Plant foods, plasma enterolactone and breast cancer - with a focus on estrogen receptor status and genetic variation

Sonestedt, Emily

2008

LIST OF PAPERS

This doctoral dissertation is based on the following original papers, which in the text will be referred to by their Roman numerals.

I. Sonestedt E, Ericson U, Gullberg B, Peñalvo JL, Adlercreutz H, Wirfält E: Variation in fasting and non-fasting serum enterolactone concentrations in women of the Malmö Diet and Cancer cohort. *Eur J Clin Nutr* 2008 Aug;62(8):1005-9

II. Sonestedt E, Borgquist S, Ericson U, Gullberg B, Landberg G, Olsson H, Wirfält E: Plant foods and oestrogen receptor α- and β-defined breast cancer: observations from the Malmö Diet and Cancer cohort. *Carcinogenesis* 2008 Nov;29(11):2203-9

III. Sonestedt E, Borgquist S, Ericson U, Gullberg B, Olsson H, Adlercreutz H, Landberg G, Wirfält E: Enterolactone is differently associated with estrogen receptor β-negative and -positive breast cancer in a Swedish nested case-control study. *Cancer Epidemiol Biomarkers Prev* 2008 Nov;17(11):3241-51

IV. Sonestedt E, Ivarsson M.I.L, Harlid S, Ericson U, Gullberg B, Carlson J, Olsson H, Adlercreutz H, Wirfält E: Polymorphisms in the estrogen receptor α and β genes, plasma enterolactone and breast cancer. (submitted)

ABSTRACT

Diets high in fibre have previously been associated with decreased risk of breast cancer in the Malmö Diet and Cancer cohort. Several potent compounds may exist in high-fibre diets that might protect against breast cancer, for example lignans. Plant lignans are converted to enterolactone by the gut microflora and are similar in structure to estrogens. Enterolactone may interact with the estrogen receptors (ERs) and, therefore, inhibit the effect of estrogens.

The aim of this doctoral project was to prospectively examine if plant food intakes and enterolactone blood concentrations were associated with breast cancer, and to examine if the association differed depending on ER status of the tumours or variation in the ER genes. Information (including high validity dietary data) from the Malmö Diet and Cancer cohort with baseline examinations from 1991 to 1996 was used. Among 15,773 women, 45-73 years at baseline, without prevalent cancer, 544 women were diagnosed with breast cancer until 31 December 2004.

High intakes of high-fibre bread were associated with decreased risk of breast cancer. When restricting the analyses to individuals with suggested more stable food habits, a decreased breast cancer risk were also observed with high intakes of fruit, berries and vegetables. High-fibre bread and fruit and berries were the main dietary determinants of enterolactone concentration. In addition, obesity and smoking was associated with lower enterolactone concentrations. High enterolactone concentrations (>16 nmol/L) was associated with decreased breast cancer risk. When stratifying for fibre intake, a decreased breast cancer risk with high enterolactone concentration was only observed among individuals with high fibre intakes. Relatively high variation of enterolactone within and between individuals was observed; the association between enterolactone and risk of breast cancer is likely attenuated. The reduced breast cancer risk with high enterolactone concentrations was only observed for ERα-positive and ERβ-negative tumours. Breast cancer risk was not significantly associated with any of the selected polymorphisms in the ERα and ERβ genes. However, the protective effect of high enterolactone concentration was seen in subgroups of the individuals with specific genetic variants, and a tendency towards an interaction between a polymorphism in intron 3 of the ERα gene and enterolactone concentrations was observed.

In conclusion, a high-fibre diet including high-fibre bread, fruit, berries and vegetables will likely reduce the risk of breast cancer in middle aged and older women. The protective effect of a high-fibre diet might be due to its content of lignans.

SAMMANFATTNING

Varje år insjuknar ca 7000 kvinnor i Sverige i bröstcancer och antalet ökar för varje år. Fortfarande är kunskapen om orsakerna till bröstcancer relativt okända, men många av de kända riskfaktorerna är kopplade till en hög halt av könshormoner i blodet. Faktorer i kosten som kan motverka effekten av könshormonerna är därför av stort intresse.

Under 1990-talet inbjöds alla invånare i Malmö födda mellan 1923-1950 till Malmö Kost Cancer studien. Av dem som kallades deltog ca 40 % i studien, 28 098 personer varav 17 035 kvinnor. I studien användes en mycket omfattande metod för att ta reda på deltagarnas kostintag. De fick också fylla i ett frågeformulär om livsstilsfaktorer; man mätte längd och vikt, och samlade in blodprover. Tillsammans ger detta unika möjligheter att undersöka sambandet mellan kost och cancer.

Det finns en teori om att en fiberrik kost skyddar mot bröstcancer; man har tidigare funnit ett samband mellan högt intag av kostfiber och minskad risk för bröst cancer i Malmö Kost Cancer studien. Flera ämnen som man identifierat i fiberrika livsmedel kan ha en speciellt skyddande effekt mot cancer, bland annat lignaner som förekommer rikligt i frön, bär, fullkornsprodukter, nötter, grönsaker och frukt.

Lignanerna i kosten är overksamma. För att de ska få effekt och kunna tas upp i kroppen måste de omvandlas av tarmbakterierna till s.k. enterolignaner, främst enterolakton. Därför är det viktigt att ha en väl fungerande tarmflora. Studier har visat att användning av antibiotika gör att halterna av enterolakton i blodet minskar kraftigt och halten kan vara reducerad upp till ett år eller ännu längre.

Enterolakton anses vara en s.k. fytoöstrogen då de till strukturen liknar östrogen, det kvinnliga könshormonet. En hög halt av östrogen är en riskfaktor för bröstcancer. Östrogen bidrar till att celler och cancertumörer växer genom att binda till östrogenreceptorerna som är proteiner i cellerna. Det är inte helt klarlagt hur enterolakton verkar, men man har sett att även enterolakton kan interagera med östrogenreceptorerna. De kan troligen hämma effekten av östrogen då de anses tävla med östrogen om östrogenreceptorern. Det finns dessutom två typer av östrogenreceptorer: α och β, som finns både i normala celler och i cancerceller. Cancertumörer med receptorer skiljer sig mycket från dem utan receptorer. Därför är det viktigt att studera dessa olika typer av tumörer var för sig istället för att undersöka bröstcancer som en enda sjukdom.

Östrogenreceptor α har varit känd länge. Tillväxten av tumörer med höga halter av östrogenreceptorer stimuleras av östrogen, därför behandlar man dessa tumörer med ämnen som hämmar östrogenet. Östrogenreceptor β upptäcktes relativt nyligen och är inte lika väl studerad. Eftersom östrogenreceptor β verkar hämma effekten av östrogenreceptor α är det viktigt att även ta hänsyn till halten av östrogenreceptor β i tumörerna. Man har på senare tid funnit att enterolakton främst binder till östrogenreceptor α.

Generna som kodar för östrogenreceptorerna varierar mer eller mindre i de flesta befolkningar. Genetisk variation kan ha betydelse eftersom detta kan medföra att receptorernas funktion är förändrad; samverkan mellan östrogenreceptorerna och enterolakton kan därför vara annorlunda.

Man tror att det krävs en viss mängd fiber från sädeskorn i kosten för att lignanerna ska ha någon effekt på cancerrisken, eftersom fibrer tillsammans med lignaner skyddar mot bröstcancer. Detta medför att källor till lignaner som inte bidrar med fibrer (t.ex. kaffe, te, juice och vin) inte heller ger något skydd mot cancer. Därför kan intaget av lignaner ha större betydelse i Nordeuropeiska befolkningar där man konsumerar en stor del fullkornscerealier.

I detta doktorandprojekt har vi undersökt sambandet mellan fiberrika livsmedel, enterolakton och bröstcancer. Flera andra studier, som har undersökt sambandet mellan kostfaktorer och bröstcancer, har tagit hänsyn till hur mycket av östrogenreceptor α som det finns i tumörerna. Det unika med den här studien är att vi även har tagit hänsyn till halten av östrogenreceptor β. Andra stora fördelar är att vi har kunnat mäta halten av enterolakton i blodet innan kvinnorna har fått bröstcancer, att vi har haft tillgång till mycket detaljerade data avseende kostintaget hos dessa kvinnor, samt att vi har haft möjlighet att undersöka genetisk variation i generna som kodar för östrogenreceptorena.

Bland de kvinnor som ingick i studien och som inte tidigare hade haft cancer insjuknande 544 kvinnor fram till 31 december 2004. Bland de fiberrika livsmedlen fann vi att främst det fiberrika brödet visade ett skyddande samband. De kvinnor som åt mest fiberrikt bröd hade 25 % minskad risk för att utveckla bröstcancer jämfört med dem som åt minst fiberrikt bröd. Det verkade även som att de kvinnor som hade ett högt intag av frukt, bär och grönsaker hade en minskad risk för bröstcancer.

För att undersöka om det var en hög exponering av lignaner som till viss del förklarade varför en fiberrik kost skulle skydda mot bröstcancer mätte vi halten av enterolakton i blodet hos 366 individer som senare fick bröstcancer och 733 individer utan bröstcancer. När vi delade in kvinnorna i två grupper beroende på halten av enterolakton såg vi att de med en hög halt av enterolakton hade en minskad risk för att insjukna i bröstcancer jämfört med dem med en låg halt. Det skyddande sambandet såg vi främst hos individer som hade ett högt fiberintag, vilket styrker teorin om att det krävs en viss halt av fiber i kosten för att lignanerna ska ha någon effekt.

En hög halt av enterolakton var kopplat till fiberrik kost, speciellt ett högt intag av fiberrikt bröd, frukt och bär. Rökning och övervikt var däremot kopplat till en lägre halt av enterolakton. En stor del av variationen mellan olika individer kan man inte förklara med dessa faktorer, vilket bl.a. visar på hur viktig bakteriefloran i tarmen är för att förmedla kostens inflytande.

Vi undersökte även variationen av enterolakton i blodet genom att mäta halten vid upprepade tillfällen hos 21 kvinnor. Vi fann en ganska stor variation, vilket gör att sambandet mellan enterolakton och bröstcancer kan bli försvagat. Vi ansåg ändå att enterolaktonhalten i blodet var användbart för att ranka individer eftersom ett relativt stort antal individer var med i vår studie.

Det skyddande sambandet med enterolakton såg vi främst bland tumörer som uttryckte östrogenreceptor α men inte östrogenreceptor β. Då östrogenreceptor β allmänt tycks hämma effekten av östrogenreceptor α kan detta tolkas som att dessa tumörer är mer känsliga för den antiöstrogena effekten av enterolakton. Vårt fynd bidrar med en ökad kunskap om lignanernas betydelse för hälsan.

Vi fann att en genetisk variation i genen som kodar för östrogenreceptor α tycktes modifiera sambandet mellan enterolakton och bröstcancer. Studierna visar att för personer med denna genetiska variation kan det vara extra viktigt att ha en hög halt av enterolakton i blodet.

Sammantaget visar studierna att ett högt intag av fiberrikt bröd och frukt, bär och grönsaker tycks minska risken för att insjukna i bröstcancer. Då en hög halt av enterolakton i blodet är kopplat till en kost rik på fiberrikt bröd och frukt och bär, kan den minskade risken för bröstcancer som vi ser hos individer med ett högt intag av fiber delvis bero på kostens innehåll av lignaner. Rökning och övervikt var däremot kopplat till en lägre halt av enterolakton. En hög halt av enterolakton i blodet var kopplad till en minskad risk för bröstcancer, speciellt den typ av bröstcancer som uttrycker östrogenreceptor α men saknar östrogenreceptor β. Sambandet mellan enterolakton och bröstcancer kan även bero på genetisk variation, då vissa individer tycks vara speciellt mottagliga för den skyddande effekten av en hög enterolaktonhalt.

ABBREVIATIONS

AF-1	Activation function 1
BMI	Body mass index
CI	Confidence interval
CV	Coefficient of variance
DAG	Directed acyclic graphs
EPIC	European Prospective Investigation into Cancer and Nutrition
ERα	Estrogen receptor alpha
ERβ	Estrogen receptor beta
ER (-)	Estrogen receptor negative
ER (+)	Estrogen receptor positive
ERE	Estrogen responsive element
ESR1	Estrogen receptor α gene
ESR2	Estrogen receptor β gene
FFQ	Food frequency questionnaire
GC/MS	Gas chromatography/mass spectrometry
HR	Hazard ratio
HWE	Hardy Weinberg equilibrium
IARC	International Agency for Research on Cancer
ICC	Intraclass correlation coefficient
ICD	International Classification of Diseases
LC/MS	Liquid chromatography/mass spectrometry
LD	Linkage disequilibrium
MAF	Minor allele frequency
MDC	Malmö Diet and Cancer
nmol/L	Nanomol/liter
PR	Progesterone receptor
OR	Odds ratio
SHBG	Sex hormone-binding globulin
SNP	Single nucleotide polymorphism
SYNE	Spectrin repeat containing nuclear envelope 2
TR-FIA	Time-resolved fluoroimmunoassay
UTR	Untranslated region
µg	Microgram
ng	Nanogram

INTRODUCTION

Breast cancer is the most common cancer among Swedish women. Although its aetiology is still relatively unclear, high exposure of sex hormones is an established risk factor for breast cancer. Several dietary factors have been hypothesised to be involved in the development of breast cancer; factors inhibiting the effect of estrogens are of specific interest.

Diets high in fibre and low in fat have previously been associated with decreased risk of breast cancer in the Malmö Diet and Cancer (MDC) cohort. There are several potent compounds in high-fibre diet that might protect against breast cancer, for example lignans.

BACKGROUND

Breast cancer

Carcinogenesis and definition of breast cancer

Cancer is characterised by uncontrolled cell growth and cell death. Carcinogenesis is the process by which normal cells develop into cancer cells, and it is initiated by a series of mutations in specific genes: activation of oncogenes and inactivation of tumour suppressor genes. Gene expression can also be altered without changing the DNA sequence. Such epigenetic changes may act as surrogates for the inactivation of these genes (1). These alterations contribute to tumour progression, with increased rates of normal cell transformation into cancer cells, uncontrolled cell growth and uncontrolled cell death.

Normal breast growth and development are regulated by hormones (e.g., estrogens, progesterone and androgens) and growth factors (2). The majority of breast cancers develop from the inner layer of luminal epithelial cells comprising the ducts. Breast tumours can be either invasive or non-invasive; non-invasive tumours are referred to as *in situ* carcinoma (2).

Descriptive epidemiology

Breast cancer is the most frequent cancer among Swedish women. In 2006, 7059 new cases of breast cancer were detected among women, representing 29.4 % of all cancers diagnosed in women (3). Breast cancer seems to be more common in the south of Sweden, demonstrating an age-standardised rate of 165.4 per 100,000 in Skåne (south part of Sweden) (178.8 in Malmö) compared to 142.5 in Sweden (3). There has been an increased incidence of cancer during the last two decades, with a 1.3 % annual increase (Figure 1). At the age of 75, 9.8 % of women in Sweden will be diagnosed with breast cancer during their life (3).

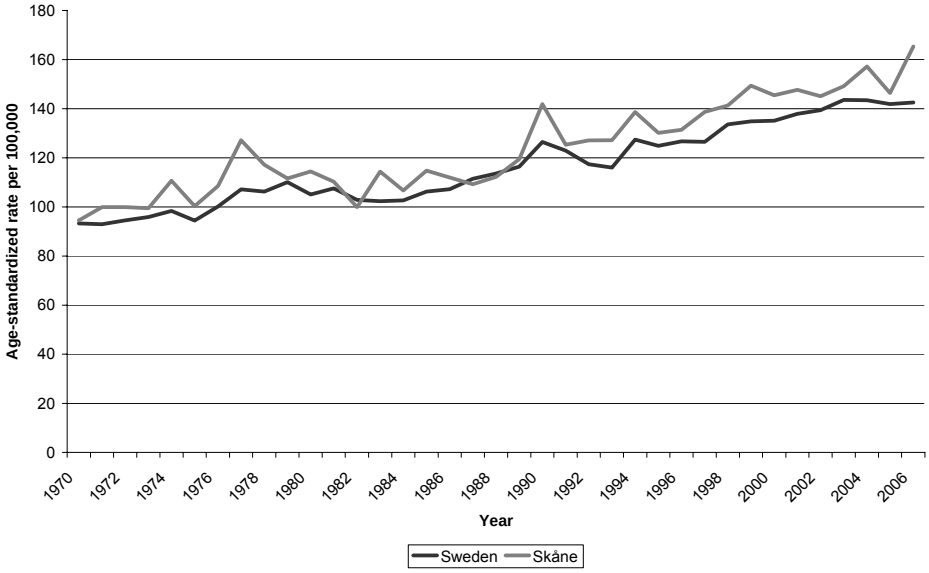

Figure 1. Age-standardised rates of incidence breast cancer per 100,000 in Sweden and Skåne

Breast cancer is also the most common cancer in women worldwide, although there are large differences in breast cancer incidence recorded from different countries. The highest incidences are reported in northern and western Europe, North America, Australia and New Zealand, as well as in the southern regions of South America (4).

Risk factors for breast cancer

Genetic factors

Family history of breast cancer is a major risk factor. The breast cancer risk is doubled for first-degree relatives compared to women in the general population (5). Familial cases are especially prominent before menopause (5). According to a study in Denmark, Sweden and Finland, the twin of a monozygotic twin with breast cancer has a five-fold increased risk of breast cancer compared to that of a monozygotic twin without breast cancer (6). A few high-risk genes accounting for familial breast cancer have been identified: BRCA1, BRCA2, PTEN and TP53. However, germline mutations in these high-penetrance genes explain only a minority of the familial cases, and only 5-10 % of all breast cancers (7). A single nucleotide polymorphism (SNP) is defined as a change in a single base pair that occur in more than 1% of the subjects in a population. More than 10 million SNPs in the human genome have been reported. According to the common disease-common variant hypothesis, common genetic variants exist in the human genome that influence susceptibility to complex polygenic diseases; however, each variant has only minor effects on an individual's risk for the disease.

Recently, a genome-wide association study identified genetic variants in five novel independent loci that showed strong and consistent association with breast cancer. Four of these loci contain genes that are assumed to be causative (FGFR2, TNRC9, MAK3K1 and LSP1). However, the study observed more genetic variants significant at the $p<0.05$ level than those expected, indicating that there are more alleles that influence susceptibility to breast cancer (8).

The Scandinavian twin study estimated that heredity might account for about 27 % of breast cancers in Scandinavian women (6). In a study assessing 9.6 million individuals comprising all Swedes born after 1934 and their parents, 25 % of breast cancers were estimated to be accounted for by heredity (9).

Migration studies suggest that environmental differences (rather than genetics) account for the differences in incidence worldwide (10). For example, a study showed that Asian-American women born in western countries had a 60 % increased risk compared to Asian-American women born in eastern countries (11). There are several possible reasons for this difference including a western lifestyle with a high-fat, low-fibre diet, high alcohol intake, and reproductive factors leading to high hormone levels.

Age

The incidence of breast cancer generally increases with age. However, the incidence curve for age at diagnosis of breast cancer is not constant throughout life. The slope decreases significantly during and shortly after menopause (the cessation of menstruation) (2). However, in the last decades the incidence of postmenopausal breast cancer in Sweden increased dramatically, mainly due to mammography screening (12); the decline in incidence after menopause is not that obvious as before. Hormone concentrations differ between pre- and postmenopausal women, as the only source of female sex hormones in postmenopausal women comes from adrenal androgens that are peripherally converted to estrogens. Many risk factors influence hormone levels; as risk factors for breast cancer have been shown to differ between pre- and postmenopausal women, they should be considered separately.

Reproductive factors and hormones

Estrogen exposure is an established risk factor for breast cancer, as high concentration of circulating endogenous estrogens have been associated with increased breast cancer risk (13-15). In addition, evidence from epidemiological studies shows that current use of exogenous estrogens in the form of menopausal hormone therapy increases the risk of breast cancer and that the risk increases for each year of use. However, the effect is reduced after cessation of use, and the increased risk of breast cancer almost disappears after 5 years (16). A large clinical trial indicated that menopausal hormone therapy is associated with increased risk of post-menopausal breast cancer (17). Reproductive factors that are characterised with prolonged estrogen exposure (e.g., early age at menarche, older age at birth of first child and late age at menopause) has also been associated with an increased risk of breast cancer (13); thus it seems that over-exposure to sex hormones during one's lifetime contributes to tumour development (18).

Estrogen is not an initiator of breast carcinogenesis; however, certain metabolites of estrogen can bind to DNA and induce mutations (19). Estrogen may also increase the risk of breast cancer by stimulating breast epithelial cell proliferation, thus rendering the breast tissue more susceptible to carcinogens (19). Many other lifestyle factors are involved in breast cancer via their influence on estrogen concentrations.

Lifestyle factors

According to the World Cancer Research Fund, which has reviewed all relevant epidemiological studies, there is convincing evidence that high alcohol intake, obesity, and body height increase the risk of postmenopausal breast cancer. The evidence is weaker (although ranked as probable) for physical inactivity, abdominal fat and adult weight gain as risk factors for postmenopausal breast cancer (20). Obesity seems to be a risk factor mainly for postmenopausal breast cancer (21). This is mainly related to the higher circulating levels of sex hormones in obese women compared to those who are lean, as much of the circulating estrogen is derived from aromatisation of androgen in peripheral adipose tissue among postmenopausal women. In premenopausal women, an inverse association between obesity and risk of breast cancer has been reported in several studies; however, the evidence for decreased risk associated with obesity among premenopausal women is more limited than the increased risk ascertained for postmenopasal breast cancer (20). The mechanisms are speculative; one explanation for this observation is that obese women are less likely to ovulate, which reduces their number of lifetime ovulations and alters their blood levels of sex hormones (22).

There are also several risk markers that are not necessarily causal factors. For example, breast cancer is more frequent among women with higher education. This is probably related to lifestyle factors including alcohol intake, use of menopausal hormone therapy and late age at birth of the first child.

Diet

Many dietary factors have the ability to influence several events during carcinogenesis, including DNA damage, DNA repair, apoptosis, proliferation and differentiation (20). In 1981, Doll and Peto estimated that 10 to 70 percent of all cancers in the USA are related to diet (23). However, the World Cancer Research Fund found no convincing or probable evidence for any specific food group being associated with the risk of breast cancer (20). Studies examining dietary fat and breast cancer have produced many conflicting results (24-26). Although not consistent across epidemiological studies, a meta-analysis suggests a moderate breast cancer protective effect for vegetables, but not for fruits (27). However, pooled analyses of eight prospective studies demonstrated a weak non-significant decreased risk associated with high intakes of fruit and vegetables (28). The ambiguous results between diet and disease observed in epidemiological studies can, to a large extent, depend on measurement errors in dietary assessment. In addition, dietary exposure is a very complex factor. Genetic factors may also explain inconsistencies, since cancer risk associated with a dietary factor may be modulated by different genetic traits (29).

Low-fat and high-fibre diets in the MDC cohort are associated with a low incidence of breast cancer in postmenopausal women (30). Unlike other countries where fruit and vegetables contribute to most of the fibre intake, grains are a major source of fibre in Nordic countries (31,30,32). Therefore, the specific plant food sources of fibre may be of particular interest when examining the association between food components and breast cancer.

Dietary factors that can influence hormone levels may be important for breast cancer prevention. Plausible mechanisms explaining the protective effect of fibre-rich foods are associated with their influence on the bioavailability and activity of estrogens. Fibre might, for example, influence the enterohepatic recirculation of estrogens resulting in reduced levels of circulating estrogens (33,34). Some of the protective effects against cancer associated with a high-fibre diet may be due to other bioactive components and phytochemicals, for example phytoestrogens, antioxidants and folate.

Enterolactone

Phytoestrogens are estrogen-like substances found in plants that may have health benefits and possibly protect against diseases like breast and prostate cancer (35). The major classes of phytoestrogens are isoflavones, coumestans and lignans. Isoflavones (mainly genistein and daidzein) are predominantly found in soybeans and legumes, and coumestans (mainly coumestrol) are found in different kinds of sprouts; isoflavones and coumestans are, therefore, probably not consumed in high amounts in Sweden (36). Fibre-rich foods (like whole grains, seeds and berries) contain lignans (37,38), which are converted to enterolignans by the intestinal microflora (39). In addition, some flavonoids also have estrogenic effects and might be classified as phytoestrogens (40).

Plant lignans

Lignans are diphenolic compounds and have been long known for their existence in various plants. Plant lignans play an important role in plant defence, as they have antibacterial, antifungal, antiviral, insect antifeedant and antioxidant properties (41). Some lignans have clinical use; for example, Podophyllotoxin has been used to treat genital warts.

In 1979, lignans were detected in humans. A cyclic pattern of lignan levels was reported during the menstrual cycle (42), and they were first believed to be endogenous hormones. The structures of the two enterolignans were thereafter identified as enterolactone (trans-2,3-bis (3-hydroxybenzyl)-γ-butyrolactone) and enterodiol (2,3-bis (3-hydroxybenzyl)-butane-1,4-diol) (43,44).

The observation that urinary excretion of enterolactone was increased in vegetarian women compared to omnivorous women, and that the amount of fibre in the diet was correlated with urinary enterolactone excretion (45) (46), suggested the existence of precursors in fibre-rich foods. Since then, several precursors of enterolignans have been identified. Secoisolariciresinol (47) and matairesinol (48) were the first plant lignans identified in foods. Recently, several other precursors of enterolignans have been identified: pinoresinol, syringaresinol, lariciresinol (39), sesamin (49,50), lignins (51) and other lignans including 7'-hydroxymatairesinol and arcigenin (52,39).

Lignans are predominantly found in the fibre component of plants. The highest concentrations are found in seeds, such as flaxseed and sesame seed. Other sources of plant lignans are whole grain cereals, legumes, berries, fruits, vegetables, wine, coffee and tea (53,37,54,55).

Several studies have tried to estimate the intake of lignans. A Dutch study estimated that 37 % of lignans were obtained from beverages and 9 % from bread (38). The main sources of lignans in Finland were estimated to be seeds, cereals, fruit, berries and vegetables (56). However, there are difficulties in estimating lignan intakes as many enterolignan precursors occur in plants, and there is a wide variation of lignans in foods. In addition, diet methods are often not designed to estimate lignan intakes.

Thompson *et al.* examined the production of enterolignans from plant foods after *in vitro* fermentation with intestinal bacteria. They found that oilseeds produced the highest amounts of enterolignans (approximately 20,000 µg/100g), followed by dried seaweeds, legumes, whole grain cereals and cereal brans, (400-900 µg/100g) and vegetables and fruit (50-150 µg/100g) (57).

Diets enriched in flaxseeds have been observed to duplicate the concentration of serum enterolactone (58). Another study found that absorption was substantially improved with crushed or milled flaxseeds (59). A diet enriched in sesame seeds also led to increased enterolactone concentrations (49), and similar urinary excretion of enterolignans after supplementation with sesame seeds or with whole flaxseed was observed (60). Intervention studies with whole grains also resulted in increased enterolactone concentration (61,62). During a 12-week intervention study in North Karelia, Finland, investigating the effect of a diet low in fat and high in vegetables, fruit and berries, the median concentration of enterolactone rose from 12.2 to 19.5 nmol/L (63).

Formation of enterolignans and the importance of microflora

Plant lignans are biologically inactive, and for biological activation they need to be released from the plant matrix and converted to enterolignans (64). Plant lignans in foods exist predominantly as glycosides. The metabolic processes that convert lignans to enterolignans include deglucosylation, ring cleavage, demethylation, dehydroxylation and oxidation. Anaerobic bacteria in the intestinal tract (mainly the proximal colon) are responsible for this conversion (65), and several bacterial strains have been identified to be responsible for this conversion (66-68). Studies have indicated that both the amount of the responsible bacteria and the composition of the bacterial strains are important for conversion of plant lignans to enterolignans (68).

Enterolignan absorption seems to occur in the large intestine (69). Enterolactone appears in the plasma 8-10 h after consumption of plant lignans (70), and after a single dose of flaxseed, blood concentrations were still increasing after 24 hours (71). Enterolactone is generally the lignan that exhibits the highest concentration in blood (72), and enterolignan that most commonly examined in epidemiological studies. However, supplementation with flaxseed primarily increases enterodiol concentrations (71).

Once absorbed, enterolactone is conjugated to glucuronic acids and sulphate (73) and in the circulation enterolactone may bind sex hormone-binding globulin (SHBG) (74,75) and compete for binding of endogenous hormones. Enterolactone can be further metabolised (76,77), although the metabolic fates of lignans in general are poorly understood. However, the resultant metabolites have been suggested to be bioactive (78,79). Lignans are excreted either in urine or bile; lignans in the intestine can be reabsorbed, and thereby undergo enterohepatic circulation. Some plant lignans are also excreted in the urine, raising questions regarding the putative health effects of plant lignans (80).

Mechanism of action

Similar to many estrogens, enterolactone is a diphenolic compound (Figure 2). A high circulating concentration of estrogens is an established risk factor for breast cancer, and enterolactone may decrease

estrogen exposure by influencing the SHBG concentration (81) and aromatase activity (82). It may also modulate the activity of estrogen receptors (ERs) (83,84,79). A study from Finland showed activation of ER-mediated transcription for enterolactone with preference for ERα; activation through ERβ required a higher enterolactone concentration. However, their data suggests tissue and cell type-specific activity of enterolactone (79). In addition, non-hormonal mechanisms for the breast cancer protective effects of enterolactone have been suggested, including antioxidative effects (85), and effects of apoptosis and inhibitory effect on angiogenesis (86).

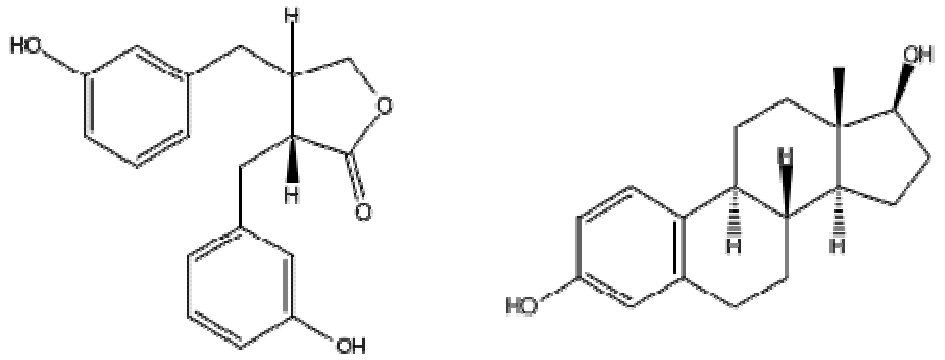

Figure 2. Chemical structures of enterolactone and 17beta-estradiol (Source: ChemBioFinder.com version 2.0.0.26)

Epidemiological studies

Case-control studies have shown decreased breast cancer risks associated with high circulating enterolignan concentrations (46,87-90), but prospective cohort studies have, however, demonstrated results that are less clear (91-97) (Table 1).

Olsen *et al*. found an inverted U-shaped curve for the associated between enterolactone and breast cancer in Danish postmenopausal women (95). A nested case-control in the north of Sweden found the opposite relationship and observed increased risks of breast cancer at very high and very low serum levels of enterolactone (92). However, the majority of the women with high enterolactone concentrations were premenopausal. Thus, other factors (e.g., genetic factors) may have played a greater role in that particular subgroup of women. In a recent study examining a Dutch cohort of postmenopausal women, enterolactone concentrations were not related to breast cancer risk (96), and a nested case-control study comprising 206 breast cancer cases conducted in Finland showed no protective effects associated with high enterolactone concentrations (94).

Several factors may have contributed to the contradictory results obtained for enterolactone and breast cancer in epidemiological studies:

- Age distribution
  As dietary factors may differ in their association with breast cancer depending on menopausal status, the contradicting results may be due to the poorly determined menopausal status of the women.

- Enterolactone measurement error
  Measurement error may lead to misclassification of the enterolactone concentration. This may be due to the low reliability of enterolactone concentrations in the population, resulting in poor precision for one single measurement. Misclassification may also be due to high frequency of antibiotic users, resulting in short-time low concentrations of enterolactone.

Table 1. Circulating enterolactone concentrations and postmenopausal breast cancer risk in prospective studies

Reference	Country	Subjects (age)	Enterolactone measurement (method)	Enterolactone concentrations	Results
den Tonkelaar, 2001 (91)	Netherlands	88 cases/268 controls (50-64 y)	2 overnight urinary samples 1 year apart (TR-FIA)		OR=1.43 (0.79-2.59) for 3rd vs. 1st tertile
Hultén, 2002 (92)	Sweden	248 cases/492 controls (Mean 51 y)	One plasma (TR-FIA)	Median: 18.6 nmol/L	Below 12.5th percentile: OR=1.6 (1.0-2.6) Above 87.5th percentile: OR=1.8 (1.4-4.3)
Grace, 2004 (93)	UK	Spot urine: 114 cases/219 controls Serum: 97 cases/187 controls (45-75 y)	Spot urine (GC/MS) Serum (LC/MS)	Median: 3.8 ng/ml (serum)	Spot urine: OR=0.98 (0.85-1.13) for doubling of level Serum: OR=1.00 (0.82-1.20) for doubling of level
Zeleniuch-Jacquotte, 2004 (98)	US	228 cases/228 controls (postmenopausal)	Non-fasting serum (TR-FIA)	14.5 nmol/l (cases) 14.3 nmol/l (controls)	OR=1.0, 1.3. 1.2, 1.3, 1.0 (over quartiles)
Kilkkinen, 2004 (94)	Finland	206 cases/215 controls (mean, 48 y)	One fasting serum (TR-FIA)	Median: 17.9 nmol/L	All subjects: OR=1.00, 1.67, 1.71, 1.30 Postmenopausal: OR=1.00 (ref), 1.26, 1.22
Olsen, 2004 (95)	Denmark	381 cases/381 controls (50-64 y)	One non-fasting plasma (TR-FIA)	Median: 28.2 nmol/L	IRR=0.93 (0.86-1.01) per 20 nmol/L higher concentration
Verheus, 2007 (96)	Netherlands	296 cases/296 controls (mean, 59 y)	Plasma (LC/MS)	2.71 (cases) 2.65 (controls)	ENL: OR= 0.97 for 3rd vs. 1st tertile END: OR=0.91 for 3rd vs. 1st tertile
Ward, 2008 (97)	UK	237 cases/952 controls (45-75 y)	Urine (GC/MS) Serum (LC/MS)	5.82 ng/ml (cases) 5.00 ng/ml (controls)	

TR-FIA= Time-resolved fluoroimmunoassay
GC/MS= gas chromatography/mass spectrometry
LC/MS=liquid chromatography/mass spectrometry

- Too low and different concentrations of enterolactone

  The conflicting results may be related to concentrations that are too low in several of these populations. Furthermore, the study may include a homogenous study population resulting in narrow range of exposure, and therefore a limited ability to observe any association.

- Different sources of enterolactone precursors

  Dietary fibre complex, including its associated lignans, seems to protect against various diseases (35). Coffee, tea, fruit juice, and wine increase plasma enterolactone without adding any fibre to the diet (99), and the protective effect of enterolactone may not be observed in populations where these food groups are major contributors to enterolactone concentrations. Furthermore, high amounts of fibre seem to be present in the diet for enterolactone to show a protective association. Therefore, in order to interpret the results obtained in epidemiological studies, information on dietary intake will be valuable.

- Heterogeneity of the subtypes of breast cancer

  The conflicting results among studies may also be explained by differences in tumour biology. Breast cancer is a heterogeneous disease with a wide range of different characteristics. There are specific types of breast cancer that are more sensitive to hormonal factors. Therefore, separate analysis of the different types of breast cancer is important.

- Genetic factors

  The occurrence of genetic factors, that influence the effect of enterolactone and estrogens, may differ across populations.

Reliability of enterolactone measurements

Because many large-scale epidemiological studies are only able to collect single blood samples, it is important to carefully examine the reliability of using one sample when classifying individuals according to their blood concentrations. Moreover, it is important to estimate the variability of the biomarkers in each study population for which an epidemiological study will be performed because the variability across studies is not always comparable.

The reliability can be estimated by the intraclass correlation coefficient (ICC), which is defined as the ratio of the between-person to the total variability (i.e. an estimate of the proportion of the variation in the exposure explained by the variation between groups relative to the total variation). A reliable biomarker that is useful for epidemiological studies is characterised by a high ICC value, while a low ICC (close to zero) means either low between-person variability or high within-person variability.

To account for the large variation, it is recommended that the sample size be increased. However, if the ICC in a population is known, this estimate can be used to correct risk estimates for random within-person measurement error (100). This procedure has been used in some epidemiological studies, for example the study performed by Stumpf (101). Biomarkers with low ICC often result in attenuation of the relationship between the exposure and the disease (102). Hankinson *et al*. suggested that relative risks would be substantially attenuated when the ICC is less than 0.65. However, because of several sources of misclassification, even an ICC of 0.65 may be too low (103).

A study utilizing fasting blood collected once a week for 3 weeks among 20 university students in Finland estimated an ICC of 0.77 for enterolactone (104). Another study examining non-fasting blood noted plasma enterolactone concentration to be relatively stable over a 2-year period, with an ICC of 0.55 (105). In an

Table 2. Determinants of circulating enterolactone concentrations

Reference	Country	Subjects, age	Dietary assessment method	Enterolactone measurement (method)	Dietary determinants	Non-dietary determinants
Johnsen, 2004 (106)	Denmark	857 women (50-64 y)	192-item FFQ	One non-fasting plasma (TR-FIA)	Whole grains↑ Leafy vegetables↑ Cabbage↑ Coffee↑	Smoking↓ High BMI↓ Frequent bowel movements↓
Kilkkinen, 2001 (107)	Finland	1168 men/1212 women (25-64 y)	38-item FFQ	One 4h fasted serum (TR-FIA)	Men: whole-grain products↑ Fruit and berries↑ Women: vegetables↑	Men: constipation↑ Women: age↑ Constipation↑ Smoking↓ Weight↓
Horner, 2002 (99)	USA	115 women/78 men (20-40 y)	3-day food record	Two fasting plasma (TR-FIA)	Vegetables↑ Fibre↑ Carbohydrates↑ Vegetable protein↑ Caffeine↑, Alcohol↑	Females↑ Age↑ BMI↓
Milder, 2007 (108)	Netherlands	331 adenomatous polyps cases /306 controls (19-75 y)	178-item FFQ	One non-fasting (LC/MS)	Total lignan intake↑ Fibre↑, Fruits↑ Whole-grain wheat bread↑ Nuts and seeds↑ Wine↑, Beer↓	Age↑ Weight↓ Smoking↓ Frequency of defecation↓
Lampe, 1999 (109)	USA	49 men/49 women (18-37 y)	5-day diet records	Three 24-h urine (GC/MS)	Fruit ↑ Fibre ↑ Fibre from grains ↑	
Hultén, 2002 (92)	Sweden	248 breast cancer cases /492 controls (Mean: 51 y)		One plasma (TR-FIA)		smoking↓
Vanharanta, 2002 (110)	Finland	100 men (mean 59 y)	5-day food record	12 h fasting (TR-FIA)	Soluble fibre↑ Insoluble fibre↑ Fruit and berries↑ Vegetables↑, Cereals ↑	

TR-FIA= time-resolved fluoroimmunoassay
GC/MS= gas chromatography/mass spectrometry
LC/MS= liquid chromatography/mass spectrometry
FFQ= food frequency questionnaire

American study, the correlation coefficient for two enterolactone concentration measurements obtained from 12-hour fasting blood collected on subsequent days was 0.84 (99).

A Danish study comprising six healthy postmenopausal women consuming three low-lignan standardised meals during three separate 24-h periods reported large within-day variations in serum enterolactone (coefficient of variance (CV) of 31 %) among women consuming a low-lignan diet. They concluded that fasting blood samples are preferable to non-fasting samples (111). However, a low within-day variation in enterolactone was observed in a study of pigs eating either a high-lignan or low-lignan diet (69).

Some studies have investigated the number of days required to estimate the average enterolactone concentration. The Danish study by Hausner *et al.* concluded that five random blood samples were required to precisely estimate the average concentration with 50 % and 80 % confidence interval (CI) (111). The study evaluating 20 university students in Finland concluded that three fasting samples were needed to estimate the serum concentration within ± 50 % with 80 % CI (104).

Determinants of enterolactone concentrations

Dietary and lifestyle determinants of enterolactone concentration have been described for several populations (107,99,106,108) (Table 2). A Danish nested case-control study comprising 857 postmenopausal women and using a food frequency questionnaire showed that whole grains, cabbage, leafy vegetables, and coffee were major dietary determinants of enterolactone concentration. They also identified Body Mass Index (BMI), smoking, and frequency of bowel movements as predictors of enterolactone concentration (106). Kilkkinen *et al.* found associations between enterolactone concentration and age, BMI, smoking, and consumption of vegetables among 1,212 Finnish women using a 38-food item questionnaire (107). In American women, vegetables, fibre, coffee, and alcohol were positively correlated with plasma enterolactone (99).

The capacity of the gut microflora is recognised as a very influential factor in the formation of enterolactone from plant lignans (112). The use of antibiotics is known to reduce the amount of bacteria in the gut and, subsequently, the concentration of enterolactone in the blood. A Finnish study indicated that the blood concentrations of enterolactone can be influenced by antibiotic use up to 12-16 months before blood collection; the women that had used antibiotics at least once during the preceding year (41 % of all women) had on average 15 % lower enterolactone concentration (113). The concentration was associated with the number of treatments and the time from the last administration. In a subsequent paper, Kilkkinen *et al.* reported that lignan intakes (assessed via 24-h dietary recall) were positively associated with serum enterolactone concentrations ($r=0.19$, $P<0.0001$) among those who had not used antibiotics during the preceding year. However, serum enterolactone concentration increased only slightly with increasing lignan intake in antibiotic users (114). However, another study found no differences in enterolactone concentration (24-h urine samples) between antibiotic users during the preceding year (43 % of the women) and nonusers (115). Other factors (for example, fermented milk) may also potentially influence the ability of colon bacteria to convert plant lignans to mammalian lignans.

Genetic factors

Several polymorphisms in genes that are proposed to influence the action of enterolactone and estrogens have been studied, for example hormone-metabolizing genes: CYP17 (encodes an enzyme that catalyzes a rate-limiting step in estradiol biosynthesis), CYP19 (encodes the enzyme aromatase that are responsible for conversion of androgens to estrogens) and COMT (involved in the conjugation and inactivation of catechol estrogens). Several carcinogen-metabolizing genes (e.g., CYP1A1, CYP1B1) have also been studied (22).

However, studies examining the modulating effect of the genotype on the association between enterolactone and breast cancer are lacking. A case-controls study with 267 premenopausal cases and 573 controls showed a significant modifying effect of a polymorphism in CYP17 on the association between enterolactone concentration and breast cancer. High enterolactone concentration was only significantly related to decreased breast cancer risk in A2A2 carriers (116), an allele that has been suggested to enhance the amount of endogenous hormone levels (117).

Because enterolactone is hypothesized to interact with the ERs and compete with the binding of estrogens to ERs, is it plausible that functional polymorphisms in these genes may modulate the cancer risk associated with enterolactone concentrations.

Estrogen receptors

Breast cancer is a heterogeneous disease with a range of different characteristics. Sporadic breast cancer is often subdivided based on molecular expression and the different subtypes show different clinical behaviours (118). A central theme in this subdivision is the expression pattern of ERs.

Estrogens stimulate the growth and development of both normal and neoplastic mammary epithelial cells, mainly by interacting with ERs. ERs are members of the nuclear hormone receptor family, which are ligand-activated transcription factors that regulate gene expression. Two types of ERs have been identified: ERα and ERβ. These ERs are encoded by different genes and seem to have diverse effects. The ERα gene (ESR1) is located on chromosome 6q25.1 and spans ~300 kb, including 8 exons. The ERβ gene (ESR2) is located on chromosome 14q23.2 and spans ~61 kb, including 8 exons. Several alternative promoters have been characterized, and several transcript variants have been identified for these genes (119). For example, ERβcx that is identical to the full-length ERβ protein except that exon 8 is replaced by 26 unique amino acid residues.

Structure and function

ER activation usually occurs when the ligand enters the cell and binds to its receptor. In addition to estrogens, ERs are activated by several synthetic ligands, for example tamoxifen, as well as several natural estrogen-like compounds. The ER-ligand complex undergoes a conformational change, forming a dimer and binds to the estrogen responsive element (ERE) in the promoters of target genes together with co-regulating proteins, resulting in transcriptional activation.

The receptors consist of several functional domains. The DNA-binding domain show high sequence homology (97 %) between ERα and ERβ, and as expected both receptors bind ERE with similar affinity. The ligand-binding domain also shares a high degree of homology between the receptors, and ERα and ERβ show similar affinity for estradiol, but different affinity to other ligands. The activation function 1 (AF-1) sequence, a ligand-independent transactivation domain, is structurally and functionally different between the two receptors (15 % similarity in amino acid sequence). Due to the different exon 8, the variant ERβ, ERβcx, lack the residues important for ligand binding and has a poor binding affinity for estradiol (120).

Estrogen receptor status in normal tissues and in tumours

ERα expression in breast tumours is used to select patients most likely to benefit from endocrine therapy and to provide prognostic information. The majority of patients with ERα positive (+) tumours benefit from endocrine therapy, and the predictive accuracy increases when progesterone receptor (PR) expression is also

taken into account, since PR is a downstream marker of functional ER signalling. ERα (+) tumours are associated with favourable prognostic features, including evidence of tumour cell differentiation and a lower rate of cell proliferation (121).

ERβ was recently identified (122); although a treatment predictive value of ERβ has been reported (123) (124,125) it is still not used in the clinical setting. The biological function of ERβ is not yet fully understood, but it is suggested that ERα and ERβ differ in their interaction with other proteins (126,127) and that ERβ may have a negative modulatory effect on ERα (128,126). Microarray analysis has shown that several genes were induced by either ERα or ERβ (129). Furthermore, the function of ERα is suppressed by dimerisation with ERβ (130), and the variant protein ERβcx show preference towards heterodimerisation with ERα, rather then with ERβ, and inhibit the binding of ERα to DNA (120).

Both ERα and ERβ are expressed in normal human epithelial breast cells (131). While there is high expression of ERβ in normal breast tissue, the expression of this receptor appears to be reduced during carcinogenesis. ERβ has, therefore, been suggested to be a tumour suppressor.

Although very simplified, some types of phytoestrogens appear to prefer ERα (lignans) (79), while others prefer ERβ (isoflavones and coumestrol) (83,84).

Enterolactone, plant foods and estrogen receptor-defined breast cancer

Since the effect of estrogens is mediated through ERs, the association between estrogen-related factors and breast cancer may differ depending on the ER status of the tumours. Conditions that influence estrogen levels are believed to have a stronger influence on the development of ER (+) breast cancer (132) and, consequently, factors related to reproduction tend to be associated with increased risk of ER (+) but not ER (-) tumours (133). In fact, the increase in breast cancer incidence in Western countries over the last 20 years is mainly due to an increase in ERα (+) tumours (134). Users of menopausal hormone therapy often have ERα (+) tumours. However, increased incidence of breast cancer in long-term menopausal hormone therapy users is generally not associated with an increase in cancer mortality.

Studies have also shown that the association between breast cancer and obesity or alcohol consumption vary depending on hormone receptor status (132,135). Furthermore, studies have suggested that ERα (+) tumours are related to high fat intake (136,132,137).

Few studies have investigated the association between fibre or plant foods and ER-defined breast cancer. Increased consumption of fruit and vegetables were associated with a decreased risk of ERα (-) breast cancer in the Nurses Health Study (138). A Danish cohort study found that fruit and vegetable intakes were differentially associated with ERα (+) and ERα (-) breast cancer, with a protective effect observed only for ERα (-) tumours (139). On the other hand, a large case-control study found that the inverse associations between fruit and vegetables and breast cancer were stronger with ERα (+) tumours than ERα (-) tumours (140). A large Swedish cohort study did not detect any heterogeneity in breast cancer risk associated with fibre intake across hormone receptors status (32).

A Danish study observed a protective effect of enterolactone with ERα (-) tumours (95). However, that study suffered from limited power as only 80 cases exhibiting ERα (-) tumours were identified. Results from the Swedish mammography cohort showed no heterogeneity in the association between lignan intake and breast cancer across ERα/PR subtypes (141). However, a large prospective study conducted in France, including 1180 cases with known ER and PR status, showed that the inverse association between risk of breast cancer and lignan intake (calculated from food records) was only observed among ERα (+) and PR (+) tumours (142).

Development of estrogen receptor-positive and -negative breast tumours

ER (+) and ER (-) tumours have been suggested to represent biologically different cancer types (143), stemming from different carcinogenesis and tumour development pathways, although the factors responsible for carcinogenesis remain rather unclear.

According to the cancer stem cell hypothesis, mammary carcinogenesis is driven by tumour stem cells, which are derived from mutated stem or progenitor cells (144). Dontu (2004) demonstrate three subtypes of breast cancer based on cell of origin. Type 1 arises from the most primitive ERα (-) stem/early progenitor cells. This subtype is poorly differentiated, is more aggressive and has a poor prognosis. Less than 10 % of the cells are positive for ERα in these tumours and the risk for ERα (-) breast cancer is not increased by menopausal hormone therapy and not reduced by treatment with antiestrogens. Type 2 also arises from ERα (-) stem/early progenitor cells. The mutations allow for differentiation of a subset of tumour cells into ERα (+) cells, and these tumours contain 10-80 % ERα (+) cells. However, menopausal hormone therapy does not significantly increase the risk. Type 3 arise from ERα (+) progenitor cells, are more well-differentiated and have the best prognosis. These tumours respond to antiestrogen treatment and menopausal hormone therapy increases the risk of breast cancer development in this subset of patients. The stem cell model provides an explanation for heterogeneity within tumours. It also provides an explanation for the heterogeneity in different patients (143).

Another model (the clonal evolution hypothesis) suggests that the tumour phenotype is primarily determined by acquired genetic and epigenetic events (145). According to this model, ER (-) breast cancer can evolve from ER (+) breast cancer (146).

Genetic variation in the estrogen receptor genes

Genetic variation in the ER genes might alter the expression of the genes and the functions of the proteins. Very few SNPs in the ER genes have been reported to be functional and examined in relation to breast cancer. Although a rare single missense coding variant, rs934077, causing the G77S substitution has been reported, there is no common non-synonymous SNP (a SNP that cause a change in the amino acid sequence) in any of the genes; however, SNPs in non-coding sequences may play important regulatory roles. Polymorphisms in these genes have been suggested to influence breast cancer risk, and common polymorphisms in these genes, both in the introns and promoter regions of these genes, could influence their transcription and the function of the encoded proteins. Previous studies have sought to find an association between polymorphisms in ERS1 or ERS2 and breast cancer with variable results (31-33).

The variant allele (C) of the SNP rs2234693 (also known as IVS1-401 T/C, IVS1-397T/C, c.454-397 T>C and PuvII) has been found to produce a binding site for the transcription factor B-myb, and has been suggested to increase transcription or produce other ERα isoforms with different properties (147). A few studies have investigated the interaction between rs2234693 and risk factors for breast cancer. We are not aware of any study that has investigated the modulatory effect of polymorphisms in the ER genes on the association between enterolactone and breast cancer. It has, however, been suggested that this polymorphism modify the effect of estrogen on breast cancer risk. A study from a population-based cohort study in the Netherlands found that women with the combination of the T allele and high estradiol levels had significantly higher risk of breast cancer compared to women with the CC genotype and low estradiol levels (148). Wedrén *et al*. observed that the association between a ESR1 haplotype (containing the T allele of rs2234693) and breast cancer grew stronger when they considered women with high BMI (149). As BMI

is the most important determinant of estrogen levels in postmenopausal women, the results would indicate that the ESR1 variation is more influential in the presence of higher estrogen levels.

Zhang *et al*. examined the joint effect of an ESR2 polymorphism and endogenous estrogen exposure on breast cancer risk in Chinese women. They found a 3-4 fold increased risk among women with the CG or GG genotype of rs1256054 (a synonymous polymorphism in exon 7) combined with high levels of estrogens (150). Recently, a large study with 5,789 breast cancer cases and 7,761 controls found an interaction between a haplotype consisting of four SNPs and estrone levels (151).

AIMS

General aim

The general aim of this doctoral project was to clarify the previous finding that high-fibre diets are associated with decreased risk of breast cancer among postmenopausal women of the Malmö Diet and Cancer cohort, by examining if the intake of plant foods and blood concentrations of enterolactone were associated with risk of breast cancer, especially according to estrogen receptor status in tumours and genetic variation.

Specific aims

1. To study the variation of fasting and non-fasting enterolactone concentrations in middle-aged healthy women in order to determine the reliability of one sample for use in epidemiological studies (Paper I)

2. To examine the association between intake of plant foods and incidence of breast cancer (Paper II)

3. To examine the association between intake of plant foods and incidence of estrogen receptor α and β-defined breast cancer (Paper II)

4. To identify the major dietary and lifestyle determinants of enterolactone concentrations (Paper III)

5. To examine the association between enterolactone concentrations and risk of breast cancer (Paper III)

6. To examine the association between enterolactone concentrations and risk of estrogen receptor α and β-defined breast cancer (Paper III)

7. To examine the association between polymorphisms in estrogen receptor α and β genes and risk of breast cancer (Paper IV)

8. To examine if polymorphisms in estrogen receptor α and β genes modify the association between enterolactone concentrations and risk of breast cancer (Paper IV)

SUBJECTS AND METHODS

Malmö Diet and Cancer cohort

The Malmö Diet and Cancer (MDC) cohort is a prospective population-based cohort study conducted in the third largest city in Sweden (about 250 000 inhabitants). The MDC study was initiated and planned in collaboration with the International Agency for Research on Cancer (IARC), Lyon, France, the Swedish Cancer Society, the Swedish Medical Research Council, and the Faculty of Medicine, Lund University, Sweden. The main goal of the MDC study was to investigate the relationship between diet and cancer incidence, with a focus on whether a diet high in fat and total calories but low in vitamin and fibre increases the risk of certain cancers, such as breast, colon, rectum, pancreas, ovary, endometrium and prostate cancer (152).

Baseline examinations were undertaken from March 1991 to October 1996. In 1991, all men and women living in Malmö and born between 1926 and 1945 (n=53,325) were invited to participate in the study. In May 1994, the background population was extended to include women born between 1923 and 1950 and men born between 1923 and 1945 (n=74,138). A letter of invitation and an information campaign including advertisements in newspapers and television, as well as posters in public places was used for recruitment. No economic compensation was offered, only gifts such as T-shirts, pens and plastic bags. Limited Swedish language skills and mental incapacity were the only exclusion criteria.

Participants visited the study centre on two occasions. During the first visit, trained project staff provided participants (in groups of 6-8) with information on the background and aims of the project, gave detailed instructions about the dietary data collection procedure, and distributed the dietary questionnaire and menu book and the extensive lifestyle and socioeconomic questionnaire. The questionnaire included items such as: (1) education, (2) previous and current occupations, including physical and psychological conditions at work, (3) country of birth, (4) social network and support, (5) leisure-time physical activity, (6) sleeping habits, (7) use of tobacco, (8) alcohol habits, (9) previous and current diseases, (10) food habit change in the past, (11) diseases among close relatives, (12) regularly used medications, (13) oral contraceptive use, and (14) reproductive factors (age at menarche, age at menopause, parity, breast feeding and miscarriage).

Nurses conducted anthropometrical measurements (weight, height, waist and hip circumference, lean body mass and body fat mass), measured the blood pressure and collected blood samples. At the second visit (after approximately 2 weeks), individual interviews were conducted by trained dietary interviewers to complete the diet history and to check the correctness of the completed questionnaires.

Of the background population (n=74,138), 17 individuals could not be identified and 3,017 died or moved before they received the first letter of invitation. Among those that came to the study, 224 died before they completed the baseline examination and 1,975 were excluded due to language problems and mental incapacity. Of the 68,905 individuals that were classified as eligible, 28,098 individuals (11,063 men and 17,035 women) completed the questionnaire, anthropometric measurements and dietary assessment, and thus comprised the cohort. In total, 5,082 joined spontaneously (community invitation) and 23,016 were recruited by invitation letters, resulting in a participation rate of 40.8 % (38.3 % for men and 42.6 % for women). A more detailed description of the cohort has been described elsewhere (152,153). Ethical permission for the study was obtained (LU 51-90).

In 1993, the MDC cohort became an associated member of the European Prospective Investigation into Cancer and Nutrition (EPIC) organised by the IARC, WHO, Lyon, France (154). EPIC covers a large cohort of 520,000 individuals from 23 centres in 10 Western European countries (155).

Reproducibility of the questionnaire

In 1994, 232 randomly selected participants were invited to complete the questionnaire a second time, three weeks after the first invitation. In total, 211 responded, and 209 were complete participants. The agreement between answers in the two questionnaires was high for most factors. Kappa coefficients among the women were as follows: born in Sweden (yes/no), 1.00; education (3 categories: ≥10 years in school, ≥ 12 years in school, university/university college), 0.84; smoking (3 categories: never smoker, current smoker, ex-smoker), 0.94; alcohol (3 categories: nothing in the last year, something last year but not last month, something in the last month), 0.77; dietary change in the past (yes/no), 0.68 (156).

The biobank

The blood from each participant was separated into fractions; 10 ml was used for the serum sample (stored at -80ºC) and 30 ml was used to purify mononuclear leucocytes (-140ºC), granulocytes (-80ºC), erythrocytes (-80ºC) and plasma (-80ºC). In August 1995, buffy coats were stored (-140 ºC) instead of mononuclear leucocytes and granulocytes. Instrument variability, yield, and the purity of the blood cell fraction and quality of the stored blood fraction are presented in the quality control program (157,158).

Representativity of the cohort

The frequency of individuals born outside of Sweden was lower compared to that in a study of the same population with higher response rates (74.6 %). The data also suggest that the MDC cohort comprises a higher frequency of individuals with better health, although the socio-demographic structure and prevalence of smoking and obesity were equivalent in the two studies. Mortality was higher in non-participants during both recruitment and follow-up. Prior to recruitment, non-participants displayed a lower cancer incidence. During recruitment, cancer incidence was higher among non-participants (153).

Design of the reliability study (I)

Study participants came for blood collection five times during May-June 2005 (three non-fasting and two fasting samples). For the fasting samples the participants had to fast beginning at 2400 the day before. Dietary intake was assessed by the same method used during the baseline examinations. A questionnaire was used to collect information on lifestyle factors, regular drug use and occasional drug use during the previous three months. Current drug usage was recorded in the menu book.

Study populations

Reliability study (Paper I)

One hundred women born between 1940 and 1950 were randomly selected from the MDC cohort and were invited by mail and telephone. Out of these, 26 responded and 21 women participated in all blood collections.

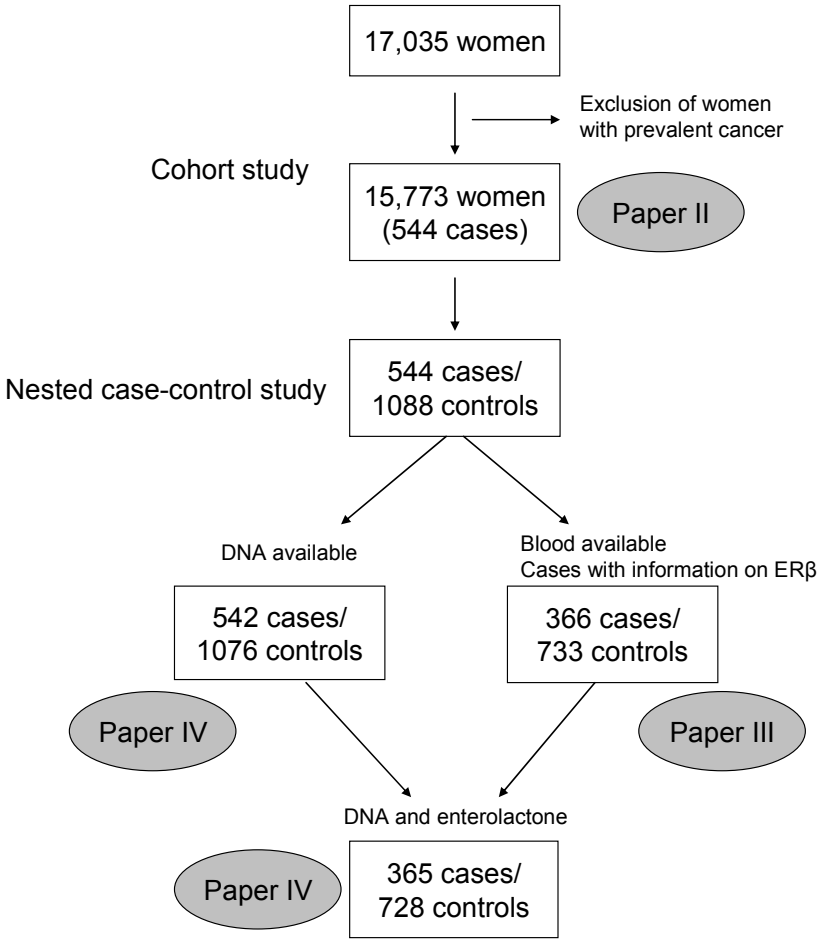

Figure 3. Study populations for Paper II-IV

Cohort Study (Paper II)

Women with prevalent cancers at baseline, except cervical cancer *in situ*, were excluded, leaving 15,773 women for analysis. At the end of follow-up in Dec 31st 2004, during an average of 10.3 years, 544 women in this population were diagnosed with incident invasive breast cancer. The participants contributed personal time from the date of enrolment until the time of invasive breast cancer diagnosis, death, migration from Sweden, or the end of follow-up (Dec 31st 2004), whichever occurred first.

Nested case-control study (Paper III-IV)

For each of the 544 cases (until the end of follow-up, Dec 31st 2004), two controls were selected and matched based on age (± 3 months) and date of blood collection (± 1 month) from the women in the cohort at risk (alive, without breast cancer and living in Sweden) at the time of case diagnosis.

In Paper III, a sub-sample of the nested case-control study was included, i.e., cases with determined ERβ status based on immunohistochemical analyses of tumour tissues (n=370) and their two matched controls (n=740). Blood samples were not available for 11 subjects, resulting in 366 cases and 733 controls for analysis of the enterolactone concentration.

In Paper IV, only individuals with available DNA for genotyping were included (542 cases and 1076 controls). Information on both polymorphisms and enterolactone concentrations were obtained for 365 cases and 728 controls.

Dietary assessment methodology

Dietary assessment method

The dietary assessment method used was a modified dietary history method specifically designed for use in the MDC study (159) that combined: (1) a 7-day menu book that collected information on cooked lunches and dinner meals and cold beverages (including alcoholic beverages), medications, natural remedies and dietary supplements, and (2) a 168-item dietary questionnaire covering foods regularly consumed during the past year, including coffee, tea, cacao, buns and cakes, biscuits and rusks, soft bread and crisp bread, edible fat, food consumed on bread, other cheeses, raw vegetables, breakfast cereals with yoghurt and milk, fruits and berries, fruit soups, sweets, ice cream and snacks. The participants estimated frequencies of food intake, and usual portion sizes were assessed using a booklet containing 48 photographs. Thereafter, during a 1-hour interview, the participants were asked questions about food choices, food preparation practices and portion sizes of the foods collected in the menu book (using a more extensive book of photos). An interviewer also checked the menu book and dietary questionnaire for overlapping information, as well as for very high reported intakes. Furthermore, the total consumption of broader groups of certain types of food was checked for reasonable values, as it is easy to over-report one's intake when many different food items constitute a broader food groups (e.g., bread, crisp bread, fruits and vegetables). A total of 17 interviewers conducted the dietary interviews.

The choice of diet methodology was guided by the need to assess the total diet, with a specific emphasis on total fat in an elderly urban population. The eating habits of this group were expected to be fairly regular and commonly include cooked sit-down meals.

In September 1994, the routines used for coding dietary data at data entry were slightly altered in order to shorten the interview time (160). No changes in the dietary methodology were made as such (menu book and questionnaire). This alteration of routines during data entry included standardised (instead of individualised) portion sizes for a few food items, and standardized (instead of individualized) recipes for a few dishes. Evaluation of this change in routines did not reveal any major influence on the ranking of individuals, but absolute intakes of energy and fat seemed to be affected (160).

The average daily intake of foods (grams per day) was calculated based on the information available in the menu book (and interview) and the questionnaire. Food intake was converted to nutrient intake data using the MDC Food and Nutrient Database, which was specifically developed for the MDC study and originated from PC KOST2-93 of the Swedish National Food Administration.

Relative validity

The relative validity of the dietary method was examined among 105 women and 101 men; 18 days of weighed food records (3 days every second month) collected over one year was used as the reference method. Energy-adjusted Pearson correlations for women were between 0.50 and 0.80 for most of the food groups (Table 3) (161,162).

Reproducibility

The reproducibility of the dietary method was examined among 126 men and 115 women who were administered the dietary method one year apart. Energy-adjusted Pearson correlations in women were between 0.60 and 0.80 and for most of the food groups (Table 3) (163).

Dietary variables

The plant foods (g/day) (Paper II-III) were categorised into 10 food groups: *vegetables*, *fruits and berries*, *fruit juices*, *boiled potato*, *fried and deep fried potatoes*, *cereals* (grains, cereals and flours), *low-fibre bread* (< 6 % fibre for soft bread, < 10 % for crisp bread, and < 10 % for biscuits and rusks), *high-fibre bread* (≥ 6 % fibre for soft bread, ≥ 10 % for crisp bread, and ≥ 10 % for biscuits and rusks), *rice and pasta* and *nuts*. *Fruit, berries and vegetables* were also combined into one variable.

The other dietary variables were *fibre* (g/day) (Paper I-III), *energy* (kcal), *alcohol* (g/day), and percentage of non-alcohol energy (E %) contributed by *protein*, *fat*, and *carbohydrates*. *Fermented milk*, *wine*, *beer*, *spirits*, *tea*, and *coffee* (Paper III) were selected based on their known lignan content (37) or their potential influence on the microflora (164).

Table 3. Relative validity and reproducibility of the dietary method among women

	Relative validity[1]	Reproducibility[2]
Protein	0.53	0.54
Fat	0.69	0.52
Carbohydrates	0.70	0.49
Fibre	0.69	0.70
Fruits	0.77	0.81
Vegetables	0.53	0.76
Potatoes	0.51	0.43
Cereals	0.73	0.61
Bread	0.58	0.65
Rice and pasta	0.24	0.23
Milk	0.84	0.70
Meat products	0.92	0.57
Fish	0.70	0.22
Added fats and oils	0.66	0.46
Beer	0.74	0.68
Wine	0.63	0.79
Spirits	0.67	0.83

[1]Energy-adjusted Pearson correlations between daily intakes estimated by the MDC method and the reference method.
[2]Energy-adjusted Pearson correlations between daily intakes estimated by the MDC method at the start and the end of a 12 month period.

Methodological variables

A variable was created for the *seasons of data collection*: winter (December - February), spring (March-May), summer (June-August) and fall (September-November). Moreover, a variable was created for the 17 *interviewers* that conducted the dietary interviews, as well as a variable for the *method version* for coding of dietary data (old/new).

Lifestyle and background variables

The extensive standardised questionnaire was used to collect information on lifestyle, demographic, socioeconomic and reproductive factors.

Age and sex

Information on *age* and *sex* was obtained via the person-identification number. In Sweden, each person is assigned a unique 10-digit number at birth: six digits indicate the date of birth and one identifies the gender.

Education

Educational status was categorised based on the type of education attained: elementary, primary and secondary, upper secondary, further education without a degree, and a university degree.

Nationality

Country of birth was divided into born in Sweden versus not born in Sweden.

Physical activity

Leisure time physical activity was obtained from a list of different physical activities in the questionnaire (18 items) that were adapted from the Minnesota Leisure Time Physical Activity Instrument (165). Participants were asked to estimate the number of minutes per week, and for each of the four seasons, they spent peforming 18 different physical activities. The duration of each activity was multiplied by an intensity factor, creating a leisure time physical activity score. The score was separated into quartiles. *Household activities* were divided into four categories: 0-9 h, 10-19 h, 20-29 h, and ≥ 30 hours per week.

Smoking

Smoking habits were categorised into current smokers (including irregular smoking), ex-smokers and non-smokers.

Alcohol

Alcohol consumption was divided into four categories. Individuals with no consumption of alcohol in the menu book and who indicated no consumption of alcohol during the previous year in the questionnaire were

categorised as zero-consumers. The other subjects were categorised into three groups according to their alcohol consumption: <15 g alcohol per day (low), 15-30 g (medium) and >30 g (high).

Reproductive factors

Parity was aggregated into five answer categories: no children, one child, two children, three children, and four or more children. *Age at birth of first child* was valuated based on the participant's year of birth and the year of first child's birth, and was divided into four categories: no children, <24 years, 24-30 years, and more than 30 years. *Age at menopause* was divided into five categories: <45 years, 45-49 years, 50-55 years, > 55 years, and those who reported no cessation of menses at baseline.

Use of medication

Information on *current use of menopausal hormone therapy* (yes/no) and *current use of antibiotics* (yes/no) were obtained from 1) two open-ended questions "Which prescription drugs and non-prescription drugs do you use on a regular basis?" in the questionnaire and 2) the open-ended item for listing drug use in the 7-day menu book recorded during the consecutive days after blood collection (166). All pharmacologic agents were classified according to the 1997 version of the Anatomic Therapeutic Chemical classification system. Menopausal hormone therapy was defined by the codes G03C (estrogens), G03D (progestagens) and G03F (progestagens and estrogens in combination). Antibiotics were defined by the codes A07AA, D01A, D01BA, D06A, D07C, G01AA, J01, J02AA, J04AB, L01D, R02AB and S01A.

Food habit change in the past

Dietary change in the past (yes/no) was derived from the questionnaire item "Have you substantially changed your eating habits because of illness or for some other reasons?"

We previously observed that individuals who reported past food habit change were more obese than individuals with no reported change, and that a higher education level, living alone, not being born in Sweden and being retired were associated with past food habit change. Moreover, different health-related behaviours were associated with past food habit change, including smoking and alcohol habits (167).

Anthropometric measures

Weight (kg) was measured to the nearest 0.1 kg by trained project staff members using a balance-beam scale with subjects wearing light clothing and no shoes; *height* (cm) was measured with a fixed stadiometer calibrated in centimetres. Trained project staff members also measured the *waist* (midway between the lowest rib margin and iliac crest) and *hip circumferences* (horizontally at the level of greatest lateral extension of the hips).

BMI (kg/m^2) was defined as weight (kg) divided by height in square meters (m^2). A three-category variable was created with subjects categorised as normal weight (<25), overweight (25-29) or obese (≥ 30) according to the WHO classification (168).

Bioelectric impedance analysis was used to estimate body composition (BIA 103 JRL Systems, MI, USA; single-frequency analyzer). The algorithm used to estimate body fat from impedance was supplied by the manufacturer. *Body fat percentage* was calculated based on the estimated body fat mass.

Ascertainment of breast cancer cases (II-IV)

It has been compulsory since 1958 for every clinician to report all newly detected cancer cases to the cancer registry. The cancers are also reported separately by pathologists and cytologists for surgically removed tissues. The regional cancer registries in Sweden (Stockholm-Gotland, Uppsala-Örebro, Linköping, Lund-Malmö, Göteborg and Umeå) send information regarding newly detected cases and corrections concerning previous reports once a year to the National Cancer Registry. These registries have been estimated to be almost 100% complete.

Tumours are classified according to the International Classification of Diseases (ICD) system. Incident breast cancer cases (ICD-7 code 170) were identified via linkage of the study population (using the unique personal identification number) to the Swedish Cancer Registry and the Southern Swedish Regional Cancer Registry. The National Tax Board provided information on vital status.

Laboratory analyses of enterolactone (I, III, IV)

Enterolactone concentrations (nmol/L) were determined at the Folkhälsan Research Center in Helsinki, Finland, by time-resolved fluoroimmunoassay, as previously reported (169,170). Plasma samples (200µl) were hydrolyzed overnight with sulphatase and β-glucuronidase (pure enzyme), and unconjugated enterolactone was extracted with diethyl ether. Sample extracts were diluted in assay buffer and analyzed by AutoDELFIA 1235 Automatic Immunoassay System. ^{3}H-estradiol glucuronide was used as an internal standard for the plasma controls. The mean values for the recovery results were employed for all samples to correct for losses during hydrolysis and extraction. All analyses utilizing the automatic instrument were carried out in duplicate. Three different control samples were used for each 96-well plate analyzed.

Estrogen receptor status assessments (II-III)

Tissue microarray

The tissue microarray technique is a well-documented method used for high-throughput tissue screening and is now the preferred method for evaluating large tumour materials (171,172). To construct tissue microarrays, two 0.6 mm tissue cores were collected from each tumour block and arranged in a recipient block using a manual tissue arrayer (Beecher Inc., Sun Prairie, WI, USA). Each recipient block contained approximately 200 cores corresponding to 100 patients.

Immunohistochemistry

For immunohistochemical analyses, slides of 4 µm thick sections were cut from the recipient block. The slides were then processed in an automatic immunohistochemistry staining machine. ERα (prediluted anti-ER 6F11, Ventana, Tucson, Arizona, USA) and ERβ (1:25 EMR02, Novocastra, Newcastle upon Tyne, UK) (173) antibodies were used. The ERβ antibody was previously validated by comparing the results obtained using immunohistochemical method with those obtained by western blot (125).

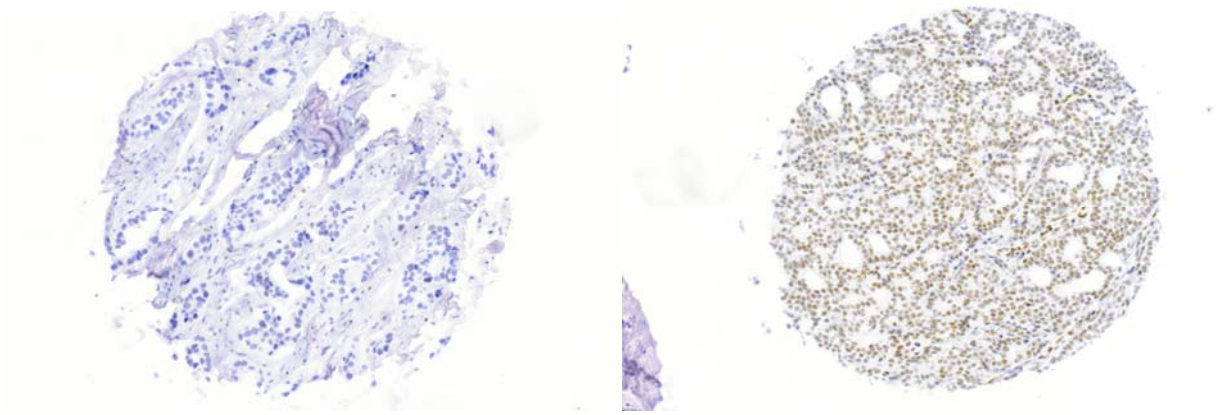

Figure 4. ERβ negative (left) and ERβ positive (right) breast cancer after immunohistochemical staining with the ERβ antibody

Tumours were grouped into categories according to the expression of ERα and ERβ (0-1%, 2-10%, 11-50% and 51-100% positive nuclei). One individual evaluated the hormone receptor status of all breast tumours in a standardised manner. All arrays were independently evaluated twice and, in the case of a discrepancy, a third examination was performed followed by a final decision. The tumours were classified as positive (+) and negative (-) using the clinically established cut-off value of 10% positive nuclei (Figure 4). ERα and ERβ were estimated for 450 and 370 cases, respectively. For the remaining cases, adequate tumour samples were not available, either due to surgery performed at other hospitals or an insufficient amount of tumour material left for histopathological evaluation. These cases were categorised as unknown (Table 4).

Table 4. Distribution of estrogen receptor α and β status among the 544 breast cancer cases

	ERα (-)	**ERα (+)**	**ERα unknown**	**Total**
ERβ (-)	31	155	2	188
ERβ (+)	21	159	2	182
ERβ unknown	6	78	90	174
Total	58	392	94	544

Selection of single nucleotide polymorphisms and genotyping (Paper IV)

An extensive search using the Entrez SNP and HapMap genome browser databases were conducted for SNPs in and immediately up- and downstream of the ESR1 and ESR2 genes (Chr.6: 152,069,700 – 152,516,519 and Chr.14: 63,769,607 – 63,850,218).

Only SNPs that destroy or create CpG-sites were selected. The SEQUENOM MassARRAY® Designer software was used for multiplex SNP analysis design. A total of 103 SNPs in ESR1 and 18 SNPs in ESR2 were analyzed on 47 or 95 control samples (depending on the heterozygosity frequency).

DNA was extracted from granulocytes or buffy coat fraction. The genotyping was conducted on a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (SEQUENOM MassArray) using iPLEX reagents according to the manufacturers protocol. Monomorphic SNPs, SNPs in strong linkage disequilibrium (LD) to another SNP, or where the assay failed were excluded, leaving 77 SNPs in ESR1 and 11 SNPs in ERS2. These SNPs were analysed on all cases and controls with extracted DNA (542 cases and 1076 controls).

The number of SNPs from ESR1 was limited by excluding SNPs located > 4kb upstream of the gene. SNPs with minor allele frequency (MAF) <0.20 among controls, as well as SNPs with missing values >5% were also excluded. The polymorphisms were tested for adherence to Hardy-Weinberg equilibrium, and SNPs with p<0.05 among controls were excluded. After these exclusions, 19 SNPs in ESR1 and 5 SNPs in ESR2 remained (Figure 5, Figure 6). All 5 SNPs in ESR2 (rs915057, rs1269056, rs1256033, rs3020450 and rs3020443) were kept (Figure 6).

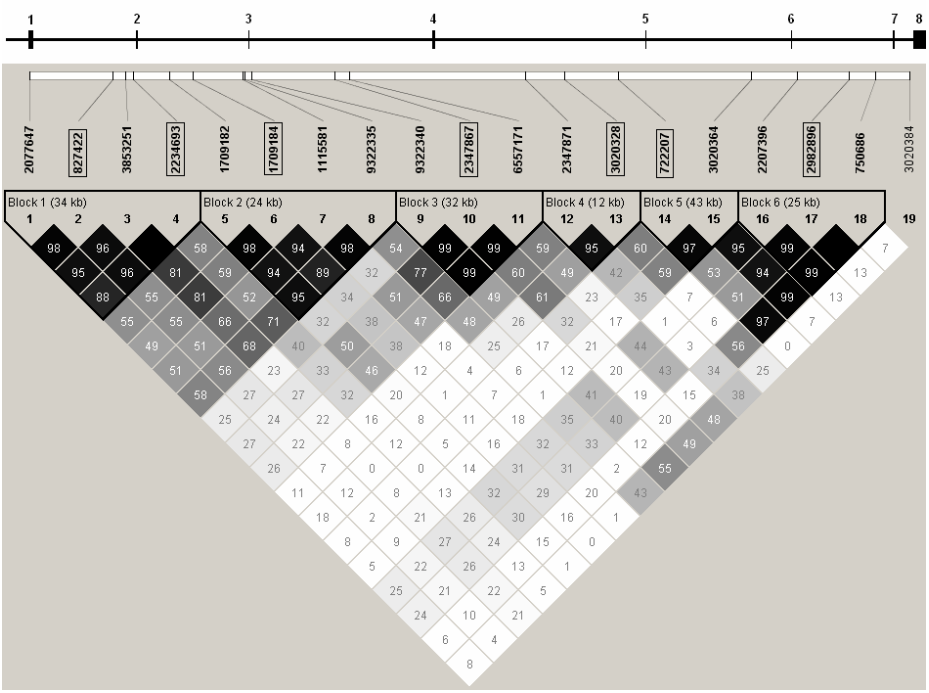

Figure 5. Schematic illustration of ESR1 and LD structure of the 19 initially selected SNPs. The eight exons in the gene are marked with vertical lines. D' values are shown, and the seven SNPs selected for this study are marked.

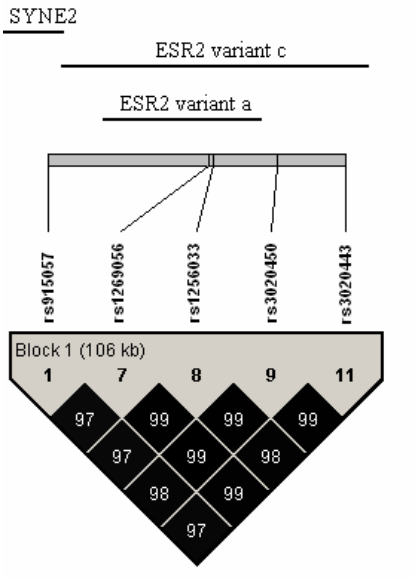

Figure 6. LD structure of the five selected SNPs in ESR2. The location of ESR2 transcription variant a (accession code NM_001437), ESR2 transcription variant c (NM_001040276) and SYNE2 (spectrin repeat containing nuclear envelope 2) are also shown.

To capture the main sequence variation in ESR1, we used the Haploview software "Tagger" for selection among the remaining SNPs. Pair-wise LD between markers was estimated using the Haploview program and LD blocks were created according to the "solid spine of LD" method (174). Six SNPs in ESR1 were selected from each block that was created (rs827422, rs1709184, rs2347867, rs3020328, rs72207 and rs2982896). In addition, rs2234693 was also included as this SNP has been suggested to be functional (147) (Figure 5).

Statistical methods

The nominal level of α=0.05 (i.e., $p < 0.05$) were considered significant in two-sided tests.

Statistical power

Sufficient statistical "power" (80% and α=0.05) to observe a risk gradient of 1 to 1.75 over quintiles of nutrient intake (given a true risk gradient of 1 to 3, and a validity coefficient of the dietary variable of 0.6) was reached when 283 cancer cases accumulated (in 1999).

Paper I

Reliability of enterolactone concentrations (Aim 1)

All calculations were performed for $\log_e$ transformed enterolactone concentrations. Pearson's correlation coefficients were determined for fasting and non-fasting samples. Estimates of between-subject, within-subject, total variance as well as ICC with 95% CIs were obtained using SPSS with reliability analysis. The number of repeated measurements (in different weeks) needed to estimate the individual underlying within-month enterolactone mean concentration was calculated using the following formula: $n=Z*CV_{intra-individual}/D)^2$ with a precision (D) of 50 % and with 80 % CI (Z=1.28).

Paper II

Plant foods and breast cancer (Aim 2)

A small number (0.01) was added to the dietary variables to handle zero intakes before transformation in order to normalise the distribution. The food variables were energy-adjusted by regressing the food variable for total energy intake, and individuals were divided into quintiles depending on their residual ranking (residual method). Due to low consumption (i.e., the 25th percentile included zero consumers) of several food groups (fruit juice, fried and deep fried potato, rice and pasta and nuts), individuals who reported no consumption were categorised as zero consumers, and the other individuals were divided into tertiles according to their energy-adjusted food group intakes (residual method).

Hazard ratios (HR) and 95% CI for food groups associated with breast cancer were estimated using Cox proportional hazard regression. Person-years of follow-up were included as the time-dependent variable. The basic model was adjusted for age and stratified based on the method version utilized (to avoid any undue influence of the different coding strategies used for the dietary data). The multivariate model was extended to adjust for total energy, season of data collection, and diet interviewer, as well as established breast cancer risk factors and potential confounding variables: weight, height, educational status, smoking habits, leisure time physical activity, hours of household activities, alcohol consumption, age at menopause, parity, and current use of menopausal hormone therapy. Missing values among these variables were recorded as separate categories to avoid exclusion of individuals from the analysis.

We also repeated the analysis excluding individuals below 55 years of age at baseline (to further ensure that the analyses only included postmenopausal women) and those with reported dietary change in the past (in order to exclude individuals more likely to have unstable food habits) (175). The test of trend was calculated as a linear trend over the quintile values.

Plant foods and estrogen receptor α- and β-defined breast cancer (Aim 3)

HRs of ER-defined breast cancer (i.e., ERα (+), ERα (-), ERβ (+), ERβ (-) as well as ERα (+)/ERβ (-) and ERα (+)/ERβ(+)) associated with each food group were estimated using the basic and the multivariate model. The number of cases of ERα-/ERβ- (n=31) and ERα-/ERβ+ (n=21) was very small, and therefore not considered separately. Individuals with the opposite (e.g., ERα (-) when ERα (+) were examined) and unknown receptor status participated in the follow-up until date of diagnosis.

Paper III

Determinants of enterolactone concentration (Aim 4)

All statistical tests were performed for $\log_e$ transformed enterolactone concentrations to normalise the distribution. Partial correlations between enterolactone and dietary variables ($\log_e$ transformed) and body composition measures were computed among the controls (n=733), controlling for age, date of blood collection, diet method version, and total energy. Thereafter, in an exploratory analysis, all food group variables, alcohol variables (wine, beer and spirits), tea, coffee, body fat %, smoking status, educational status, and leisure time physical activity were included in a linear regression model, followed by stepwise backwards elimination of variables with $p>0.10$. The model was adjusted for age, date of blood collection, method version, and total energy.

Enterolactone and risk of breast cancer (Aim 5)

The individuals were divided into quartiles based on the distributions of enterolactone among controls. Conditional logistic regression was used to estimate the odds ratios (ORs) and 95 % CI for enterolactone using the lowest quartile as the reference.

The lowest quartile of enterolactone may include both women with long-term low plasma enterolactone concentrations as well as those with low concentrations due to the use of antibiotics. Because of the expected heterogeneity of the lowest quartile, additional analyses using the 2nd quartile as the reference group were performed.

We evaluated the associations with and without adjustment for potential confounders (i.e., weight, height, educational status, smoking habits, leisure time physical activity, household activities, alcohol habits, age at menopause, parity, age at birth of first child, and current use of menopausal hormone therapy). Missing values among these variables were recorded as separate categories to avoid the exclusion of individuals.

We conducted a series of sensitivity analyses. The analyses were repeated excluding individuals below 50 years of age at baseline (in an attempt to exclude premenopausal women) and individuals diagnosed with breast cancer within one year after blood collection (to ensure that the enterolactone concentrations were not influenced by preclinical cancer). We also repeated the analyses excluding individuals who reported dietary change in the past (to include only those individuals more likely to have stable food habits and, consequently, more stable long-term enterolactone concentrations).

By using the ICC of 0.48 from the reliability study (175), the risk estimates were corrected for within-person variation in plasma enterolactone concentrations over one month according to the following equation $OR_{observed}=OR_{true}^{ICC}$ (176).

Enterolactone and estrogen receptor defined breast cancer (Aim 6)

Unconditional multinomial logistic regression was used to estimate OR and 95 % CI for enterolactone separately for each type of ER-defined breast cancer (ERα (-), ERα (+), ERβ (-), ERβ (+), ERα (+)/ERβ (-) and ERα (+)/ERβ (+)) adjusted for age, date of blood collection, weight, height, education, and current use of menopausal hormone therapy. Breast cancers that were ERα(-)/ERβ(-) (n=31) and ERα(-)/ERβ(+) (n=21) were not analyzed separately due to the small number of cases. We divided the individuals into two categories (low/high) based on the median enterolactone concentration obtained for controls. We used Wald's test to examine the heterogeneity between different tumour subgroups regarding their association with enterolactone.

Paper IV

Polymorphisms in ESR1 and ESR2 and breast cancer (Aim 7)

Associations (ORs and 95 % CI) between genotypes and breast cancer were calculated using conditional logistic regression, with individuals homozygous for the major allele as the reference category. The trend across genotypes using an additive genetic model was also examined. The analyses were repeated for only those women greater than 50 years of age, as well as those greater than 55 years of age, at baseline (in an attempt to exclude premenopausal women, as hormonal risk factors have been suggested to be more closely related to postmenopausal breast cancer). We used Haploview (174) to examine the association between haplotypes formed by the seven SNPs in ESR1 and five SNPs in ESR2 with breast cancer.

Polymorphisms in ESR1 and ESR2, enterolactone and breast cancer (Aim 8)

Unconditional logistic regression was used to calculate the ORs and 95% CI for the joint effect of enterolactone (low vs. high) and ESR1 and ESR2 genotypes (homozygous for the major allele vs. heterozygous or homozygous for the minor allele), adjusting for the matching variables (age and date of blood collection). The associations were evaluated with and without adjustment for potential confounders (i.e., weight, height, educational status, smoking habits, leisure time physical activity, alcohol habits and current use of menopausal hormone therapy).

Interactions between enterolactone and SNPs were evaluated by introducing a multiplicative term for the specific genotype and the dichotomous enterolactone variable. The analyses were repeated excluding women below 50 years of age at baseline and individuals diagnosed with breast cancer within 1 year after blood collection (to ensure that the enterolactone concentrations were not influenced by preclinical cancer).

RESULTS

Paper I

Reliability of enterolactone concentrations (Aim 1)

The median interval between the first and last blood collection (i.e., non-fasting samples) was 33 days (range 19-46 days), with a median of 7 days (range 4-23 days) between the two fasting measurements. A higher Pearson's correlation coefficient was estimated for fasting blood samples (0.74) than for non-fasting samples (0.62 for the first and second non-fasting samples, and 0.42 for the first and last non-fasting samples (Figure7)).

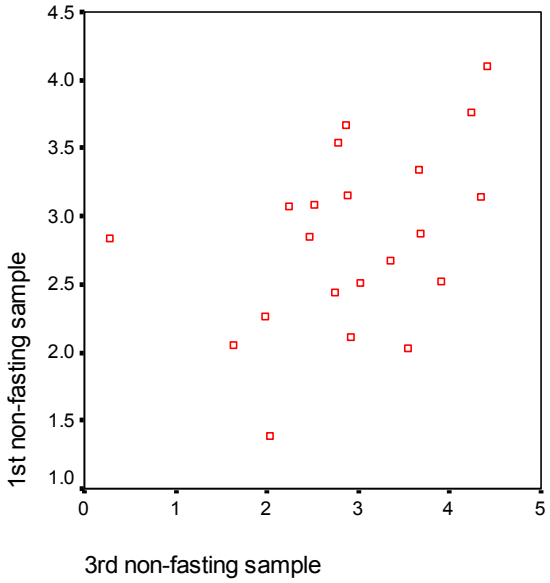

Figure 7. Enterolactone concentrations (ln-transformed, nmol/L) for the first and last blood collection for the 21 individuals

The fasting samples also demonstrated a slightly higher ICC than the non-fasting samples (0.66 vs. 0.48), with a slightly lower within-subject CV (59% vs. 71%). Three samples for the fasting and four samples for the non-fasting would be required to estimate the mean within-month concentration (to within ± 50% with 80% CI) for individuals in this population (Table 5).

Table 5. Enterolactone concentrations, CV, ICC and number of samples to estimate mean levels of enterolactone

Variables	Non-fasting	Fasting
Enterolactone concentration (nmol/L), median	18	22
CV (within-subject)	71%	59%
CV (between-subject)	67%	89%
ICC (95% CI)	0.48 (0.22-0.72)	0.66 (0.35-0.84)
Number of samples[1]	4	3

[1]The number of samples required to estimate individual mean level of enterolactone concentration within 50% of the true mean and with 80% CI

Paper II

Plant foods and breast cancer (Aim 2)

The highest quintile of fibre was associated with a non-significant decreased risk of breast cancer (HR, 0.82; 95% CI, 0.62-1.09) compared to the lowest quintile (Table 6). Among the food groups, the highest quintile of high-fibre bread was significantly associated with a 25 % decreased risk of breast cancer incidence compared to the lowest quintile. The association with fibre was more pronounced when restricting the analysis to women with no reported dietary change in the past (HR, 0.66; 95% CI, 0.46-0.95, for the highest vs. the lowest quintile). Moreover, fruit, berries and vegetables were also associated with a decreased risk when the analysis was restricted to individuals with no reported dietary change in the past (HR, 0.65; 95% CI, 0.46-0.93, for the highest vs. the lowest quintile, p for trend=0.03).

Table 6. Hazard ratios (95% CI) of invasive breast cancer by quintiles of intake[1]

	Quintiles of intake					
Food groups	1	2	3	4	5	*p*
Fibre						
All women						
Median intakes (g/day)	12	16	18	21	26	
Cases/follow-up	107/31058	103/31812	105/32386	124/32859	105/33805	
HR (95 % CI)	1.00	0.91 (0.70-1.20)	0.89 (0.68-1.18)	1.02 (0.78-1.34)	0.82 (0.61-1.09)	0.40
Exclusion of women with dietary change in the past						
Cases/follow-up	94/26281	74/26386	84/25383	87/23279	52/19607	
HR (95 % CI)	1.00	0.76 (0.56-1.04)	0.88 (0.65-1.20)	0.98 (0.72-1.33)	0.66 (0.46-0.95)	0.21
Fruit, berries, vegetables						
All women						
Median intakes (g/day)	118	288	370	463	629	
Cases/follow-up	119/31175	101/31976	96/32501	117/32927	111/33341	
HR (95 % CI)	1.00	0.81 (0.62-1.06)	0.72 (0.55-0.95)	0.86 (0.66-1.13)	0.78 (0.59-1.03)	0.18
Exclusion of women with dietary change in the past						
Cases/follow-up	98/25797	81/25369	72/25292	84/24146	56/20332	
HR (95 % CI)	1.00	0.83 (0.62-1.12)	0.70 (0.51-0.95)	0.84 (0.62-1.14)	0.65 (0.46-0.93)	0.03
High-fibre bread						
All women						
Median intakes (g/day)	0	9	19	34	65	
Cases/follow-up	115/31101	108/31736	98/32394	115/33067	108/33622	
HR (95 % CI)	1.00	0.87 (0.67-1.13)	0.74 (0.56-0.97)	0.82 (0.63-1.07)	0.75 (0.57-0.98)	0.04
Exclusion of women with dietary change in the past						
Cases/follow-up	90/24690	108/25582	70/24749	77/23394	72/22522	
HR (95 % CI)	1.00	0.84 (0.63-1.14)	0.71 (0.52-0.98)	0.81 (0.59-1.10)	0.75 (0.54-1.03)	0.09

[1]Adjusted for season of data collection, diet interviewer, method version, age, total energy, weight, height, educational status, smoking habits, leisure time physical activity, hours of household activities, alcohol consumption, age at menopause, parity and current use of menopause hormone therapy.

Plant foods and estrogen receptor-defined breast cancer (Aim 3)

High intake of fried potato was associated with an increased risk of ERβ (-) breast cancer (p for trend=0.01). The highest tertile of fried potato was associated with an increased risk compared to individuals reporting no consumption of fried potato (HR, 1.61; 95% CI, 1.11-2.33). None of the other food groups were significantly associated with any of the ER-defined breast cancers (Table 7). The results were similar when adjusting for potential confounding factors. However, there was a tendency towards an inverse association between high-fibre bread and ERα (+) breast cancer and ERβ (+) breast cancers.

Table 7. The risk trends of estrogen receptor characterized invasive breast cancer by quintiles[1] of total fibre and plant foods[2]

	All women (n=544)	ERα (+) (n=392)	ERα (-) (n=58)	ERβ (+) (n=182)	ERβ (-) (n=188)
Food groups	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Fibre	0.98 (0.92-1.04)	0.98 (0.91-1.05)	1.09 (0.90-1.31)	1.00 (0.90-1.11)	1.01 (0.91-1.12)
Vegetables	0.98 (0.92-1.04)	1.01 (0.94-1.09)	0.99(0.82-1.18)	0.96 (0.87-1.07)	1.05 (0.95-1.17)
Fruits and berries	0.99 (0.93-1.05)	0.98 (0.91-1.05)	1.07 (0.89-1.28)	1.03 (0.92-1.14)	0.95 (0.86-1.05)
Fruit, berries, vegetables	0.97 (0.92-1.03)	0.99 (0.92-1.06)	1.03 (0.86-1.24)	1.00 (0.90-1.10)	0.98 (0.89-1.09)
Fruit juices[1]	1.04 (0.97-1.12)	1.05 (0.97-1.14)	1.03 (0.83-1.29)	1.08 (0.95-1.22)	0.97 (0.65-1.43)
Boiled potatoes	0.99 (0.93-1.05)	0.98 (0.91-1.05)	1.07 (0.89-1.29)	1.01 (0.91-1.12)	0.98 (0.88-1.08)
Fried potatoes[1]	1.05 (0.98-1.13)	1.08 (0.99-1.17)	0.97 (0.77-1.21)	0.96 (0.85-1.09)	1.17 (1.03-1.31)
Low-fibre bread	1.02 (0.96-1.08)	1.00 (0.93-1.07)	1.02 (0.85-1.22)	0.99 (0.89-1.10)	0.96 (0.87-1.06)
High-fibre bread	0.96 (0.90-1.02)	0.95 (0.89-1.02)	1.03 (0.85-1.24)	0.93 (0.84-1.03)	0.98 (0.88-1.08)
Cereals	0.97 (0.92-1.03)	0.97 (0.90-1.04)	0.92 (0.77-1.11)	0.95 (0.85-1.05)	1.03 (0.93-1.14)
Rice and pasta[1]	0.99 (0.92-1.07)	0.99 (0.90-1.08)	1.15 (0.91-1.45)	1.07 (0.94-1.22)	1.02 (0.90-1.16)
Nuts[2]	1.01 (0.94-1.09)	1.02 (0.93-1.12)	1.18 (0.95-1.47)	0.99 (0.87-1.14)	1.09 (0.96-1.23)

The HR (95% CI) indicates the risk associated with an increase of each intake category.
[1]Due to many zero-reporters four variables are not quintile intake but have four categories
[2]Adjusted for age and method version

Paper III

Median enterolactone concentration was 14.5 nmol/L (range 0.3-334) among the cases and 16.1 nmol/L (range 0.3-115) among the controls (Figure 8).

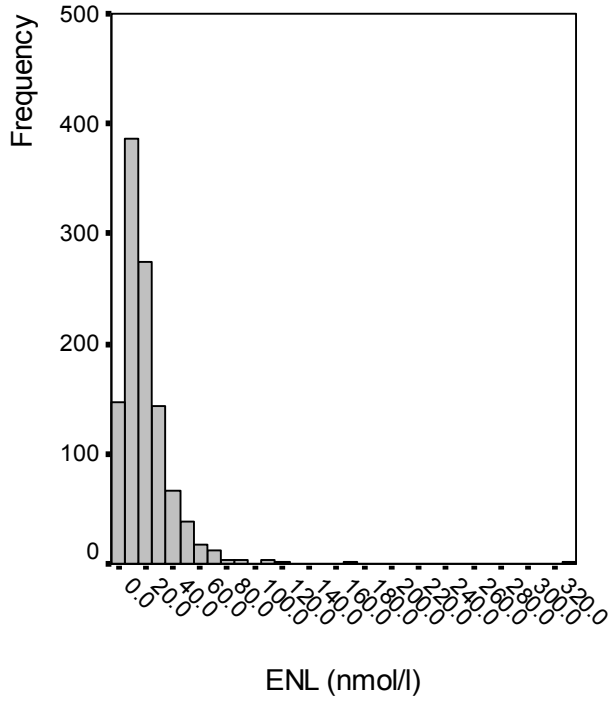

Figure 8. Distribution of enterolactone concentrations among the subjects (n=1099)

Determinants of enterolactone concentrations (Aim 4)

Individuals with a university degree had a 55 % higher enterolactone concentration than individuals with elementary degree as their highest educational level (p<0.001). Non-smokers had a 38 % higher enterolactone concentration than smokers (p<0.001). Individuals consuming 15-30 g of alcohol/day had a 54 % higher enterolactone concentrations than zero-consumers (p=0.012). In contrast, a tendency towards lower enterolactone concentrations was observed among individuals consuming more than 30 g of alcohol/day compared with those consuming 15-30 g/day (p=0.06). Leisure time physical activity was positively associated with enterolactone concentration (p for trend=0.02) (Table 8).

Table 8. Mean plasma concentrations of enterolactone (nmol/L) according to participant characteristics among the controls (n=733)

Participant characteristics	No. of subjects	Mean	95 % CI	p-value[1]
Age (years)				0.23
44-50	162	22.2	19.6-24.8	
50-55	159	20.1	17.5-22.7	
55-60	170	19.4	16.8-21.9	
60-65	132	18.3	15.5-21.2	
65-70	65	16.1	12.1-20.2	
70-75	45	22.1	17.0-27.2	
Educational status				<0.001
Elementary	305	17.2[2]	15.4-19-1	
Primary and secondary	226	20.0	17.8-22.2	
Upper secondary	46	19.9	15.1-24.7	
Further education without a degree	46	20.5	15.7-25.2	
University degree	106	26.6[2]	23.4-29.8	
Smoking status				<0.001
Current smokers	200	16.2[2]	13.9-18.5	
Ex-smokers	197	19.2	16.9-21.5	
Non-smokers	335	22.4[2]	20.6-24.1	
Alcohol consumption				0.007
Zero-consumers	55	14.9[2]	10.5-19.3	
< 15 g/day	561	20.0	18.6-21.3	
15-30 g/day	100	22.9[2]	19.7-26.2	
> 30 g/day	16	13.5	5.4-21.6	
BMI (kg/m^2)				0.001
≤ 25	367	22.0[2]	20.3-23.7	
25-30	245	18.7	16.7-20.8	
>30	121	15.4[2]	12.5-18.3	
Leisure time physical activity				0.005
Quartile 1	183	16.7[2]	14.3-19.1	
Quartile 2	182	20.1	17.7-22.5	
Quartile 3	168	22.9[2]	20.4-25.4	
Quartile 4	197	19.9	17.6-22.2	
Current use of menopausal hormone therapy				0.91
No	553	19.6	18.3-21.0	
Yes	133	19.7	16.9-22.6	
Current use of antibiotics				0.79
No	718	19.7	18.5-20.9	
Yes	15	24.2	15.8-32.6	

[1]Test of differences in means ($\log_e$-transformed) using the general linear model, adjusted for age and date of blood collection, [2]indicates statistically significant differences in means at the 0.05 level using Tukey's multiple comparison test within each variable

Enterolactone concentrations were positively correlated with fibre intake (r=0.25, p<0.001). Several food sources of fibre: vegetables, fruit and berries, high-fibre bread, and nuts (r=0.10-0.17, p<0.01), as well as

alcohol intake (especially wine) were positively correlated with the enterolactone concentration when examined in separate models. BMI and body fat % were negatively correlated with the enterolactone concentration in plasma (Table 9).

In the multivariate linear regression model, fruit and berries and high-fibre bread were significantly positively associated, and body fat % was significantly negatively associated with enterolactone concentrations. A university degree and the 3rd quartile of leisure time physical activity were positively associated, while current smoking was negatively associated with enterolactone concentrations. Fruit and berries, high-fibre bread, body fat %, a university degree, current smoking, and 3rd quartile of physical activity explained 11.4 % of the variation in enterolactone concentrations.

Table 9. Partial correlations between enterolactone concentrations, and dietary variables (grams per day) and body composition among the controls (n=733)

	Separate models[1]		Multivariate model[2]	
Variables	Coefficient	*p*-value	Coefficient	*p*-value
Fibre	0.25	<0.001	0.22[3]	<0.001
Vegetables	0.10	0.008		N.S.
Fruit and berries	0.17	<0.001	0.13	<0.001
Fruit juices	0.04	0.26		N.S.
Boiled potatoes	-0.003	0.95		N.S.
Fried and deep fried potatoes	-0.04	0.23		N.S.
Cereals	0.05	0.19		N.S.
Low-fibre bread	-0.06	0.10		N.S.
High-fibre bread	0.16	<0.001	0.13	0.001
Rice and pasta	0.06	0.09		N.S.
Nuts	0.10	0.008	0.07	0.06
Fermented milk	0.03	0.36		N.S
Tea	0.05	0.15		N.S.
Coffee	0.03	0.41		N.S.
Alcohol	0.10	0.007		
Wine	0.09	0.01		N.S
Beer	0.07	0.07		N.S
Spirits	0.003	0.94		N.S
BMI	-0.11[4]	0.004		
Body fat (%)	-0.14[4]	<0.001	-0.14	<0.001

[1]Adjusted for age, date of blood collection, method version, and total energy. [2]Stepwise backwards linear regression model with smoking status, educational status, and leisure-time physical activity included in the model, adjusted for age categories, date of blood collection, method version, and total energy; N.S.= eliminated variables with p>0.10. [3]Model not adjusted for fibre-rich foods. [4]Adjusted for age and date of blood collection.

Enterolactone and breast cancer (Aim 5)

A tendency towards reduced breast cancer risk was observed for the highest compared to the lowest quartile of enterolactone concentrations (OR, 0.81; 95 % CI, 0.55-1.20, *p* for trend=0.14 in the multivariate model). The heterogeneity of the lowest quartile (likely including individuals with low concentrations due to antibiotic use) was apparent as individuals in the lowest quartile reported similar fibre intakes as individuals in the 2nd quartile (Table 10). In additional analyses (not shown in Paper III), when stratifying for fibre intake, we observed decreased breast cancer risk associated with high enterolactone concentrations only among those individuals with high fibre intake (Table 11).

When the analyses included only women over 50 years of age and cases with breast cancer diagnosis more than 1 year after the baseline examinations, a significant association was observed for the 4th quartile compared to the 2nd quartile (p=0.04). When enterolactone concentrations were divided into two categories

(low/high), high enterolactone concentrations were associated with decreased breast cancer risk compared to low enterolactone concentrations (OR, 0.75; 95 % CI, 0.58-0.98). When the risk estimates were corrected for the within-person variation in enterolactone over one month (ICC=0.48), the risk estimates were lower. The OR of 0.75 for high enterolactone compared to low enterolactone concentrations was 0.55 after correction.

Table 10. Odds ratios (95 % CI) [1] for breast cancer across quartiles of plasma enterolactone concentration

	Quartiles of enterolactone concentration				
	1 (<8.4 nmol/L)	2 (8.5-16.1 nmol/L)	3 (16.1-26.1 nmol/L)	4 (>26.1 nmol/L)	*p*
All women					
Fibre intake (g/day)	17.6	17.2	18.9	19.7	<0.01
Cases/controls	100/183	104/183	78/185	84/182	
OR (95 % CI)	1.00	1.03 (0.70-1-52)	0.70 (0.46-1.06)	0.81 (0.55-1.20)	0.14
Women over 50 years + exclusion of cases with diagnosis within 1 year					
Cases/controls	73/145	80/133	66/134	57/136	
OR (95 % CI)	1.00	1.25 (0.78-2.00)	0.89 (0.55-1.45)	0.73 (0.45-1.19)	0.13
OR (95 % CI) (2^{nd} quartile as ref)	0.80 (0.50-1.28)	1.00	0.71 (0.44-1.15)	0.59 (0.35-0.97)	0.04[2]

[1]Adjusted for weight, height, educational status, smoking habits, leisure time physical activity, household activities, alcohol habits, age at menopause, parity, age at birth of first child, and current use of menopausal hormone therapy. [2]Test of trend from quartile 2 to 4

Table 11. Odds ratios (95 % CI) for breast cancer across quartiles of plasma enterolactone concentration in strata of fibre intake

	Quartiles of enterolactone concentration				
	1 (<8.4 nmol/L)	2 (8.5-16.1 nmol/L)	3 (16.1-26.1 nmol/L)	4 (>26.1 nmol/L)	*p*
Low fibre intake					
Median fibre intake (g/day)	14.3	14.6	15.2	15.1	
Cases/controls	52/103	58/112	33/90	36/65	
OR (95 % CI)[1]	1.00	1.04 (0.65-1.65)	0.71 (0.42-1.19)	1.06 (0.62-1.81)	0.70
OR (95 % CI)[2]	1.00	1.03 (0.61-1.68)	0.74 (0.42-1.29)	0.96 (0.53-1.72)	0.57
High fibre intake					
Median fibre intake (g/day)	22.3	21.9	22.6	22.8	
Cases/controls	48/80	46/71	45/95	48/117	
OR (95 % CI)[1]	1.00	1.11 (0.66-1.88)	0.77 (0.46-1.28)	0.69 (0.42-1.13)	0.07
OR (95 % CI)[2]	1.00	0.97 (0.54-1.73)	0.58 (0.33-1.01)	0.63 (0.36-1.10)	0.04

[1]Adjusted for age and date at baseline. [2]Adjusted for weight, height, educational status, smoking habits, leisure time physical activity, household activities, alcohol habits, age at menopause, parity, age at birth of first child, and current use of menopausal hormone therapy

Enterolactone and estrogen receptor-defined breast cancer (Aim 6)

The reduced breast cancer risk associated with high enterolactone concentrations was only observed for ERα (+) (OR, 0.75; 95 % CI, 0.58-0.98) and ERβ (-) tumours (OR, 0.60; 95 % CI, 0.42-0.84). The risk was significantly different for ERβ (-) and ERβ (+) tumours (*p* for heterogeneity = 0.04) (Table 12).

The reduced risk for ERβ (-) tumours in association with high enterolactone concentrations was also observed when analyzing ORs across quartiles of enterolactone (1.00 (ref), 1.13, 0.50, 0.78; *p* for trend=0.045), whereas the risk of ERα (+) tumours across quartiles of enterolactone concentrations did not reach significance (1.00 (ref), 1.06, 0.72, 0.78; *p* for trend=0.07). The risk for ERβ (+) tumours across quartiles of enterolactone concentrations (1.00 (ref), 1.03, 0.99, 0.90; *p* for trend=0.63) and ERα (-) tumours (1.00 (ref), 1.21, 0.71, 1.11; *p* for trend=0.90) were not statistically significant.

As the majority of tumors were ERα (+), high enterolactone concentrations were also associated with a decreased risk of ERα (+)/ERβ (-) tumours (OR, 0.58; 95 % CI, 0.35-0.85).

Table 12. Odds ratios (95 % CI)[1] for estrogen receptor defined breast cancer across plasma enterolactone concentration (low/high)

	Enterolactone concentration		
	Low (<16.1 nmol/L)	High (>16.1 nmol/L)	*p* heterogeneity
All women			
Cases/controls	204/366	162/367	
OR (95 % CI)	1.00	0.75 (0.58-0.98)	
ERα status			
ERα (-) tumours			
Cases/controls	27/366	25/367	
OR (95 % CI)	1.00	0.86 (0.48-1.54)	
ERα (+) tumours			
Cases/controls	175/366	134/367	
OR (95 % CI)	1.00	0.73 (0.55-0.97)	0.62[2]
ERβ status			
ERβ (-) tumours			
Cases/controls	113/366	72/367	
OR (95 % CI)	1.00	0.60 (0.43-0.85)	
ERβ (+) tumours			
Cases/controls	91/366	90/367	
OR (95 % CI)	1.00	0.94 (0.67-1.33)	0.04[3]
Combined ERα and ERβ status			
ERα (+)/ERβ (-) tumours			
Cases/controls	93/366	59/367	
OR (95 % CI)	1.00	0.59 (0.41-0.86)	
ERα (+)/ERβ (+) tumours			
Cases/controls	82/366	75/367	
OR (95 % CI)	1.00	0.90 (0.63-1.28)	0.08[4]

[1]Adjusted for age, date of blood collection, weight, height, smoking status, educational status, and current use of menopausal hormone therapy. [2] Test of heterogeneity between ERα (-) and ERα (+) tumours. [3] Test of heterogeneity between ERβ (-) and ERβ (+) tumours. [4] Test of heterogeneity between ERα (+)/ERβ (-) and ERα (+)/ERβ (+) tumours

PAPER IV

Polymorphisms in estrogen receptor genes and breast cancer (Aim 7)

No statistically significant association with breast cancer was observed for the selected polymorphisms with the exception of individuals with the rs2982896 CT genotype, who demonstrated a statistically significant 26 % increased risk of breast cancer compared to those with the CC genotype.

There was no major difference when the analyses were restricted to women over 50 years of age, as well as over 55 years of age at baseline. Moreover, there were no associations with any of the haplotypes formed by the seven SNPs in ESR1 and five SNPs in ESR2.

Polymorphisms in estrogen receptor genes, enterolactone and breast cancer (Aim 8)

High enterolactone concentrations were statistically significantly associated with a decreased risk of breast cancer among individuals harbouring the ESR1 rs1709184 AA genotype, rs2347867 AG and GG genotype, rs72207 GA and AA genotypes, and ESR2 rs3020443 AC and CC genotypes.

A tendency towards an interaction between a polymorphism in intron 3 (rs2347867) and enterolactone was observed (p=0.07). For this polymorphism, no association between breast cancer and enterolactone concentration was found among those homozygous for the major allele (AA) (p=0.93), whereas an inverse association among carriers of the minor allele was found (AG and GG) (p=0.007).

Table 13. Odds ratios (95 % CI) for breast cancer according to the joint effect of enterolactone concentration (low/high) and genotype of ESR1 and ESR2 polymorphisms [1]

		Enterolactone concentration					
		Low		High			
	Genotype	Cases/ controls	OR (95 % CI)	Cases/ controls	OR (95 % CI)	*p*	*p* interaction
ESR1							
rs827422	TT	54/83	1.00 (ref)	41/84	0.75 (0.45-1.25)	0.27	0.87
	TC/CC	144/268	0.82 (0.45-1.22)	112/262	0.65 (0.43-0.98)	0.13	
rs2234693	TT	69/101	1.00 (ref)	44/99	0.65 (0.41-1.04)	0.08	0.37
	TC/CC	133/258	0.75 (0.52-1.09)	116/267	0.63 (0.43-0.92)	0.25	
rs1709184	AA	77/118	1.00 (ref)	54/139	0.60 (0.39-0.91)	0.02	0.14
	AG/GG	125/241	0.80 (0.56-1.14)	105/226	0.71 (0.49-1.03)	0.48	
rs2347867	AA	70/156	1.00 (ref)	72/156	1.03 (0.69-1.53)	0.93	0.07
	AG/GG	131/202	1.44 (1.01-2.07)	86/209	0.92 (0.63-1.34)	0.007	
rs3020328	TT	109/196	1.00 (ref)	83/188	0.79 (0.56-1.13)	0.19	0.85
	TC/CC	93/163	1.03 (0.73-1.45)	76/176	0.78 (0.54-1.11)	0.15	
rs72207	GG	96/176	1.00 (ref)	80/169	0.87 (0.60-1.25)	0.38	0.35
	GA/AA	106/181	1.07 (0.76-1.52)	78/195	0.73 (0.51-1.05)	0.04	
rs2982896	CC	112/221	1.00 (ref)	88/204	0.85 (0.61-1.20)	0.31	0.43
	CT/TT	90/136	1.31 (0.92-1.86)	72/156	0.91 (0.64-1.31)	0.07	
ESR2							
rs915057	CC	68/127	1.00 (ref)	51/115	0.83 (0.53-1.29)	0.38	0.75
	CT/TT	132/229	1.08 (0.75-1.56)	107/245	0.82 (0.56-1.19)	0.07	
rs1269056	GG	66/117	1.00 (ref)	47/112	0.74 (0.47-1.17)	0.19	0.81
	GA/AA	137/244	0.99 (0.69-1.44)	112/252	0.79 (0.54-1.15)	0.12	
rs1256033	GG	58/106	1.00 (ref)	47/108	0.80 (0.50-1.27)	0.33	0.90
	GA/AA	142/246	1.06 (0.72-1.55)	111/250	0.81 (0.55-1.20)	0.08	
rs3020450	GG	90/166	1.00 (ref)	74/163	0.84 (0.57-1.22)	0.33	0.68
	GA/AA	111/195	1.05 (0.74-1.49)	86/201	0.79 (0.55-1.13)	0.10	
rs3020443	AA	102/194	1.00 (ref)	90/189	0.91 (0.64-1.28)	0.58	0.24
	AC/CC	100/167	1.14 (0.81-1.61)	70/175	0.76 (0.53-1.10)	0.03	

[1]Adjusted for age and date of blood collection

The inverse association among individuals with the AG and GG genotype was also observed when analyzing ORs across quartiles of enterolactone (1.00 (ref), 0.98, 0.54, 0.71; *p* for trend=0.03). Carriers of the minor allele (AG or GG) in combination with low enterolactone blood concentrations had an increased risk of breast cancer (OR: 1.44; 95 % CI: 1.01-2.07) compared to individuals homozygous for the major allele (Table 13).

None of the other selected polymorphisms seemed to modify the association between enterolactone and breast cancer. Nevertheless, we found that carriers of the rs2234693 C allele in combination with high enterolactone concentrations had a significantly decreased risk of breast cancer (OR: 0.63; 95 % CI: 0.43-0.92) compared to women with the TT genotype and low enterolactone concentrations. We also observed decreased risks for the combination of high enterolactone concentrations and the rs827422 TC and CC genotypes (OR: 0.65; 95 % CI: 0.43-0.98), as well as the rs1709184 AA genotype (OR: 0.60; 95 % CI: 0.39-0.91).

The results were similar when they were adjusted for potential confounding factors; however, significantly decreased risks were then also observed for the combination of high enterolactone concentrations and the rs2234693 TT genotype (OR: 0.59; 95 % CI: 0.39-0.96), as well as the rs1709184 AG and GG genotypes (OR: 0.66; 95 % CI: 0.45-0.97).

Moreover, the results were similar when women less than 50 years of age at baseline and individuals diagnosed with breast cancer within one year after blood collection were excluded; however, only the joint effect of rs2234693 and high enterolactone remained statistically significant (OR: 0.63; 95 % CI: 0.41-0.96). Although the sample size was reduced by 23 %, the tendency towards an interaction between rs2347867 and enterolactone remained (*p*=0.07).

METHODOLOGICAL CONSIDERATIONS

Observational studies can be perceived as natural experiments; the individuals assigned themselves to the different exposure levels. However, without randomisation, systematic errors may occur, and bias is a major concern in observational epidemiological studies.

There are many sources of random and systematic errors (bias). The probability of misclassification may be equivalent for all groups (non-differential misclassification) or may vary between groups (differential misclassification). A low magnitude of random error (non-differential misclassification) provides high precision; a low magnitude of systematic errors (differential misclassification) provides high validity. The impact of errors and biases can be limited by using an appropriate study design and data collection method, as well as by using appropriate statistical methods during the analysis.

Study design

The prospective design (in which exposure information is collected before diagnosis of disease) has many advantages compared to the retrospective design (in which information is collected after diagnosis). In addition, the cohort study design, in which individuals are free of disease and are followed until diagnosis or the end of follow-up, has many advantages compared to case-control studies. Differential measurement error is of particular concern in case-control studies. Therefore, by using a prospective cohort design, the selection and recall bias that may be introduced into a retrospective case-control study are minimised. These errors are likely to be particularly serious for dietary information (177). For example, because the information regarding diet is collected before cancer diagnosis in a prospective study, the reported exposure is not influenced by the diagnosis.

Cancer status may influence the concentration of biomarkers. Cancer itself may influence the biomarker concentration, or the cancer diagnosis may influence diet and lifestyle habits and, therefore, influence dietary biomarkers. Since the blood was drawn before cancer diagnosis in MDC, the enterolactone concentrations were, therefore, not influenced by the diagnosis. In the sensitivity analyses, we also excluded cases with diagnoses within one year after blood collection (to ensure that enterolactone concentrations were not influenced by preclinical cancer).

As an individual's genotype status is permanent throughout his or her lifetime, cancer diagnosis does not influence the genotype, and gene-disease associations can be analysed using a case-control design. It is, however, very important to carefully select the controls (from the population at risk) so that the probability of selection bias is minimised. However, when examining gene-environmental interactions prospective data regarding exposure are required.

Measurement errors and misclassification of exposure

Diet

Measurement errors in the dietary data may introduce dietary exposure misclassification, leading to attenuated risk estimates and reducing statistical power for detecting significance relationships. This can be further complicated by misreporting in specific groups, leading to systematic errors. Doubly labelled water enabled confirmation that underreporting biases are associated with both social desirability and education in women (178).

All dietary methods are prone to errors, and measurement error for dietary intake is a major concern. Studies have, however, shown larger errors with food frequency questionnaires than with 7-day diaries (179). For instance, in the EPIC-Norfolk cohort, significant relationships between saturated fat and breast cancer were detected using data from detailed food records, but not data from questionnaires (180). To draw conclusions from epidemiological studies examining nutrition and to clarify controversial issues regarding diet and breast cancer, dietary assessment methods of the highest quality are needed (181,182,179). The use of dietary biomarkers is also essential.

The MDC study used a modified dietary history method consisting of a quantitative diet questionnaire and a food diary. In addition, all participants underwent a 1 h dietary interview. This method has been shown to possess high relative validity for the dietary data compared to the reference method. However, there is a correlation in the measurement errors between the diet assessment methods (179).

Energy-adjusted Pearson correlations in women were: 0.69 for fibre, 0.77 for fruit, 0.53 for vegetables, 0.51 for potato, 0.73 for cereals, 0.58 for bread and 0.24 for rice and pasta (161,162). Fruits and berries were mainly reported in the questionnaire. Vegetables intake was reported both in the menu book (dinner and lunch meals) and in the questionnaire (breakfast and between meals). Bread intake was mainly reported in the questionnaire. The dietary method has some advantages compared to other methods. Individuals were required to specify the brand of the bread and the portion sizes in order to accurately specify the fibre content of the bread, and this information was confirmed during the 1 h dietary interview. Fibre was estimated for the whole diet; fibre estimation is (in contrast to food groups) dependent on the quality of the nutrient database.

Flaxseeds and sesame seeds contain very high concentrations of plant lignans. It is very difficult to estimate seed intake because, for example, they may be present in bread. Consequently, we were unable to obtain information on seed intake in the MDC. However, in the reliability study, we specifically asked for information regarding flaxseed consumption. None of the subjects reported flaxseed consumption, but the two subjects exhibiting the highest enterolactone levels reported that they might have consumed flaxseeds in bread. The correlation between fibre and enterolactone became positive ($r=0.36$) upon exclusion of four subjects with the highest enterolactone levels, indicating high intake of low-fibre, lignan-rich foods among these subjects.

Dietary factors can be treated as absolute intakes or as relative intakes. Foods tend to be correlated with total energy (Table 14), and by adjusting for total energy intake you treat the foods as relative intakes. This is relevant as the same amount of food may have a smaller impact in a larger, high-consuming person than a smaller person who requires only a low amount of energy. We used energy-adjusted (by the residual method) food variables, and therefore examined the relative dietary composition, not the absolute amount, of foods. One goal of energy adjustment in nutrition epidemiological studies is to reduce the influence of

reporting errors (183), as energy adjustments have been shown to substantially reduce the measurement error regarding the reported intake of protein or total energy (184).

Table 14. Spearman correlation coefficients between energy, BMI and intake of foods groups among women in the MDC cohort (n=15,773)

	BMI	Alcohol	Fibre	Vegetables	Fruit	Boiled potato	Low-fibre bread	High-fibre bread	Nuts	Meat	Milk
Energy	-0.13	0.14	0.51	0.17	0.17	0.22	0.42	0.10	0.16	0.34	0.26
BMI	-	-0.14	-0.05	-0.01	0.05	0.03	-0.06	-0.01	-0.07	0.06	0.08
Alcohol		-	-0.03	0.12	-0.05	-0.07	-0.07	0.04	0.22	0.11	-0.20
Fibre			-	0.54	0.63	0.09	0.09	0.47	0.12	0.06	0.08
Vegetables				-	0.35	-0.03	-0.10	0.21	0.12	0.09	-0.04
Fruit					-	-0.02	-0.11	0.19	0.05	-0.06	0.01
Boiled potato						-	0.11	-0.02	-0.07	0.23	0.12
Low-fibre bread							-	-0.39	0.001	0.16	0.07
High-fibre bread								-	0.02	-0.03	-0.01
Nuts									-	0.02	-0.08
Meat										-	0.05

Longitudinal studies often repeat the assessment of diet during the follow-up period in order to capture long-term dietary exposure during disease-free years. However, there seems to be only a small effect incurred by the additional measurements (185-187). Furthermore, we suspect that dietary habits are fairly stable in this age group.

Cancers are believed to have long latency periods (20). The time at which, during an individual's lifetime, the exposure is of greatest importance remains unclear. In fact, the dietary habits from the previous decade as well as those at the time of disease presentation could both be important. Therefore, although we have estimated current dietary habits, we are also interested in long-term exposure.

The availability of information on past food habit change is an advantage of the MDC cohort, although it is self-reported. When individuals with reported past dietary change were excluded from the analyses, the observed association between fibre and breast cancer was more pronounced, and the highest quintile was significantly associated with a decreased risk of breast cancer. High intake of fruit, berries and vegetables were also associated with a decreased risk, but only when the analysis was restricted to individuals without past dietary change.

Thus, the reason for observing a more pronounced relationship between diet and breast cancer when excluding individuals with past dietary change (in this study) may be due to unstable diets among those with reported dietary change, and, consequently, misclassification among dietary exposure categories in relation to tumour development. Dietary habits before a lifestyle change may be positively related to the development of disease, an association that may be reversed after the dietary habits are changed. This information may aid in the interpretation of many previous epidemiologic studies that have produced null results, despite animal and ecological studies indicating strong dietary associations.

However, the dietary change variable was very crude, potentially associated with many confounding factors, and prone to errors. For instance, changers were more obese than non-changers (167), and obesity is associated with postmenopausal breast cancer in many studies (188). The high frequency of past food habit change among obese subjects might be due to dieting behaviour. In addition, obesity is often associated with underreporting in studies using self-reported dietary data (189). A self-reported dietary change for an obese individual may, however, indicate a real change in food habits in an attempt to lose weight, which is restrained eating. It may also indicate an overall change in behaviour patterns, including other health

behaviours, as indicated by a previous study. It is also plausible that the perceived dietary change may not always be a real change (190).

Few studies have investigated the clinical implications of underreporting. Rosell *et al.* investigated the effect of misreporting on the association between dietary factors and metabolic syndrome. Underreporters had a higher prevalence of metabolic syndrome (greater waist circumference and higher systolic and diastolic blood pressure) than those individuals who did not underreport. The association between fasting insulin concentrations and intakes of polyunsaturated fats, omega-6 fats, and fat obtained from milk products were stronger in underreporters than in non-underreporters, which indicates that inaccurate data can introduce spurious associations (191).

Macdiarmid *et al.* assessed the relationship between dietary fat and sugar consumption in men and women with different BMIs. Among women, the inclusion of low energy reporters completely reversed the relationship between consumption of high-fat, sweet foods and BMI due to reduced reporting of these products by obese women (192). Inverse associations between BMI and the reported energy intake have been reported to be reduced by approximately 20 % when controlling for low energy dieting (193). Studies have also used more sophisticated approaches and examined social desirability and social approval, or restraining behaviours concurrent with diet assessment. For instance, Hebert *et al.* have reported that important factors in predicting underreporting of energy intake include dissatisfaction with body image and body fat percentage (194,178).

Enterolactone

Misclassification of the enterolactone concentration can occur for various reasons, including degradation during storage time and laboratory errors. Blood was stored at -80ºC. Storage time does not generally seem to influence the concentration of enterolactone, and storage time was not correlated with enterolactone concentration in our study.

Measurement error due to short-term variation or long-term changes in the biomarker being examined within subjects can have a significant impact in an epidemiological study (176). Similar to many other studies, a single measurement for enterolactone was used in our studies. However, the purpose of epidemiological studies is to rank individuals (i.e., classify into quartiles) according to their blood concentration, not to estimate the mean blood concentration of single individuals. We are therefore more interested in reliability (estimated by the ICC) than reproducibility (estimated by the number of samples required to estimate the individual mean level). Therefore, high within-person variability may be a minor problem if the biomarker also shows high between-person variability (195).

In Paper I, we estimated the short-time variation in a small study sample over 1 month. A moderate estimate of the reliability (ICC) was observed using one measurement. However, the accuracy associated with an ICC of this magnitude is comparable to the validation coefficients of foods and nutrients observed in most validation studies. The validation coefficients for nutrients in the MDC methodology are, however, generally higher. Compared to isoflavonoids (another group of phytoestrogens), enterolactone concentrations are suggested to be relatively stable (105).

Because the validity coefficient is dependent on the variance of the true biomarker in the population, the validity study can be applied to the study population if replicates are sampled over the entire period for which the true biomarker is intended to relate. Furthermore, specimen handling, storage and analytical techniques must demonstrate variability in the reliability study similar to the parent study (176).

However, we measured the variation in a small number of women (n=21) in 2005, not during the baseline examination. Although the women in the reliability study were collected from the MDC cohort, variability in enterolactone concentration may have differed between the two time periods, as lifestyle between the two periods may differ, and therefore led to diverse sources of variation. In addition, the youngest women were selected for the reliability study and, as there were few individuals in the reliability study, low precision was encountered. However, a similar variation in enterolactone concentrations, i.e., median, 18 nmol/L (range, 1.3-82) vs. median, 16 nmol/L (range, 0.3-115), was observed in the reliability study and case-control study, respectively. Thus, the reliability study provided a crude estimate of the variation in blood enterolactone that occurs during 1 month. Ultimately, blood (both fasting and non-fasting) should have been collected at the time of baseline examination for a sample of the participants, followed by additional blood samples from these same individuals for several years.

The use of antibiotics can influence the enterolactone concentration for up to one year. Unfortunately, complete information on the history of antibiotic use was lacking. Only 2 % of the individuals in the MDC cohort reported antibiotic use. Moreover, this information was (for the majority of women) recorded after blood collection, and therefore did not influence the enterolactone concentration. According to a recent report regarding antibiotic use in Sweden (SWEDRES), the current number of antibiotic users is higher (i.e., approximately 26 % of individuals in southern Sweden collected at least one antibiotic prescription in 2007) (196).

It is possible that low enterolactone concentration is due to exposure to factors possessing antibacterial effects (e.g., intermittent antibiotic use and high alcohol consumption). These factors decrease enterolactone concentrations over a short period of time, and the measured enterolactone concentration does not reflect long-term enterolactone exposure.

Non-differential misclassification of a dichotomous exposure tends to bias the association towards the null. However, this may not be the case when there are more than two exposure categories. Misclassification for nonadjacent categories (e.g., from highest to lowest) may result in obscure relationships. Because we suspect that there may be misclassification across the quartiles of enterolactone, we believe it more appropriate to categorise enterolactone concentration into a dichotomous variable. The decision to divide individuals according to the median concentration of enterolactone is also supported by the similar breast cancer risk estimates among individuals in the 1st quartile and 2nd quartiles. In addition, even if the median split is a crude and conservative approach for estimating exposure, the risk associated with breast cancer is significantly different between the two halves.

Genetics

Hardy Weinberg equilibrium (HWE) can be used to detect signs of genotyping error or confounding due to population admixtures. We excluded SNPs with HWE values below $p<0.05$. This is a more conservative cut-off than many other studies evaluating a large number of SNPs have utilized. Therefore, some SNPs would be classified as not being in HWE by chance.

Another concern is whether we have captured the SNPs that account for the different functionality of the ERs. The SNPs were systematically selected from the promoter, introns, exons, and 3'UTR in ESR1 and ESR2. We used a haplotype-tagging SNPs approach to selection SNPs in ESR1. This approach makes it possible to identify SNPs that account for the major sequence variation from a large number of SNPs. However, we only analysed a selected number of SNPs in ERS1 and ESR2 (i.e, SNPs that destroy or create CpG-sites), and rare variants (< 0.20 MAF) were excluded. Of the 1400 reported SNPs in ESR1, 400 had >5% heterozygosity frequency, of which 100 created or destroyed CpG sites, with considerably fewer SNPs

present in ESR2. Therefore, there might be other functional polymorphisms in these genes that were not included in our analyses.

Misclassification of other lifestyle factors

In Paper III, even though both lifestyle factors and enterolactone were measured with various degrees of error, we observed the expected association, i.e., a low concentration among smokers and overweight and obese subjects. Weight and height were measured by trained persons in a standardised way, as self-reported anthropometric measurements may be prone to errors.

Alcohol intake was based on the alcohol consumption reported by participants over the course of 7 days. Zero-consumers were identified in the questionnaire based on no reported alcohol use during these 7 days, as well as no reported consumption of alcohol during the previous year. Correlation coefficient was estimated in the validity study to 0.74 for beer, 0.63 for wine and 0.67 for spirits. The participants also estimated their average intake of wine, beer and spirits during the previous year in the questionnaire. However, none of these alcohol sources were correlated with enterolactone in the multivariate model using these estimates.

In the MDC study, leisure-time physical activity was estimated using a very extensive method that some participants found difficult to complete. We suspect that this variable may be associated with some misclassification, which could explain why we did not observe the highest enterolactone concentration in the 4^{th} quartile of physical activity.

In MDC, information only for current use (from menu book and questionnaire) of menopausal hormone therapy was available. There is, however, high agreement between the use reported in the menu book and questionnaire, indicating the high validity of this question item in MDC (166). However, studies have shown that current use is the main predictor, although risk seems to increase with duration. If information on duration was also obtained, we would suspect even higher risk estimates.

Misclassification of disease status

Breast cancer status

The cancer registry in Sweden is almost 100 % complete. In 2006, 100 % of the cases had cytologically or histologically verified breast cancer. We should, however, pay attention to loss of follow-up as some individuals moved out of Sweden during follow-up, and therefore information is lacking with regard to whether they became diagnosed cases or not. However, these numbers are small (approximately 0.5 % of the individuals in the cohort). Mammography attendance differs in relation to socioeconomic factors. Therefore, detection of breast cancer is suspected to be higher in high socioeconomic groups.

There may be undiagnosed breast cancers at baseline, and these individuals should be excluded from the analyses. However, for the sensitivity analyses, we excluded cases that received a diagnosis within one year after the baseline examinations.

We have only included invasive breast cancer, not *in situ*, as it is uncertain whether *in situ* cancer progress into invasive cancer or not. Inclusion of *in situ* cancer may obscure true associations between diet and disease.

Estrogen receptor status

Misclassification of the receptor status may occur among cases. Two cores of tumour materials were used for the tissue microarray. In the majority of duplicate cores, ER expression did not differed. If a difference was found, the evaluation was based on the core exhibiting the highest expression. One person evaluated the hormone receptor status of all breast tumours in a standardised way, thereby eliminating inter-observer variation. All arrays were evaluated twice independently and, in the case of discrepancy, a third examination was performed, followed by a final decision. This strategy reduced the potential intra-observer bias.

There is considerable variation in the reported efficacy of ERβ antibodies for immunohistochemical evaluation (197), contributing to the controversial view of ERβ. There are many antibodies available, and there is a lack of consensus regarding which antibody and what cut-off value to use to determine ERβ positivity. The ERβ antibody used in the present study (EMR02) was a commercial, but less established antibody. It was validated by comparing the immunohistochemical method with western blot using five different breast cancer cell lines (125). The antibody is, however, only able to detect full-length ERβ, as it demonstrates specific reactivity with a 17-amino acid sequence present in the full-length ERβ; several isoforms (for example ERβcx) cannot be detected by this antibody. We used a cut-off value of 10 % positive nuclei for both ERα and ERβ, and observed a higher frequency of ERβ-negative tumours compared to other studies (173,123), indicating specific antibody recognition.

The numbers of cases with missing ER status (especially ERβ status) was relatively high resulting in decreased statistical power in examining the association between diet and ER-defined breast cancer. Individuals lacking ER status data were, however, not significantly different from other cases according to major risk factors (i.e., age, weight, height and use of menopausal hormone therapy).

Misclassification of menopausal status

As some hormonal risk factors are specifically related to postmenopausal breast cancer, is it important to restrict the analyses to those individuals that are truly postmenopausal. The questionnaire item “what year did your menses cease” was well answered, although some women found it difficult to relate the answer to surgery (hysterectomy or oophorectomy) or hormone therapy use. The information available on hysterectomy and oophorectomy from Malmö University Hospital does not cover the entire relevant time period for the entire MDC cohort. Instead, we used an age-dependent definition of menopausal status. Natural menopause was estimated in a group of women (n=2,898) in the MDC cohort without surgery and without hormone therapy. The median age among these women was 50 years of age (range, 46-54). In the absence of information on menstrual history, 50 years of age has been shown to be a reasonable proxy for defining menopause (198).

Age at the time of diagnosis, however, is more important than age at baseline. At the end of follow-up, all women were over 50 years of age. The cases were 45-73 years old at baseline, and the mean age at diagnosis was 63 years of age (range 46-81, inter-quartile range 57-69). The breast cancer diagnoses occurred 0-13 years (mean of 6 years) after baseline examinations. The mean age at the time of breast cancer diagnosis was 63 years of age. Therefore, we have reason to believe that very few women were premenopausal at the time of diagnosis.

In Paper II-IV, we used 50 years and/or 55 years as cut-offs for the sensitivity analyses. In Paper II, the risk estimates for the plant food groups were essentially the same when the analyses were restricted to individuals > 55 years at baseline. In Paper III, a more apparent inverted U-shaped curve was observed between the enterolactone concentration and the risk of breast cancer when individuals greater than 50 years

of age were excluding (as well as 55 years of age), and a significant association was observed for the 4th quartile compared to the 2nd quartile. In Paper IV, there were no major differences in the observed associations with polymorphisms when individuals below 50 or 55 years of age were excluded. When examining the joint effect of enterolactone and polymorphisms on the risk of breast cancer we observed similar results when women below 50 years at baseline and individuals diagnosed with breast cancer within one year after blood collection were excluded; however, only the joint effect of rs2347867 and high enterolactone remained statistically significant. Although the sample size was reduced by 23 %, there were still a tendency towards an interaction between rs2347867 and enterolactone (p=0.07).

Confounding factors

Since dietary factors tend to co-vary with other lifestyle and socio-economic factors, other factors could confound the association between diet and risk of disease. Unexposed individuals differ in many other aspects (apart from the exposure of interest) from the exposed individuals.

A confounding factor has to be associated with the disease (a real cause of the disease or a marker of a real cause) apart from its association with exposure (that is, the confounding factor has to be associated with disease among unexposed individuals). The confounder must also be associated with exposure among the source population of cases. Furthermore, it must not be affected by the exposure or the disease, especially not be an intermediate factor between exposure and disease (i.e., in the causal chain).

The decision for selection of confounders must be based on prior knowledge. Selection of confounders using statistical procedures based on significance tests can be particularly misleading (199). Causal diagrams (DAGs, directed acyclic graphs) link variables by arrows representing direct causal effects. They can be used to identify variables that must be measured and controlled for. They also provide a method for the critical evaluation of traditional epidemiologic criteria for confounding factors (200).

Unmeasured confounding factors may also exist. Using a surrogate (marker) for the unmeasured confounder, we can alleviate the impact of unmeasured confounding, for example age (a marker of aging) and education (a marker of other lifestyle factors). Confounders may also be imperfectly measured, causing residual confounding even when major potential confounders are included in the model.

We selected covariates based on a biological model, with risk factors for breast cancer based on previous knowledge (i.e., weight, height, physical activity, alcohol consumption, age at menopause, parity, and use of menopausal hormone therapy) and variables associated with lifestyle and food habits which may function as surrogates for the actual cause (educational status and smoking habits). A confounding factor must be associated with the exposure examined in the study. In Paper II, there are 11 models (food variables) and it would be inappropriate to evaluate the covariate for each model as this may lead to inclusion of different covariates in each model.

In Paper IV, we did not adjust for additional covariates when examining the association between polymorphisms and breast cancer because alleles are randomly segregated at the time of gamete formation, and lifestyle factors are generally not able to influence the genotype (201).

In Paper II, adjustments for potential confounders generally produced stronger associations. In Paper III, adjustments for confounding factors only slightly changed the risk estimate for enterolactone. In Paper IV, the results were similar when adjusting for potential confounding factors. However, we observed significant decreased risks for the joint effect between two additional polymorphisms and enterolactone.

One concern is the adjustment for obesity, i.e, whether obesity is a confounder or an intermediate factor in the causal chain from enterolactone to breast cancer. Obesity is associated with enterolactone, and we suspect that obesity itself can cause the decrease in enterolactone concentrations. However, if we use enterolactone as a marker of a high-fibre diet, adoption of this diet may have resulted in weight change.

For example, in analyses from the Nurses Health Study, the strong association between a low-carbohydrate score and diabetes completely disappeared when adjustments were made for BMI (202). No associations were observed between a low carbohydrate-diet score and risk of type 2 diabetes when stratified for obesity, and therefore no effect-modification of obesity.

One limitation is that there was no information concerning family history of breast cancer in MDC. However, familial breast cancer mainly affects younger women; therefore, mainly sporadic breast cancer occurs in this age group (5).

Interaction

The association between an exposure and the disease may depend on the presence of other factors. In statistical models, effect-modification is used as a synonym for interaction. However, for a biological interaction there must be a causal interaction. From the public health perspective, it is usually the stable factors (e.g., genetics) that are regarded as the effect-modifiers.

In Paper IV, we found a tendency towards an interaction between a polymorphism in intron 3 (rs2347867) and enterolactone (p=0.07). Interaction can be evaluated either as a heterogeneity of effect or by using a joint effect model (sufficient-cause model). When assessing the heterogeneity of effects, the interpretation is based on another factor increasing or decreasing susceptibility to developing the disease, i.e., that there is a stronger association in subgroups of individuals that are exposed to a particular factor compared to the other individuals. That is, the slopes of the associations are not parallel when the interaction is present. We found no association between breast cancer and enterolactone concentration among those homozygous for the major allele (AA) (p=0.93), whereas there was an inverse association among carriers of the minor allele (AG and GG) (p=0.007). That is, the effect of enterolactone was modified by genetic factors, as the association with enterolactone was stronger in individuals harbouring the minor allele than in those individuals with the major allele.

The joint effect model, on the other hand, compares the observed joint effects (the combined effects) of two factors with the expected joint effect. Carriers of the minor allele (AG or GG) in combination with low enterolactone concentrations had an increased risk of breast cancer (OR: 1.44; 95 % CI: 1.02-2.07) compared to individuals with low enterolactone concentrations and homozygous for the major allele.

Sometimes a confounder is also an effect-modifier. If the risk differs between the strata of another factor, then it is incorrect to adjust for that factor. Instead, risk stratified for the effect-modifier can be presented. Fibre may be an effect-modifier on the enterolactone-breast cancer association, as the effect of enterolactone is suspected to be observed only if fibre intake is high. Therefore, the association between enterolactone and breast cancer risk should not be adjusted for fibre, but instead analysed in strata of fibre intake. In fact, when stratifying for fibre intake, we observed decreased breast cancer risk in association with high enterolactone concentrations only among individuals with high fibre intake.

Statistical considerations

Type 1 (α) errors are referred to as false positives; we used α=0.05. When a large number of tests is performed, the probability of type 1 errors increases, and mass significance is a major concern. Type 2 (β) errors are referred to as false negatives.

In Paper II, we examined the associations of 11 food groups with breast cancer, and some of the results may have occurred by chance. However, the results were in line with the hypothesis, i.e., that high-fibre bread and fruit, berries and vegetables are the strongest predictors of breast cancer. However, none of these associations were observed in the crude model, and the CI was relativity wide; thus, the precision of the risk was not specific.

In Paper III, we examined a single exposure factor (enterolactone). However, many lifestyle and dietary factors where examined when tested for correlation with the enterolactone concentration, and some of these factors might have occurred by chance.

In Paper IV, we investigated numerous associations between polymorphisms, enterolactone and breast cancer. The only SNP selected for an *a priori* hypothesis was rs2234693. Although we tried to reduce the number of tests by reducing the number of SNPs, significant associations between enterolactone and breast cancer in strata of various genotypes, and the tendency towards an interaction between the SNP and enterolactone might have occurred by chance. After adjusting for multiple comparisons, none of these associations remained significant. We note, however, that all of the observed significant associations were in line with our original hypothesis that high enterolactone is associated with protection against cancer.

In Paper II and III, we examine multiple outcomes (i.e., subtypes of breast cancer) with cases categorized as either negative or positive according to the expression of ERs in the tumour. By using multinomial (polychotomous) logistic regression when analysing the risk with more than one disease category the fact that the same controls are in all analyses could be taken into account.

External validity

External validity refers to the ability to generalise the results to individuals outside of the study population. All individuals born between 1923 and 1950 and living in Malmö were invited to participate in the MDC. With a participation rate of approximately 40 %, there are, however, concerns about the representatively of the study population. Data suggest that the MDC cohort consists of a higher frequency of individuals with better health, although the socio-demographic structure and prevalence of smoking and obesity were equivalent to those obtained in a study from the same cohort with higher response rates. As expected, mortality was higher among non-participants both during recruitment and during follow-up (153). Selection of participants for the cohort may be affected by the level of exposure; however, the association between exposure and disease are not influenced by which subjects that participate in the study.

FINDINGS AND IMPLICATIONS

Among plant foods, high intakes of high-fibre bread were associated with a decreased risk of breast cancer. In addition, when combining intake of fruit, berries and vegetables and restricting the analyses to individuals with suggested more stable food habits, we also observed a decreased breast cancer risk with high intakes of these items. High-fibre bread, and fruit and berries were the main dietary determinants of enterolactone concentration, and high concentrations of enterolactone were associated with a decreased risk of breast cancer.

There is a large inconsistency between studies regarding plant food intakes and the incidence of breast cancer. There may, however, be a causal relationship in studies that do not observe any association. The high relative validity data used in the MDC may explain why we were able to observe associations with breast cancer. Studies indicate that whole grains show protective associations with several forms of cancer (203); high-fibre bread is a major source of whole grains in the Western diet (204).

Apart from food sources, other predictors of enterolactone levels were observed, including low degree of obesity, non-smoking, physical activity and higher education. The determinants were similar to those obtained in other studies investigating Nordic women. Smoking and obesity have been suggested to directly influence enterolactone, as well as indirectly due to their influence on food habits. Education level does not directly influence enterolactone concentrations, but is associated with other socio-economic factors and many health behaviours and food choices. The differences in means across quartiles of physical activity could in a similar way be explained by other health-related factors, since physical activity is associated with various health behaviours and food choices. Thus, a multifaceted complex of many factors arises, making it difficult to single out and identify specific predictors of enterolactone levels. Our goal was to estimate predictors of enterolactone concentration, and we clearly showed that enterolactone is associated with various health-related factors. Our results support the use of enterolactone as a biomarker of a healthy lifestyle (63), including a diet containing high amounts of whole grains, fruits, and vegetables, without smoking and obesity.

We could only explain 12 % of the variation in enterolactone concentrations between subjects. Therefore, unmeasured factors probably explain most of this variation. The capacity of the gut microflora is known to be a very influential factor during the formation of enterolactone from plant lignans (112).

Enterolactone and breast cancer

We observed an inverted U-shaped association between enterolactone and the risk of breast cancer, especially when only women over 50 years were included in the analysis. There was no significant trend across quartiles, but a significant trend was observed from the 2nd to the 4th quartile, and a significant relationship was detected when enterolactone concentrations were categorised into dichotomous exposure levels. Moreover, we observed a moderate to large variation in enterolactone concentrations over the course of one month. Therefore, the association observed between enterolactone and breast cancer is likely attenuated. The inverted U-shaped relationship between enterolactone and breast cancer has been previously

documented in a Danish prospective study that was conducted for a very similar population (95), and a case-control study in Finland observed a similar inverted U-shaped association among postmenopausal women (88). In summary, our data supports the hypothesis of an inverted U-shaped relation between enterolactone and breast cancer. However, the question is whether the observed association is due to measurement errors as low enterolactone concentration may not reflect the long-term exposure (because of factors with antibacterial effect).

Enterolactone is a biomarker; however, the question remains: a marker of what? The enterolactone concentration present in blood can be used in several ways: as a 1) measure of enterolactone internal exposure *per se,* as it has its own biological effects, 2) biomarker of dietary lignan intake, 3) biomarker of a healthy lifestyle, with a diet containing high amounts of whole grains, fruits and vegetables, and a non-smoking lifestyle without obesity, or 4) marker of the capacity of the microflora.

In principal, plant lignans lack the bioactivity that has been observed for enterolactone. For lignans to be active, they must be transformed by the gut microflora. Complete information regarding lignan content is not available, making it difficult to estimate individual lignan intake based on dietary assessment (205). Furthermore, new precursors of enterolactone have recently been identified, resulting in an underestimation of lignan exposure. In addition, the amount of lignans present in the same food group varies largely, as few food items contain large amounts of lignans, making it difficult to estimate lignan intake based on regular dietary assessment methods. Although a large difference in lignan content exists between foods, amount of absorbed lignans do not display a linear correspondence with the amounts present in foods. Microflora is the major determinant of lignan absorption and account for most of the variation observed between individuals. As there is a large variation in the amount of absorbed lignan, intake of lignan-rich foods may not correspond to the internal lignan exposure.

Fibre, enterolactone and the importance of the microflora

According to our hypothesis, lignans in addition to fibre are negatively associated with disease. Dietary fibre (mainly cereal fibre), including its associated lignans, seems to protect against diseases, and the protective effect of enterolactone may mainly be observed in populations where cereals are major contributors to the enterolactone concentrations. We observed a decreased risk of breast cancer in association with high enterolactone concentrations, especially among women with high intakes of fibre. This further supports the theory that sufficient amounts of fibre must be present in the diet in order for the enterolactone to have a beneficial effect.

Fibre-rich foods contain enterolignan precursors. Moreover, fibre stimulates the microflora (as prebiotics), and relatively high intakes of fibre-rich foods (especially from cereals) are needed to maintain an active intestinal microflora (63), which potentially increase the conversion of plant lignans to enterolignans.

Cereal fibre seems to reduce enterohepatic circulation of estrogens, and therefore plasma estrogen concentration. Fat has the opposite effect on enterohepatic estrogen reabsorption (206), as well as on lignan bioavailability and the production of enterolignans, and separation of these independent effects is difficult. Although we did not observe any correlation between fat and enterolactone, previous studies have observed that a high-fat diet reduces the protective effect of lignans.

Obesity is negatively correlated with enterolactone. Although very speculative, the association between increased breast cancer risk and high obesity may be an indirect result of the reduction in enterolactone, or overall dysfunction of the microflora among these individuals. The recent observation that obese individual have different composition of their microflora (207) contributes to the view asserting the importance of the microflora, especially in mediating the effect of diet.

Fermented products may also facilitate gut fermentation (208). In fact, a Dutch study (209), and results from the MDC study (210), indicate that a diet high in fermented milk and fibre may protect against breast cancer. Although uncertainties regarding the causality, recent observations that antibiotic use was associated with increased risk of breast cancer (211,212) further points towards the importance of a balanced microflora in maintaining optimal health. The widespread use of antibiotics could have a major impact on public health.

Food matrix and food synergy

During recent years, it has been proposed that nutrition research should focus on foods and less on specific food compounds. This is because whole foods contain more than the specific food components, and information is lost when each component is examined separately. Therefore, a focus on nutrients tends to oversimplify a complex system. This is referred to as food synergy. Whole grains are one example of a food that seems to have more beneficial health effect than the sum of each specific compound present in this food (213).

Lignans are one of several thousand bioactive compounds that are provided by a normal diet. Although we tried to evaluate the effect of enterolactone *per se*, it is difficult to separate the effect of enterolactone as enterolactone concentration is correlated with many other factors (although we adjusted for several potential confounders). Enterolactone is associated with many characteristics of a healthy lifestyle, and together these factors reduce the risk of breast cancer.

It is important to keep in mind that all foods come from living organisms in which the components interact to produce biological systems, and this is relevant when the food is consumed. The naturally occurring constituents and their interactions are called the food matrix (214). The context in which the lignans operate seems very crucial. Lignans originate from the fibre-complex present in foods. Fibre and lignans are an example of food components that interact to produce health effects. Lignan sources that are not sources of fibre (for example wine, tea, coffee and juice) may not be associated with the same health effects.

Plant foods, enterolactone and estrogen receptor-defined breast cancer

Enterolactone was mainly associated with ERα (+) and ERβ (-) tumours. The main determinants of enterolactone, fruit and berries, showed an inverse association mainly with ERβ (-) tumours (non-significant); high-fibre bread was mainly inversely associated with ERα (+) and ERβ (+) tumours (non-significant). Few studies considered ERβ; however, a recent observation for material from the MDC cohort suggested that obesity is associated with ERα (+) and ERβ (-) tumours (215).

ERβ has been suggested to have a regulatory role in ERα activity (126), and new insights regarding the interaction between the two ERs reveal that ERβ acts by antagonizing ERα by modulating a very specific subset of estrogen-stimulated genes and actively prevents ERα-stimulated cell growth (216). A recent study showed that enterolactone has a preference for ERα (79). Our results might be explained by the lack of ERβ-mediated inhibitory effects in ERα(+)/ERβ (-) tumours, thus causing the tumours to be more susceptible to the antiestrogenic influence of enterolactone.

Genetic variations

High enterolactone concentrations were associated with breast cancer protection in subgroups of the population possessing specific genetic variants. We were, however, unable to detect statistically significant interactions between enterolactone concentrations and any of the selected polymorphisms with regard to the risk of breast cancer. The only SNP that showed a tendency to modify the association between enterolactone and breast cancer was rs2347867 in ESR1. It appears that carriers of the minor allele are particularly receptive to the protective effect of enterolactone, as we observed an increased risk of breast cancer in individuals with low enterolactone blood concentrations. By increasing their enterolactone levels, these individuals should attain the same level of breast cancer risk exhibited by individuals homozygous for the major allele.

The SNP is located in intron 3, and its function is unknown. It might have a causal modifying effect or be in strong LD with other functional SNPs. In fact, it is close to and in strong LD (D'=0.99) with another polymorphism (rs6557171) located in a region with high homology to the corresponding ESR1 gene in both rat and mouse. This high agreement may indicate that this region contains functional segments that are conserved among species.

The variant allele (C) of rs2234693 (also known as PuvII) has been found to produce a binding site for the transcription factor B-myb, and has been suggested to increase transcription or produce other ERα isoforms with different properties (147). We found that carriers of the C allele in combination with high enterolactone concentrations had a significantly lower risk of breast cancer compared to women with the TT genotype and low enterolactone concentrations. In other words, individuals with this polymorphism, which might enhance transcription of ESR1, in combination with high enterolactone exposure, demonstrated a decreased risk of breast cancer. However, no evidence for an interaction between the polymorphism and enterolactone concentration was detected (p=0.37).

FUTURE CHALLENGES

Large prospective studies are needed to study the association between dietary factors and breast cancer. Although highly valid, the diet history method used during the MDC baseline examinations was time-consuming and expensive. Therefore, new high validity methods to assess diet and energy expenditure need to be developed for the future. These methods should preferable focus on food diary techniques, not food frequency questionnaires. Questions regarding dieting behaviour and dietary change in the past should preferably be included.

Future epidemiological studies need to consider the complexity of the diet and should focus on foods and dietary patterns instead of separate nutrients and bioactive compounds in order to account for food synergy. There is also a need for reliable biomarkers of dietary exposures. The dietary assessments utilised in epidemiological studies may encounter difficulties in separating whole grains from refined products, as many people find it difficult to accurately identify whole grain foods (217). However, biomarkers of whole grain intakes are under development (for example alkylresorcinols (rye and wheat) (218) and avenanthramides (oat) (219)), and should in future studies be used to confirm the role of whole grain intakes in cancer prevention.

There are still large uncertainties regarding the association between enterolactone and breast cancer. Future studies need to reduce measurement errors in enterolactone estimation, mainly by excluding individuals using antibiotics and estimating enterolactone exposure based on repeated measurements.

To interpret the results of enterolactone across the study populations, determinants of enterolactone need to be identified; studies with highly valid dietary data are needed to accurately identify the major determinants.

To further explain the inconsistent findings regarding the association between enterolactone and breast cancer, other effect-modifiers of the association must be identified (e.g., smoking, fat intake and BMI). Furthermore, estrogen levels in individuals with regard to fat intake, fibre intake, and enterolactone concentrations must be determined in order to clarify the effects of enterolactone.

Potential effect-modification of the diet-disease association by genetic factors must also be clarified. Several of the selected polymorphisms present in ER genes have not been examined together with environmental factors with regard to breast cancer risk. As these genes contain several alternative promoter sites and produce several transcription variants and isoforms, more functional regions in the genes and in the promoter regions could be characterised using bioinformatic tools, as well as functional studies of the transcript levels associated with different genotypes. Other polymorphisms in the genes that encode proteins involved in estrogen synthesis and metabolism, as well as in SHBG, also exist, and are extremely interesting candidates as effect-modifiers in the association between enterolactone and breast cancer.

There is also a need for additional large studies that take the ER status of the tumours into account, especially ERβ status using validated ERβ antibodies, when examining the association between dietary factors and breast cancer. The observation that the protective effect of enterolactone in breast cancer risk was more evident in tumours lacking the expression of ERβ will need to be confirmed by further studies.

CONCLUSIONS

1. We observed relatively high variation of enterolactone within and between individuals over 1 month; the association between enterolactone and risk of breast cancer is likely attenuated due to the moderate reliability of a single measurement of enterolactone.

2. Among the plant foods, high intakes of high-fibre bread were associated with decreased risk of breast cancer. Also, when restricting the analyses to individuals with suggested more stable food habits, we observed a decreased breast cancer risk with high intakes of fruit, berries and vegetables.

3. There was no strong evidence for difference in incidence according to the ERα and ERβ status of breast cancer. However, high intake of fried potato was associated with an increased risk of ERβ (-) breast cancer, and there was a tendency of an inverse association for high-fibre bread among ERα (+) breast cancer and ERβ (+) breast cancer.

4. High-fibre bread and fruit and berries were the main dietary determinants of enterolactone concentration. In addition, obesity, smoking was associated with low enterolactone concentrations and high education and physical activity was associated with high enterolactone concentrations in this population. This observation supports other studies suggesting that enterolactone is a marker of a healthy lifestyle. However, these factors could only explain a minor part of the variation in enterolactone concentration between subjects, which points towards the importance of the microflora in mediating the effect of the diet.

5. A tendency towards a reduced breast cancer risk was observed for the highest compared to the lowest quartile of enterolactone concentration. When divided the individuals in two groups according to enterolactone concentration, a high concentration (>16 nmol/L) was associated with a 25% significant decreased risk of breast cancer. The heterogeneity of the lowest quartile (likely including individuals with low concentrations due to antibiotic use) appear obvious as individuals in the lowest quartile reported similar fibre intakes as individuals in the 2^{nd} quartile. When stratifying for fibre intake, we observed decreased breast cancer risk with high enterolactone concentration only among individuals with high fibre intake. This supports the theory that an association between enterolactone and breast cancer are more likely to be observed in populations consuming a high-fibre diet.

6. The reduced breast cancer risk with high enterolactone concentrations was only observed for ERα (+) and ERβ (-) tumors. The risk was significantly different for ERβ (-) and ERβ (+) tumors.

7. Breast cancer risk was not significantly associated with any of the selected polymorphisms in ESR1 and ESR2.

8. The protective effect of high enterolactone concentration was seen in subgroups of the individuals with specific genetic variants. There was a tendency towards an interaction between a polymorphism in intron 3 of ESR1 (rs2347867) and enterolactone concentration. There was no association between breast cancer and enterolactone concentration among those homozygous for the major allele. However, we found an inverse association among carriers of the minor allele.

ACKNOWLEDGEMENTS

Without the support and help from all the wonderful people around me this dissertation wouldn't have looked as it does now. In particular I would like to thank:

All the 28, 028 individuals in the Malmö diet and cancer cohort for taking you time to participate in this study.

Everyone that planned the study and developed the dietary method. Because of you we have one of the best dietary methods in the world.

Each person that was involved during the baseline examinations. I can only imagine all the work that has been done.

All staff at the clinical research unit that helped us during the reliability study.

Elisabet Wirfält, my supervisor, for you great support, for letting me take large responsibility in my own research and for believing in me. Thank you for you outstanding comments on my manuscripts and for all the discussions in the wonderful field of nutrition epidemiology.

Bo Gullberg for all your statistical advice and other discussions regarding research and the manuscripts.

Ulrika Ericson, also a PhD student and my room-mate at CRC. Thank you for always answering my questions, especially regarding the baseline examinations.

The other people in the research group of nutrition epidemiology, Peter Wallström and Isabel Holmbäck.

Joyce Carlson, for all interesting and stimulating discussions in the field of genetics.

Signe Borgquist and Göran Landberg, for letting us use the receptor data and for the comments on my manuscripts.

Håkan Olsson, for the contribution of your great knowledge in cancer biology.

Herman Adlercreutz, co-author, for all your research regarding lignans during the last decades. Can you imagine, I was hardly born when you came up with the hypothesis regarding fibre and lignans.

Malin Ivarsson, for all genotyping and work with the selection of SNPs, and for the comments on the fourth manuscript.

Sophia Harlid, for your work with the SNPs and the comments on the forth manuscript.

Carl Bryngelsson, for the handling of blood samples and Anders Dahlin for the handling of data files.

Göran Berglund, for making the Malmö diet and cancer study possible and for making it possible for me to start working in the project.

Marju Orho-Melander, for the opportunity to continue doing great research.

Charlotta Roos, for the good time at the NuGo week in Potsdam in September.

Alicja Wolk, for suggesting phytoestrogens as the subject of my ex-work at the Karolinska Institute.

All the people in Malmö student choir, for all the joy we have each Monday, at all parties and during the travels. Thanks for a great time in Berlin although it was the weekend before I handed in my dissertation to the printers.

My mother, father and sisters for always being so proud of me. I know from who I've got my interest in science!

Anders, the love of my life, for all your love and support. You are the best! You have the most amazing thoughts and I hope you soon will get the opportunity to be a PhD student yourself. I promise you, you won't regret it!

REFERENCES

1. Feinberg, A.P., Ohlsson, R. and Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet*. 2006, 7, 21-33.

2. Kufe, D.W., Holland, J.F., Frei, E. and American Cancer Society *Cancer medicine 6*. Decker, Hamilton, Ont. ; London (2003).

3. *Cancer incidence in Sweden 2006*. The National Board of Health and Welfare, Stockholm (2007).

4. Bray, F., McCarron, P. and Parkin, D.M. The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res*. 2004, 6, 229-39.

5. Pharoah, P.D., Day, N.E., Duffy, S., Easton, D.F. and Ponder, B.A. Family history and the risk of breast cancer: a systematic review and meta-analysis. *Int J Cancer*. 1997, 71, 800-9.

6. Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000, 343, 78-85.

7. Palacios, J., Robles-Frias, M.J., Castilla, M.A., Lopez-Garcia, M.A. and Benitez, J. The molecular pathology of hereditary breast cancer. *Pathobiology*. 2008, 75, 85-94.

8. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007, 447, 1087-93.

9. Czene, K., Lichtenstein, P. and Hemminki, K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer*. 2002, 99, 260-6.

10. Stanford, J.L., Herrinton, L.J., Schwartz, S.M. and Weiss, N.S. Breast cancer incidence in Asian migrants to the United States and their descendants. *Epidemiology*. 1995, 6, 181-3.

11. Ziegler, R.G., Hoover, R.N., Pike, M.C., Hildesheim, A., Nomura, A.M., West, D.W., et al. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst*. 1993, 85, 1819-27.

12. Chia, K.S., Reilly, M., Tan, C.S., Lee, J., Pawitan, Y., Adami, H.O., et al. Profound changes in breast cancer incidence may reflect changes into a Westernized lifestyle: a comparative population-based study in Singapore and Sweden. *Int J Cancer*. 2005, 113, 302-6.

13. Key, T., Appleby, P., Barnes, I. and Reeves, G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst*. 2002, 94, 606-16.

14. Onland-Moret, N.C., Kaaks, R., van Noord, P.A., Rinaldi, S., Key, T., Grobbee, D.E., et al. Urinary endogenous sex hormone levels and the risk of postmenopausal breast cancer. *Br J Cancer*. 2003, 88, 1394-9.

15. Zeleniuch-Jacquotte, A., Shore, R.E., Koenig, K.L., Akhmedkhanov, A., Afanasyeva, Y., Kato, I., et al. Postmenopausal levels of oestrogen, androgen, and SHBG and breast cancer: long-term results of a prospective study. *Br J Cancer*. 2004, 90, 153-9.

16. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet*. 1997, 350, 1047-59.

17. Rossouw, J.E., Anderson, G.L., Prentice, R.L., LaCroix, A.Z., Kooperberg, C., Stefanick, M.L., et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *Jama*. 2002, 288, 321-33.

18. Pike, M.C. Reducing cancer risk in women through lifestyle-mediated changes in hormone levels. *Cancer Detect Prev*. 1990, 14, 595-607.

19. Yager, J.D. and Davidson, N.E. Estrogen carcinogenesis in breast cancer. *N Engl J Med*. 2006, 354, 270-82.

20. World Cancer Research Fund and American Institute for Cancer Research *Food, nutrition, physical activity, and the prevention of cancer: a global perspective*. AICR, Washington DC (2007).

21. Lahmann, P.H., Hoffmann, K., Allen, N., van Gils, C.H., Khaw, K.T., Tehard, B., et al. Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). *Int J Cancer*. 2004, 111, 762-71.

22. Hankinson, S.E. and Hunter, D. Breast cancer. In Hunter, D., Trichopoulos, D. and Adami, H.-O. (eds.), *Textbook of cancer epidemiology*. Oxford University Press, New York, N.Y.,(2002) vol., pp. 599 s.

23. Doll, R. and Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst*. 1981, 66, 1191-308.

24. Boyd, N.F., Martin, L.J., Noffel, M., Lockwood, G.A. and Trichler, D.L. A meta-analysis of studies of dietary fat and breast cancer risk. *Br J Cancer*. 1993, 68, 627-36.

25. Hunter, D.J., Spiegelman, D., Adami, H.O., Beeson, L., van den Brandt, P.A., Folsom, A.R., et al. Cohort studies of fat intake and the risk of breast cancer--a pooled analysis. *N Engl J Med*. 1996, 334, 356-61.

26. Boyd, N.F., Stone, J., Vogt, K.N., Connelly, B.S., Martin, L.J. and Minkin, S. Dietary fat and breast cancer risk revisited: a meta-analysis of the published literature. *Br J Cancer*. 2003, 89, 1672-85.

27. Gandini, S., Merzenich, H., Robertson, C. and Boyle, P. Meta-analysis of studies on breast cancer risk and diet: the role of fruit and vegetable consumption and the intake of associated micronutrients. *Eur J Cancer*. 2000, 36, 636-46.

28. Smith-Warner, S.A., Spiegelman, D., Yaun, S.S., Adami, H.O., Beeson, W.L., van den Brandt, P.A., et al. Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. *Jama*. 2001, 285, 769-76.

29. Ordovas, J.M. and Mooser, V. Nutrigenomics and nutrigenetics. *Curr Opin Lipidol*. 2004, 15, 101-8.

30. Mattisson, I., Wirfalt, E., Johansson, U., Gullberg, B., Olsson, H. and Berglund, G. Intakes of plant foods, fibre and fat and risk of breast cancer--a prospective study in the Malmo Diet and Cancer cohort. *Br J Cancer*. 2004, 90, 122-7.

31. Bingham, S.A., Day, N.E., Luben, R., Ferrari, P., Slimani, N., Norat, T., et al. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet*. 2003, 361, 1496-501.

32. Suzuki, R., Rylander-Rudqvist, T., Ye, W., Saji, S., Adlercreutz, H. and Wolk, A. Dietary fiber intake and risk of postmenopausal breast cancer defined by estrogen and progesterone receptor status--a prospective cohort study among Swedish women. *Int J Cancer*. 2008, 122, 403-12.

33. Adlercreutz, H., Hockerstedt, K., Bannwart, C., Bloigu, S., Hamalainen, E., Fotsis, T., et al. Effect of dietary components, including lignans and phytoestrogens, on enterohepatic circulation and liver metabolism of estrogens and on sex hormone binding globulin (SHBG). *J Steroid Biochem*. 1987, 27, 1135-44.

34. Monroe, K.R., Murphy, S.P., Henderson, B.E., Kolonel, L.N., Stanczyk, F.Z., Adlercreutz, H., et al. Dietary fiber intake and endogenous serum hormone levels in naturally postmenopausal Mexican American women: the multiethnic cohort study. *Nutr Cancer*. 2007, 58, 127-35.

35. Adlercreutz, H. Lignans and human health. *Crit Rev Clin Lab Sci*. 2007, 44, 483-525.

36. Keinan-Boker, L., Peeters, P.H., Mulligan, A.A., Navarro, C., Slimani, N., Mattisson, I., et al. Soy product consumption in 10 European countries: the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr*. 2002, 5, 1217-26.

37. Mazur, W. Phytoestrogen content in foods. *Baillieres Clin Endocrinol Metab*. 1998, 12, 729-42.

38. Milder, I.E., Arts, I.C., van de Putte, B., Venema, D.P. and Hollman, P.C. Lignan contents of Dutch plant foods: a database including lariciresinol, pinoresinol, secoisolariciresinol and matairesinol. *Br J Nutr*. 2005, 93, 393-402.

39. Heinonen, S., Nurmi, T., Liukkonen, K., Poutanen, K., Wahala, K., Deyama, T., et al. In vitro metabolism of plant lignans: new precursors of mammalian lignans enterolactone and enterodiol. *J Agric Food Chem*. 2001, 49, 3178-86.

40. Miksicek, R.J. Commonly occurring plant flavonoids have estrogenic activity. *Mol Pharmacol*. 1993, 44, 37-43.

41. Kwon, M., Davin, L.B. and Lewis, N.G. In situ hybridization and immunolocalization of lignan reductases in woody tissues: implications for heartwood formation and other forms of vascular tissue preservation. *Phytochemistry*. 2001, 57, 899-914.

42. Setchell, K.D.R. and Adlercreutz, H. The excretion of two new phenolic compounds (180/442 and 180/410) during the human menstrual cycle and in pregnancy *J Steroid Biochem*. 1979, 11, xv-xvi.

43. Setchell, K.D., Lawson, A.M., Mitchell, F.L., Adlercreutz, H., Kirk, D.N. and Axelson, M. Lignans in man and in animal species. *Nature*. 1980, 287, 740-2.

44. Stitch, S.R., Toumba, J.K., Groen, M.B., Funke, C.W., Leemhuis, J., Vink, J., et al. Excretion, isolation and structure of a new phenolic constituent of female urine. *Nature*. 1980, 287, 738-40.

45. Adlercreutz, H., Fotsis, T., Heikkinen, R., Dwyer, J.T., Goldin, B.R., Gorbach, S.L., et al. Diet and urinary excretion of lignans in female subjects. *Med Biol*. 1981, 59, 259-61.

46. Adlercreutz, H., Fotsis, T., Heikkinen, R., Dwyer, J.T., Woods, M., Goldin, B.R., et al. Excretion of the lignans enterolactone and enterodiol and of equol in omnivorous and vegetarian postmenopausal women and in women with breast cancer. *Lancet*. 1982, 2, 1295-9.

47. Axelson, M., Sjovall, J., Gustafsson, B.E. and Setchell, K.D. Origin of lignans in mammals and identification of a precursor from plants. *Nature*. 1982, 298, 659-60.

48. Borriello, S.P., Setchell, K.D., Axelson, M. and Lawson, A.M. Production and metabolism of lignans by the human faecal flora. *J Appl Bacteriol*. 1985, 58, 37-43.

49. Penalvo, J.L., Heinonen, S.M., Aura, A.M. and Adlercreutz, H. Dietary sesamin is converted to enterolactone in humans. *J Nutr*. 2005, 135, 1056-62.

50. Liu, Z., Saarinen, N.M. and Thompson, L.U. Sesamin is one of the major precursors of mammalian lignans in sesame seed (Sesamum indicum) as observed in vitro and in rats. *J Nutr*. 2006, 136, 906-12.

51. Begum, A.N., Nicolle, C., Mila, I., Lapierre, C., Nagano, K., Fukushima, K., et al. Dietary lignins are precursors of mammalian lignans in rats. *J Nutr*. 2004, 134, 120-7.

52. Saarinen, N.M., Warri, A., Makela, S.I., Eckerman, C., Reunanen, M., Ahotupa, M., et al. Hydroxymatairesinol, a novel enterolactone precursor with antitumor properties from coniferous tree (Picea abies). *Nutr Cancer*. 2000, 36, 207-16.

53. Mazur, W., Fotsis, T., Wahala, K., Ojala, S., Salakka, A. and Adlercreutz, H. Isotope dilution gas chromatographic-mass spectrometric method for the determination of isoflavonoids, coumestrol, and lignans in food samples. *Anal Biochem*. 1996, 233, 169-80.

54. Mazur, W.M., Wahala, K., Rasku, S., Salakka, A., Hase, T. and Adlercreutz, H. Lignan and isoflavonoid concentrations in tea and coffee. *Br J Nutr*. 1998, 79, 37-45.

55. Mazur, W.M., Uehara, M., Wahala, K. and Adlercreutz, H. Phyto-oestrogen content of berries, and plasma concentrations and urinary excretion of enterolactone after a single strawberry-meal in human subjects. *Br J Nutr*. 2000, 83, 381-7.

56. Valsta, L.M., Kilkkinen, A., Mazur, W., Nurmi, T., Lampi, A.M., Ovaskainen, M.L., et al. Phyto-oestrogen database of foods and average intake in Finland. *Br J Nutr*. 2003, 89 Suppl 1, S31-8.

57. Thompson, L.U., Robb, P., Serraino, M. and Cheung, F. Mammalian lignan production from various foods. *Nutr Cancer*. 1991, 16, 43-52.

58. Tarpila, S., Aro, A., Salminen, I., Tarpila, A., Kleemola, P., Akkila, J., et al. The effect of flaxseed supplementation in processed foods on serum fatty acids and enterolactone. *Eur J Clin Nutr*. 2002, 56, 157-65.

59. Kuijsten, A., Arts, I.C., van't Veer, P. and Hollman, P.C. The relative bioavailability of enterolignans in humans is enhanced by milling and crushing of flaxseed. *J Nutr*. 2005, 135, 2812-6.

60. Coulman, K.D., Liu, Z., Hum, W.Q., Michaelides, J. and Thompson, L.U. Whole sesame seed is as rich a source of mammalian lignan precursors as whole flaxseed. *Nutr Cancer*. 2005, 52, 156-65.

61. Juntunen, K.S., Mazur, W.M., Liukkonen, K.H., Uehara, M., Poutanen, K.S., Adlercreutz, H.C., et al. Consumption of wholemeal rye bread increases serum concentrations and urinary excretion of enterolactone compared with consumption of white wheat bread in healthy Finnish men and women. *Br J Nutr*. 2000, 84, 839-46.

62. Jacobs, D.R., Jr., Pereira, M.A., Stumpf, K., Pins, J.J. and Adlercreutz, H. Whole grain food intake elevates serum enterolactone. *Br J Nutr*. 2002, 88, 111-6.

63. Stumpf, K., Pietinen, P., Puska, P. and Adlercreutz, H. Changes in serum enterolactone, genistein, and daidzein in a dietary intervention study in Finland. *Cancer Epidemiol Biomarkers Prev*. 2000, 9, 1369-72.

64. Eeckhaut, E., Struijs, K., Possemiers, S., Vincken, J.P., Keukeleire, D.D. and Verstraete, W. Metabolism of the lignan macromolecule into enterolignans in the gastrointestinal lumen as determined in the simulator of the human intestinal microbial ecosystem. *J Agric Food Chem*. 2008, 56, 4806-12.

65. Glitso, L.V., Mazur, W.M., Adlercreutz, H., Wahala, K., Makela, T., Sandstrom, B., et al. Intestinal metabolism of rye lignans in pigs. *Br J Nutr*. 2000, 84, 429-37.

66. Wang, L.Q., Meselhy, M.R., Li, Y., Qin, G.W. and Hattori, M. Human intestinal bacteria capable of transforming secoisolariciresinol diglucoside to mammalian lignans, enterodiol and enterolactone. *Chem Pharm Bull (Tokyo)*. 2000, 48, 1606-10.

67. Clavel, T., Henderson, G., Alpert, C.A., Philippe, C., Rigottier-Gois, L., Dore, J., et al. Intestinal bacterial communities that produce active estrogen-like compounds enterodiol and enterolactone in humans. *Appl Environ Microbiol*. 2005, 71, 6077-85.

68. Jin, J.S., Zhao, Y.F., Nakamura, N., Akao, T., Kakiuchi, N., Min, B.S., et al. Enantioselective dehydroxylation of enterodiol and enterolactone precursors by human intestinal bacteria. *Biol Pharm Bull*. 2007, 30, 2113-9.

69. Bach Knudsen, K.E., Serena, A., Kjaer, A.K., Tetens, I., Heinonen, S.M., Nurmi, T., et al. Rye bread in the diet of pigs enhances the formation of enterolactone and increases its levels in plasma, urine and feces. *J Nutr*. 2003, 133, 1368-75.

70. Rowland, I., Faughnan, M., Hoey, L., Wahala, K., Williamson, G. and Cassidy, A. Bioavailability of phyto-oestrogens. *Br J Nutr*. 2003, 89 Suppl 1, S45-58.

71. Nesbitt, P.D., Lam, Y. and Thompson, L.U. Human metabolism of mammalian lignan precursors in raw and processed flaxseed. *Am J Clin Nutr*. 1999, 69, 549-55.

72. Rowland, I.R., Wiseman, H., Sanders, T.A., Adlercreutz, H. and Bowey, E.A. Interindividual variation in metabolism of soy isoflavones and lignans: influence of habitual diet on equol production by the gut microflora. *Nutr Cancer*. 2000, 36, 27-32.

73. Adlercreutz, H., Fotsis, T., Lampe, J., Wahala, K., Makela, T., Brunow, G., et al. Quantitative determination of lignans and isoflavonoids in plasma of omnivorous and vegetarian women by isotope dilution gas chromatography-mass spectrometry. *Scand J Clin Lab Invest Suppl*. 1993, 215, 5-18.

74. Martin, M.E., Haourigui, M., Pelissero, C., Benassayag, C. and Nunez, E.A. Interactions between phytoestrogens and human sex steroid binding protein. *Life Sci*. 1996, 58, 429-36.

75. Schottner, M., Spiteller, G. and Gansser, D. Lignans interfering with 5 alpha-dihydrotestosterone binding to human sex hormone-binding globulin. *J Nat Prod*. 1998, 61, 119-21.

76. Jacobs, E., Kulling, S.E. and Metzler, M. Novel metabolites of the mammalian lignans enterolactone and enterodiol in human urine. *J Steroid Biochem Mol Biol*. 1999, 68, 211-8.

77. Niemeyer, H.B., Honig, D., Lange-Bohmer, A., Jacobs, E., Kulling, S.E. and Metzler, M. Oxidative metabolites of the mammalian lignans enterodiol and enterolactone in rat bile and urine. *J Agric Food Chem*. 2000, 48, 2910-9.

78. Niemeyer, H.B. and Metzler, M. Oxidative metabolites and genotoxic potential of mammalian and plant lignans in vitro. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2002, 777, 321-7.

79. Penttinen, P., Jaehrling, J., Damdimopoulos, A.E., Inzunza, J., Lemmen, J.G., van der Saag, P., et al. Diet-derived polyphenol metabolite enterolactone is a tissue-specific estrogen receptor activator. *Endocrinology*. 2007, 148, 4875-86.

80. Smeds, A.I., Saarinen, N.M., Hurmerinta, T.T., Penttinen, P.E., Sjoholm, R.E. and Makela, S.I. Urinary excretion of lignans after administration of isolated plant lignans to rats: the effect of single dose and ten-day exposures. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2004, 813, 303-12.

81. Low, Y.L., Dunning, A.M., Dowsett, M., Folkerd, E., Doody, D., Taylor, J., et al. Phytoestrogen exposure is associated with circulating sex hormone levels in postmenopausal women and interact with ESR1 and NR1I2 gene variants. *Cancer Epidemiol Biomarkers Prev*. 2007, 16, 1009-16.

82. Brooks, J.D. and Thompson, L.U. Mammalian lignans and genistein decrease the activities of aromatase and 17beta-hydroxysteroid dehydrogenase in MCF-7 cells. *J Steroid Biochem Mol Biol*. 2005, 94, 461-7.

83. Kuiper, G.G., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., van der Saag, P.T., et al. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. *Endocrinology*. 1998, 139, 4252-63.

84. Mueller, S.O., Simon, S., Chae, K., Metzler, M. and Korach, K.S. Phytoestrogens and their human metabolites show distinct agonistic and antagonistic properties on estrogen receptor alpha (ERalpha) and ERbeta in human cells. *Toxicol Sci*. 2004, 80, 14-25.

85. Prasad, K. Antioxidant Activity of Secoisolariciresinol Diglucoside-derived Metabolites, Secoisolariciresinol, Enterodiol, and Enterolactone. *Int J Angiol*. 2000, 9, 220-225.

86. Bergman Jungestrom, M., Thompson, L.U. and Dabrosin, C. Flaxseed and its lignans inhibit estradiol-induced growth, angiogenesis, and secretion of vascular endothelial growth factor in human breast cancer xenografts in vivo. *Clin Cancer Res*. 2007, 13, 1061-7.

87. Ingram, D., Sanders, K., Kolybaba, M. and Lopez, D. Case-control study of phyto-oestrogens and breast cancer. *Lancet*. 1997, 350, 990-4.

88. Pietinen, P., Stumpf, K., Mannisto, S., Kataja, V., Uusitupa, M. and Adlercreutz, H. Serum enterolactone and risk of breast cancer: a case-control study in eastern Finland. *Cancer Epidemiol Biomarkers Prev*. 2001, 10, 339-44.

89. Dai, Q., Franke, A.A., Jin, F., Shu, X.O., Hebert, J.R., Custer, L.J., et al. Urinary excretion of phytoestrogens and risk of breast cancer among Chinese women in Shanghai. *Cancer Epidemiol Biomarkers Prev*. 2002, 11, 815-21.

90. Piller, R., Chang-Claude, J. and Linseisen, J. Plasma enterolactone and genistein and the risk of premenopausal breast cancer. *Eur J Cancer Prev*. 2006, 15, 225-32.

91. den Tonkelaar, I., Keinan-Boker, L., Veer, P.V., Arts, C.J., Adlercreutz, H., Thijssen, J.H., et al. Urinary phytoestrogens and postmenopausal breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2001, 10, 223-8.

92. Hulten, K., Winkvist, A., Lenner, P., Johansson, R., Adlercreutz, H. and Hallmans, G. An incident case-referent study on plasma enterolactone and breast cancer risk. *Eur J Nutr*. 2002, 41, 168-76.

93. Grace, P.B., Taylor, J.I., Low, Y.L., Luben, R.N., Mulligan, A.A., Botting, N.P., et al. Phytoestrogen concentrations in serum and spot urine as biomarkers for dietary phytoestrogen intake and their relation to breast cancer risk in European prospective investigation of cancer and nutrition-norfolk. *Cancer Epidemiol Biomarkers Prev*. 2004, 13, 698-708.

94. Kilkkinen, A., Virtamo, J., Vartiainen, E., Sankila, R., Virtanen, M.J., Adlercreutz, H., et al. Serum enterolactone concentration is not associated with breast cancer risk in a nested case-control study. *Int J Cancer*. 2004, 108, 277-80.

95. Olsen, A., Knudsen, K.E., Thomsen, B.L., Loft, S., Stripp, C., Overvad, K., et al. Plasma enterolactone and breast cancer incidence by estrogen receptor status. *Cancer Epidemiol Biomarkers Prev*. 2004, 13, 2084-9.

96. Verheus, M., van Gils, C.H., Keinan-Boker, L., Grace, P.B., Bingham, S.A. and Peeters, P.H. Plasma phytoestrogens and subsequent breast cancer risk. *J Clin Oncol*. 2007, 25, 648-55.

97. Ward, H.A., Chapelais, G., Kuhnle, G.G., Luben, R., Khaw, K.T. and Bingham, S. Breast cancer risk in relation to urinary and serum biomarkers of phytoestrogen exposure in the EPIC-Norfolk study. *Breast Cancer Res*. 2008, 10, R32.

98. Zeleniuch-Jacquotte, A., Adlercreutz, H., Shore, R.E., Koenig, K.L., Kato, I., Arslan, A.A., et al. Circulating enterolactone and risk of breast cancer: a prospective study in New York. *Br J Cancer*. 2004, 91, 99-105.

99. Horner, N.K., Kristal, A.R., Prunty, J., Skor, H.E., Potter, J.D. and Lampe, J.W. Dietary determinants of plasma enterolactone. *Cancer Epidemiol Biomarkers Prev*. 2002, 11, 121-6.

100. Rosner, B., Spiegelman, D. and Willett, W.C. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am J Epidemiol*. 1992, 136, 1400-13.

101. Stumpf, K. Serum enterolactone as biological marker and in breast cancer: from laboratory to epidemiological studies [dissertation]. *Folkhälsan Research Center, Institute for Preventive Medicine*. University of Helsinki, Helsinki, (2004) vol.

102. Vineis, P. Sources of variation in biomarkers. In Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N. (eds.), *Application of biomarkers in cancer epidemiology*. IARC Sci Publ, Lyon,(1997) vol., pp. 59-71.

103. Hankinson, S.E., Manson, J.E., Spiegelman, D., Willett, W.C., Longcope, C. and Speizer, F.E. Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period. *Cancer Epidemiol Biomarkers Prev*. 1995, 4, 649-54.

104. Stumpf, K. and Adlercreutz, H. Short-term variations in enterolactone in serum, 24-hour urine, and spot urine and relationship with enterolactone concentrations. *Clin Chem*. 2003, 49, 178-81.

105. Zeleniuch-Jacquotte, A., Adlercreutz, H., Akhmedkhanov, A. and Toniolo, P. Reliability of serum measurements of lignans and isoflavonoid phytoestrogens over a two-year period. *Cancer Epidemiol Biomarkers Prev*. 1998, 7, 885-9.

106. Johnsen, N.F., Hausner, H., Olsen, A., Tetens, I., Christensen, J., Knudsen, K.E., et al. Intake of whole grains and vegetables determines the plasma enterolactone concentration of Danish women. *J Nutr*. 2004, 134, 2691-7.

107. Kilkkinen, A., Stumpf, K., Pietinen, P., Valsta, L.M., Tapanainen, H. and Adlercreutz, H. Determinants of serum enterolactone concentration. *Am J Clin Nutr*. 2001, 73, 1094-100.

108. Milder, I.E., Kuijsten, A., Arts, I.C., Feskens, E.J., Kampman, E., Hollman, P.C., et al. Relation between plasma enterodiol and enterolactone and dietary intake of lignans in a Dutch endoscopy-based population. *J Nutr*. 2007, 137, 1266-71.

109. Lampe, J.W., Gustafson, D.R., Hutchins, A.M., Martini, M.C., Li, S., Wahala, K., et al. Urinary isoflavonoid and lignan excretion on a Western diet: relation to soy, vegetable, and fruit intake. *Cancer Epidemiol Biomarkers Prev*. 1999, 8, 699-707.

110. Vanharanta, M., Voutilainen, S., Nurmi, T., Kaikkonen, J., Roberts, L.J., Morrow, J.D., et al. Association between low serum enterolactone and increased plasma F2-isoprostanes, a measure of lipid peroxidation. *Atherosclerosis*. 2002, 160, 465-9.

111. Hausner, H., Johnsen, N.F., Hallund, J. and Tetens, I. A single measurement is inadequate to estimate enterolactone levels in danish postmenopausal women due to large intraindividual variation. *J Nutr*. 2004, 134, 1197-200.

112. Bowey, E., Adlercreutz, H. and Rowland, I. Metabolism of isoflavones and lignans by the gut microflora: a study in germ-free and human flora associated rats. *Food Chem Toxicol*. 2003, 41, 631-6.

113. Kilkkinen, A., Pietinen, P., Klaukka, T., Virtamo, J., Korhonen, P. and Adlercreutz, H. Use of oral antimicrobials decreases serum enterolactone concentration. *Am J Epidemiol*. 2002, 155, 472-7.

114. Kilkkinen, A., Valsta, L.M., Virtamo, J., Stumpf, K., Adlercreutz, H. and Pietinen, P. Intake of lignans is associated with serum enterolactone concentration in Finnish men and women. *J Nutr*. 2003, 133, 1830-3.

115. Horn-Ross, P.L., Barnes, S., Kirk, M., Coward, L., Parsonnet, J. and Hiatt, R.A. Urinary phytoestrogen levels in young women from a multiethnic population. *Cancer Epidemiol Biomarkers Prev*. 1997, 6, 339-45.

116. Piller, R., Verla-Tebit, E., Wang-Gohrke, S., Linseisen, J. and Chang-Claude, J. CYP17 genotype modifies the association between lignan supply and premenopausal breast cancer risk in humans. *J Nutr*. 2006, 136, 1596-603.

117. Haiman, C.A., Hankinson, S.E., Spiegelman, D., Colditz, G.A., Willett, W.C., Speizer, F.E., et al. The relationship between a polymorphism in CYP17 with plasma hormone levels and breast cancer. *Cancer Res*. 1999, 59, 1015-20.

118. Simpson, P.T., Reis-Filho, J.S., Gale, T. and Lakhani, S.R. Molecular evolution of breast cancer. *J Pathol*. 2005, 205, 248-54.

119. Hirata, S., Shoda, T., Kato, J. and Hoshi, K. Isoform/variant mRNAs for sex steroid hormone receptors in humans. *Trends Endocrinol Metab*. 2003, 14, 124-9.

120. Ogawa, S., Inoue, S., Watanabe, T., Orimo, A., Hosoi, T., Ouchi, Y., et al. Molecular cloning and characterization of human estrogen receptor betacx: a potential inhibitor ofestrogen action in human. *Nucleic Acids Res*. 1998, 26, 3505-12.

121. Anderson, W.F., Chu, K.C., Chatterjee, N., Brawley, O. and Brinton, L.A. Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the surveillance, epidemiology, and end results database. *J Clin Oncol*. 2001, 19, 18-27.

122. Kuiper, G.G., Enmark, E., Pelto-Huikko, M., Nilsson, S. and Gustafsson, J.A. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A*. 1996, 93, 5925-30.

123. Gruvberger-Saal, S.K., Bendahl, P.O., Saal, L.H., Laakso, M., Hegardt, C., Eden, P., et al. Estrogen receptor beta expression is associated with tamoxifen response in ERalpha-negative breast carcinoma. *Clin Cancer Res*. 2007, 13, 1987-94.

124. Murphy, L.C. and Watson, P.H. Is oestrogen receptor-beta a predictor of endocrine therapy responsiveness in human breast cancer? *Endocr Relat Cancer*. 2006, 13, 327-34.

125. Borgquist, S., Holm, C., Stendahl, M., Anagnostaki, L., Landberg, G. and Jirstrom, K. Oestrogen receptors alpha and beta show different associations to clinicopathological parameters and their co-expression might predict a better response to endocrine treatment in breast cancer. *J Clin Pathol*. 2008, 61, 197-203.

126. Saji, S., Hirose, M. and Toi, M. Clinical significance of estrogen receptor beta in breast cancer. *Cancer Chemother Pharmacol*. 2005, 56 Suppl 1, 21-6.

127. Duffy, M.J. Estrogen receptors: role in breast cancer. *Crit Rev Clin Lab Sci*. 2006, 43, 325-47.

128. Matthews, J. and Gustafsson, J.A. Estrogen signaling: a subtle balance between ER alpha and ER beta. *Mol Interv*. 2003, 3, 281-92.

129. Stossi, F., Barnett, D.H., Frasor, J., Komm, B., Lyttle, C.R. and Katzenellenbogen, B.S. Transcriptional profiling of estrogen-regulated gene expression via estrogen receptor (ER) alpha or ERbeta in human osteosarcoma cells: distinct and common target genes for these receptors. *Endocrinology*. 2004, 145, 3473-86.

130. Pettersson, K., Delaunay, F. and Gustafsson, J.A. Estrogen receptor beta acts as a dominant regulator of estrogen signaling. *Oncogene*. 2000, 19, 4970-8.

131. Pelletier, G. and El-Alfy, M. Immunocytochemical localization of estrogen receptors alpha and beta in the human reproductive organs. *J Clin Endocrinol Metab*. 2000, 85, 4835-40.

132. Kushi, L.H., Potter, J.D., Bostick, R.M., Drinkard, C.R., Sellers, T.A., Gapstur, S.M., et al. Dietary fat and risk of breast cancer according to hormone receptor status. *Cancer Epidemiol Biomarkers Prev*. 1995, 4, 11-9.

133. Althuis, M.D., Fergenbaum, J.H., Garcia-Closas, M., Brinton, L.A., Madigan, M.P. and Sherman, M.E. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. *Cancer Epidemiol Biomarkers Prev*. 2004, 13, 1558-68.

134. Yasui, Y. and Potter, J.D. The shape of age-incidence curves of female breast cancer by hormone-receptor status. *Cancer Causes Control*. 1999, 10, 431-7.

135. Suzuki, R., Ye, W., Rylander-Rudqvist, T., Saji, S., Colditz, G.A. and Wolk, A. Alcohol and postmenopausal breast cancer risk defined by estrogen and progesterone receptor status: a prospective cohort study. *J Natl Cancer Inst*. 2005, 97, 1601-8.

136. Harlan, L.C., Coates, R.J., Block, G., Greenberg, R.S., Ershow, A., Forman, M., et al. Estrogen receptor status and dietary intakes in breast cancer patients. *Epidemiology*. 1993, 4, 25-31.

137. Jain, M. and Miller, A.B. Tumor characteristics and survival of breast cancer patients in relation to premorbid diet and body size. *Breast Cancer Res Treat*. 1997, 42, 43-55.

138. Fung, T.T., Hu, F.B., Holmes, M.D., Rosner, B.A., Hunter, D.J., Colditz, G.A., et al. Dietary patterns and the risk of postmenopausal breast cancer. *Int J Cancer*. 2005, 116, 116-21.

139. Olsen, A., Tjonneland, A., Thomsen, B.L., Loft, S., Stripp, C., Overvad, K., et al. Fruits and vegetables intake differentially affects estrogen receptor negative and positive breast cancer incidence rates. *J Nutr*. 2003, 133, 2342-7.

140. Gaudet, M.M., Britton, J.A., Kabat, G.C., Steck-Scott, S., Eng, S.M., Teitelbaum, S.L., et al. Fruits, vegetables, and micronutrients in relation to breast cancer modified by menopause and hormone receptor status. *Cancer Epidemiol Biomarkers Prev*. 2004, 13, 1485-94.

141. Suzuki, R., Rylander-Rudqvist, T., Saji, S., Bergkvist, L., Adlercreutz, H. and Wolk, A. Dietary lignans and postmenopausal breast cancer risk by oestrogen receptor status: a prospective cohort study of Swedish women. *Br J Cancer*. 2008.

142. Touillaud, M.S., Thiebaut, A.C., Fournier, A., Niravong, M., Boutron-Ruault, M.C. and Clavel-Chapelon, F. Dietary lignan intake and postmenopausal breast cancer risk by estrogen and progesterone receptor status. *J Natl Cancer Inst*. 2007, 99, 475-86.

143. Dontu, G., El-Ashry, D. and Wicha, M.S. Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol Metab*. 2004, 15, 193-7.

144. Ponti, D., Zaffaroni, N., Capelli, C. and Daidone, M.G. Breast cancer stem cells: an overview. *Eur J Cancer*. 2006, 42, 1219-24.

145. Polyak, K. Breast cancer: origins and evolution. *J Clin Invest*. 2007, 117, 3155-63.

146. Zhu, K., Bernard, L.J., Levine, R.S. and Williams, S.M. Estrogen receptor status of breast cancer: a marker of different stages of tumor or different entities of the disease? *Med Hypotheses*. 1997, 49, 69-75.

147. Herrington, D.M., Howard, T.D., Brosnihan, K.B., McDonnell, D.P., Li, X., Hawkins, G.A., et al. Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein. *Circulation*. 2002, 105, 1879-82.

148. Onland-Moret, N.C., van Gils, C.H., Roest, M., Grobbee, D.E. and Peeters, P.H. The estrogen receptor alpha gene and breast cancer risk (The Netherlands). *Cancer Causes Control*. 2005, 16, 1195-202.

149. Wedren, S., Lovmar, L., Humphreys, K., Magnusson, C., Melhus, H., Syvanen, A.C., et al. Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. *Breast Cancer Res*. 2004, 6, R437-49.

150. Zheng, S.L., Zheng, W., Chang, B.L., Shu, X.O., Cai, Q., Yu, H., et al. Joint effect of estrogen receptor beta sequence variants and endogenous estrogen exposure on breast cancer risk in Chinese women. *Cancer Res*. 2003, 63, 7624-9.

151. Cox, D.G., Bretsky, P., Kraft, P., Pharoah, P., Albanes, D., Altshuler, D., et al. Haplotypes of the estrogen receptor beta gene and breast cancer risk. *Int J Cancer*. 2008, 122, 387-92.

152. Berglund, G., Elmstahl, S., Janzon, L. and Larsson, S.A. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med*. 1993, 233, 45-51.

153. Manjer, J., Carlsson, S., Elmstahl, S., Gullberg, B., Janzon, L., Lindstrom, M., et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. *Eur J Cancer Prev*. 2001, 10, 489-99.

154. Riboli, E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol*. 1992, 3, 783-91.

155. Slimani, N., Kaaks, R., Ferrari, P., Casagrande, C., Clavel-Chapelon, F., Lotze, G., et al. European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. *Public Health Nutr*. 2002, 5, 1125-45.

156. Manjer, J., Elmstahl, S., Janzon, L. and Berglund, G. Invitation to a population-based cohort study: differences between subjects recruited using various strategies. *Scand J Public Health*. 2002, 30, 103-12.

157. Pero, R.W., Olsson, A., Berglund, G., Janzon, L., Larsson, S.A. and Elmstahl, S. The Malmo biological bank. *J Intern Med*. 1993, 233, 63-7.

158. Pero, R.W., Olsson, A., Bryngelsson, C., Carlsson, S., Janzon, L., Berglund, G., et al. Quality control program for storage of biologically banked blood specimens in the Malmo Diet and Cancer Study. *Cancer Epidemiol Biomarkers Prev*. 1998, 7, 803-8.

159. Callmer, E., Riboli, E., Saracci, R., Akesson, B. and Lindgarde, F. Dietary assessment methods evaluated in the Malmo food study. *J Intern Med*. 1993, 233, 53-7.

160. Wirfalt, E., Mattisson, I., Johansson, U., Gullberg, B., Wallstrom, P. and Berglund, G. A methodological report from the Malmo Diet and Cancer study: development and evaluation of altered routines in dietary data processing. *Nutr J*. 2002, 1, 3.

161. Elmstahl, S., Riboli, E., Lindgarde, F., Gullberg, B. and Saracci, R. The Malmo Food Study: the relative validity of a modified diet history method and an extensive food frequency questionnaire for measuring food intake. *Eur J Clin Nutr*. 1996, 50, 143-51.

162. Riboli, E., Elmstahl, S., Saracci, R., Gullberg, B. and Lindgarde, F. The Malmo Food Study: validity of two dietary assessment methods for measuring nutrient intake. *Int J Epidemiol*. 1997, 26 Suppl 1, S161-73.

163. Elmstahl, S., Gullberg, B., Riboli, E., Saracci, R. and Lindgarde, F. The Malmo Food Study: the reproducibility of a novel diet history method and an extensive food frequency questionnaire. *Eur J Clin Nutr*. 1996, 50, 134-42.

164. Parvez, S., Malik, K.A., Ah Kang, S. and Kim, H.Y. Probiotics and their fermented food products are beneficial for health. *J Appl Microbiol*. 2006, 100, 1171-85.

165. Taylor, H.L., Jacobs, D.R., Jr., Schucker, B., Knudsen, J., Leon, A.S. and Debacker, G. A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis*. 1978, 31, 741-55.

166. Merlo, J., Berglund, G., Wirfalt, E., Gullberg, B., Hedblad, B., Manjer, J., et al. Self-administered questionnaire compared with a personal diary for assessment of current use of hormone therapy: an analysis of 16,060 women. *Am J Epidemiol*. 2000, 152, 788-92.

167. Sonestedt, E., Wirfalt, E., Gullberg, B. and Berglund, G. Past food habit change is related to obesity, lifestyle and socio-economic factors in the Malmo Diet and Cancer Cohort. *Public Health Nutr*. 2005, 8, 876-85.

168. *Obesity: preventing and managing the global epidemic*. World Health Organization (WHO), Geneva (2000).

169. Adlercreutz, H., Wang, G.J., Lapcik, O., Hampl, R., Wahala, K., Makela, T., et al. Time-resolved fluoroimmunoassay for plasma enterolactone. *Anal Biochem*. 1998, 265, 208-15.

170. Stumpf, K., Uehara, M., Nurmi, T. and Adlercreutz, H. Changes in the time-resolved fluoroimmunoassay of plasma enterolactone. *Anal Biochem*. 2000, 284, 153-7.

171. Camp, R.L., Charette, L.A. and Rimm, D.L. Validation of tissue microarray technology in breast carcinoma. *Lab Invest*. 2000, 80, 1943-9.

172. Hoos, A. and Cordon-Cardo, C. Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations. *Lab Invest*. 2001, 81, 1331-8.

173. Fuqua, S.A., Schiff, R., Parra, I., Friedrichs, W.E., Su, J.L., McKee, D.D., et al. Expression of wild-type estrogen receptor beta and variant isoforms in human breast cancer. *Cancer Res*. 1999, 59, 5425-8.

174. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005, 21, 263-5.

175. Sonestedt, E., Gullberg, B. and Wirfalt, E. Both food habit change in the past and obesity status may influence the association between dietary factors and postmenopausal breast cancer. *Public Health Nutr*. 2007, 10, 769-79.

176. White, E. Effects of biomarker measurement error on epidemiological studies. In Toniolo, P. (ed.), *Application of biomarkers in cancer epidemiology*. IARC scientific publications, Lyon,(1997) vol. 142, pp. 73-93.

177. Willett, W. *Nutritional epidemiology*. Oxford University Press, New York ; Oxford (1998).

178. Hebert, J.R., Ebbeling, C.B., Matthews, C.E., Hurley, T.G., Ma, Y., Druker, S., et al. Systematic errors in middle-aged women's estimates of energy intake: comparing three self-report measures to total energy expenditure from doubly labeled water. *Ann Epidemiol*. 2002, 12, 577-86.

179. Day, N., McKeown, N., Wong, M., Welch, A. and Bingham, S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol*. 2001, 30, 309-17.

180. Bingham, S.A., Luben, R., Welch, A., Wareham, N., Khaw, K.T. and Day, N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet*. 2003, 362, 212-4.

181. Wareham, N.J., Hennings, S.J., Byrne, C.D., Hales, C.N., Prentice, A.M. and Day, N.E. A quantitative analysis of the relationship between habitual energy expenditure, fitness and the metabolic cardiovascular syndrome. *Br J Nutr*. 1998, 80, 235-41.

182. Wong, M.Y., Day, N.E. and Wareham, N.J. Measurement error in epidemiology: the design of validation studies II: bivariate situation. *Stat Med*. 1999, 18, 2831-45.

183. Willett, W.C., Howe, G.R. and Kushi, L.H. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr*. 1997, 65, 1220S-1228S; discussion 1229S-1231S.

184. Kipnis, V., Subar, A.F., Midthune, D., Freedman, L.S., Ballard-Barbash, R., Troiano, R.P., et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*. 2003, 158, 14-21; discussion 22-6.

185. Goldbohm, R.A., van 't Veer, P., van den Brandt, P.A., van 't Hof, M.A., Brants, H.A., Sturmans, F., et al. Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur J Clin Nutr*. 1995, 49, 420-9.

186. Hu, F.B., Stampfer, M.J., Rimm, E., Ascherio, A., Rosner, B.A., Spiegelman, D., et al. Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. *Am J Epidemiol*. 1999, 149, 531-40.

187. Larsson, S.C. *Diet and gastrointestinal cancer : one-carbon metabolism and other aspects*, Stockholm, (2006).

188. Lahmann, P.H., Lissner, L., Gullberg, B., Olsson, H. and Berglund, G. A prospective study of adiposity and postmenopausal breast cancer risk: the Malmo Diet and Cancer Study. *Int J Cancer*. 2003, 103, 246-52.

189. Lissner, L. Measuring food intake in studies of obesity. *Public Health Nutr*. 2002, 5, 889-92.

190. Arnold, M.S., Funnell, M.M., Herman, W.H., Brown, M.B., Merritt, J.H., Fogler, J.M., et al. Discrepancies between perceived dietary changes and 4-day food records in older adults with diabetes. *J Am Diet Assoc*. 1996, 96, 705-7.

191. Rosell, M.S., Hellenius, M.L., de Faire, U.H. and Johansson, G.K. Associations between diet and the metabolic syndrome vary with the validity of dietary intake data. *Am J Clin Nutr*. 2003, 78, 84-90.

192. Macdiarmid, J.I., Vail, A., Cade, J.E. and Blundell, J.E. The sugar-fat relationship revisited: differences in consumption between men and women of varying BMI. *Int J Obes Relat Metab Disord*. 1998, 22, 1053-61.

193. Ballard-Barbash, R., Graubard, I., Krebs-Smith, S.M., Schatzkin, A. and Thompson, F.E. Contribution of dieting to the inverse association between energy intake and body mass index. *Eur J Clin Nutr*. 1996, 50, 98-106.

194. Hebert, J.R., Ma, Y., Clemow, L., Ockene, I.S., Saperia, G., Stanek, E.J., 3rd, et al. Gender differences in social desirability and social approval bias in dietary self-report. *Am J Epidemiol*. 1997, 146, 1046-55.

195. Li, J.H., Awad, A.B., Fink, C.S., Wu, Y.W., Trevisan, M. and Muti, P. Measurement variability of plasma beta-sitosterol and campesterol, two new biomarkers for cancer prevention. *Eur J Cancer Prev*. 2001, 10, 245-9.

196. Swedish Institute for Infectious Disease Control (Smittskyddsinstitutet) and Swedish Strategic Programme against Antibiotic Resistance (STRAMA) *Swedres 2007 : a report on Swedish antimicrobial utilisation and resistance in human medicine*, Solna (2008).

197. Skliris, G.P., Parkes, A.T., Limer, J.L., Burdall, S.E., Carder, P.J. and Speirs, V. Evaluation of seven oestrogen receptor beta antibodies for immunohistochemistry, western blotting, and flow cytometry in human breast tissue. *J Pathol*. 2002, 197, 155-62.

198. Morabia, A. and Flandre, P. Misclassification bias related to definition of menopausal status in case-control studies of breast cancer. *Int J Epidemiol*. 1992, 21, 222-8.

199. Greenland, S. and Neutra, R. Control of confounding in the assessment of medical technology. *Int J Epidemiol*. 1980, 9, 361-7.

200. Greenland, S., Pearl, J. and Robins, J.M. Causal diagrams for epidemiologic research. *Epidemiology*. 1999, 10, 37-48.

201. Davey Smith, G. and Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003, 32, 1-22.

202. Halton, T.L., Liu, S., Manson, J.E. and Hu, F.B. Low-carbohydrate-diet score and risk of type 2 diabetes in women. *Am J Clin Nutr*. 2008, 87, 339-46.

203. Jacobs, D.R., Jr., Marquart, L., Slavin, J. and Kushi, L.H. Whole-grain intake and cancer: an expanded review and meta-analysis. *Nutr Cancer*. 1998, 30, 85-96.

204. Lang, R. and Jebb, S.A. Who consumes whole grains, and how much? *Proc Nutr Soc*. 2003, 62, 123-7.

205. Lampe, J.W. Isoflavonoid and lignan phytoestrogens as dietary biomarkers. *J Nutr*. 2003, 133 Suppl 3, 956S-964S.

206. Aubertin-Leheudre, M., Gorbach, S., Woods, M., Dwyer, J.T., Goldin, B. and Adlercreutz, H. Fat/fiber intakes and sex hormones in healthy premenopausal women in USA. *J Steroid Biochem Mol Biol*. 2008.

207. Ley, R.E., Turnbaugh, P.J., Klein, S. and Gordon, J.I. Microbial ecology: human gut microbes associated with obesity. *Nature*. 2006, 444, 1022-3.

208. Fooks, L.J. and Gibson, G.R. Probiotics as modulators of the gut flora. *Br J Nutr*. 2002, 88 Suppl 1, S39-49.

209. Van 't Veer, P., van Leer, E.M., Rietdijk, A., Kok, F.J., Schouten, E.G., Hermus, R.J., et al. Combination of dietary factors in relation to breast-cancer occurrence. *Int J Cancer*. 1991, 47, 649-53.

210. Wirfalt, E., Mattisson, I., Gullberg, B., Olsson, H. and Berglund, G. Fat from different foods show diverging relations with breast cancer risk in postmenopausal women. *Nutr Cancer*. 2005, 53, 135-43.

211. Kilkkinen, A., Rissanen, H., Klaukka, T., Pukkala, E., Heliovaara, M., Huovinen, P., et al. Antibiotic use predicts an increased risk of cancer. *Int J Cancer*. 2008, 123, 2152-5.

212. Tamim, H.M., Hanley, J.A., Hajeer, A.H., Boivin, J.F. and Collet, J.P. Risk of breast cancer in relation to antibiotic use. *Pharmacoepidemiol Drug Saf*. 2008, 17, 144-50.

213. Jacobs, D.R., Jr. and Steffen, L.M. Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *Am J Clin Nutr*. 2003, 78, 508S-513S.

214. Jacobs, D.R., Jr. and Tapsell, L.C. Food, not nutrients, is the fundamental unit in nutrition. *Nutr Rev*. 2007, 65, 439-50.

215. Borgquist, S., Jirstrom, K., Anagnostaki, L., Manjer, J. and Landberg, G. Anthropometric factors in relation to different tumor biological subgroups of postmenopausal breast cancer. *Int J Cancer*. 2008.

216. Lin, C.Y., Strom, A., Li Kong, S., Kietz, S., Thomsen, J.S., Tee, J.B., et al. Inhibitory effects of estrogen receptor beta on specific hormone-responsive gene expression and association with disease outcome in primary breast cancer. *Breast Cancer Res*. 2007, 9, R25.

217. Jacobs, D.R., Jr., Meyer, H.E. and Solvoll, K. Reduced mortality among whole grain bread eaters in men and women in the Norwegian County Study. *Eur J Clin Nutr*. 2001, 55, 137-43.

218. Ross, A.B., Kamal-Eldin, A. and Aman, P. Dietary alkylresorcinols: absorption, bioactivities, and possible use as biomarkers of whole-grain wheat- and rye-rich foods. *Nutr Rev*. 2004, 62, 81-95.

219. Chen, C.Y., Milbury, P.E., Collins, F.W. and Blumberg, J.B. Avenanthramides are bioavailable and have antioxidant activity in humans after acute consumption of an enriched mixture from oats. *J Nutr*. 2007, 137, 1375-82.