From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

CANCER AND CALCIUM

Epidemiological studies of Cancer Incidence and Survival

Cat Halthur

Karolinska Institutet

Stockholm 2011

Karolinska Institutet
Dep. of Medical Epidemiology and Biostatistics (MEB)
SE-171 77 Stockholm
SWEDEN

ISBN 978-91-7457-276-6

Akademisk avhandling som med tillstånd av Karolinska Institutet framlägges till offentlig granskning för avläggande av Medicine Doktorsexamen fredagen den 8 april 2011 klockan 9.00 i hörsal E 525, MTC, Karolinska Institutet, Theorellsväg 1, Stockholm.

Tryck: Larserics Digital Print AB

To my Family

Abstract

Studies have shown that there could be an association between dietary calcium and cancer, and more specifically it has been suggested that serum calcium is involved in the etiology of cancers of the prostate. In order to explore this association we performed epidemiological studies of the association between pre-diagnostic serum calcium and prostate cancer incidence and survival. We also studied temporal trends in the survival in Hodgkin's Lymphoma in order to evaluate the advances in treatment routines.

In our relative survival analysis of Hodgkin's Lymphoma we could see that those diagnosed in the later periods of time had considerably higher 1-, 5- and 10-year survival. This improvement in survival was especially prominent for those aged 51-65 years. Despite this, the long-term survival is still low in the older age groups.
We conclude that the recent improvements in treatment strategies in Hodgkin's Lymphoma have considerably improved survival in all ages. However, age is still an important factor indicating the need of further progress in diagnosis and treatment of older patients.

The key aim of this thesis was to study whether serum calcium is involved in the etiology and prognosis of cancers of the prostate. In our studies we did not find serum calcium to be associated with incidence of prostate cancer, incidence of fatal prostate cancer, prostate cancer-specific mortality, nor relative survival in prostate cancer. We did however find a small but significant association with incidence of prostate cancer in a stratified analysis of those men who entered the cohort at a young age, and with a high body mass index. In a descriptive analysis of the variance of serum calcium in correlation with other factors, we found that serum calcium was significantly associated with age, season of screening and estrogen related factors.
We conclude that though we did not find any association between prediagnostic serum calcium and prostate cancer in our study at large, our stratified analyses together with the descriptive analysis of variance makes it plausible that the association between calcium and cancer found in other studies, partly could be confounded by a mediating, if not causative factor, involving the mechanisms of calcium homeostasis, such as; sun exposure, vitamin D level, level of sex hormones, body constitution, or insulin levels.
Further studies of this mechanism in general, and its association with prostate cancer risk specifically, would be of interest when further exploring the association between cancer and calcium.

Work included in the thesis

List of Papers

Paper I Reproductive history, lifestyle factors and season as determinants for serum calcium concentrations in women
Martin Almquist, Anne-Greth Bondeson, Lennart Bondeson, **Cat Halthur**, Johan Malm and Jonas Manjer *The Scandinavian Journal of Clinical & Laboratory Investigation* **2008**, *68(8)*, 777–85

Paper II Serum calcium and the risk of prostate cancer
Cat Halthur, Anna L V Johansson, Martin Almquist, Johan Malm, Henrik Grönberg, Jonas Manjer, Paul W Dickman *Cancer Causes and Control* **2009 Sep**, *20(7)*, 1205-14

Paper III Prostate cancer mortality in Swedish men: Is serum calcium an important factor?
Cat Halthur, Jonas Manjer, Johan Malm and Paul W Dickman *Submitted to Cancer Causes and Control* **2010 Dec**

Paper IV Progress in Hodgkin lymphoma: A population-based study on patients diagnosed in Sweden 1973-2005
Cat Halthur, Jan Sjöberg, Sigurdur Y Kristinsson, Ola Landgren, Paul W Dickman and Magnus Björkholm *Submitted to Blood* **2010 Dec**

Supplementary analysis of variance

Similar analyses as those of the serum calcium variance in women (Paper I) was also performed for the men in the Malmö Preventive Project cohort. The results from this unpublished original descriptive analysis can be found in Chapter 8.

Contents

1 Introduction

Epidemiology is the study of the causes, distribution, and control of disease in populations.[A] This is the definition of epidemiology, but to me epidemiology is the magnifying glass in medicine; a tool that can detect potential risks or possible beneficial aspects in medicin.

Every other day a new risk factor for cancer is announced in the daily media. If it is not cellular phones or snuff (Swedish *'snus'*), it's potato chips or ketchup. Not rarely, what is proclaimed as a risk factor one day, is the next day declared to be a preventive factor. When it comes to milk[B] studies have shown that milk is good for the bone health, and might also prevent different diseases such as colorectal cancer,[1] but studies have also shown that it might increase the risk of other diseases such as prostate cancer.[2]

In epidemiology we often begin with a suspicion of an association, and wish to examine this further. That is, a correlation between a factor and an outcome, commonly a disease. My thesis consists of four articles that all look at cancer from an epidemiological point of view.

The heart of this thesis is the potential association between serum calcium and prostate cancer, both incidence and survival. In addition to this we wanted to study temporal trends in survival in Hodgkin's disease. I will explain why we thought these associations would be interesting to study, and also explain how and why we used the epidemiological and statistical methods that we did. To do so, I will start by presenting some clinical background of the different variables in these studies.

[A] Definition by the American Heritage Dictionary of the English language

[B] Not only milk, but dairy products in general

2 Background

Contents

2.1 Calcium

Calcium is a chemical element and an essential mineral for all mammals. Calcium can be found in e.g. green vegetables, nuts, and beans, but our main dietary source of calcium is dairy products such as milk and cheese. In the body, calcium is primarily found in the skeleton, but it is also important for cellular function, nerve signaling, muscular function, and blood coagulation.

Dietary calcium

Several epidemiological studies have found an association between dietary intake of calcium and disease, both positive and negative associations. However, when looking at the dietary intake of calcium one will automatically mainly study the intake of dairy products, seeing as this is the main source of calcium in our diet. However, dairy products also contain other components than calcium, which possibly could be an intermediar in the association between dairy products and disease. Other factors associated with dairy product intakemay have a *confounding effect*,[A] e.g. lactose intolerance. Though studies have found lactose intolerance to be associated with disease, it is difficult to know if this is a result of genetic co-variation, diet and/or ethnicity, seeing as lactose intolerance is more prevalent in e.g. the Mediterranean and Asian countries.

[A] For more information, please see page 15

Calcium homeostasis

Since the calcium balance is so important for the body, it is controlled by a complex and very strict system. The body can regulate the intestinal uptake, and also release calcium from the skeleton through temporary degradation of bone. The bone mass in the body is constantly undergoing degradation and regeneration in a process called bone turnover. This system, also known as the calcium homeostasis, is why the level of calcium in the blood is believed to be more or less unaffected by the dietary intake of calcium.

Serum calcium

Even though the association between dietary calcium and the blood level of calcium is disputed, studies have shown a possible association between serum calcium level and consequent disease, such as breast cancer.[3]

Free or ionized calcium In order to estimate the amount of calcium in the blood, one can measure the amount in the serum. Either you analyze *total* calcium, or you measure the level of free, unbound calcium, i.e. *ionized* calcium (Ca^{2+}). The level of *total* calcium can be misleading in individuals with an anomaly in the amount of calcium binding protein due to other malignancies. The analysis measuring *ionized* calcium was less common in the 70-ies why old cohorts seldom have information on Ca^{2+}-levels. Ionized calcium is considered the most accurate and/or relevant measure of calcium, but, total calcium is generally considered a fair estimate of serum calcium in healthy men.[4] It is however, possible to calculate the free calcium level from the total calcium level based on the concentration of the protein albumin, since approximately 40 percent of calcium in blood is bound to this protein. In article III we have calculated an estimated level of free calcium in those with data on serum albumin, using the formula:[5]

Corrected calcium (mmol/L) = Total Calcium (mmol/L) + 0.02 [40 (g/L) - albumin (g/L)]
where 40 is an average amount of serum albumin in g/L

Vitamin D

Parathyroid hormone (PTH) is a hormone produced in the parathyroid gland, partly responsible for the release of calcium in the blood, both by increasing absorption in the intestine as well as controlling bone resorption. PTH has been found to be associated with cancers such as breast cancer.

Vitamin D is a group of fat-soluble steroids who, together with Paratyhroid Hormone, is involved in maintaining a balance between calcium and phosphates in the blood. It exists mainly in two forms; D_2 (ergocalciferol) and D_3 (cholecalciferol). Vitamin D_2 is mainly attained through dietary intake of vegetables. Vitamin D_3 is partly attained through ingestion of dietary products like fish and nuts, but most part is attained through the production in skin at exposure to ultraviolet B light from the sun. The blood level of vitamin D therefore, in general, tends to vary with season. Elderly or dark skin is known to produce lower amounts of vitamin D. Vitamin D_3 is considered much more potent than vitamin D_2. Vitamin D deficiency was long considered

as a rare disease showing symptomatically as Rickets or Osteomalacia, but recent studies have shown possible associations with a number of diseases, e.g. depression and cancer.[6] It is possible that some of the diseases known to be more prevalent at higher latitudes, and among e.g. African Americans are a result of a consequently low level of vitamin D due to skin color in combination with little sun light exposure. Sun light exposure, the main source of vitamin D, was recently found to be protective against prostate cancer[7]. Vitamin D has also been found to vary significantly with season of screening, even in countries with little seasonal difference.[8]

2.2 CANCER

When cells divide, they sometimes get damaged. All cells are naturally programmed to die (i.e. go into apoptosis) in cases of error. However, some cells have a malfunction in their control system, allowing them not only to live, but sometimes even to continue to divide. If the body does not detect this, or is disabled to liquidate the erroneous cells, these cells can form a tumor. Cancer is a malignant tumor; that is a tumor that not only consists of erroneous cells, but actually continues to grow and damage surrounding tissue. Cancer is one of the leading causes of death in the world, with almost 7.6 million fatalities in 2008 and this number is expected to grow to over 11 million in 2030.[9,10] The most common cancer worldwide is lung cancer, with more than 1.6 million new cases and almost 1.4 million fatalities in 2008.[9] More than 30 percent of all cases are believed to be preventable.

Prostate cancer

The prostate is a gland in men situated near the rectum. The gland is responsible for producing 30 percent of the seminal fluid, and is believed to help make the semen alkaline. Prostate cancer is the fourth most common cancer worldwide and the second most common cancer in men.[9] Many possible risk factors have been identified, such as endogenous factors (e.g. ethnicity and genes), but also exogenous factors such as diet and latitude of residence.[11] Despite this the etiology of prostate cancer for most parts still remains unknown. In the late 80-ies a promising method to detect prostate cancer was introduced, namely the screening of Prostate Specific Antigen. Through this the number of new cases diagnosed per year increased greatly, but despite this it is unclear whether screening of prostate cancer improve survival.[12]

Prostate-Specific Antigen (PSA) is a protein produced by the cells of the prostate gland. PSA is present in small quantities in the serum of healthy men, but is often elevated in the presence of prostate disorders such as cancer. The protein is believed to help liquefy the semen, and also help dissolving the cervical mucus cap.

Breast cancer

Breast cancer is the second most common cancer in the world, and the most common cancer in women. In 2008 there were more than a million new cases worldwide. The

mortality is not as high as in lung cancer, but still account for almost half a million deaths every year.[9] Some identified risk factors are malfunctioning genes, such as the BRCA genes, but most established risk factors involve the female hormone system, ranging from hormone replacement therapy (including oral contraceptives) to obesity.[B] Some protective factors include parity (number of children) and breast feeding. Other suggested factors associated with breast cancer are serum calcium (risk factor),[3] and vitamin D (protective effect).[13]

Hodgkin's Lymphoma

Hodgkin's Lymphoma (HL) is a specific cancer affecting the lymphocytes (white blood cells). Since HL affects the lymphatic nodes, its effect is systemic. In 2008 there were almost 70,000 cases reported worldwide and almost 30,000 fatalities.[9] Thanks to advances in treatment the cancer is considered highly curable. However, seeing as the cancer requires a systemic therapeutic treatment there is a high risk of adverse effects caused by the treatment itself such as cerebral and cardiovascular damage as well as occurrence of new cancers.

Non-Hodgkin lymphoma is a group name for lymphomas *not* categorized as Hodgkin's. These lymphomas range from almost indolent to highly aggressive cancers.

Colon cancer

During fetal development the colon grows from two different origins and is joined somewhere in the middle. This conjunction is not seen with the eye, but studies have shown that the etiology of cancer might differ depending on which part of the colon it is situated in. Usually colon cancer is grouped together with rectal cancer and then referred to as colorectal cancer. Known risk factors are smoking, and alcohol, but also fatty food, and diet low in fiber. Obesity and/or lack of exercise is also a known risk factor. One theory is that the colon develops tumors if it absorbs mal nutrients and toxins, but another theory is that an under stimulated colon develop cancer. Milk and dairy products have been found to be *negatively* associated with colon cancer, i.e. have a *protective* effect.

2.3 OTHER VARIABLES

Body mass index (BMI)

The Body mass index is a rough measure of body constitution. It is calculated by taking the weight in kilograms divided by the height in centimeters squared. Though not exact, this estimate of body constitution is commonly used as a proxy for (measure of) either fat deposit, and/or physical fitness. However, body builders and similar athletes tend to have a BMI indicating overweight, whereas e.g. swimmers tend to have a BMI indicating underweight, why the individual measure should be interpreted with caution. Those with a BMI over 30 are commonly considered obese, and those with a BMI between 25 and 30

[B] Fat tissue is known to produce estrogens

are considered to have an overweight. Underweight is set at a BMI of either 18.5 or 20. In my articles I have used the latter.

Socioeconomic status (SES)

Socioeconomic status, sometimes referred to as socioeconomic index, is a proxy for (measure of) living standards and economic welfare. In the cohort *Malmö Preventive Project*[C] it was measured as manual or non-manual work, and not related to income or economic situation as such. However, this may be estimated to be a fair proxy for economic status in the 70-ies when the cohort was collected.

Age

The estimated variables and covariates in a study are very often affected by age, why almost every study takes this into consideration in some way. However, age in epidemiological studies can stand for different things. Either its age at entry to a cohort, age when a sample is drawn, the age at the exposure to a risk factor or a health promoting factor, age at onset of a disease, or a combination of several of these.

Estrogen related factors

Diseases related to biological mechanisms often tend to be affected by hormones such as estrogen (female sex hormone). The level of estrogen is of course affected by the intake of oral contraceptives (OC) and hormone replacement therapy (HRT), both with estrogen or estrogen-like components. However, the hormone level is also affected by menarche (age at first menstruation) and menopause, why the age of first and last period can summarize a woman's production of and exposure to her self-produced estrogens. Also how much a woman has breastfed (a factor known to prevent e.g. breast cancer) and parity (number of pregnancies) describe a woman's estrogen levels. Another, less obvious factor affecting the estrogen levels include amount of fat tissue and insulin level.

Sex

The risk of a disease sometimes differ dependant on if the individual is a man or a woman. The strength of an association can also differ depending on this factor. The biological difference between men and women is sometimes important, especially when estrogen related factors are of relevance. However, sometimes the difference between risk is dependent on the different roles of women and men in society, and not the sex per se, i.e. the *gender*.

Marital status

Marital status is usually classified as unmarried, married, or widowed. This factor is especially relevant when the studied disease is related to psychological wellbeing, and then act as a proxy for (marker of) social support.

[C] For more information please see page 11

Smoking

Smoking is a well known risk factor for many diseases. Not only does it leave toxins in the lung, and damage the tissue of the lungs, but the nicotine also affect the blood system and nerve system. It also puts a stress on the immune system. Depending on the proposed affect of smoking it is sometimes measured in detail. Either the subjects state the number of cigarettes per day during different periods of their lifes and we thereafter calculate number of packages per year/month. However, this method is sensitive to e.g. the honesty of the respondent. Another measure is a blood sample where the nicotine derivates and other components are measured. This is not sensitive to self estimation of number of cigarettes, and is also not sensitive to variations in how much smoke a study participant absorbs from each cigarette. However, the measurement generally only reflects the status at the time when the blood sample is drawn. If the exact amount of smoke is less important one can categorize the variable e.g. as never/ever smoker, or as non-smoker, current smoker and former smoker. This is a rough estimate of amount of cigarette components an individual has absorbed, and does not take into consideration that potential difference between someone who used to smoke in their young adulthood but never since, someone who has smoked for 35 years and just stopped, and someone who started smoking just a year ago.

Alcohol consumption

Alcohol is a factor known to be associated with different diseases in different ways. Some studies have seen that alcohol increase the risk of disease as it is a toxin. Other studies have seen positive effects of alcohol. It is possible that the source of alcohol is relevant, since some studies have found a beneficial effect of e.g. a glass of red wine. It is not clear if this association is a result of the alcohol in it-self, other components in the wine, the psycological effect on the mind, or simply a proxy for a specific lifestyle and/or a high living standard. Seeing as heavy alcohol consumption can be shameful to admit, and therefore difficult to assess through a simple question, one can use different types of systematic general questions. One of these systems are the Michigan Alcoholism Screening Test (MAST) [14] and the Malmö modified MAST (Mm-MAST) [15] with specially constructed questions including attitudes towards alcohol. One of the questions included in the Mm-MAST is e.g. *Do you drink a couple of drinks (or beers) a day to relax?*. These questions are then graded according to a set system and the result is interpreted as an estimate of general alcohol consumption.

2.4 Background of associations

Serum calcium and prostate cancer

Studies have shown that dairy products might be associated with an increased risk of prostate cancer, and some have shown that there may be an association between dietary calcium intake and risk of prostate cancer [16–21]. Giovannucci et al hypothesized that a

high intake of calcium causes cancer by forcing the body to lower the levels of the tumor suppressing hormone vitamin D ($1,25(OH)_2D$) in order to prevent hypercalcemia.[22]

Studies have still not been able to determine an association between high calcium intake and low levels of vitamin D. Even if dietary calcium affects vitamin D levels, this mechanism does not have to involve a pathway affecting the measurable amounts of serum calcium seeing as the serum calcium is very tightly regulated by the calcium homeostasis mechanism.[23]

Not many studies have examined the association between serum calcium and risk of prostate cancer, but in 2008 Skinner et al found indications of a (non-significant) positive association between serum calcium and prostate cancer in an American cohort.[24] They also saw an association between serum calcium and incidence of *fatal prostate cancer*, i.e., the incidence of prostate cancers that caused death.[24] This finding of an association with fatal prostate cancer was later supported by an additional study where they speculate that serum calcium is implicated in the etiology of only aggressive prostate cancer, and that previous studies showing little or no association between serum calcium and prostate cancer, was diluted by the inclusion of incident cases of indolent prostate cancer.[25]

Long-term survival in Hodgkin's Lymphoma

The treatment routines in patients with Hodgkin's Lymphoma has undergone great improvements and now include; more accurate radiotherapy, effective combinations of chemotherapy, immunotherapy, improved staging procedures and important developments in supportive measures for myelosuppression, infectious and other complications. Hodgkin's Lymphoma has become a highly curable disease, especially for those diagnosed with an early-stage disease. However, despite major advances over the past decades in the treatment of patients with Hodgkin's Lymphoma, the survival is affected by secondary complications such as; infections, primary cancers at other sites, and cardiovascular and cerebrovascular disease.

2.5 Aim

The purpose of this thesis is to improve human health by enabling further prevention and more effective treatment of cancer through knowledge gained from four specific epidemiological studies of cancer incidence and survival. Specifically, we will study whether serum calcium is involved in the etiology of cancers of the prostate, and the association between pre-diagnostic levels of serum calcium and survival of men diagnosed with prostate cancer. An additional specific aim is to evaluate effectiveness of treatment through studying temporal trends in survival of patients diagnosed with Hodgkin's lymphoma,

3 Material

Contents

3.1 Study populations

Malmö Preventive Project

In paper I-III, as well as in the the supplementary analysis of variance of serum calcium, in Chapter 8, we have looked at an unusually large cohort called the Malmö Preventive Project (MPP) (in Swedish called *Malmö Förebyggande Medicin*). Malmö is the third largest city in Sweden with almost 280.000 residents, and is situated in the south. The Malmö Preventive Project was started in 1974 in Malmö and invited selected birth-year cohorts born in Malmö between the years 1920 and 1945. In 1984 the study expanded to also include the women of selected birth-year cohorts between the years 1926 and 1949. More than 33.000 men and women participated, resulting in a participation rate of more than 70 percent. In paper I we study the 10.902 women of the cohort, and in paper II and III, as well as the supplementary analysis in Chapter 8, we study the 22.444 men of the cohort.

The original purpose of the cohort study was to evaluate general health with a special focus on cardiovascular disease, and the participants were given a health examination, asked to fill in a questionnaire on lifestyle factors etc, and also informed about the importance of healthy eating and regular exercise.[26]

Cases with Hodgkin's Lymphoma In paper IV the study population was all those diagnosed with Hodgkin's Lymphoma in Sweden between 1973 and 2005, i.e. 6,136 men and women. This data was retrieved from the National Cancer Registry. We thereafter compared the survival in those diagnosed during four different time-periods, between which the treatment routines had progressed in order to reduce treatment-related morbidity.

3.2 VARIABLES

Prostate cancer

In paper II the outcome is *incidence of prostate cancer*. In December 2006, 1.539 of the men had been diagnosed with prostate cancer. Cases were identified by record linkage with the nationwide Swedish Cancer Registry and the Southern Swedish Regional Cancer Register.

Fatal Prostate Cancer

In paper III, parts of the study have *incidence of fatal prostate cancer* as the outcome, measured as prostate cancer diagnosis that later was followed by prostate cancer related death. In December 2008 the number of prostate cancer diagnoses had increased to 1.753, and out of these 270 where considered fatal. In paper III we also study overall mortality measured as *excess mortality* among those previously diagnosed with prostate cancer. In this group there were 612 fatalities. Information on death, and cause of death, was obtained through the Swedish Cause of Death Registry.

Hodgkin's Lymphoma In paper IV we studied the overall mortality in all those diagnosed with Hodgkin's Lymphoma in Sweden between 1973 and 2005.Data was retrieved from the National Cancer Registry. **The outcome** was *excess mortality* compared to the Swedish population in general.

Serum Calcium

In paper I-III we study serum calcium measured continuously over a period of almost 15 years. During this time the type of instrument routinely used at the department of clinical chemistry was exchanged several times. A phenomenon common in laboratory analyses are drift, i.e. the analysis instrument tend to get out of tune, little by little with each analysis or day that passes. This is easily dealt with by re-calibrating the instrument at intervals using e.g. a reference sample. Some of the instruments used in the analysis of serum calcium in our study had quite large variation, which could be a consequence of infrequent calibration. However, it also turned out that the baseline measures of the different instruments differed, and no data on reference sample was available. This meant that the mean serum calcium levels before 1980 is different than that of those analyzed after 1980. Similar to the problem with drift, this would not present a problem unless the proportion of cases and non-cases differed between these time periods. However, the subjects were invited to the cohort by birth-year, and seeing as prostate cancer is a disease highly related to age, this led to a difference between the periods, not only in age, but also in incidence of prostate cancer.

The Categorization of Serum Calcium In paper I and II we categorized the variable serum calcium by quartiles. Per definition, quartiles are all subjects sorted from lowest to highest, and then divided into four evenly large groups. However, seeing as the mean of the

serum calcium differed between instruments, a simple quartilation could not be performed. Instead of stratifying the following analyses on period of serum calcium measurement, we corrected for the measurement interval by making two separate quartile definitions; one for samples analyzed before October 1980 (71.7 percent of all subjects), and one for those analyzed after October 1980. The two subsets of the data were thus categorized from 1 to 4 according so that each individual's serum calcium was valued in relation to the others in the same subset. The variable for each subset was then concatenated and treated as one variable of serum calcium. This enabled us to look at the association between higher and lower serum calcium, not affected by the variation between periods of analysis. The numerical value for each of the four quartile groups was 1, 2, 3 and 4 respectively.

When several subjects has the same value as a cut-off value between quartiles, the group-size can differ somewhat. As seen in the tables in paper II and III, the difference in group-size between the quartiles is prominent. This is caused by the low number of unique values in the original variable. Instead of randomly assigning the individuals with the cut-off value to either the quartile in question, or the quartile beneath this, we chose to continue with uneven groups to prevent risks of miss-classification.

In paper III, where we study the association between serum calcium and mortality, some of the analyses are performed on the full cohort and other parts only on those with prostate cancer diagnosis. We thus did one quartile categorization for the full cohort, and one for only the prostate cancer cases in order to prevent the risk that a weak association could disappear in the graininess of the data.

Albumin corrected calcium The MPP cohort also included level of serum albumin level (in g/L) for 76 percent of the men in the cohort (n=17,096 out of 22,439 men), and for 83 percent of those with a prostate cancer diagnosis (n=1,457 out of 1,755). This measure was used in order to calculate the level of free calcium using the formula below.[5]

Corrected calcium (mmol/L) = Total Calcium (mmol/L) + 0.02 [40 (g/L) - albumin (g/L)]
where 40 is an average amount of serum albumin in g/L

Other variables

The data from the Malmö Preventive project cohort also include the variables: **Smoking status** (categorized as never, current and former), and **alcohol consumption** (categorized as low, medium and high, derived from a value according to the Mm-MAST system[A]), both from the questionnaire on life-style parameters. **Body Mass Index**, calculated from height and weight measured by a nurse at baseline examination at entry to the study. **Season of screening**, i.e. time of baseline examination, categorized as winter, spring, summer and fall. **Socioeconomic index** or status, and **marital status**, retrieved from the 1980 Swedish population and housing census.[27]

[A] For more information, please see page 8

4 Methods

Contents

4.1 Directed Acyclic Graphs (DAG)

When discussing epidemiological studies we often start with a suspected association between two variables, one is often a risk factor, referred to as *the exposure*, and the other is often a disease, referred to as *the outcome*. We also commonly have other co-varying factors we may or may not be able to include in the analysis. Sometimes these covariates are very important and must be controlled for in the analysis, but this may be difficult to tell from start.

In order to facilitate the decision on what variables to stratify on, adjust for, condition on and in other ways control for, and also to elucidate what methods to apply in doing so, it is possible to use some type of graphical diagram such as a Directed Acyclic Graph (see figure 4.1). This is a visual aid, and though it is far more complex than my explanation here, I will use it as a mean to visualize parts of the projects included in this thesis.

The outcome we wish to study is commonly denoted **(d)** for **disease** and is the commonly the illness we wish to find the source of. **The exposure** is the risk factor (or health promoting factor) we wish to study, and is commonly denoted **(e)** for **exposure**. In a DAG there are also **other factors** or variables potentially associated with the exposure and/or the outcome. Although some of these covariates, such as **(a)**, may be irrelevant for the association at focus, some of these should (if possible) be adjusted for in the analysis, either by restriction, matching or by including them in the statistical model. One of the factors in this DAG is believed to partly cause the exposure *(e)*, but also cause the outcome *(d)*. Such a factor is called *a confounder* and is denoted **(c)**. It is very important to adjust for such variables in the analyses, since they otherwise can cause false associations, or block parts

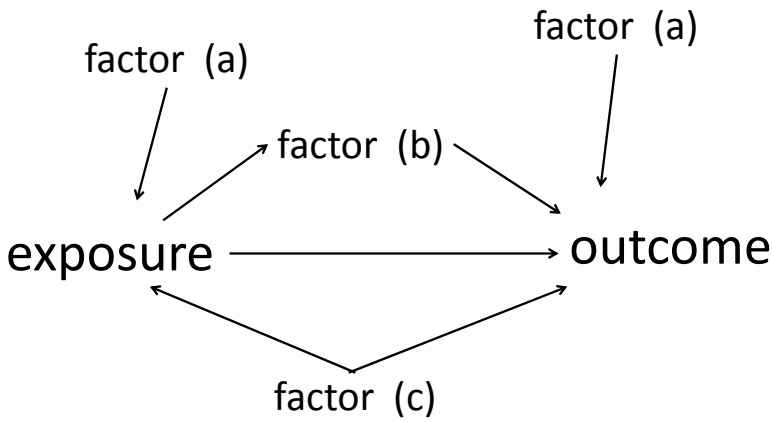


Figure 4.1. Example of a DAG

of a true association. However, sometimes the confounding effect in it-self is of interest, such as e.g. sex and/or body mass index in paper II. In the DAG (figure 4.1) we also have a factor, **(b)**, which is caused by the exposure *(e)*, and is thereafter partly causing the outcome *(d)*. We call this **an intermediar**, as it lie in the causal pathway between (e) and (d). If we wish to study only the isolated association between the exposure and the outcome, not accounting for the causal effect that go through the intermediar, we must adjust for this variable in the analyses. However, if we wish to get an estimate of the full association between the exposure and the outcome, including the part of the association that goes through the intermediar, we must *not* adjust for this variable! There are also variables related to the exposure and the outcome in a manner that makes them important *not* to adjust for. One example is a collider, i.e a variable believed to be caused by the exposure as well as the outcome. Adjustment for a collider risk creating a false association, or wipe out (parts of) a true association.

Causality Although a DAG contains arrows with a distinct direction, epidemiology with its statistical methods, can only show an association, and not causality. This means that we can never prove with epidemiology the direction of an association, and not truly know that there are no forgotten confounders.

4.2 ANALYSIS OF VARIANCE

In paper I and in the supplementary analysis of serum calcium variance in men, in Chapter 8, we wanted to study the association between different lifestyle factors and the level of serum calcium. We compared means between different subgroups of the data by performing one-way ANOVA's complemented with Bonferroni corrected tests.

ANOVA

If we wish to study the variance between two groups, we can compare the means by doing a t-test. However, if there are several groups, we could do an ANOVA (Analysis Of Variance) where we compare the variance in each group with the variance from the sample means. If a group variable is significant in an ANOVA, we may try and figure out more specifically which of the groups that is responsible for the difference. However, performing multiple t-tests will always risk presenting false positive correlations (i.e. type I error). To try and adjust for this one can do a Bonferroni correction.

Bonferroni correction Instead of comparing each p-value to the chosen significance level (α), one can chose to do a Bonferroni adjusted test, i.e. compare the p-value for each null hypothesis to a value corrected for multiple-testing: α/n (where n is the number of tests/hypotheses one has tested for). The reason for doing a Bonferroni correction of the p-values is to avoid finding false positive associations due to chance.[A]

Logistic regression

Even though ANOVA is a simple and robust analysis, it does not take into consideration variance that is correlated to several covariates at the same time. In the analyses of variance of serum calcium (paper I and supplementary analysis in Chapter 8), the association between age and serum calcium was prominent. Many of the other factors in the analysis could very well be covariates to age, why multiple adjustments could be of relevance. In order to adjust for other variables it is possible to perform a logistic regression analysis, with the dependant variable (the exposure) transformed to a dichotomous variable. We therefore transformed the exposure variable (serum calcium) to a dichotomous variable, where an individual with serum calcium above the means was given the value 1 and the other where given the value 0. The odds of having a value above the means were then calculated for each subgroup and compared to a reference group. In the supplementary analysis of variance, chapter 8, we used the group with the lowest value as reference. The choice of reference group in logistic analyses does not alter the association as such, but merely the numerical output. The choice of reference group is therefore simply a reflection of what the authors believe is the best when it comes to interpretation of the odds ratios.

4.3 SURVIVAL ANALYSIS

Cox proportional hazard regression

Cox regression (or Proportional hazard regression) is a survival analysis where we calculate the probability of events per time-at-risk and compare these for different subgroups of an exposure in relation to a reference group.

[A] Bonferroni correct for type I errors, but at the cost of an increase in type II errors. If it is important not to miss potential associations one can do a Bonferroni-Holm test, where you do a, less extreme, step-wise correction of p-values

In paper II we looked at the difference in risk of being diagnosed with prostate cancer in four groups of serum calcium, i.e. we calculated the hazard of cancer for each calcium quartile and estimated the hazard ratios (HR),with the lowest quartile as reference group. The time-at-risk was calculated from baseline examination until date of diagnosis (or death/end of follow-up).

In paper III we estimated the hazard ratios for *incidence of prostate cancer later followed by death* per calcium quartile with time-at-risk measured from baseline examination until date of diagnosis. In paper III we also calculated the hazard ratios for *prostate cancer specific death* per calcium quartile. We did this in two ways; first for the full cohort with time-at-risk measured from baseline examination until date of death (or end of follow-up), and secondly in only those with a previous prostate cancer diagnosis with time-at-risk measured from date of diagnosis until date of death (or end of follow-up). The timescale was time since entry (or time-in-study) for all Cox analyses in this thesis, adjusted for age at entry.

Relative survival

Cause-specific mortality relies on the accurate classification of cause-of-death. In population-based studies of cancer patient survival it is often more relevant to study excess mortality, i.e. the difference between all-cause mortality, and the mortality that would have been expected if the participants had not been diagnosed with cancer.[28] Expected mortality can be estimated from population life-tables, commonly stratified by age, sex, and calendar period. Excess mortality provides a measure of mortality associated with a diagnosis of cancer regardless of whether the excess mortality is directly or indirectly caused by the cancer. The excess mortality can thereafter be modeled using Poisson regression[29] and adjusted for potential confounding factors.

In paper III we wanted to see if there was an association between serum calcium and excess mortality among men diagnosed with prostate cancer. We therefore compared the mortality in different strata's of our cohort, to that seen in the Swedish population at whole. The excess mortality was thereafter modeled using Poisson regression, adjusting for potential confounding factors including age at baseline examination and age at diagnosis.

In paper IV we wanted to examine if the long-term survival in patients diagnosed with Hodgkin's Lymphoma had improved with the changes in treatment routines. We therefore compared the excess mortality in different age groups and different periods of time. By restricting the time-to-event in the Poisson regression model we could study the association both at 1-, 5- and 10-year survival.

4.4 STRATIFIED ANALYSIS

In paper II we could see evidence of a potential interacting effect of body mass index and age on the association between serum calcium and incidence of prostate cancer. We chose to perform a stratified analysis, by dividing the cohort into four subsets of the data and then analyzing each strata with Cox regression analysis. We also adjusted for age at baseline as a continuous variable within each Cox model.

4.5 ANALYSIS OF EFFECT OF MISSINGNESS

In epidemiological studies we are sometimes presented with missing data. Sometimes the amount of data affected by missingness is worryingly large. This is especially prominent in multivariate analyses with missing data in many different variables in several different groups. One could then create a complete data by coding the missingness in the individuals with missing data.

In paper II we chose to explore the effect of the missingness by performing several analyses. First we looked at only subjects with complete information on all covariates. Secondly we modeled the missing values as a separate category, i.e. the missing indicator method. Thirdly, and finally, we wanted to see if imputing extreme values would lead to a difference in association between serum calcium and prostate cancer incidence. The variables with substantial amount of missingness were alcohol status and socioeconomic index (SEI). We imputed extreme values generating four models; Alcohol imputed as low consumer and SEI imputed as manual worker, alcohol as low consumer and SEI as non-manual worker, alcohol as high consumer and SEI as manual worker, and alcohol as high consumer and SEI as non-manual worker. All these models were adjusted for age at entry, smoking status, marital status and BMI with missing as separate categories. Neither of these models differed significantly from each other, why we chose to continue with missing categorized as a separate category.

In paper I we did not perform any analyses sensitive to the value of the variable groups, and thus used the missing indicator method, but in **paper III** we chose not to replace the missing values, i.e. complete case analysis.

5 Summary of Papers

Contents

5.1 Paper I and Variance analysis of serum calcium

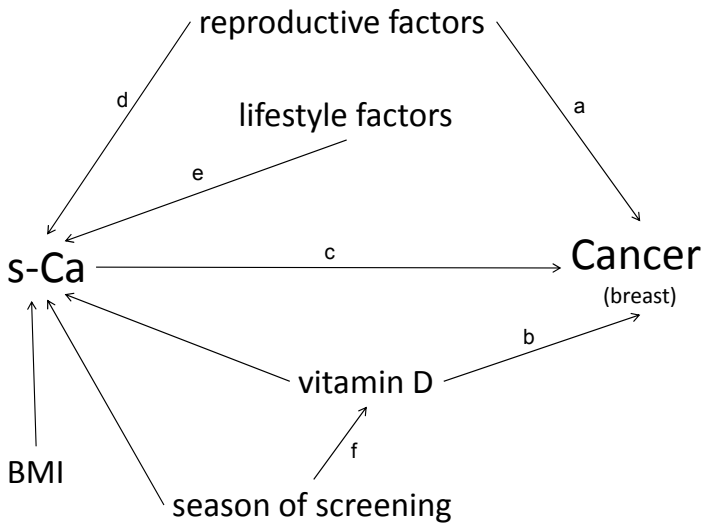


Figure 5.1. DAG of the association between serum calcium and breast cancer

Background

The association between serum calcium (s-Ca) and cancer (of the breast) can schematically be described as in figure 5.1. The association between reproductive factors and risk of breast cancer, **(a)**, has since long been well established, and there was studies showing an association between vitamin D and/or PTH and breast cancer, **(b)**.[30,31] In 2007 Almquist et al showed a possible association between prediagnostic serum calcium and risk of breast cancer among the women in the Malmö Preventive Project, **(c)**.[3] Seeing as both reproductive factors, **(d)**,[32–35] and lifestyle factors, **(e)**,[36] had been associated with serum calcium, we wanted to further explore the potential determinants of serum calcium in women in the Malmö cohort. This study also included screening season, a factor believed to affect vitamin D, **(f)**.[37] We also wanted to study the same covariates (apart form the reproductive factors) among the men of the Malmö Cohort, in a supplemented analysis of variance (see Chapter 8 on page 43). This study also included the additional variables of socioeconomic status and marital status.

Aim

> "The aim of our study was to examine serum calcium concentrations in relation to reproductive history, selected lifestyle factors and screening season in a large population-based cohort study comprising 8,114 women."
>
> Additional aim for the supplementary analysis: "The aim of our study was to examine serum calcium concentrations in relation to selected lifestyle factors and screening season in a large population-based cohort study comprising 16.882 men."

Material

The study population was men and women from the cohort *the Malmö Preventive project*[A]. All participants had their serum levels of calcium measured, as well as height and weight, and the screening season was also registered. They were asked to fill out a questionnaire on life-style factors; such as smoking and alcohol consumption, and reproductive factors; such as hormone-replacement therapy and age of menopausal status (when applicable). Prevalence of cancer was collected from national registries, and so was data on Socioeconomic index as well as marital status for the men. **The outcome** was *serum calcium*[B] level as a continuous variable. **The exposure**, i.e. *the potential co-varying factors* were ; age at screening, body mass index at entry to cohort, smoking status, alcohol consumption, season of screening, and prevalent cancer.[C] In the women we also studied age at menarche, menopausal status, number of children, and use of oral contraceptives or hormone-replacement therapy.[D]. In the men we also studied socio economic index and

[A] For more information, please see page 11
[B] For more information, please see page 4
[C] For more information, please see Section 2.3
[D] For more information, please see page 7

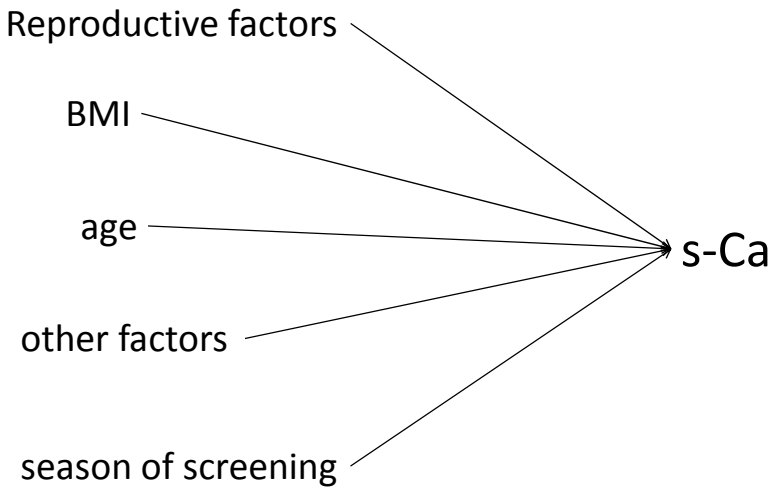


Figure 5.2. DAG for paper I

marital status. Since any of these variables could co-vary with each other, all were potential **confounders** for each individual association.

Methods

Missing values were categorized as a separate category.[E] Differences in mean calcium concentrations between different categories of the studied factors were studied using ANOVA followed by a Bonferroni corrected test.[F] In the preparatory work before the supplementary analysis of serum calcium variance in men (Chapter 8), the mean was found to be notably different for the men analyzed before 1980, and those analyzed after. In order to conduct an ANOVA to compare the variance in the men, with that found in the women in paper I, we chose to conduct the ANOVA analysis only on those analyzed before 1980. Odds of having a serum calcium above means was calculated using logistic regression in an unadjusted model as well as adjusted models.[G]

Results

We found indications of an association between serum calcium and **age**, as well as a strong association between serum calcium and **season** in which the serum sample is drawn. In men there seem to be higher levels of serum calcium for those screened in springtime, and

[E] For more information, please see page 19
[F] For more information, please see page 45
[G] For more information, please see page 46

lower levels for those screened during fall. Among the women however, there seem to be higher levels in those screened both in spring, and in autumn.

Estrogen related factors, such as oral contraceptive use and hormone-replacement therapy was both found to be associated with a lower level of serum calcium. Having given birth to at least one child was also found to be associated with lower level of serum calcium. Age of menarche was not found to be associated with serum calcium.

Body mass index was found to be strongly associated with serum calcium *in women*, with higher levels for those with an "abnormal" BMI. This association was *not* found in the men. Smoking and alcohol consumption was generally not found to be associated with serum calcium, although there were indications of higher levels in former smokers in men, and lower serum levels in those with missing data on alcohol consumption. Previous cancer diagnosis was not found to be associated with serum calcium in our cohort. Socio economic index and marital status was only studied in men, but was not found to be associated with serum calcium.

Discussion and conclusion

In both paper I and the supplementary analysis of variance of serum calcium in men, in Chapter 8, we see an association between age and serum calcium. In paper I we speculate that this association may be partly explained by the difference in hormonal levels in women between ages. This may be the case also in the men, but this result strengthen the theory of age being an independent factor. Body Mass Index, was found to be associated with serum calcium in women, but not the men. This we believe strengthen the theory that the association between Body Mass Index and serum calcium is mediated by estrogen and estrogen related factors such as PTH seeing as the hormonal mechanisms are different between sexes. Smoking status was not found to be associated with serum calcium in women, nor in the men, apart from a possible positive association in those said to be former smokers. Alcohol was found to be associated with serum calcium only in the ANOVA analysis in the women, and for those with missingness in the men. This show that there may be a co-varying factor related to the missingness of alcohol consumption. The estrogen related factors were, as expected, found to be *negatively* associated with serum calcium, although age at menarche proved not to affect the level. Screening season has in other studies been found not to be associated with serum calcium according to studies,[37] but in our studies we did see an association. Strangely enough the association was different between men and women. In women both spring and fall was found to be associated with an increased serum calcium, not following the theory of a sun related factors such as vitamin D. However, in the men, the association differed between seasons with a higher level during spring, and a lower level for those screened during fall. A lowered level of serum calcium did thus correlate in time with an increased vitamin D level.
In conclusion, these studies show that all factors related to hormonal levels could have a potentially confounding effect when studying serum calcium, if related to the outcome.

5.2 PAPER II - SERUM CALCIUM AND INCIDENT OF PROSTATE CANCER

Background

Studies had suggested a possible association between high dietary intake of dairy products and increased risk of prostate cancer (PCa). Giovannucci et al[22] had speculated that the biological pathway of this mechanism included a lowering of the levels of *vitamin D*[H], a hormone believed to prevent the development of cancer.

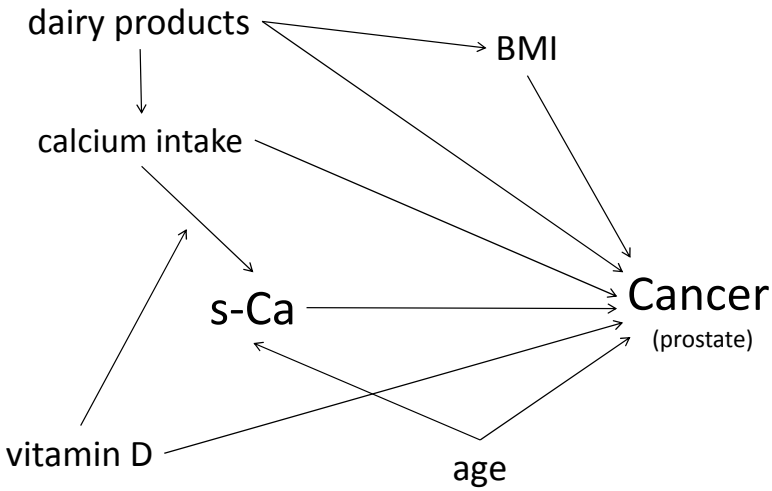


Figure 5.3. DAG of association between serum calcium and prostate cancer

Although many had studied the association between prostate cancer and dietary calcium such as dairy products, as well as its association with vitamin D, none had looked at serum calcium and subsequent incidence of prostate cancer.

Aim

> "Our primary aim was to investigate the incidence of prostate cancer in relation to the prediagnostic serum calcium levels in a prospective study of 22,391 healthy Swedish men. An additional aim was to examine whether this association was modified by age and BMI."

Material

The study population was all men in the cohort textitthe Malmö Preventive Project[I]. The exclusion criterion was prevalence of cancer at entry to cohort. Seeing as the participants

[H] For more information, please see page 4
[I] For more information, please see page 11

were quite young at entry to cohort, the prostate cancer in general debuted years after baseline examination. Serum calcium had therefore been analyzed years prior most diagnosis since all participants were healthy at the baseline examination in the 70ies.

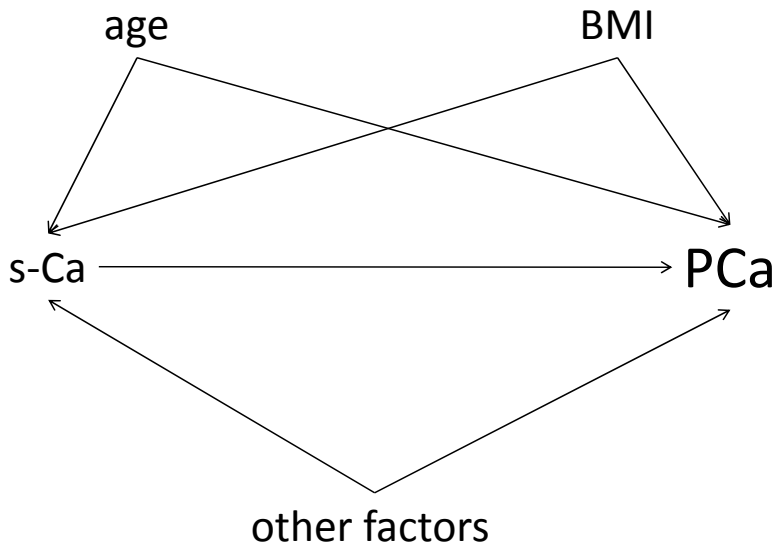


Figure 5.4. DAG for paper II

The outcome was incidence of *prostate cancer*,[J] i.e. cases reported to the national cancer registry and/or to the local tumor register at Oncologic Center during follow-up. **The exposure** was *serum calcium*[K] drawn and analyzed at the baseline examination at entry to the cohort (beginning in the 70-ies) and then grouped into serum calcium *quartiles*.[L] **The confounders** were age and Body Mass Index (BMI) at entry to cohort. There was also information on lifestyle factors collected through a questionnaire around the time of entry to the cohort. The variables *smoking status, marital status, socioeconomic status*, and *alcohol consumption* had been considered by some studies as potential confounders, and were therefore included in the fully adjusted analyses.[M]

Methods

In order to estimate the incidence of prostate cancer we did *Cox proportional hazard regression*.[N] The timescale was time since entry to cohort (in days), but adjusted for age as a

[J] For more information, please see page 5
[K] For more information, please see page 3
[L] For more information, please see page 12
[M] For more information, please see section 2.3
[N] For more information, please see page 17

continuous variable. We also had a fully adjusted model, adjusting for BMI and aforementioned lifestyle factors.

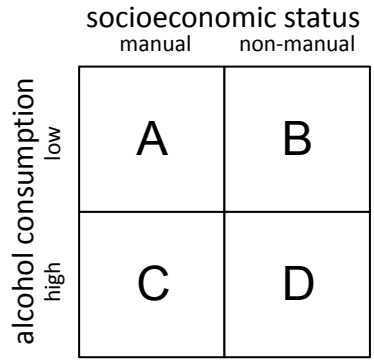


Figure 5.5. Diagram illustrating the four statistical models included in the sensitivity analysis

Missingness A number of study participants had a substantial amount of missing data, especially for the variables socioeconomic index (n=1,137) and alcohol consumption (n=2,132). In order to estimate the potential impact of this missingness we first did a standard multiple adjusted model including only those with complete information on all covariates. Given the large amount of lost data why we wished to categorize the missing values, why we did an analysis using the missing indicator method. It is generally recommended not to classify missing as a separate category when the reason for the missingness is unknown. We therefore decided to do a sensitivity analysis where we imputed extreme values.[O] This third analysis also included all subject, but here we chose to impute extreme values instead of the missing in four different ways as seen in figure.5.5. This comparison showed no difference in association between the models, why we chose to treat the missing data as a separate category in the other analyses in this paper.

Interaction In order to study the three-way interaction between age, body mass index and serum calcium on the association with prostate cancer risk we performed a stratified analysis,[P] dividing the cohort into four subcohorts as seen in figure 5.6 and adjusting for age at entry as a continuous variable within each Cox model.

Results

We did not see any indications of an association in general, but in the stratified analysis there was a small association between serum calcium and risk of prostate cancer for those

[O] For more information, please see section 4.5
[P] For more information, please see page 19

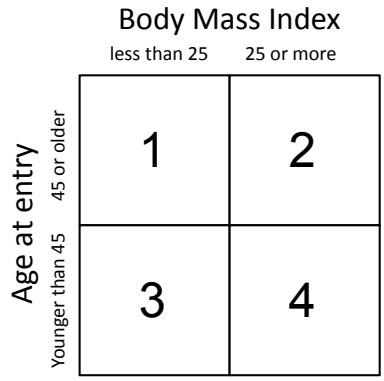


Figure 5.6. Diagram of the four models in the stratified analysis of the modifying effect of bmi and age

with a BMI above 25 and younger than 45 years (Section B in figure 5.6). In this stratum we found that those in the highest quartile of serum calcium had a hazard ratio of 0.63 (with the lowest serum calcium quartile as reference) with a confidence interval between 0.40 and 0.99. The difference between the quartiles was not very prominent, nor were there any indications of a trend (e.g. lower hazard ratio for each increase in serum calcium quartile).

Conclusion

This study does not support the hypothesis that high serum calcium level is a risk factor for prostate cancer. On the contrary, the data imply that high prediagnostic serum levels of calcium in young overweight men may be a marker for a *decreased* risk of developing prostate cancer. These results do not proclaim that there truly is an association, but rather that the positive association found in other studies is not seen here, seeing as the association in our stratum of "young" and "overweight" is *negative* with increasing serum calcium. Another conclusion is that the potential three-way interaction effect between age, bmi and serum calcium with prostate cancer risk may indicate that there is an unaccounted confounder related to these, such as vitamin D, hormonal level, adiposity or similar.

5.3 Paper III - Prostate cancer mortality . . .

Background

After submitting paper II, Skinner et al[24] published a paper on serum calciums association with incidence of prostate cancer, but also looking at incidence of *fatal* prostate cancer. They saw that the association between serum calcium and fatal prostate cancer was stronger than that between serum calcium and incident prostate cancer. They believed this supported the theory that there was a true association, merely diluted by indolent tumors in the analyses not adjusted for the severity of prostate cancer.

Aim

> "The purpose of this study was to further study our cohort of 22,390 healthy Swedish men to determine if serum calcium is a risk factor for fatal prostate cancer or associated with poorer survival among men diagnosed with prostate cancer."

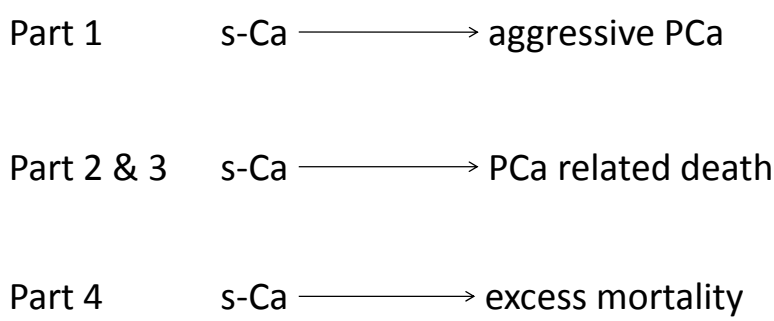


Figure 5.7. Schematics of the associations in paper III

Material

The study population was based on the men in the cohort *the Malmö Preventive Project*.[Q] In two of our four analyses our study population was all men in the MPP cohort (excluding those with prevalent cancer). The other two analyses looked only at those men diagnosed with prostate cancer during the 35 years of follow-up.

In part 1 of the paper **the outcome** was *incidence of fatal prostate cancer*, measured as prostate cancer diagnoses that later was followed by prostate cancer related death. In part 2 and 3 the outcome was *prostate cancer-specific death*. In part 4 the outcome was overall mortality measured as *excess mortality*. That is, mortality in comparison to the

[Q] For more information, please see page 11

Swedish population in general. **The exposure** was *serum calcium quartile*, where the quartile definitions for part 1 and 2 were for the full cohort, and the definitions for part 3 and 4 were for only the prostate cancer cases. Potential **confounders**[R] were *age at baseline examination, body mass index, smoking status*, and *socioeconomic index*. Seeing as we had date of death as end point (and not date of diagnosis) in our excess mortality analysis, we were able to adjust for the possible confounding effect of *age at diagnosis*. It turned out that age of diagnosis was a strong confounder in the excess mortality analysis. Analyses in part 3 and 4 were therefore, in addition to the variables in the models in part 1 and 2, also adjusted for age at diagnosis, apart from the addition of *period of diagnosis*.

Methods

Skinner et al[24,25] chose to study incidence of fatal prostate cancer, where an event was defined as the diagnosis of prostate cancer, given that the subject later (or at the same time) passed away (with prostate cancer specified as a cause of death). That is, the person time was calculated from entry to cohort, until diagnosis of "death-causing"-cancer.

In our analyses we chose to study the mortality in four different ways:

Part 1; incidence of prostate cancer later followed by cause-specific death in the full cohort, i.e. hazard ratios for incidence of fatal prostate cancer (as Skinner et al did) through Cox regression of association with serum calcium quartiles, in the full cohort, with time-at-risk calculated from baseline examination until date of diagnosis (or death/end of follow-up).

Part 2; cause specific mortality in the full cohort, i.e. hazard ratios for prostate cancer-specific death through Cox regression of association with serum calcium quartiles, in the full cohort, with time-at-risk calculated from baseline examination until date of death (or end of follow-up).

Part 3; cause specific mortality in those with a prostate cancer diagnosis, i.e. hazard ratios for prostate cancer-specific death through Cox regression of association with serum calcium quartiles, for those with prostate cancer diagnosis, with time-at-risk calculated from date of prostate cancer diagnosis until date of death (or end of follow-up).

Part 4; relative survival for those with prostate cancer diagnosis, i.e. Excess Mortality Ratios through Poisson regression analysis of the excess mortality in different stratas of those with prostate cancer diagnosis. Time-at-risk was calculated from date of prostate cancer diagnosis until date of death (or end of follow-up).

Analyses 1 and 2 were adjusted for age at baseline, BMI, smoking, and SEI. Analyses 3 and 4 were also adjusted for age and period of diagnosis.

Results

We found no evidence of an association between prediagnostic serum levels of calcium and mortality due to prostate cancer in either of our analyses; **1**, fatal cancer incidence rate ratio 0.85, 95% CI: 0.56-1.31, **2**, cause-specific mortality rate ratio (full cohort) - 1.02, 95% CI: 0.73-1.42, **3**, cause-specific mortality rate ratio (among men with prostate cancer) - 1.04,

[R] For more information, please see Section 2.3

95% CI: 0.74-1.46, and **4**, excess mortality rate ratio (among men with prostate cancer) - 0.98, 95% CI: 0.61-1.60, Highest vs lowest quartile.

Conclusion

In this study we did not find any evidence of prediagnostic serum calcium being a risk factor for aggressive prostate cancer. We did not find any association with excess mortality, cause-specific mortality, nor incidence of fatal prostate cancer. This strengthens our previous conclusion that possibly serum calcium in itself is not a risk factor for prostate cancer.

5.4 Paper IV – Long-term survival in Hodgkin's Lymphoma

Background

Treatment of Hodgkin lymphoma (HL) has over the last 40 years improved patient survival considerably, and HL is now considered highly curable. However, seeing as the cancer requires a systemic therapeutic treatment there is a high risk of secondary disease caused by the treatment itself, mainly cardiovascular disease. This tends to affect the long-term survival negatively, making it difficult to assess and evaluate any beneficial effects of changes in treatment routines.

Aim

> "Our aim was to assess trends in patient survival and long-term [and short-term] excess mortality among all patients, regardless of clinical trial enrolment, during this 33-year period starting when curative treatment principles were well established."

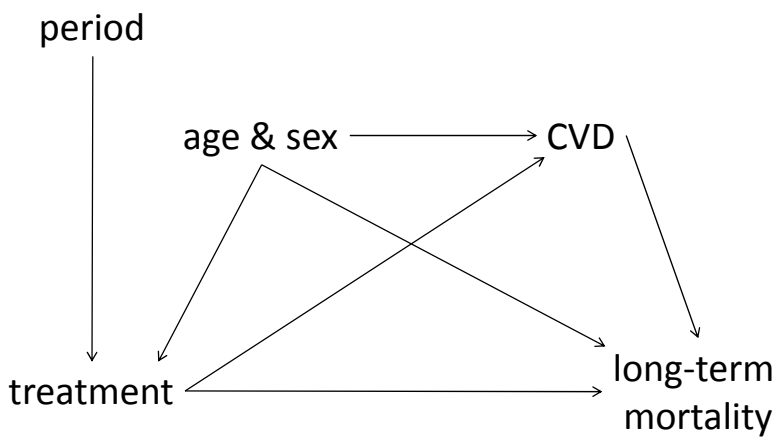


Figure 5.8. DAG of the paper IV

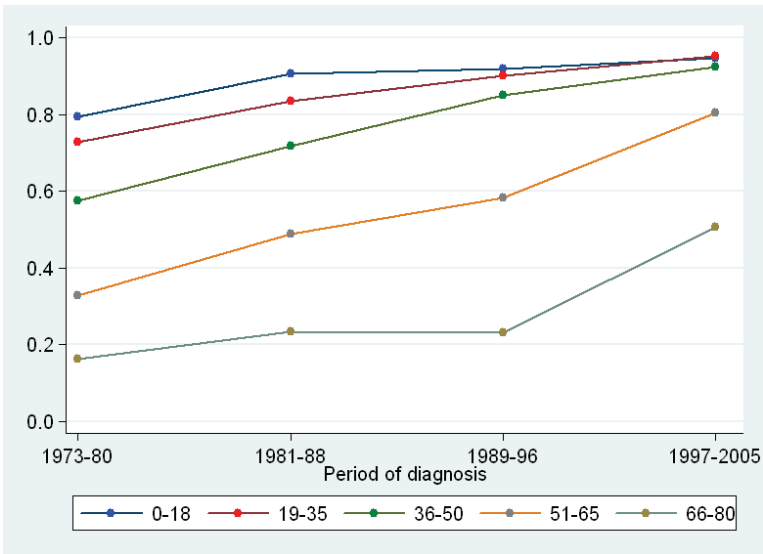
Material

The study population was all those diagnosed with Hodgkin's' Lymphoma in Sweden between 1973 and 2005, i.e. 6,136 men and women. We have thereafter compared the survival in those diagnosed during four different time-periods, between which the treatment routines had progressed in order to reduce treatment-related morbidity. **The outcome** was overall mortality, as seen in figure 5.8, measured in *excess mortality* compared to the

Swedish population in general. **The exposure** was treatment routine measured as *period of diagnosis*. Potential **confounders** were age and sex.

Methods

We assessed relative survival analyses, i.e. calculated Excess Mortality Ratios through Poisson regression analysis of the excess mortality in different strata's of age and period of diagnosis. We thus compared the excess mortality in different age groups and different periods of time (by sex). By restricting the time-to-event in the Poisson regression model we could study the association both at 1-, 5- and 10-year survival, i.e. time-at-risk was calculated from date of Hodgkin's Lymphoma diagnosis and for 1, 5 and 10 years following this.



	Calender Period			
	1973-1980	1981-1988	1989-1996	1997-2005
Age category	RSR 95% CI	RSR 95% CI	RSR 95% CI	RSR 95% CI
0-18	0.79 (0.70 to 0.86)	0.91 (0.84 to 0.95)	0.92 (0.86 to 0.95)	0.95 (0.90 to 0.97)
19-35	0.73 (0.68 to 0.77)	0.83 (0.79 to 0.87)	0.90 (0.87 to 0.93)	0.95 (0.92 to 0.97)
36-50	0.58 (0.51 to 0.63)	0.72 (0.65 to 0.77)	0.85 (0.79 to 0.89)	0.92 (0.85 to 0.97)
51-65	0.33 (0.28 to 0.38)	0.49 (0.42 to 0.56)	0.58 (0.50 to 0.66)	0.80 (0.72 to 0.87)
66-80	0.16 (0.12 to 0.21)	0.23 (0.18 to 0.30)	0.23 (0.17 to 0.31)	0.51 (0.28 to 0.75)

Figure 5.9. One of the findings in paper IV: 10-year survival in Hodgkin's Lymphoma

Results

Relative survival had improved for all age groups, especially for those aged 51-65 years (From RSR 0.58 to 0.80 in the last period, as seen in figure 5.9). The groups aged 0-18, 19-35 had very little excess mortality after 5-years, and also the group aged 36-50 had in the last period reached the high levels of 5-year survival as the younger. Though the 10-year survival had improved greatly, survival is still low in the older age groups.

Conclusion

We conclude that the recent improvements in treatment strategies in Hodgkin's Lymphoma has considerably improved survival in all ages. However, age is still an important factor indicating the need of further progress in diagnosis and treatment of especially older patients.

6 Discussion

Contents

6.1 Calcium and prostate cancer

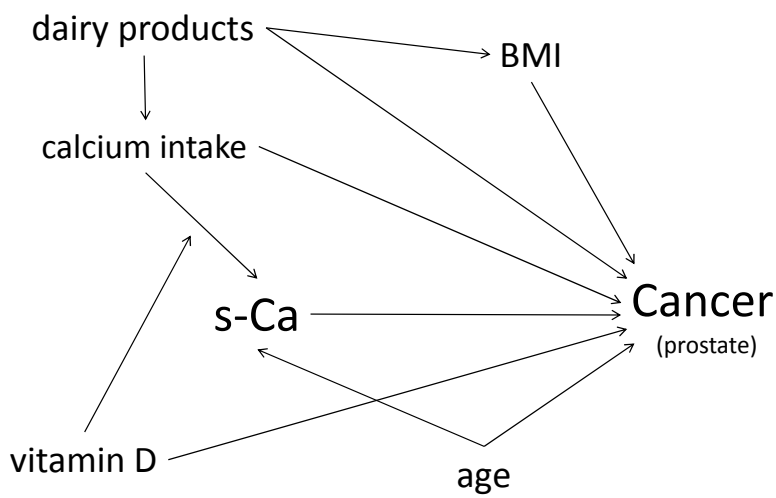


Figure 6.1. Simplified DAG of the association between serum calcium (s-Ca) and Cancer (of the prostate)

The main purpose of this thesis was to study whether serum calcium is involved in the etiology of cancer of the prostate. Many studies had suggest that a high intake of dietary calcium, such as dairy products, could increase the risk of prostate cancer.[16,18–21] In 1998

Giovannucci et al[22] presented the hypothesis that a high intake of calcium causes cancer by forcing the body to lower the levels of the tumor suppressing hormone vitamin D_3 in order to prevent hypercalcemia. This could very well be true, seeing as there e.g. is a difference in prostate cancer incidence and survival between African Americans and White Americans[38]. Studies had also shown that not only colon and skin, but also breast and prostate tissue have the enzymatic machinery to produce the active form of Vitamin D_3, also known as calcitriol ($1,25(OH)_2D$).[39]

In paper II we wished to explore the association between calcium and cancer, by studying prediagnostic serum calcium and the association with prostate cancer. Seeing as serum calcium is believed to be tightly regulated it did not have an apparent active role in the proposed pathway between dietary calcium and cancer. However, by looking at a very large cohort, we believed we would be able to detect even small variations in serum calcium between those who developed prostate cancer, and those who didn't. Or more precise, we believed we could see a difference in prostate cancer incidence between groups of different serum calcium level.

We didn't. Well, actually we could see a small association in a subgroup of young and overweight men, but this association was *negative*, i.e. showing a possible *protective* effect of high serum calcium, i.e. the opposite of what was anticipated. This association was weak, and does not imply that there is a *true negative* association. It does however suggest that, seeing as the positive association found in other studies is not seen here, the association between dietary calcium intake and prostate cancer is less likely to act through an effect of a measurable increase in the amount of serum calcium, but rather through other factors associated with diet, such as e.g. other components in dairy products.[40]

Serum calcium could be a marker for the dietary intake of calcium, and it is possible that the source of calcium in the diet of Swedish men differ from other study populations in regard to e.g. consumption of sour cream and milk. In 2010 Newmark et al proposed that the association between dairy products and prostate cancer is not due to the calcium, but instead a result of the high content of phosphate.[41] This does well correlate with our results, that serum calcium in itself is not associated with prostate cancer. It could instead act as a possible marker for e.g. high dietary intake of phosphate through dairy products.

However, during the time when our first paper on prostate cancer and calcium was about to be published, another group of researchers studied the same association (serum calcium and prostate cancer) in an American cohort.[24] They, as opposed to us, did find a positive association in the full cohort. They also showed that this association may be more prominent in those with an aggressive prostate cancer.[24,25]

In the 70-ies researchers discovered that at presence of prostate infection and prostate cancer, one could see elevated serum levels of a protein later denoted PSA, prostate-specific antigen.[42] PSA is (amongst other things) an agent in semen, optimizing fertilization. Serum analysis of level of PSA is routinely used as an aid in early detection of prostate abnormalities such as prostate cancer. Although PSA is not part of any national screening program in Sweden, it is offered free of charge in national health care for anyone

requesting it. However, though sensitive for prostate cancer, it is not significant, why not all of the men with elevated PSA have prostate cancer. There are also many cases where men with PSA elevation at biopsy are found to have prostate cancer, but at such an early stage, or even pre-stage cancer, that the risk of them ever developing severe cancer is highly unlikely. Skinner et al suggested that PSA screening risked introducing misclassification by increasing the diagnoses of indolent cancer, and thus risk diluting a true association, which would explain the significant association they found between serum calcium and incidence of fatal prostate cancer.[24,25]

In paper III we therefore wanted to explore the possibility of an association between serum calcium and *aggressive* prostate cancer. We did this both by looking at the association between serum calcium and incidence of *fatal prostate cancer*, but also looking at the association between serum calcium and the risk of *prostate cancer specific death*. We also figured that if there was an association between serum calcium and the aggressiveness of a tumor, we should be able to, in the men who developed prostate cancer, see a difference in *over-all excess mortality* between groups with different level of serum calcium.

We didn't. Not even in our body-mass-index-and-age-stratified analyses could we see any signs of an association, neither positive nor negative. This is suggesting that either the mediating effect of body mass index and age on the association between serum calcium and prostate cancer was a finding by chance, or possibly prediagnostic serum calcium does not predict prognosis in prostate cancer, but merely the incidence of diagnosis of indolent tumors.

One of the critiques against our first prostate cancer study (paper II) was that we looked at total calcium, and not ionized calcium. We therefore calculated the amount of free calcium based on the level of albumin. Albumin is the protein that is believed to bind a substantial amount of calcium in the blood. By estimating the amount of albumin-bound calcium and subtracting that from the amount of total calcium, it is believed we get a fair estimate of the level of free calcium in the blood.[43] But nor these analyses did show an association with prostate cancer mortality.

We also considered the possibility that, though we specifically looked at mortality (as cause specific mortality and excess mortality among prostate cancer cases) our results could be diluted by indolent cancers diagnosed through Prostate Specific Antigen screening.[44] Seeing as we did not have information on why and how the cases were diagnosed with prostate cancer, we separated our cohort in two subsets, one for those analyzed before 2000, and one for those analyzed after. PSA analysis is not in any screening program in Sweden, but was introduced in the late 90-ies for anyone requesting it free of charge. We did not see any difference in association between those diagnosed pre and post the introduction of PSA test in Sweden.

Paper I explored the association between potential confounders and serum calcium in women. In a supplementary study of serum calcium variance, in Chapter 8, I looked at the association between potential confounders and serum calcium but *in men*, in case there was

a difference in association between sexes. The results from this previously unpublished exploratory study can be found in Chapter 8 on page 43, and include similar statistical analyses as in paper I. In cases where the determinant variable is a continuous variable one may prefer a general linear regression analysis. Just as in logistic regression it is easy to include other covariates in the model as opposed to e.g. the ANOVA. Another benefit is that in linear regression the resulting estimates are in the same unit as the determinant variable and therefore sometimes more clinically interpretable. We therefore also performed a linear regression analysis of serum calcium in men as a complement for the supplementary analysis of variance (data not shown). This analysis confirm the same associations as the logistic regression analysis and the Bonberroni corrected tests.

In paper I we saw a strong effect of screening season on serum calcium. We did not adjust for screening season in paper II and paper III simply because the season of prediagnostic screening is not likely to predict neither the incidence of nor the survival in prostate cancer, and thus not present a confounding effect on the association with serum calcium. However, if the seasonal effect means that we did not estimate the serum calcium correctly due to large variation it could introduce a misclassification bias by poor estimates of the serum calcium quartile. A simple adjustment for screening season in the model, could resolve parts of such a problem. However, if a true association was undetectable due to vast variation, a screening season-stratified quartilation, much like what was applied when dealing with the instrumental variation, could be enough to reveal a true association in a future analysis.

In the "trend analysis" in paper II we modeled the serum calcium quartile variable as a categorical, with the values 1 through 4 imputed. This was suboptimal, since it's always better if the difference between the values of two groups (the increment) better represent the proportional change from one group to another. However, seeing as the quartilations were done in two steps, imputing the mean value would have been misguiding. As mentioned, we did not find a trend or any indication of an association in these analysis.

In paper II and III a potential confounder is age. Serum calcium decrease with age, and since the incidence of prostate cancer increase with age those entering the cohort at an older age (and who therefore have a higher prediagnostic level of serum calcium) are more likely to become cases within the follow-up of 35 years, than those entering the study at a very young age. Although the effect of screening age on the incidence of prostate cancer decreases to a minimum as the cohort progress, this could lead to a false positive association between serum calcium and prostate cancer if not adjusted for. When looking at the survival, we also have age of diagnosis as a potential confounder. In the analyses where incidence of prostate cancer diagnosis (followed by cause-specific death) is the outcome, we cannot adjust for this variable. However, we did not see an association in the analyses where we could adjust for this.

In order to be able to compare our results with those by Skinner and colleagues, we did a cox analysis looking at diagnosis of prostate cancer later followed by death caused by cancer. In the article we state:

> "Men stopped being at risk on 31 December 1998 but we continued to follow the men until 31 December 2008 in order to determine whether incident cases of prostate cancer were 'fatal' and avoid possible misclassification of fatal cases as non-fatal cases due to lack of follow-up. If we did not make this adjustment then men diagnosed with prostate cancer during later years (e.g., 2008) would be less likely to be classified as cases of fatal prostate cancer since they would have a considerable shorter time at risk in which the cancer could cause their death"

However, an unadjusted analysis would not lead to a misclassification of a case, because, in the analysis we compare the number of *cases* (i.e. PCa diagnoses later followed by PCa-specific death) with *non-cases*. That is, those diagnosed with prostate cancer who does not die from their cancer before end of follow-up, are considered non-cases regardless of whether or not we have a cut-off in 1998. This is because we do not make any difference between a nonfatal prostate cancer and a cohort participant not diagnosed with cancer. This adjustment does however make us lose some true cases of fatal prostate cancer, that is those diagnosed and diseased between 1998 and 2008. So why do we make this adjustment? Well, there is a difference in person-time. We do not want yet not manifested true cases to contribute person-time to the group of non-cases. This is why we chose to not account for *any* person-time after 1998.

In 1996 Leifsson et al found a positive association between serum calcium and excess mortality in the general population.[45] They found that most part of this excess mortality in those with higher serum calcium was attributable to cardiovascular disease, but almost 30 percent was believed to be attributed to other malignant disorders such as cancer. This could strengthen the belief that serum calcium is positively associated with the severity of prostate cancer. In paper III we compared the excess mortality between serum calcium quartiles in men diagnosed with prostate cancer in order to assess if serum calcium is associated with the prognosis in prostate cancer. If there is a general association between serum calcium and mortality in our population we would not be able to adjust for this in our relative survival analysis, and thus these analysis would be invalid. However, Leifsson et al found serum calcium to be *positively* associated with mortality,[45] and Skinner et al also found it to be *positively* associated, but with fatality in prostate cancer.[25] If either or both these *positive* associations existed in our cohort, we should have at least detected an association. This could mean that in our data, there *is* a positive association with prostate cancer, but hidden by a *negative* association with mortality in general. Now, that is on the contrary to what Leifsson et al found in their aforementioned study. There could also be a true *negative* association with prostate cancer in our cohort, similar to what we found in the subset in paper II, but thus hidden by a positive association with overall mortality in the general population, in accordance with the results of Leifsson et al. However, if so, we should have been able to see a difference in the analysis of incidence of fatal prostate cancer in the full cohort.

A future study of interest would be to either study the association between serum cal-

cium and over-all mortality in the full cohort, or even in those free from cancer. If such an analysis show an association in our cohort it would be interesting to study the difference in mortality in those diagnosed with prostate cancer compared to those in the full cohort, and see if this excess mortality differ between groups of serum calcium.

One explanatory hypothesis of an association between calcium and cancer involves hormones. In paper I we could see that most estrogen related factors were related to serum calcium in a sense that high estrogen levels seem to be associated with lower levels of serum calcium. However, a high Body Mass Index was related to higher level of serum calcium, though fat tissue is known to produce estrogen as well as leptin.[46] This effect is however known to be more distinct after menopause. The association with body mass index was not seen in the men analyzed in the supplementary analysis of variance but estrogens have been found to be associated also with prostate cancer.[40] We could also see that there was a difference between men and women in the association of serum calcium and season of screening. It is not farfetched that the hormonal differences between sexes are mediating in the association between serum calcium and season and body mass index. Hormonal levels could also explain the results in paper II where we found a potential three-way interaction effect between age, BMI and serum calcium with prostate cancer. Both season of screening, vitamin D level, body mass index, fat mass/adiposity, insulin levels and skin color are factors correlated with each other, and many of these factors have been associated with estrogen level, PTH level, serum calcium and risk of cancer.[47,48] Further studies of these associations in general, and their association with prostate cancer risk specifically, could elucidate the potential association between cancer and calcium.

6.2 Temporal trends and survival in Hodgkin's Lymphoma

This thesis also includes Paper IV, where we study the difference in short-term and long-term survival in Hodgkin's Lymphoma between different time periods. The reason for this was to see if the recent changes in treatment strategies of Hodgkin's has decreased the amount of treatment-related morbidity and mortality. It is difficult to assess progress in treatment of a disease where many get cured, but die from side effects from the treatment itself. One can look at cause-specific mortality, as we did in paper III when the outcome was prostate cancer specific mortality, but that requires that the death registry has a reliable interpretation of cause-of-death. When it comes to Hodgkin's Lymphoma the disease-related mortality include causes such as *cardiovascular disease*, *primary cancer*, or *infection*. A correct interpretation of cause-specific death would therefore be difficult to make, which is why we chose to study excess mortality instead. By studying the difference in excess mortality between periods of time we could get an estimate of the temporal trends. Here we could see that the changes in routines seem to have made a difference in long-term survival for all age-groups, even though the excess mortality is still quite high especially for those diagnosed at a higher age.

7 Conclusion

One specific aim of this thesis was to evaluate effectiveness of treatment through studying temporal trends in survival of patients diagnosed with Hodgkin's lymphoma.
We conclude that the recent advances in treatment in, and diagnosis of, patients with Hodgkin's Lymphoma has considerably improved short-term, as well as long-term survival in all ages. However, age is still an important factor indicating that extra care and caution should be put in the decision of current and future treatment strategies in and clinical care of older patients.

The key aim of this thesis was to study whether serum calcium is involved in the etiology of cancers of the prostate, and if there is an association between pre-diagnostic levels of serum calcium and survival of men diagnosed with prostate cancer.
In our studies we did not find serum calcium to be associated with incidence of prostate cancer, incidence of fatal prostate cancer, prostate cancer-specific mortality, nor relative survival in prostate cancer.
We did however find a small but significant association with incidence of prostate cancer in a stratified analysis of those men with a high body mass index, and entering the cohort at young age. This together with the serum calcium's strong correlation with season of screening and estrogen related factors, makes it possible that the association between dietary calcium and cancer could partly be explained by a mediating, if not causative, effect involving the mechanisms of calcium homeostasis such as; sun exposure, vitamin D level, level of sex hormones, body constitution or insulin level.
Further studies of this mechanism in general, and its association with prostate cancer risk specifically, would be of interest when exploring the association between cancer and calcium.

8 Supplementary analysis of variance

Determinants for Serum Calcium concentrations in men

(unpublished original work)

BACKGROUND: Several studies have examined the possible association between pre-diagnostic serum calcium and the risk of cancer, such as breast cancer and prostate cancer. The causal pathway is unclear and many factors are believed to affect both serum calcium as well as cancer risk.

OUTLINE: In order to explore potential confounders' individual association with calcium, we analysed the difference in serum calcium levels between different groups in a subset of 16,882 healthy Swedish men from the Malmö Preventive Project, started in the 70-ies.

METHODS & MATERIAL: All participants had their serum levels of calcium measured, as well as height and weight. They were also asked to fill out a questionnaire on life-style factors such as smoking and alcohol consumption. Socioeconomic data as well as marital status and prevalence of cancer was collected from national registries after given consent. Variance in serum calcium was analysed in a one-way ANOVA complemented with Bonferroni corrected tests. Odds of having a serum calcium above means was calculated using logistic regression in an unadjusted model as well as adjusted models.

RESULTS: We found a significant difference in serum levels of calcium between groups examined at different ages (ANOVA p=0.000). Bonferroni corrected tests showed significant difference in those examined at ages 38-44 (p=0.001) compared to means of those younger than 38. However, Odds Ratio of having serum calcium level above mean was 1.11 (95% CI: 1.01-1.22, p=0.035, crude model) for those aged 48 or older (comp. to younger than 38), and this increased further when fully adjusted (1.42, 95% CI: 1.25-1.62, p=0.000). We found a significant difference in means in groups screened in different seasons (ANOVA p=0.000), and Bonferroni corrected tests showed significant p-values for those examined in spring and fall (winter as reference). Logistic regression showed significantly higher Odds Ratio for those examined in spring compared to those examined in winter (OR 1.22, 95% CI: 1.12-1.33, p=0.000, fully adj.). The corresponding Odds Ratios for individuals examined in fall (comp. to winter) was (0.79, 95% CI: 0.72-0.86,

p=0.000). We did not find evidence of an association with Body Mass Index, Prevalent cancer, Socio Economic Index nor Marriage status. We found a small but significant association with serum calcium in smokers (ANOVA p=0.018), and Odds Ratio of 1.07 for former smokers compared to never-smokers (95% CI: 1.01-1.17, p=0.023, fully adjusted model). However, Bonferroni corrected p-value for this group was not significant (p=0.173, ref. never-smokers). We found a significant Odds ratio for the alcohol consumption strata for "missing data" compared to non-drinkers (OR 0.66, 95% CI: 0.57-0.75, fully adjusted model). However, the one-way analysis of variance showed a low but not significant p-value for ANOVA analysis of alcohol consumption (p=0.070) as well as Bonferroni corrected p-value for those missing data (p=0.058, ref non-drinkers).

CONCLUSION: We found indications of an association between serum calcium and age, as well as a strong association between serum calcium and season in which the serum sample is drawn. Alcohol consumption, Body Mass Index, Smoking status, Marriage status and Socio Economic Index was not found to be associated with serum calcium in our data.

Table 8.1. Mean serum calcium levels in relation to life-style factors, and season

Variable	No.	Mean	95% CI	Bonferroni t-test (p-value)	ANOVA (p-value)
Age at screening					
<38	3053	2.418	2.329-2.508	(ref)	
38-44	3558	2.410	2.307-2.512	0.001	
45-47	6664	2.414	2.329-2.500	0.280	
48 or old	3608	2.419	2.336-2.502	1.000	0.000
Body Mass Index					
less than	892	2.413	2.322-2.504	(ref)	
20-24	8954	2.415	2.325-2.506	1.000	
25-29	5997	2.415	2.327-2.503	1.000	
30 or mor	1036	2.416	2.327-2.504	1.000	0.850
Smoking status					
never	5363	2.414	2.324-2.504	(ref)	
ex	8583	2.417	2.326-2.508	0.173	
current	2937	2.412	2.329-2.495	0.978	0.018
Alcohol consumption					
Low	6775	2.417	2.329-2.504	(ref)	
Medium	6632	2.415	2.325-2.506	1.000	
High	1344	2.414	2.312-2.515	1.000	
missing	2132	2.411	2.327-2.494	0.058	0.070
Prevalent cancer					
No	6755	2.415	2.326-2.504	(ref)	
Yes	128	2.425	2.312-2.539	0.191	0.191
Screening season					
winter	6082	2.417	2.330-2.504	(ref)	
spring	2964	2.429	2.340-2.519	0.000	
summer	4203	2.415	2.326-2.504	1.000	
fall	3634	2.401	2.310-2.493	0.000	0.000
Socio Economic Index					
Manual wo	6923	2.416	2.328-2.505	(ref)	
Non-manua	6931	2.414	2.326-2.502	1.000	
Self empl	1476	2.412	2.318-2.506	0.928	
Other	666	2.420	2.321-2.519	1.000	
missing	887	2.415	2.323-2.507	1.000	0.206
Marriage status					
unmarried	2329	2.416	2.321-2.511	(ref)	
married	2458	2.415	2.328-2.503	1.000	
divorced	1921	2.413	2.320-2.505	1.000	
widowed	106	2.399	2.300-2.498	0.528	
missing	69	2.435	2.346-2.524	0.858	0.079

Table 8.2. Crude and adjusted odds ratios (OR) for high (>2.415 mmol/l) vs. low (<=2.415 mmol/L) serum calcium levels from multiple regression analysis, in relation to life-style factors, and season of baseline examination

Variable	Crude			Age adjusted			Fully adjusted[a]		
	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Screening age									
younger than 38	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
38-44	0.99	0.90-1.09	0.891	0.99	0.90-1.09	0.891	1.03	0.93-1.14	0.529
45-47	0.98	0.90-1.07	0.685	0.98	0.90-1.07	0.685	0.99	0.91-1.09	0.881
48 or older	1.11	1.01-1.22	0.035	1.11	1.01-1.22	0.035	1.42	1.25-1.62	0.000
Body Mass Index									
less than 20	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
20-24	1.03	0.90-1.19	0.662	1.03	0.90-1.18	0.666	1.06	0.92-1.22	0.424
25-29	1.07	0.93-1.23	0.341	1.07	0.93-1.23	0.354	1.11	0.96-1.28	0.168
30 or more	1.01	0.84-1.21	0.943	1.00	0.83-1.20	0.993	1.03	0.86-1.24	0.727
Smoking status									
never	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
former	1.08	1.01-1.16	0.022	1.08	1.01-1.15	0.036	1.09	1.01-1.17	0.023
current	0.97	0.88-1.06	0.482	0.96	0.87-1.05	0.367	0.96	0.87-1.05	0.373
Alcohol consumption									
low	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
medium	1.01	0.94-1.08	0.81	1.02	0.95-1.09	0.613	1.02	0.95-1.09	0.596
high	0.95	0.84-1.07	0.397	0.96	0.86-1.09	0.545	0.96	0.85-1.08	0.480
missing	0.92	0.84-1.02	0.119	0.71	0.62-0.81	0.000	0.66	0.57-0.75	0.000
Prevalent cancer									
no	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
yes	1.08	0.76-1.52	0.682	1.06	0.75-1.51	0.728	1.08	0.76-1.53	0.672
Screening season									
winter	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
spring	1.25	1.14-1.36	0.000	1.24	1.13-1.35	0.000	1.22	1.12-1.33	0.000
summer	0.96	0.89-1.04	0.363	0.96	0.89-1.04	0.337	0.92	0.84-0.99	0.037
fall	0.83	0.76-0.90	0.000	0.82	0.76-0.89	0.000	0.79	0.72-0.86	0.000
Socio Economic Index									
manual worker	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
non-manual	0.95	0.89-1.02	0.168	0.95	0.89-1.02	0.146	0.96	0.90-1.03	0.299
self employed	0.89	0.79-1.00	0.044	0.89	0.79-1.00	0.041	0.90	0.80-1.00	0.060
other	0.96	0.82-1.13	0.617	0.95	0.81-1.12	0.531	0.96	0.82-1.13	0.651
missing	0.94	0.81-1.08	0.356	0.93	0.81-1.07	0.316	0.91	0.79-1.05	0.209
Marriage status									
unmarried	1.00	(ref)	-	1.00	(ref)	-	1.00	(ref)	-
married	0.99	0.91-1.09	0.884	0.99	0.90-1.08	0.756	0.99	0.90-1.08	0.808
divorced	0.93	0.83-1.06	0.274	0.93	0.82-1.05	0.265	0.92	0.81-1.04	0.191
widowed	1.11	0.75-1.64	0.598	1.10	0.74-1.62	0.639	1.09	0.74-1.61	0.671
missing	1.28	0.79-2.07	0.312	1.29	0.80-2.09	0.293	1.31	0.80-2.14	0.281

[a] Adjusted for age at baseline examination (younger than 38, 38-44, 45-47, 48 or older), Body Mass Index (less than 20, 20-24, 25-29, 30 or more), smoking status (never, former, current), Socio Economic Index (manual worker, non-manual worker, self employed, other, missing), Prevalent cancer (yes/no), Season of baseline examination (winter, spring, summer, fall)

9 Acknowledgments

I would like to acknowledge:

Henrik Grönberg, who both in the role as chair of the department, as well as co-supervisor gave me the possibility to conduct the studies in this thesis.

Paul Dickman, who took me under his wings and miraculously managed to find not only publishable but also rewarding and edifying studies!

Hans-Olov Adami, who was the chair of the department when I started at MEB, and who also was the PI of my original doctoral project, for his visions for the department and epidemiology at large.

Pär Sparén, my former supervisor, who in collaboration with my former co-supervisor first lured me to the department.

Keng-Ling Wallin, my former co-supervisor, for your great scientific standards, for your mentorship, your knowledge, your pedagogic skills and your friendship! Academia should miss you!

Juni Palmgren and **Christina Hultman**, for being role models in academia, with heart and brains in synergy!

Anna Torrång, **Anna Johansson** and especially **Sandra Eloranta** for excellent statistical, intellectual and emotional support!

Zack Ysof and **Gastón Nuñez** for superb technical support and smiles!

Marie Reilly and **Trung Nam Tran**, for their intelligent drive and scientific spark in an intriguing project that sadly was put to sleep.

Jonas Manjer, **Martin Almquist** and **Johan Malm** in Malmö for offering me the unique possibility to study the Malmö Preventive Project, and of course for excellent co-authorship.

Alla män och kvinnor i Malmö Förebyggande medicin, för utan ert deltagande och engagemang hade vi inte kunnat studera sambandet mellan kalcium och cancer!

Ylva Trolle-Lagerros, for being there, encouraging and insightful when I needed it the most! I am forever greatful!

Catherine Tuvblad, for teaching me the most important fundaments of research!

Leila Nyrén, for being a joy to work with, in our former project, and for always cheering up my day!

The former **Doctoral Council** as well as the former **Future council** at KI and of course **Harriet Wallberg-Henriksson**, for unique insights, knowledge and good times!

Past and present colleagues at MEB, especially **Sanna Green, Marie Jansson , Eva Dannetun, Andreas Jacks, Gudrun Jonasdottir Bergman, Lisen Arnheim, Amina Said, Camilla Lagerberg, Ove Strind, Åsa Klint, Anna Olsén, Denny Rönngren, Monica Leu, Arvid Sjölander, Junmei Miao Jonasson, Fatima Azerkan, Ninoa Malki, Ulrika Ericsson, Nathalie Ylitalo, Katarina Bälter, Gunilla Sonnebring, Ann Almqvist, Connie Nordlund** and **Camilla Ahlqvist**.

All the wonderful baristas at Nesta, especially **Mia, Erik, Max**, and **Sebastian**, and **Sofie and her splendid colleagues** at Espresso House, Storgatan. Thanks to you I not only had great coffee, I also met my wonderful past, present, and future (?) colleagues; (among others) **Annika-Mimmis-mamma**, **Jurist-Marcus**, **Cykelbuds-Markus**, **Jockepjocke**, and last but not least, **toaster-Anna** - my best caféragg ever!

I want to thank **Jessica**, **Viktor**, **Fredrik**, and **@allaandra** who have distracted me and given me #pepp in equal amounts! <3

Alla ni fina, underbara på SB! Ingen nämnd, och absolut ingen glömd! Utan er så. . . <3

Raffaella Crinelli, for being an excellent researcher, a remarkable friend and an extraordinary witch!

All of my wonderful friends, especially; **Emma & Marcus, Tina, Sanja, Rita, Sofia, Anna, Fredrik & Boel, Andreas & Milena, Ekan,** and **Jessica & Johan**, for thoughts, laughs, and love!

My in-laws; **Lena**, **Janne & Ulla-Carin**, **Tessa**, with **Malva**, **Isaac** & **Elias**, **Axel** & **Magnus**, **Frida**, and of course **Wilma**, for all the love and support!

My family: **Magnus**, and his family **Natalia**, **Cristina** & **Lucas**, **Katrin & Lars-Ingmar**,

Margareta & Micke, **Olof**, **Martin & Johan**, **Helene & Morgan**, **Hasse**, **Caroline** & **Rickard**, and **Berno**, for all the help over the years and for being supportive!

My father, **Pelle**, and my mother and mentor, **Tina**, for all the love and help through out the years!

Lina, for being so close yet so far away, and for knowing me better than I do myself! I owe you my life!

Tobias, my beautiful husband, for being absolutely awesome at LaTeX, and for being my best friend in the whole world!

My smart, wonderful and fantastic daughters, **Nea** & **Lin**, for being 'my *best* Nea and Lin', and for putting everything in the right perspective! I love you all so much!

Bibliography

[1] Cox, B.; Sneyd, M. J. *Am J Epidemiol* **2011**, *173*, 394–403.

[2] Chan, J. M.; Gann, P. H.; Giovannucci, E. L. *J Clin Oncol* **2005**, *23*, 8152–8160.

[3] Almquist, M.; Manjer, J.; Bondeson, L.; Bondeson, A.-G. *Cancer Causes Control* **2007**, *18*, 595–602.

[4] Nilsson-Ehle, P. *Laurells klinisk kemi i praktisk medicin*; Studentlitteratur AB, Sweden, 2003.

[5] Unnamed Author, *Br Med J* **1977**, *1*, 598.

[6] Thacher, T. D.; Clarke, B. L. *Mayo Clin Proc* **2011**, *86*, 50–60.

[7] Bodiwala, D.; Luscombe, C. J.; French, M. E.; Liu, S.; Saxby, M. F.; Jones, P. W.; Fryer, A. A.; Strange, R. C. *Cancer Lett* **2003**, *200*, 141–148.

[8] Kristal-Boneh, E.; Froom, P.; Harari, G.; Ribak, J. *Eur J Epidemiol* **1999**, *15*, 237–244.

[9] IARC, *Section of Cancer Information* (**4/3/2011**).

[10] WHO, *Fact sheet No 297*, 2011.

[11] Grönberg, H. *Lancet* **2003**, *361*, 859–864.

[12] Gjertson, C. K.; Albertsen, P. C. *Med Clin North Am* **2011**, *95*, 191–200.

[13] Engel, P.; Fagherazzi, G.; Boutten, A.; Dupr, T.; Mesrine, S.; Boutron-Ruault, M.-C.; Clavel-Chapelon, F. *Cancer Epidemiol Biomarkers Prev* **2010**, *19*, 2341–2350.

[14] Selzer, M. L. *Am J Psychiatry* **1971**, *127*, 1653–1658, MAST.

[15] Kristenson, H.; Trell, E. *Br J Addict* **1982**, *77*, 297–304, MAST Mm-MAST.

[16] Chan, J. M.; Stampfer, M. J.; Ma, J.; Gann, P. H.; Gaziano, J. M.; Giovannucci, E. L. *Am J Clin Nutr* **2001**, *74*, 549–554.

[17] Chan, J. M.; Giovannucci, E.; Andersson, S. O.; Yuen, J.; Adami, H. O.; Wolk, A. *Cancer Causes Control* **1998**, *9*, 559–566.

[18] Tseng, M.; Breslow, R. A.; Graubard, B. I.; Ziegler, R. G. *Am J Clin Nutr* **2005**, *81*, 1147–1154.

[19] Giovannucci, E.; Liu, Y.; Stampfer, M. J.; Willett, W. C. *Cancer Epidemiol Biomarkers Prev* **2006**, *15*, 203–210.

[20] Kesse, E.; Bertrais, S.; Astorg, P.; Jaouen, A.; Arnault, N.; Galan, P.; Hercberg, S. *Br J Nutr* **2006**, *95*, 539–545.

[21] Gao, X.; LaValley, M. P.; Tucker, K. L. *J Natl Cancer Inst* **2005**, *97*, 1768–1777.

[22] Giovannucci, E. *Cancer Causes Control* **1998**, *9*, 567–582.

[23] Bonjour, J.-P.; Chevalley, T.; Fardellone, P. *Br J Nutr* **2007**, *97*, 611–616.

[24] Skinner, H. G.; Schwartz, G. G. *Cancer Epidemiol Biomarkers Prev* **2008**, *17*, 2302–2305.

[25] Skinner, H. G.; Schwartz, G. G. *Cancer Epidemiol Biomarkers Prev* **2009**, *18*, 575–578.

[26] Berglund, G.; Eriksson, K. F.; Israelsson, B.; Kjellström, T.; Lindgärde, F.; Mattiasson, I.; Nilsson, J. A.; Stavenow, L. *J Intern Med* **1996**, *239*, 489–497, MPP Malm Preventive Project.

[27] *Subset of FOB80 from Statistics Sweden*, Statistiska Centralbyrn SCB, rebro, Sweden, www.scb.se.

[28] Dickman, P. W.; Adami, H.-O. *J Intern Med* **2006**, *260*, 103–117.

[29] Dickman, P. W.; Sloggett, A.; Hills, M.; Hakulinen, T. *Stat Med* **2004**, *23*, 51–64.

[30] Bertone-Johnson, E. R.; Chen, W. Y.; Holick, M. F.; Hollis, B. W.; Colditz, G. A.; Willett, W. C.; Hankinson, S. E. *Cancer Epidemiol Biomarkers Prev* **2005**, *14*, 1991–1997.

[31] McCarty, M. F. *Med Hypotheses* **2000**, *54*, 475–482.

[32] Young, M. M.; Nordin, B. E. *Proc R Soc Med* **1967**, *60*, 1137–1138.

[33] Marshall, R. W.; Francis, R. M.; Hodgkinson, A. *Clin Chim Acta* **1982**, *122*, 283–287.

[34] Pitkin, R. M.; Reynolds, W. A.; Williams, G. A.; Hargis, G. K. *J Clin Endocrinol Metab* **1978**, *47*, 626–632.

[35] Zittermann, A.; Schwarz, I.; Scheld, K.; Sudhop, T.; Berthold, H. K.; von Bergmann, K.; van der Ven, H.; Stehle, P. *J Clin Endocrinol Metab* **2000**, *85*, 95–101.

[36] Jorde, R.; Sundsfjord, J.; Bnaa, K. H. *Eur J Epidemiol* **2001**, *17*, 1117–1123.

[37] Krall, E. A.; Sahyoun, N.; Tannenbaum, S.; Dallal, G. E.; Dawson-Hughes, B. *N Engl J Med* **1989**, *321*, 1777–1783.

[38] Morton, R. A. *Urology* **1994**, *44*, 637–645.

[39] Holick, M. F. *J Cell Biochem* **2003**, *88*, 296–307.

[40] Raimondi, S.; Mabrouk, J. B.; Shatenstein, B.; Maisonneuve, P.; Ghadirian, P. *Prostate* **2010**, *70*, 1054–1065.

[41] Newmark, H. L.; Heaney, R. P. *Nutr Cancer* **2010**, *62*, 297–299.

[42] Rao, A. R.; Motiwala, H. G.; Karim, O. M. A. *BJU Int* **2008**, *101*, 5–10.

[43] Morton, A. R.; Garland, J. S.; Holden, R. M. *Semin Dial* **2010**, *23*, 283–289.

[44] Kvåle, R.; Auvinen, A.; Adami, H.-O.; Klint, A.; Hernes, E.; Møller, B.; Pukkala, E.; Storm, H. H.; Tryggvadottir, L.; Tretli, S.; Wahlqvist, R.; Weiderpass, E.; Bray, F. *J Natl Cancer Inst* **2007**, *99*, 1881–1887.

[45] Leifsson, B. G.; Ahrn, B. *J Clin Endocrinol Metab* **1996**, *81*, 2149–2153.

[46] Reid, I. R. *Osteoporos Int* **2008**, *19*, 595–606.

[47] Alemzadeh, R.; Kichler, J.; Babar, G.; Calhoun, M. *Metabolism* **2008**, *57*, 183–191.

[48] Norman, A. W. *Am J Clin Nutr* **1998**, *67*, 1108–1110.