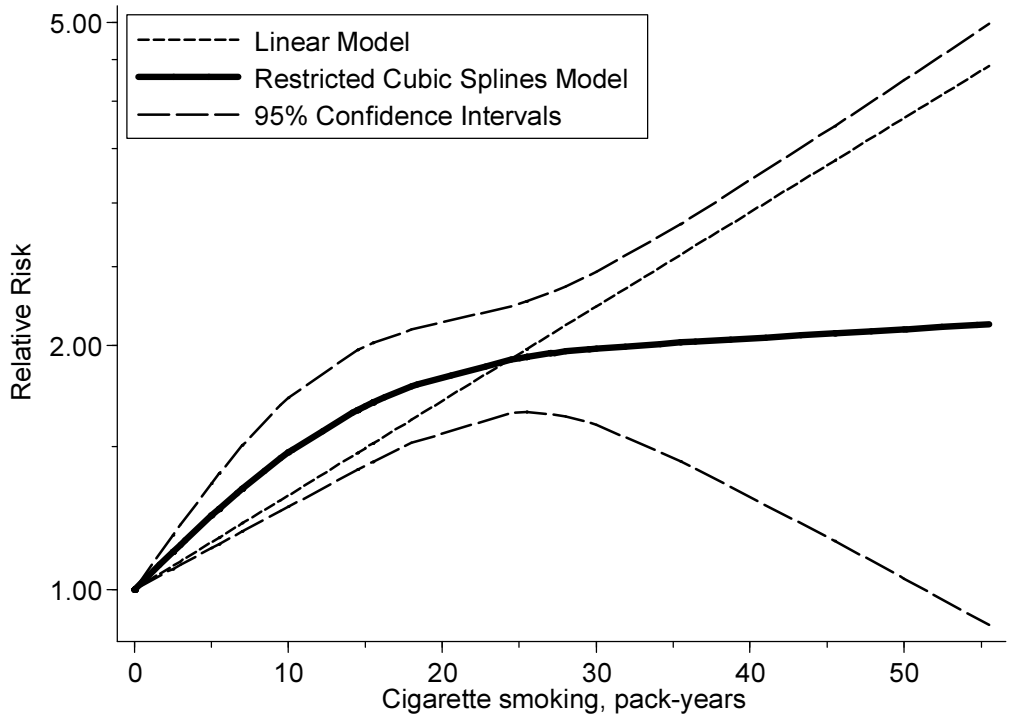

Figure 15. Non-linear dose-response relationship between pack-years of cigarette smoking and relative risk of rheumatoid arthritis based on summary of available evidence.

Relative risk (solid line) and 95% confidence interval (long dashed lines) from the restricted cubic splines model. The short dashed line represents the RR from the linear model. Estimates reported in the table are based on the median value of each category.

The dose-response analysis was supported by a comparison of the highest vs. the lowest categories of pack-years reported in each study. All studies reported as the reference level never smokers, while the median of the highest category ranged between 15 and 55 pack-years. The analysis showed a twofold increase of RA risk comparing the highest category of pack-years to never smokers (RR=2.02, 95% CI: 1.75-2.33).

There was no evidence of heterogeneity ($P_{heterogeneity} = 0.32$) or publication bias (p-value from the Egger's regression asymmetry test was 0.10).

6 DISCUSSION

6.1 METHODOLOGICAL CONSIDERATIONS

The objective of an epidemiological study is to obtain an accurate estimate of the association of an exposure on the occurrence of a disease in the source population of the study. Moreover, a further objective is to obtain an estimate that can be generalized to relevant target populations.

Accuracy in the estimation implies that the value of the parameter that is the object of measurement is estimated with little error. The errors in the estimation process are classified as random or systematic. Below follows a discussion of the types of errors that may have influenced the results of the papers included in this thesis.

6.1.1 Random error

Random error is often referred to as chance or random variation and can be defined as variability in the estimates that cannot readily be explained. A main component of random error is the process used to select the specific study subjects, called sampling. There is always a random variation in the estimate between samples.

An estimate with little random error (also called chance or random variation) may be described as precise. Statistical precision is often taken as the inverse of the variance of the measurements and therefore can be improved by increasing the sample size. In this thesis, precision was measured through 95% confidence intervals. If the confidence interval does not include the null value (RR=1), it is unlikely that the observed association is due to chance. The main results presented in this thesis were statistically significant, indicating that chance is an unlikely explanation for the observed associations.

In **Figure 16**, the statistical power has been plotted against different hazard ratios. From the graph it is possible to notice how the power rapidly decreases for hazard ratios that are near the null value.

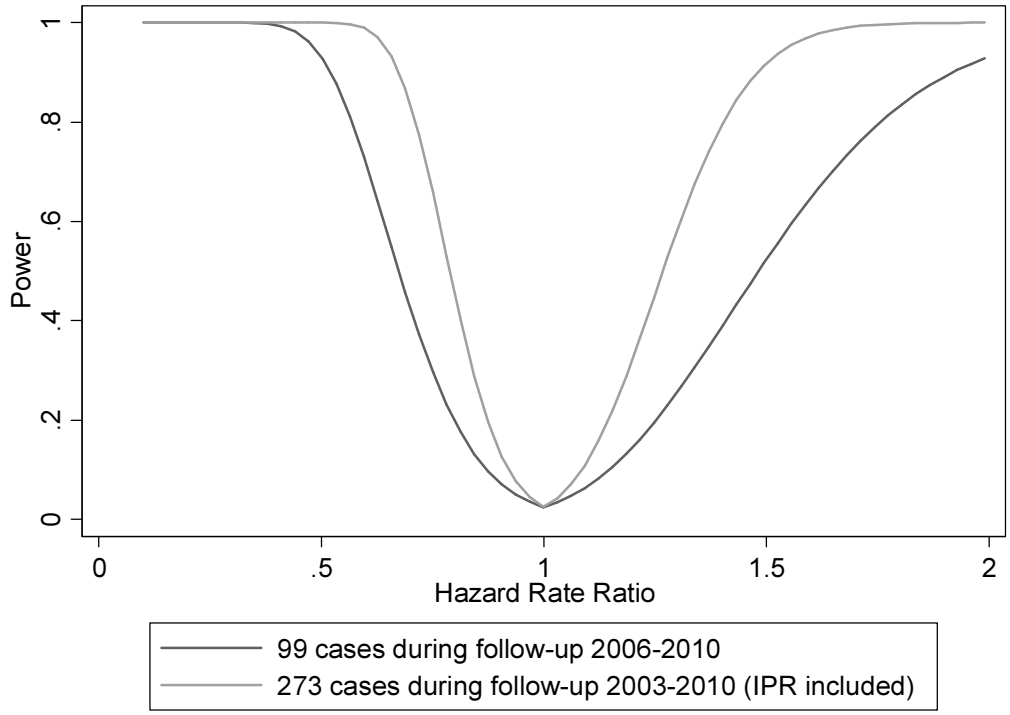

Figure 16. Statistical power associated with given hazard ratios according to number of cases in two sensitivity analyses from the Appendix of Paper IV.

6.1.2 Systematic error

Systematic errors in estimates are commonly referred to as biases. The validity of a study is usually separated into two components: internal validity and external validity (or generalizability). Most of the violations to internal validity can be classified as confounding, selection bias, and information bias.

6.1.2.1 Confounding

Confounding may be considered as the alteration of the effect of the exposure under study due to the effect of another factor that is mixed with the real exposure effect. The distortion introduced by a confounding factor can be large, and can lead to overestimation or underestimation of an effect. Formally, a factor can be considered a confounder of an association if (**Figure 17**):

- Is an independent risk factor for the disease
- Is associated with the exposure under study in the source population
- It cannot be an intermediate step in the causal path between the exposure and the disease.[172]

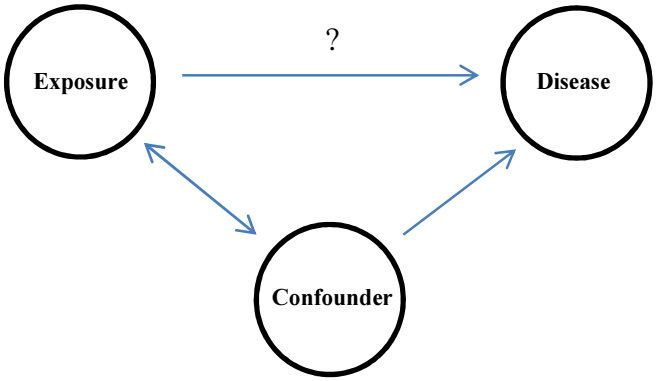

Figure 17. Directed Acyclic Graph (DAG) of the concept of a confounder.

In order to control for confounding, the potential confounders are added to the estimation model to adjust the estimate.

In **Paper I, III, IV** and **V**, many possible confounders have been taken into consideration and have been adjusted for in the model. However, only confounders that changed the beta coefficient by more than 5% were included in the final model, whereas the other variables were excluded. Age was accounted in the model as the time-scale, thus no assumption was made regarding the association between age and RA, since age was modeled in a non-parametric way and therefore error due to mismodeling was minimized.

Smoking was the main confounder of the associations presented in **Paper III, IV** and **V**. A possible interaction was evaluated in **Paper III** between alcohol consumption and smoking, but the interaction was not statistically significant. However, a difference in risk of RA was observed when the analysis was stratified by smoking status.

Confounding variables may be affected by measurement error. This error could lead to residual confounding, meaning that the association is still confounded even after the confounder is adjusted for in the analysis. Moreover, the possibility that unmeasured confounding may have influenced the observed results should be taken into consideration. For example, it is not possible to rule out that the findings were observed because of confounding by family history of RA, a potential confounder of the associations under study in this thesis, due to both shared genes and shared environmental factors. The inability to adjust for family history of RA is a limitation of these studies.

6.1.2.2 Selection bias

Selection bias is a distortion of the effect that may result from the procedures that have been used to select the study subjects or from the factors that influence the decision to participate in a study. A common consequence is a difference in the relation between exposure and disease among those who participated to the study and those who were eligible.[172]

Selection bias is a minor problem in prospective cohort studies because exposed and unexposed subjects are free from the disease at the time of enrollment. However, selection bias could arise if there is a difference in completeness of follow-up among exposed and unexposed. The virtually complete follow-up of participants in the SMC through linkage to various population-based registries minimizes this possibility.

6.1.2.3 Information bias

Information bias arises when the exposure or the outcome are not measured correctly. Errors in measurement can be caused by the observer (observer bias), the study participants (responder bias), or the instrument (e.g. the questionnaire structure). For categorical variables, information bias is often called misclassification, and can be classified as differential or nondifferential.

Misclassification is nondifferential when error in measurement of the exposure is not related to the outcome, or when error in the classification of the disease is not related to the exposure. Bias introduced by nondifferential misclassification can be predicted in direction for a binary exposure or outcome, towards the null value. However, when the exposure or the outcome have more than two levels, the bias could be either towards or away from the null value.[173]

Differential misclassification can bias the results in either direction. The most common differential misclassification is recall bias. Recall bias occurs in case-control studies, where cases are likely to recall the exposure differently than the healthy controls. For example, cases could over-report the consumption of certain foods considered unhealthy, while they could underestimate consumption of healthy foods in order to explain the causes of their disease. In this scenario, the bias would overestimate the association between diet and disease.

Misclassification of the exposure

The main exposures analyzed in this thesis (**Paper I, III, IV** and **V**) have been assessed using the 1997 questionnaire. Measurement error in dietary data derives from normal within-person variation in intakes over time and from errors associated with self-reporting. Due to the prospective design of the studies, in which all the participants were free from RA at the start of the follow-up, a possible potential misclassification of the exposure is likely to be nondifferential. Nevertheless, RA onset cannot usually be considered occurring at the same time of the date of diagnosis as there is likely substantial lag time between symptom onset and disease diagnosis. Therefore there is the possibility that women with early symptoms of RA could have changed their habits, leading to differential misclassification. However, the early assessment of exposure and covariates, 6 years before the start of the follow-up, should have helped in minimizing the differential misclassification.

Study participants could have under-reported their consumption of alcohol, since alcohol drinking can be considered a socially undesirable behavior. However, high validity of alcohol intake estimates indicated that the FFQ is a valid and reliable instrument of assessment. Similar bias could have occurred for physical activity, of which women could tend to over-report, since it is considered a healthy behavior. Also

for physical activity, the high validity suggested that this misclassification was relatively small.

Misclassification of the outcome
A possible source of nondifferential misclassification of the outcome may be the inclusion of prevalent cases as incident cases. In this situation, the misclassification is nondifferential, since the error is not related to the exposure of interest. This situation has been extensively evaluated through the use of simulations, and the consistency of the results supported that even if this bias was present, it had a small effect on the estimates.

6.1.2.4 External validity (generalizability)

The Swedish Mammography Cohort is a population-based cohort, where participants were invited from the general population of central Sweden. The response rate was 74% in 1987 and 70% in 1997. Since the 1997 questionnaire was sent only to women who answered to the 1987 questionnaire, only approximately 52% of the eligible women answered the 1997 questionnaire. Although this response rate could seem low, the study cohort well represents the entire Swedish population of middle-aged and elderly women, in terms of age, BMI and educational level distribution (**Table 6**).
However, results from the SMC, a cohort of middle aged and elderly women, cannot be generalized to younger women or to men. Moreover, the Swedish population is mainly Caucasian and the generalizability of these results to other ethnic population should be made with care.

Table 6. Comparison of the Swedish Mammography Cohort (SMC) with the Swedish population in 1997 for women aged 48-83 years, regarding age distribution, educational level, and body mass index (data from the Official Statistics Sweden). All values are percentages.

Characteristics	SMC 1997	Female Population, aged 48-83 years, 1997
Age groups, years		
Total, n	38 984	1 633 520
48-54	28.5	27.6
55-59	17.9	15.0
60-64	14.5	12.7
65-69	13.5	12.7
70-74	12.4	12.7
75-79	10.1	12.1
80-83	2.9	7.3
Education, ages 48-74 years		
Total, n	33 914	1 316 743
≤12 years	78.9	78.7
>12 years	20.5	19.9
Body Mass Index (>25 kg/m^2), by age groups		
45-54	37.6	38.8
55-64	45.7	47.4
65-74	49.7	52.0
75-84	43.0	42.3

6.1.3 Meta-analysis

Meta-analysis is a statistical tool to summarize published results regarding an association of interest. However, meta-analysis is affected by many limitations, mainly related to the quality of the summarized studies.

First, some meta-analysis, including **Paper II** of this thesis, summarizes results from both prospective cohort and case-control studies. However, these two types of designs have many differences, from the different statistical estimate (odds ratio in case-control studies, hazard ratio in prospective cohort studies) to different biases. For example, case-control studies could be affected by recall bias and selection bias, while prospective cohort studies are usually not.

Studies included in a meta-analysis usually account for confounding in different ways. Although results from the maximally adjusted model are usually selected, unmeasured confounding may still be present and it may have different effects on the results in the different studies. Moreover, the difference in the number of covariates adjusted for could depend on the study population. In fact, subjects from different studies differ in geographical location, gender, race, dietary habits, and many others aspects.

A quality score, the Newcastle–Ottawa Quality Assessment Scale (NOQAS) for cohort and case-control studies, was used in **Paper II** to assess the quality of each study included in the meta-analysis based on the selection of the study groups, the comparability of the groups, and the ascertainment of exposure and outcome. All studies included scored from moderate to high. Nevertheless, this score does not take into account the differences in study design. In **Paper II** a sensitivity analysis was performed by stratifying for study design and results were slightly different.

From a statistical point of view, the influence of differences between the studies is evaluated by testing heterogeneity. No heterogeneity was observed in the meta-analysis presented in **Paper II**. However, the limited number of studies (only 10) could have prevented the detection of heterogeneity.[160] In addition, publication bias was not detected, but the same problem related to power could explain the lack of statistical significance of the Egger test.

6.2 MAIN FINDINGS AND GENERAL DISCUSSION

6.2.1 Cigarette smoking and rheumatoid arthritis

Cigarette smoking is one of the few established risk factors for RA. Findings from this thesis support the role of smoking in developing RA, indicating a twofold increased risk among current smokers. Moreover, all characteristics of smoking were associated with risk of RA. In particular, **Paper I** highlighted that not only heavy smokers were at higher risk of RA, but also light smokers. In fact, women with a median intensity of 5 cigarettes per day had a 2.3 times higher risk of RA compared to never smokers.

Former smokers were also at higher risk of RA. Results showed that even women who stopped smoking 15 years before baseline in 1997 had a doubled risk of developing RA compared to never smokers. Moreover, age at quitting smoking did not play a role in reducing the risk of RA.

The results presented in **Paper I** are not surprising if seen from a pathogenic point of view. In fact, cigarette smoking acts as a trigger of the process called citrullination, that in later stages (even after several years) can lead to RA development. The trigger process could also explain why smoking cessation did not decrease RA risk compared to never smokers. However, this biological mechanism has been mainly related to ACPA positive RA, while less is known about the mechanism of smoking in ACPA negative RA. Moreover, studies have shown that the association between smoking and RA is less strong in ACPA negative and RF negative RA cases.[63] A limitation of this thesis, and in particular of **Paper I**, is the lack of stratification for RA subtypes (ACPA and RF status), due to a limited power.

In **Paper I**, cigarette smoking was assessed only at a single occasion, using the questionnaire sent in 1997. Repeated measurement of the exposure could help to avoid possible nondifferential misclassification. In fact, the analyses regarding smoking status in 2009 (when a third questionnaire was sent to women living in the study area who were still alive in 2009) showed that approximately 50% of the women had quit smoking, probably due to medical advice.

In **Paper II**, a meta-analysis regarding smoking and RA was conducted. In particular, the published estimates of the association between lifelong exposure to cigarette smoking and risk of RA were summarized. Results from the meta-analysis showed similar results to **Paper I**. The risk of RA increased up to 20 pack-years of cigarette smoking and then stabilized to a doubled risk for higher consumption. The risk was also statistically significant for light consumption, 1 to 10 pack-years.

6.2.2 Alcohol consumption and rheumatoid arthritis

Results from **Paper III** showed that RA development was inversely associated with alcohol consumption among women in the SMC. In particular, women who consumed more than 4 glasses of alcohol per week (median of 6 glasses per week) had a 37% lower risk compared to infrequent drinkers (<1 glass per week or never).

Despite that the category associated with the lowest RA risk was the one with the highest alcohol consumption in the SMC, conclusions referred to a moderate alcohol consumption. The word moderate was chosen because women in the SMC drank little alcohol. In fact, the median value of the upper category was 6 glasses per week, less than one glass per day, and only 1.4% of women consumed more than two glasses per day.

Conclusions on an inverse association of alcohol consumption with RA were supported by the analysis of beer, wine, and liquor consumption separately. Results, even when not statistically significant, showed a similar decrease in risk for all three types of alcoholic beverages.

Consistent long-term alcohol consumption of more than 3 glasses per week over a period of more than 10 years was associated with a halved risk of RA. It should be kept in mind that levels of alcohol consumption estimated in 1987 and 1997 were different. The median alcohol consumption in 1987 was 13.25 grams per week (~1 glass), while in 1997 was 30.85 (~2 glasses). The difference can be attributed to an overall increase in alcohol consumption among Swedish women. It is unlikely that the structure of the 1997 questionnaire caused this difference, even if the number of questions used to assess alcohol intake in the two questionnaires was different (in 1997 there was an additional question regarding the consumption of fortified wine with an alcohol volume of >18%).

Studies preceding the publication of **Paper III** were not consistent regarding the role of alcohol in the development of RA. Some case-control studies showed an inverse association between alcohol consumption and RA,[46 49 72 73] while prospective cohort studies showed no association.[70 71] The difference could be attributed to differences in the study design, in particular considering that case-control studies could have been affected by recall bias. However, previous prospective cohort studies had several methodological problems that could explain the discrepancy in results with case-control studies and with **Paper III** of this thesis. In fact, one study[70] did not adjust the analyses for cigarette smoking, an important confounder of the association, and another had a very limited number of cases (n=89).[71] Moreover, a following study,[74] together with two meta-analyses that summarized the published evidence,[75 76] also showed an inverse association.

6.2.3 Long-chain n-3 PUFAs and rheumatoid arthritis

Results from **Paper IV** showed an inverse association between long-chain n-3 PUFAs and RA. In particular, women with a dietary intake of more than 0.21 grams per day in both 1987 and 1997 had a halved risk of RA compared to women with a lower intake in both years.

The intake of 0.21 grams per day was the cut-off used in the threshold model and was also the first quintile of the long-chain n-3 PUFAs distribution in the study population. This intake level corresponds to at least one serving per week of fatty fish (~50 grams of salmon) or four servings per week of lean fish (~500 grams of cod). However, while this fish consumption was medium-low in the SMC, it can be considered medium-high in other (non-Swedish) populations where fish consumption is substantially lower.[174]

Average long-chain n-3 PUFAs intake in the SMC was different in 1987 and in 1997 (0.25 grams per day vs. 0.41 grams per day).[142] These estimates are not directly comparable, since they are based on a different number of food items (67-items in 1987, 96-items in 1997), and this could explain the difference. However, it is possible that the increase in the average long-chain n-3 PUFA intake is related to the increasing public attention given to health benefits associated with fatty fish consumption, a main source of long-chain n-3 PUFAs. According to statistics from the Swedish Board of Agriculture, per capita consumption of salmon (the most commonly consumed fatty fish), as well as of canned and processed fish, substantially increased (+43 and 15 %, respectively) between the late 1980s and the late 1990s in Sweden, although total fresh fish consumption decreased (−10 %).[175]

The observed threshold effect observed in **Paper IV** is similar to the threshold effect observed in a meta-analysis of studies on fish, fish oil and coronary heart disease risk.[176] That meta-analysis observed a threshold level of 0.25 grams per day of long-chain n-3 PUFAs, similar to the level reported in Paper IV. A similar level of intake can be obtained by eating fish twice a week, as recommended by the Swedish Food Administration and the American Dietary Guidelines Advisory Committee.

A limitation of **Paper IV** was the lack of detailed information regarding fish oil supplements. From the 1997 questionnaire only information regarding intake (yes/no) and number of capsules per week was available. However, these data were not sufficient to estimate an intake of long-chain n-3 PUFAs from supplements, since information about type of fish oil supplements, and therefore information on the content of long-chain n-3 PUFAs, was not asked. Results showed a lack of association between fish oil and risk of RA, but the number of cases was limited (n=17). In the main analysis, women who consumed fish oil supplements were excluded. In this way, the possibility of spurious association due to a mixed effect of long-chain n-3 PUFAs from supplements and dietary long-chain n-3 PUFAs was minimized.

The association between fish consumption and risk of RA was also evaluated. There was a moderate inverse association between fish consumption in 1997 and risk of RA, although not statistically significant. However, after adjustment for long-chain n-3 PUFAs, the inverse association disappeared. A similar scenario was observed when

analyzing long-term consumption of fish. There was a 30% decrease in RA risk among women with consumption of one or more serving of fish per week, compared to women with a lower consumption at both time points. However, after adjusting for long-chain n-3 PUFAs the association disappeared. These results indicate that the potential protective effect of fish against RA is completely mediated by its content of long-chain n-3 PUFAs.

6.2.4 Physical activity and rheumatoid arthritis

Physical activity is an important health related habit. Many are the articles promoting physical activity to prevent obesity and its comorbidities. Physical activities can be divided in activities performed during work hours (including housework), leisure time activities and leisure-time inactivity, including sitting, watching TV and reading. The activities that are easier to modify are leisure-time activity and inactivity.

The analysis of leisure-time activity in the SMC showed that combined exercise of one hour or more per week and walking 20 or more minutes per day was associated with a 35% decreased risk of RA. It was not possible to evaluate higher levels of leisure-time activities due to a lack of power, since only 7% of women in the cohort walked more than 1.5 hours per day and only 11.3% exercised more than 5 hours per week.

Separate models were used to evaluate the association of each type of physical activity with RA risk. Results from a model with all the activities mutually adjusted were similar, but shifted towards the null value. None of the relative risk estimates from any single activity reached statistically significance. The reason could be related to both a limited power and to nondifferential misclassification of the exposure. In fact, women who consider physical activity a healthy behavior may tend to over-report it.[177]

Physical activity was also evaluated by calculating total energy expenditure. Twenty-four hour total energy expenditure was not associated with the risk of RA. However, the calculation of total energy expenditure presented in **Paper V** was only a rough estimate based on information regarding only physical activity and number of sleeping hours. A more precise estimate of this variable would have included the thermic effect of food and more advanced techniques to assess physical activity during the day.[178]

Assessing physical activity using a questionnaire can lead to high levels of misclassification due to over-reporting, as mention before, but also to inability to assess precise duration and intensity of these activities. For example, women who reported 20 to 40 minutes of walking per day could have walked at a leisurely pace or walked very fast. However, validity of the estimates of leisure-time activity and inactivity using the questionnaire was relatively high, as compared to 7-day activity records.[144]

7 CONCLUSIONS

In this thesis the etiology of rheumatoid arthritis (RA) has been investigated. Results from the Swedish Mammography Cohort showed:

- Cigarette smoking is a risk factor for RA. Not only heavy smokers, but also light smokers were at higher risk of developing RA as compared to never smokers. Smoking cessation can help in reducing the risk of RA, but former smokers who stopped smoking even more than 15 years before baseline still had a doubled risk compared to never smokers

- Moderate alcohol consumption was associated with reduced risk of RA among Swedish women. This finding was not related to a specific type of alcohol.

- Long-term analysis of alcohol consumption based on data from 1987 and 1997 showed a stronger inverse association compared to estimates based only on data from 1997. The risk of developing RA was halved among women with a long-term alcohol consumption of more than 3 glasses per week.

- Long-chain n-3 polyunsaturated fatty acids (PUFAs) dietary intake was inversely associated with the risk of RA. The analyzed model showed a threshold effect of long-chain n-3 PUFA dietary intake.

- The risk of developing RA was halved among women with a long-term dietary intake of $\geq$0.21 grams per day of long-chain n-3 PUFAs. Long-term consumption of fish was also associated with a decreased risk of RA. However, the inverse association disappeared after adjusting for long-chain n-3 PUFAs, indicating this nutrient was the main cause of the protective effect against RA attributed to fish consumption.

- Leisure-time activity was inversely associated with risk of RA. The analysis of single activities showed a moderate decreased risk for physical activities (walking, exercising, working, housework) and a moderate increased risk for physical inactivity (watching TV/sitting), although none of the estimates were statistically significant.

8 FUTURE RESEARCH

The results presented in this thesis contribute to the knowledge about the etiology of rheumatoid arthritis. Further research should include:

- Large, prospective population-based cohort studies to evaluate the etiology of the different sub-types of rheumatoid arthritis.

- Large, prospective population-based cohort studies to evaluate the association between dietary factors and risk of rheumatoid arthritis, possibly with collection of blood and urine to measure intake of nutrients.

- Previous studies have presented different results depending on the method of preparation of foods. Future studies should take into account different methods of food preparation, such as raw, cooked, fried, grilled, etc., when analyzing the association between foods and risk of rheumatoid arthritis.

- Studies to address the association between high alcohol consumption and risk of developing RA.

- Studies to address the potentially important difference between daily consumption of a small quantity of alcohol compared to consumption of the corresponding moderate amount on one weekly occasion.

- Studies to investigate the potential interaction between environmental factors, such as alcohol consumption or diet, and genes.

9 ACKNOWLEDGEMENTS

Many people have contributed to this thesis and have helped me in "surviving" my PhD studies and for this reason I would like to thank them.

First of all I would like to thank all women of the Swedish Mammography Cohort: without your collaboration and patience this work would have not been possible.

Alicja Wolk, my main supervisor: you believed in me when I moved to Sweden even though my English was quite bad. I have worked hard to repay your trust and I hope that you are proud of my achievements.

Johan Askling, my co-supervisor: thank you for helping me in the definition of RA cases and for always suggesting new analyses and how to interpret the results from our studies.

Matteo Bottai, my co-supervisor: your enthusiasm for biostatistics is amazing and has rekindled my interest in statistics.

Rino Bellocco, my mentor: thank you for convincing me to start a PhD here in Sweden. Without your effort I would have not achieved all this.

My Italian (and a bit nerd) colleagues: *Nicola Orsini*, always saying that Stata is the best statistical software, while *Alessio Crippa* replies that R is ten times better (I am sorry but I will continue to use SAS), *Andrea Discacciati*, the official Stata helper (to be honest half of the graphs in this thesis have required your supervision), and *Andrea Bellavia*, the master of Laplace regression. Thank you guys for always making me laugh, particularly when I was homesick.

Bettina Julin, my "old" roommate, fadder, and dear friend: you have helped me in so many different ways. Since the first days you helped me with all the bureaucracy and during the years I have asked you so many times to translate my Swedish documents! Moreover, you explained me several Swedish folkways: I have been very lucky for having you as a roommate.

All my colleagues at the unit of NutEpi: *Niclas* (thank you for all the work that you always do with our data), *Alice* (it has been great to collaborate with you as a co-author), *Holly* (thank you so much for the English revision of this thesis), *Iffat* (my new roommate), *Agneta, Jinjin, Camilla, Susanne, Sanna, Ann, Laura, Viktor, Otto, Anna, Mimi, Carola, Thanasis, Katica, and Louise*.

Daniela Mariosa, PhD student at MEB and dear friend: you have a radiant personality and I am really missing our Sunday morning swims when we chatted between laps. Come back soon from North Carolina!

Heather and Måns Jonasson (and of course Max and Leo as well!): thank you for your friendship, for all the dinners and for your joie de vivre. You are an inspiration for me when I think about how to raise my little girl. Grazie di cuore!

My mum and dad: thank you for "kicking me" out of home when I was too afraid to move to Sweden, so far away from you. I really hope that you are proud of me. (La mia mamma e il mio papá: grazie per avermi "cacciata" di casa quando ero troppo spaventata di trasferirmi in Svezia, così lontano da voi. Spero che siate orgogliosi di me).

My grandparents, nonno Francesco, nonna Carina e nonna Carmellina: thank you for being so proud of me and for loving me so much. (Grazie per essere così orgogliosi di me e di amarmi così tanto).

My new family, *the Tettamanti*: Elena and Enrico, thank you for accepting me in your family as a daughter; Ele and Riky, I really love playing with you; Patty, Roby, Francy, Lele, Checco, for being great siblings-in-law. (La mia nuova famiglia, i Tettamanti: Elena e Enrico, grazie per avermi accettato in famiglia come una figlia; Ele e Riky, mi piace proprio tanto giocare con voi; Patty, Roby, Francy, Lele, Checco, per essere dei grandi cognati)

My little angel *Giada*: I am still wondering how it is possible that you are so perfect! You are so inquiring and always happy, you are the joy of my life.

And last but not least, my husband, *Giorgio Tettamanti*: this is all your fault!!! (and I love you so much for this!)

10 REFERENCES

1. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet* 2009;373(9664):659-72.
2. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *Lancet* 2010;376(9746):1094-108.
3. Symmons DP. Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. *Best Pract Res Clin Rheumatol* 2002;16(5):707-22.
4. Furneri G, Mantovani LG, Belisari A, Mosca M, Cristiani M, Bellelli S, et al. Systematic literature review on economic implications and pharmacoeconomic issues of rheumatoid arthritis. *Clin Exp Rheumatol* 2012;30(4 Suppl 73):S72-84.
5. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;43(1):30-7.
6. Jacobsson LT, Jacobsson ME, Askling J, Knowler WC. Perinatal characteristics and risk of rheumatoid arthritis. *BMJ* 2003;326(7398):1068-9.
7. Feitsma AL, Worthington J, van der Helm-van Mil AH, Plant D, Thomson W, Ursum J, et al. Protective effect of noninherited maternal HLA-DR antigens on rheumatoid arthritis development. *Proc Natl Acad Sci U S A* 2007;104(50):19966-70.
8. van der Helm-van Mil AH, Huizinga TW. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. *Arthritis Res Ther* 2008;10(2):205.
9. van Oosterhout M, Bajema I, Levarht EW, Toes RE, Huizinga TW, van Laar JM. Differences in synovial tissue infiltrates between anti-cyclic citrullinated peptide-positive rheumatoid arthritis and anti-cyclic citrullinated peptide-negative rheumatoid arthritis. *Arthritis Rheum* 2008;58(1):53-60.
10. Feldmann M, Brennan FM, Maini RN. Rheumatoid arthritis. *Cell* 1996;85(3):307-10.
11. Choy EH, Isenberg DA, Garrood T, Farrow S, Ioannou Y, Bird H, et al. Therapeutic benefit of blocking interleukin-6 activity with an anti-interleukin-6 receptor monoclonal antibody in rheumatoid arthritis: a randomized, double-blind, placebo-controlled, dose-escalation trial. *Arthritis Rheum* 2002;46(12):3143-50.
12. McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol* 2007;7(6):429-42.
13. Cohen SB, Dore RK, Lane NE, Ory PA, Peterfy CG, Sharp JT, et al. Denosumab treatment effects on structural damage, bone mineral density, and bone turnover in rheumatoid arthritis: a twelve-month, multicenter, randomized, double-blind, placebo-controlled, phase II clinical trial. *Arthritis Rheum* 2008;58(5):1299-309.
14. Schett G. Erosive arthritis. *Arthritis Res Ther* 2007;9 Suppl 1:S2.
15. Muller-Ladner U, Kriegsmann J, Franklin BN, Matsumoto S, Geiler T, Gay RE, et al. Synovial fibroblasts of patients with rheumatoid arthritis attach to and invade normal human cartilage when engrafted into SCID mice. *Am J Pathol* 1996;149(5):1607-15.
16. Schett G, Firestein GS. Mr Outside and Mr Inside: classic and alternative views on the pathogenesis of rheumatoid arthritis. *Ann Rheum Dis* 2010;69(5):787-9.

17. Klareskog L, Forsum U, Scheynius A, Kabelitz D, Wigzell H. Evidence in support of a self-perpetuating HLA-DR-dependent delayed-type cell reaction in rheumatoid arthritis. *Proc Natl Acad Sci U S A* 1982;79(11):3632-6.
18. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;30(11):1205-13.
19. Rieck M, Arechiga A, Onengut-Gumuscu S, Greenbaum C, Concannon P, Buckner JH. Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J Immunol* 2007;179(7):4704-10.
20. Klareskog L, Ronnelid J, Lundberg K, Padyukov L, Alfredsson L. Immunity to citrullinated proteins in rheumatoid arthritis. *Annu Rev Immunol* 2008;26:651-75.
21. Gullick NJ, Scott DL. Co-morbidities in established rheumatoid arthritis. *Best Pract Res Clin Rheumatol* 2011;25(4):469-83.
22. Gabriel SE. Why do people with rheumatoid arthritis still die prematurely? *Ann Rheum Dis* 2008;67 Suppl 3:iii30-4.
23. Radner H, Smolen JS, Aletaha D. Impact of comorbidity on physical function in patients with rheumatoid arthritis. *Ann Rheum Dis* 2010;69(3):536-41.
24. Symmons DP, Gabriel SE. Epidemiology of CVD in rheumatic disease, with a focus on RA and SLE. *Nat Rev Rheumatol* 2011;7(7):399-408.
25. Nicola PJ, Maradit-Kremers H, Roger VL, Jacobsen SJ, Crowson CS, Ballman KV, et al. The risk of congestive heart failure in rheumatoid arthritis: a population-based study over 46 years. *Arthritis Rheum* 2005;52(2):412-20.
26. Curtis JR, Xie F, Chen L, Spettell C, McMahan RM, Fernandes J, et al. The incidence of gastrointestinal perforations among rheumatoid arthritis patients. *Arthritis Rheum* 2011;63(2):346-51.
27. Smitten AL, Simon TA, Hochberg MC, Suissa S. A meta-analysis of the incidence of malignancy in adult patients with rheumatoid arthritis. *Arthritis Res Ther* 2008;10(2):R45.
28. Doran MF, Crowson CS, Pond GR, O'Fallon WM, Gabriel SE. Frequency of infection in patients with rheumatoid arthritis compared with controls: a population-based study. *Arthritis Rheum* 2002;46(9):2287-93.
29. Stein J, Reichert S, Gautsch A, Machulla HK. Are there HLA combinations typical supporting for or making resistant against aggressive and/or chronic periodontitis? *J Periodontal Res* 2003;38(5):508-17.
30. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology (Oxford)* 2002;41(7):793-800.
31. Pedersen JK, Kjaer NK, Svendsen AJ, Horslev-Petersen K. Incidence of rheumatoid arthritis from 1995 to 2001: impact of ascertainment from multiple sources. *Rheumatol Int* 2009;29(4):411-5.
32. Costenbader KH, Chang SC, Laden F, Puett R, Karlson EW. Geographic variation in rheumatoid arthritis incidence among women in the United States. *Arch Intern Med* 2008;168(15):1664-70.
33. Eriksson JK, Neovius M, Ernestam S, Lindblad S, Simard JF, Askling J. Incidence of rheumatoid arthritis in Sweden: a nationwide population-based assessment of incidence, its determinants, and treatment penetration. *Arthritis Care Res (Hoboken)* 2013;65(6):870-8.
34. Maetzel A, Li LC, Pencharz J, Tomlinson G, Bombardier C. The economic burden associated with osteoarthritis, rheumatoid arthritis, and hypertension: a comparative study. *Ann Rheum Dis* 2004;63(4):395-401.

35. Backman CL. Employment and work disability in rheumatoid arthritis. *Curr Opin Rheumatol* 2004;16(2):148-52.
36. Kavanaugh A. Economic consequences of established rheumatoid arthritis and its treatment. *Best Pract Res Clin Rheumatol* 2007;21(5):929-42.
37. Kalkan A, Hallert E, Bernfort L, Husberg M, Carlsson P. Costs of rheumatoid arthritis during the period 1990-2010: a register-based cost-of-illness study in Sweden. *Rheumatology (Oxford)* 2014;53(1):153-60.
38. Mishan EJ. Microeconomic Theory - a Mathematical Approach - Henderson,Jm, Quandt,Re. *Economica* 1959;26(104):365-66.
39. Symmons D MC, Pfleger B. The global burden of rheumatoid arthritis in the year 2000. 2006.
40. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31(3):315-24.
41. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;62(9):2569-81.
42. Criswell LA, Saag KG, Mikuls TR, Cerhan JR, Merlino LA, Lum RF, et al. Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. *Ann Rheum Dis* 2006;65(9):1163-7.
43. Hutchinson D, Shepstone L, Moots R, Lear JT, Lynch MP. Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA. *Ann Rheum Dis* 2001;60(3):223-7.
44. Olsson AR, Skogh T, Wingren G. Comorbidity and lifestyle, reproductive factors, and environmental exposures associated with rheumatoid arthritis. *Ann Rheum Dis* 2001;60:934-39.
45. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum* 2004;50(10):3085-92.
46. Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, et al. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. *Arthritis Res Ther* 2006;8(4):R133.
47. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis* 2003;62(9):835-41.
48. Uhlig T, Hagen KB, Kvien TK. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. *J Rheumatol* 1999;26(1):47-54.
49. Hazes JM, Dijkmans BA, Vandenbroucke JP, de Vries RR, Cats A. Lifestyle and the risk of rheumatoid arthritis: cigarette smoking and alcohol consumption. *Ann Rheum Dis* 1990;49(12):980-2.
50. Kallberg H, Ding B, Padyukov L, Bengtsson C, Ronnelid J, Klareskog L, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. *Ann Rheum Dis* 2011;70(3):508-11.
51. Krishnan E, Sokka T, Hannonen P. Smoking-gender interaction and risk for rheumatoid arthritis. *Arthritis Res Ther* 2003;5(3):R158-62.

52. Mikuls TR, Sayles H, Yu F, Levan T, Gould KA, Thiele GM, et al. Associations of cigarette smoking with rheumatoid arthritis in African Americans. *Arthritis Rheum* 2010.
53. Too CL, Yahya A, Murad S, Dhaliwal JS, Larsson PT, Muhamad NA, et al. Smoking interacts with HLA-DRB1 shared epitope in the development of anti-citrullinated protein antibody-positive rheumatoid arthritis: results from the Malaysian Epidemiological Investigation of Rheumatoid Arthritis (MyEIRA). *Arthritis Res Ther* 2012;14(2):R89.
54. Voigt LF, Koepsell TD, Nelson JL, Dugowson CE, Daling JR. Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. *Epidemiology* 1994;5(5):525-32.
55. Yahya A, Bengtsson C, Lai TC, Larsson PT, Mustafa AN, Abdullah NA, et al. Smoking is associated with an increased risk of developing ACPA-positive but not ACPA-negative rheumatoid arthritis in Asian populations: evidence from the Malaysian MyEIRA case-control study. *Mod Rheumatol* 2012;22(4):524-31.
56. Symmons DP, Bankhead CR, Harrison BJ, Brennan P, Barrett EM, Scott DG, et al. Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England. *Arthritis Rheum* 1997;40(11):1955-61.
57. Avila MH, Liang MH, Willett WC, Stampfer MJ, Colditz GA, Rosner B, et al. Reproductive factors, smoking, and the risk for rheumatoid arthritis. *Epidemiology* 1990;1:285-91.
58. Carlens C, Hergens MP, Grunewald J, Ekbom A, Eklund A, Hoglund CO, et al. Smoking, use of moist snuff, and risk of chronic inflammatory diseases. *Am J Respir Crit Care Med* 2010;181(11):1217-22.
59. Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med* 2006;119(6):503 e1-9.
60. Criswell LA, Merlino LA, Cerhan JR, Mikuls TR, Mudano AS, Burma M, et al. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. *Am J Med* 2002;112(6):465-71.
61. Heliovaara M, Aho K, Aromaa A, Knekt P, Reunanen A. Smoking and risk of rheumatoid arthritis. *J Rheumatol* 1993;20(11):1830-5.
62. Karlson EW, Lee IM, Cook NR, Manson JE, Buring JE, Hennekens CH. A retrospective cohort study of cigarette smoking and risk of rheumatoid arthritis in female health professionals. *Arthritis Rheum* 1999;42(5):910-7.
63. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis* 2010;69(1):70-81.
64. Klareskog L, Padyukov L, Alfredsson L. Smoking as a trigger for inflammatory rheumatic diseases. *Current Opinion in Rheumatology* 2007;19:49-54.
65. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006;54(1):38-46.
66. Mandrekar P, Catalano D, White B, Szabo G. Moderate alcohol intake in humans attenuates monocyte inflammatory responses: inhibition of nuclear regulatory factor kappa B and induction of interleukin 10. *Alcohol Clin Exp Res* 2006;30(1):135-9.

67. Lu B, Solomon DH, Costenbader KH, Keenan BT, Chibnik LB, Karlson EW. Alcohol consumption and markers of inflammation in women with preclinical rheumatoid arthritis. *Arthritis Rheum* 2010;62(12):3554-9.
68. Waldschmidt TJ, Cook RT, Kovacs EJ. Alcohol and inflammation and immune responses: summary of the 2005 Alcohol and Immunology Research Interest Group (AIRIG) meeting. *Alcohol* 2006;38(2):121-5.
69. Pedersen M, Stripp C, Klarlund M, Olsen SF, Tjonneland AM, Frisch M. Diet and risk of rheumatoid arthritis in a prospective cohort. *J Rheumatol* 2005;32(7):1249-52.
70. Cerhan JR, Saag KG, Criswell LA, Merlino LA, Mikuls TR. Blood transfusion, alcohol use, and anthropometric risk factors for rheumatoid arthritis in older women. *J Rheumatol* 2002;29(2):246-54.
71. Heliovaara M, Aho K, Knekt P, Impivaara O, Reunanen A, Aromaa A. Coffee consumption, rheumatoid factor, and the risk of rheumatoid arthritis. *Ann Rheum Dis* 2000;59(8):631-5.
72. Kallberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, et al. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. *Ann Rheum Dis* 2009;68(2):222-7.
73. Maxwell JR, Gowers IR, Moore DJ, Wilson AG. Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. *Rheumatology (Oxford)* 2010;49(11):2140-6.
74. Bergstrom U, Jacobsson LT, Nilsson JA, Wirfalt E, Turesson C. Smoking, low formal level of education, alcohol consumption, and the risk of rheumatoid arthritis. *Scand J Rheumatol* 2013;42(2):123-30.
75. Jin Z, Xiang C, Cai Q, Wei X, He J. Alcohol consumption as a preventive factor for developing rheumatoid arthritis: a dose-response meta-analysis of prospective studies. *Ann Rheum Dis* 2013.
76. Scott IC, Tan R, Stahl D, Steer S, Lewis CM, Cope AP. The protective effect of alcohol on developing rheumatoid arthritis: a systematic review and meta-analysis. *Rheumatology (Oxford)* 2013;52(5):856-67.
77. Booth FW, Gordon SE, Carlson CJ, Hamilton MT. Waging war on modern chronic diseases: primary prevention through exercise biology. *J Appl Physiol* 2000;88(2):774-87.
78. Warburton DE, Charlesworth S, Ivey A, Nettlefold L, Bredin SS. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. *Int J Behav Nutr Phys Act* 2010;7:39.
79. Khan KM, Thompson AM, Blair SN, Sallis JF, Powell KE, Bull FC, et al. Sport and exercise as contributors to the health of nations. *Lancet* 2012;380(9836):59-64.
80. van den Ende CHM, Breedveld FC, le Cessie S, Dijkmans BAC, de Mug AW, Hazes JMW. Effect of intensive exercise on patients with active rheumatoid arthritis: a randomised clinical trial. *Annals of the Rheumatic Diseases* 2000;59(8):615-21.
81. Han A, Robinson V, Judd M, Taixiang W, Wells G, Tugwell P. Tai chi for treating rheumatoid arthritis. *Cochrane Database Syst Rev* 2004(3):CD004849.
82. Stenstrom CH, Minor MA. Evidence for the benefit of aerobic and strengthening exercise in rheumatoid arthritis. *Arthritis Rheum* 2003;49(3):428-34.
83. de Jong Z, Munneke M, Lems WF, Zwinderman AH, Kroon HM, Pauwels EKJ, et al. Slowing of bone loss in patients with rheumatoid arthritis by long-term high-intensity exercise - Results of a randomized, controlled trial. *Arthritis and Rheumatism* 2004;50(4):1066-76.

84. de Jong Z, Munneke M, Zwinderman AH, Kroon HM, Jansen A, Ronday KH, et al. Is a long-term high-intensity exercise program effective and safe in patients with rheumatoid arthritis? Results of a randomized controlled trial. *Arthritis Rheum* 2003;48(9):2415-24.
85. de Jong Z, Vlieland TPMV. Safety of exercise in patients with rheumatoid arthritis. *Current Opinion in Rheumatology* 2005;17(2):177-82.
86. Hagen KB, Byfuglien MG, Falzon L, Olsen SU, Smedslund G. Dietary interventions for rheumatoid arthritis. *Cochrane Database Syst Rev* 2009(1):CD006400.
87. Di Giuseppe D, Wolk A. Diet and rheumatoid arthritis development: what does the evidence say? *International Journal of Clinical Rheumatology* 2014.
88. Jiang G, Li B, Liao X, Zhong C. Poultry and fish intake and risk of esophageal cancer: A meta-analysis of observational studies. *Asia Pac J Clin Oncol* 2013.
89. Wu S, Feng B, Li K, Zhu X, Liang S, Liu X, et al. Fish consumption and colorectal cancer risk in humans: a systematic review and meta-analysis. *Am J Med* 2012;125(6):551-9 e5.
90. Kolahdooz F, van der Pols JC, Bain CJ, Marks GC, Hughes MC, Whiteman DC, et al. Meat, fish, and ovarian cancer risk: Results from 2 Australian case-control studies, a systematic review, and meta-analysis. *Am J Clin Nutr* 2010;91(6):1752-63.
91. Li YH, Zhou CH, Pei HJ, Zhou XL, Li LH, Wu YJ, et al. Fish consumption and incidence of heart failure: a meta-analysis of prospective cohort studies. *Chin Med J (Engl)* 2013;126(5):942-8.
92. Vrablik M, Prusikova M, Snejdrlova M, Zlatohlavek L. Omega-3 fatty acids and cardiovascular disease risk: do we understand the relationship? *Physiol Res* 2009;58 Suppl 1:S19-26.
93. Benito-Garcia E, Feskanich D, Hu FB, Mandl LA, Karlson EW. Protein, iron, and meat consumption and risk for rheumatoid arthritis: a prospective cohort study. *Arthritis Res Ther* 2007;9(1):R16.
94. Di Giuseppe D, Wallin A, Bottai M, Askling J, Wolk A. Long-term intake of dietary long-chain n-3 polyunsaturated fatty acids and risk of rheumatoid arthritis: a prospective cohort study of women. *Ann Rheum Dis* 2013.
95. Linos A, Kaklamani VG, Kaklamani E, Koumantaki Y, Giziaki E, Papazoglou S, et al. Dietary factors in relation to rheumatoid arthritis: a role for olive oil and cooked vegetables? *Am J Clin Nutr* 1999;70(6):1077-82.
96. Linos A, Kaklamanis E, Kontomerkos A, Koumantaki Y, Gazi S, Vaiopoulos G, et al. The effect of olive oil and fish consumption on rheumatoid arthritis--a case control study. *Scand J Rheumatol* 1991;20(6):419-26.
97. Rosell M, Wesley AM, Rydin K, Klareskog L, Alfredsson L. Dietary fish and fish oil and the risk of rheumatoid arthritis. *Epidemiology* 2009;20(6):896-901.
98. Shapiro JA, Koepsell TD, Voigt LF, Dugowson CE, Kestin M, Nelson JL. Diet and rheumatoid arthritis in women: a possible protective effect of fish consumption. *Epidemiology* 1996;7(3):256-63.
99. Kontogianni MD, Panagiotakos DB, Pitsavos C, Chrysohoou C, Stefanadis C. Relationship between meat intake and the development of acute coronary syndromes: the CARDIO2000 case-control study. *Eur J Clin Nutr* 2008;62(2):171-7.
100. Kaluza J, Wolk A, Larsson SC. Red meat consumption and risk of stroke: a meta-analysis of prospective studies. *Stroke* 2012;43(10):2556-60.
101. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Intake of fat, meat, and fiber in relation to risk of colon cancer in men. *Cancer Res* 1994;54(9):2390-7.

102. Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. *Int J Cancer* 2006;119(11):2657-64.
103. Merlino LA, Curtis J, Mikuls TR, Cerhan JR, Criswell LA, Saag KG. Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum* 2004;50(1):72-7.
104. Boeing H, Bechthold A, Bub A, Ellinger S, Haller D, Kroke A, et al. Critical review: vegetables and fruit in the prevention of chronic diseases. *Eur J Nutr* 2012;51(6):637-63.
105. Slavin JL, Lloyd B. Health benefits of fruits and vegetables. *Adv Nutr* 2012;3(4):506-16.
106. Cerhan JR, Saag KG, Merlino LA, Mikuls TR, Criswell LA. Antioxidant micronutrients and risk of rheumatoid arthritis in a cohort of older women. *Am J Epidemiol* 2003;157(4):345-54.
107. Formica MK PJ, Rosenberg L, McAlindon TE. Lifestyle factors associated with the development of rheumatoid arthritis (RA): results for the Black Women's Health Study (BWHS). *Arthritis Rheum* 2001;44 Suppl 9: S376.
108. Mikuls TR, Cerhan JR, Criswell LA, Merlino L, Mudano AS, Burma M, et al. Coffee, tea, and caffeine consumption and risk of rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum* 2002;46(1):83-91.
109. Karlson EW, Mandl LA, Aweh GN, Grodstein F. Coffee consumption and risk of rheumatoid arthritis. *Arthritis Rheum* 2003;48(11):3055-60.
110. Garabrant DH, Dumas C. Epidemiology of organic solvents and connective tissue disease. *Arthritis Res* 2000;2(1):5-15.
111. Yang CS, Wang ZY. Tea and cancer. *J Natl Cancer Inst* 1993;85(13):1038-49.
112. Cleland LG, Hill CL, James MJ. Diet and arthritis. *Baillieres Clin Rheumatol* 1995;9(4):771-85.
113. Cantorna MT, Hayes CE, DeLuca HF. 1,25-Dihydroxycholecalciferol inhibits the progression of arthritis in murine models of human arthritis. *J Nutr* 1998;128(1):68-72.
114. Andjelkovic Z, Vojinovic J, Pejnovic N, Popovic M, Dujic A, Mitrovic D, et al. Disease modifying and immunomodulatory effects of high dose 1 alpha (OH) D3 in rheumatoid arthritis patients. *Clin Exp Rheumatol* 1999;17(4):453-6.
115. Costenbader KH, Feskanich D, Holmes M, Karlson EW, Benito-Garcia E. Vitamin D intake and risks of systemic lupus erythematosus and rheumatoid arthritis in women. *Ann Rheum Dis* 2008;67(4):530-5.
116. Hiraki LT, Munger KL, Costenbader KH, Karlson EW. Dietary intake of vitamin D during adolescence and risk of adult-onset systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2012;64(12):1829-36.
117. Huang Z, Rose AH, Hoffmann PR. The role of selenium in inflammation and immunity: from molecular mechanisms to therapeutic opportunities. *Antioxid Redox Signal* 2012;16(7):705-43.
118. Heliovaara M, Knekt P, Aho K, Aaran RK, Alfthan G, Aromaa A. Serum antioxidants and risk of rheumatoid arthritis. *Ann Rheum Dis* 1994;53(1):51-3.
119. Knekt P, Heliovaara M, Aho K, Alfthan G, Marniemi J, Aromaa A. Serum selenium, serum alpha-tocopherol, and the risk of rheumatoid arthritis. *Epidemiology* 2000;11(4):402-5.
120. McCord JM. Free radicals and inflammation: protection of synovial fluid by superoxide dismutase. *Science* 1974;185(4150):529-31.
121. Merry P, Winyard PG, Morris CJ, Grootveld M, Blake DR. Oxygen free radicals, inflammation, and synovitis: and synovitis: the current status. *Ann Rheum Dis* 1989;48(10):864-70.

122. Costenbader KH, Kang JH, Karlson EW. Antioxidant intake and risks of rheumatoid arthritis and systemic lupus erythematosus in women. *Am J Epidemiol* 2010;172(2):205-16.
123. Comstock GW, Burke AE, Hoffman SC, Helzlsouer KJ, Bendich A, Masi AT, et al. Serum concentrations of alpha tocopherol, beta carotene, and retinol preceding the diagnosis of rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheum Dis* 1997;56(5):323-5.
124. Karlson EW, Shadick NA, Cook NR, Buring JE, Lee IM. Vitamin E in the primary prevention of rheumatoid arthritis: the Women's Health Study. *Arthritis Rheum* 2008;59(11):1589-95.
125. Turner S, Cherry N. Rheumatoid arthritis in workers exposed to silica in the pottery industry. *Occup Environ Med* 2000;57(7):443-7.
126. Sluiscremer GK, Hessel PA, Hnizdo E, Churchill AR. Relationship between Silicosis and Rheumatoid-Arthritis. *Thorax* 1986;41(8):596-601.
127. Klockars M, Koskela RS, Jarvinen E, Kolari PJ, Rossi A. Silica Exposure and Rheumatoid-Arthritis - a Follow-up-Study of Granite Workers 1940-81. *British medical journal* 1987;294(6578):997-1000.
128. Steenland K, Sanderson W, Calvert GM. Kidney disease and arthritis in a cohort study of workers exposed to silica. *Epidemiology* 2001;12(4):405-12.
129. Lundberg I, Alfredsson L, Plato N, Sverdrup B, Klareskog L, Kleinau S. Occupation, Occupational Exposure to Chemicals and Rheumatological Disease - a Register-Based Cohort Study. *Scandinavian Journal of Rheumatology* 1994;23(6):305-10.
130. Gabriel SE, Crowson CS, O'Fallon WM. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955-1985. *Arthritis and Rheumatism* 1999;42(3):415-20.
131. Symmons DPM, Barrett EM, Bankhead CR, Scott DGI, Silman AJ. The Incidence of Rheumatoid-Arthritis in the United-Kingdom - Results from the Norfolk Arthritis Register. *Brit J Rheumatol* 1994;33(8):735-39.
132. Karlson EW, Deane K. Environmental and Gene-Environment Interactions and Risk of Rheumatoid Arthritis. *Rheum Dis Clin N Am* 2012;38(2):405-+.
133. Merlino LA, Cerhan JR, Criswell LA, Mikuls TR, Saag KG. Estrogen and other female reproductive risk factors are not strongly associated with the development of rheumatoid arthritis in elderly women. *Semin Arthritis Rheum* 2003;33(2):72-82.
134. Pikwer M, Bergstrom U, Nilsson JA, Jacobsson L, Berglund G, Turesson C. Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. *Ann Rheum Dis* 2009;68(4):526-30.
135. Brun JG, Nilssen S, Kvale G. Breast-Feeding, Other Reproductive Factors and Rheumatoid-Arthritis - a Prospective-Study. *Brit J Rheumatol* 1995;34(6):542-46.
136. Berglin E, Kokkonen H, Einarsdottir E, Agren A, Dahlqvist SR. Influence of female hormonal factors, in relation to autoantibodies and genetic markers, on the development of rheumatoid arthritis in northern Sweden: a case-control study. *Scandinavian Journal of Rheumatology* 2010;39(6):454-60.
137. Harris H, Hakansson N, Olofsson C, Julin B, Akesson A, Wolk A. The Swedish mammography cohort and the cohort of Swedish men: Study design and characteristics of 2 population-based longitudinal cohorts. *OA Epidemiology* 2013;1(2):16.
138. Bergström L KE, Hagman U, et al. The food composition database KOST: the National Administration's information system for nutritive values of food. *Vår Föda* 1991;43(439-447).

139. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* 1986;124(1):17-27.
140. Friberg E, Wolk A. Long-term alcohol consumption and risk of endometrial cancer incidence: a prospective cohort study. *Cancer Epidemiol Biomarkers Prev* 2009;18(1):355-8.
141. Wolk A, Ljung H, Vessby B, Hunter D, Willett WC. Effect of additional questions about fat on the validity of fat estimates from a food frequency questionnaire. Study Group of MRS SWEA. *Eur J Clin Nutr* 1998;52(3):186-92.
142. Wallin A, Di Giuseppe D, Burgaz A, Hakansson N, Cederholm T, Michaelsson K, et al. Validity of food frequency questionnaire-based estimates of long-term long-chain n-3 polyunsaturated fatty acid intake. *Eur J Nutr* 2013.
143. Messerer M, Johansson SE, Wolk A. The validity of questionnaire-based micronutrient intake estimates is increased by including dietary supplement use in Swedish men. *J Nutr* 2004;134(7):1800-5.
144. Orsini N, Bellocco R, Bottai M, Hagstromer M, Sjostrom M, Pagano M, et al. Validity of self-reported total physical activity questionnaire among older women. *Eur J Epidemiol* 2008;23(10):661-7.
145. Hansson LM, Galanti MR. Diet-associated risks of disease and self-reported food consumption: how shall we treat partial nonresponse in a food frequency questionnaire? *Nutr Cancer* 2000;36(1):1-6.
146. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
147. Baecklund E, Iliadou A, Askling J, Ekbom A, Backlin C, Granath F, et al. Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. *Arthritis Rheum* 2006;54(3):692-701.
148. Knight A, Sandin S, Askling J. Increased risk of autoimmune disease in families with Wegener's granulomatosis. *J Rheumatol* 2010;37(12):2553-8.
149. Neovius M, Simard J, Sundstrom A, Jacobsson L, Geborek P, Saxne T, et al. Generalisability of clinical registers used for drug safety and comparative effectiveness research: coverage of the Swedish Biologics Register. *Ann Rheum Dis* 2011;70(3):516-9.
150. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;24(11):659-67.
151. Hernandez-Carcia C, Vargas E, Abasolo L, Lajas C, Bellajdell B, Morado IC, et al. Lag time between onset of symptoms and access to rheumatology care and DMARD therapy in a cohort of patients with rheumatoid arthritis. *Journal of Rheumatology* 2000;27(10):2323-28.
152. Cox DR. Regression Models and Life-Tables. *J R Stat Soc B* 1972;34(2):187-+.
153. Schoenfeld D. Partial Residuals for the Proportional Hazards Regression-Model. *Biometrika* 1982;69(1):239-41.
154. Pearson K. Report on certain enteric fever inoculation statistics. *British medical journal* 1904;2:1243-46.
155. Glass GV. Primary, secondary, and meta-analysis of research. *Educational Researcher* 1976;10:3-8.
156. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* 1992;135(11):1301-9.
157. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med* 2012;31(29):3821-39.

158. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21(4):589-624.
159. Jackson D, White IR, Riley RD. A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biom J* 2013;55(2):231-45.
160. Gavaghan DJ, Moore RA, McQuay HJ. An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain* 2000;85(3):415-24.
161. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34.
162. Orsini N, Bellocco R, Greenland S. Generalized least squares for trend estimation of summarized dose-response data. *Stata J* 2006;6(1):40-57.
163. White IR. Multivariate random-effects meta-analysis. *Stata J* 2009;9(1):40-56.
164. McKnight B, Cook LS, Weiss NS. Logistic regression analysis for more than one characteristic of exposure. *Am J Epidemiol* 1999;149(11):984-92.
165. Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking Intensity, Duration, and Cessation, and the Risk of Rheumatoid Arthritis in Women. *American Journal of Medicine* 2006;119(6):503.e1-03.e9.
166. Criswell LA, Merlino LA, Cerhan JR, Mikuls TR, Mudano AS, Burma M, et al. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: Results from the Iowa Women's Health Study. *American Journal of Medicine* 2002;112(6):465-71.
167. Di Giuseppe D, Orsini N, Alfredsson L, Askling J, Wolk A. Cigarette smoking and smoking cessation in relation to risk of rheumatoid arthritis in women. *Arthritis Res Ther* 2013;15(2):R56.
168. Olsson Reckner A, Skogh T, Wingren G. Comorbidity and lifestyle, reproductive factors, and environmental exposures associated with rheumatoid arthritis. *Annals of the Rheumatic Diseases* 2001;60(10):934-39.
169. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis* 2003;62(9):835-41.
170. Yahya A, Bengtsson C, Lai TC, Larsson PT, Mustafa AN, Abdullah NA, et al. Smoking is associated with an increased risk of developing ACPA-positive but not ACPA-negative rheumatoid arthritis in Asian populations: Evidence from the Malaysian MyEIRA case-control study. *Modern Rheumatology* 2012;22(4):524-31.
171. Hutchinson D, Shepstone L, Moots R, Lear JT, Lynch MP. Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA. *Annals of the Rheumatic Diseases* 2001;60(3):223-27.
172. Rothman K, Greenland S, Lash TL. *Modern epidemiology*. 3 ed ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
173. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *American journal of epidemiology* 1990;132(4):746-8.
174. Wallin A, Di Giuseppe D, Orsini N, Patel PS, Forouhi NG, Wolk A. Fish consumption, dietary long-chain n-3 fatty acids, and risk of type 2 diabetes: systematic review and meta-analysis of prospective studies. *Diabetes Care* 2012;35(4):918-29.
175. The Swedish Board of Agriculture Statistics Database.

176. Mozaffarian D, Rimm EB. Fish intake, contaminants, and human health: evaluating the risks and the benefits. *JAMA* 2006;296(15):1885-99.
177. Rzewnicki R, Vanden Auweele Y, De Bourdeaudhuij I. Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutr* 2003;6(3):299-305.
178. Levine JA. Measurement of energy expenditure. *Public Health Nutr* 2005;8(7A):1123-32.