Risk characterization of familial cancer using the Swedish Family-Cancer Database with a special reference to breast cancer

Pauli Vaittinen

Stockholm 2003

Center for Nutrition and Toxicology
Department of Biosciences at Novum
Karolinska Institutet
SE-141 57 HUDDINGE, Sweden

ABSTRACT

This thesis reports on epidemiological studies on family related cancer in Sweden. The Family-Cancer Database was constructed in 1996 from several national registries. Family relations were available from the Generation register, cancer cases from the Cancer registry and socioeconomic information from Censuses. The sources were linked together using personal ID-numbers and then unidentified. A particular advantage of the database is that the contribution of both parental lineages on cancer risk can be examined. The main method was to calculate relative incidence measures using birth cohorts to compare the study groups selected by index cases, probands.

Cancer risk in young and middle aged offspring was increased approximately 1.1 times from the usual risk when one of the parents had cancer, if both parents had cancer, the risk was 1.3-1.4 times higher. When offspring cancer risk was analyzed by index cancer, same as in parent, the risk elevated to 2-3 times for many cancer sites like colorectum, skin, melanoma and endocrine glands, or even to 4-8 times for uterus, testis and thyroid.

Analysis of familial cancer risks between non-index sites provides etiologic understanding on genetic and environmental risk factors. Novel findings associated parental-offspring sites of pancreas-breast, breast-testis and uterus-nervous system. For these, the familial relative risks were modest, but increased in those whose parents were diagnosed before ages 50. Mutations in known cancer-related genes may explain some of these findings. For melanoma, pancreatic and liver cancer, environmental factors are important etiologic factors and may contribute to the observed familial effects.

The molecular genetic explanation may be that rare dominant single genes increase susceptibility at many sites, or that overlapping sets of genes control susceptibility at multiple sites.

The proportion of familial breast cancer among all breast cancers diagnosed before age 54 was 8.7%. The familial relative risk was about 1.8, but is likely to decrease to about 1.5 or less in the aging population. The higher familial relative risks were evident in young women, being 4.0 when both the mothers and their daughters were diagnosed at ages before 40 years. In mothers and daughters, ovarian cancer risk was increased in combination with breast cancer. Breast cancer was studied in three of the papers.

Cancer cases can be divided into familial cases and sporadic cases, individuals in these groups are at high risk, medium risk or low risk. The high risk cases in the familial group may have hereditary causes. If new cancer genes are found they could affect both the familial and sporadic groups, especially with low penetrance or gene-environment interaction as expected features. Family studies are useful to distinguish both hereditary and environmental risk factors and their possible action together.

Keywords: familial cancer, genetic epidemiology, register study, standardized relative risk.

ISBN 91-7349-723-1

This is **the Internet version** of my published thesis; it is revised because of

- Small grammar errors which do not change the intended or basic meaning of the sentence.
- Better formulations on familial proportions of cancer when combining different cancers on page 39, marked with *italic type.*
- One additional concluding remark on use of the term familial cancer on page 50, marked with *italic type*.
- More readable reference list with help of the updated software EndNote 5 and Br J Cancer style.

When referring to this version of the thesis please mention the Internet address and accessed date, for example: http://diss.kib.ki.se/2003/… (accessed day moth year).

På spikningsdagen
6 November 2003
Pauli Vaittinen

LIST OF PAPERS

The thesis is based on the following papers.

A. Hemminki K, **Vaittinen P.**
National database of familial cancer in Sweden.
Genet Epidemiol. 1998;15:225-236.

B. **Vaittinen P,** Hemminki K.
Familial cancer risks in offspring from discordant parental cancers.
Int J Cancer. 1999 Mar 31;81:12-19.

C. Hemminki K, **Vaittinen P.**
Familial breast cancer in the family-cancer database.
Int J Cancer. 1998 Jul 29;77:386-391.

D. **Vaittinen P,** Hemminki K.
Risk factors and age-incidence relationship for contralateral breast cancer.
Int J Cancer. 2000 Dec 15;88:998-1002.

E. Hemminki K, **Vaittinen P,** Easton D.
Familial cancer risks to offspring from mothers with 2 primary breast cancers: leads to cancer syndromes.
Int J Cancer. 2000 Oct 1;88:87-91.

CONTENTS

GENERAL BACKGROUND

1. Introduction

This project to study **familial cancer or family related cancer** on register based information started in 1996. A newly computerized Swedish national register containing family relations, parents – offspring, had become available. Preliminary calculations using general population and cancer data showed that the linkage to the national cancer registry would produce sufficient numbers of cases in the younger generation to permit statistical comparisons. The pilot study (paper A) on familial cancer showed that parental cancer had an effect on offspring cancer risk.
To recall the scene: hereditary factors in the induction of cancer play a limited role (The Swedish Cancer Committee: Cancer causes and prevention, 1992) and the research on cancer genes had got more publicity (Scientific American, Sept 1996).

Later, the Family-Cancer Database has been used for several purposes:

a) to analyze for familial/genetic predisposition to cancer at defined sites and between sites
b) to model for patterns of inheritance and for dissection of genetic and environmental etiologic components
c) to search for and measure other impact factors/mechanisms in familial and sporadic cancer and
d) to provide basic data for decision-making about collection of tissue specimens for mapping of cancer susceptibility genes.

The articles included in the thesis fall into the categories a) and c) in other words:
to estimate the degree of familial susceptibility for different cancers and to study factors which may affect the risk of familial cancer in contrast with sporadic cancer. The approach was to use nationally registered data to define study groups that were followed up for cancer. The main selection was to decide if a person had a relative with or without cancer. The relative who selected the persons into the study groups was called proband, index case or index person. **Proband** is the family member through whom the family became medically interesting (Med-Dictionary,

2002). The procedure to include persons and cases into study groups should not by default influence the risk calculations. But generally the idea that there would be a third variable or unknown connection, that could influence cancer incidence or death, cannot be abandoned in health statistics (Lomborg, 2002). An investigator dealing with applied statistics proceeds from the conviction that probabilistic judgements express certain properties of the phenomena under study, asserting a set of conditions under which a relationship exists independently of the investigator (Gnedenko, 1976) in introduction.

Familial cancer will be treated in the context of **familial factors**, which describe shared genetic and family influences. Family members share not only similar genes they also live under similar rules, surroundings, socioeconomic conditions and circumstances.

Some relevant terms has to be explained because they introduce the background and help with the formulations in my study; and because epidemiological terminology is not always uniform, alternative explanations or stresses may turn up within this study.

Terms: susceptibility to risk

A **risk factor** is anything that has an influence, raising or lowering, of a person's chance/risk/probability to get diseased. It can be lifestyle, environmental exposure, genetic make-up, and family history of disease or infection. Some individuals are less sensitive to a risk factor than others and having a single risk factor does not mean that one will get for example cancer, so susceptibility is individual (Rockhill, 2001).

Susceptibility means that a person is vulnerable to infection or disease if he/she has been exposed. On the other hand many infections result in a loss of susceptibility due to the immune system memory. The risk of disease and severity thereof is modified by age. In theory that risk is assumed to influence a person in two ways: by accumulating or by programming which latter means that the biological system responds differently to exposure later in life compared to earlier (Hall et al., 2002). There seems to be a need to define susceptibility in various risk contexts coupled with their purposes. Efforts to protect persons at higher risk may fail due to different interpretations (Parkin & Balbus, 2000). What is **at risk** or what is the significance of findings might rely on the use of statistics in **risk estimation** (Bailar & Bailer, 1999).

Susceptibility genes, that can cause cancer when mutated**,** have been identified primarily through studies of unusual cancer cases (Li, 1995). These genes also operate in subgroups of common cancers like breast, ovary, colon and melanoma. For other cancers such as prostate there are linkages to chromosomal regions - susceptibility loci – as leads to candidate genes (Simard et al., 2003; Simard et al., 2002; Wiklund et al., 2003). Late onset and multitude of contributing factors complicate the task (Peters & Ostrander, 2001).

Most cancers are considered to be **sporadic** with single occurrences in families and without documented causes other than gender (sex) and old age. Some of the sporadic cases might by chance have been exposed for cancer causing radiation or chemicals. When there are found new groups with systematic exposure, the group members might be considered as predisposed or susceptible for the actual disease. One current estimate for sporadic cancer is about 90%. No more than 2% of cancer arises in families with strong hereditary susceptibility connected to specific rare cancer types or syndromes. Up to 10% of cancer cases show moderate familial clustering of a common cancer or cancer pattern. This is in accordance with my study (see RESULTS)

Penetrance is the proportion of individuals who will appear with the disease phenotype given that they had the genotype with the mutated gene known to contribute to the disease. Hence penetrance is an estimated probability derived from a temporal information (Berry et al., 2002; Iversen & Chen, 2002). The distinction between sporadic and hereditary cancer is not absolute; modifying factors/genes can affect the susceptibility. There are sporadic cancers with significant hereditary component where the genes involved are expected to have low penetrance which make them very hard to locate (Ruivenkamp et al., 2002).

More relative epidemiological terms

Epidemiology studies disease patterns and their determinants in a population in order to give a general validity to the conclusions. An epidemiological study involves people with and without disease and calculation of comparable risk estimates (Potter, 2001). This procedure is not always successful and may result in contradictory results in risk factor epidemiology (Breslow, 2003).

Incidence is the number of new cases of a disease that occur during a specific time period in a population at risk for developing the disease. It can be expressed as the annual number of new cases, then indicating the volume of new patients, or as **incidence rate**, the number of disease onsets divided by the person time at risk. It is usual to calculate the person time in years. If 100000 persons were followed for one year and the new cases were summed up one would get a number which is the rate of new cases per 100000 person-years (= new cases / risk time in years * 10^5).

Prevalence is the number of existing or prevalent cases in the population who have the disease at a given point in time, or more seldom, period of time. For many diseases prevalence describes the number of persons requiring care (Parkin et al., 2001). Cancer patients alive five years after the diagnosis can be considered to be cured (or without extra need of medical care) if their death rates are similar to those in the general population. For several cancers the death rates are higher even after five or ten years (Stenbeck et al., 1995). Advances in early detection, treatment and survival have increased the number of persons with cancer who in the future are at risk for a second tumor. The practices of classification and coding multiple independent cancers have developed consequently (Crocetti et al., 1996).

Relative risk (RR) is the ratio between two proportions or two rates: the disease rate among the exposed to the disease rate among the unexposed, or related/referenced to a more general rate. Alternative reference rates could be obtained from the nationally published rates (not used in this study). If a general validity is aspired the rates or proportions ought to be made comparable with respect to relevant factors such as age, sex, time period and socioeconomic class before forming the ratio RR. The procedure to not explicitly compare but control for social class, age or sex in disease is called **standardization** and can be done in different ways (taken up in the methods section). On the other hand if social class or other differences might influence the conclusions they should be presented separately and not only in the standardized form.

RR measures association or relative effect in a population, which in my study is divided into those called familial susceptible and the rest, based on the registered information. The following example clarifies what kinds of relative risks can be calculated. Let R_1, R_0 and R denote the rates (or risks) of the disease in the familial group, the rest group and the whole group, respectively. The relative risk can in a routine manner be expressed as $RR=R_1/R_0$ using the unexposed group as denominator.

In a population based setting I would prefer to use the whole group (exposed + unexposed) as the reference group because it can give more reliable estimates and add comparability to other studies. Then the relative risk among exposed can be expressed as $RR_1 = R_1/R$ and the relative risk among the rest (unexposed) as $RR_0 = R_0/R$, the whole group is left at the usual/general relative risk with RR = R/R = 1.00. Using some proxy values from the study, calculations from the second version of the database, we get: RR_1=1.10 meaning that the familial exposed group had 10% higher risk than the usual risk to get cancer. While RR_0=0.98 indicates that the unexposed group had 2% lower risk than the usual risk.

Bias can be defined as any difference between the true value and the actually obtained value due to all causes other than random variability in measurement and in sampling. In other words it is a 'systematic' error or deviation in results or conclusions (Clarke & Oxman, 2001). Bias means "different". **Confounding** or an unaccounted variable may introduce differences between the study groups, which do not reflect the true differences/conclusions from the used factors/variables (Ahlbom & Norell, 1990). Statistical theory is used to control and measure the uncertainty in estimation. It results in security in terms of statistical significance, which normally and unfortunately does not comprise bias and confounding. A related problem is how to pass the insecurity of the group level exposure information to risk estimation and its presentation. **Confidence interval (CI)** is calculated on the basis of the assumed distribution of cases in the (target) population. It is the likely range for the true value and it is situated around the estimated value, usually at the 95% confidence. The simplest way to add confidence is to apply the 99% confidence level (Fleiss, 1986).

The Poisson distribution is used to model the number of cancers as rare events occurring in a given time period. The number of events counts up to 0, 1, 2, etc. To characterize the Poisson distribution we can construct one so that outcome 0 event has the probability 5%. Then, it can be calculated from the Poisson density function that the mean of this distribution will be 3.0 events, outcome 6 events happens to have probability 5% too and the cumulative probability for outcomes 7 and over is less than 4%. **Figure 1** shows the outcome probabilities, for outcomes 11 events and over the probabilities are too small to be visible on the present scale.

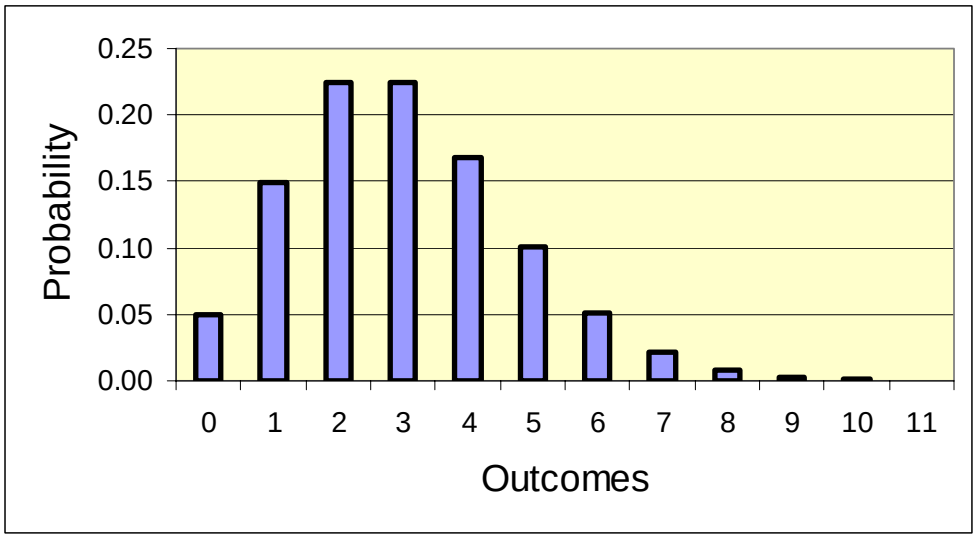

Figure 1. Poisson probabilities for outcomes with mean 3.0 events.

P-value is the probability that the observed findings could have come up due to chance under the conditions given or asserted for the case distribution. More formally it is the probability for a studied variate **X** to assume a value more extreme or equal than the observed value by chance: Probability(**X** >= **X**observed).

Confidence interval and p-value are interchangeable: getting a p-value less than 0.05 is the same as having a 95% confidence interval that does not overlap zero. Compared with the p-value the confidence interval is more informative because it gives both the likely range for an effect and an indication of the uncertainty in its estimation (dos Santos Silva, 1999).

If the significance calculations/tests fail, for example due to bias, one can find oneself in the **type I error, α-error** or **false alarm** situation at the "going rate" of 5% when there in reality was no effect. Bias should always be considered as a possible explanation for new findings that are statistically significant (Sterne & Davey Smith, 2001).

Large **sample size** can give statistical significance to findings that are not clinically or practically significant. Low increases in relative risks, like RR=1.10 or less, can obtain statistical significance when common parental cancers select large groups of offspring, while a rare parental cancer that produces a high score in offspring risk without statistical significance, might possess clinical interest (Froehlich, 1999).

Genetic epidemiology

Because this thesis deal with the familial distribution of a trait, susceptibility to cancer, understanding its possible genetic basis can be placed into the inter-disciplinary field called genetic epidemiology. To distinguish an entity from its parents, genetics and epidemiology, genetic epidemiology has been characterized as having three focuses. One on population-based studies, second on the joint effects of genes and environment and third on incorporating the underlying biology of the disease into its conceptual models, both descriptive and mechanistic (Thomas, 2000). By population-based it is meant that the study design is aimed to test hypothesis that can be generalized to populations at large rather than to narrow sub-populations or inbred pedigrees.

Mechanistic models include the two-mutation clonal-expansion model, which was initially developed to describe the genetics of retinoblastoma (Knudson, 1971; Moolgavkar & Knudson, 1981). Another such model is the breast tissue-aging model (Pike et al., 1983). These types of models provide a framework for testing effects of potential risk factors over time using incidence data with or without family information (Rosner & Colditz, 1996). A challenging field, another project, would be to assess the predictive capacity of individual risk models used in medical counseling. For example the Gail model, which is used for predicting individual breast cancer risks under certain conditions (Euhus, 2001; Gail et al., 1989).

2. From General to Familial Cancer

A glossary definition of cancer is an abnormal, potentially unlimited and disorderly new tissue growth, neoplasm. Natural selection is involved in populations of any organism with hereditary variation in reproductive capacity. Genotypes that reproduce faster are more likely to dominate in later generations. At the cellular level this can be applied to the population of cells. If a somatic mutation creates a variant that reproduces or proliferates faster, then the mutant clone will tend to take over the site. Anti-tumor mechanisms may stop this step from further development by either repairing or killing the potential tumor cells. Other mutations can affect the stability of the entire genome. Cancer is said to depend on these three types of mutations

developing in stages and affecting a target tissue (Strachan & Read, 1999), in chapter on Cancer genetics. Because tumor emancipation from growth control involve changes in many genes and proceed along various pathways, each tumor can be regarded as a unique individual (Klein, 2000).

One genetic event can be a mutation in oncogenes, tumor suppressor genes, mismatch repair genes or angiogenesis promoting genes (vascual stimulation) and is either inherited or somatically acquired. Mutations are also occurring in cells during developmental growth, before and during tumorigenesis. These component mutations will subsequently be detectable throughout the tumor. It is believed that most cancers have accumulated thousands of mutations, most of them in non-coding DNA (Tomlinson et al., 2002).

Solid tumors develop in **stages** operationally defined as initiation, promotion and progression. During progression host processes support tumor development and cells acquire the ability to invade first locally and then distally. The stage of the cancer refers to how much it has spread, what is its extent. Another way to classify a tumor is by its **grade**, which refers to how closely it resembles the tissue of origin, by its differentiation (Lakhani, 2003).

Cancer as multi-step process

Around 1990 Fearon and Vogelstein suggested how the genes mutated in the order of the histological stages during the tumor progression from adenoma to carcinoma. The steps required involved the mutational activation of an oncogene coupled with the loss of function in several tumor suppression genes (Fearon & Vogelstein, 1990; Vogelstein et al., 1988).

Models of carcinogenesis were originally developed to provide a framework to view and to ask questions about the process behind (Moolgavkar, 1994). The basic assumptions in the biologically based model are now that critical mutations in stem cells start to accumulate when after these cells transform into a malignant state either spontaneously or as a response to a promoter (Moolgavkar, 2002). The number of genetic changes that is required to convert normal human cells into cancer cells has been extrapolated to at least four to six. The two-hit model for retinoblastoma would spell that at least two events were needed (DiCiommo et al., 2000). Immortalization is an essential feature of a tumor cell. Forcing this last barrier is regulated by telomere

shortening and by the retinoblastoma and p53 tumor-suppressor pathways (Hahn & Weinberg, 2002). Genome instability in tumors involves other chromosomal regions than the cancer gene loci (Hanahan, 2000). The steps of the process towards cancer can represent genetic mutations and epigenetic changes that alter the behavior of cells (Ponder, 2001). Epigenetic lesions involve environmental factors modulating changes in the cellular carcinogenic pathway. The phenotype from a certain mutation is not always the same because of this interaction (Gunter, 2001).

There are two theoretical models of tumor evolution, which are called the "inverted pyramid" and the "nexus model" (Ilyas et al., 1999). In the inverted pyramid model there is a large degree of interdependence between the steps, for example event A requires event C or D to proceed towards the next stage in tumorigenesis. By this way different pathways are formed where the capital letters present mutations at each step. Colorectal cancer is the classical example of the genetic pathway for multi-step carcinogenesis. There are five stages needed for normal tissue to convert via mild, moderate and severe dysphasia, abnormality in growth or development, to locally invading tumor and finally to metastasizing tumor. In the nexus model the mutations are not interdependent, meaning that there would be a set of possible mutations at each level and it would not make a difference via which mutation the next level was reached.

These are the basic models that can be made more complicated by introducing interacting genes, modifier genes or bystander events. Example of bystander effect could be that genomic instability caused by increased mutation rate arises as a consequence of a mutation selected for a completely other reason, for which the endpoint, genomic instability and cancer, was a harmful side effect. Understanding the interactions between mutations may affect the therapy and prevention. If it were known that one specific mutation is needed for the further stage then that mutation would be the target for screening and gene therapy.

Generally it is thought that familial hereditary cancer is one step ahead but not different in other ways. In familial hereditary cancer the mutation could already be present in the genome. The two-stage model for carcinogenesis provides an early framework within the rare hereditary cancers may not be different from the sporadic (Weber et al., 1983)**.**

Except the rare familial predisposition where germline mutations play the key role it is believed that the combined action of environmental factors and individual susceptibility determines an individual's risk of cancer (Minamoto et al., 1999).

Gene-environment interactions

Before the era of molecular genetics, in 1951, it was asserted by Barbara McClintock that elements in (maize) cells were capable to move from one location to another and alter the cell's instructions, thus explaining fundamental processes in cellular organizations (Briggs & Chudley, 2000). It was hypothesized that there were controlling factors in genome that affected gene expression. From the late 1980s and onwards the molecular genetic research has revealed new genetic mechanisms underlying cancers, defining it as a genetic disease at the molecular level.

The relative roles of genetic and environmental factors in causation of cancer are causing confusion (Hoover, 2000). Geographic differences in cancer incidence, varying trends over time and studies of migrant populations implicate environmental factors as the major cause. Many responsible carcinogens have been identified e.g. tobacco, radiation, occupational toxins, infectious agents as human papilloma virus (HPV) and aflatoxin and dietary compositions. Based on these findings the widely accepted estimate is that 80-90 percent of human cancer is due to environmental factors.

In a study of 21000 Swedish twins it was suggested that individuals might possess a genetic susceptibility to cancer in general (Ahlbom et al., 1997). A Finnish twin study estimated the inherited genetic factors to account for 18% (95% CI: 4-32%) of the variation in overall cancer risk (Verkasalo et al., 1999). One more recent population based twin study showed large effects of hereditary factors at a few common cancer sites (Lichtenstein et al., 2000). The estimated proportions of heritable risk factors were 35% for colorectal cancers, 27% for breast cancers and 42% for prostate cancers. These three cancers were the only to show significant hereditary effects in the population of 90000 Nordic twins, which number is not high enough for more precise conclusions. Nevertheless it can be argued that 1/3 of all cancer could be due to heritable factors, if the model and data behind that study were correct, unchangeable.

A part of the heritable effects might relate to genetic modification of environmental risk factors as well as the other way, environmental modification of the genetic risk factors i.e. interactions between genes and environment. More innovative study designs and larger studies with exposure information would be needed to reveal susceptibility genes and environmental factors participating in such interactions (Brennan, 2002). Nontraditional approaches like case-only design could be used to assess the possible interaction itself (Yang et al., 1997). Multiple factors are normally leading to the concept of interaction (Botto & Khoury, 2001).

Major cancer causes and familiality

The three major factors causing or contributing to human carcinogenesis are considered to be cigarette smoking, infections, and nutrition and dietary factors (Sugimura, 2000). Descriptive epidemiological data support an environmental etiology for many cancers, although the cause itself was unknown (Higginson & Muir, 1977).

In the *Harward Report on Cancer Prevention* from 1996 cancer mortality risks from established causes were estimated. Cancer deaths attributable to tobacco and diet/obesity got 30% each, 60% of total cancer deaths; the percentage due to family history of cancer was estimated to five. Sedentary lifestyle, alcohol and socioeconomic status together accounted for 11% of cancer deaths. The Harward report focused on cancer prevention at the population level. Attributable fraction is used as a measure of public health significance (Begg, 2001).

While lung cancer has strong **environmental determinants**, specially smoking, genetic factors may still have an influence. About 90% of lung cancers worldwide are attributable to smoking, while less than 20% of smokers develop lung cancer. Mildly elevated risks for tobacco-related cancers are found in persons with a history of these cancers in relatives (Wunsch-Filho et al., 2002). The list of tobacco-associated malignancies is long. The organs which are associated with a higher cancer risk and which are not in direct contact with smoke include stomach, kidney, bladder, pancreas and colorectum (Giovannucci, 2001). In the 2003 IARC (International Agency for Research on Cancer) monograph (volume 83) on involuntary smoking colorectal cancer is not on that list but both uterine cervix cancer and myeloid leukemia are. The

IARC working group had also considered studies with evidence for synergy between smoking and some occupational causes, alcohol use and HPV infection.
There are two ways how genetics or familiality comes into the picture: Danish twin and adoption studies have suggested that genetic factors can play a role in smoking behavior (Lawlor, 2002; Osler et al., 2001). Another question is why some 85% of smokers do not get lung cancer. There may exist protective alleles or mechanisms in the case of lung cancer as found with some other diseases, for example with coronary heart disease (Potter, 2001).

It has been estimated that about 15% of the worldwide cancer incidence 1990 were attributed to **infections**, the figure being higher for developing countries and lower for developed countries. Infections with hepatitis B and C, HPV, human immunodeficiency virus (HIV) and Helicobacter pyroli were among the more common viruses/bacteria, which were considered (Pisani et al., 1997).

Nutrition and dietary carcinogens can be divided into micro-components, genotoxic or protective vitamin-related, and macro-components including the total calorie intake. In the multi-step process of carcinogenesis it can easily be appreciated that various genitoxic substances cause genetic alterations in many cells. Examples of dietary carcinogens are smoke from grilling, mycotoxins and nitrites. The macro category consists of nutrients whose excess intake causes tumor promotion. These include salt, fat and total calorie intake when resulting in fat deposits (Sugimura, 2000).
Although there are inconsistencies, the risks of breast, prostate, colon, lung and liver cancers are commonly associated with dietary patterns, which could modify the incidence and behavior of tumors. The inconsistencies may reflect the multi-factorial nature of cancer (Milner, 2002). At the macro level, high intake of fruits and vegetables is considered appropriate to lower the risk of cancer (Gerber et al., 2002).
The relatively recent setbacks concerning beta-carotene and lung cancer and antioxidant pills call for cautiousness in drawing oversimplified conclusions on diet related issues (Albanes et al., 1996; Lancet, 2002). More general considerations about the understanding and validity of epidemiological studies have also been addressed (Smith & Ebrahim, 2001; Smith & Ebrahim, 2002).

It can be stated that dietary issues are complicated as seen in two more examples, with heterocyclic amines and acrylamide where the cancer risks have not been established (Augustsson et al., 1999; Mucci et al., 2003).

Work-related cancer may not be easy to connect with families, when occupations currently more seldom than before run in families. In a Nordics study it was established that farmers and gardeners had about 20% lower cancer risk than people in general, while waiters, painters, beverage and tobacco workers had 30-60% higher cancer risk (Andersen et al., 1999). Concerning the related factor socioeconomic status it has been found that high socioeconomic status is associated with excess cancer in colon, breast, ovary and skin melanoma in most female populations. An excess of colon cancer and melanoma was also observed in high-class men (Faggiano et al., 1997). It is notable that colon, breast and prostate cancers have been causally associated with physical activity, which had a “convincing to probable” protective role (Friedenreich, 2001).

Standardized incidence (SIR) and standardized mortality ratios (SMR) were used to study breast cancer and the effect of socioeconomic differences and spouses’ occupation in Denmark (Danoe et al., 2003). The risks were highest in academics (with risk ratios 1.39 and 1.29 respectively) and lowest in women in agriculture with risk ratios around 0.76. Breast cancer incidence increased by 38% in women aged 50-64, during the study period of 1970-95. All social classes contributed to the increase. The time trends in social distribution indicate that breast cancer is becoming more frequent.

Lifestyle factors can be studied in family settings using married couples. The cancer risk of cancer patient's spouse was assessed and compared to the general population cancer risk in a smaller Israeli study (Walach et al., 1998). The spouse risk was estimated to be 1.5-1.9 times higher for colon, prostate and female breast cancers.

The Family–Cancer Database has also been used to study spouses' cancer risk. The spouse's risk for the same cancer was increased for lung and stomach cancers, and for melanoma when diagnosed before age 50. Shared lifestyle and environmental factors among couples seem to explain only a small proportion of cancer cases (Hemminki et al., 2001). Residental radon exposure has been connected to lung cancer, and

suggestions have been made towards other cancers especially leukemia, kidney cancer and melanoma (Axelson, 1995; Pershagen et al., 1994).

Familial cancers may be hard to find in low risk populations. On the other extreme high incidence areas may explore familial aggregation although predominantly studied and attributed to environmental factors (Chang-Claude et al., 1997).
Cancer incidence increases by age or when there has been enough time for a tumor to develop. Elderly people may have a slower tumor growth than younger persons have, permitting them to die with, rather than of, the cancer (Burns & Leventhal, 2000).

Because dietary habits to some extent run in families it cannot be excluded that a part of the observed familial cancer risk in this study may be explained by nutrition and diet. Smoking habits and infections have familial patterns. When aging runs in families it can contribute to that several cancers will get the familial label.
There is a chance that a shared family environment explains a familial character and it is especially important for behavioral attributes. Generally it is needed more than a familial tendency to prove that a non-Mendelian character is under genetic control (Strachan & Read, 1999), in chapter on Complex diseases: theory and results. Stomach cancer (gastric cancer) can illustrate the complexity. One's risk for this cancer depends on multiple factors: H. pyroli infection, genetic make up or polymorphism and diet (McKee, 2001). To the shared environmental factors can the behavior when seeking health care be added (Hippisley-Cox et al., 2002).

To conclude it cannot be excluded that a part of the familial extra risk that goes over to the offspring/relatives in my study might be composed of the major causes/common risk factors above, influencing families. Dietary habits, physical activity, housing conditions and preferences in seeking health care can be mentioned as candidates.

3. Familial Cancer

The term **familial cancer** is interpreted broadly: people with familial cancer have one or more other cancer cases in their family. To what extent familial cancer is caused by genes, environment or chance is not outlined by the term. Familial cancer is just cancer or predisposition towards cancer that run in families. Because about 1 of 3 people develop cancer in their lifetime it is not necessarily so that several cancers in the same family imply the hereditary case. Nevertheless, strong familial aggregation should be a prerequisite for further investigation into genetic etiology among other shared exposures and risk factors in environment and lifestyle (Liang & Beaty, 2000). If the familial cases within a family or a pedigree had additional characteristics such as younger age at diagnosis, tendency to bilateral disease or autosomal dominant inheritance pattern in several generations then **hereditary cancer** or **inherited cancer predisposition** would be the likely explanation. One more cardinal feature of hereditary cancer is the excess of multiple primary cancers including specific tumor combinations in affected close relatives (Lindblom & Nordenskjold, 2000; Lynch et al., 1995). It is a matter of view and estimation if a mutated gene should cause a substantially higher cancer risk, more than 10 times, to be in the genetic susceptibility or if also a modest increased risk caused by low-risk genes were included (Goldgar, 2002). That is one reason for the broad estimate that 5-10% of overall cancer has a hereditary component, where an inherited genetic mutation, which is present in all cells of an individual, plays a contributing role.

Familial cancer syndromes

The vast majority of cancer mutations occur in an individual's cancer cells. About 1% of all cancer arises in individuals who have a cancer mutation in all cells from the germline and as a part a known cancer syndrome. Sporadic and hereditary cancers are alike because somatic mutations found in sporadic cases are often found altered in the inherited cancer genes (Fearon, 1997). The more than 20 known familial cancer syndromes can be divided into three main types according to inheritance. The first and most common type is autosomal dominant disorders, which means that the condition will be expressed in individuals who have one copy of the defective gene on a non-

sex chromosome. The second type is autosomal recessive disorders and the third category is syndromes with uncertain mode of inheritance (Lindor et al., 1998).

1) Autosomal dominant: which includes breast/ovarian cancer with mutated BRCA1 gene – breast cancer gene 1 in the chromosomal location 17q21 and breast/other cancer with BRCA2 gene in 13q12-13. Two colon cancer syndromes are included, namely, familial adenomatous polyposis - FAP and hereditary nonpolyposis colorectal cancer - HNPCC with several involved genes. Retinoblastoma, a tumor that develops in the retinal cell at early age and is connected with the RB1 gene in location 13q14. Li-Fraumeni syndrome - with p53 gene mutation in 17p13 has been investigated widely, but not found in all cases. The Multiple Endocrine Neoplasia syndrome - MEN can involve certain tumors in thyroid and pancreas.
2) Autosomal recessive: some very unusual syndromes like Bloom syndrome and Xeroderma pigmentosum belong to this group.
3) Uncertain mode of inheritance: Hodgkin's disease, pancreatic cancer, testicular cancer and carcinoid have been observed to have familial component. All syndromes in this category are rare.

Lifetime risk of breast cancer is about 12% (or 1 to 8), making it the most frequent female cancer, with nearly 30% of all female cancer in the United States and Sweden. Nearly 10% of all breast cancer is accounted for a positive family history (in our data 9%). Because the most important breast cancer susceptibility genes, BRCA1 and 2, account for less than 5% of cases, other genes are expected to be found.

Colorectal cancer is the second frequent cancer with 12% of all cancer and about equal distribution in men and women. Moderate familial clustering is found to be present in up to 20% of cases. There is a portion of 3-7% of cases, which show highly penetrant autosomal dominant pattern (FAP and HNPCC). Familial adenomatous polyposis has been recognized clinically for over 100 years, but the disease gene was isolated first in 1991. HNPCC is also called the Lynch Syndrome. There are strict conditions called "Amsterdam criteria" for susceptible families: (a) at least three family members with colorectal cancer, two of them are first degree relatives, (b) at least two affected generations and (c) at least one case diagnosed before the age 50 (Lynch et al., 1997).

Similar canonical definition exits for the Li-Fraumeni syndrome, but with sarcoma, a connective tissue tumor, diagnosed before 45 years of age together with a second

sarcoma and any cancer at early age in other relatives. Carriers of p53 mutations are at higher risk to develop sarcoma and breast tumors (Chompret, 2002).

Melanoma occurrence is on the rise in the Western Caucasian populations, with a current lifetime risk of 1%. One of ten individuals with melanoma has inherited the predisposition, which manifests in an autosomal dominant way. Inherited alterations of the tumor suppressor gene p16 have been found in a minor part of these cases.

Studies dealing with familial genetic cancer risks

The estimated risks associated with specific mutations are almost entirely based on specific populations or families selected because they had high incidence of a certain cancer. Cancer risks may differ outside such families, even if they share the same susceptibility genes. Other genes or lifestyle factors that raise cancer risk may be present in high-risk families (Begg, 2002).

Segregation analysis deals with statistical fitting of a general model with Mendelian options to the inheritance pattern of a trait in a pedigree. It tests the putative pattern of inheritance using distinct phenotypes. Many traits are however complex and lead to other approaches (Lander & Schork, 1994).

Studies that try to identify disease genes are broadly grouped as linkage studies and association studies. **Linkage studies** examine co-segregation of genetic markers and the disease in interesting pedigrees. They are effective in finding gene variants with strong effects. An **association study** can have a case-control design where people with the mutant gene allele are considered exposed and those carrying the normal gene allele form the unexposed group. These studies can detect disease genes with weaker effects, but confounding bias can easily arise from differences between cases and controls (Ahsan et al., 2002).

Finding genes or familial factors by case-control studies, linkage analysis including positional cloning have been successful mostly in Mendelian cases. But the importance in terms of public health promotion has been limited (Tomatis, 2002). Epidemiological perspective is still important to view how genetic and non-genetic disease risk factors work (Risch, 2000).

Other/earlier population-based studies

NCI, USA: the National Cancer Institute had been assembling family histories since 1971 in the course of collecting human milk samples. Using the mothers as probands the cancer information could be obtained for the relatives. About two thousand cancers were found in the selected 32000 persons. Familial clustering was not significant when all cancer was taken together. But there were more cancers than expected in specific sites including breast, ovary, corpus uteri, skin, lung and colorectum. If the mother had breast cancer the risk was increased by 1.5 in their daughters. If a sister had breast cancer the risk to other sisters was increased three times (Albert & Child, 1977).

Utah, USA: the Utah Population Database has been used since 1980 to examine the familial nature of cancer. Its core is the genealogy of Utah Mormon pioneers and their descendants. There were approximately 10000 mainly unrelated founding pioneers, which is large as founder size. Genetic studies comparing European populations have shown that the Mormon allele frequencies are similar to those of northern Europeans. Techniques named HLA loci (human leukocyte antigen) and electrophoretic loci showed the correlation, while red cell antigen loci corresponded less closely (McLellan et al., 1984).

The Utah Cancer registry was made statewide in 1966. The linkage to the population database in 1992 gave 42000 matching individuals, covering 36% of the individuals in the Utah Cancer registry. The rate of record linkage was lower for females than for males because of name changes, despite the use of maiden name in a probabilistic record linkage scheme (Cannon-Albright et al., 1994).

Familial relative risks in first-degree relatives were calculated using cohorts. The highest familial risks were found for thyroid and colon cancers and leukemia, with 4-8 times the expected risk value. Also breast, prostate and tobacco-related cancers (here lung, larynx, lip and cervix) showed significant familial associations. Early age at onset increased the relative's risk of the same site cancer especially with colorectum, melanoma, prostate and brain/nervous system (Goldgar et al., 1994).

Cancer incidences in Mormons and non-Mormons were compared in a Utah study (Lyon et al., 1980). The cancer sites associated with smoking like lung, larynx, esophagus and urinary bladder showed 50% lower incidence in Mormons than in non-

Mormons. Rates of breast, cervix and ovarian cancers were low in Mormons too. These differences could be explained by Mormon standpoints on sexual behavior, tobacco and drugs use. The observed lower incidence of colorectal cancer was left unexplained.

The genealogical index of familiality was developed to measure the degree of familial clustering. Genealogy refers to family lineage. The idea was to employ the coefficient of kinship defined as the probability of randomly selected homologous genes from two individuals to be identical taken by a common ancestor. In a later Utah study on familial associations between cancer sites they reported that three cancer sites – gallbladder, pancreas and uterus – did not show excess cancer incidence with any other cancer site out of the total 24 sites. This suggests environmental causes or low penetrance genetic causes for those cancers. The other 21 cancers showed significant excess incidence on different levels (Thomas et al., 1999).

Laredo, Texas: a similar study to that in Utah was conducted in Laredo (Weiss et al., 1986). About 80% of the city population from 1870-1981 were included from church records of birth, marriage and death totaling to over 300000 individuals. The Mexican-American study population had generally low cancer risk. Laredoans are predominantly Roman Catholic and have been geographically isolated by waterless rangeland towards the rest of the United States. Statistically significant excess of familial cancer was observed in overall cancer and in breast cancer. Site combinations, which manifested excess familial risk, involved breast cancer and some digestive system cancers. The method they used for comparisons of cancer risks was to calculate stratified standardized incidences selected by probands, basically the same method that I have used.

In all the **Nordic countries** it is possible to study disease/cancer incidences among relatives to patients with the disease, because the existing national civil registration systems, cancer registers and hospital discharge registers (Moller et al., 2002). The main limitation for generation studies might be how many patients are eligible with respect to the follow-up time and comparable calendar time window.

For Danes civil registration numbers were assigned from 1968 and onwards. The links using these numbers to biological/legal parents were established later but cover now Danes born 1968 and onwards (Vaeth, 2003).

Familial cancer has been studied in Denmark on several occasions: nearly 6000 children who had cancer at ages 0-14 were linked to their parents. No generally increased risk for cancer was found in these parents during the follow-up period of 1943-89. The national incidence rates were used for the comparisons (Olsen et al., 1995).

Familial aggregation of colorectal cancer was analyzed up to age 60 at diagnosis. Parents were found to have 80% excess risk while siblings had the double risk. The study population was relatively small, consisting of about 6000 relatives who had developed 325 colorectal cancers during the follow-up (Carstensen et al., 1996).

One more Danish example is the study of Hodgkin's disease and family structure among children and young adults. Siblings of cases were found to be at increased risk for the disease (Westergaard et al., 1997).

PRESENT STUDY

4. Aim of the project

To estimate the degree of familial susceptibility, familial risk, for different cancers and to study factors which may affect the risk for familial cancer in contrast with sporadic cancer.

This means calculation of familial relative risks for cancer with different methods and in varying situations. The study groups are selected by relatives having a certain type of cancer, diagnosed at certain age (or ultimately completed with other familial features distinguishing study groups).

The susceptibility is clear with the familial cancer syndromes. The biological/other plausibility is therefore turned to the cancers, which are considered next to sporadic. The study is population-based and quantitative relying on numerical results. Breast cancer was studied more intensively because it is frequent, complex, has relatively low age of onset and easy to follow up due to good survival.

5. Sources of registered subjects

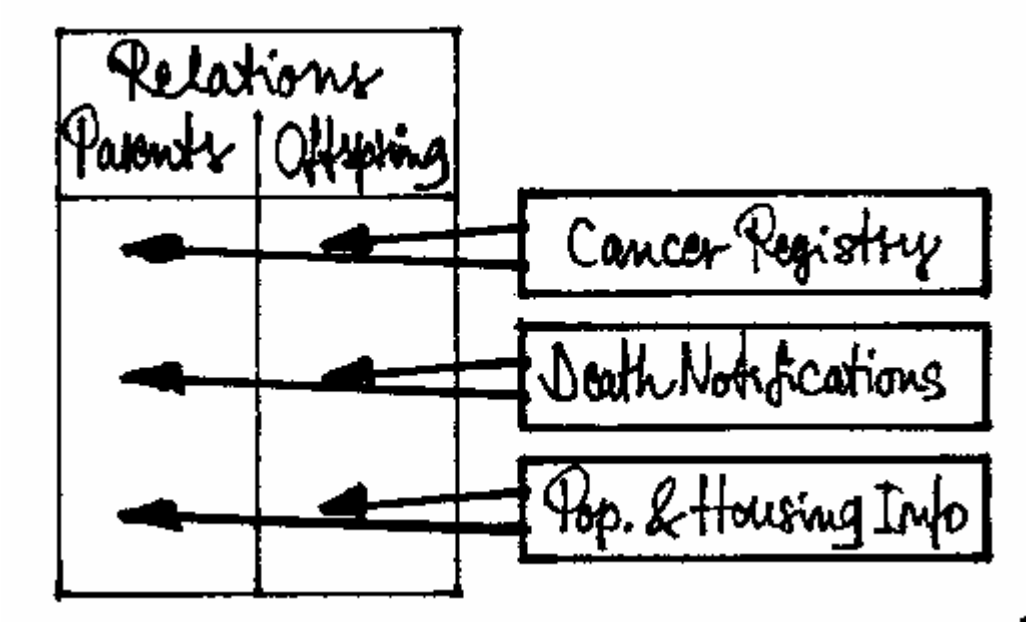

Figure 2. Construction of the Family-Cancer Database. Information from national registers was linked to the relations file. The product database was unidentified.

The Family–Cancer Database was formed from separate sources using the personal identification number as the key when linking together the parts. **Figure 2** shows to the left the Second-generation file, where the rows consist of the identity codes of child, his mother and father. Children/offspring are the unique index persons on the file. One person is only once in the offspring place but might exist several times in the parent place. If the file is sorted by mother, the family relations from mothers are explored, and the mother is called the proband. The same procedure from fathers will give the paternal family structures. It is critical how complete the family file is, how many offspring have both their parents and how many parents have all their children there. The family-relation file is steering the successful linkage of subsequent information to family members from the cancer file and the census files. Each source will be described shortly starting with the linkage key.

The system with **personal ID-numbers** was introduced in Sweden 1947. The last digit in the ID-number was introduced in 1967 to control that the other digits like year month and day of birth were not changed by mistake. However the control does not detect if certain two of them change places, for example that the 2-year digits change place with the 2-month digits. This shortage is a mathematical consequence of the Modula 10 method explained in the information brochure from the Swedish Tax Administration (Skattemyndigheten).

Because **longitudinal studies** that use register data measured over time have become more popular Statistics Sweden (SCB) has taken measures on the problem of changes in ID-numbers. During 1968-2000 there have been in average about 2200 yearly changes. The most changes refer to immigrants who come with new information. It is not unusual that the old ID-number has been valid for several years before the replacement had taken place. Most of these changes can be traced back in registers, except some 4000 persons who were assigned recycled ID-numbers from other persons who had died before 1975 (Alm Stenflo, 2002).

Second Generation Register (later Multi-Generation Register)

The best reference to the Multi-Generation Register and its development is the report 2003:5.1 from Statistics Sweden, which is also available on internet (SCB, 2003).

The Second Generation Register was maintained computerized by Statistics Sweden since 1994. The core of the register was the index persons (=offspring) born in 1941-1992. Most of the data on living persons was obtained from the population registration system in the autumn of 1992. Persons who were born in Sweden 1961-1992, but had died or emigrated during the same period completed the first source from the register of births. Similar completions for birth years 1941-1960 were done from the 1960 Population and Housing Census (FoB60).

The basic condition for the index persons to have their parents in the system was that the parents were registered on population records with their personal ID-numbers since the introduction, year 1947. Another significant year was 1968 when the Population Register was computerized and migration information was included at Statistics Sweden (SCB2001:5, 2001; SCB2001:6, 2001).

In 2000 the name of the register was changed to Multi-Generation Register and yearly updates were implemented. The register included all known relations between parents and children received on population records from the National Tax Board (Riksskatteverket). The target population of the register was widened to index persons who were born in 1932 or later and population registered for some period since 1961.

When the civil registration number was introduced in 1947 the personal files included information on parents for index persons who were at ages 0 to 15. That is why the earliest possible birth year for an index or second-generation person became 1932. If that 15-year-old person had elder brothers or sisters they would not been registered as belonging to the same family. They would possibly been registered later as parents. There is seldom information on parents of an immigrant if he/she was 18 or older when moving into Sweden (SCB2001:5, 2001).

The number of offspring who have information of their parents differs between subpopulations of the register. In the 2002 Multi-Generation Register 85% of all index persons/offspring have mother and 83% have father, those born in Sweden in 1950 or later are almost complete with percentages 99 and 98 respectively. The subgroup born in Sweden and deceased 1961-2002 have mothers and fathers to 76 and 64 percent (SCB, 2003).

As the earlier register name suggests the first usage for the register was to make it possible to produce official statistics on second-generation immigrants and native citizens. Other uses have been to bring together common-law spouses into families in

population statistics, likewise to bring together home living children with their parents. For research the register has been used to study disease, accident risks and education levels in families after linkage to information from other sources (Hemminki et al., 1999).

In Sweden about 2% all live births lead to forensic paternity investigation. In 25% of the cases the tested man was not the father of the child. The qualified guess from the National Board of Forensic Medicine is that 1-2% of the children born in Sweden have another father than other people believe (Naturvetaren, 2000).

The National Swedish Cancer Registry

The Swedish Cancer Registry was established in 1958. At the beginning there was one central registry which collected and coded manual cancer reports under the cover of the National Medical Board (Medicinalstyrelsen). Since the middle 1970s the six medical regions have the responsibility for the primary data collection, coding and registration. The National Board of Health and Welfare (Socialstyrelsen, SoS) maintains the national registry. The coding practices may vary since the decentralization; on the other hand the registries may have more local support. Yearly coding meetings are arranged to coordinate the coding practices between the regions. The coding is done applying the International Classification of Diseases (ICD), presently according to ICD-O/2 where O stands for Oncology. The actual ICD-code is translated back to the 7^{th} version to keep comparability in this respect. Unlike the other Nordic cancer registries the Swedish does not collect cases from the cause-of-death registry (Engeland et al., 1993).

Completeness of the registration has been studied. The undernotification was estimated to 4 % of total cancer incidence in the Stockholm region for the year 1978. The unreported cases were found in the Swedish cause-of-death registry and in the regional in-patient register (Mattsson, 1984; Mattsson et al., 1985).

In a quality control study on multiple primary malignancies it was found that 94% out of 209 diagnoses were correctly reported and registered (Frodin et al., 1997).

Reporting of cancer to the regional cancer registries is done by two separate forms, clinical report and pathology/cytology report. Less than 2% of cancer cases are registered on the basis of one report only (EpC/SoS, 2002).

Information from population and housing censuses

Since 1960 Sweden has carried out national censuses on both population and housing as coordinated total surveys, which took place every fifth year until 1990. Both questionnaires to public and already registered information were used. The scopes of information and variable definitions have varied between the censuses. The purpose has been to provide statistical information on population, households and housing flats for planning, research and public purposes (SCB, 2000).

One pioneer register study was conducted when the Cancer Registry for the years 1961-73 and the 1960 Population and Housing Census were merged to create the Swedish Cancer-Environment Register at the late 1970s. Using sampling techniques it was estimated that non-response in the census and incorrect ID-matching together accounted for 1.7% of the records. No indications were found that erroneously matched persons would charge any specific study group (Eklund & Wiklund, 1979).

The Linked Product: Family-Cancer Database

The Family-Cancer Database was updated for its fourth version in January 2001. Here the structure and main developments with the updates will be covered from the first version. The process started in 1995 with applications for permissions to link together data from national registers. At that time the register contained children born in 1941-1992 who were alive at the end of the construction year of the register, 1992. The restriction to living children caused the loss of about 100000 children. Some 80000 were not included because of lack of one or both parents.

The Second Generation register was linked by the unique ID-number to the Cancer Registry. Cancer cases were diagnosed in 1958-1994. Selected socioeconomic background data was linked from the Swedish Population and Housing Census 1960. After the linkages Statistics Sweden replaced the ID-number by a non-informative unique sequence number thus making the Family-Cancer Database unidentified. **Table I** shows the updates of the database.

Table I. Updates of the Swedish Family-Cancer Database

Version	Offspring born in	Years of diagnosis	Covered individuals	Used in paper
1	1941-92	1958-92	6 million	A
2	1941-94	1958-94	7 million	B,C
3	1935-96	1958-96	9 million	D,E
4	1932-98	1958-98	10 million	
	Subsequent updates have been done later			

In the first version of the database there were about 6 million individuals, as parents and offspring or both. The population of children in the early birth cohorts of 1941-1955 was about 1.5 million. These offspring, together with their biological parents, were chosen as subjects in our first papers on familial cancer at ages 15-51.

The first version included offspring born in Sweden who were living at the end of 1992. This limitation was due to how the second-generation register was created.

For the second version the birth register 1961-92 was used to complete with offspring who were not in the first version because they had died or they had emigrated during the period. Censuses 1960 and 1970 were used in a similar way to complete with offspring born 1941-60.

The third version extended the offspring birth years to earlier cohorts, born 1935-40, although with only those living at the year of update 1997. Migration information from 1987 and onwards was incorporated to improve the follow up. The fourth version extended the coverage of the database further.

In practically every register there is some incorrect or lacking information. A careful approximation with a uniform error rate of 2% in all the three sources will give 2% error in family relations, 2% in diagnosis validity and 2% in other information like socioeconomic status. If these errors draw the result to one direction (which is not likely but possible) then the result would vary in the range +/- 6%.

6. Methods

Cancer incidence varies greatly by age and sex, time period and country. While the general cancer incidence curve grows exponentially with age, there are many contrasts to this rule. For example childhood cancer, which has its highest rates at ages 0-4 with about 25 cases per 100000 children and year or Hodgkin's disease with an even and low rate throughout the adulthood (EpC/SoS, 2002). Different cancers reach their incidence maximum at different ages. For breast cancer the highest incidence (over 300 cases per 100000 women and year in Sweden) occurs at ages 60-69. Among men the most common form of cancer, in prostate, reaches its highest incidence (about 1000 per 100000) at ages 75-84. Despite the differences in age distribution there are about equally many women and men who get cancer in their 60s but somewhat more women than men who get the disease at a younger age.

Standardization with direct and indirect method

The first method to cope with the incidence differences caused by age distributions in the studied populations is the **method of direct standardization**. It is applied when cancer incidence is compared between countries or between time periods within a country. The standard used is a predefined age structure, the special weights given to for example the 18 five-year age classes from 0-4 to 80-84 and 85+. In the IARC publication Cancer Incidence in Five Continents the age structure of the World standard population was applied to weight the five-year age-specific incidence rates into a single measure called the age standardized incidence rate to facilitate comparisons between countries (IARC, 1997). The World standard population reflects the age structure of the world population with its pyramid form. Another age standard population is the European standard, which might be more suitable for Swedish conditions because its even course of weighs (7%) for the five-year age classes between 5 and 54 years.

The ratio between two directly standardized rates is called the **comparative incidence figure - CIF** (Estève et al., 1994). CIF is a relative incidence estimate, a relative risk (RR), and it can be inaccurate (which is reflected in larger confidence intervals) when imprecise age-specific estimates are weighted upwards. This is not likely to occur

when populations and cancer cases derive from countries with similar age structures and registration practices.

It is possible to standardize directly by more factors, for example age and socioeconomic status. The World standard weights can be applied for age and within each age class the socioeconomic classes can be given certain predefined weights, for example 10, 30 and 60 percent for high, middle and low class. This means that the number of the predefined class weighs for (socioeconomic status, age) will be 54 (=3*18). This many classes as standard introduce complications. To make meaningful comparisons would namely require that all the 54 classes in the compared populations should have some minimum amount of person-years at risk to ensure that all single status and age specific incidence estimates were reliable. There are two ways to go on with the non-ideal situation: 1) to restrict the comparison to those age and socioeconomic classes, which are present in both populations or 2) to make the standard classes fewer and broader by merging similar classes together. The first cure will usually work out properly. The second cure might be difficult to accomplish when the incidence of different cancers varies both by age and social class. That is why we come to the next method, when it becomes more likely that some classes under observation will lack cases or person-years.

Indirect method of standardization brings about the alternative comparison**.** Here the study groups are usually smaller for example not containing complete sets of age groups. The interest is to compare the study groups towards some reference experience using age-specific (or any class) rates of the reference population as standards. The expected numbers (E) for the study groups are calculated using the reference rates on the observed risk time. The numbers E are hypothetical; they would occur if the study groups had the specific rates of the reference population (Miettinen, 1985).

This method was chosen as the main method in present thesis and the rates from the whole database were (usually) applied as reference. The relative risk between observed numbers (O) and expected numbers, O/E, is termed the **standardized incidence ratio - SIR**. Poisson distribution for O was assumed in the calculated confidence limits (Estève et al., 1994). The method of SIR will normally produce shorter confidence intervals than the direct method would. The main disadvantage

with this method is that one study group is not formally comparable with another study group, instead the comparisons are only valid towards the reference experience of the study. Because the reference experiences differ from study to study it is not always easy to meta-analyze several results.

Age-Period-Cohort models (APC-models)

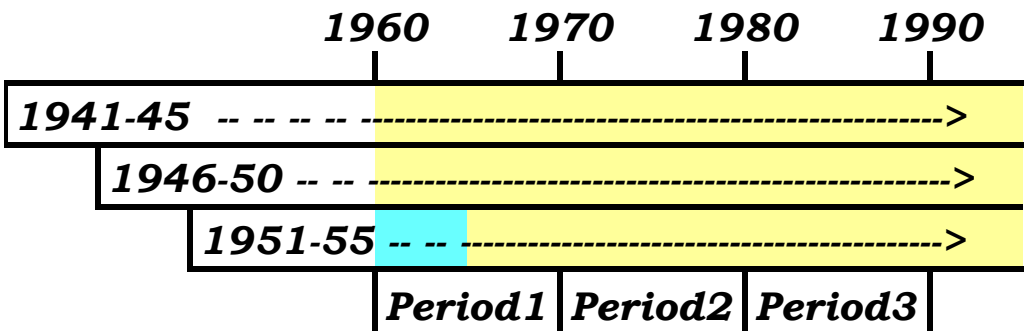

Figure 3. Age-Period-Cohort model. Five-year birth cohorts are followed up for adult cancer through three periods. These three birth cohorts represent the offspring in paper A.

As cancer rates display considerable variation the general approach is to study this variation along three critical dimensions: age at diagnosis, year or period of its occurrence and the birth cohort where the individuals belong (Wilmoth, 1998). Once this basic description is available the analysis continues with specific influences that are assumed to be responsible for the observed incidence patterns in terms of age, period and cohort. The specific influences can include familiality of cancer i.e. to study the cancer incidence patterns of offspring whose parents had certain type of cancer. Other relevant factors like socioeconomic status, place of living can be controlled for. It is assumed that the APC-parameters reflect the real effects of an exposure. It has been found that period effects are of importance in breast cancer incidence in the Nordic countries, therefore calling for more attention on APC time trend analysis (Rostgaard et al., 2001).

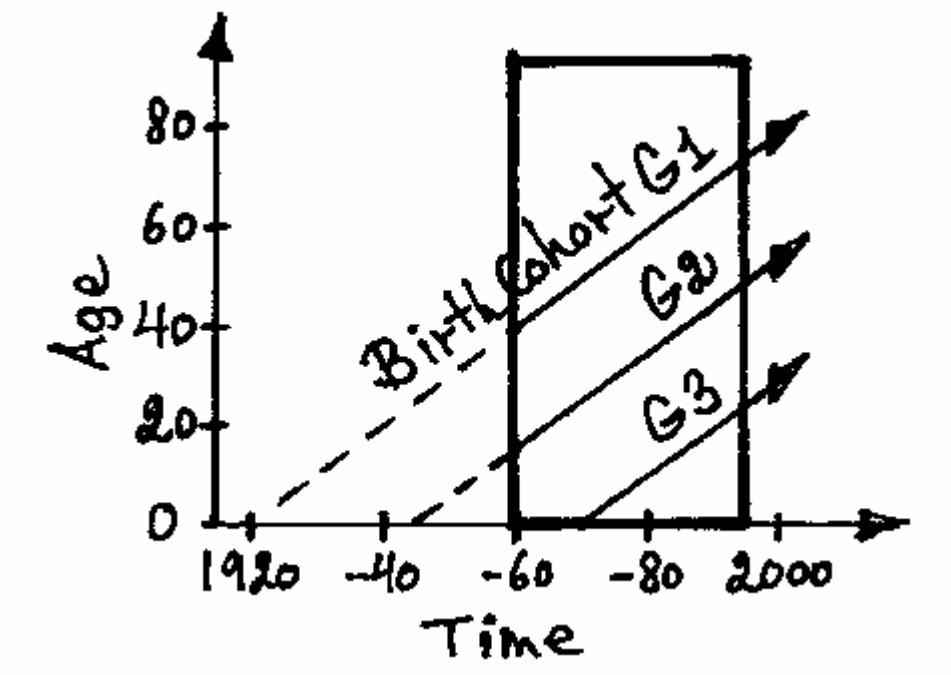

Figure 4. Lexis Diagram. The birth cohorts G1, G2, G3 illustrate how different generations enter and pass the follow-up window.

In the Lexis diagram of **figure 4** the first diagonal line represents the lifetime of individuals born during 1920 (cohort G1). The dotted part of the line marks the lifetime when the individuals were not under follow up for cancer. The second line diagonally under, represents cohort G2, born 1944, belonging to the possible second generation. The cohort G3 was followed for cancer from age zero. These three cohorts visualize different generations in the Family-Cancer Database. From the diagram we can read that the period of diagnosis for certain age and cancer, is delayed from generation G1 to G2 by 24 years (G2-G1), giving that time for conditions to change.

Poisson regression model and procedure GENMOD

A typical use of the GENMOD procedure in the SAS program package is to perform Poisson regression (SAS, 2000). The procedure was used in paper D.

In regression a dependent variable is regressed on a set of independent variables.

In a Poisson regression model the incidence rate (incidences per person-years) can be expressed by Rate = $\exp(\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots)$ or by $\log(\text{Rate}) = (\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots)$ after taking natural logarithm on the both sides of the equal. We introduce Counts, expected number of events, as Counts = person-years * Rate and get

$\log(\text{Counts}) = \log(\text{person-years} * \text{Rate})$

$\log(\text{Counts}) = \log(\text{person-years}) + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$

After modification of an exponential formula we have ended up with a linear expression. The natural log is called the link function and specifies the relationship between the mean Counts and the explanatory factors X. The number of events is assumed to be Poisson distributed with the mean Counts. The βs are the unknown parameters/regression coefficients to be estimated by the statistical (in this case maximum likelihood) procedure. The relationship is termed as log-linear, because log(Counts) is linear in the factors X. Log(person-years) is called the offset variable, it is a regression variable with the regression coefficient (β) fixed as one. The explanatory factors X are used to classify the data to the procedure input as a multi-way contingency table containing the cell sums for events and exposures (measured in person-years).

Interpretation of the parameters: to describe the estimated effect of a certain factor Xj with the parameter βj is to calculate exp(βj) with the corresponding confidence interval. Exp(βj) indicates, when βj is positive, the increase in risk or odds when the factor Xj is forwarded one step and all the other X are kept fixed. If the Xj was age in years then exp(10βj) would give an estimate of change in risk for a 10 year increase in age. If the Xi was parity then the step from null to one is not explained only by parity change but needs to be incorporated with other changing factors like mother's age at childbirth (i.e. there is an interaction and term $\beta_3 X_i X_j$ should be inserted to the formula).

Note that if too broad classes are used for critical factors like age then the Poisson assumption may fail and the procedure can produce unpredictable and false results. Regression models are often useful approximations in many applications, but reality is often so complicated that one cannot know the true model, or the model is only based on the available factors that can be measured (and not the factors that would really be needed). Parameter estimates may depend strongly on particular values or be seriously distorted by a single incorrect data value when put into the "black box". It is important with reasonability checks and control of extreme data values. The procedure output gives some help with model testing and fit.

Potential drawbacks of the study

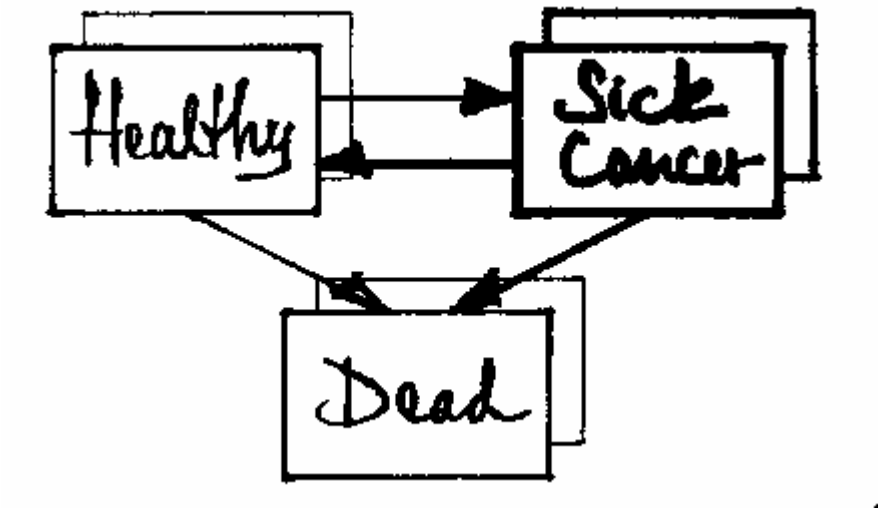

Figure 5. Multiple state model specified in terms of states and transitions. The modeled situation is influenced/conditioned by exposures, screening activities, diagnostic procedures, survival time and generation belongings…

Cancer is a complex disease where environmental, occupational and other hazards can play mixing roles. Precision of diagnosis, early detection and criteria for reporting have changed over time, which means that cancer incidences (trends) are not very reliable over time. Because there is variability and uncertainty at many levels it is possible that cancer risk estimates are not captured by the conventional statistical confidence bounds (Bailar & Gornik, 1997). There is variability in data on different generations. **Figure 5** presents an overview but the factors influencing would depend on if the person belonged to a parent or offspring generation, was a male or female etc. In the analysis the choice of statistical model can affect the inference.

Some bias in epidemiological studies occurs because of imperfect sampling or classification procedures (Vineis & McMichael, 1998). Here will follow some reflections upon selection bias, information bias, intra-individual variation and confounding.

Examples of **selection bias** are hospital settings where patients have many diseases; their values may not reflect those with a single disease. Other "classical" examples are the 'healthy worker effect' and the 'healthy migrant effect' (Razum et al., 2000). This kind of bias could be applied to my study if for example non-severe cancers (cancer *in situ* or not) can vary both in registration and seeking hospital care intensity along

social class or family basis. Differential misclassification occurs when the sensitivity or specificity of the outcome measurements for the exposed (familial) group differ from those for the rest group (dos Santos Silva, 1999) in chapter 2. Selection bias can also arise from better survival, if those who survive longer or a certain time period are more likely to get registered for the second decease/cancer, also the first registration may be affected because of screening.

Detection bias can be considered as a form of **information bias**. For example if the chance to be defined as diseased is conditional on the available clinical services/information which can differ among people belonging different risk groups. Screening activities and the detection of early stage or benign tumors varies and may affect the results in the studied groups. Selection and information biases cannot be excluded in family situations.

Intra-individual variation can occur over time and differing biological sampling or diagnosis procedures. If the Swedish Cancer Registry were set up using the procedures of today, including the autopsy frequency, the result would not be the same, thus affecting this study if repeated, and how the familial groups in different generations would be composed.

Because cancer risk varies strongly by age it may not always be enough to report or study incidence in 5-year intervals; the age difference in such groups can be 4 years causing that cases could belong to different age classes if the ages were chosen in another way. Period can act confounding when the study groups are spread over several periods. In 20 years from 1970 to 1990 the age-standardized incidence has changed dramatically for many cancers. **Table II** shows the approximate incidence changes. Lung cancer is shown separately for males and females because there was a clear difference in trend.

Table II. Changes in the age standardized cancer incidences from 1970 to 1990, measured by cases per 100000 in the Swedish standard population according to Census 1970 (FoB70)

ICD-7	Cancer site	cases per 100000 1970	1990	Change
151	Stomach	28	15	-46%
153	Colon	28	30	7%
163	Lung M	38	42	11%
163	Lung F	9	18	100%
170	Breast F	80	110	38%
171	Cervix	19	10	-47%
177	Prostate	80	110	38%
178	Testis	3	5	67%
190	Skin	10	17	70%
191	Melanoma	7	17	143%

Source: Cancer Incidence in Sweden 1970 and 1990
Standard population 1970: as if the age structure from 1970 was prevailed

Cancers in stomach and cervix have decreased in incidence by 46-47%. Breast and prostate cancers have increased by 38%. The highest increases were noted for the cancers of melanoma 140%, female lung 100%, skin 70% and testis 67%. Colon cancer had the same age standardized incidence as stomach cancer in 1970, but because of divergent development, colon cancer incidence was twice the stomach incidence in 1990. Male lung cancer has had decreasing incidence since the early 1980s, but taken from 1970 to 1990 there was an increase of 11%.

What can also be confounding is that the above incidence changes do not reveal how the trends behave in different age classes. Interventions like screening or other campaigns for new life styles are age/period dependent.

Changes in coding practices, autopsy frequency and improvements in diagnostic practices affect the trends (EpC/SoS, 2002). Overdiagnosis of clinically insignificant prostate cancer through PSA (prostate specific antigen) testing has been approximated to 30% or more (Etzioni et al., 2002). Other cancer sites where screening and testing activities have been in progress are cervix, breast and skin melanoma. It cannot be excluded that the information and tendency for testing goes in families and affect the familial rates in two ways: raising the rates in the younger generation and/or shortening the time between diagnoses in the two generations, especially in the beginning of a campaign.

Decreased age at onset in successive generations has been reported for several diseases, including some cancers (Hsu et al., 2000). The familial explanation of genetic anticipation has to be seen in competition with other possible explanations, from changing environmental risks, the quality and availability of medical care including mammography, to improper selection of probands or at-risk periods.

Table III indicates two observations. At which ages a cancer parent will presumably be when diseased and if that age was considered young or old for the cancer type in question. Secondly the relative survival, comparison between cancer patients and general population, shows which cancers are less fatal and generate more risk time for second primary cancers. Median ages at diagnosis, the typical age, for any cancer were 72 for males and 69 for females. These ages can be compared with mean ages, which were lower: 69 and 67 years respectively.

Table III. Swedish cancer statistics 1999, selected common sites percent of all cancer, among males (M), among females (F) median age at diagnosis and relative survival after 5 and 10 years

ICD-7	Cancer site	Percent of cancer	Median age	Relative survival % 5-year	10-year
177	Prostata M	31	73	73	47
170	Breast F	29	62	84	74
153-154	Colorectum	12	74	58	48
162	Lung	6	69	13	7
191	Skin	6	79	89	85
181	Urinary organs	5	74	70	65
190	Melanoma	4	61	88	83
200	non-Hodgkin	3	70	56	42
193	Brain&Nervous	3	55	48	61
172	Corpus uteri F	6	69	83	79
151	Stomach	2	75	19	17
180	Kidney	2	70	55	42
157	Pancreas	2	71	2	1.5
175	Ovary F	4	64	44	37
178	Testis M	1	34	97	96
140-209	All cancer M		72	56	44
140-209	All cancer F		69	64	57
	Source: Cancer i siffror 2001, Cancerfonden och Socialstyrelsen				

Family structures and sizes have changed a lot during the study period.

Fertility or number of children per female has varied between 2.5 and 1.5 in 1950-2000. It went up in two occasions, to 2.5 in 1964 and to 2.1 in 1990-91. It was low between 1975-85, but went down to its lowest level 1.5 children per female in 1997-99 (SCB, 2001).

In 1999 about 76% of the children in ages 0-17 were living in families with married parents or parents living together, 20% were living with their single mothers and 4% with their single fathers. Low education and living in big cities contributed to the higher frequency of single parent families. The proportion of children in ages 0-17 in the Swedish population has varied, in 1940 it was about 25%, in 1950 it was 27% when after lowering to the present 22% or slightly less. Death rates among 0-1 year old infants and 2-17 years old children have gone down to one third of the 1970 level in the latest 30 years. During the same period the parental age at the birth of the first child has increased by 4 years.

Earlier it was much more usual than today that one parent had died in a family. Around 1950 about 10% of all 16 year old children had one parent dead while in 1999 that percentage had gone down to about 3% (SCB, 2002).

RESULTS AND COMMENTS

The proportion of familial cancer can be calculated as the percentage from the number of certain cancer in offspring who had a parent with that cancer and the number of all that cancer in offspring. Before the proportion is calculated it should be controlled that the offspring at risk for the familial cancer include the same age groups as the total group, all offspring at risk for that cancer. Because the selection of the familial group by cancer parents lead to different age structures depending on the type of cancer and when it was diagnosed, while the total group does not change.

The proportions of familial cancer were approximated to 11% in prostate cancer, 9% in female breast cancer, 5% in lung cancer, 5% in uterine cancer and 2-3% or less for most of the other cancers in an early article (Hemminki & Vaittinen, 1999). The second version of the Family-Cancer Database was used and the ages at diagnosis for the offspring cancer were in the interval 15-53 years.

The straightforward calculation on the latest version of the database (the fifth) with ages at diagnosis 15-68 shows that the above proportions hold with the margin of one percentage unit. What can be added is that the familial proportions of colon and rectum cancers seem to increase.

Combining two parental cancers: if mother or father had colorectal cancer the familial proportions of offspring cancer for both colon and rectum cancers were 8%. Combining skin and skin melanoma cancers in parents will give familial proportions of about 4% in offspring. For female cancers the familial proportion of ovarian cancer raises from 2.3% to 7% when it is combined with parental breast cancer.

Overall the familial proportion of the same cancer in offspring, taken from the parents, was 3-4%. The familial proportion of a cancer does not explain about the cancer risk when compared with the sporadic cancer, but it is a reflection of how frequent a cancer is.

Now, as a proposal, it is the time to move to the reprinted papers or to the concluding remarks or to read the next as a complement to the papers. I consider paper A as introduction, papers B and C as important for understanding familial cancer and papers D and E as focusing on breast cancer related details and how register based information can be used in the analysis.

7. On papers A, B, C, D, E

A. National database of familial cancer in Sweden

This paper was one of the first papers published from the project. The first came out in December 1997 and was called "Effect of paternal and maternal cancer on cancer in the offspring: a population-based study" (Hemminki & Vaittinen, 1997).
The idea was to present the main familial effects observed in the first version of the Family-Cancer database. The information in the database was restricted in several ways. The children/offspring born 1941 and later were alive at the end of the year 1992, which was also the last year of cancer follow-up. The effect of the fatal cancers like those in lung, stomach, pancreas and liver might be underestimated because dead persons were missing in the database. People with these cancers must have survived a longer time or maybe they had a more benign form of the disease. Another restriction was that offspring generation was too young for middle age or late onset cancers. Only ages 17-37 were commonly followed up for offspring cancer in all the three 5-year birth cohorts of 1941-55.

The birth cohort-wise relative risks (RRs) in the offspring grouped by their parents' cancer status were presented. The offspring group who had parents without cancer diagnosis was used as reference. The RRs for offspring with both parents with cancer were around 1.3–1.4. The clue to hereditary cancer, that younger birth cohorts who had younger parents with cancer would have higher familial risks was observed only for sons.
The main cancer sites with elevated familial risks in sons from any parental cancer were colorectum, testis and melanoma with RRs in 1.9-1.4 for offspring, when both parents had cancer. Similarly for daughters the main affected cancer sites were breast, cervix uteri and ovary with RRs around 1.5-1.3. The total elevation of offspring cancer risks was based on about 200 extra cancer cases in sons and about 400 extra cancer cases in daughters.
The final part of the analysis was to see which combinations of parental cancer would give the most effects to children’s cancer risk. The results from site-specific parental cancers are summed up in **table IV** on the next page.

The approximate numbers of excess cancer cases in offspring (in parenthesis) are given to make it easier to compare the relative importance of the excess risk.

Table IV. Relative risk and number of excess cases of offspring cancer, RR (excess N), from parental cancer using offspring with no cancer in both parents as reference (RR=1.00)

Paternal Cancer	Maternal Cancer				
	Colorectum	Lung	Breast	Other cancer	No cancer
Colorectum	**2.6 (14)**	1.4 (1)	**1.7 (14)**	**1.4 (22)**	1.1 (42)
Lung	1.5 (5)	2.1 (5)	1.1 (2)	1.3 (15)	1.1 (39)
Prostate	**1.9 (18)**	1.4 (2)	1.3 (15)	**1.3 (33)**	1.1 (81)
Melanoma/skin	1.6 (3)		**2.4 (15)**	1.1 (2)	1.1 (18)
Other cancer	1.0 (-1)	1.4 (6)	**1.3 (27)**	**1.3 (67)**	1.1 (181)
No cancer	0.9 (-33)	1.0 (-1)	1.05 (34)	1.0 (-1)	1.00
	bolded = significant at the 95% confidence level				

The two parental cancer combinations (father-mother) that gave offspring relative risk higher than two were colorectum-colorectum (RR=2.6) and melanoma/skin-breast (RR=2.4). Combinations of parental cancers that included prostate or lung cancer produced increased RR between 1.3 and 2.1, although the results with lung cancer were not significant due to small numbers.

Conclusions from Paper A: Familial risk of offspring cancer increased specially if both parents had cancer. Certain combinations of parental cancers increased the offspring risk more than other cancers, for example when cancers of colorectum, breast, melanoma or prostate were diagnosed in parents.

B. Familial cancer risks in offspring from discordant parental cancers

This paper was based on the second version of the family-cancer database. The covered ages at diagnosis in offspring were 15-53 and the database was completed with those family members who had died after 1960. Among the more fatal cancers were those in liver and pancreas to show increased familial relative risks.

The analysis starts with an overview answering the question which of the other cancers than the cancer which father or mother had showed increased risks in offspring. These cancers are called non-index or discordant cancers. Direct method of age standardization was used to keep the comparability between the relative risks

between offspring cancers. RRs to discordant cancers in offspring of the magnitude 1.2 or higher were observed from parental cancers of rectum, liver, pancreas, uterus, ovary, kidney, and melanoma. Maternal melanoma gave the RR of 1.66 in sons but no observed increase in daughters.

When coming to the site-specific cancers in offspring the method of direct standardization would not apply because of small numbers or no cases in many age classes. So the indirect method was applied. It uses the observed weights of person-years in study groups together with a reference experience to calculate the expected number. The standardized incidence ratios (SIRs) are not comparable between study groups because these can be composed of very different ages or birth cohorts and may thus weight very differently.

Somewhat outside of the scope of the title of paper B but inseparable in fact is the effect of index cancers to offspring cancer risk. **Table V** presents the significant findings.

Table V. Concordant cancer sites, from parent to offspring
Parental cancers that elevated the risk of same cancer in offspring

Concordant cancer site	Offspring SIR by father	by mother	
Colon	1.8	2.2	**effect seen from both**
Rectum	2.0	2.1	**parents**
Breast	2.5	1.9	
Melanoma	2.4	3.0	
Nervous system	1.6	1.9	
Thyroid	7.8	10	
Other endocrine	3.3	2.4	
Lung	2.0	ns	**effect seen from father**
Prostate	1.9	-	
Testis	5.1	-	
Kidney	2.9	ns	
Skin	3.0	ns	
Cervix	-	2.0	**effect seen from mother**
Uterus	-	4.4	
Ovary	-	2.7	
Lymphoma	ns	1.8	
Leukemia	ns	2.9	

ns = not significant
reference rates from the Family-Cancer Database

Missing from the above list are the index cancers in oral region, stomach, liver, pancreas, bladder and myeloma. These cancers were rare but some of them became significant when taking offspring by father and mother together. In the selected/studied offspring groups the risk for index cancer was at least twice the normal risk for 8 different cancer sites. Thyroid, testis and uterus gave the highest scores, between 10 and 4, among them.

Table VI shows combined results. The comparable incidence figure or CIF, based on direct age standardization, was used to rank the excess risk for non-index cancers. Familial relative risk measured with CIF of at least 1.2 was taken as the limit to pass to the table. For example rectum cancer in mothers elevated the overall non-index cancer risk in sons by CIF=1.27. Further analysis of offspring cancer risk from maternal rectum cancer with the measure SIR gave increased cancer risks in prostate and lymphoma. It should be noted that other parental cancers gave significant effects too, but they were lower, such were lung cancer with CIF=1.14 from father to sons and breast cancer with CIF=1.13 from mother to sons.

Table VI. Discordant cancer sites, from parent to offspring
Parental cancers that elevated the risk of other cancer in offspring

Parent	Parental cancer site	Relative risk to non-index cancer > 1.2	Offspring cancer risk non-index cancer site	SIR	N
Mother	Rectum	sons	prostate	3.9	4
			lymphoma	1.5	32
Mother	Liver	sons	pancreas	2.5	7
			cervix	1.7	32
			prostate	4.0	4
Father	Pancreas	daughters	breast	1.3	89
			uterus	2.2	9
Mother	Uterus	sons,daughters	colon	1.9	36
			bladder	1.6	18
			nervous syst.	1.5	65
Mother	Ovary	daughters	breast	1.8	183
Father	Kidney	sons	bladder	1.9	16
Mother	Melanoma	sons	skin	2.9	9
Indication of elevated cancer risks, method: direct standardization (CIF)			Observed relative risks, method indirect standardization (SIR)		

There was an **age effect** from early onset parental cancers to offspring cancers. When the parental age at cancer diagnosis was grouped as 15-49 and 50-99, it was observed that the younger onset parent group mediated generally two-three fold offspring risks, in some cases up to 10-20 fold risks. These high relative risks suggest connection to cancer syndromes. For example paternal colon cancer with age at diagnosis 15-49 gave FRR=10 in offspring colon cancer and FRR=20 in offspring uterus cancer.

Early onset maternal cancer gave extra high risks, over 10 fold, in offspring index cancers of rectum and uterus.

Conclusions from Paper B: The familial cancer risk from parent to offspring was pronounced at the index cancer site. A few across site cancer connections like those of colon-rectum, kidney-bladder, melanoma-skin and thyroid-other endocrine glands gave elevated risks at the same level as the index cancers.

Early onset effects might be due to cancer syndromes where more than 10 fold cancer risks compared with the normal are found. Parental-offspring cancers of colorectum, uterus and kidney were observed to be involved.

Lung cancer has been found to account for nearly half of the excess cancers after treatment for cervical cancer (Kleinerman et al., 1995). The probable link between lung and cervix cancers was cigarette smoking.

C. Familial breast cancer in the family-cancer database

The paper goes into more detailed study of breast cancer using both mothers and daughters as probands to select the familial group. The daughters to mothers who had breast cancer form one susceptibility group for familial breast cancer. In the same way the mothers with a daughter who had beast cancer form another familial study group. As in paper B the second version of the Family-Cancer database was used.

When mothers with a breast cancer daughter were compared to all mothers it was observed that the risk to get breast cancer in the familial group was higher than the general risk in all ages. Relatively for mothers, familial to normal, the RR was about 3 at age 40, about 2 at age 50 and then leveling off towards 1.5 at age 70.

For daughters the possible ages to analyze were limited to between 15 and 53. The decreasing familial relative risk by age was observed, but with slightly lower values: at age 25 the RR was 4, at 35 it was 2.1 and already at 50 it was 1.5.

These results can be shown in a more concentrated way by taking the mothers by their age at breast cancer diagnosis and calculating the relative risk in three age groups for their daughters. This is done in the next table and shows the pattern of relative breast cancer risk decreasing from 4.0 to 1.5, the age effect.

Table VII. Age specific familial relative risk of breast cancer in daughters by mother's age at breast cancer diagnosis

Mother's age at diagnosis	Daughter's risk (RR, N) by age					
	15 - 39		40 - 44		45- 53	
	RR	**N**	**RR**	**N**	**RR**	**N**
less than 40	**4.0**	19	1.9	3	2.5	3
40 - 49	**2.8**	71	**2.2**	36	**1.8**	32
50 - 59	**2.5**	87	**2.8**	84	**1.6**	60
60 or over	**2.2**	119	**1.8**	136	**1.5**	168
	bolded = 95% significant reference rates from all daughters in the database					

Breast cancer had also connection with other cancers. Maternal breast cancer increased daughters' risk for cancers in ovary with RR=1.4, melanoma 1.2 and lymphoma 1.5. Other cancers in mothers that were observed to have effect on daughter's risk for breast cancer were uterus with RR=1.4 and ovarian cancer with RR=1.8.

There are known susceptibility genes for both ovarian cancer and melanoma namely BRCA2 and p16. Sporadic pancreatic cancer has been diagnosed in some of these susceptibility families (Klein et al., 2001).

Very similar relative risks as in **table VII** have been observed for familial prostate cancer with age classes up to 60, 60-64 and 65 and over for proband's age at diagnosis (Lesko et al., 1996).

Discussion on Paper C: Family history of breast cancer is considered to be one of the strongest risk factors for breast cancer (Arver et al., 2000). In addition the age at onset can indicate the possible genetic subtype (Claus et al., 1990). Mutations in the BRCA1 and BRCA2 genes are known to predispose families for breast and ovarian cancer. In a British study, it was estimated that 6% of breast cancer patients younger than 50 were BCRA1/2 mutation carriers, while among the older patients, 50 and over, 1.3% were carriers. Using the patients' mutation prevalences with current

penetrance estimates gave that the general population BRCA1/2 mutation frequency could be 0.23% and about equally contributed from the two genes (Peto et al., 1999). Because mutations in the known breast cancer genes account for a small proportion of breast cancer risk it is expected that more breast cancer genes are to be found (Nathanson & Weber, 2001). In an Icelandic segregation analysis of 389 pedigrees the familial aggregation of breast cancer was best explained by a high-risk allele which among the homozygotes (with genotype AA) would contribute to the early onset breast cancer. The homozygote carriers of this putative high-risk allele were estimated to be 2.6 % of the Icelandic population, and all carriers (AA+AB) being as many as 29% (Baffoe-Bonnie et al., 2000).

Other genes: p53 gene, which is known from Li-Fraumeni syndrome that shows familial clusters of early onset cancers in breast, soft tissue sarcomas, brain tumors and leukemias. BRCA1 and p53 genes are located in the same chromosome, number 17, but in different arms. Somatic p53 mutations are found in 20-40% of sporadic breast cancer.

Current estimated probabilities that a woman of age 20 would get breast cancer by age 50 are 1.7%, 3.7% and 8% for woman with 0, 1, and 2 affected first degree relatives. Converted to relative risks they would yield 1.0, 2.2 and 4.7 keeping the group with unaffected relatives as reference. Similarly when lifetime, from age 20 to 80, was considered the relative risks were less dramatically rising by the number of affected first degree relatives: RRs were 1.0, 1.7 and 2.7. The lifetime probability for breast cancer, corresponding now to RR=1.0, no affected relative, was estimated to 7.8 percent (Lancet, 2001).

D. Risk factors and age-incidence relationship for contralateral breast cancer

The idea of the paper was to study the subgroups of women who developed a second cancer in connection to breast cancer, and specially if the two cancers were both breast cancers. The third version of the Family-Cancer database was used. Because the laterality of breast cancer was first coded in 1970 we followed up women born 1900-59 for breast and other cancers from year 1970 to 1996.

Term contralateral refers to the opposite side; unilateral cancer affects only one side while bilateral breast cancer would affect both sides equally also in time. Ipsilateral

pertains to the remaining breast when a unilateral mastectomy has removed one breast (Med-Dictionary, 2002).

Among the 72000 women who had had a primary breast cancer as a first cancer had 11% developed a second cancer in breast or other organs. The breast cancer was called contralateral if it had been diagnosed at least 6 months after the first unilateral breast cancer. The 11% of the second tumors were distributed as 3.5% in breast contralaterally, 2.0% in breast synchronously (within 5 months from the first breast cancer) and 5.6% in other organs. So, breast cancer and other cancer were equally frequent when diagnosed as second cancers after breast cancer.

When other cancer was diagnosed as the first cancer, in 190000 women, 6.8% had developed a second cancer, distributed as 1.2% in breast and 5.6% in other organs.

From the incidence curves for the first cancer in females it can be observed that

1) Other cancer is more usual than breast cancer in all age groups.
2) The incidence of breast cancer follows an exponential growth curve until menopause through the ages 20 to 49, after which the incidence is increasing by age but in a slow, stabilizing manner. Other cancer taken as a whole behaves exponentially through all ages up to the age 80. Typically, at ages after 50 other cancer is 2-4 times as common as breast cancer.

Breast cancer was more common as second cancer after breast cancer than after other cancer (females who had non-breast cancer as first cancer were generally older too).

The incidence curves for second cancers were similar to the first cancers if the first cancer was not the same cancer: 2^{nd} other cancer (1^{st} breast) was similar to 1^{st} other cancer and 2^{nd} breast cancer (1^{st} other) was similar to 1^{st} breast cancer in both form and level of the incidence curve.

The main result was a strikingly different pattern with the contralateral breast cancer as the 2^{nd} cancer. In ages 30 to 49 the incidence was as high as 800 cases per 100 000 person-years, that is about 10 times the normal breast cancer incidence. At ages after menopause the incidence was about 450 cases per 100 000 person-years, still about twice the normal incidence.

The high incidence of contralateral breast cancer was also observed in breast cancer families, also the 2^{nd} breast cancer (after other cancer) had an elevated level compared to the normal level. The familial incidence was shaky due to small numbers.

We used Poisson regression to analyze the about 2500 contralateral breast cancer cases with respect to available risk factors in the database. Due to small number of cases the class variables were given broad classes, for the same reason interaction terms were omitted (the classes would have been even wider). The regression model was log(incidence) = intercept + period + birth cohort + familiality + age at first diagnosis.

According to the analysis the significant risk factors were early age at first diagnosis, age 25-34 compared to 75-84 gave 2.5 times higher risk and the interval from the 1st breast cancer, when 0.5-5 years compared to 11-26 years gave 1.3 higher risk. The familial cases were only 147 in number but gave an increased risk of 1.5 compared with the rest considered as sporadic cases. Other risk factors we tested with the model, birth cohort, age at 1st birth and parity did not contribute with significant differences between their classes, although these variables had significant trends. For example the risk for contralateral breast cancer was elevated by birth cohort from 1900-19, via 1920-39, to 1940-59.

Conclusions from Paper D: Contralateral breast cancer had emerged in 3.5% of the women who had breast cancer as the first cancer, out of these (3.5%) familial history of breast cancer was found in 5.8%. The familial cases did not differ from the sporadic cases except for somewhat higher incidence, about 1.5 fold.

E. Familial cancer risks to offspring from mothers with 2 primary breast cancers: leads to cancer syndromes

The third version of the Family-Cancer database was used to analyze cancer risks in offspring whose mothers were registered for two primary bilateral breast cancers. Multiple cancers in one individual have been recognized as a clue to hereditary cancer or syndrome. We wanted to study if new cross-site connections would show up in the offspring group with breast cancer gene susceptibility and how the risk varied by age.

There were about 4700 mothers with bilateral breast cancer who selected the study group of about 9700 offspring. The cancer incidences of all offspring in the database were used as the comparison rates to calculate the relative risks with SIR. The maximal follow-up age for offspring cancer was 61 years.

In daughters the risk of breast cancer was elevated 3.0 times while the ovarian cancer had 1.8 times the all offspring risk. When taken by early age of onset the breast cancer risk was 7.1 fold in ages 20-34 for daughters whose mothers had the age of onset

under 50. Also daughters' melanoma risk showed an early age component. On the contrary ovarian cancer showed a middle age component with SIR=3.3 at ages 45-61. Taking the offspring cancer by histological type we found that glandular epithelium (testis cancer) was elevated in sons to SIR=2.2 and squamos-cell carcinoma in daughters to SIR=1.8. Generally the sons' cancer risk was not elevated in the study group while daughters' risk was nearly doubled.

Conclusions from Paper E: Squamous-cell carcinomas were the second group of cancers, after breast-ovarian cancers, which were observed to have an elevated risk in daughters whose mothers had bilateral breast cancer. This might suggest immunosupression as one mechanism.

CONCLUDING REMARKS

Family related cancer or familial cancer consists of inherited susceptibility, environmental and behavioral attributes. Taken together, parental cancer from one parent raised the offspring cancer risk with 10-15%. If both parents had cancer the offspring risk was elevated by 30-40% from the common risk. Because these categories of offspring with cancer parents were in clear minority the effect of taking away the familial exposure would not lower the common cancer incidence more than 2-3%. Despite persons with no registered parents had higher cancer incidence than those with registered parents. This can be due to that their parents had died, because dead persons (as parents) are missing in a higher extent in the Family-Cancer Database and the relations file.

Cancer is generally a disease of the elderly, with a late onset, but in the hereditary cancer it usually has an early onset.
In my study many familial cancers like breast and melanoma cancers show early age effects, their relative risks are higher for young age groups when compared with same sporadic cancer. It seems to me that there is a connection with the raising incidence trends with these cancers, i.e. the raising incidence can explain a part of the early age effect. This suggests other than only genetic causes.

I propose that 2-3% of the common cancer incidence as a conservative estimate when taking the "strictly" familial cancer: for example cancers diagnosed at ages 75+ should not by default be included in the familial probands or cases, unless there were several similar cases. Secondly, for example socioeconomic status may not be a familial feature at first hand.

Cancer cases can be divided into familial cases and sporadic cases; individuals in these groups are at high risk, medium risk or low risk. The high risk cases in the familial group may have hereditary causes. If new cancer genes are found they could affect both the familial and sporadic groups, especially with low penetrance or gene-environment interaction, as expected features. Family studies are useful to distinguish both hereditary and environmental risk factors and their possible action together.

There is a need that statisticians cooperate with genetic and other experts, although statistical method is not always easy to apply and may be felt as a constraint among eager researchers.

8. Acknowledgements etc.

I want to thank all the CNT staff at Novum, Huddinge: professor Kari Hemminki for driving the project, also the next: Kamila, Dan, Rajiv, Mairon and the Chinese group specially Dong and the India connection and others in seminars and in the IKEA inspired coffee rooms.
Then in turn the Epidemiologiskt Centrum (EpC) and specially there those with the Cancer Registry, still working there, but not forgetting those who were there before and on the regional branches.

Statistics Sweden for good cooperation and service with information and data.
Cancerfonden for the financial support of the project.

Then to my stats teachers in Uppsala, who might by occasion come over to this report, I would answer that stats was not boring, at least not boring enough to continue. Then those bearing beard and others not mentioned.

THINKABLE 1

Barbara McClintock was the Winner of the 1983 Nobel Prize in medicine for her discovery of mobile genetic elements.

Risk assessment involves four steps: hazard/risk identification, dose-response modeling, exposure assessment and risk characterization (Bailar & Bailer, 2001).
Suited here, the familial aggregation of cancer is the identified hazard. Dose-response modeling would sort out how the plausible biological steps and common environmental exposures work in families. Exposure assessment would measure the genetic or environmental load. **Risk characterization** integrates the parts in the risk assessment process into an estimation of the effect of familial cancer in a given Swedish registered population, including accompanying uncertainties.

Risk characterization involves describing and categorizing the uncertainty and variability in risk estimates in order to support public and private decision making (RTI International).

THINKABLE 2

First: if you can get infected, it is biology!
Heard from National Public Radio

Second: epidemiology is hard, but genetic epidemiology might be harder.
Freely after professor Angus Macdonald

Third: Epidemics appear and often disappear without traces, when a new culture period has started; thus with leprosy, and the English sweat. The history of epidemics is therefore the history of disturbances of human culture.
Rudolph Virchow, 1821-1902

Conclusion: is cancer epidemic in a modern society?

REFERENCES

Ahlbom, A., Lichtenstein, P., Malmström, H., Feychting, M., Hemminki, K. & Pedersen, N.L. (1997). Cancer in twins: genetic and nongenetic familial risk factors. *J Natl Cancer Inst*, **89,** 287-93.

Ahlbom, A. & Norell, S. (1990). *Introduction to modern epidemiology*. Epidemiology Resources Inc. MA, USA.

Ahsan, H., Hodge, S.E., Heiman, G.A., Begg, M.D. & Susser, E.S. (2002). Relative risk for genetic associations: the case-parent triad as a variant of case-cohort design. *Int J Epidemiol*, **31,** 669-678.

Albanes, D., Heinonen, O.P., Taylor, P.R., Virtamo, J., Edwards, B.K., Rautalahti, M., Hartman, A.M., Palmgren, J., Freedman, L.S., Haapakoski, J., Barrett, M.J., Pietinen, P., Malila, N., Tala, E., Liippo, K., Salomaa, E.R., Tangrea, J.A., Teppo, L., Askin, F.B., Taskinen, E., Erozan, Y., Greenwald, P. & Huttunen, J.K. (1996). Alpha-Tocopherol and beta-carotene supplements and lung cancer incidence in the alpha-tocopherol, beta-carotene cancer prevention study: effects of base-line characteristics and study compliance. *J Natl Cancer Inst*, **88,** 1560-70.

Albert, S. & Child, M. (1977). Familial cancer in the general population. *Cancer*, **40,** 1674-9.

Alm Stenflo, G. (2002). Personnummerbyten 1970-2000. Statistics Sweden.

Andersen, A., Barlow, L., Engeland, A., Kjärheim, K., Lynge, E. & Pukkala, E. (1999). Work-related cancer in the Nordic countries. *Scand J Work Environ Health*, **25,** 1-116.

Arver, B., Du, Q., Chen, J., Luo, L. & Lindblom, A. (2000). Hereditary breast cancer: a review. *Semin Cancer Biol*, **10,** 271-88.

Augustsson, K., Skog, K., Jagerstad, M., Dickman, P. & Steineck, G. (1999). Dietary heterocyclic amines and cancer of colon, rectum, bladder, and kidney: a population-based study. *Lancet*, **353,** 703-7.

Axelson, O. (1995). Cancer risks from exposure to radon in homes. *Environ Health Perspect.*, **103,** 37-43.

Baffoe-Bonnie, A.B., Beaty, T.H., Bailey-Wilson, J.E., Kiemeney, L.A., Sigvaldason, H., Olafsdottir, G., Tryggvadottir, L. & Tulinius, H. (2000). Genetic epidemiology of breast cancer: segregation analysis of 389 Icelandic pedigrees. *Genet Epidemiol*, **18,** 81-94.

Bailar, J.C., 3rd & Bailer, A.J. (1999). Risk assessment--the mother of all uncertainties. Disciplinary perspectives on uncertainty in risk assessment. *Ann N Y Acad Sci*, **895,** 273-85.

Bailar, J.C., 3rd & Bailer, A.J. (2001). Environment and health: 9. The science of risk assessment. *Cmaj*, **164,** 503-6.

Bailar, J.C., 3rd & Gornik, H.L. (1997). Cancer undefeated. *N Engl J Med*, **336,** 1569-74.

Begg, C.B. (2001). The search for cancer risk factors: when can we stop looking? *Am J Public Health*, **91,** 360-4.

Begg, C.B. (2002). On the use of familial aggregation in population-based case probands for calculating penetrance. *J Natl Cancer Inst*, **94,** 1221-6.

Berry, D.A., Iversen, E.S., Jr., Gudbjartsson, D.F., Hiller, E.H., Garber, J.E., Peshkin, B.N., Lerman, C., Watson, P., Lynch, H.T., Hilsenbeck, S.G., Rubinstein, W.S., Hughes, K.S. & Parmigiani, G. (2002). BRCAPRO validation,

sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *J Clin Oncol*, **20,** 2701-12.
Botto, L.D. & Khoury, M.J. (2001). Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. *Am J Epidemiol*, **153,** 1016-20.
Brennan, P. (2002). Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis*, **23,** 381-7.
Breslow, N. (2003). Are statistical contributions to medicine undervalued? *Biometrics*, **59,** 1-8.
Briggs, K.J. & Chudley, A.E. (2000). Genetic landmarks through philately--Barbara McClintock. *Clin Genet*, **58,** 280-3.
Burns, E.A. & Leventhal, E.A. (2000). Aging, immunity, and cancer. *Cancer Control*, **7,** 513-22.
Cannon-Albright, L.A., Thomas, A., Goldgar, D.E., Gholami, K., Rowe, K., Jacobsen, M., McWhorter, W.P. & Skolnick, M.H. (1994). Familiality of cancer in Utah. *Cancer Res*, **54,** 2378-85.
Carstensen, B., Soll-Johanning, H., Villadsen, E., Söndergaard, J. & Lynge, E. (1996). Familial aggregation of colorectal cancer in the general population. *Int J Cancer*, **68,** 428-35.
Chang-Claude, J., Becher, H., Blettner, M., Qiu, S., Yang, G. & Wahrendorf, J. (1997). Familial aggregation of oesophageal cancer in a high incidence area in China. *Int J Epidemiol*, **26,** 1159-65.
Chompret, A. (2002). The Li-Fraumeni syndrome. *Biochimie*, **84,** 75-82.
Clarke, M. & Oxman, A. (2001). Cochrane Reviewers' Handbook Glossary 4.1.4 [updated March 2001]. http://www.cochrane.dk/cochrane/handbook/handbook.htm (accessed 17 October 2003).
Claus, E.B., Risch, N.J. & Thompson, W.D. (1990). Age at onset as an indicator of familial risk of breast cancer. *Am J Epidemiol*, **131,** 961-72.
Crocetti, E., Lecker, S., Buiatti, E. & Storm, H.H. (1996). Problems related to the coding of multiple primary cancers. *Eur J Cancer*, **32A,** 1366-70.
Danoe, H., Andersen, O., Ewertz, M., Holm Petersen, J. & Lynge, E. (2003). Socioeconomic status and breast cancer in Denmark. *Int. J. Epidemol.*, **32,** 218-224.
DiCiommo, D., Gallie, B.L. & Bremner, R. (2000). Retinoblastoma: the disease, gene and protein provide critical leads to understand cancer. *Semin Cancer Biol*, **10,** 255-69.
dos Santos Silva, I. (1999). *Cancer epidemiology: principles and methods*. International Agency for Research on Cancer (IARC): Lyon, France.
Eklund, G. & Wiklund, K. (1979). Kvalitetskontroll av personidentiteten i cancer-miljöregistret. *Statistisk tidskrift 1979:5***,** 373-376, 407-408.
Engeland, A., Haldorsen, T., Tretli, S., Hakulinen, T., Horte, L.G., Luostarinen, T., Magnus, K., Schou, G., Sigvaldason, H., Storm, H.H. & et al. (1993). Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. A collaborative study of the five Nordic Cancer Registries. *APMIS Suppl*, **38,** 1-124.
EpC/SoS. (2002). Cancer Incidence in Sweden 2000. Centre for Epidemiology. National Board of Health and Welfare. Official statistics of Sweden. Health and diseases 2002:5. Published at internet from http://www.sos.se/fulltext/ (2002-06-04): Stockholm.

Estève, J., Benhamou, E. & Raymond, L. (1994). *Statistical Methods in Cancer Research, Volume IV, Descriptive Epidemiology*. Vol. 128. IARC Scientific Publication. IARC: Lyon.

Etzioni, R., Penson, D.F., Legler, J.M., di Tommaso, D., Boer, R., Gann, P.H. & Feuer, E.J. (2002). Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst*, **94,** 981-90.

Euhus, D. (2001). Understanding mathematical models for breast cancer risk assessment and counseling. *Breast J*, **Jul-Aug;7,** 224-32.

Faggiano, F., Partanen, T., Kogevinas, M. & Boffetta, P. (1997). Socioeconomic differences in cancer incidence and mortality. *IARC Sci Publ*, **138,** 65-176.

Fearon, E. & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, **61,** 759-67.

Fearon, E.R. (1997). Human cancer syndromes: clues to the origin and nature of cancer. *Science*, **278,** 1043-50.

Fleiss, J.L. (1986). *The design and analysis of clinical experiments*. Wiley series in probability and mathematical statistics. John Wiley & Sons.

Friedenreich, C.M. (2001). Physical activity and cancer prevention: from observational to intervention research. *Cancer Epidemiol Biomarkers Prev*, **10,** 287-301.

Frodin, J.E., Ericsson, J. & Barlow, L. (1997). Multiple primary malignant tumors in a national cancer registry-- reliability of reporting. *Acta Oncol*, **36,** 465-9.

Froehlich, G.W. (1999). What is the chance that this study is clinically significant? A proposal for Q values. *Eff Clin Pract*, **2,** 234-9.

Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C. & Mulvihill, J.J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*, **81,** 1879-86.

Gerber, M., Boutron-Ruault, M.C., Hercberg, S., Riboli, E., Scalbert, A. & Siess, M.H. (2002). [Food and cancer: state of the art about the protective effect of fruits and vegetables]. *Bull Cancer*, **89,** 293-312. resum.htm.

Giovannucci, E. (2001). An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*, **10,** 725-31.

Gnedenko, B. (1976). *The Theory of Probability*. Mir Publishers: Moscow.

Goldgar, D.E. (2002). Population aspects of cancer genetics. *Biochimie*, **84,** 19-25.

Goldgar, D.E., Easton, D.F., Cannon-Albright, L.A. & Skolnick, M.H. (1994). Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst*, **86,** 1600-7.

Gunter, C. (2001). The molecular genetics of cancer: down the rabbit hole. *Hum Mol Genet*, **10,** 655-6.

Hahn, W.C. & Weinberg, R.A. (2002). Modelling the molecular circuitry of cancer. *Nat Rev Cancer*, **2,** 331-41.

Hall, A.J., Yee, L.J. & Thomas, S.L. (2002). Life course epidemiology and infectious diseases. *Int J Epidemiol*, **31,** 300-1.

Hanahan, D. (2000). Benefits of bad telomeres. *Nature*, **406,** 573-4.

Hemminki, K., Dong, C. & Vaittinen, P. (2001). Cancer risks to spouses and offspring in the Family-Cancer Database. *Genet Epidemiol*, **20,** 247-57.

Hemminki, K., Nilsson, Å. & Vaittinen, P. (1999). Andragenerationsregister ny resurs för cancerforskningen. *Läkartidningen*, **96**.

Hemminki, K. & Vaittinen, P. (1997). Effect of paternal and maternal cancer on cancer in the offspring: population-based study. *Cancer Epidemiol Biomarkers Prev*, **6,** 993-7.
Hemminki, K. & Vaittinen, P. (1999). Familial cancers in a nationwide family-cancer database: age distribution and prevalence. *Eur J Cancer*, **35,** 1109-17.
Higginson, J. & Muir, C.S. (1977). [Determination of the importance of environmental factors in human cancer: the role of epidemiology]. *Bull Cancer*, **64,** 365-84.
Hippisley-Cox, J., Coupland, C., Pringle, M., Crown, N. & Hammersley, V. (2002). Married couples' risk of same disease: cross sectional study. *Bmj*, **325,** 636.
Hoover, R.N. (2000). Cancer--nature, nurture, or both. *N Engl J Med*, **343,** 135-6.
Hsu, L., Zhao, L.P., Malone, K.E. & Daling, J.R. (2000). Assessing changes in ages at onset over successive generation: an application to breast cancer. *Genet Epidemiol*, **18,** 17-32.
IARC. (1997). Cancer Incidence in Five Continents, Parkin, D., Whelan, S., Ferlay, J., Raymond, L. & Young, S. (eds), Vol. VII. IARC: Lyon.
Ilyas, M., Straub, J., Tomlinson, I.P. & Bodmer, W.F. (1999). Genetic pathways in colorectal and other cancers. *Eur J Cancer*, **35,** 335-51.
Iversen, E.S., Jr. & Chen, S. (2002). Population-calibrated gene characterization: cancer penetrance among BRCA1/2 mutation carriers. Institute of statistics and decision sciences (ISDS), Duke University: Durham, NC, USA.
Klein, A.P., Hruban, R.H., Brune, K.A., Petersen, G.M. & Goggins, M. (2001). Familial pancreatic cancer. *Cancer J*, **7,** 266-73.
Klein, G. (2000). Better understanding of the biology of cancer cells (article in Swedish). In *Läkartidningen*, Vol. 97. pp. 3260-4.
Kleinerman, R.A., Boice, J.D., Jr., Storm, H.H., Sparen, P., Andersen, A., Pukkala, E., Lynch, C.F., Hankey, B.F. & Flannery, J.T. (1995). Second primary cancer after treatment for cervical cancer. An international cancer registries study. *Cancer*, **76,** 442-52.
Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*, **68,** 820-3.
Lakhani, S. (2003). The molecular pathology of human familial cancer. In *5th Course in Cancer Genetics, European School of Genetic Medicine*. www.eurogene.org/etext/cancgen/lectures/Lakhani.htm.
Lancet. (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet*, **358,** 1389-99.
Lancet. (2002). MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*, **360,** 23-33.
Lander, E.S. & Schork, N.J. (1994). Genetic dissection of complex traits. *Science*, **265,** 2037-48.
Lawlor, D. (2002). Smoking: a genetic condition? *Int J Epidemiol*, **31,** 280-1.
Lesko, S.M., Rosenberg, L. & Shapiro, S. (1996). Family history and prostate cancer risk. *Am J Epidemiol*, **144,** 1041-7.
Li, F.P. (1995). Identification and management of inherited cancer susceptibility. *Environ Health Perspect*, **103 Suppl 8,** 297-300.
Liang, K.Y. & Beaty, T.H. (2000). Statistical designs for familial aggregation. *Stat Methods Med Res*, **9,** 543-62.

Lichtenstein, P., Holm, N., Verkasalo, P., Illiado, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. & Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer. *N Engl J Med*, **343,** 78-85.

Lindblom, A. & Nordenskjold, M. (2000). The biology of inherited cancer. *Semin Cancer Biol*, **10,** 251-4.

Lindor, N., Greene, M. & Program, t.M.F.C. (1998). A concise handbook of family cancer syndromes. *J Natl Cancer Inst*, **90,** 1039-71.

Lomborg, B. (2002). How healthy is the world? *Bmj*, **325,** 1461-6.

Lynch, H.T., Fusaro, R.M. & Lynch, J. (1995). Hereditary cancer in adults. *Cancer Detect Prev*, **19,** 219-33.

Lynch, H.T., Smyrk, T. & Lynch, J. (1997). An update of HNPCC (Lynch syndrome). *Cancer Genet Cytogenet*, **93,** 84-99.

Lyon, J.L., Gardner, J.W. & West, D.W. (1980). Cancer incidence in Mormons and non-Mormons in Utah during 1967--75. *J Natl Cancer Inst*, **65,** 1055-61.

Mattsson, B. (1984). Cancer registration in Sweden: studies on completeness and validity of incidence and mortality registers. Karolinska Institute: Stockholm.

Mattsson, B., Rutqvist, L.E. & Wallgren, A. (1985). Undernotification of diagnosed cancer cases to the Stockholm Cancer Registry. *Int J Epidemiol*, **14,** 64-9.

McKee, M. (2001). Epidemiology in the 21st century. The challenges ahead. *Eur J Public Health*, **11,** 241-2.

McLellan, T., Jorde, L.B. & Skolnick, M.H. (1984). Genetic distances between the Utah Mormons and related populations. *Am J Hum Genet*, **36,** 836-57.

Med-Dictionary. (2002). On-line Medical Dictionary. cancerweb.ncl.ac.uk.

Miettinen, O. (1985). *Theoretical Epidemiology*. John Wiley & Sons, Inc.

Milner, J.A. (2002). Strategies for cancer prevention: the role of diet. *Br J Nutr*, **87 Suppl 2,** 265-72.

Minamoto, T., Mai, M. & Ronai, Z. (1999). Environmental factors as regulators and effectors of multistep carcinogenesis. *Carcinogenesis*, **20,** 519-27.

Moller, B., Fekjaer, H., Hakulinen, T., Tryggvadottir, L., Storm, H., Talbäck, M. & Haldorsen, T. (2002). Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention*, **11 (suppl 1),** S1-S96.

Moolgavkar, S. (2002). 2001 Award Winners. In *RISK newsletter*, Vol. First quarter 2002. pp. 7.

Moolgavkar, S. & Knudson, A., Jr. (1981). Mutation and cancer: a model for human carcinogenesis. *J Natl Cancer Inst*, **66,** 1037-52.

Moolgavkar, S.H. (1994). Biological models of carcinogenesis and quantitative cancer risk assessment. *Risk Anal*, **14,** 879-82.

Mucci, L., Dickman, P., Steineck, G., Adami, H. & Augustsson, K. (2003). Dietary acrylamide and cancer of the large bowel, kidney, and bladder: absence of an association in a population-based study in Sweden. *Br J Cancer.*, **88,** 84-9.

Nathanson, K.L. & Weber, B.L. (2001). "Other" breast cancer susceptibility genes: searching for more holy grail. *Hum Mol Genet*, **10,** 715-20.

Naturvetaren. (2000). DNA-tekniken hittar rätt pappa. Interview of Bertil Lindblom from the Forensic Genetic department, Linköping, Sweden. In *Naturvetaren*, Vol. 14/2000. pp. 16-17.

Olsen, J.H., Boice, J.D., Seersholm, N., Bautz, A. & Fraumeni, J.J.F. (1995). Cancer in the parents of children with cancer. *New Engl. J. Med.*, **333,** 1594-9.

Osler, M., Holst, C., Prescott, E. & Sorensen, T.I. (2001). Influence of genes and family environment on adult smoking behavior assessed in an adoption study. *Genet Epidemiol*, **21,** 193-200.
Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. (2001). Estimating the world cancer burden: Globocan 2000. *Int J Cancer*, **94,** 153-6.
Parkin, R.T. & Balbus, J.M. (2000). Variations in concepts of "susceptibility" in risk assessment. *Risk Anal*, **20,** 603-11.
Pershagen, G., Akerblom, G., Axelson, O., Clavensjö, B., Damber, L., Desai, G., Enflo, A., Lagarde, F., Mellander, H. & Svartengren, M. (1994). Residential radon exposure and lung cancer in Sweden. *N Engl J Med.*, **330,** 159-64.
Peters, M.E. & Ostrander, E.A. (2001). Prostate cancer: simplicity to complexity. *Nat Genet*, **27,** 134-5.
Peto, J., Collins, N., Barfoot, R., Seal, S., Warren, W., Rahman, N., Easton, D., Evans, C., Deacon, J. & Stratton, M. (1999). Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst*, **91,** 943-9.
Pike, M., Krailo, M., Henderson, B., Casagrande, J. & Hoel, D. (1983). 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature*, **303,** 767-70.
Pisani, P., Parkin, D., Munoz, N. & Ferlay, J. (1997). Cancer and infection: estimates of the attributable fraction in 1990. *Cancer Epidemiol Biomarkers Prev*, **6,** 387-400.
Ponder, B.A. (2001). Cancer genetics. *Nature*, **411,** 336-41.
Potter, J.D. (2001). At the interfaces of epidemiology, genetics and genomics. *Nat Rev Genet*, **2,** 142-7.
Razum, O., Zeeb, H. & Rohrmann, S. (2000). The 'healthy migrant effect' - not merely a fallacy of inaccurate denominator figures. *Int J Epidemiol*, **29,** 191-2.
Risch, N.J. (2000). Searching for genetic determinants in the new millennium. *Nature*, **405,** 847-56.
Rockhill, B. (2001). The privatization of risk. *Am J Public Health*, **91,** 365-8.
Rosner, B. & Colditz, G.A. (1996). Nurses' health study: log-incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst*, **88,** 359-64.
Rostgaard, K., Vaeth, M., Holst, H., Madsen, M. & Lynge, E. (2001). Age-period-cohort modelling of breast cancer incidence in the Nordic countries. *Stat Med*, **20,** 47-61.
Ruivenkamp, C.A., van Wezel, T., Zanon, C., Stassen, A.P., Vlcek, C., Csikos, T., Klous, A.M., Tripodis, N., Perrakis, A., Boerrigter, L., Groot, P.C., Lindeman, J., Mooi, W.J., Meijjer, G.A., Scholten, G., Dauwerse, H., Paces, V., van Zandwijk, N., van Ommen, G.J. & Demant, P. (2002). Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet*, **31,** 295-300.
SAS. (2000). SAS OnlineDoc, version 8. SAS Institute.
SCB2001:5. (2001). Bakgrundsfakta till Befolknings- och Välfärdsstatistik 2001:5 Flergenerationsregistret - The Multi-Generation Register. Statistics Sweden.
SCB2001:6. (2001). Bakgrundsfakta till Befolknings- och Välfärdsstatistik 2001:6 Utredning om komplettering och utvidgning av Flergenerationsregistret. Statistics Sweden.
SCB. (2000). Earlier population and housing censuses 1960-1990. Statistics Sweden. Internet: http://www.scb.se/statinfo/2000/Be0205.asp (accessed April 18 2002).

SCB. (2001). Varför föds det så få barn? Statistics Sweden, available from http://www.scb.se/publkat/befolkning/fodsfabarn.asp (accessed 10 August 2001).
SCB. (2002). Barnfamiljerna. Statistics Sweden, available from http://www.scb.se/publkat/befolkning/.
SCB. (2003). Multi-Generation Register 2002. A description of contents and quality. Statistiska centralbyrån/Statistics Sweden.
Simard, J., Dumont, M., Labuda, D., Sinnett, D., Meloche, C., El-Alfy, M., Berger, L., Lees, E., Labrie, F. & Tavtigian, S. (2003). Prostate cancer susceptibility genes: lessons learned and challenges posed. *Endocr Relat Cancer*, **10,** 225-59.
Simard, J., Dumont, M., Soucy, P. & Labrie, F. (2002). Perspective: prostate cancer susceptibility genes. *Endocrinology*, **143,** 2029-40.
Smith, D. & Ebrahim, S. (2001). Epidemiology -- is it time to call it a day? (editorial). *Int J Epidemiol. 2001Jun*, **30,** 624-5.
Smith, G. & Ebrahim, S. (2002). Data dredging, bias, or confounding (editioral). *BMJ. 2002 Dec 21*, **325,** 1437-8.
Stenbeck, M., Rosen, M. & Holm, L.E. (1995). Cancer survival in Sweden during three decades, 1961-1991. *Acta Oncol*, **34,** 881-91.
Sterne, J.A. & Davey Smith, G. (2001). Sifting the evidence-what's wrong with significance tests? *Bmj*, **322,** 226-31.
Strachan, T. & Read, A. (1999). *Human Molecular Genetics*. Wiley.
Sugimura, T. (2000). Nutrition and dietary carcinogens. *Carcinogenesis*, **21,** 387-95.
Thomas, A., Cannon-Albright, L., Bansal, A. & Skolnick, M.H. (1999). Familial associations between cancer sites. *Comput Biomed Res*, **32,** 517-29.
Thomas, D.C. (2000). Genetic epidemiology with a capital "E". *Genet Epidemiol*, **19,** 289-300.
Tomatis, L. (2002). Primary prevention protects public health. *Ann N Y Acad Sci*, **982,** 190-7.
Tomlinson, I., Sasieni, P. & Bodmer, W. (2002). How many mutations in a cancer? *Am J Pathol*, **160,** 755-8.
Vaeth, M. (2003). How should we design and analyse register-based studies of disease incidence among relatives to patients with the disease? www.math.uio.no/avdc/workshops/norevet01-vaeth.html: Aarhus, Denmark.
Walach, N., Novikov, I., Milievskaya, I., Goldzand, G. & Modan, B. (1998). Cancer among spouses: review of 195 couples. *Cancer*, **82,** 180-5.
Weber, W., Ott, J., Gencik, A. & Muller, H. (1983). Familial cancer: genetically determined? (review). *Anticancer Res*, **3,** 133-42.
Weiss, K.M., Chakraborty, R., Smouse, P.E., Buchanan, A.V. & Strong, L.C. (1986). Familial aggregation of cancer in Laredo, Texas: a generally low-risk Mexican-American population. *Genet Epidemiol*, **3,** 121-43.
Verkasalo, P.K., Kaprio, J., Koskenvuo, M. & Pukkala, E. (1999). Genetic predisposition, environment and cancer incidence: a nationwide twin study in Finland, 1976-1995. *Int J Cancer*, **83,** 743-9.
Westergaard, T., Melbye, M., Pedersen, J.B., Frisch, M., Olsen, J.H. & Andersen, P.K. (1997). Birth order, sibship size and risk of Hodgkin's disease in children and young adults: a population-based study of 31 million person-years. *Int J Cancer*, **72,** 977-81.

Wiklund, F., Jonsson, B.A., Goransson, I., Bergh, A. & Gronberg, H. (2003). Linkage analysis of prostate cancer susceptibility: confirmation of linkage at 8p22-23. *Hum Genet*, **112,** 414-8.

Wilmoth, J.R. (1998). Age-Period-Cohort Models in Demography. *Démographie: Analyse et Synthèse*.

Vineis, P. & McMichael, A.J. (1998). Bias and confounding in molecular epidemiological studies: special considerations. *Carcinogenesis*, **19,** 2063-7.

Vogelstein, B., Fearon, E., Hamilton, S., Kern, S., Preisinger, A. & Leppert, M. (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med*, **319,** 525-32.

Wunsch-Filho, V., Boffetta, P., Colin, D. & Moncau, J.E. (2002). Familial cancer aggregation and the risk of lung cancer. *Sao Paulo Med J*, **120,** 38-44.

Yang, Q., Khoury, M.J. & Flanders, W.D. (1997). Sample size requirements in case-only designs to detect gene- environment interaction. *Am J Epidemiol*, **146,** 713-20.