Familial risks for cancer with reference to lung cancer

Xinjun Li

Stockholm 2004
To Qianren and Daniel
1. ABSTRACT

Familial aggregation of cancer may be due either to environmental factors shared by family members, or to shared genes. Familial clustering has been an avenue to the understanding of the etiology of cancer and has been a basis for clinical decisions and counseling, as well as guiding the identification of cancer-related genes.

The nation-wide Swedish Family-Cancer Database was created by linking the Statistics Sweden ‘Multi-Generation Register’ with the Swedish Cancer Registry, national census data and death notifications. All Swedes born after 1931 and their parents are included in the Database, which encompasses over 10 million individuals. Cancer cases were retrieved from the Swedish Cancer Registry from the years 1958, including over 1 million primary cancers and in situ tumors (paper I). The Family-Cancer Database was used as the source of family and cancer data, according to the premise that comparisons can be done based on different probands: parental probands are informative of dominant effects in offspring, whereas analysis of risks between siblings who lack affected parents provides clues about possible recessive effects.

With specific reference to lung cancer, in paper II, we used the Family-Cancer Database to examine the time trends of lung cancer in Sweden by histological type, with specific reference to gender, birth cohort, years of diagnosis (period), and age. The incidence rates of squamous cell carcinoma and other subtypes in men peaked in the period 1980-1990, and then decreased slightly thereafter. In contrast, the incidence rates of adenocarcinoma have continued to increase throughout the period. In women, all subtypes except squamous cell carcinoma have been increasing linearly. The incidence ratio of male to female for all cases of lung cancer was 2.8. For squamous cell carcinoma, the ratio was 12.4 in the beginning and 3.6 at the end. For adenocarcinoma, the ratio was stable at about 1.5 for the whole period.

In paper III, the standardized incidence ratio was 1.87 (95% CI 1.66-2.10) for all offspring (0–66 years old) when parents had lung cancer. The proportion of familial affect was 6.39%, and the population-attributable risk of all lung cancers was 2.97%. Lung cancer in offspring was associated with parental rectal, cervical, kidney, urinary
bladder and endocrine gland tumors. Age-specific, histology-specific familial risks for lung cancer were analyzed in paper IV. For offspring diagnosed before 50 years, the SIRs for histological types of lung cancer between offspring and parents were 1.98 for adenocarcinoma (offspring) and small cell/large cell carcinoma (parents), 2.66 for squamous cell carcinoma and small/large cell carcinoma (offspring/parents), 3.54 for small cell carcinoma and squamous cell carcinoma, and 3.70 for large cell carcinoma and adenocarcinoma. At a young age, risks between siblings were higher than those between offspring and parents. The SIR ratio (sibling risk/offspring risk) was 2.92 for all lung cancers. The early onset familial component accounted for 29.4% of familial adenocarcinoma and 33.3% of familial small and large cell carcinoma. The proportion was lowest for squamous cell carcinoma (13.3%).

In papers III and V, for multiple primary lung cancers we also found that risk of lung cancer was increased after upper aerodigestive tract, breast, cervical, kidney, urinary bladder or squamous cell skin cancer, or non-Hodgkin’s lymphoma, Hodgkin’s disease or leukemia, through all follow-up periods in men and women. Patients had an increased familial risk when their first-degree relatives were diagnosed with two lung cancers compared to those whose first-degree relatives were diagnosed with one lung cancer, and vice versa.

In summary, we suggested that there is an equal sensitivity of both genders to tobacco-induced lung cancer. Familial history of lung cancer was associated with increased risk of lung cancer; the population-attributable fraction of familial lung cancer was 2.97%. A large proportion of lung cancers before 50 years of age appear to be heritable and are probably due to a high-penetrant recessive gene or genes that predispose to tobacco carcinogens. Familial risks for multiple primary lung cancers have also suggested that there is an inherited susceptibility.

Key words: familial cancer, population-based, hereditary factors, familial risk, lung neoplasms.
2. MAIN REFERENCES

This thesis is based on the following papers, which will be referred to by their Roman numerals:


(IV) Li X, Hemminki K. Inherited predisposition to early onset lung cancer according to histological type. Int J Cancer 2004;112:451-57. ⁴

(V) Li X, Hemminki K. Familial multiple primary lung cancers: a population-based analysis from Sweden. Lung Cancer, in press. ³

The published articles are reprinted with permission of the respective publishers: ¹© Taylor & Francis Ltd., ²© Lippincott Williams & Wilkins, ³© 2003 Elsevier Ltd., and ⁴© John Wiley & Sons, Inc.
CONTENTS

1. ABSTRACT ..................................................................................................................... 1

2. MAIN REFERENCES ........................................................................................................ 3

3. INTRODUCTION ............................................................................................................ 5
   3.1 Hereditary cancer ..................................................................................................... 5
   3.2 Familial clustering of cancer .................................................................................... 7
   3.3 Genetic epidemiology in twins ............................................................................... 10
   3.4 Population-based studies on familial risks for cancer ........................................... 11
      3.4.1 Familial risks for cancer .................................................................................. 11
      3.4.2 Second cancers ............................................................................................... 13
      3.4.3 Sibling risks .................................................................................................... 14
      3.4.4 Shared environmental effects ......................................................................... 16
      3.4.5 Familial risks for cancer by site and histopathology, and gender effects ........ 18
      3.4.6 Migrant study .................................................................................................. 19
      3.4.7 Population-attributable fraction ..................................................................... 20
   3.5 Familial risks for lung cancer .................................................................................... 21
      3.5.1 Risk factors and incidence .............................................................................. 22
   3.6 Familial risks for lung cancer .................................................................................... 23

4. AIMS OF THE STUDY ................................................................................................... 25

5. METHODS ....................................................................................................................... 26
   5.1 Setting ...................................................................................................................... 26
   5.2 The Swedish Family-Cancer Database ................................................................ 26
   5.3 Patients and subjects ............................................................................................. 28
   5.4 Methods used in the study ..................................................................................... 29
      5.4.1 Incidence rates ................................................................................................. 29
      5.4.2 Standardized incidence ratios ......................................................................... 29
      5.4.3 Sibling risks .................................................................................................... 30
      5.4.4 The kappa test ................................................................................................ 31
      5.4.5 The population-attributable fraction .............................................................. 31
      5.4.6 Poisson regression .......................................................................................... 31
   5.5 Study design, subjects and methods of follow-up ................................................... 32
      5.5.1 Papers I and III ............................................................................................... 32
      5.5.2 Paper II ........................................................................................................... 33
      5.5.3 Papers IV and V ............................................................................................. 33
   5.6 Statistical analysis .................................................................................................... 34
   5.7 Statistical program .................................................................................................. 36

6. RESULTS AND DISCUSSION ....................................................................................... 37
   6.1 The Swedish Family-Cancer Database ................................................................ 37
   6.2 Lung cancer histology and incidence trends (Papers II and IV) .......................... 45
   6.3 Familial risks for lung cancer (Papers III, IV and V) ............................................. 47
   6.4 Sibling risk (Paper IV) ............................................................................................ 49
   6.5 Multiple primary lung cancers (Papers III and V) ................................................ 50

7. CONCLUSIONS ............................................................................................................. 52

8. ACKNOWLEDGEMENTS ........................................................................................... 54

9. REFERENCES ............................................................................................................... 56
3. INTRODUCTION

3.1 Hereditary cancer

Cancer has a genetic component in its etiology. It has been estimated that primary genetic factors play a major etiological role in approximately 5% to 10% of the total cancer burden. Of these, there are more than 20 cancer-associated syndromes for which the genetic basis is known and the genes have been identified. Collectively, the syndromes affect about 1 per cent of cancer patients. Although rare, the inherited cancer syndromes are of vast practical and biological importance, and knowledge of a genetic cancer syndrome allows for counselling. From a clinical standpoint, the identification of patients at high risk presents an opportunity to study the value of screening tests, to develop new treatment approaches, and to investigate new prevention and control measures such as treatment with inhibitory drugs and gene therapy. From a scientific standpoint, the identification of a syndrome presents an opportunity to isolate a specific gene and, thus, to define a specific genetic mechanism of cancer development.

Knowledge about hereditary cancer in humans has advanced at a rapid rate. To a great extent, this has been due to advances in molecular genetics with the discovery of an increasing variety of cancer-prone germline mutations that are etiologically linked to a vast number and variety of hereditary cancer syndromes. In a predominantly hereditary cancer syndrome, the cancer is expected to be present in multiple generations, because the presence of one copy of the mutation is sufficient to lead to pathology. Since the affected members carry one mutant and one normal allele of the disease gene, the risk of each offspring inheriting the mutation is thus 50%. Parents of patients with a recessive disorder are (most) often heterozygotes and therefore not affected themselves. Typically, recessive disorders often emerge in consanguineous mating, since this increases the probability that both parents will be carriers of the same mutation. In a recessive disorder, offspring of carriers of a recessive mutation have a 25% chance of developing the disease. In families with typical dominant disorders, half of the children are affected and the risk to children of unaffected family members is low. However, this general rule is somewhat modified by the fact that the
penetrance is reduced for many cancer-causing mutations. The cumulative incidence of disease (penetrance) may depend on age, sex and other factors. Sometimes these are even an obvious explanation for nonpenetrance, e.g., breast and ovarian cancer where male gene carriers remain unaffected, or in families with prostate cancer where females do not show any symptoms although they may transmit the disease to their offspring. However, in most cases the reason for a variable penetrance is unknown, and it may be determined by other genes or environmental agents.

**Table 3.1** Selected familial cancer syndromes

<table>
<thead>
<tr>
<th>Familial cancer syndrome</th>
<th>Cloned gene(s)</th>
<th>Type of tumor observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial retinoblastoma</td>
<td><strong>RB1</strong></td>
<td>Retinoblastoma, osteogenic sarcoma</td>
</tr>
<tr>
<td>Li-Fraumeni syndrome</td>
<td><strong>TP53</strong></td>
<td>Brain tumors, sarcomas, leukemia, breast cancer</td>
</tr>
<tr>
<td>Familial adenomatous polyposis</td>
<td><strong>APC</strong></td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>von Hippel-Lindau syndrome</td>
<td><strong>VHL</strong></td>
<td>Renal cancers, hemangioblastomas, pheochromocytoma</td>
</tr>
<tr>
<td>Wilms tumor</td>
<td><strong>WT1</strong></td>
<td>Pediatric kidney cancer</td>
</tr>
<tr>
<td>Familial melanoma</td>
<td><strong>CDKN2A</strong></td>
<td>Melanoma, pancreatic cancer, others</td>
</tr>
<tr>
<td>Hereditary non-polyposis colorectal cancer</td>
<td><strong>MSH2,</strong></td>
<td>Colorectal cancer, extracolonic cancer</td>
</tr>
<tr>
<td></td>
<td><strong>MLH1</strong></td>
<td></td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td><strong>BRCA1,</strong></td>
<td>Breast and ovarian cancer</td>
</tr>
<tr>
<td></td>
<td><strong>BRCA2</strong></td>
<td></td>
</tr>
<tr>
<td>Cowden syndrome</td>
<td><strong>PTEN</strong></td>
<td>Breast, thyroid and endometrial cancer</td>
</tr>
<tr>
<td>Multiple endocrine neoplasia type 1</td>
<td><strong>MEN1</strong></td>
<td>Parathyroid and pituitary adenomas, islet cell tumors, carcinoid</td>
</tr>
<tr>
<td>Neurofibromatosis type 1</td>
<td><strong>NF1</strong></td>
<td>Neurofibromas, sarcomas, gliomas</td>
</tr>
</tbody>
</table>

To date, many genes have been identified in which heritable constitutional mutations predispose to cancer development (Table 3.1) 3-6. The first cancer predisposition gene to be isolated was **RB1**, the gene for hereditary retinoblastoma, which was cloned a decade after cytogenetic studies of retinoblastoma had revealed deletions of chromosome 13 in normal and tumor tissues from retinoblastoma 1. The most prevalent hereditary syndromes associated with tumors are breast and ovarian cancer, hereditary non-polyposis colorectal cancer and familial adenomatous polyposis. However, each type of cancer in a family may be caused by mutations in one of several different genes. For example, for breast cancer a number of genes have been implicated. Two genes, **BRCA1** and **BRCA2**, have been isolated and in addition there is at least one or more yet unidentified breast cancer gene. Several other autosomal
dominant syndromes also show an increased risk of breast cancer, such as Li-Fraumeni syndrome and Cowden syndrome.

Additionally, it has become increasingly evident that the hereditary cancer syndromes often indicate an increased risk of cancer at many sites other than the ‘index’ sites, although with lower risk in the nonindex than in the index site. Germline mutations in the BRCA1 and BRCA2 genes can cause an inherited susceptibility to breast and ovarian cancer, with a penetrance of approximately 80% and 25–60% for breast and ovarian cancer, respectively, but they also confer smaller increased relative risks of certain other cancers, such as pancreatic, prostate and male breast cancer. Further examples are Li-fraumeni syndrome and hereditary non-polyposis colorectal cancer.

Autosomal dominant hereditary cancer syndromes account for a significant minority of common cancers. Population-based epidemiological studies have shown that only 15–20% of the observed familial clustering of breast cancer occurs in families that carry a strongly predisposing BRCA1 or BRCA2 mutation. In principle, the remaining 80–85% of familial risks may have a genetic or an environmental origin, but evidence from studies of breast cancer in twins indicates that susceptible women contribute to a high proportion, and perhaps even the majority of overall breast cancer incidence. This suggests the existence of less penetrant variants, or perhaps other genes that affect some of the same pathways. Similar results have been observed for colorectal cancer, where approximately 20–50% of familial colorectal cancers have been attributable to mutations in the tumor suppressor genes that are responsible for hereditary non-polyposis colorectal cancer, corresponding to 1 to 2.5% of all colorectal cancer among 0- to 61-year-old individuals. In addition, 1% or less of colorectal cancer was due to mutations in the APC gene which is responsible for familial adenomatous polyposis.

3.2 Familial clustering of cancer

The study of familial clustering of cancer has been fundamental to the understanding of heritable components in cancer, and the discovery of the genes. It has been established that there are familial risks between cancer sites which cannot be explained by shared environmental factors. Hereditary cancer syndromes have
been described for each of the cancers considered (Table 3.1). Familial risks, as defined as being between first-degree relatives, have ranged between 1.5 and 10 for the site-specific cancers that are sufficiently common for quantitative analysis 11-16.

Studies of site-specific familial aggregation of cancer have provided important leads in understanding the genetics underlying cancer development. Familial aggregation demonstrated both in rare cancers, such as bilateral retinoblastoma, and in more common cancers, such as breast and colon, has led to the identification of cancer genes (e.g. RB1, BRCA2, and APC) 3. Cancers of different sites also aggregate in families, such as in Li-Fraumeni syndrome and breast and ovarian cancer. Some family clusters of cancer have been well-defined, to the point where successful linkages have been made with particular genetic loci (e.g. p53, BRCA1) 3. While shared genetic defects can explain some family clusters, shared environments, high frequency/low penetrant susceptibility genes, and gene-environment interactions may also be operating in familial aggregation of cancer risk.

Familial clustering of cancer has been studied by four approaches: 1) by the clinical identification of probands; 2) by analytical epidemiological studies; 3) by twin studies; and 4) by carrying out population-based studies.

Traditionally, familial clustering of cancer has been studied in the clinical setting where probands and their multiple affected family members have been identified 2, 17. Ascertainment of cases in multi-generation families and the availability of biological specimens are usually optimal, or at least the best possible in the clinical settings 11. Clinical knowledge about the general features of inherited cancer syndromes include the following 4-6:

- A family history with several affected close relatives.
- An early age of onset compared to sporadic cases of the same disease.
- Multiple primary tumors.
- A high rate of cancer within a family.

This approach has also been productive in terms of understanding cancer genetics. Many forms of cancer in which a single gene poses a high risk have been identified. Of the 4,700 dominant and 2,800 recessive human genetic traits known in the early
1990s, some 440 were single-gene traits in which cancer was a complication; many of them were extremely rare, with only a few families identified worldwide. Clinical observation probably works for dominant diseases when these pose a risk of between 10 and 100, or more, above the population rate of the disease. For recessive conditions, clinical observation is less sensitive, and most of the results on recessive conditions have come from isolated populations with high rates of consanguineous marriage.

The accuracy and completeness of family history data, however, must be taken into account in using family history to assess individual risk in clinical practice, and in identifying families appropriate for cancer research. A reported family history may be erroneous, or a person may be unaware of relatives with cancer. In addition, decreasing family sizes, common polygenic cancers, and premature deaths are a major challenge to clinically based studies. Efficient pooling of clinical patient series may overcome the problems of sample size, but in the clinical definition of hereditary non-polyposis colorectal cancer syndrome, for example, use is made of the “Amsterdam criteria”, which include: 1) at least three relatives with verified colorectal cancer; 2) at least one must be a first degree relative to the other two; 3) familial adenomatous polyposis (FAP) is excluded; 4) at least two successive generations affected, and 5) one case of colon cancer at < 50 years of age. These criteria may have resulted in false-positive families, based on chance aggregation of tumors, as well as false-negative families, due to the low probability of finding 3 cases of colorectal cancer within small families. The smaller the family or the less pedigree information available, the less likely it is that the definition of hereditary non-polyposis colorectal cancer will be fulfilled. In addition, the Amsterdam criteria do not account for extracolorectal cancers, in particular, those of the endometrium. This could skew the results and hamper comparisons between different series. Although the identification of individuals and families at high risk is important for the design and implementation of efficacious primary and secondary disease prevention strategies, the cost effectiveness and the legal and ethical implications of genetic screening, diagnosis, and counselling in clinical practice have yet to be determined in population-based samples.
Another approach to studying familial cancer has been to analyze cancer risks of the relatives of the index case in analytical (traditional) epidemiological studies. Whereas the numbers of cases may be large in such studies, the certainty of information, usually based on recall, tends to be less certain.

### 3.3 Genetic epidemiology in twins

Studies of twins offer the third approach to the genetic epidemiology of cancer. These studies offer a particularly powerful approach to estimating the components of cancer risk that are attributable to genotype and environment. Monozygotic (identical) twins are derived from the fission of a single fertilized egg and thus inherit identical genetic material. By contrast, dizygotic (fraternal) twins are derived from two distinct fertilized eggs and thus have the same genetic relationship as full siblings, although they may be more “biologically” related because of sharing the same prenatal intrauterine experience. Thus, by comparing the similarity of monozygotic and dizygotic pairs, the evidence for genetic factors being involved may be provided. If there is twin similarity not accounted for by genetic effects, this indicates that shared environmental effects, e.g. shared childhood experiences such as diet, contribute to variance in the trait. Despite the large amount of genetic information that exists in twin populations, because of the rareness of twinning, such populations have only been used to a limited extent in cancer epidemiology.

In the Nordic countries, there are well-known twin and cancer registries. A cancer study was carried out by pooling data from the Swedish, Finnish and Danish twin registries, allowing a joint analysis. Data from 90,000 twins were combined to assess the cancer risks at 28 sites for co-twins of twins with cancer. Increases in risk for co-twins of affected twins were detected for several sites, including stomach, colorectal, lung, breast and prostate cancer. Structural equation modeling was used to determine the relative importance of heritable and environmental effects in cancers at 11 sites.

The results from model fitting are presented in Table 3.2. For stomach cancer, heritability was estimated to account for 28%, shared environmental effects for 10%, and non-shared environmental effects for the remaining 62% of the variation in liability. Statistically significant heritability estimates, where the 95% confidence
interval did not include zero, were detected for cancers of the colorectum (35%), breast (27%), and prostate (42%). The estimates for the shared environmental effects ranged from 0 to 20%, but none was statistically significant. There were no significant differences between the sexes at any of the sites. The structural equation modeling carried out can accommodate both dominant and recessive Mendelian modes and polygenic modes of inheritance. Thus, the results on heritability summarize the total genetic effects. All the above sites have also shown a familial effect in population-based studies from Utah and Sweden \(^{12, 15, 20}\). If the range of genetic effects for colorectal, breast and prostate cancer of 27–42% turns out to be true, then there must be major gaps in our understanding of the genetic basis of these diseases.

### Table 3.2 Heritable and environmental effects for cancers among Swedish, Danish, and Finnish twins \(^9\)

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Proportion of variance attributed to Heritable effects</th>
<th>Shared environment effects</th>
<th>Nonshared environment effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach</td>
<td>0.28</td>
<td>0.10</td>
<td>0.62*</td>
</tr>
<tr>
<td>Colorectum</td>
<td>0.35*</td>
<td>0.05</td>
<td>0.60*</td>
</tr>
<tr>
<td>Pancreas</td>
<td>0.36</td>
<td>0</td>
<td>0.64*</td>
</tr>
<tr>
<td>Lung</td>
<td>0.26</td>
<td>0.12</td>
<td>0.62*</td>
</tr>
<tr>
<td>Breast</td>
<td>0.27*</td>
<td>0.06</td>
<td>0.67*</td>
</tr>
<tr>
<td>Cervix uteri</td>
<td>0</td>
<td>0.20</td>
<td>0.80*</td>
</tr>
<tr>
<td>Corpus uteri</td>
<td>0</td>
<td>0.17</td>
<td>0.82*</td>
</tr>
<tr>
<td>Ovary</td>
<td>0.22</td>
<td>0</td>
<td>0.78*</td>
</tr>
<tr>
<td>Prostate</td>
<td>0.42*</td>
<td>0</td>
<td>0.58*</td>
</tr>
<tr>
<td>Bladder</td>
<td>0.31</td>
<td>0</td>
<td>0.69*</td>
</tr>
<tr>
<td>Leukemia</td>
<td>0.21</td>
<td>0.12</td>
<td>0.66*</td>
</tr>
</tbody>
</table>

*The 95%CI does not include 0.00, i.e. the estimate is significant.

### 3.4 Population-based studies on familial risks for cancer

#### 3.4.1 Familial risks for cancer

Population-based studies require that all cancers are registered and family relationships can be reconstructed. The power is in large numbers and unbiased risk estimates. These advantages in turn allow estimation of familiality at multiple sites.
Population-based studies have been carried out in only a few geographic areas, of which those on the Mormon population in Utah have been based on existing genealogy \textsuperscript{12}. Regarding cancer in Denmark and Iceland, cases have been obtained from the nationwide cancer registry and family relationships have been constructed from other national registers \textsuperscript{21}.

![Proportion of cancer susceptibility due to genetic factors](image)

**Cancer site**

**Figure 3.1** Estimates of proportion of cancer susceptibility due to genetic factors \textsuperscript{22}

Population-based studies have provided reliable quantitative estimates of familial risks, particularly when the family relationships and cancers in relatives have been confirmed \textsuperscript{12, 14, 15}. Using the Swedish Family-Cancer Database, Czene et al. applied structural equation modeling to derive estimates of the importance of genetic and environmental effects for 15 common cancers \textsuperscript{22}. In their study, the structural equation of liability (D) to cancer for one individual was included in genetic effects (G), shared environmental effects (S), childhood shared environmental effects (F) and nonshared environmental effects (E). Thus, the equation was written as: D=G+S+F+E. The correlations between any pair of relatives for environmental and genetic effects were...
set to fixed values according to their degree of genetic and environmental relationship. In the results, inherited genetic factors accounted for 1–53% of causation of cancer (Figure 3.1) \(^22\). The shared environmental effects ranged from 0% (cervix) to 15% (stomach). The shared environmental effects in childhood were most important in testicular cancer, stomach cancer and in situ cancer of the cervix. However, in comparison to twin studies (Table 3.2) \(^9\), the heritability estimates were lower in this population-based study. The reason for the difference of estimates is the inability of population-based studies to fully cover low penetrant and polygenic effects. Low penetrant gene effects will make familial patterns difficult to observe, but they affect monozygotic twins.

### 3.4.2 Second cancers

Second primary malignancies are defined as malignant neoplasms that occur apart from the first primary cancer. The occurrence of second primary cancers in particular individuals has intrigued clinicians and scientists for more than a century. Due to improved diagnostic procedures and/or treatment, the time from diagnosis of the first cancer until the patient’s death has increased. The number of people experiencing multiple cancer diagnoses will probably increase. This makes it more important to evaluate the risk of different secondary cancers – in order to be able to prevent them, or at least to detect them as early as possible. An increased occurrence of second primary cancers can result from 1) intensive medical surveillance after first diagnosis; 2) therapy-induced exposure to chemical or physical carcinogens; or 3) shared environmental, hereditary and immunological factors between the first and second cancers \(^{23-25}\).

Multiple primary cancers have provided much important information about environmental carcinogens and genetic problems. Thus, a second cancer offers an interesting opportunity for the study of risk factors of cancer, including heritable factors. Many known cancer syndromes, such as BRCA-related breast cancers and mismatch repair gene-related non-polyposis colorectal cancers and multiple endocrine neoplasia are characterized by an increased risk of multiple primary cancers \(^{24-26}\).

Population-based systematic analysis of second primary cancer has been studied using the nationwide cancer registries in USA, Denmark, Finland and Sweden \(^{27-30}\). The
Swedish Cancer Registry has clear instructions about the reporting of multiple primary malignancies, and a re-evaluation of 209 multiple primary tumors found 98% of second primary cancers to be correctly classified. It is thus unlikely that intensive medical surveillance is causing diagnostic misclassification. Table 3.3 lists the results of the risks for concordant second primary cancers in Sweden. Remarkably high risks for second cancers were noted for nose, squamous cell skin, connective tissue and leukemia. These data provide evidence that the shared exposure to environmental or inherited risk factors leads to multiple cancers at a specific site. Familial subsequent primary cancers of colon, breast and melanoma have been analyzed recently, and the data suggest that patients with a second cancer include a subgroup with a strong genetic predisposition to cancer – which cannot often be predicted by a family history. Such risks would be typical of a polygenic model of carcinogenesis.

Table 3.3 Risk of subsequent concordant primary cancer (Sweden) 

<table>
<thead>
<tr>
<th>Initial cancer site</th>
<th>Follow-up interval (years)</th>
<th>0-9</th>
<th>10-38</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>SIR</td>
<td>95%CI</td>
</tr>
<tr>
<td>Oral, etc.</td>
<td>190</td>
<td>8.2</td>
<td>7.1</td>
</tr>
<tr>
<td>Colon</td>
<td>549</td>
<td>3.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Nose</td>
<td>5</td>
<td>31.8</td>
<td>10.0</td>
</tr>
<tr>
<td>Breast</td>
<td>3598</td>
<td>3.8</td>
<td>3.7</td>
</tr>
<tr>
<td>Female genitals</td>
<td>8</td>
<td>6.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Testis</td>
<td>13</td>
<td>5.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Kidney</td>
<td>43</td>
<td>1.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>270</td>
<td>2.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Melanoma</td>
<td>267</td>
<td>8.7</td>
<td>7.7</td>
</tr>
<tr>
<td>Skin, squamous cell</td>
<td>1029</td>
<td>15.9</td>
<td>14.9</td>
</tr>
<tr>
<td>Connective tissue</td>
<td>25</td>
<td>20.3</td>
<td>13.1</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>111</td>
<td>3.8</td>
<td>3.1</td>
</tr>
<tr>
<td>Leukemia</td>
<td>128</td>
<td>6.9</td>
<td>5.8</td>
</tr>
</tbody>
</table>

O=observed; SIR=standardized incidence ratio; CI=confidence interval.

3.4.3 Sibling risks

More than 450 single-gene traits associated with specific neoplasms have been identified in clinical and epidemiological studies. Most of these family syndromes are autosomal dominant disorders, such as neurofibromatosis 1 and 2, von Hippel-
Lindau disease, familial breast and ovarian cancer caused by mutations in \textit{BRCA1} and \textit{BRCA2}, Li-Fraumeni syndrome, and retinoblastoma. However, for recessive conditions, clinical observation is less sensitive, and most of the results on recessive conditions have come from isolated populations with high rates of consanguineous marriage. Formal epidemiological studies have had little importance in defining new familial traits. If the higher sibling risks are genetic, they could be due either to autosomal recessive or X-linked recessive genes. An X-linked susceptibility allele would be predicted to cause an increased risk only in male siblings of male cases, whereas autosomal recessive alleles would affect both sexes. Risks to siblings are of interest since high risks, in comparison with risks to parents or offspring, may indicate a recessive or X-linked component to the disease. Recently, an X-linked locus for testicular cancer was mapped \textsuperscript{34}.

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Parent only</th>
<th>Sibling only</th>
<th>SIR ratio (sibling/parent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>SIR</td>
<td>O</td>
</tr>
<tr>
<td>Colorectum \textsuperscript{(1)}</td>
<td>269</td>
<td>2.23</td>
<td>114</td>
</tr>
<tr>
<td>Lung</td>
<td>138</td>
<td>1.69</td>
<td>71</td>
</tr>
<tr>
<td>Breast</td>
<td>963</td>
<td>1.61</td>
<td>728</td>
</tr>
<tr>
<td>Ovary</td>
<td>78</td>
<td>3.48</td>
<td>35</td>
</tr>
<tr>
<td>Prostate</td>
<td>215</td>
<td>2.55</td>
<td>170</td>
</tr>
<tr>
<td>Testis</td>
<td>10</td>
<td>4.33</td>
<td>24</td>
</tr>
<tr>
<td>Kidney</td>
<td>34</td>
<td>1.92</td>
<td>22</td>
</tr>
<tr>
<td>Melanoma</td>
<td>105</td>
<td>2.33</td>
<td>86</td>
</tr>
<tr>
<td>Nervous system</td>
<td>81</td>
<td>1.86</td>
<td>44</td>
</tr>
<tr>
<td>Endocrine glands</td>
<td>22</td>
<td>2.18</td>
<td>17</td>
</tr>
<tr>
<td>Non-Hodgkin's lymphoma</td>
<td>29</td>
<td>1.68</td>
<td>23</td>
</tr>
<tr>
<td>Leukemia</td>
<td>24</td>
<td>1.64</td>
<td>17</td>
</tr>
</tbody>
</table>

Bold type: 95\%CI does not include 1.00. O=observed; SIR=standardized incidence ratio; CI=confidence interval.

(1): adenocarcinoma only.

Despite the fact that at least six recessive hereditary cancer syndromes have been identified \textsuperscript{3,4}, there have been no population-based studies on recessive cancers from other groups, which probably explains the current lack of knowledge in this area. The detection of recessive conditions is difficult because the cases appear apparently randomly in pedigrees, but often reveal consanguinity on closer inspection. Population geneticists have raised questions about the relatively small number of known human recessive syndromes. In different species of experimental animals,
recessive traits predominate – in contrast to humans where dominant traits are more common. It is not impossible that this is an observation bias because of difficulties in identifying a recessive pattern in humans. The recessive mode of inheritance can be tested in segregation analysis, but such analyses have not been carried out on unselected population-based datasets.

We have done systematic comparison of cancer risks between parents and offspring (offspring risk) for putative dominant effects, and between siblings (sibling risk) for putative recessive or X-linked effects, based on the Family-Cancer Database. The results are summarized in Table 3.4. Among the 13 cancer sites, all offspring and sibling risks were significantly elevated. Sibling risks were high at many sites when parents were affected, suggesting strong recessive effects for, e.g., prostate, testicular and kidney cancer, and leukemia. For prostate cancer, linkage studies have suggested X-linked loci for susceptibility genes and segregation analysis has provided evidence either for X-linked or recessive inheritance. For kidney cancer, the VHL gene is a main susceptibility gene and it may be responsible for some of the dominant effects observed, but the results suggest the presence of autosomal recessive susceptibility genes for renal cancer. At some sites, a minor recessive component may exist, including colorectal, lung, cervical and endocrine gland tumors and non-Hodgkin’s lymphoma.

3.4.4 Shared environmental effects
Family members share many environmental factors such as lifestyle – including diet and habits – which may increase or decrease exposure to cancer-causing or protective factors. From twin studies (Table 3.2), shared environmental effects were estimated to be 0–20%, but none of these values were statistically significant. From the Swedish Family-Cancer Database, the model revealed the significant contribution of the environment shared among family members to many cancer types. The highest proportion of cases with shared environment was found for gastric and lung cancer between spouses. Studies on genetically unrelated spouses appear particularly rewarding because the couples have shared decades of life experience until diagnosis of cancer. Experience from the Swedish Family-Cancer Database suggests that lifestyle-related factors explain (at least in part) the familial aggregation observed in lung and cervix cancers, for example. However, spouse concordance will not
detect sharing of environment early in life. Migrant studies from Sweden show that
the cancer patterns of immigrants are largely set during the two first decades of life \(^{48, 49}\); thus, spouse correlation would only detect strong effects that spouses share. This is
completely consistent with the assumption that certain dietary habits increase the
chances of developing stomach and intestinal cancer, and that smoking increases the
chances of developing cancer of the lung and bladder \(^{40, 41}\).

**Table 3.5** Risk of cancer in siblings, by age difference \(^{35}\)

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Siblings ages&lt;5 years</th>
<th>SIR ratio</th>
<th>95%CI</th>
<th>Siblings ages&lt;5 years</th>
<th>SIR ratio</th>
<th>95%CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>SIR (A)</td>
<td>95%CI</td>
<td>O</td>
<td>SIR (B)</td>
<td>95%CI</td>
<td></td>
</tr>
<tr>
<td>Colorectum (1)</td>
<td>64</td>
<td>2.55</td>
<td>1.97</td>
<td>3.26</td>
<td>41</td>
<td>2.38</td>
<td>1.71</td>
</tr>
<tr>
<td>Lung</td>
<td>45</td>
<td>2.19</td>
<td>1.60</td>
<td>2.94</td>
<td>26</td>
<td>2.04</td>
<td>1.33</td>
</tr>
<tr>
<td>Breast</td>
<td>349</td>
<td>1.58</td>
<td>1.42</td>
<td>1.75</td>
<td>367</td>
<td>1.85</td>
<td>1.66</td>
</tr>
<tr>
<td>Cervix</td>
<td>3</td>
<td>1.31</td>
<td>0.25</td>
<td>3.87</td>
<td>7</td>
<td>2.53</td>
<td>1.00</td>
</tr>
<tr>
<td>Ovary</td>
<td>19</td>
<td>3.78</td>
<td>2.27</td>
<td>5.91</td>
<td>16</td>
<td>3.51</td>
<td>2.00</td>
</tr>
<tr>
<td>Prostate</td>
<td>100</td>
<td>3.43</td>
<td>2.79</td>
<td>4.17</td>
<td>64</td>
<td>3.83</td>
<td>2.95</td>
</tr>
<tr>
<td>Testis</td>
<td>13</td>
<td>11.53</td>
<td>6.11</td>
<td>19.77</td>
<td>11</td>
<td>6.78</td>
<td>3.37</td>
</tr>
<tr>
<td>Kidney</td>
<td>14</td>
<td>4.46</td>
<td>2.43</td>
<td>7.50</td>
<td>8</td>
<td>3.75</td>
<td>1.60</td>
</tr>
<tr>
<td>Melanoma</td>
<td>52</td>
<td>3.99</td>
<td>2.31</td>
<td>4.06</td>
<td>34</td>
<td>2.15</td>
<td>1.49</td>
</tr>
<tr>
<td>Nervous system</td>
<td>17</td>
<td>1.43</td>
<td>0.83</td>
<td>2.29</td>
<td>27</td>
<td>2.43</td>
<td>1.60</td>
</tr>
<tr>
<td>Thyroid gland, nonmedullary</td>
<td>1</td>
<td>1.22</td>
<td>0.00</td>
<td>7.01</td>
<td>5</td>
<td>6.61</td>
<td>2.08</td>
</tr>
<tr>
<td>Endocrine glands</td>
<td>13</td>
<td>4.20</td>
<td>2.23</td>
<td>7.20</td>
<td>4</td>
<td>1.39</td>
<td>0.36</td>
</tr>
<tr>
<td>Non-Hodgkin's lymphoma</td>
<td>6</td>
<td>1.05</td>
<td>0.38</td>
<td>2.30</td>
<td>17</td>
<td>3.30</td>
<td>1.92</td>
</tr>
<tr>
<td>Leukemia</td>
<td>14</td>
<td>4.06</td>
<td>2.20</td>
<td>6.81</td>
<td>3</td>
<td>1.03</td>
<td>0.19</td>
</tr>
<tr>
<td>All</td>
<td>710</td>
<td>2.03</td>
<td>1.89</td>
<td>2.19</td>
<td>630</td>
<td>2.13</td>
<td>1.97</td>
</tr>
</tbody>
</table>

Bold type: 95%CI does not include 1.00. No twins included. O = observed; SIR = standardized incidence ratio; CI = confidence interval.
P-value: the probability of chi-squared distribution based on the test statistic \(X^2\) with 1 degree of freedom.
(1): adenocarcinoma only.

Shared environment could also be assessed in effects of shared childhood. Czene et al.
found significance in shared childhood environment for cancer of the testis, stomach,
cervix in situ, endocrine glands, and also melanoma \(^{22}\). Similar effects of shared
environment in childhood were also assessed by comparing cancer risks in siblings
from the Family-Cancer Database who were born close or far apart \(^{35, 50}\). Risks for
siblings were analyzed as a function of their age difference, assuming that
environmental factors are stronger among siblings of similar age (Table 3.5) \(^{35}\). For
example, data on the effect of age differences suggest that in testicular cancer,
melanoma, endocrine tumors and leukemia, shared childhood environment may
influence familial risk. Exposure to sun is a plausible common risk factor for
melanoma \(^{51}\), and childhood infections a possible risk for leukemia \(^{52, 53}\), but no
obvious environmental factors are known for testicular and endocrine tumors.
3.4.5 Familial risks for cancer by site and histopathology, and gender effects

Increased familial risks for cancer have been reported for a number of cancers according to sites. But histology-specific risks have not been examined, except that tobacco smoking causes multiple histological types of lung cancer, human papilloma virus causes both squamous cell carcinoma and adenocarcinoma of the cervix, and solar ultraviolet light predisposes individuals to all forms of skin neoplasia, squamous cell carcinoma, basal cell carcinoma and melanoma\(^{54-56}\). Among heritable cancers, a syndrome may encompass a wide variety of tumors \(^{57}\). Li-Fraumeni syndrome features the presentation of sarcomas, pheochromocytomas, brain tumors and breast cancers, and the histology of brain tumors includes many types – of which astrocytomas and medulloblastomas are the most common \(^{58}\). In syndromes affecting the intestines, hereditary non-polyposis colorectal cancer causes adenocarcinomas and multiple endocrine neoplasia I causes carcinoid tumors \(^{59-61}\). Using the Swedish Family-Cancer Database, the histology-specific analysis of familial risks was 2.07 for specific histologies and 2.00 for any histology \(^{62}\). Although this difference was small, concordant histology consistently showed a higher risk than did any histology. For some histopathologies, the familial risk was especially high. Familial risks for serous papillary cystadenocarcinoma of the ovary, papillary thyroid cancer, and low-grade astrocytoma exceeded 4.0. For signet-ring cell gastric cancer and various forms of ovarian cancer and skin squamous cell carcinoma, the familial risk was over 3.0. Histopathology-specific familial risks were also notably high for hepatocellular carcinoma (2.48), large cell carcinoma (2.29) and adenocarcinoma of the lung (2.18), and clear cell carcinoma of the kidney (2.73). Familial risks for cancer by histological types were also analyzed at other sites, for example, breast, testis, skin and melanoma etc. \(^{63-68}\). Many of the findings were novel and could be revealed only by applying codes for specific histopathology.

The question of gender differences in familial cancer may be informative of models of inheritance – in so far as familial effects are due to heritable causes of cancer. Both X- and Y-chromosome linked heritable effects would show sex preference in transmission of the malignant phenotype \(^{36,69}\). However, practically nothing is known about X- or Y-linked diseases where cancer is a manifestation, although linkage studies on prostate and testicular cancer have mapped loci on the X-chromosome \(^{38,70}\). Another aspect of sex-specific familial risks would be the mode of interaction with the
background rate, i.e. whether the sex of high or low incidence responds differently to a familial effect. The nationwide Swedish Family-Cancer Database was used to analyze familial risks for male and female offspring in concordant paternal and maternal cancer. Sex ratios for familial cancer were derived for cancer at 15 sites shared by men and women. At 14 sites, the sex ratio (male/female) for familial relative risk ranged between 0.78 and 1.41, with no evidence of sex preference, suggesting that sex chromosomes do not contribute to a noticeable extent to familial risks of cancers that occur in both sexes.

3.4.6 Migrant study
All common cancers are complex diseases of environmental and heritable etiology. Migrant studies are interventions where individuals or families move or are forced to move from one environment to another. Their great value has been in guiding the etiological search for causes of cancer to environmental factors. The classical migrant studies on Japanese immigrants to the USA and on European immigrants to Australia have been strong arguments for the predominant environmental etiology of cancer. In Sweden, close to 1 million people, 10% of the population, is foreign born, having immigrated from a number of countries at various periods of time, reflecting epochs in world history. Using the Swedish Family-Cancer Database, the probability of developing cancer was analyzed among first-generation and second-generation immigrants in Sweden to demonstrate the influence of the nurturing environment. For the first-generation immigrants who had come to Sweden in their 20s, many of the cancer rates were similar to those found in the original homeland. This suggests that the chances of developing cancer are in most cases "fixed" during a person’s early years, both through the direct effect of the environment and by virtue of the fact that lifestyle patterns are established during this phase of life. For immigrants from the Nordic region, the rest of Europe and the USA, the total frequency of cancer was similar to that in Sweden, while immigrants from other parts of the world – particularly Asia – showed a relatively low risk of developing the disease.

The study of second-generation immigrants showed that the differences that exist between first-generation immigrants, owing to differences in their childhood and adolescent environment and lifestyles, largely disappear in the Sweden-born second generation. In most cases, they have adopted the Swedish “cancer pattern”.

19
further strengthens the indications that the crucial patterns in the risk of later developing cancer are established during childhood and adolescence. The great challenge and opportunity presented by studies such as those involving immigrants is to explain how environmental factors and lifestyle patterns can determine the development of different forms of cancer.

Migrant studies could also answer the question of population differences in genetic susceptibility to cancer, because of the differential distribution of genotypes in populations \(^{79,80}\). For many cancer sites, however, there are still marked variations in incidence within countries, even when their populations are genetically fairly homogeneous \(^{81,82}\).

### 3.4.7 Population-attributable fraction

Population-based studies can also be used to estimate the population-attributable fraction (PAF) of familial risk, i.e. the proportion of a particular cancer that is related to familial clustering and that could be gained if familial cancer could be prevented. Using the Swedish Family-Cancer Database, Hemminki et al. calculated familial PAFs for 28 neoplasms among 0 to 66-year-old offspring. The results showed higher PAFs of 20.6% for prostate cancer, and 10.6% for breast cancer (Table 3.6) \(^{83}\). The PAFs corresponding to the heritability estimates derived from twin studies (Table 3.2) for prostate and breast cancer were 42 and 27%, respectively \(^9\), and were thus 2–5 times higher than the present estimates. One reason for the difference between the PAF and the twin estimates is the inability of population-based studies to fully cover low penetrant and polygenetic effects. Low penetrant gene effects will make familial patterns difficult to observe, but they affect monozygotic twins. In practice, familial cancers cannot be completely prevented, but PAF shows the weighting of familial causes compared to other causes of cancer \(^{84}\).

On an individual level, PAFs were also calculated to measure the proportion of cancer that can be ascribed to socioeconomic differences and educational factors and \(^{85,86}\), which might involve science and health policy issues. One practical question is the relative contribution of these factors to the overall burden of cancer. The population-attributable risk approach takes into account the magnitude of the relative risk that is associated with an exposure, along with the likelihood of exposure in the general
Because of the interactions between exposures, the combined population-attributable risks for cancers that have a well-characterized set of important risk factors can exceed 100%. For example, population-attributable risk estimates for lung cancer indicate that active smoking is responsible for 90% of lung cancer cases. Occupational exposure to carcinogens accounts for approximately 13 to 47% of lung cancer cases, while radon causes 10% of lung cancers, and radon, outdoor air pollution, asbestos and other factors are also contributory risks for lung cancer. Table 3.6 also lists the PAFs due to socioeconomic status and educational level for 9 sites of cancers from the Swedish Family-Cancer Database. The differences in cancer occurrence between socioeconomic groups showed that PAF for lung cancer was over 50% for men and women, and for men other tobacco-related sites, the kidney and bladder, showed PAFs between 18.0% and 32.3%. For educational factors, the PAF was 13.8% for men and 16.7% for women, and it was highest, over 50%, for stomach cancer in both genders and for cervical cancer in women. A similar result was also found in a cohort study on the lifestyle and health of Norwegian-Swedish women.

### Table 3.6 Population-attributable fraction (%) due to family history, socioeconomic status and educational level

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Family history Proportion Total PAF (1)</th>
<th>Socioeconomic status (2) Men</th>
<th>Socioeconomic status (2) Women</th>
<th>Educational level (3) Men</th>
<th>Educational level (3) Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach</td>
<td>4.49</td>
<td>1.5</td>
<td>29.5</td>
<td>15.1</td>
<td>51.2</td>
</tr>
<tr>
<td>Liver</td>
<td>2.25</td>
<td>0.8</td>
<td>33.0</td>
<td>26.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Lung</td>
<td>7.61</td>
<td>3.8</td>
<td>52.6</td>
<td>50.2</td>
<td>37.7</td>
</tr>
<tr>
<td>Breast</td>
<td>14.56</td>
<td>10.6</td>
<td>40.1</td>
<td>16.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Cervix</td>
<td>2.97</td>
<td>2.4</td>
<td>29.4</td>
<td>29.4</td>
<td>52.8</td>
</tr>
<tr>
<td>Prostate</td>
<td>18.58</td>
<td>20.6</td>
<td>4.4</td>
<td>4.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Kidney</td>
<td>3.46</td>
<td>1.9</td>
<td>18.0</td>
<td>9.9</td>
<td>17.9</td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>4.59</td>
<td>2.0</td>
<td>32.3</td>
<td>25.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Melanoma</td>
<td>4.27</td>
<td>2.7</td>
<td>32.2</td>
<td>17.2</td>
<td>15.4</td>
</tr>
</tbody>
</table>

(1): the PAF of cases with a family history of cancer was estimated as follows: proportion of cases with a family history x (familial SIR-1)/familial SIR. (2), (3): the PAFs were calculated based on age-adjusted incidence rates using the formula: \( \frac{(I - Io)}{It} \), where \( It \) is the age-adjusted incidence for a defined group and \( Io \) is the incidence for the group with lowest significant standardized incidence ratio.

### 3.5 Familial risks for lung cancer

At the beginning of the twentieth century, lung cancer was a very rare disease, but rates in most European countries and North America have increased so dramatically...
that lung cancer can be considered to have been a major epidemic of the former century 94, 95. Nowadays, lung cancer is the most common fatal cancer in the world 96, 97, and in Sweden it ranks fourth in incidence after prostate, colorectal and squamous cell skin cancer in men, and after breast, colorectal and endometrial cancer in women 98.

3.5.1 Risk factors and incidence

Many risk factors have been identified in lung cancer 83, 87-92, 99-104. More than 80% of all cases of lung cancer are related to cigarette smoking, and there is a relatively strong dose-response relationship between cigarette smoking and the development of lung cancer 99, 105. Environmental tobacco smoke contains many carcinogens, and has been associated with an increased risk of lung cancer in nonsmokers 99, 102-104. Other important risk factors for lung cancer among nonsmokers and smokers are occupational exposure to asbestos, arsenic compounds, coal tar and other agents 83, 87-92, 99, 100. Air pollution, residential radon and personal and family history have been suspected to be causes of lung cancer 83, 90, 91, 99, 101. Diet, in particular low vegetable and fruit consumption, have also been considered to be a cause of lung cancer 99.

The incidence of lung cancer is rising in women and declining in men as a result of changed smoking behavior 94, 105, 106. The observation that the risk of lung cancer is greater in women than in men exposed to equivalent amounts of tobacco smoke 107-109 is not supported by recent studies from Europe and USA, which have concluded that the risk is similar in both sexes 110-112. Patterns of histological types of lung cancer have also changed over time. Increasing rates of adenocarcinoma and decreasing incidence of squamous cell carcinoma have previously been reported in men in USA 113, whereas all lung cancers of all histological types have been increasing in women. A similar situation has been observed in Europe 105, 114. The changing histological patterns in lung cancer probably relate to changes in the composition of cigarettes 105, 114-117. With the introduction of the filter tip and low-tar cigarettes, is has been suggested that the reduction in risk may be negligible if smokers smoke more cigarettes, take larger puffs or inhale more deeply in order to satisfy their craving for nicotine. Filter cigarettes remove larger particles in cigarette smoke, thus reducing the deposition of these particles in the central airways where squamous cell carcinoma
occurs, while exposing the peripheral part of the lung (where adenocarcinoma occurs) to a proportionately higher amount of smoke carcinogens 94, 106, 114, 117.

### 3.5.2 Familial risks for lung cancer

As only 10% of cigarette smokers develop lung cancer 118, it has been suggested that genetic factors may contribute. Segregation analysis of lung cancer also supports the existence of heritable effects 119-121. The recent twin studies and family-based studies have suggested that familial risk of lung cancer may be due to genetic factors and shared environmental factors, implying aggregation of smokers in families 9, 22, 84.

Lung cancer has been one of the main cancer types used for studies on the effects of genetic polymorphisms. The underlying assumption in such studies is that lung cancer has a hereditary etiology 122, 123. Studies in molecular biology have begun to elucidate the role of genetic factors in modifying an individual’s risk for lung cancer 124, 125. There is increasing evidence that genes coding for carcinogen metabolism and DNA repair contribute to familial aggregation of lung cancer 126-129. The first reports on genetic influences in lung cancer susceptibility appeared in the early 1980s and showed that first-degree nonsmoking relatives of lung cancer patients had an increased risk of lung cancer compared to nonsmokers with no affected relatives 130, 131. Models for lung tumorigenesis have also been developed in experimental animals 132, 133. However, the role of hereditary factors in tumor development has been less well understood in lung cancer than in many other human neoplastic diseases. No distinct familial forms of the common types of lung cancer have been defined. No consistent effect on histology has emerged. Many tumor suppressor genes have been identified as being responsible for hereditary cancer syndromes, but none of them have been singled out specifically for lung cancer. On the other hand, lung cancer may occur as a manifestation in cancer syndromes that predominantly predispose to other cancers, such Li-Fraumeni, Bloom and xeroderma pigmentosum syndromes 4, 118.

Regarding men and women who do not smoke, the pattern of lung cancer occurrence in families of nonsmoking lung cancer patients differs from that in families of smoking lung cancer patients 134, 135. This suggests the presence of a high-risk gene that contributes to early-onset lung cancer in a population in which the probands are
nosmokers. Nevertheless, these results still do not clarify whether the evidence supporting a familial association suggests that the etiology of lung cancer includes shared genes or shared environment, or both. Certainly, environmental factors such as passive smoking have some influence on occurrence of lung cancer in the family. However, many other non-environmental mechanisms may provide additional explanations. Such mechanisms might include common genetic polymorphisms of carcinogen-metabolizing enzymes, mutations of tumor suppressor genes, or variability in DNA repair activity.

Few studies have evaluated the familial risk of lung cancer in parent-offspring relations according to the histological types and age at diagnosis. No distinct familial histology of lung cancer has emerged. No population-based study has been conducted to assess the role of family history in the development of multiple primary lung cancers.
4. AIMS OF THE STUDY

The thesis is a part of population-based study on familial risks for cancer based on the Swedish Family-Cancer Database.

The specific aims of the study were:
1. To evaluate the quality of the Swedish Family-Cancer Database
2. To estimate the effects of birth cohort, period, and age on the time trends of lung cancer incidence rates.
3. To analyze the gender-specific incidence of different histological types of lung cancer in Sweden, in order to find evidence for the changing histological patterns in lung cancer and the gender-related difference in sensitivity to tobacco-induced lung cancer.
4. To observe the association of lung cancer with malignant disease, in general and for specific sites, by comparing family history of malignant diseases according to probands.
5. To determine genetic predisposition to lung cancer by the features of age of cancer diagnosis, identification of certain types of cancer within families, and identification of individuals with multiple primary cancers.
6. To determine genetic patterns of lung cancer according to probands.
5. METHODS

5.1 Setting

Statistic Sweden has maintained a nationwide family database, the ‘Second Generation Register’ since 1995. This covers offspring (second generation) born after 1941 with their biological parents, and it has been renamed to the ‘Multi-Generation Register’. A scheme for linking the Multi-Generation Register with the Swedish Cancer Registry, national census data and death notifications is presented in Figure 5.1. In this thesis, we use the terms “the Swedish Family-Cancer Database”, “Family-Cancer Database” and “the Database” interchangeably.

![Diagram of data linkage]

**Figure 5.1** Linkage of data for the Swedish Family-Cancer Database

5.2 The Swedish Family-Cancer Database

The Swedish Family-Cancer Database was created in 1995. Each person is assigned a unique technical identification number (which is different from the national identification number, i.e. their “personal number”), allowing construction of families through the mother, for example. The Swedish Family-Cancer Database was
constructed in four expanded versions in 1996, 1998, 2000 and 2002 (Table 5.1). Initially, the database included offspring born in Sweden in 1941, along with their biological parents, as families, i.e. 6 million individuals. It was expanded in 1998 to 9.6 million individuals, and it reached over 10.2 million by the year 2002. This expansion covered offspring born after 1931, along with their parents. The number of cancers in the second generation increased from 20,000 in 1996 to 320,000 in 2000. In the parental generation, the increase was from 500,000 to 969,000 invasive cancers.

Table 5.1 The updated editions of the Swedish Family-Cancer Database

<table>
<thead>
<tr>
<th>Version</th>
<th>Updated year (year)</th>
<th>Offspring born (year)</th>
<th>Ages of offspring (years)</th>
<th>Cancer Registry (year)</th>
<th>No. of individuals (million)</th>
<th>No. of families (million)</th>
<th>No. of cancer cases (million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1996</td>
<td>1941-1994</td>
<td>0-53</td>
<td>1958-1994</td>
<td>6.0</td>
<td>1.3</td>
<td>0.6</td>
</tr>
<tr>
<td>IV</td>
<td>2002</td>
<td>1932-2000</td>
<td>0-68</td>
<td>1958-2000</td>
<td>10.2</td>
<td>3.3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The Fourth Swedish Family-Cancer Database, updated in 2002, includes people born in Sweden after 1931 with their biological parents, totalling over 10.2 million individuals (Table 5.2).

Table 5.2 Number of individuals and cancer cases in the Swedish Family-Cancer Database, 1958-2000

<table>
<thead>
<tr>
<th>Population</th>
<th>First primary invasive cancer cases</th>
<th>First primary in situ cases</th>
<th>All cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offspring</td>
<td>7888119</td>
<td>188627</td>
<td>319874</td>
</tr>
<tr>
<td>Son</td>
<td>3740079</td>
<td>77444</td>
<td>84296</td>
</tr>
<tr>
<td>Daughter</td>
<td>3548040</td>
<td>111183</td>
<td>235578</td>
</tr>
<tr>
<td>Parents</td>
<td>6468813</td>
<td>768335</td>
<td>968875</td>
</tr>
<tr>
<td>Father</td>
<td>3191892</td>
<td>390393</td>
<td>422946</td>
</tr>
<tr>
<td>Mother</td>
<td>3276921</td>
<td>377942</td>
<td>545929</td>
</tr>
<tr>
<td>Total individuals</td>
<td>10210536</td>
<td>802053</td>
<td>1019014</td>
</tr>
<tr>
<td>Men</td>
<td>5220666</td>
<td>407060</td>
<td>440818</td>
</tr>
<tr>
<td>Women</td>
<td>4989870</td>
<td>394993</td>
<td>578196</td>
</tr>
</tbody>
</table>

Parents’ ages were not limited, but offspring were 0 to 68 years of age (Table 5.2). The Database includes cancers from the nationwide Swedish Cancer Registry from
the years 1958-2000. These included 968,875 first primary cancers among the parental (first) generation and 319,874 first primary cancers among the offspring (second) generation. Regarding the Family-Cancer Database, it is worth pointing out that the parents were registered at the time of birth of the child. Thus, it is possible to track biological parents regardless of divorces and remarriages.

5.3 Patients and subjects

The Swedish Cancer Registry is based on compulsory notification of cases. The completeness of cancer registration is currently considered to be close to 100%; the percentage of cytologically or histologically verified cases of cancers has been close to 100%. A 4-digit diagnostic code according to the 7th revision of the International Classification of Diseases (ICD-7) was used. The following ICD-7 codes were pooled: upper aerodigestive tract cancer, codes 161 (larynx) and 140-148 (lip, mouth, tongue, pharynx), except for code 142 (salivary gland), and leukaemia, codes 204-207 (leukemias), 208 (polycythemia vera) and 209 (myelofibrosis). Basal cell carcinoma of the skin is not registered in the Cancer Registry. According to the ICD-7 classification, lymphomas are classified as lymphomas irrespective of the site at which they occur. These codes have been used since the start of cancer registration in Sweden (WHO/HS/CANC/24.1 Histology Code). These codes describe histology in main subgroups, such as adenocarcinoma and squamous cell carcinoma; and this classification is referred to as histology throughout the study.

For lung cancer, for example, ICD7 codes of 162 and 163 were used. The histological classification for lung cancer in the Cancer Registry was changed at the end of 1985. The original code of 196 ‘unspecified’ was then divided into small cell carcinoma (code 186) and large cell carcinoma (code 196). Before 1986, there were three main histological subtypes of lung cancer in the Cancer Registry: adenocarcinoma (code 096), squamous cell carcinoma (code 146) and unspecified lung cancer (code 196). After 1986, codes 186 and 196 have been equally common in the Cancer Registry. We limited papers III–V as to histological type of incidence, familial risk, and kappa statistics of lung cancer diagnosed since 1986 and divided the histological types of lung cancer into five subtypes: adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma and other types (Table 5.3).
Table 5.3 Classification of the main types of lung cancer according to histology and SNOMED

<table>
<thead>
<tr>
<th>Type</th>
<th>PAD</th>
<th>SNOMED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td>096</td>
<td>81403</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>146</td>
<td>80703</td>
</tr>
<tr>
<td>Small cell carcinoma</td>
<td>186</td>
<td>80433</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>196</td>
<td>80203</td>
</tr>
</tbody>
</table>

PAD: patho-anatomic diagnosis.  
SNOMED: systematized nomenclature of medicine.

From 1993 onwards, ICD-O-2/ICD with histopathological data according to the Systematized Nomenclature of Medicine (SNOMED, http://snomed.org) was used in the Swedish Cancer Registry. For example, in Table 5.3, the histopathological types were used for lung cancer. We refer to this classification as ‘SNOMED’ or histopathology.

5.4 Methods used in the study

5.4.1 Incidence rates
All tumors incidence rates were based on the data in the Swedish Family-Cancer Database, and they were essentially similar to rates in the Swedish Cancer Registry. In the study (Papers I-V), age-adjusted incidence rates were calculated based on the European standard population.

5.4.2 Standardized incidence ratios
Standardized incidence ratios (SIRs) were used to measure cancer risks for offspring according to occurrence of cancers in their families. SIRs were calculated for offspring whose parent or sibling had the same, concordant cancer, i.e. using parents or sibling as probands. Follow-up was started for each offspring at birth, immigration or January 1, 1958, whichever came last. Follow-up was terminated on diagnosis of first cancer, death, emigration, or the closing date of the study, December 31, 1996 (the 2nd edition of the Swedish Family-Cancer Database, Paper II). SIRs were
calculated as the ratio of observed (O) to expected (E) number of cases. The expected numbers were calculated from 5-year age-specific, sex-specific, tumour type-specific, period-specific (5 year bands), socioeconomic status-specific (6 groups: agriculture, manual worker, blue worker, professional, self-employed and others) and residential area-specific (3 groups: large cities, south, middle and north Sweden) standard incidence rates for all offspring lacking a family history of analyzed cancer. Confidence intervals (95%CI) were calculated assuming a Poisson distribution.

The following terms are used: (1) concordant cancer: same affected site between parent-offspring or sibling-sibling; (2) discordant cancer: different affected site between parent-offspring or sibling-sibling; (3) reverse order analysis: when an association is found between two sites, it can be tested in reverse order, e.g. SIR for offspring breast cancer by parental myeloma is 1.81, but SIR for offspring myeloma by parental breast cancer is only 0.89, casting doubt on the association. Because of different age distributions of the offspring and parental populations, the comparisons do not necessarily agree but, if they do, an increased risk is unlikely to be due to chance; the terms of risks according to the probands are defined as follows:

a) Cancer risks in offspring by parental cancer only (“parent only”): cancer risks in offspring who had a parent affected with cancer, but no sibling affected with cancer, were referred to it as “offspring risk”.

b) Cancer risks in offspring by sibling cancer only (“sibling only”): cancer risks in offspring who had a sibling affected with cancer, but no parent affected with the cancer, were referred to it as “sibling risk”.

c) Cancer risks in offspring by both parental and sibling cancer (“parent and sibling”): cancer risks in offspring who had both a parent and a sibling affected with the cancer. Sibling risk/offspring risk ratio: SIR in b) divided by SIR in a).

5.4.3 Sibling risks

Risks for siblings were calculated using the cohort method, which has been described and discussed elsewhere. In the cohort method we simply define a cohort of individuals with at least one affected sibling, and compute the incidence rates in this cohort over the period from 1958 to the closing date of the study. Note that in a family with two or more siblings affected, each affected individual is included in the cohort.
(as the sibling of an affected individual), so that the observed cancers in the cohort are not independent. Thus, a family with 2 affected siblings will contribute 2 observed cancers, whereas it is in fact only one independent event, leading to relative risk confidence intervals that are too narrow.

5.4.4 The kappa test
The kappa index of agreement for categorical data was used to compare the agreement against that which might be expected by chance \(^{148, 149}\). The following equation is used: (observed agreement - expected agreement)/(1 - expected agreement). The kappa can assume values between -1 and 1; 0 indicates complete chance occurrence, and -1 or 1 show a complete concordance or discordance. Various scales to assess the significance of kappa have been proposed: values between 0.41 and 0.60 are considered to be moderate agreement, and values 0.61–0.80 are considered to be substantial agreement \(^{148, 149}\).

5.4.5 The population–attributable fraction
The population–attributable fraction (PAF) of cases with a family history of cancer was estimated as follows: proportion of cases with a family history x (familial SIR-1)/familial SIR, as defined by Miettinen \(^{150, 151}\). Family history can be defined through different probands, parents, children, and siblings, their combinations, or all first-degree relatives. SIR was estimated as for relative risk, i.e. the expected numbers were calculated for those lacking a family history.

5.4.6 Poisson regression
The Poisson regression model was used to analyze trends in the incidence of cancer to quantify the effects of one or several variable(s) on another variable \(^{152-154}\). This method transforms the underlying incidence rate to a regression function that describes the relationship between predictor variables such as age, period of diagnosis year, birth cohort, and other variables that are estimated directly by maximum likelihood techniques. The statistical significance of each effect was examined by maximum likelihood procedure. Relative risks were calculated from exponentiated regression coefficients.
5.5 Study design, subjects and methods of follow-up

5.5.1 Papers I and III

Study design and subjects
The third edition of the Swedish Family-Cancer Database was created in the year 2000. This register was linked by each unique national registration number to the Cancer Registry from the years 1958–1998. Cancer registration is considered to be close to 100%. The Database contains offspring who were born in Sweden after 1931, with their biological parents registered at the time of birth. Some mothers and a larger number of fathers were born in the 1800s, with the oldest father being born in 1864. A 4-digit diagnostic code in accordance with the ICD-7 and subsequent ICD classifications are available. Cancers are also recorded according to the first or subsequent primary cancer, and cancer in situ.

For lung cancer in paper III, ICD-7 codes 162 and 163 were used. Histology of lung cancer was divided by histological type of lung cancer into three subtypes: adenocarcinoma, squamous cell carcinoma and other types. Family history information was collected on all first-degree relatives (parents, siblings, and children), but only the parent-offspring relationship was used in the study. All individuals, parents and offspring were included in the analysis of second events. To be included, the first malignancies had to have a median survival time of at least 2 years, which was not met by esophageal, gastric, liver, pancreatic and lung cancers. At least 5 cases had to be observed for men or women.

Follow-up
In paper I, follow-up was started for each offspring at birth, immigration or January 1, 1958, whichever came last. Follow-up was terminated on diagnosis of first cancer, death, emigration, or the closing date of the study (December 31, 1998).

In paper III, follow-up was started for each offspring at birth, immigration or January 1, 1961, whichever came last. Follow-up was terminated on diagnosis of first cancer, death, emigration, or the closing date of the study, December 31, 1998. Second primary lung cancer was followed from the diagnosis of the first primary cancer and ending with the date of diagnosis of a second primary cancer, date of death, date of
emigration or December 31, 1998, whichever came first. Even synchronous second cancers were included and the follow-up time was divided into three periods, namely less than 1, 1 to 10 and more than 10 years, allowing assessment of the effect of follow-up time.

5.5.2 Paper II
Study design and subjects
The second edition of the Swedish Family-Cancer Database was created in 1998. This register was linked by each unique national registration number to the Cancer Registry from the years 1958–1996. The Database contains offspring who were born in Sweden after 1934, with their biological parents registered at the time of birth, totaling over 9.6 million individuals. ICD7 codes 162 and 163 were used for lung cancer. The histological classification of lung cancer was divided into three subtypes: adenocarcinoma, squamous cell carcinoma and other types.

Follow-up
Person years were calculated for the living individuals in each period. 5-year age-specific and period-specific incidence rates of lung cancer by histological types were calculated in 4 periods of diagnosis (1958–1969, 1970–1979, 1980–1989 and 1990–1996). Relative weights were given according to the 1970 European standard population to give the age-standardized incidence rates. The number of cases was determined in each 10-year age band (15–24, 25–34…, 85+) for each sex. In offspring, all cases were divided further into 9 age-groups (15–19, 20–24 …, 55–61).

5.5.3 Papers IV and V
Study design and subjects
The fourth edition of the Swedish Family-Cancer Database was created in the year 2002. This register was linked by each unique national registration number to the Cancer Registry from the years 1958–2000. We limited offspring diagnosed between years 1991 and 2000 to eliminate any possibility of bias (paper IV). Histological type of lung cancer was divided into five subtypes: adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma and other types. Family history information was collected on all first-degree relatives (parents, siblings, and children). All tumor incidence rates were based on the data in the Swedish Family-Cancer
Database. The risk of lung cancer was calculated for offspring whose first-degree relatives presented with lung cancer and their risk was compared to the rate of lung cancers among persons with no family history. For each individual, information on socio-economic status and residential area was also included.

**Follow-up**

In paper IV, follow-up was started for each offspring at birth, immigration or January 1, 1991, whichever came last. Follow-up was terminated on diagnosis of first cancer, death, emigration, or the closing date of the study (December 31, 2000). When more than 2 affected offspring were found in any family, they were counted as independent events.

In paper V, follow-up was started for each offspring at birth, immigration or January 1, 1961, whichever came last. Follow-up was terminated on diagnosis of first cancer, death, emigration, or the closing date of the study (December 31, 2000). For histology-specific analysis, we limited the number of lung cancer patients diagnosed after 1986 and divided the histological types of lung cancer into five subtypes: adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma and other types. For second primary lung cancer, follow-up was started from the diagnosis of the first lung cancer and ended with the date of diagnosis of a second primary cancer, date of death, date of emigration or December 31, 2000, whichever came first. Follow-up time was divided into three periods, namely less than 1, 1 to 10 and more than 10 years, allowing assessment of the effect of follow-up time.

**5.6 Statistical analysis**

**Paper I**

Overall and site-specific SIRs were calculated for concordant cancer in offspring. The SIRs were further stratified according to age at diagnosis, both in parents and offspring. Incidence rates were calculated based on the Family-Cancer Database and the Swedish Cancer Registry, and both rates were adjusted to the standard European population.

**Paper II**
Histology-specific incidence trends of lung cancer were calculated separately for men and women, based on the data in the Swedish Family-Cancer Database. Poisson regression was used to quantify the effects of variables while adjustment was made for age and period of diagnosis. Other predictor variables included in this statistical analysis were socioeconomic status (4 groups: farmers, professionals, blue collar workers and others), and areas of living (two groups: big city-Stockholm, Malmö, Göteborg, and other areas). The records for cells with no person-years of observation were omitted. One last level was chosen as a baseline category.

Paper III
Histology-specific SIRs for lung cancer were calculated in offspring when parents presented with any cancer. The SIRs were stratified further according to age at diagnosis, both in parents and offspring. The SIRs for second lung cancers were calculated separately for men and women, starting follow-up from the diagnosis of any first primary tumor. The population-attributable proportion of cases with a family history of lung cancer was estimated.

Paper IV
Age-specific, histology-specific SIRs were calculated separately for lung cancer in offspring by parental and sibling probands. Age-specific incidence rates of lung cancer by histological types and family history (first-degree relatives with/without lung cancer) were based on the Swedish Family-Cancer Database. Sibling risks (without an affected parent) were calculated.

Kappa measures of agreement between histological types of lung cancer in offspring/sibling and first-degree relatives were calculated. For simplicity, in the present report only results with positive values of kappa are presented.

Paper V
Age-specific SIRs were calculated for lung cancer separately for men and women when a first-degree relative presented with one or two lung cancers. The SIRs for second lung cancers were calculated separately for men and women according to family history, starting follow-up from the diagnosis of first primary lung cancer. Incidence rates of first and second primary lung cancer were calculated separately for
men and women according to family history. Kappa measures of agreement between histological types of first and second primary lung cancers of lung cancer patients were calculated using the method described for paper IV.

5.7 Statistical program

The statistical analyses in the Swedish Family-Cancer Database study were performed using SAS 8.20 statistical software. This is licensed to HIS, Karolinska Institutet.
6. RESULTS AND DISCUSSION

6.1 The Swedish Family-Cancer Database

The Swedish Family-Cancer Database is the largest population-based dataset using for studies on familial risks for cancer. In the Database, the coverage of families and cancers is practically complete and the estimates of familial risks are free from ascertainment and recall biases, which often beset family studies.

**Missing linkage (Paper I)**

Due to the Second Generation Register, some limitations have been introduced into the Database. One is that the data are from those born in 1932 and later, causing truncation at persons aged 66 years and younger, for example, in the third edition of the database (1958-1998). The second limitation is that the Second Generation Register lacks information on those born in 1932 or later who died before the 1990s.

According to Table 6.1, 2.2% of the more than 7 million offspring had cancer, and 97.6% had a link to at least one parent. The linkage was 98.8% among those alive as of the end of 1998. Among the 216,550 individuals who were deceased, the linkage existed among 71.3%. There was a shortage of parental information on 15,560 deceased offspring who had been diagnosed with cancer; this was 7.2% of the deceased offspring and implied that 9.8% of all offspring with cancer had no links to parents. Offspring who died before 1960 are missing from the Database.

<table>
<thead>
<tr>
<th>Offspring</th>
<th>Total number of offspring</th>
<th>Offspring with cancer</th>
<th>Offspring linked to parent</th>
<th>Offspring with cancer not linked to parent</th>
</tr>
</thead>
<tbody>
<tr>
<td>All offspring</td>
<td>7017638</td>
<td>157779</td>
<td>2.2</td>
<td>6846587</td>
</tr>
<tr>
<td>Living offspring</td>
<td>6801088</td>
<td>102153</td>
<td>1.5</td>
<td>6692209</td>
</tr>
<tr>
<td>Deceased offspring</td>
<td>216550</td>
<td>55626</td>
<td>25.7</td>
<td>154378</td>
</tr>
</tbody>
</table>

Table 6.1 Number of offspring in the Family-Cancer Database in 1958–1998

The problem of linking deceased offspring to their parents – which has existed throughout the history of the Second Generation Register – has been partially
resolved, as less than 10% of the deceased offspring with cancer now lack a link to a parent. We compared the familial risks for concordant cancer in offspring diagnosed before 1991 and those diagnosed in 1991 or later (Table 6.2), and the results showed no overall difference between the two periods, the risks being 1.99 and 2.01, respectively. However, there were some differences for individual sites, which were overshadowed by the common familial cancers that showed no difference between the periods.

Table 6.2 Risks for concordant cancer in offspring diagnosed before 1991 and later

<table>
<thead>
<tr>
<th>Parents cancer</th>
<th>Before 1991</th>
<th></th>
<th></th>
<th></th>
<th>1991 and later</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>SIR</td>
<td>95%CI</td>
<td>O</td>
<td>SIR</td>
<td>95%CI</td>
<td></td>
</tr>
<tr>
<td>Upper aerodigestive tract</td>
<td>8</td>
<td>1.62</td>
<td>0.69 3.20</td>
<td>13</td>
<td>1.33</td>
<td>0.70 2.28</td>
<td></td>
</tr>
<tr>
<td>Stomach</td>
<td>7</td>
<td>0.68</td>
<td>0.27 1.41</td>
<td>55</td>
<td>1.91</td>
<td>1.44 2.49</td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>113</td>
<td>2.07</td>
<td>1.70 2.48</td>
<td>227</td>
<td>2.12</td>
<td>1.85 2.41</td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>32</td>
<td>2.12</td>
<td>1.45 2.99</td>
<td>71</td>
<td>1.75</td>
<td>1.36 2.20</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>1</td>
<td>0.35</td>
<td>0.00 2.03</td>
<td>28</td>
<td>1.69</td>
<td>1.12 2.45</td>
<td></td>
</tr>
<tr>
<td>Pancreas</td>
<td>4</td>
<td>2.11</td>
<td>0.55 5.46</td>
<td>30</td>
<td>1.64</td>
<td>1.10 2.34</td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>36</td>
<td>1.68</td>
<td>1.18 2.33</td>
<td>253</td>
<td>1.90</td>
<td>1.67 2.15</td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td>766</td>
<td>1.89</td>
<td>1.76 2.03</td>
<td>1312</td>
<td>1.83</td>
<td>1.74 1.94</td>
<td></td>
</tr>
<tr>
<td>Cervix</td>
<td>60</td>
<td>1.94</td>
<td>1.48 2.50</td>
<td>32</td>
<td>1.88</td>
<td>1.28 2.65</td>
<td></td>
</tr>
<tr>
<td>Endometrium</td>
<td>35</td>
<td>3.24</td>
<td>2.25 4.50</td>
<td>62</td>
<td>2.60</td>
<td>1.99 3.34</td>
<td></td>
</tr>
<tr>
<td>Ovary</td>
<td>45</td>
<td>2.27</td>
<td>1.65 3.04</td>
<td>80</td>
<td>3.46</td>
<td>2.74 4.31</td>
<td></td>
</tr>
<tr>
<td>Prostate</td>
<td>32</td>
<td>2.17</td>
<td>1.48 3.07</td>
<td>493</td>
<td>2.52</td>
<td>2.30 2.75</td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>15</td>
<td>1.05</td>
<td>0.59 1.73</td>
<td>50</td>
<td>1.93</td>
<td>1.43 2.55</td>
<td></td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>50</td>
<td>1.87</td>
<td>1.39 2.47</td>
<td>81</td>
<td>1.72</td>
<td>1.36 2.14</td>
<td></td>
</tr>
<tr>
<td>Melanoma</td>
<td>127</td>
<td>2.67</td>
<td>2.23 3.18</td>
<td>116</td>
<td>2.32</td>
<td>1.92 2.78</td>
<td></td>
</tr>
<tr>
<td>Skin, squamous cell</td>
<td>15</td>
<td>1.64</td>
<td>0.91 2.71</td>
<td>54</td>
<td>2.60</td>
<td>1.95 3.40</td>
<td></td>
</tr>
<tr>
<td>Nervous system</td>
<td>81</td>
<td>1.63</td>
<td>1.30 2.03</td>
<td>82</td>
<td>1.77</td>
<td>1.41 2.20</td>
<td></td>
</tr>
<tr>
<td>Thyroid gland</td>
<td>48</td>
<td>8.31</td>
<td>6.13 11.02</td>
<td>22</td>
<td>7.53</td>
<td>4.71 11.41</td>
<td></td>
</tr>
<tr>
<td>Endocrine glands</td>
<td>34</td>
<td>2.49</td>
<td>1.72 3.48</td>
<td>29</td>
<td>2.26</td>
<td>1.51 3.25</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1509</td>
<td>1.99</td>
<td>1.89 2.09</td>
<td>3090</td>
<td>2.01</td>
<td>1.94 2.09</td>
<td></td>
</tr>
</tbody>
</table>

Bold type: 95%CI does not include 1.00.

We also compared age-specific incidence rates of cancers in the Swedish Family-Cancer Database with the Swedish Cancer Registry (Figure 6.1). For all cancers, the incidences in the Family-Cancer Database and the Cancer Registry were very similar up to about age 70, but thereafter the rate in the Database dropped relative to that in the Cancer Registry. The reason for the deviation in rates among the elderly is likely to be either population selection or a preferential loss of older individuals diagnosed with cancer. However, the almost superimposable age-incidence curves up to age 70
between the Cancer Registry and the Family-Cancer Database show that the rates from the two sources are fully comparable.

Figure 6.1 Age-specific incidence for cancer in Family-Cancer Database and in the Cancer Registry, 1958–1998

Bias and confounding
In this thesis, a nationwide Family-Cancer Database has been formed by combining a national dataset on families with the Swedish Cancer Registry and used in papers I to V. All cancer data have been medically verified and both cancer and family data have a practically complete national coverage, which should minimize the selection bias in our studies.

Diagnosis of cancer in a family member, particularly in a sibling, is likely to alert the remaining relatives to search for symptoms in themselves and to seek medical advice. Increased medical awareness may also result in earlier diagnosis of cancer. In papers III and V, it cannot be excluded that a history of lung cancer might result in lead-time bias, i.e. earlier diagnosis of subsequent cancers. Thus, an apparent excess lead-time bias would be noted in cancers for which screening methods are available. It is
sometimes impossible to distinguish second cancers as independent primaries from the apparent second cancers as recurrences from the first primary tumour. The Swedish Cancer Registry has clear instructions about the reporting of multiple primary malignancies, and a re-evaluation of 209 multiple primary cancers found 98% of second malignancies to be correctly classified. Almost all cancer notifications registered at the Swedish Cancer Registry bear a histological or cytological verification, and diagnostic misclassification is unlikely. In papers III and V, we also have at least three important technical considerations to do with subsequent cancers. One is that second cancers were included after the first malignancies had to had a median survival of at least two years and a diagnosis that was different by topology (paper III). Topology involved a different anatomical site (by ICD7). Another consideration is that we divided the follow-up time into three periods, < 1, 1–10 and > 10 years (papers III and V). The period involving > 10 years of follow-up may reveal the effects due to radiotherapy and chemotherapy, prompting comparison of the SIRs between the two periods. The third consideration is that the estimation of SIRs is based on adjustment for residence and socioeconomic status, which may be the potential confounding factors, and this has been ignored by most studies on second cancers (papers III and V).

Birth order and family size may affect cancer risk: genetic risk factors, early-onset cancers or other inherited diseases may limit the reproductive period of the parents and show higher risks for small families because of selection. Various socioeconomic and cultural factors are relevant to large families and these could also be related to risk of cancer. However, using the Database analysis, Hemminki et al. have found that excluding two cancer sites (breast and melanoma), birth order and family size, has no major effect on the risk for common cancers.

Lack of information of possible confounding factors is a weakness in most register-based cohort studies, and also in our study using the Swedish Family-Cancer Database. The major problem in the work for this thesis, especially in papers II, III, IV and V regarding familial risks for lung cancer, is lack of detailed smoking information and of certain important lifestyle factors. Thus without this information, a familial/genetic effect cannot be clearly separated from the smoking effect. However, there are at least five ways to consider this limitation. Firstly, the unique aspect in the studies is that the
data on family relationships and cancers were obtained from registered sources of practically complete coverage. Secondly, since smoking is the single most important risk factor for lung cancer and there is a lag period of 20 to 50 years from the initiation of smoking until the development of lung cancer. The effect of stopping smoking is the stabilization of, rather than decrease in, the absolute lung cancer risk. In study IV, the higher risk of lung cancer at ages of early onset (< 50 years) is reasonable when compared to the risks between spouses (Table 6.4), that it could not be explained by smoking effects only. Thirdly, we assessed the level of environmental effect (among family members) by comparing cancer risks in spouses (Table 6.4) and in siblings (Table 3.5) born close or far apart. The risk between spouses showed that the degree of environmental sharing does not exceed an SIR of 1.24. Among siblings, for lung cancer, there was no difference (p = 0.73) in risks by age difference. Additionally, familial risks for full siblings and half siblings have also been used to evaluate the shared environmental or genetic factors: if environmental factors were important for familial aggregation the risk should not have been halved. For example, for breast cancer, half sisters showed an excess of familial risks that was exactly half that of full sisters, in line with the low environmental contribution to the familial clustering of breast cancer. Fourthly, shared environmental effects for common cancers have been suggested in twin and family studies. For example, for lung cancer, the shared environmental effects (such as shared smoking habits or environmental smoke) explain no more than about 10%. Lastly, smoking shows social class dependence, and some effect of smoking was most likely controlled by the adjustment for socioeconomic status, period and region of residence. Over 90% of lung cancer patients are smokers, and we have no reason to believe that the proportion is less among familial cancers. Most epidemiological data are less than 90% accurate, anyway.

In the Swedish Family-Cancer Database, one problem has been the introduction of the personal registration number, which took place in Sweden in 1947, to foreign-born individuals and to children in the early period. However, the proportion of offspring without links to parents in the current Database is so small that bias is unlikely in family studies.
Additionally, in this thesis, the highest age of the offspring generation was 68 years, while in the parental generation there was no such truncation. Censoring bias due to observed ages in offspring may give rise to observed anticipation, and truncation bias due to parents who were diagnosed before the starting year of the cancer registry in 1958. However, statistical tests for age-at-onset anticipation with affected parent-child pairs have been described previously and showed no difference between parent-offspring pairs. In the construction of the Swedish Family-Cancer Database in paper I, the controls that we carried out showed data of good quality. In paper IV, in order to make the familial risks comparable with the age of the offspring population, we calculated the risks for offspring when parental age was limited to 68 years or younger to match the age range of the offspring population.

To eliminate possible ascertainment bias, in paper IV the follow-up for offspring was from years 1991 to 2000 in order to ensure complete linkage between parents and offspring. The completeness of Swedish cancer registration is considered to be close to 100%, and the percentage of cytologically and histologically verified cases of cancer is also close to 100%.

**External validity**

External validity, or the possibility of generalizing the findings to populations other than the one under study, relies on knowledge about the characteristics of the study population, and of the effect of exposure in different settings.

The familial risk for 20 cancer sites from the Swedish Family-Cancer Database (paper I) and Utah Cancer Database is listed in Table 6.3. The Utah Cancer Database was matched to the Utah Genealogic Database with the Utah Cancer Registry to determine the frequency of cancer in the first-degree relatives (parents, siblings and offspring) of those cases. Cancer diagnosis was prior to 80 years. In the Swedish Family-Cancer Database, the familial risks shown in the table are the risks for cancer in offspring when a parent with concordant cancer was diagnosed at ages 0 to 66 years. There is similarity in risks across cancer sites, with a mean value of 2.11 for Utah and 1.72 for the Swedish offspring. Of the 20 familial sites that were positive in both studies, the risk estimates were in agreement in 12 sites (60%). Among the remaining 8 sites, the Utah estimates were higher than for the Swedish Database for
upper aerodigestive tract, stomach, colon, lung, breast, kidney and leukemia, whereas, the Swedish Database was higher for endometrial cancer. Compared with the Utah Population Database, the results from the Swedish Family-Cancer Database are in reasonable agreement.

Table 6.3 Familial risk for cancer in Sweden (paper I) and Utah

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>Sweden (offspring) (1)</th>
<th>Utah (total) (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper aerodigestive tract</td>
<td>1.21</td>
<td>2.80</td>
</tr>
<tr>
<td>Stomach</td>
<td>1.05</td>
<td>2.09</td>
</tr>
<tr>
<td>Colon</td>
<td>1.72</td>
<td>2.67</td>
</tr>
<tr>
<td>Rectum</td>
<td>1.54</td>
<td>1.78</td>
</tr>
<tr>
<td>Liver</td>
<td>0.97</td>
<td>2.13</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1.20</td>
<td>1.25</td>
</tr>
<tr>
<td>Lung</td>
<td>1.39</td>
<td>2.55</td>
</tr>
<tr>
<td>Breast</td>
<td>1.69</td>
<td>1.83</td>
</tr>
<tr>
<td>Cervix</td>
<td>1.69</td>
<td>1.74</td>
</tr>
<tr>
<td>Endometrium</td>
<td>2.12</td>
<td>1.32</td>
</tr>
<tr>
<td>Ovary</td>
<td>2.34</td>
<td>2.05</td>
</tr>
<tr>
<td>Prostate</td>
<td>2.20</td>
<td>2.21</td>
</tr>
<tr>
<td>Testis</td>
<td>4.69</td>
<td>8.32</td>
</tr>
<tr>
<td>Kidney</td>
<td>1.37</td>
<td>2.45</td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>1.72</td>
<td>1.53</td>
</tr>
<tr>
<td>Melanoma</td>
<td>2.57</td>
<td>2.10</td>
</tr>
<tr>
<td>Nervous system</td>
<td>1.39</td>
<td>1.96</td>
</tr>
<tr>
<td>Thyroid gland</td>
<td>7.47</td>
<td>8.57</td>
</tr>
<tr>
<td>Non-Hodgkin's lymphoma</td>
<td>1.45</td>
<td>1.68</td>
</tr>
<tr>
<td>Leukemia</td>
<td>1.12</td>
<td>4.00</td>
</tr>
<tr>
<td>All sites</td>
<td>1.72</td>
<td>2.11</td>
</tr>
</tbody>
</table>

(1): Cancers diagnosed in offspring were 0-66 years of age.
(2): Cancers diagnosed prior to age 80 years are limited.

Familial risks for all concordant cancer (according to the fourth edition of the Family-Cancer Database)

From the Family-Cancer Database updated in 2002, we followed up cancer cases from the years 1991 to 2000 in order to ensure complete linkage between parents and offspring and 112,216 offspring cancers was included. Familial risks were calculated for 89,576 offspring when parent was affected by a concordant cancer for sites in which > 5 familial pairs were found (Table 6.4)\textsuperscript{35}. All cancers showed a familial effect; SIRs for familial risk ranged from 3.95 for esophageal cancer and Hodgkin’s disease to 1.67 for breast cancer and 1.66 for bladder cancer.
Table 6.4 Risks for cancer in offspring and spouses \(^{35}\)

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Cancer in offspring by parental cancer</th>
<th>Cancer in husband by wife's cancer</th>
<th>Cancer in wife by husband's cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>SIR</td>
<td>O</td>
</tr>
<tr>
<td>Upper aerodigestive tract</td>
<td>27</td>
<td>1.55</td>
<td>32</td>
</tr>
<tr>
<td>Esophagus</td>
<td>7</td>
<td>3.95</td>
<td>3</td>
</tr>
<tr>
<td>Stomach</td>
<td>59</td>
<td>2.05</td>
<td>187</td>
</tr>
<tr>
<td>Colorectum (^{(1)})</td>
<td>524</td>
<td>1.95</td>
<td>1003</td>
</tr>
<tr>
<td>Liver</td>
<td>31</td>
<td>1.66</td>
<td>77</td>
</tr>
<tr>
<td>Pancreas</td>
<td>33</td>
<td>1.70</td>
<td>77</td>
</tr>
<tr>
<td>Lung</td>
<td>292</td>
<td>1.83</td>
<td>458</td>
</tr>
<tr>
<td>Breast</td>
<td>1418</td>
<td>1.67</td>
<td>36</td>
</tr>
<tr>
<td>Kidney</td>
<td>57</td>
<td>1.82</td>
<td>79</td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>103</td>
<td>1.66</td>
<td>151</td>
</tr>
<tr>
<td>Melanoma</td>
<td>159</td>
<td>2.43</td>
<td>78</td>
</tr>
<tr>
<td>Skin, squamous cell</td>
<td>69</td>
<td>2.56</td>
<td>110</td>
</tr>
<tr>
<td>Nervous system</td>
<td>102</td>
<td>1.78</td>
<td>63</td>
</tr>
<tr>
<td>Thyroid gland, nonmedullary</td>
<td>10</td>
<td>2.78</td>
<td>6</td>
</tr>
<tr>
<td>Endocrine glands</td>
<td>33</td>
<td>2.05</td>
<td>19</td>
</tr>
<tr>
<td>Non-Hodgkin's lymphoma</td>
<td>71</td>
<td>1.86</td>
<td>71</td>
</tr>
<tr>
<td>Hodgkin's disease</td>
<td>8</td>
<td>3.95</td>
<td>1</td>
</tr>
<tr>
<td>Myeloma</td>
<td>21</td>
<td>3.31</td>
<td>13</td>
</tr>
<tr>
<td>Leukemia</td>
<td>52</td>
<td>1.80</td>
<td>58</td>
</tr>
<tr>
<td>All</td>
<td>3076</td>
<td>1.81</td>
<td>2522</td>
</tr>
</tbody>
</table>

Bold type: 95%CI does not include 1.00. Underlining=borderline significance.

\(^{(1)}\) adenocarcinoma only.

Additionally, we assessed the level of environmental effect by comparing cancer risks in spouses (Table 6.4) \(^{35}\). In the familial setting, any reference to heritable effects needs to exclude or quantify the contribution of environmental effects. The present data on spouse risks show that the degree of environmental sharing does not exceed an SIR of 1.24, and can only be noted for cancers with known strong environmental risk factors: lung, gastric and skin cancer and melanoma. Our earlier data also showed sharing for genital and early-onset pancreatic cancers \(^{40, 41}\). Thus, for most other sites, heritability is likely to be the main contributor. However, spouse concordance will not detect environmental sharing early in life. Migrant studies from Sweden show that the cancer patterns of immigrants are largely set during the two first decades of life \(^{48, 49}\); thus, spouse correlation would only detect strong effects that spouses share. With these caveats, familial risks were shown at every cancer site presented. Although we have no possibility of discussing to what extent known susceptibility genes explain the observed aggregation, we can refer to published work on some specific cancer
types. For breast cancer, known genes are assumed to explain about one quarter of the familial clustering 164-166.

6.2 Lung cancer histology and incidence trends (Papers II and IV)

Follow-up of the Swedish Family-Cancer Database covered years 1991–2000 from the Swedish Cancer Registry in paper IV: 2,813 lung cancers in sons and 2,477 in daughters of ages 0–68 years were diagnosed. The histological types of lung cancer were divided into five subtypes: adenocarcinoma, squamous cell carcinoma (SCC in Figure 6.2), large cell carcinoma, small cell carcinoma, and other types (Figure 6.2). Adenocarcinoma accounted for 28% and 23% in sons and fathers, respectively, and squamous cell carcinoma accounted for 22% and 29%. In women the most common histological type was adenocarcinoma, accounting for 35% and 40% in mothers and daughters, respectively.

![Figure 6.2 Distribution of histological types of lung cancer in offspring and parents](Image)

The second edition of the Swedish Family-Cancer Database included a total of 45,297 lung cancer cases between 1958 and 1996 in paper II. The incidence rates of all lung cancer, squamous cell carcinoma (SCC in Figure 6.3) and other subtypes in men peaked in the period 1980–1990, and then decreased slightly thereafter. In contrast, in
women the incidence rates of adenocarcinoma have continued to increase throughout the period (Figure 6.3), with all the subtypes except squamous cell carcinoma increasing linearly. The incidence ratio of male to female for all lung cancer was 2.8. For squamous cell carcinoma, the ratio was 5.3 for the whole period: 12.4 in the beginning and 3.6 at the end. For adenocarcinoma, the ratio was stable at about 1.5 for all the whole period. The Poisson regression analysis of lung cancer by histological subtypes showed that the highest relative risk of adenocarcinoma was between 1980 and 1989, a decade later than for the other subtypes. The birth cohort of 1940–1949 was at the highest risk of adenocarcinoma, whereas for squamous cell carcinoma the cohort of 1920–1929 was at the highest risk.

Figure 6.3 Age-specific incidence trends of lung cancer in men and women by histology

We used the population-based Database to examine the time trends of lung cancer in Sweden by histological type and to fit a parametric model in order to identify the components of age, period of diagnosis, and birth cohort as determinants of the observed time trends. Even though our Database lacked personal history of smoking, our data could support a compatible result with international experience in explanation of trends in tobacco use and lung cancer incidence. As there was a clear
history of proportion of smoking trends in the Swedish population\textsuperscript{167, 168}, and in the European Union in the 1990s, paradoxically, Sweden had the lowest proportion of male smokers and the highest proportion of female smokers\textsuperscript{169}, and also a clear history of tobacco consumption\textsuperscript{169, 170}.

The decrease in incidence rates for squamous cell carcinoma among men was probably due to a decrease in the percentage of smokers after the 1960s (45% in the 1960s, declining towards 20% in the 1990s) and the change to low-tar filter cigarettes in Sweden\textsuperscript{167, 168, 171}. The use of filter cigarettes was more common in the new series of patients than in the older one, and this change has been suggested to contribute to the increase in adenocarcinomas\textsuperscript{172}. Comparing the cross-sectional incidence trends of lung cancer in men and women with the history of cigarettes in Sweden, the incidence peaks are in line with the market penetration of low-tar filter cigarettes, but shifted 1-2 decades in time. Also, the stable incidence ratio between adenocarcinomas in males and females is reasonably consistent with the prevalence of smokers and their daily consumption a few decades ago. We have suggested that modern cigarettes are associated with a markedly higher risk of adenocarcinoma than of squamous cell carcinoma, and an equal sensitivity of men and women to tobacco-induced lung cancer. Our results agree with similar suggestions from studies in Denmark\textsuperscript{110}, Germany and Italy\textsuperscript{111}, and also supported by a recent cohort study from USA\textsuperscript{112}.

### 6.3 Familial risks for lung cancer (Papers III, IV and V)

Familial clustering of cancer may be due to environmental factors shared by family members or to shared genes\textsuperscript{9, 40, 84}. To search for evidence for a genetic predisposition in lung cancer, we analyzed familial risks for lung cancer by family history, age at onset, histology, and multiple primary tumors. The results are described in papers III, IV and V.

Altogether, 4,524 cases of lung cancer in offspring were observed in paper III. The SIR was 1.87 (95% CI 1.66–2.10) compared with those without family history. 6.39% of lung cancer patients had parents who were affected by lung cancer, and the population-attributable risk of all lung cancer was 2.97%. Lung cancer in offspring
was associated with parental rectal, cervical, kidney, urinary bladder and endocrine gland tumors.

In paper IV, follow-up of the Swedish Family-Cancer Database covered years 1991 to 2000 of the Swedish Cancer Registry. Among lung cancer cases, 187 men and 168 women was diagnosed with a first degree relative with lung cancer. For offspring diagnosed before 50 years of age, the SIRs for histological types of lung cancer between offspring and parents were 1.98 for adenocarcinoma (offspring) and small cell/large cell carcinoma (parents), 2.66 for squamous cell carcinoma -small/large cell carcinoma (offspring/parents), 3.54 for small cell carcinoma and squamous cell carcinoma, and 3.70 for large cell carcinoma and adenocarcinoma. The highest SIR for lung cancer in offspring by parental family history was shown in individuals aged 35–39 years (SIR = 2.48, 95%CI 0.89–5.44) and 50–54 years (SIR = 2.20, 95%CI 1.64–2.88). The overall kappa value for the agreement of histological types of lung cancer between offspring and parents was 0.03.

Among the first-degree relatives in paper V, patients had an increased familial risk when their first-degree relatives were diagnosed with two lung cancers relative to those whose first-degree relatives were diagnosed with one lung cancer. The SIRs for all lung cancer were 1.70 for one and 5.23 for two lung cancers in probands, and the highest SIR (11.36) was seen for lung cancers in women whose age at diagnosis was less than 60 years and whose first-degree relatives had two lung cancers.

Our study has provided evidence concerning familial risks in lung cancer by histological types and age of onset in first-degree relatives. The detailed analysis of age-specific familial risks helped to unravel evidence of a heritable component in lung cancer. These data support and strengthen previous epidemiological studies. The results of the kappa values were close to 0 for histological types in parent-offspring, indicating that the histological types of lung cancer are not genetically determined.

Besides family history, early age of onset of cancer and multiple primary cancers are also hallmarks of hereditary cancer. In paper IV, the results showed that there was a distinct early onset component in familial risk, and in paper V, the highest SIR was
shown for lung cancer at an age of diagnosis of less than 60 years when there were first-degree relatives with two lung cancers.

Familial aggregation of lung cancer may be environmental in part, due to shared smoking habits between family members, as has been shown in spouse correlation for lung cancer. In the 1960s, investigators attempted to examine the smoking habits of family members. The results indicated that the excess risk of lung cancer in case relatives compared to control relatives occurred irrespective of the relative’s smoking history. Also, the familial risk estimate in that study was greater for nonsmokers than for smokers, suggesting that the risk was not simply due to shared smoking habits in families of lung cancer patients.

6.4 Sibling risk (Paper IV)

Risk ratio for histological types of lung cancer in offspring of affected parents and among siblings (siblings/parents) is shown in Figure 6.4. Parental age was limited to 68 years or younger to match the age range of the offspring population. The SIR ratio was 1.20 for all lung cancer and 2.92 at ages less than 50 years. For histological types, the ratio was 3.44, 1.43, 3.20 and 2.55 for adenocarcinoma, squamous cell carcinoma (SCC in Figure 6.4), small cell and large cell carcinoma, respectively, at ages less than 50 years. The early onset familial component accounted for 29.4% of familial adenocarcinoma and 33.3% of familial small and large cell carcinoma. The proportion was smallest for squamous cell carcinoma (13.3%).

The higher risk among siblings than among offspring of affected parents before the age of 50 years calls for an explanation, e.g. that it could be associated with sharing of environmental effects (such as shared smoking habits or environmental smoke). However, the shared environmental effects have not explained more than about 10% of concordance for lung cancer in twin and family studies. These data were generated without information on smoking habits; however, had shared smoking habits been important, the concordance would surely have been higher. In fact, the data suggest that smoking is aggregated in families by a pattern dominated by the father rather than by the sibling. A second alternative explanation would be X-linked inheritance. If such a disease was X-linked, it would affect mainly men. For
lung cancer, a large X-linked effect is unlikely because a recent study from this
database has shown that there is no gender effect in familial lung cancer\textsuperscript{71}. A further
alternative explanation would be a major change in smoking habits over time. Apparent recessive effects could be mimicked by a large increase in the number of smokers in a short time period\textsuperscript{121}. However, such changes are unlikely to have taken place in Sweden\textsuperscript{179}. The data on other cancers in siblings with lung cancer suggest that the risks are related to smoking by interaction with the susceptible genotype. Our results support previous evidence of Mendelian inheritance in early-onset lung cancer\textsuperscript{134, 173-176}.

![Figure 6.4](image)

**Figure 6.4** Risk of lung cancer in offspring of parental and sibling probands

### 6.5 Multiple primary lung cancers (Papers III and V)

Lung cancer patients have a poor survival and it is not meaningful to follow second cancers during this short survival period; instead, we reported data on second lung cancers following first primary cancers of at least moderate survival\textsuperscript{30}. In paper III, we analyzed 2,569 cases of second primary lung cancer in men (father and son) and 1,282 in women (mother and daughter) with any primary cancer. The risk of lung cancer was over 10 after cancer of the small intestine in men and after anal cancer in
women during the first year. Lung cancer was even increased after upper aerodigestive tract, breast, cervical, kidney, urinary bladder and squamous cell skin cancer, non-Hodgkin’s lymphoma, Hodgkin’s disease and leukemia, through all follow-up periods in men and women.

The first systematic analysis of familial multiple lung cancer was analyzed in paper V using the Swedish Family-Cancer Database. A total of 267 patients with first and second primary lung cancers were included, the mean follow-up time being 3.7 years (ranging from less than 1 year to 31 years). The incidence of second primary lung cancer was 9-fold compared to that of first primary lung cancer. Compared with all individuals, patients with a family history were at a significantly increased risk for subsequent primary lung cancer, both among men (SIR = 9.89, 95%CI 4.48–18.66) and women (SIR = 17.86, 95%CI 5.63–42.00). The corresponding SIRs in patients without a family history were 2.04 (95%CI 1.75–2.36) and 5.10 (95%CI 3.99–6.43) for men and women, respectively.

The difference in familial risk of lung cancer between genders was also shown in paper V: the incidence of second lung cancer was approximately three times higher for men than for women. Even the SIRs for first and multiple primary lung cancers were higher for women than for men. A higher familial risk was also shown for women when probands had multiple primary lung cancer, but the number of cases was small.

The increased risk of lung cancer in men and women with skin cancer and non-Hodgkin’s lymphoma could result from a depressed immune function since there is a large excess of these malignancies in immunosuppressed patients. Studies on patients with Hodgkin’s disease, a relatively early-onset malignancy, who received radiotherapy and chemotherapy have shown a significant risk of lung cancer. Tobacco smoking is a shared factor that could explain the association of esophageal and urinary bladder cancers with lung cancers. The study also suggests that an inherited susceptibility predisposes a small proportion of lung cancer patients to second primary lung cancer, and vice versa, and that multiple primary lung cancers are associated with a familial risk.
7. CONCLUSIONS

The objective of this thesis was to perform a quality analysis of the Swedish Family-Cancer Database. This Database has unique attributes for family studies. All cancers are registered and medically verified, and family relationships can be reconstructed. The power of the Database is in its large numbers and in not being sensitive to selection or reporting biases.

Systematic comparison of all cancer risks was carried out, based on different probands. Parental probands are informative of dominant effects in offspring, whereas analysis of risks between siblings who lack affected parents provides clues about possible recessive effects. From the results, we have suggested that familial clustering for all cancers at most sites is heritable, and caused by dominant effects.

The overall age-adjusted incidence rates showed different incidence patterns for different histological types of lung cancer. The change of histological types of lung cancer can be explained in part by differences in smoking patterns, including changes in the prevalence of smoking and the use of low-tar and filter cigarettes. Modern cigarettes induced lung adenocarcinoma and squamous cell carcinoma in the proportion 1:0.6, and there was equal sensitivity in both genders to tobacco-induced lung cancer.

Risk of second primary lung cancers was increased in men and women after smoking and lifestyle-related sites, and after skin cancer, non-Hodgkin’s lymphoma and Hodgkin’s disease. Patients with multiple lung cancer were more likely to have a family history of lung cancer, and vice versa. This strongly suggests the presence of some inherited susceptibility.

Familial risk of lung cancer was found to be highest for adenocarcinoma and large cell carcinoma, and there was only a slightly lower risk for squamous cell and small cell carcinoma.

Lung cancer was associated with parental rectal, cervical, kidney, urinary bladder and endocrine gland cancer. The population-attributable fraction of familial lung cancer
was 2.97%. At a young age, risks between siblings were higher than those between offspring and parents. Up to the age of 68 years, approximately 1.7% of lung cancers were heritable and probably due to a high-penetrant recessive gene or genes that predispose to tobacco carcinogens.
8. ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all colleagues and friends who supported and helped me to complete this study. Specifically, I wish to thank:

Kari Hemminki, my supervisor, for giving me the possibility of working in a creative and highly stimulating research environment, for his excellent supervision, for sharing his broad scientific knowledge and for support and encouragement throughout the study.

Kamila Czene, co-author in many papers, for the great time we have had working together, for always offering help without even the slightest delay, and for her skilful assistance in writing.

Charlotta Granström, Pauli Vaittinen, Chuanhui Dong, Perti Mmutanen, Yongwen Jiang, Hong Zhang and Jianguang Ji, for being nice and helpful colleagues in statistics, and friends.

Dan Segerbäck, Asta Försti and Rajiv Kumar, for always contributing to a pleasant atmosphere and for helping with the sending of proofs.

Krystyna Håkansson, for the excellent help in computer techniques and for a great and stimulating time working together.

Guogang Xu, my classmates, we met each other since 1985, and then studied and worked together in Shanghai Medical University, China and in Karolinska Institute, Sweden.

I would like to thank all former and present members of the group: Gabriela Bacova, Natalia Kotova, Malgorzata Strozyk, Xin Ma, Chunyan Zhao, Sabrina Angelini, Petra Berggren, Johanna Smeds, Heli Rajaniemi, Erika Nogradi, Somali Sanyal, Sanna Tiikkaja, Mikko Koskinen, Frank le Curieux, Huaiyu Zheng and Zhengzhong Zhang.

Kristina Bergholm, Ylva Svanberg and the Department of Management at Biosciences, for their continuous help through my studies and work at Novum.

Jie Chen, Saimei Yu and Hengjin Dong, my tutors of Master of Public Health in Shanghai Medical University, for their continuous encouragement and parent-like love in my life. We had a very happy meeting in Stockholm at Dong’s Ph.D. party when Chen was working at WHO.

Tong Zhou, my classmates, we met each other since 1985, and then studied and worked together in Shanghai Medical University, China and in Umeå University, Sweden.

Taiyi Jin and his family, for much help and concern both in China and in Umeå, Sweden.

Birgitta Åström, for her hearty well-wishing during my stay at Umeå University.

To STINT (The Swedish Foundation for International Cooperation in Research and Higher Education) and the Department of Public Health and Epidemiology, Umeå University, for their support of my Master of Public Health studies, and for affording me the chance to watch the Nobel Awards.

My Chinese friends, Yaofeng Zhao, Peipei Zhao, Yu Shi, Zhong He, Sicheng Wen, Jing Ding, Qinyang Wu, Wengpeng Niu, Yan Li, Weihua Zhang, Li Xu, Xiaolei Zhou, Pu
Zhang, Kejun Li, Shujing Zhang, Xiaofeng Zhang, Li Xing, Jinjin Pei and his family, Qing Lu, Xiaohong Yu, Zhiwen Liu, Shushun Li, Wei Zhou, Zhihao Zhao, Qi Chen and her family, Hong Fang and her family, Weili Zhan and her family and Chaoyong Zhu for social support and enthusiastic encouragement.

Thanks also to little gentlemen Kevin (Weichu Zhao), Haoran (Haoran Wen) and Robin (Ziyue Zhou) for sharing their happiness with Daniel.

Zhihong Bao, Guifang Sun, Guiqin Wu, Tai Yao, and Yumei Wen and colleagues at the Department of Research and Development, Shanghai Medical University, for many years enjoying in medical scientific research and development management.

My parents and my family members in China for constant love. My parents-in-law and my sister-in-law for support and help, especially for their taking care of Daniel during our period of inexperience.

And last but not least, my wife Qianren Jin, my life companion, for her endless love, tolerance, happiness, kindness; and for giving me a wonderful family and life.

My son, Daniel for filling my life with beauty and joy every day.

My greatest thanks to you all!

This study was supported by grants from the Swedish Cancer Society, the Cancer Fund and the King Gustaf V's Jubilee Fund.
9. REFERENCES

32. Dong C, Hemminki K. Multiple primary cancers of the colon, breast and skin (melanoma) as models for polygenic cancers. *Int J Cancer* 2001;92:883-887.


