

From DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND  
BIostatISTICS

Karolinska Institutet, Stockholm, Sweden

# EFFECTS OF KINSHIP AND TIMING IN FAMILIAL CANCER

Myeongjee Lee



**Karolinska  
Institutet**

Stockholm 2015

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Myeongjee Lee, 2015

ISBN 978-91-7549-862-1

Printed by E-Print AB 2015

# Effects of Kinship and Timing in Familial Cancer

## THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Myeongjee Lee**

*Principal Supervisor:*

Prof. Marie Reilly  
Karolinska Institutet  
Department of Medical Epidemiology and  
Biostatistics

*Co-supervisor(s):*

Prof. Kamila Czene  
Karolinska Institutet  
Department of Medical Epidemiology and  
Biostatistics

Dr. Paola Rebora  
University of Milano-Bicocca  
Department of Health Sciences

*Opponent:*

Assoc. Prof. Karin Ekström Smedby  
Karolinska Universitetssjukhuset  
Department of Medicine

*Examination Board:*

Assoc. Prof. Staffan Nilsson  
Chalmers University of Technology  
Department of Mathematical Sciences  
Division of Mathematical Statistics

Prof. Juni Palmgren  
Karolinska Institutet  
Department of Medical Epidemiology and  
Biostatistics

Prof. Leif Bergkvist  
Uppsala University  
Centre for Clinical Research, Västerås



To my dearly beloved family

사랑하는 나의 가족에게



## ABSTRACT

Family cancer history is one of the most important risk factors for cancer. We investigated in depth the effect of family cancer history, in particular, the effects of kinship and timing on cancer incidence and prognosis using Swedish population-based registers.

In **Study I**, we investigated how the type of kinship and sex affect the familial risk of adult chronic lymphatic leukemia (CLL). We found the highest relative risks for sisters of affected women and sons had a much higher risk than daughters if the affected parent was the mother.

In **Study II**, we developed a unified model for familial risk by extending a Cox regression model to estimate the detailed effects of all specific kinships, using data from all family members simultaneously. This enabled a formal comparison of the risk to different relatives. We illustrated the method with applications to adult leukemia and non-Hodgkin's lymphoma and found sisters of female patients had significantly higher risk for both cancers than other relatives.

We next investigated how the risk pattern of four main cancers in Sweden (colorectal cancer, breast cancer, prostate cancer and melanoma) changes with age and elapsed time from diagnosis of the same cancer in a sibling. Results were presented graphically in **Study III**. For all four cancers, siblings of cancer patients had higher cancer incidence at all ages compared to siblings of cancer-free individuals. Relative risks were especially high in siblings who were young when the first cancer was diagnosed in the family. The relative risks were relatively constant up to 20 years after the cancer diagnosis in siblings for all cancers except prostate cancer, where the hazard ratio decreased steeply during the first few years. We found evidence that this may be due to a screening effect for prostate cancer while there was no evidence of a screening effect in breast cancer.

In **Study IV**, we examined how family cancer history affects prognosis for patients with several major cancers in Sweden and we further investigated for cancers whose prognosis has been affected by family cancer history whether these effects are associated with histological type or subtype or tumor stage at cancer diagnosis. For breast and prostate cancer patients, family cancer history played a protective role in cancer survival, and this may be associated with medical surveillance of family members. However, daughters or sisters of ovarian cancer patients had poorer cancer survival, which is consistent with a higher proportion of diagnosed later stage tumors and of an aggressive histological type of ovarian cancers.

In conclusion, our investigation in this thesis used Swedish population-based registers to highlight the effect of family cancer history on several outcomes, including cancer risks for family members and cancer patient prognosis. Our findings provide evidence that could help tailor screening programs for relatives of cancer patients and inform the design of genetic biomarker studies.

## LIST OF SCIENTIFIC PAPERS

- I. Rebora P, Lee M, Czene K, Valsecchi MG, Reilly M. **High risks of familial chronic lymphatic leukemia for specific relatives: signposts for genetic discovery?** *Leukemia* 2012; **26**: 2419-2421.
- II. Lee M, Rebora P, Valsecchi MG, Czene K, Reilly M. **A unified model for estimating and testing familial aggregation.** *Statist. Med.* 2013; **32**: 5353-5365.
- III. Lee M, Czene K, Rebora P, Reilly M. **Patterns of changing cancer risks with time since diagnosis of a sibling.** *Int. J. Cancer* 2015; **136(8)**: 1948-56.
- IV. Lee M, Reilly M, Lindström LS, Czene K. **Differences in survival for patients with familial and sporadic cancer.** *Manuscript*.

## RELATED PUBLICATION

(Not included in thesis)

Lindström LS, Li J, Lee M, Einbeigi Z, Hartman M, Hall P, Czene K. **Prognostic information of a previously diagnosed sister is an independent prognosticator for a newly diagnosed sister with breast cancer.** *Ann Oncol.* 2014; **25(10)**: 1966-72.



# CONTENTS

1	INTRODUCTION.....	5
2	BACKGROUND.....	6
2.1	Cancer Risk and Family Cancer History .....	8
2.2	Cancer Survival and Family Cancer History .....	9
3	AIMS .....	11
4	DATA SOURCES.....	12
4.1	Swedish Multi-Generation Register.....	12
4.2	Cancer Register.....	13
4.3	Cause of Death Register .....	13
4.4	Other Registers .....	14
5	STUDY POPULATION AND STUDY DESIGN .....	15
5.1	Study I-Study III.....	15
5.1.1	Family Cluster .....	16
5.2	Study IV .....	17
5.3	Consideration of Bias .....	19
6	STATISTICAL METHODS.....	20
6.1	Survival analysis.....	20
6.1.1	Cox regression model.....	20
6.1.2	A unified model (main aim of Study II) .....	21
6.1.3	Poisson regression .....	24
6.1.4	Flexible parametric model .....	24
6.2	Multinomial logistic regression .....	25
7	RESULTS AND INTERPRETATION .....	26
7.1	Study I : High risks of familial CLL for specific relatives: signposts for genetic discovery? .....	26
7.1.1	Results .....	26
7.1.2	Interpretation .....	26
7.2	Study II : A unified model for estimating and testing familial aggregation .....	27
7.2.1	Results .....	27
7.2.2	Interpretation .....	28
7.3	Study III : Patterns of changing cancer risks with time since diagnosis of a sibling.....	29
7.3.1	Results .....	29
7.3.2	Interpretation .....	32
7.4	Study IV : Differences in survival for patients with familial and sporadic cancer.....	33
7.4.1	Results .....	33
7.4.2	Interpretation .....	34
8	CONCLUSIONS AND FUTURE PERSPECTIVES .....	37
9	ACKNOWLEDGEMENTS.....	39
	REFERENCES .....	41

## LIST OF ABBREVIATIONS

CLL	Chronic Lymphatic Leukemia
FAP	Familial Adenomatous Polyposis
FIGO	Fédération Internationale de Gynécologie Obstétrique (International Federation of Gynecology and Obstetrics)
HNPCC	Hereditary Nonpolyposis Colorectal Cancer
HR	Hazard Ratio
ICD	International Classification of Disease
IRR	Incidence Rate Ratio
LRT	Likelihood Ratio Test
NHL	Non-Hodgkin's Lymphoma
OR	Odds Ratio
PIN	Personal Identity Number
PSA	Prostate-Specific Antigen
SES	Socio-Economic Status
SIR	Standardized Incidence Ratio
TNM	Tumor, Node, and Metastasis; a system of cancer staging
95% C.I.	95% Confidence Interval

# 1 INTRODUCTION

Aldred Scott Warthin, MD, PhD (1866-1931) provided the first documentation of the longest family cancer histories in 1913 [1] and again in 1925 [2]. This type of family history was updated by Weller and Hauser [3] in 1936, and Lynch [4] in 1971 and was subsequently described as a Lynch syndrome family [5]. Lynch syndrome is now one of the best-known inherited cancer syndromes among more than 200 hereditary cancer susceptibility syndromes [6].

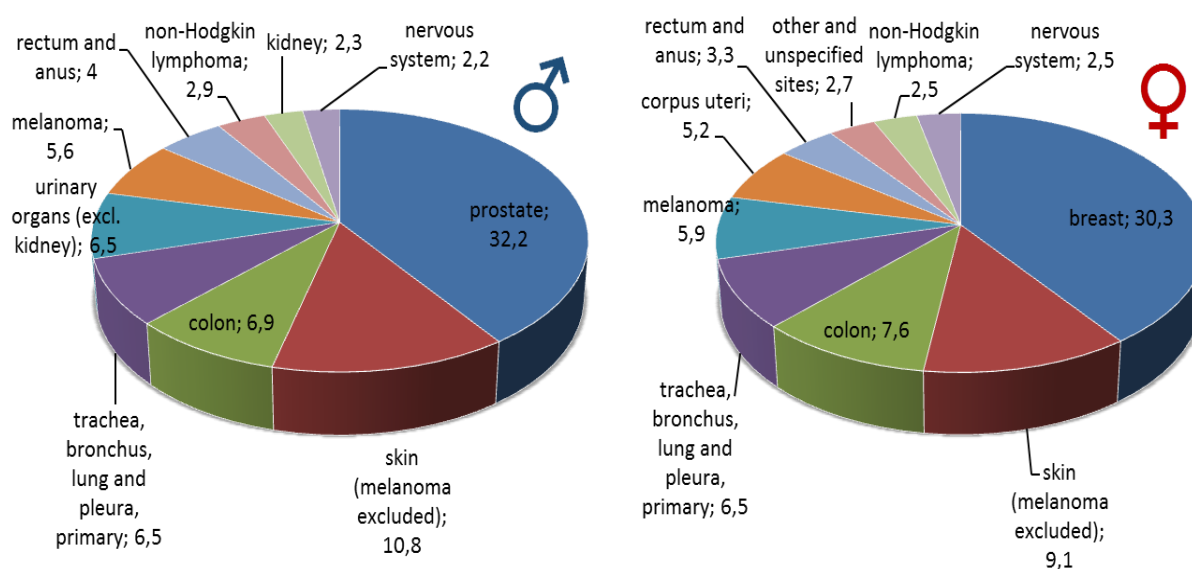
Today, where information on family cancer history is available from population based registers [7-11], inherited cancers can be more readily identified unlike Lynch syndrome which took more than 50 years. A number of studies have been carried out on the role of family cancer history, and a positive family history of cancer is considered as a surrogate for genetic susceptibility in many epidemiological studies.

We performed in-depth analyses to understand the role of family cancer history in cancer risk and cancer survival, especially the effect of kinship and timing, using the Swedish population-based registers. The first two studies aimed to investigate relationship-specific familial aggregation (**Study I**) and developed a statistical model for estimation and comparison of relationship-specific familial risks (**Study II**). We also examined the pattern of familial risk changes over time, age, and elapsed time since diagnosis (**Study III**). In **Study IV**, we focused on the association between cancer survival in cancer patients and family cancer history.

## 2 BACKGROUND

Cancer is the second most common cause of death in Sweden followed by diseases of the circulatory system [12]. Hence, understanding of cancer from various angles is crucial to improve health and to reduce preventable deaths.

**Figure 1** shows the percentage distribution of the ten most frequent cancer sites by gender in Sweden. The most common cancers are prostate cancer for males and breast cancer for females, representing 32.2 per cent and 30.3 per cent of all cancers for males and females, respectively [13]. Colorectal cancer, skin cancer (melanoma excluded), and cancer of the respiratory tract are the second, third, and fourth most common cancers in both males and females. Malignant melanoma increased the most rapidly over the last decade, 5.2 per cent per year for males and 5.3 per cent per year for females, compared to other malignant tumors [13].

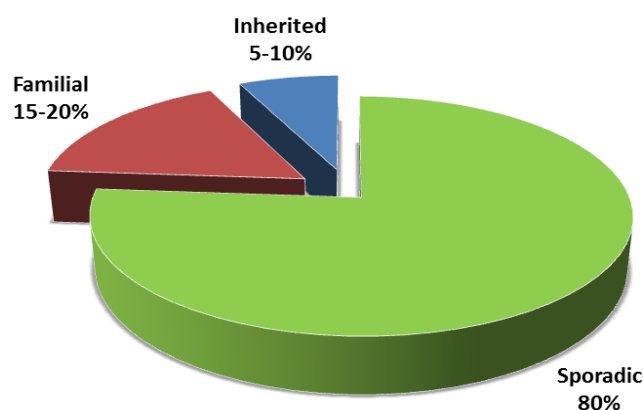


**Figure 1** The percentage distribution of the ten most frequent cancer sites, by gender. Modified from [13].

Cancers begin in gene mutations within cells. These mutations are caused by both genetic factors (such as inherited mutations and hormones) and environmental factors (such as tobacco, alcohol, diet, radiation, and infectious organisms) (see **Table 1**). Most cancers are sporadic, developed by chance or from environmental or lifestyle factors that are not passed from parent to child (see **Figure 2**) [6, 14]. Cancers with certain highly penetrant germline mutations are called inherited cancers, and these make up 5 to 10 per cent of most common cancers. So far, more than 200 inherited cancer syndromes have been investigated [6], and examples of common inherited family cancers are listed in **Table 1** [15]. A further 10-15 per

**Table 1** Selection of known mutations and their associated with cancers.

Mutations	Genes/ Risk factors	Relevant Cancer sites	Refs.	Identified cancer syndrome
<b>Germline mutations in</b>	BRCA1, BRCA2	Breast, Ovarian, Prostate, etc.	[16]	Hereditary breast and ovarian cancer syndrome
	p53, CHEK2	Breast, Leukemia, Brain, etc.	[16]	Li-Fraumeni Syndrome
	MLH1, MSH2,MSH6	Colon, Endometrial, Ovarian, etc.	[16]	Hereditary Nonpolyposis Colorectal cancer (HNPCC), Lynch syndrome
	APC	Colon	[16]	Familial Adenomatous Polyposis (FAP)
<b>Somatic mutations by</b>	CDKN2	Melanoma	[16]	Melanoma syndrome
	Alcohol	Breast, liver, oesophagus, oropharynx, larynx, etc.	[17]	
	Smoking	Lung, bladder, stomach, , oesophagus, oropharynx, larynx, etc.	[17]	
	Diet	Prostate, colorectal, breast, stomach, lung, etc.	[18]	
	Obesity	Colorectal, stomach, ovarian, breast, non-Hodgkin's lymphoma, etc.	[19]	
	Infection	Stomach, lymphoma, etc.	[20]	
	Environmental Carcinogens	Colorectal, leukemia, bladder, lung, stomach, etc.	[21]	



**Figure 2** Main type of cancer. Reproduced with permission from [6].

cent of cancers are referred to as “familial” and may be caused by multiple factors, such as chance, exposure to the same environment, inheritance of low-penetrance genes, and gene–environment interactions.

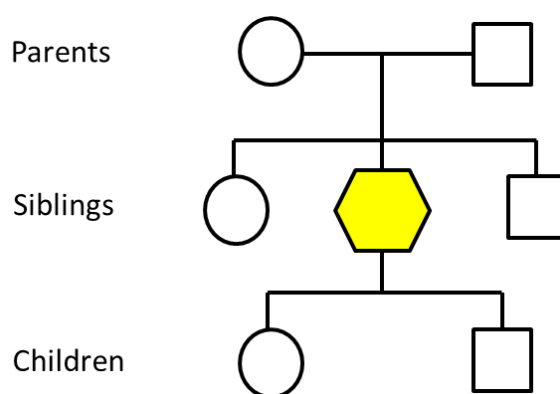
As cancers come from gene mutations, it is important to understand the mechanism by which this occurs and there are several approaches [8, 22-25]. The traditional method is clinical detection, in which probands and their affected relatives are identified [22, 26]. Use of twin data is another possible way to study the overall contribution of hereditary factors [23]. However, these two methods have similar difficulties, such as lack of sufficient numbers of cases and the possibility of selection bias [27]. An alternative way is comparing the observed and expected cases in families of patients or comparing the risk in those with a family cancer history to the risk in those with no family cancer history, in region-wide or population-based registers of cancer patients and family relations [8, 11, 24]. This method can have large

numbers at its disposal, but genetic and environmental factors are hard to separate, and it is also essential to have good infrastructure for the collection of high quality data to avoid biased estimates [28-35].

Although newly developed technologies for identifying genes are now available [36, 37], still family based studies are required prior to genotyping all individuals to improve efficiency. Susceptibility to an inherited cancer in a family may be suspected as a result of cancers in close relatives, the same type of cancer on the same side of the family, young age of affected cancer patients, or the occurrence of genetically related cancers in one family.

## 2.1 CANCER RISK AND FAMILY CANCER HISTORY

A family history of cancer, especially in first degree relatives (**Figure 3**), is one of the most important risk factors of cancer occurrence.



**Figure 3** All first degree relatives of a proband (yellow colored). Square and circle denote male and female, respectively.

**Table 2** shows common cancers associated with cancer history in a family in Sweden with familial aggregation (as SIRs) stratified by affected type of relatives [38]. Overall, cancer risks in patients with affected siblings are higher than in patients with affected parents and it increased much higher if there are more affected first-degree relatives. Taking ovarian cancer as an example, women with both mother and sister affected are more than 30 times more likely to have ovarian cancer.

Increased cancer risk among family members, especially first degree relatives, is likely due to these patients sharing more genes than the population at large, but family members also share habits and lifestyle to a greater degree than the general population. The increased cancer risk in relatives of patients can be affected by several, such as the degree of relationship with the affected relative [39, 40], kinship [38, 40], number of affected relatives [39-41], sex of relatives and sex of probands [40, 41]. The contribution of environmental factors to familial cancer risk can be analyzed of using spouses who shared habits and life style but not genetic

**Table 2** Standard Incidence ratios (SIRs) with 95% confidence interval (95% C.I.) for cancer in patients by parental/sibling/both concordant cancer. Reproduced with permission from [38].

Cancer site	Parents affected		Sibling affected		Both affected <sup>b</sup>	
	SIR	95% C.I.	SIR	95% C.I.	SIR	95% C.I.
Stomach	2.17	(1.73-2.69)	3.29	(1.41-6.52)	12.66	(1.19-46.55)
Colorectal <sup>a</sup>	1.86	(1.73-2.01)	2.87	(2.40-3.40)	5.38	(3.48-7.95)
Lung	2.09	(1.88-2.32)	3.13	(2.53-3.84)	5.06	(2.16-10.02)
Breast	1.84	(1.76-1.93)	2.03	(1.89-2.17)	3.27	(2.66-3.99)
Ovary	3.15	(2.56-3.85)	4.25	(3.01-5.84)	31.64	(11.39-69.33)
Prostate	2.45	(2.30-2.62)	4.46	(3.85-5.15)	8.62	(6.49-11.23)
Kidney	1.87	(1.44-2.38)	4.74	(3.09-6.95)	-	
Urinary bladder	1.75	(1.45-2.10)	2.02	(1.25-3.10)	8.16	(1.54-24.16)
Melanoma	2.62	(2.23-3.05)	2.93	(2.38-3.57)	8.95	(3.55-18.54)
Nervous system	1.71	(1.41-2.06)	1.81	(1.32-2.43)	6.55	(1.23-19.39)
Non-Hodgkin's lymphoma	1.82	(1.43-2.28)	2.25	(1.44-3.35)	-	
Leukemia	1.85	(1.39-2.41)	3.02	(1.79-4.78)	-	

a. Adenocarcinoma only.

b. Only familial triplets were diagnosed for any of the cancers listed.

variants. However, cancer susceptibility among spouses because of shared lifestyle is small [42, 43]. Environmental factors may also be manifested by cancer risks associated with family size or birth order [44]. For example, family size has been studied as a risk factor for melanoma because of exposure to sun during holidays spent together [45-47] and for stomach cancer because of increased risk of infection with *Helicobacter pylori* in larger families [48-50]. Lower birth weight and higher parental age are also associated with higher birth order and have been identified as risk factors for cancers such as breast cancer, melanoma, and testicular cancer [48, 51-55]. In general, lifestyles are likely to differ more between parents and offspring than between spouses and familial cancers in parents and offspring or in siblings are more likely to be due to heritable rather than environmental effects [42, 43].

As cancer is a common disease at older ages, cancer occurrence at a young age is more likely to be associated with inheritance [56-58]. Thus, age is an important factor in cancer risk, and both the age at diagnosis of a cancer patient and the age of their relatives need to be considered in the study of familial cancer risk [59]. In addition, the profile of cancer risk with elapsed time since the first diagnosis in a family is important with regard to improved counseling and optimized screening of family members of cancer patients [60].

Familial cancer risk is a measure of the clustering of cancer in family members, thereby giving the first indication of the possible involvement of heritable genes [61-63] and since it is the basis for clinical decisions, counseling and genetic testing for cancer risk [38, 64], it is deserving in-depth study.

## 2.2 CANCER SURVIVAL AND FAMILY CANCER HISTORY

The concordance of cancer-specific survival in family members has been studied [65-67] for breast cancer, prostate cancer, lung cancer, and melanoma. Children with poor parental

survival have an increased risk of poor survival compared with children with good parental survival [65, 67]. These findings might be also described by reference to genetic factors and environmental factors. Family members are more likely to have similar genetic variants associated with cancer [68-73], tumor characteristics such as tumor stage, tumor grade, and histological type [70-72], and response to systematic treatment [74-76]. Also, if there is at least one affected family member, this may affect remaining family members in terms of seeking medical surveillance, choice of treatment, and lifestyle after the cancer diagnosis [77, 78]. However, there are few systematic studies about differences in survival for patients with a family history of cancer as compared to patients without such a family history [79-82].



### **3 AIMS**

This thesis aims to provide a greater understanding of cancer etiology and prognosis in families affected by cancer, especially with regard to kinship and timing, in order to provide evidence for genetic testing, counselling and screening protocols for cancer patients and their relatives. The specific aim of each study in this thesis is described in the following.

#### **Study I:**

To examine how the type of kinship and sex affect the familial risk of adult chronic lymphatic leukemia.

#### **Study II:**

To develop a unified model for familial risk by extending a Cox regression model to jointly estimate and compare the effects of different kinships, using adult leukemia and non-Hodgkin's lymphoma as examples.

#### **Study III:**

To compare the pattern of changing risk of four main cancers in Sweden (colorectal cancer, breast cancer, prostate cancer and melanoma) with time from the index diagnosis of cancer in a sibling.

#### **Study IV:**

To investigate whether family cancer history affects the prognosis in patients diagnosed with twelve different cancers.

## 4 DATA SOURCES

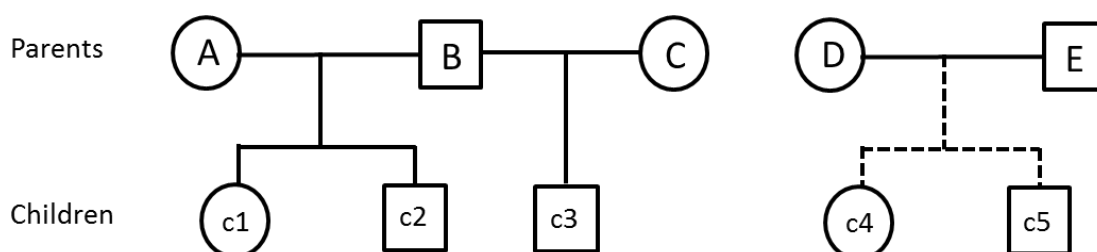
Sweden has established high-quality population-based registers, which facilitate large epidemiological studies in a wide range of fields.

### 4.1 SWEDISH MULTI-GENERATION REGISTER

This register contains all children who were born in 1932 or after and were still alive in 1961 [83]. With the introduction of a personal identity number (PIN) in Sweden in 1947, personal records were established for all persons who were registered in a parish registry [84]. Then, children who were 15 years old or younger in 1947 were recorded along with information on their biological/adoptive parents. The first computerized census, on which the Multi-Generation Register was based, was conducted in 1961. In 1991, the national registration at local level moved from the parish office to the local tax office and parental information is essentially complete for children who were alive in 1991, thanks to computerized registration.

Children are referred to as indexes, and can be entered only once, whereas parents can appear in the register as many times as their number of children. Parents need not be married to be registered. Using this register, which contains all children–parents pairs, we can also define all other family relationships, of which some examples we used in this thesis are as follows.

- Siblings are defined as full siblings when they share both biological father and mother. ((c1, c2) in **Figure 4**)
- Siblings are defined as half siblings when they share only one biological parent. This type of sibling can also be divided into paternal half siblings or maternal half siblings, according to shared parents. ((c1, c3) and (c2, c3) in **Figure 4**)
- Siblings are defined as adoptive siblings when they share the same non-biological father and non-biological mother. ((c4, c5) in **Figure 4**)



**Figure 4** An example of two families. In a relationship between A and B, two children (c1, c2) were born. In another relationship between B and C, a child was born (c3). D and E adopted two children (c4, c5) with dashed line denoting adoption. Square and circle denote male and female, respectively.

- Spouse is defined as a partner with whom an individual shares at least one biological child. ((A, B) and (B, C) in **Figure 4**)

## 4.2 CANCER REGISTER

The Swedish Cancer Registry was established in 1958 and covers the whole population in Sweden [85]. It is obligatory for Swedish health care providers (physicians, pathologists and cytologists) to report to the registry newly detected cancer cases at clinical, morphological, and other laboratory examinations and at autopsies. Therefore, all tumors in an individual are recorded independently with personal information (PIN, sex, age, place of residence) and medical information (site of tumor, histological type, stage, date of diagnosis, etc.). Although tumors have been coded in several versions of the International Classification of Disease (ICD) at other times, the codes have been available as ICD-7 codes since 1958. This is particularly important for a cancer such as leukemia, whose classification is difficult and subject to change [86, 87]. Histological type has been coded in ICD-O/2 during 1993-2004 and in ICD-O/3 from 2005. However, the old histology code according to WHO/HS/CANC/24.1 is also available for the whole period since 1958. Stage information has been collected since 2004 for all cancer sites except brain, cranial nerves, lymphoma, and leukemia. Gynecological tumors are coded according to FIGO (staging scheme developed by the International Federation of Gynecology and Obstetrics, <http://www.figo.org>) and the rest are coded according to TNM (6<sup>th</sup> edition, classification system developed by the American Joint Committee on Cancer) [88].

Almost 98 per cent of the cases are morphologically verified in the Cancer Register [89]. The overall completeness of the Cancer Register is considered high and comparable to other high-quality registers in northern Europe [90, 91].

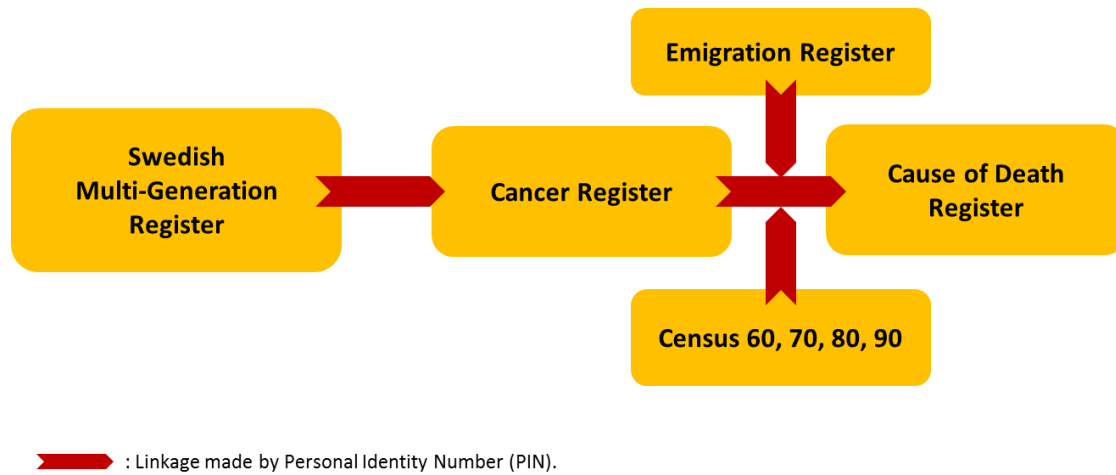
Cancers studied in this thesis are malignant tumors that were diagnosed as the first primary cancer in an individual.

## 4.3 CAUSE OF DEATH REGISTER

The Cause of Death Register was established in 1961 and is updated annually [92]. This register covers all deceased persons among Swedish residents, whether the deceased was a Swedish citizen or not and whether the death occurred in Sweden or not, and the coverage has been near to 100 per cent since 1997. The main information in the register is PIN, date of death, and underlying cause of death, with several contributory causes of death. The cause of death is recorded according to the international version of the ICD codes. Swedish death certificates have been examined with respect to the reporting of malignant neoplasm as the underlying cause of death and found to be generally reliable [93-95]. We collected cancer-specific cause of death information from the Cause of Death Register in **Study IV**.

#### 4.4 OTHER REGISTERS

Emigration information was used for censoring and was collected from Emigration and Immigration Register [96]. Socioeconomic status [37] and county of residence for matching variables were assembled from Censuses which were conducted at 1960, 1970, 1980, and 1990 [97].

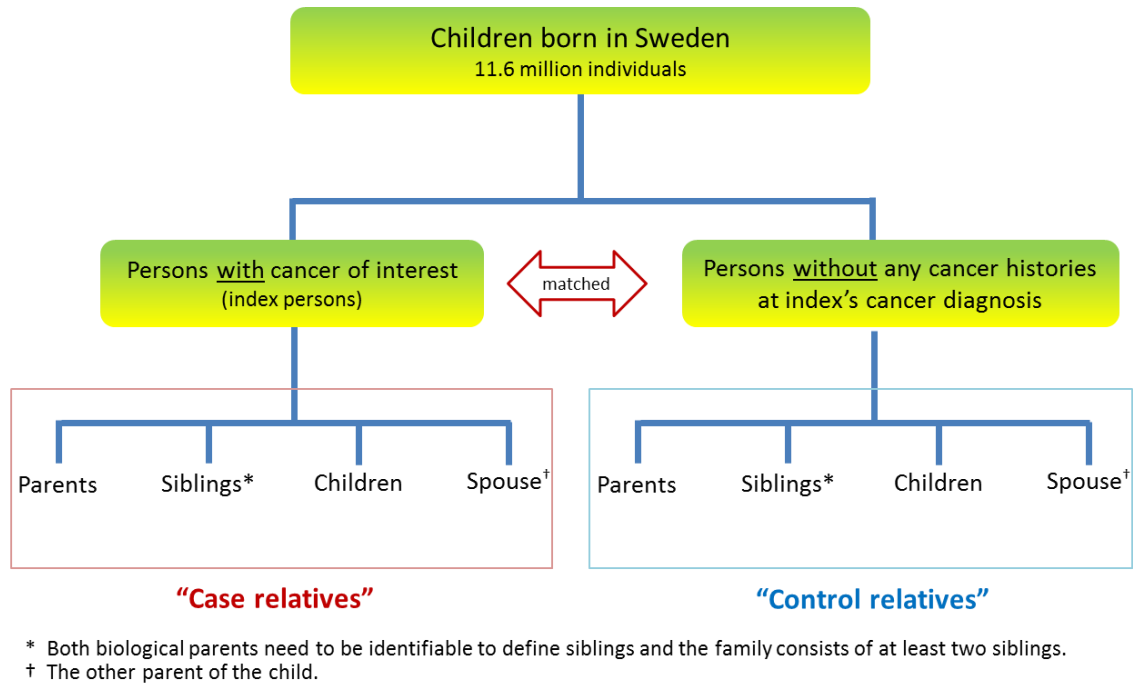


**Figure 5** Linkage of database.

From the linkage of these population-based registers, using personal identity number (**Figure 5**), we found 11.6 million children who were born in Sweden since 1932 and about 7 million children of whom both biological parents were identifiable.

## 5 STUDY POPULATION AND STUDY DESIGN

**Figure 6** shows how we defined the study population. We identified all cancer cases of interest from the linked registers and referred to these as the “index persons”.



**Figure 6** A scheme for data selection.

### 5.1 STUDY I-STUDY III

Familial clustering was measured by comparing the cancer risk in relatives of cancer patients to the cancer risk in relatives of healthy persons in this thesis. For each cancer patient, we randomly sampled at most five controls from the population using a nested matched design (see **Figure 6** and **Figure 7**). Controls should be free from the disease at the time of diagnosis of the corresponding case and were selected from that index's matched risk set, matching factors being sex, family size, year of birth, and county of residence, thus avoiding possible biases due to changing cancer incidence or differential reporting in different areas. We counted these cancer patients and their matched controls as case probands and control probands, respectively. Then we recruited relatives of these probands from the study population and defined them as case relatives (referred to as exposed relatives or affected relatives) and control relatives (referred to as unexposed relatives or unaffected relatives), respectively. Because we analyzed relatives of probands, probands with at least one relative contributed to the analysis.

In **Study I** and **Study II**, we identified patients with a cancer of interest who were diagnosed during 1958-2007 and studied (adult) chronic leukemia for **Study I** and (adult) leukemia and non-Hodgkin's lymphoma for **Study II**, respectively. Here, "adult" diseases meant that the disease was diagnosed at 15 or more than 15 years old [98]. We followed the case relatives and control relatives from birth or start of cancer registration, if that came later, until they were diagnosed with the concordant cancer or experienced other events, diagnosis of another malignant cancer, death, emigration or end of the study (31 December 2007), whichever came first.

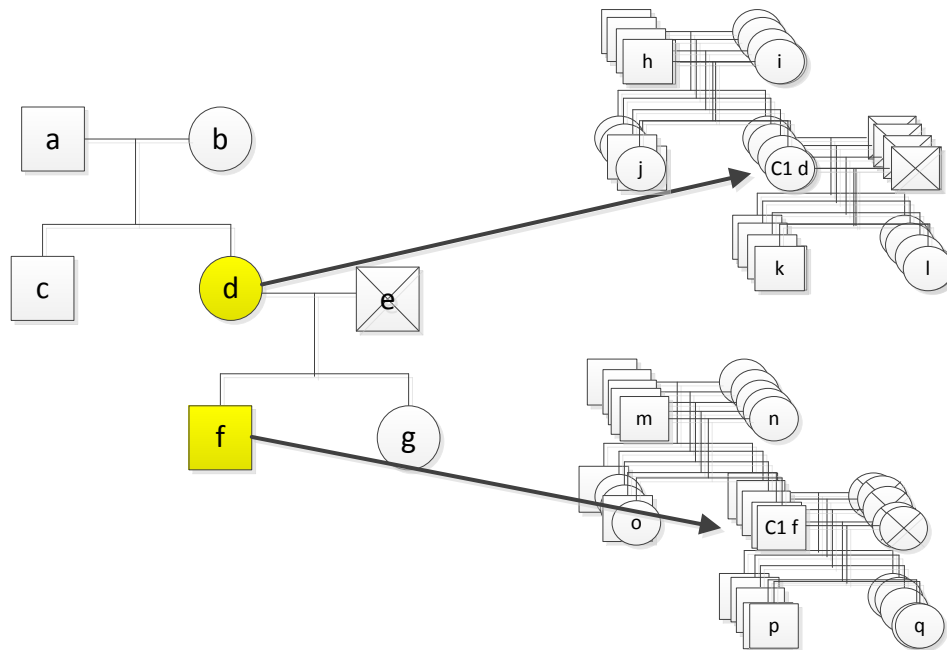
In **Study III**, we studied siblings of probands for four major cancers in Sweden, colorectal cancer, breast cancer, prostate cancer, and melanoma. Case probands were the patients who were first diagnosed with a cancer of interest as a primary in a sibship regardless of his/her age, during 1958-2009. Observation time began from the index's cancer diagnosis until a diagnosis of the same cancer and subjects did not contribute to the risk set at any time after the diagnosis of another cancer, emigration, death, or the end of the observation period (31 December 2009).

#### *Consideration of the Sampling method*

We compared cancer risks of the first degree relatives of cancer patients to the risks in first-degree relatives of selected controls (at most five per case) rather than using the whole population as reference. This sampling design has several advantages. First, minimal loss of information is expected, as all cancer cases were selected from the population-based registers and all the exposed family members were included. We randomly sampled control probands using a nested matched design, which is useful to accommodate censoring in the disease experience of the proband [99]. This design also enables us to study of elapsed time (in **Study III**), from the date of the index cancer diagnosis, whereas such time cannot be well defined for an individual in the general population. Thirdly, the computational burden of the analysis is much reduced compared to using the whole population as a reference group.

#### **5.1.1 Family Cluster**

If there are more than two index persons in a family, information on some individuals may appear several times in the analysis. Hence, some individuals may appear more than once and probably in different roles in the analysis data set. **Figure 7** presents an illustration of a matching cluster with two potential index persons ( $d$ ,  $f$ ). Each will have his or her own (matched) controls, ( $C1d$ ,  $C2d$ ,  $\dots$ ,  $C5d$ ) for  $d$  and ( $C1f$ ,  $C2f$ ,  $\dots$ ,  $C5f$ ) for  $f$ , respectively, when we take five controls for each case, and we include both sets of relatives in the data to be analyzed. Here, individual  $f$  will not only be one of the affected relatives of  $d$  but will also provide his own affected relatives  $d$ ,  $e$ ,  $g$ . In these correlated data, we defined a cluster which



**Figure 7** An example of a matching cluster with two potential probands (d,f, yellow colored) in a case family. Square and circle denote male and female, respectively.

comprises the combined set of all first-degree relatives of case probands who are in a family and all first-degree relatives of their matched control probands. Therefore, all the family members illustrated in **Figure 7** constitute a single cluster if  $d$  and  $f$  are both cases. To account for the dependence in the data, we estimated variance by two different methods in **Study II**; sandwich variance estimates regarding families as independent sampling units and the bootstrap approach [100] using the clusters defined by the relatives of the case proband and their matched controls [101]. The former approach can be directly implemented in standard statistical software but does not consider matching nor the possible replication of individuals. The latter approach accommodates the dependence in the data more accurately, and is thus more conservative.

## 5.2 STUDY IV

We investigated whether there is an association between family cancer history and cancer survival with respect to twelve cancer sites, namely stomach, colorectal, lung, breast, ovarian, prostate, kidney, bladder, melanoma, (adult) nervous system, non-Hodgkin's lymphoma, and (adult) leukemia. We identified all individuals who were diagnosed with a cancer of interest as a primary cancer in 1991-2009 (as index persons in **Figure 6**) and whose biological parents could be identified in the register. For the analysis of affected siblings, there must be at least two siblings in a family. After collecting information on cancer in parents or siblings, we divided cancer patients into two groups, familial and sporadic: familial cancer patients were those whose biological parents or siblings had the concordant cancer in the register and, all other patients were classified as sporadic. In contrast to **Study I - Study III** where we

followed relatives of cancer patients until the diagnosis of cancer, in this study we followed cancer patients for up to five years from the date of their primary cancer diagnosis and compared the survival of those with and without a family history of cancer. The choice of five years is common in cancer survivorship studies since most outcomes such as recurrence or cancer death occur within the first five years after the cancer diagnosis [102, 103]. Survival time was defined as the elapsed time from the date of cancer diagnosis until the date of cancer-specific death or the date of other censoring events (death, emigration or end of study on 31 December 2010), whichever came first.

**Table 3** presents the number of individuals and number of events (cancer incidence or cause-specific death) in each study and **Table 4** summarizes the follow-up and survival time ( $T$ ) for each study. Those individuals born before the start of cancer registration cannot be followed fully, so that our cohorts for **Study I** and **Study II** are left truncated by the start of the Cancer Register, which could yield biased estimates (see Section 5.3).

**Table 3** Numbers of study subjects for each study.

Study <sup>a</sup>	cancer	Studied subjects	case			control				
			probands	relatives	events	probands	relatives	events		
I	Adult CLL	1 <sup>st</sup> degree relatives	9 143	26 941	167	36 354	117 088	100		
II	Adult LEUK	1 <sup>st</sup> degree relatives	22 104	66 616	308	90 456	309 313	729		
	NHL		29 492	94 061	500	123 346	431 134	1397		
III	colorectal		15 590	30 808	414	75 243	151 982	866		
	Breast	[104]	24 362	35 418	1 681	109 674	175 750	3 607		
	Prostate	Siblings	14 446	20 953	1 589	63 486	103 722	2 251		
	melanoma		14 452	26 125	241	71 109	129 891	383		
	colorectal		2 348	3 874	32	11 214	13 861	866		
	Breast	half-	1 546	2 033	62	6 800	7 228	3 607		
	Prostate	Siblings	3 880	5 057	110	14 569	18 466	2 251		
	melanoma		2 593	4 186	15	12 830	15 725	383		
Study <sup>a</sup>	cancer	subjects	Parents were identifiable				At least two siblings in a family			
			familial		sporadic		familial		sporadic	
			patients	deaths	patients	deaths	patients	deaths	patients	deaths
IV	Stomach	Cancer patients, themselves	195	119	3 593	2 215	45	31	2 948	1 795
	Colorectal		2 563	778	21 365	6 760	861	272	17 468	5 427
	Lung		1 156	860	15 816	11 534	625	481	12 489	9 040
	Breast		4 974	403	46 299	3 969	3 062	223	37 676	3 240
	Ovarian		204	82	6 173	2 343	111	58	4 897	1 824
	Prostate		6 427	392	34 956	2 636	3 897	260	27 812	2 004
	Kidney		170	53	5 805	1 783	71	24	4 565	1 413
	Bladder		417	46	8 604	1 203	129	16	6 752	905
	Melanoma		583	28	16 346	1 191	402	34	13 160	925
	Nervous system		251	93	11 486	3 638	154	53	9 411	2 943
	NHL		205	42	8 820	1 827	88	25	7 000	1 428
	(adult) LEUK		206	40	6 753	1 723	84	17	5 385	1 352

a. study period is 1958-2007 for Study I and Study II, 1958-2009 for Study III, 1991-2009 for Study IV.



**Table 4** Definition of survival time.

Study	Survival time (T)		Event of interest
	Entry date, which comes later	Exit date, which comes earlier	
I, II	<ul style="list-style-type: none"> <li>• Birth</li> <li>• Start of cancer register</li> </ul>	<ul style="list-style-type: none"> <li>• The concordant Cancer diagnosis</li> <li>• Other cancer diagnosis</li> <li>• Emigration</li> <li>• Death</li> <li>• End of study (31 Dec, 2007)</li> </ul>	Concordant cancer
III	<ul style="list-style-type: none"> <li>• Index's cancer diagnosis</li> </ul>	<ul style="list-style-type: none"> <li>• The concordant Cancer diagnosis</li> <li>• Other cancer diagnosis</li> <li>• Emigration</li> <li>• Death</li> <li>• End of study (31 Dec, 2009)</li> </ul>	Concordant cancer
IV	<ul style="list-style-type: none"> <li>• Date of cancer diagnosis</li> </ul>	<ul style="list-style-type: none"> <li>• Cancer-specific death</li> <li>• Death by other causes</li> <li>• Emigration</li> <li>• End of study (31 Dec, 2010)</li> </ul>	Cancer-specific death within five years

### 5.3 CONSIDERATION OF BIAS

Although registers may cover the whole population in a country, we should consider biases due to incomplete ascertainment [105]. Our studies are based on the several high-quality Swedish population-based registers which started in different calendar years: 1932 for the Multi-Generation Register (complete linkage of parental information was made in 1991), 1958 for the Cancer Register, and 1961 for the Cause of Death Register. Left truncation at the start of the Cancer Register might result in ascertainment bias in familial risk estimates, especially in parents from the earlier birth cohort. Almost all cancer diagnoses of index persons and of their siblings are captured in the Cancer Register, except the small number of cases occurring before age 26 years, and this complete ascertainment will yield unbiased estimates of sibling relative risk [106]. By contrast, in the parents of affected children who were diagnosed with cancer prior to 1958, this will result in ascertainment bias but is non-differential because of the matched design. However, Leu *et al.* found that familial risk estimates from linkage of the Swedish Multi-Generation Register and the Cancer Register have minimal bias, owing to left truncation of cancer diagnoses before the start of cancer registration [107]. Thus, we further deemed all cases included from 1958 to have good statistical power. As for complete linkage between parents and children made in 1991, estimates may be biased because of some individuals who died before 1991 and hence who have missing parental information. However, Leu *et al.* concluded that the missing familial links due to death had little effect, except when there was differential mortality for cases with and without a family history of disease [108]. Some studies confirmed this finding by comparing estimates from the data from 1958 and from 1991 onwards [109, 110].

## 6 STATISTICAL METHODS

### 6.1 SURVIVAL ANALYSIS

Survival analysis is a tool for analyzing follow-up time ( $T$ ) from a starting point, such as birth, until the time of events of interest. Here, the event can be either death or occurrence of a disease and is defined in the context of research purpose. However, it is usually difficult to observe the follow-up time fully until the event of interest occurs, because individuals are lost to follow-up due to several reasons. For example, in following individuals from birth for a diagnosis of stomach cancer, some will go “out of view” due to emigration, death, or the end of the study. We refer to such cases as censored (see **Table 4**).

In the analysis of survival data, two functions are of central interest, survival function  $S(t)$  and hazard function  $h(t)$  [111]. Survival function is defined to be the probability that the survival time  $T$  is greater than or equal to a given time  $t$ ,

$$S(t) = P(T \geq t). \quad (1)$$

The hazard function (also referred to as hazard rate) is the (instantaneous) rate of the failure for the survivors to time  $t$  during the next instant of time.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (2)$$

The cumulative hazard function  $H(t)$  is an integration of hazard function and is associated with survival function as follows.

$$H(t) = \int_0^t h(u) du = -\log S(t) \quad (3)$$

where  $\log$  denotes the natural logarithm.

#### 6.1.1 Cox regression model

The Cox proportional hazard model is the most commonly used model in survival analysis [112]. It can be written in two parts, the baseline hazard function of  $t$ ,  $h_0(t)$ , and the effect of covariates  $X$ :

$$h(t|X) = h_0(t) \exp(X\beta). \quad (4)$$

The main assumption of this model is that the effects of covariates are multiplicatively related to the hazard function. Since no parametric form for  $h_0(t)$  is assumed, the Cox regression model is referred to as a semi-parametric model. Here, estimation of parameters is based on partial likelihood. It is essential to check the model assumption, for example, by using Schoenfeld's test [113].

In **Study I**, we stratified data by kinship, sex of relatives and sex of probands and estimated hazard ratios (HRs) as the familial aggregation of CLL using the Cox regression model, where the time scale was age. HRs were adjusted for sex, year of birth, relationship to the proband and sex of the proband. A similar Cox model was used in **Study IV** to investigate the association of cancer survival with a concordant family cancer history as a function of time since cancer diagnosis and HRs were adjusted for sex, age at diagnosis, year of diagnosis, SES, and region of cancer diagnosis.

### 6.1.2 A unified model (main aim of Study II)

We expanded the Cox regression model (4) by the inclusion of several interaction terms allowing us to estimate relationship-specific familial aggregation of the disease of interest without data stratification and test the differences using formal statistical tools such as likelihood ratio test (LRT) or Wald test.

We rephrased the above Cox regression model (4) as follows for the hazard rate  $h_{ij}$  for  $j$  th record in the  $i$  th cluster (see section 5.1.1):

$$h_{ij}(t_{ij}|X_{ij}, Z_{ij}) = h_0(t_{ij}) \exp(\beta X_{ij} + \gamma Z_{ij}) \quad (5)$$

where  $h_0(t_{ij})$  is the baseline hazard at time  $t_{ij}$ ,  $Z_{ij}$  is the binary variable for the exposure, and  $X_{ij}$  is the row of the design matrix describing the other covariates, including type of first-degree relationship (parent, sibling, child), sex of the person at risk and sex of the proband (see **Table 5**). The variable defining the kinship is categorical and so can be represented by dummy indicators such as  $X_1$  and  $X_2$  in **Table 5**, where the reference group (in this case, siblings) depends on the chosen parametrization. In below, we will omit the subscripts,  $ij$ , for simplicity as long as no confusion arises.

Including two-way interactions between  $X$  and  $Z$  in model (5) makes it possible to investigate the effects of kinship, sex of relatives and sex of probands in studies of familial diseases:

1) A model to estimate HRs to different kinships is as follows:

$$h(t|X, Z) = h_0(t) \exp\left(\beta X + \gamma Z + \sum_{k=1}^2 \delta_k X_k Z\right) \quad (6)$$

where  $X = (X_1, X_2, X_3, X_4)$  as described in **Table 5**. For this model,  $\log(\text{HR})$ , comparing hazard of case relatives to hazard of control relatives, is given by  $\gamma$  for siblings of affected siblings and  $\gamma + \delta_1$  for parents of affected children and  $\gamma + \delta_2$  for children of affected parents.

2) A model to estimate HRs of the effects of sex of relatives is below:

$$h(t|X, Z) = h_0(t)\exp(\beta X + \gamma Z + \delta_3 X_3 Z) \quad (7)$$

where  $\gamma$  provides  $\log(\text{HR})$  for female relatives of affected patients and  $\gamma + \delta_3$  for male relatives of affected patients.

3) A model to estimate HRs of the effects of sex of probands is below:

$$h(t|X, Z) = h_0(t)\exp(\beta X + \gamma Z + \delta_4 X_4 Z) \quad (8)$$

where  $\log(\text{HR})$  for relatives of affected female patients is obtained from  $\gamma$  and  $\log(\text{HR})$  for relatives of affected male patients is provided by  $\gamma + \delta_4$ .

We are able to test the significance of the effects of each risk factor such as type of relatives, sex of relatives or sex of probands by using the likelihood ratio test (LRT) because the model (5) is nested in each model (6-8).

**Table 5** Coding system for exposure and variables identifying specific familial relationships.

		Z	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	
<b>Exposure</b>	Case	1					
	Control	0					
<b>Relationship</b>	Parents		1	0			
	Siblings		0	0			
	Children		0	1			
<b>Sex of relatives</b>	Male				1		
	Female				0		
<b>Sex of probands</b>	Male					1	
	Female					0	
<b>Specific family relation</b>	(proband-relative)					Contrasts of exposed vs unexposed relatives for each relationship from model (9)	
	Sister-sister		0	0	0	0	$\gamma$
	Daughter-mother		1	0	0	0	$\gamma + \delta_1$
	Mother-daughter		0	1	0	0	$\gamma + \delta_2$
	Sister-brother		0	0	1	0	$\gamma + \delta_3$
	Brother-sister		0	0	0	1	$\gamma + \delta_4$
	Daughter-father		1	0	1	0	$\gamma + \delta_1 + \delta_3 + \varphi_{13}$
	Son-mother		1	0	0	1	$\gamma + \delta_1 + \delta_4 + \varphi_{14}$
	Mother-son		0	1	1	0	$\gamma + \delta_2 + \delta_3 + \varphi_{23}$
	Father-daughter		0	1	0	1	$\gamma + \delta_2 + \delta_4 + \varphi_{24}$
	Brother-brother		0	0	1	1	$\gamma + \delta_3 + \delta_4 + \varphi_{34}$
	Son-father		1	0	1	1	$\gamma + \delta_1 + \delta_3 + \delta_4 + \varphi_{13} + \varphi_{14} + \varphi_{34} + \psi_1$
	Father-son		0	1	1	1	$\gamma + \delta_2 + \delta_3 + \delta_4 + \varphi_{23} + \varphi_{24} + \varphi_{34} + \psi_2$

In general, by two-way and three-way products of  $(X_1, X_2, X_3, X_4)$  and inclusion of appropriate interactions between  $X_k$ s and exposure ( $Z$ ) in model (5), all relative types can be considered.

Therefore, a unified model of disease risk for first-degree relatives can be constructed as follows:

$$\begin{aligned}
h(t|X, Z) = h_0(t) \exp & \left( \beta X + \gamma Z + \sum_{k=1}^4 \delta_k X_k Z + \sum_{k \neq k', k < k'} \eta_{kk'} X_k X_{k'} \right. \\
& + \sum_{k \neq k', k < k'} \varphi_{kk'} X_k X_{k'} Z + \xi_1 X_1 X_3 X_4 + \xi_2 X_2 X_3 X_4 \\
& \left. + \psi_1 X_1 X_3 X_4 Z + \psi_2 X_2 X_3 X_4 Z \right) \quad (9)
\end{aligned}$$

where  $\eta_{12} = \varphi_{12} = 0$  because  $X_1 = 1$  for parents and  $X_2 = 1$  for children and each individual should be either a parent or a child of a proband.

Model (9) allows the estimation of  $\log(\text{HR})$  comparing exposed relatives to unexposed relatives for all specific first-degree relationships, as presented in **Table 5** (rightmost column). For example,  $\log(\text{HR})$  comparing risk to a sister of a case male proband to risk to a sister of a control male proband is  $\gamma + \delta_4$ .

Furthermore, this model (9) can be extended for familial risk for higher degree relatives [114].

As we described in **Figure 7**, there may be dependence in data because of two or more than two index persons in a family. To deal with this issue, we employed the methods of Pfeiffer *et al.* (2004) [101], using a sandwich variance estimate with families as sampling units, or the bootstrap approach [100] using the clusters defined by the relatives of the case proband and their matched controls (see section 5.1.1). With these two methods, we estimated standard errors and presented confidence intervals. We tested the difference in HRs between different kinships using the robust Wald test [115] with the sandwich variance estimates that regards families as independent sampling units. Detailed explanation can be found in the published **Paper II**.

We developed the unified model in **Study II** and applied this method to studies of adult leukemia and non-Hodgkin's lymphoma.

### 6.1.3 Poisson regression

To unravel the effects of several time scales (such as age and elapsed time) on cancer incidence in case and control relatives in **Study III**, we used the Poisson regression model. The model can accommodate more than one time scale and enables the study of time dependent effects. To implement the Poisson regression model, we first split the records of age or elapsed time and aggregated the records with the same covariate patterns (exposure status, age/elapsed time, sex, current calendar year, and number of family members), and recorded the total number of events of interest and the total time that each individual spent in each of these intervals. The number of events ( $y_i$ ) in  $i$  th interval is assumed to follow a Poisson distribution with mean  $\mu_i$  and

$$\log(\mu_i) = \mathbf{X}_i\beta + \ln(PT_i) \quad (10)$$

where  $\mathbf{X}_i$  represents the vector of covariates in interval  $i$ ,  $\beta$  represents the vector of regression coefficients, and  $PT_i$  is the total person-time that individuals spent in interval  $i$ . Various techniques developed for generalized linear models are applicable here. But, we should be careful when continuous variables are split because the choice of cutoffs for the split can be subjective.

We used the Poisson regression in **Study III** to obtain incidence rate ratios (IRRs) comparing the incidence rate in case siblings and control siblings with adjustment for current calendar year, sex, and number of siblings in models using age as time scale. In models using time since the index's diagnosis as time scale, we adjusted for current calendar year, current age, sex, and number of siblings. In the Poisson data, we plotted the smoothed hazard of each cancer from the **bshazard** package [116], whose function was based on B-splines and generalized linear mixed models [116, 117]. This package has been released in R [118].

### 6.1.4 Flexible parametric model

To study the pattern of changing relative risk of the event in the exposed group, we used the flexible parametric model rather than further splitting the data. The flexible parametric model [119] is fitted on the log cumulative hazard scale and used restricted cubic spline function [120, 121] of  $\ln t$  with  $\mathbf{k}$  knots to estimate the log baseline cumulative hazard.

$$\ln H(t|X) = \ln H_0(t) + X\beta = s(\ln t|\gamma, \mathbf{k}) + X\beta \quad (11)$$

The restricted cubic splines forced the estimate of log baseline cumulative hazard to be linear outside the first knot and the last knot, i.e., the minimum and maximum of the uncensored survival times [122]. With larger datasets (tens of thousands of observations and more), five or six knots are recommended to capture the reasonable shape of the log cumulative hazard [122]. This flexible parametric model provides an estimate of baseline hazard function. Here, the proportional assumption is still required to hold.

If the influence of covariates varies over time, the flexible parametric model can be extended to include interactions between covariates ( $x_j$ ) and spline function of  $\ln t$ ,

$$\ln H(t|X) = s(\ln t | \gamma, \mathbf{k}) + X\beta + \sum_{j=1}^D s(\ln t | \delta_j, \mathbf{k}_j)x_j \quad (12)$$

where  $D$  is the number of time dependent effects,  $\mathbf{k}_j$  denotes the knots for the  $j$  th time dependent effect with associated a parameter,  $\delta_j$ .

In **Study III**, the changing cancer risks with age and elapsed time are presented graphically as HRs. These were estimated by using five knots in spline for cumulative baseline hazard and two knots for time dependent effect using the **stpm2** package [123] in Stata (Statacorp, College Station, TX) [124].

## 6.2 MULTINOMIAL LOGISTIC REGRESSION

We used multinomial logistic regression to study the relationship between covariates and an outcome with more than two categories [125]. Assume there are three categories in the outcome variable, coded 0, 1, 2 for  $Y$ , with 0 denoting the reference outcome. If we are interested in the model with a dichotomous covariate  $x$ , such as family cancer history variable in **Study IV**, we have the following regression model:

$$P(Y = j|x) = \frac{1}{1 + \sum_{k=1}^2 \exp(x\beta_k)} \quad \text{for } j = 0 \quad (13)$$

$$P(Y = j|x) = \frac{\exp(x\beta_j)}{1 + \sum_{k=1}^2 \exp(x\beta_k)} \quad \text{for } j = 1 \text{ or } 2 \quad (14)$$

With this, we can compute the odds ratio of the  $j$  th category ( $OR_j$ ) as below: for  $j \in \{1, 2\}$

$$OR_j = \frac{P(Y = j|x = 1)/P(Y = 0|x = 1)}{P(Y = j|x = 0)/P(Y = 0|x = 0)} \quad (15)$$

$OR_j$  represents the odds ratio of the category  $j$  relative to the reference ( $Y=0$ ) within those with  $x=1$  compared to those with  $x = 0$ .

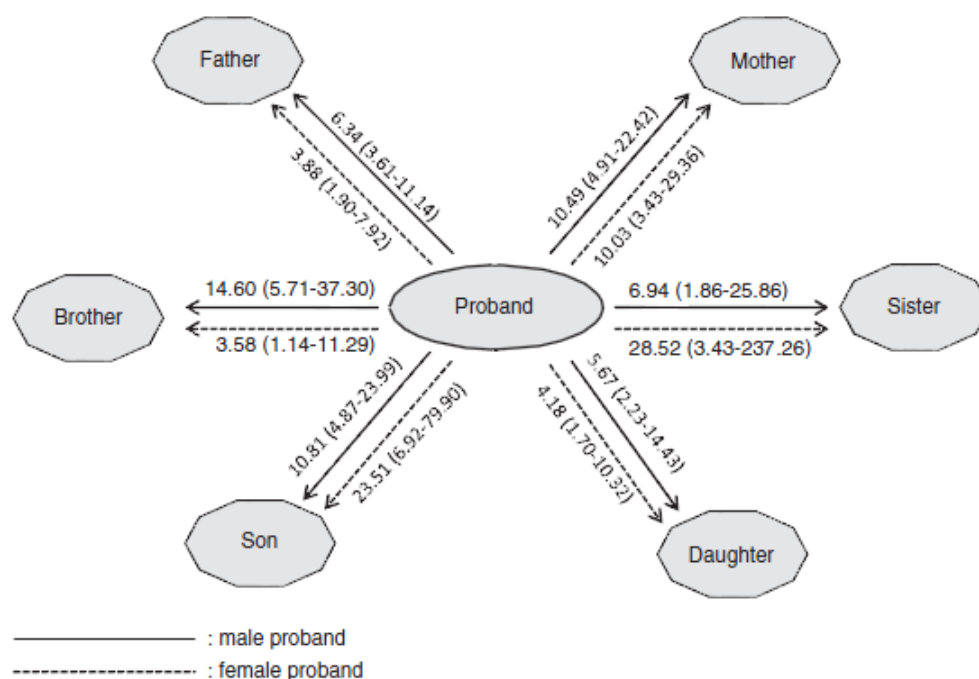
Multinomial logistic regression was implemented in **Study IV** to investigate the association between tumor stage having multiple categories and family cancer history. Here, the tumor stage is treated as outcome and the family cancer history is considered as the main exposure, while adjusting for sex, age at diagnosis, year of diagnosis, socio-economic status, and region.

## 7 RESULTS AND INTERPRETATION

### 7.1 STUDY I : HIGH RISKS OF FAMILIAL CLL FOR SPECIFIC RELATIVES: SIGNPOSTS FOR GENETIC DISCOVERY?

#### 7.1.1 Results

Overall familial aggregation of CLL is HR=7.7 (95% confidence interval (6.00-9.87)) consistent with those reported by Goldin *et al.* [126], and specific familial aggregations for each type of relative are presented using a pedigree diagram in **Figure 8**. The highest familial aggregations were for same-sex siblings of both female probands (28.52, 95% C.I. (3.43-237.26)) and male probands (14.60, 95% C.I. (5.71-37.30)). Although we excluded twins from sibling pairs to remove the latent genetic confounder, little or no difference was estimated in the results. Sons had a much higher risk than daughters, especially if mother was affected (23.51, 95% C.I. (6.92-79.90)). To assess the potential impact of shared environmental exposure, we studied spouses but found no evidence of an increased risk of CLL in the spouses of cases.



**Figure 8** Pedigree diagram presenting the hazard ratios with 95% C.I.s of CLL for different family members of male and female probands. The estimates from a function of age and are adjusted for sex, birth cohort, relationship to the proband and sex of the proband.

#### 7.1.2 Interpretation

Adult leukemia and childhood leukemia are considered as different disease entities. Childhood leukemia is mostly of the acute lymphatic type, while the majority of adult lymphatic leukemia cases are chronic. There was no evidence of familial effect in childhood leukemia [98, 127, 128]. In addition, the response to treatment is dramatically different for



childhood and adult leukemia, with a very high cure rate for childhood leukemia [129] and a poor prognosis for adults [130, 131]. We observed the striking familial aggregation for CLL, which is consistent with an inherited genetic effect that has been proposed [132, 133], with evidence accumulated from family studies and linkage studies [104]. Recent association studies have identified common variants associated with susceptibility to CLL [134-137], but these may only account for a small proportion of the genetic risk [137, 138]. The lack of increased risk for spouses in our data suggests that shared adult environment is not an important factor for familial CLL. Our observation of high risks for same-sex siblings might call for further work to assess the possible contribution of childhood environment. We propose that in studies of familial CLL, consideration of sex and kinship may be crucial to the genotyping of informative subjects for genetic discovery.

## 7.2 STUDY II : A UNIFIED MODEL FOR ESTIMATING AND TESTING FAMILIAL AGGREGATION

### 7.2.1 Results

We applied extended Cox regression models, (5)-(8) and the unified model (9) described in section 6.1.2, to studies of familial cancer risk with respect to adult leukemia and non-Hodgkin's lymphoma, from Swedish population registers. **Table 6** presents results from the

**Table 6** Estimated hazard ratios with 95% confidence intervals for risks of adult leukemia in family members.

Adult Leukemia		HR and 95% C.I. <sup>a</sup>	P-value <sup>b</sup>	Applied model
<b>Overall</b>		2.00 (1.75-2.28)		(5)
<b>Relationship</b>	Parents	1.91 (1.57-2.32)		(6)
	Siblings	3.07 (2.19-4.31)		
	Children	1.80 (1.45-2.24)	0.03 <sup>c</sup>	
<b>Sex or relatives</b>	Male	2.09 (1.75-2.49)		(7)
	Female	1.89 (1.54-2.32)	0.46 <sup>c</sup>	
<b>Sex of probands</b>	Male	2.05 (1.72-2.44)		(8)
	Female	1.93 (1.57-2.39)	0.69 <sup>c</sup>	
<b>Specific family relationship</b>	<b>(proband-relative)</b>			
	Sister-sister	5.35 (2.72-10.54)	Ref.	
	Daughter-mother	1.37 (0.83-2.27)	<0.01	
	Mother-daughter	1.61 (0.93-2.80)	<0.01	
	Sister-brother	2.03 (0.91-4.50)	0.07	
	Brother-sister	2.34 (1.09-5.05)	0.11	
	Daughter-father	1.92 (1.26-2.93)	0.02	(9)
	Son-mother	2.02 (1.40-2.91)	0.02	
	Mother-son	1.96 (1.30-2.95)	0.04	
	Father-daughter	1.56 (0.99-2.48)	0.02	
	Brother-brother	3.10 (1.77-5.41)	0.25	
	Son-father	2.13 (1.52-2.99)	0.10	
	Father-son	1.95 (1.35-2.83)	0.10	

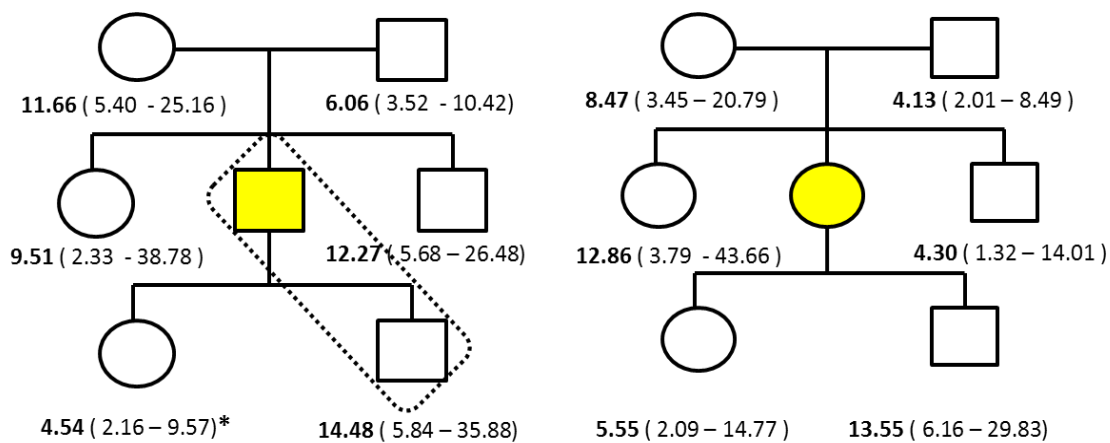
a. HRs and 95% C.I.s from a function of age and with adjustment for kinship, sex of proband, and sex of relative.

b. Based on the robust Wald test, to test the difference in HRs of kinships where the HR of sisters of female probands as reference.

c. Based on the LRT, to test significance of difference in HRs.

adult leukemia analyses. The overall HR of adult leukemia was 2.00 with 95% C.I. (1.75-2.28) based on the sandwich variance and 95% C.I. (1.64-2.39) based on the bootstrap variance. We found familial risks to be higher for siblings (3.07, 95% C.I. (2.19-4.31)) than for parents (1.91, 95% C.I. (1.57-2.32) and children (1.80, 95% C.I. (1.45-2.24), and this observed difference was statistically significant based on the likelihood ratio test (LRT) ( $p=0.03$ ). By implementing our unified model, we found that sisters of female probands have a significantly higher risk than any of the other relatives except brothers (of female probands) and a higher risk than mothers or daughters of male probands (**Table 6**). Graphical presentation with a pedigree diagram is presented in the published **Paper II**. Additional results of non-Hodgkin’s lymphoma are also published in **Paper II**.

We re-analyzed the CLL data, from **Study I** [139], using the unified model, and found no clear evidence of sex-linked genetic susceptibility [140] as the risk to daughters was not significantly different to the risk to sons of affected mothers ( $p=0.13$ ) but we detected a higher risk for sons (compared with daughters) of affected fathers ( $p=0.04$ ) (see **Figure 9**).



**Figure 9** Pedigree diagrams showing the hazard ratios of adult chronic lymphatic leukemia for different family members of male and female probands. Squares and circles denote male and female, respectively. The estimate marked with an asterisk (\*) are significantly different than the reference (risk to sons of male probands who are represented by a box with dotted borders). Significance level is 0.05.

## 7.2.2 Interpretation

### *Unified model*

To investigate how familial risk of disease varies with kinship, we proposed a unified model using an extended Cox regression model by inclusion of interaction terms using all data simultaneously. This allowed us to formally test and compare the difference of kinship-specific familial aggregation. Our method enables the effect of different kinships to be estimated and offers an improvement over traditional epidemiological investigations, which have relied on stratifying the data and conducting many separate analyses. However, our

method requires a large sample size because we are interested in the effect of the interaction terms, but this may present computational challenges in the variance estimation from the bootstrap approach. In our data, confidence intervals obtained from robust sandwich estimation and bootstrap are similar although the bootstrap approach is preferable because it considers dependence in data more accurately [101].

#### *Application to adult leukemia and non-Hodgkin lymphoma studies*

The ability to model all relationships within the family may offer deeper insight into the potential genetic and environmental factors underlying familial clustering of disease and provide a more refined tool for genetic counseling of family members, where the kinship with the affected family member is considered.

We confirmed previous findings of familial aggregation of adult leukemia and non-Hodgkin's lymphoma: both are well known as familial cancers [38, 141, 142], in which higher sibling risks compared to risks from parents may be explained by shared childhood environment as well as involvement of recessive susceptibility [38], and sex-specific and kindred-specific familial risk of NHL from simple standardized incidence ratios are reconfirmed [141]. The significantly increased risk in siblings of the same sex (particularly sisters) suggests that shared environmental risk factors between family members account for at least part of the risk [143, 144].

### **7.3 STUDY III : PATTERNS OF CHANGING CANCER RISKS WITH TIME SINCE DIAGNOSIS OF A SIBLING**

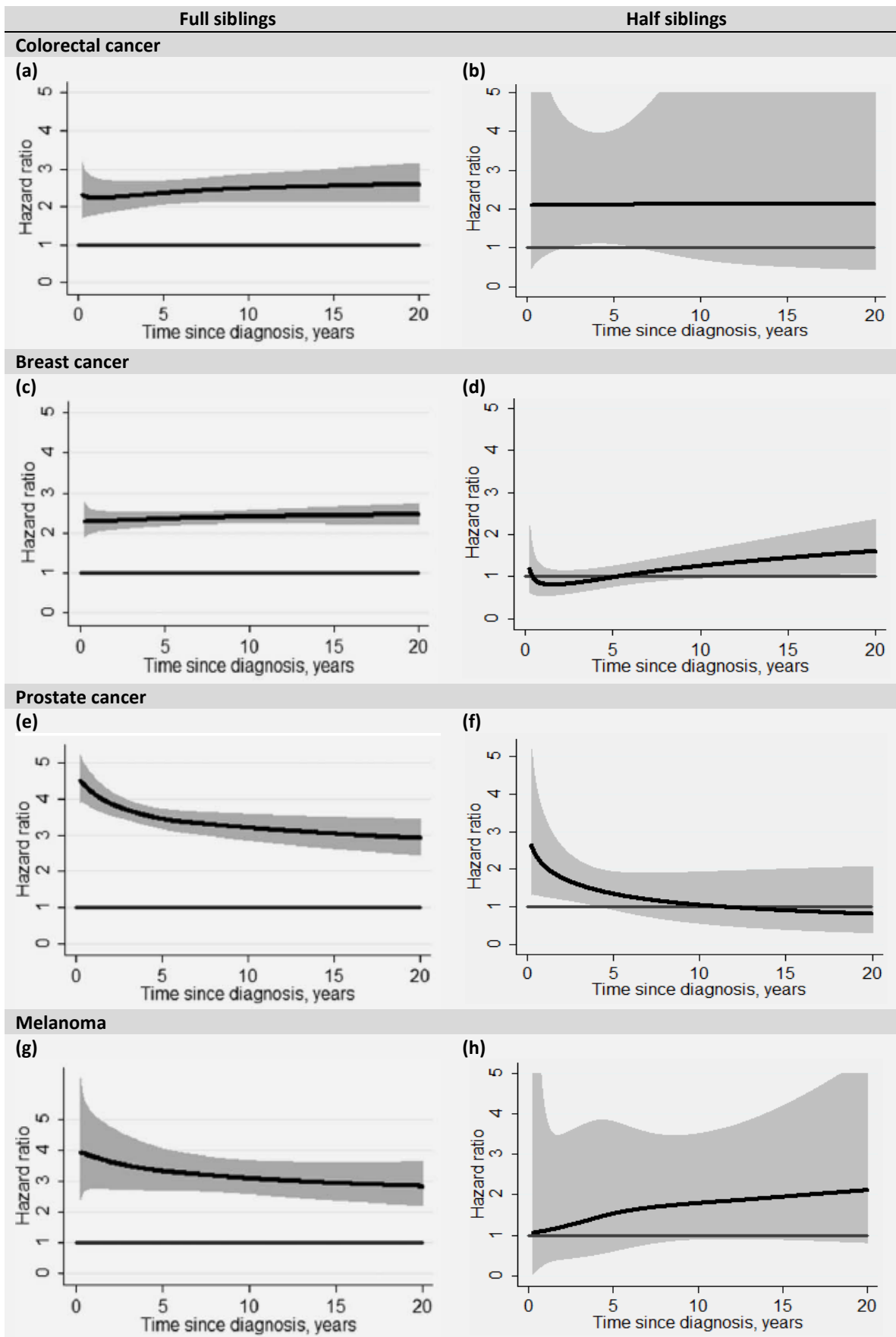
#### **7.3.1 Results**

##### *Absolute risk of cancer by age*

Case siblings had higher cancer incidence than control siblings for all cancers at all ages, with the exception of prostate cancer, for which no conclusions could be drawn for siblings younger than 40 years because no events occurred. Graphical presentation and additional details can be found in the published **Paper III**.

##### *Overall relative risk of cancer*

Overall familial aggregation of each cancer was found to be consistent regardless of the time scale, age, or elapsed time. The incidence rate ratio (IRR) with 95% confidence interval (95% C.I.) was 2.41 (2.14-2.71) for colorectal cancer, 2.37 (2.24-2.52) for breast cancer, 3.69 (3.46-3.93) for prostate cancer, and 3.20 (2.72-3.76) for melanoma. These findings were consistent with other studies [145-149].



**Figure 10** HRs with 95% C.I.s (solid lines with gray-shaded intervals) comparing siblings of cancer patients and siblings of matched controls where time scale is time since diagnosis. The straight line at hazard ratio=1 corresponds to the reference group, control siblings. Left column for full siblings and right column for half-siblings.

### *Relative risk of cancer by age*

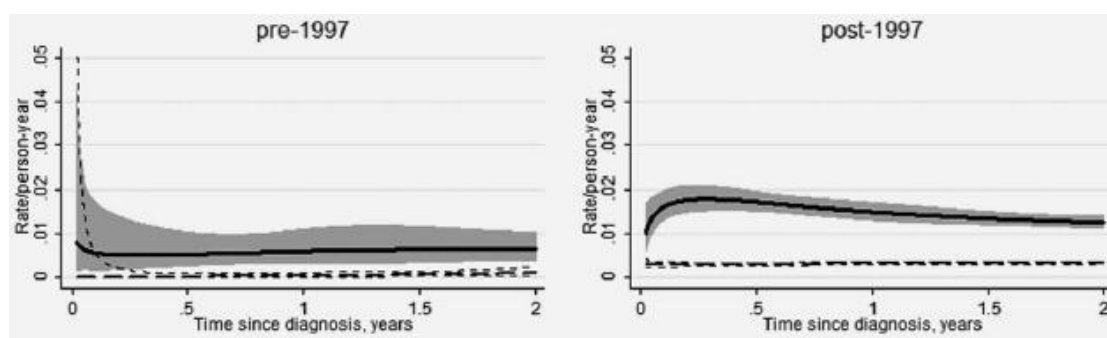
The relative risk was significantly higher in case siblings at all ages. Younger siblings had the highest IRR for colorectal cancer, breast cancer, and prostate cancer, and the IRRs for each cancer decreased with age, with melanoma having the most gradual decrease and prostate cancer decreasing steeply at age 50-55, with relatively slow decrease thereafter. The increased relative risk in case siblings was relatively stable in older ages. Figures and additional results are published in **Paper III**.

### *Relative risk of cancer by time since diagnosis of the index case*

The risks in case siblings were significantly higher than in control siblings for up to 20 years after the index person's diagnosis. The increased risk was relatively stable over time for colorectal cancer, breast cancer, and melanoma, whereas prostate cancer illustrated a steeply decreasing trend during the first few years and was stable thereafter (see (e), left column in **Figure 10**).

### *Screening effects for breast cancer and prostate cancer*

We probed the effect of screening on siblings of prostate cancer patients by comparing the risk shortly after index diagnoses in the periods pre- and post- 1997, the year of introduction of prostate-specific antigen (PSA) testing for prostate cancer in Sweden [60, 150]. We found that the incidence rate in brothers of prostate cancer patients diagnosed after 1997 increased right after the index diagnosis and reduced thereafter, a typical pattern of a "screening effect", whereas no such pattern was apparent before 1997 (see **Figure 11**). A similar comparison was conducted for breast cancer pre- and post-1990, the year of nationwide implementation of mammography screening for breast cancer in Sweden [60, 151], but no such a trend of a screening effect was observed after the cut-year. Additional figures and detailed information are published in **paper III**.



**Figure 11** Cancer incidence rate with 95% C.I.s in brothers of prostate cancer patients (solid lines with gray-shaded intervals) and their matched controls (dashed lines) with adjustment for age at the time of brother's prostate cancer diagnosis.

### *Analyses in adoptive or half-siblings*

We also analyzed familial risks in adopted or half siblings to support interpretation about our observed patterns. However, despite a reasonable number of adoptees (n=9,894) and cancer events (n=129, 55, 38, 38 of breast, prostate, colorectal, and melanoma, respectively), we found only one pair of adoptive sisters where both had breast cancer and no concordant sibships for the other cancers, and thus further investigation with adoptive siblings was not able to be conducted.

In half-siblings, overall relative risks were much lower than for full siblings, 1.97 (1.11-3.48) for colorectal cancer, 1.12 (0.90-1.39) for breast cancer, 1.55 (1.14-2.12) for prostate cancer, and 1.51 (0.79-2.89) for melanoma and some IRRs were not significant and had wide confidence intervals because of lack of statistical power. However, we found a significantly higher (though lower than for full siblings) relative risk of prostate cancer in men during the first few years after a step-brother's diagnosis (see (f), right column in **Figure 10**).

### **7.3.2 Interpretation**

In this study, we focused on siblings rather than parent-child relationships. Since siblings are more likely to be contemporaries, a health professional is able to provide helpful and realistic advice for siblings at risk. Also, much less susceptibility to bias was expected in studying siblings than parent-children pairs.

The high familial risks at younger ages for colorectal cancer was consistent with the familial risk of tumors related to hereditary non-polyposis colorectal cancer which is associated with a young age at onset [152-155]. Also, the high familial aggregation of breast cancer at younger ages is consistent with reports of a higher prevalence of mutations in the BRCA1/BRCA2 genes in young patients and a high probability of the same gene mutations in their sisters [156, 157]. The decreasing pattern in the first five years after a brother's cancer prostate cancer may be related to the screening effect, which is consistent with a lead-time bias from opportunistic screening of family members [158], and further extends to half-brothers. In contrast, no evidence of screening effect for breast cancer was observed because of its shorter lead time, similar in magnitude to the screening invitation interval in Sweden [159, 160].

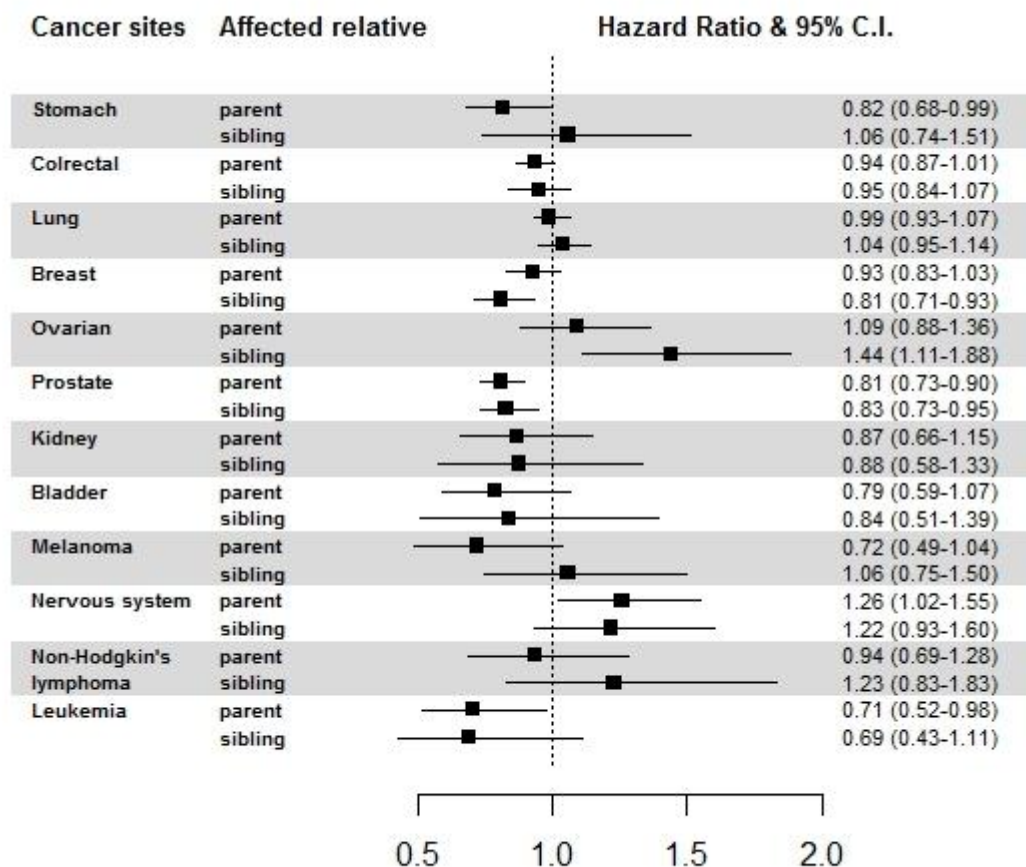
The graphical presentation of trends regarding the change of cancer risk in siblings from the first cancer diagnosis in a sibling over two important time scales (age and time since diagnosis) can help in understanding and interpreting the change in familial risk over time. Siblings of cancer patients continue to be at increased risk, even after many years, and thus continued, and perhaps more frequent, screening may be important in this subpopulation.

## 7.4 STUDY IV : DIFFERENCES IN SURVIVAL FOR PATIENTS WITH FAMILIAL AND SPORADIC CANCER

### 7.4.1 Results

We found that a family cancer history of a concordant cancer played a protective role for cancer survival in patients with breast cancer, prostate cancer or leukemia: the hazard ratio (HR) was 0.88 (95% confidence interval 0.81-0.96) for breast cancer, 0.82 (0.75-0.90) for prostate cancer, 0.70 (0.54-0.92), for leukemia. We also found that stomach cancer patients who had an affected parent had better survival than sporadic cancer patients, 0.82 (0.68-0.99). In contrast, familial cancer patients had a worse prognosis for patients with ovarian cancer and nervous-system cancer: 1.20 (1.01-1.43) and 1.24 (1.05-1.47), respectively. These HRs are presented in **Figure 12** stratified by the affected relative.

For breast cancer, prostate cancer and leukemia, the protective effect was strongest in younger patients and in breast cancer patients with an affected sibling or leukemia patients with an affected parent. Familial ovarian cancer patients with an affected sister diagnosed at younger ages had the highest relative risk, 1.75 (1.23-2.49). Additional details can be found in the manuscript for **Study IV**.



**Figure 12** HRs with 95% C.I.s comparing survival of familial cancer patients to survival of sporadic cancer patients over five years after the cancer diagnosis. The estimates are adjusted for sex, age at diagnosis, year of diagnosis, SES, and region of cancer diagnosis.

For the six cancers which had any significant association in **Figure 12**, we further examined the potential explanation for the differential survival varying with cancer sites by the proportion of aggressive histological type/subtype in familial cancer patients, differential survival within histological type/subtype, and stage at diagnosis. **Table 7** presents results of breast cancer and ovarian cancer from our data. Among breast cancer patients, most had a ductal breast cancer and ductal breast cancer patients with family cancer history had better cancer survival than sporadic ductal breast cancer patients. Serous ovarian cancer is quite aggressive and has a familial component but no survival difference was found in familial ovarian cancer patients and sporadic ovarian cancer patients. However, we observed that mucinous ovarian cancer patients with family cancer history had twice the risk of death of sporadic cancer patients. Family history of prostate cancer was associated with a reduction of 11 per cent in the odds of getting a T3 tumor (**Table 8**). No significant association was found with nodal involvement (N) nor with metastasis (M). We observed a large and significant increase in the odds of higher FIGO stages versus the lowest stage for ovarian cancer, odds ratio ranging from 2.52 to 3.22. Further details can be found in the manuscript for **Study IV**.

#### 7.4.2 Interpretation

We have demonstrated a different role of family cancer history in cancer prognosis depending on the cancer site: a protective effect for stomach cancer, breast cancer, prostate cancer, and leukemia but a poorer prognosis for ovarian cancer and nervous system cancers. No association was found for lung cancer, kidney cancer, bladder cancer, melanoma, or non-Hodgkin's lymphoma.

**Table 7** Results related to histological type/subtype : Overall 5-year mortality rate (column 3), distribution of histology by family cancer history among patients (columns 4-8) and hazard ratios comparing survival of familial cancer patients to survival of sporadic cancer patients over 5 years (column 9).

Cancer sites	histology	5-year mortality rate per 1000 person-years	n.patients				p <sup>b</sup>	HR & 95% C.I. <sup>c</sup>
			sporadic	(%)	familial	(%)		
<b>Breast<sup>a</sup></b>	all	19.8 (19.2-20.4)						
	Ductal	19.5 (18.7-20.2)	27723	(67.9)	4787	(67.3)		<b>0.86 (0.76-0.96)</b>
	Lobular	14.7 (13.3-16.2)	5240	(12.8)	1000	(14.1)	*	0.99 (0.75-1.30)
	Others	24.2 (22.7-25.8)	7865	(19.3)	1330	(18.7)	<b>0.02</b>	0.88 (0.73-1.06)
<b>Ovarian</b>	all	114.8 (110.1-119.7)						
	Serous	144.3 (136.2-152.8)	2425	(43.3)	154	(54.8)	*	1.06 (0.83-1.34)
	Mucinous	85.7 (73.3-100.2)	523	(9.3)	16	(5.7)		<b>2.09 (1.09-3.98)</b>
	Endometrioid	70.8 (61.5-81.5)	694	(12.4)	38	(13.5)		0.99 (0.50-1.97)
	Clear-cell	90.9 (75.0-110.2)	315	(5.6)	13	(4.6)		1.05 (0.37-3.02)
	Adenocarcinoma NOS	143.0 (129.5-157.9)	868	(15.5)	31	(11.0)		0.80 (0.44-1.45)
	Others	75.2 (65.8-85.9)	778	(13.9)	29	(10.3)	<b>0.002</b>	1.36 (0.73-2.52)

a. female only.

b. Chi-square test for difference in proportions between familial cancer patients and sporadic cancer patients.

\* Histological type where difference occurred between familial cancer patients and sporadic cancer patients.

c. HRs with 95% C.I.s with adjustment for sex, age at diagnosis, year of diagnosis, SES, and region of cancer diagnosis.



**Table 8** Observed counts of familial cancer patients and sporadic cancer patients by tumor extent (T) categories for prostate cancer or FIGO categories for ovarian cancer.

By TNM		T1	T2	T3	T4	$p^b$
Prostate	Sporadic (ref.)	11220	6129	2860	384	<b>0.001</b>
	Familial	3418	1769	734	130	
	OR & 95% C.I. <sup>a</sup>		0.97 (0.91-1.04)	<b>0.89</b> <b>(0.81-0.97)</b>	1.14 (0.93-1.39)	
By FIGO		I	II	III	IV	$p^b$
Ovarian	Sporadic (ref.)	565	176	827	265	<b>0.01</b>
	Familial	13	11	45	18	
	OR & 95% C.I. <sup>a</sup>		<b>2.79</b> <b>(1.21-6.40)</b>	<b>2.52</b> <b>(1.33-4.79)</b>	<b>3.22</b> <b>(1.52-6.79)</b>	

a. ORs with 95% C.I.s for familial cancer patients vs sporadic cancer patients are adjusted for sex, age at diagnosis, year of diagnosis, SES, and region of cancer diagnosis.

b. Chi-square test for difference in proportions between familial cancer patients and sporadic cancer patients.

Patients with a family history of breast cancer may derive a survival benefit from positive medical surveillance and increased awareness in breast cancer patients [160]. Familiality of lobular breast cancer [161-163] and its better prognosis compared to survival of other types of breast cancer [164] could partly explain the protective effect.

Prostate cancer patients with an affected father or brother are more likely to seek medical attention. We showed in **Study III** that the cancer incidence in full or half- brothers of prostate cancer patients increased significantly right after a cancer diagnosis in a brother, providing evidence that the introduction of PSA testing may contribute to the increased risk [165]. Thus familial prostate cancer patients have higher probability of early detection thanks to opportunistic screening such as PSA testing [150], which contributes to a protective role of family cancer history in cancer prognosis.

We observed several factors to support the poor prognosis in women with a family history of ovarian cancer. Young ovarian cancer women with an affected sister had a higher risk of death in the five years after the cancer diagnosis, which may be related to germline mutations relevant to ovarian cancer [166]. Serous ovarian cancer was more frequently detected among familial ovarian cancer patients [167], and had a poorer prognosis than other histological types [79, 168]. In our data, patients with familial mucinous ovarian cancer had increased risk of death, possibly suggesting germline mutations related to prognosis or therapy response. In addition, higher FIGO stage tumors detected in familial ovarian cancer patients may be due to lack of diagnostic testing sufficiently accurate to identify early-stage ovarian cancer [169].

Familial cancer patients with cancer of the nervous system had a poorer outcome than sporadic cancer patients, the effect being highest in patients with gliomas of uncertain origin histological type. Among leukemia patients with a family history of leukemia, lymphatic leukemia whose prognosis was better than other subtypes [170] was more likely to be observed [171-173]. This may contribute to the protective effect of family cancer history.

We provide evidence that family cancer history is a prognostic factor for cancers at some sites and that histological type of cancer also has some prognostic value. As we noted for

leukemia in **Study I**, the strong evidence of more aggressive ovarian cancers in sisters and daughters of ovarian cancer patients may be informative for genetic counseling and help to guide further molecular or genetic investigations.

## 8 CONCLUSIONS AND FUTURE PERSPECTIVES

This thesis presents in-depth analyses of familial cancer using the Swedish population based registers to address the following questions: what is the risk to specific relatives where a family member is diagnosed with cancer? how should relatives of cancer patients perceive their risk at different ages? should relatives be considered at higher risk for their remaining life time? does family cancer history play a role in cancer survival?

We examined how the type of kinship and sex affect the familial risk of adult chronic lymphatic leukemia. We developed a unified model to address this question in general by extending the Cox regression model to include interaction terms that enable estimation and formal comparison of family members. For four of the main cancers in Sweden (colorectal cancer, breast cancer, prostate cancer and melanoma), we presented the pattern of changing cancer risk with time since the first diagnosis in a sibling. Lastly, we investigated the association between family cancer history and cancer patient survival with respect to twelve familial cancers.

The main conclusions from the four studies are as follows,

- Familial aggregation in chronic lymphatic leukemia is high and it varies by sex and kinship: the highest familial risk is in sisters of affected women, followed by sons of affected mothers.
- No increased risk of CLL was found in the spouses of case probands suggesting that shared adult environment may not contribute to familial risk.
- Familial risks to all family members can be jointly estimated and compared with formal statistical tests using a unified model that accommodate correlation in the data.
- In four major cancers in Sweden, colorectal cancer, breast cancer, prostate cancer and melanoma, the increased cancer risk in siblings persists for up to 20 years after the index cancer diagnosis in the family. Moreover, the risks are higher if an individual is young when his or her sibling is firstly diagnosed with cancer.
- Brothers of prostate cancer patients have slightly higher relative risks in the first few years after the index diagnosis in the family, which may be related to the effect of opportunistic PSA screening.
- Cancer survival is associated with a concordant family cancer history for some cancer sites: stomach, breast, prostate, ovarian, nervous-system and leukemia.
- Differences in histological type/subtype contributed to differential cancer survival for familial and sporadic cancers.
- The poor prognosis for sisters and daughters of ovarian cancer patients is associated with multiple factors, including familiarity of aggressive histological type and higher tumor stage at cancer diagnosis.

The unified model developed in **Study II** can estimate kinship-specific familial aggregation without the need for data stratification. The model can be readily extended to higher degree relatives, although it needs further development to derive variance estimators to properly account for matching and familial correlation considering the different layers of dependence in the data. Today, there have been many studies of the most complex human diseases and their association with genetic variants, such as cancer, diabetes, and Alzheimer's disease [174, 175]. We believe that our unified model has potential for application to many areas, not only in cancer studies but also in studies of other diseases such as autism [176] or psychiatric problems [177], offering an improvement over traditional epidemiological investigations. A deeper understanding of the role of kinship should be given priority in the design of genetic biomarker studies.

It is important to know if/how the risk of disease changes over time as well as describing overall or lifetime increased risk. **Study III** in this thesis unraveled the effects of age and elapsed time on cancer risks in siblings. The resulting information which was presented graphically provides evidence for whether or when family members should undergo screening and/or genetic testing. We can also observe an evidence of the effect of screening especially for prostate cancer. Further linkage of family data to known risk variants or identified susceptibility markers [178] may improve risk stratification and characterize subpopulations for targeted screening.

Family cancer history is a well-known and an unavoidable risk factor for cancer. This thesis studied family cancer history from various angles and found that it is associated not only with cancer risk in relatives but also with cancer survival in patients for some cancer sites and even in cancer patients with specific histological types. This being so, it is important to understand the role of family cancer history systematically. For example, relatively high familial risks in sisters of affected female siblings for several hematological malignancies, including adult (chronic) leukemia, and non-Hodgkin's lymphoma, reflects the possible contribution of shared genetic variants [179]. Thus, it is important to consider sex and kinship in selecting informative subjects to identify such genes. As the high familial aggregation can also be caused by shared childhood environment, further analyses after the linkage of environmental exposure data can be warranted. Further investigation of familial cancer survival should include clinical details, such as biological markers and treatment and behavioral factors such as (response to) genetic counselling. Further molecular or genetic investigations of cancer patients are necessary to gain a deeper understanding of the biological mechanisms behind the survival.

My PhD project has been made possible by the infrastructure for research in Sweden, especially the accessibility of high-quality population-based registers. Thanks to such valuable resources, the results of this thesis can, I hope, leave a small footprint on the wide world of cancer epidemiology. Nowadays, many countries have established population-based registers in various fields offering more opportunities for epidemiological research to provide increased understanding of human diseases and help to provide better public health.

## 9 ACKNOWLEDGEMENTS

Now, at the end of this long journey and the start of the next one, I should like to express my appreciation to many people who have contributed along the way.

First, I thank all my supervisors, who gladly shared their knowledge, showed their passion for the research, and expanded my horizons during last four years. Your feedback at every meeting excited me.

**Marie Reilly**, my main supervisor, taught me a lot about the application of general life principles to research. I wish to learn from your constant curiosity and precision. I will keep in mind your saying, “Find the ‘röd tråd’.”

I greatly admire the insight, logical thinking, communication skills – and fashion sense – of my cosupervisor, **Kamila Czene**. Your energy and confidence encouraged me always.

Thanks to my other cosupervisor, **Paola Rebora**, who readily shared her statistical knowledge and whose quick response always gave me a lift. I will never forget your smile at our first meeting.

I am grateful to my co-authors, **Maria Grazia Valsecchi** and **Linda Sofie Lindström**, for the interesting scientific discussions we had.

Thanks to **all friends and colleagues** in our department, with whom I have been fortunate to spend the last four years. Your encouragement and your kind manners contributed to a good working environment. Thank you to my friends and Biostat group members, especially to past and present roommates.

I thank past and present **Korean researchers in KI**. One of my favorite events is having lunch with you, and I hope that our relationship continues far beyond this collaboration. I want to express special appreciation to my mentors, **Mun-Gwan Hong** and **Sohyun Lee**.

I wish to thank from the bottom of my heart **my parents and parents-in-law** for their understanding and support, particularly my mother. You have unfailingly fulfilled all my requests, large and small, from 8,000 km away, far beyond the call of duty.

Last but not the least, I would like to thank my husband and my son, **Woojoo** and **Seongjoon**, who cheer and heal me all the time. Without your endless support and love, none of this dream would have come true.



## REFERENCES

1. Warthin, A.S., *Heredity with reference to carcinoma*. Arch Intern Med, 1913. **12**: p. 546-555.
2. Warthin, A.S., *The further study of a cancer family*. J Cancer Res, 1925. **9**: p. 279-286.
3. Hauser IJ, W.C., *A further report on the cancer family of Warthin*. Am J Cancer, 1936. **27**: p. 434-449.
4. Lynch, H.T. and A.J. Krush, *Cancer Family G Revisited - 1895-1970*. Cancer, 1971. **27**(6): p. 1505-&.
5. Douglas, J.A., et al., *History and molecular genetics of Lynch syndrome in family G: a century later*. Jama, 2005. **294**(17): p. 2195-202.
6. Nagy, R., K. Sweet, and C. Eng, *Highly penetrant hereditary cancer syndromes*. Oncogene, 2004. **23**(38): p. 6445-70.
7. Carstensen, B., et al., *Familial aggregation of colorectal cancer in the general population*. Int J Cancer, 1996. **68**(4): p. 428-35.
8. Goldgar, D.E., et al., *Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands*. J Natl Cancer Inst, 1994. **86**(21): p. 1600-8.
9. Peto, J., et al., *Cancer mortality in relatives of women with breast cancer: the OPCS Study*. Office of Population Censuses and Surveys. Int J Cancer, 1996. **65**(3): p. 275-83.
10. Gudbjartsson, T., et al., *A population-based familial aggregation analysis indicates genetic contribution in a majority of renal cell carcinomas*. Int J Cancer, 2002. **100**(4): p. 476-9.
11. Hemminki, K. and C. Eng, *Clinical genetic counselling for familial cancers requires reliable data on familial cancer risks and general action plans*. Journal of medical genetics, 2004. **41**(11): p. 801-7.
12. *Center for Epidemiology. Causes of Death 2010*. Stockholm: The National Board of Health and Welfare, 2011.
13. *Center for Epidemiology. Cancer incidence in Sweden 2011*. Stockholm: The National Board of Health and Welfare, 2012.
14. Anand, P., et al., *Cancer is a Preventable Disease that Requires Major Lifestyle Changes*. Pharm Res-Dord, 2008. **25**(9): p. 2097-2116.
15. Lindor, N.M. and M.H. Greene, *The concise handbook of family cancer syndromes*. Mayo Familial Cancer Program. J Natl Cancer Inst, 1998. **90**(14): p. 1039-71.
16. Garber, J.E. and K. Offit, *Hereditary cancer predisposition syndromes*. J Clin Oncol, 2005. **23**(2): p. 276-92.
17. Irigaray, P., et al., *Lifestyle-related factors and environmental agents causing cancer: an overview*. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie, 2007. **61**(10): p. 640-58.

18. Willett, W.C., *Diet and cancer*. *Oncologist*, 2000. **5**(5): p. 393-404.
19. Calle, E.E., et al., *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults*. *The New England journal of medicine*, 2003. **348**(17): p. 1625-38.
20. Parkin, D.M., *The global health burden of infection-associated cancers in the year 2002*. *Int J Cancer*, 2006. **118**(12): p. 3030-44.
21. Belpomme, D., et al., *The multitude and diversity of environmental carcinogens*. *Environmental research*, 2007. **105**(3): p. 414-29.
22. Easton, D.F., *The inherited component of cancer*. *British medical bulletin*, 1994. **50**(3): p. 527-35.
23. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland*. *The New England journal of medicine*, 2000. **343**(2): p. 78-85.
24. Hemminki, K. and P. Vaittinen, *Familial cancers in a nationwide family cancer database: age distribution and prevalence*. *European journal of cancer*, 1999. **35**(7): p. 1109-17.
25. Cannon-Albright, L.A., et al., *Familiality of cancer in Utah*. *Cancer research*, 1994. **54**(9): p. 2378-85.
26. Lynch, H.T., R.M. Fusaro, and J. Lynch, *Hereditary cancer in adults*. *Cancer detection and prevention*, 1995. **19**(3): p. 219-33.
27. Lindelof, B. and G. Eklund, *Analysis of hereditary component of cancer by use of a familial index by site*. *Lancet*, 2001. **358**(9294): p. 1696-8.
28. Aitken, J.F., et al., *Accuracy of case-reported family history of melanoma in Queensland, Australia*. *Melanoma Res*, 1996. **6**(4): p. 313-7.
29. Soegaard, M., et al., *Accuracy of self-reported family history of cancer in a large case-control study of ovarian cancer*. *Cancer causes & control : CCC*, 2008. **19**(5): p. 469-79.
30. Kerber, R.A. and M.L. Slattery, *Comparison of self-reported and database-linked family history of cancer data in a case-control study*. *Am J Epidemiol*, 1997. **146**(3): p. 244-8.
31. Glanz, K., et al., *Underreporting of family history of colon cancer: correlates and implications*. *Cancer Epidemiol Biomarkers Prev*, 1999. **8**(7): p. 635-9.
32. Eerola, H., et al., *Familial breast cancer in southern Finland: how prevalent are breast cancer families and can we trust the family history reported by patients?* *European journal of cancer*, 2000. **36**(9): p. 1143-8.
33. King, T.M., et al., *Accuracy of family history of cancer as reported by men with prostate cancer*. *Urology*, 2002. **59**(4): p. 546-50.
34. Gaff, C.L., et al., *Accuracy and completeness in reporting family history of prostate cancer by unaffected men*. *Urology*, 2004. **63**(6): p. 1111-6.
35. Mitchell, R.J., et al., *Accuracy of reporting of family history of colorectal cancer*. *Gut*, 2004. **53**(2): p. 291-5.



36. Lee, H., P. Flaherty, and H.P. Ji, *Systematic genomic identification of colorectal cancer genes delineating advanced from early clinical stage and metastasis*. BMC medical genomics, 2013. **6**: p. 54.
37. Mo, Q., et al., *Pattern discovery and cancer gene identification in integrated cancer genomic data*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(11): p. 4245-50.
38. Hemminki, K., X. Li, and K. Czene, *Familial risk of cancer: data for clinical counseling and cancer genetics*. Int J Cancer, 2004. **108**(1): p. 109-14.
39. *Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease*. Lancet, 2001. **358**(9291): p. 1389-99.
40. Bevier, M., K. Sundquist, and K. Hemminki, *Risk of breast cancer in families of multiple affected women and men*. Breast Cancer Res Tr, 2012. **132**(2): p. 723-8.
41. Bevier, M., K. Sundquist, and K. Hemminki, *Risk of breast cancer in families of multiple affected women and men*. Breast Cancer Res Treat, 2012. **132**(2): p. 723-8.
42. Hemminki, K., C. Dong, and P. Vaittinen, *Cancer risks to spouses and offspring in the Family-Cancer Database*. Genetic epidemiology, 2001. **20**(2): p. 247-57.
43. Hemminki, K. and Y. Jiang, *Cancer risks among long-standing spouses*. Brit J Cancer, 2002. **86**(11): p. 1737-1740.
44. Bevier, M., et al., *Influence of family size and birth order on risk of cancer: a population-based study*. BMC cancer, 2011. **11**: p. 163.
45. Hemminki, K. and P. Mutanen, *Birth order, family size, and the risk of cancer in young and middle-aged adults*. Brit J Cancer, 2001. **84**(11): p. 1466-1471.
46. Veierod, M.B., et al., *Sun and Solarium Exposure and Melanoma Risk: Effects of Age, Pigmentary Characteristics, and Nevi*. Cancer Epidem Biomar, 2010. **19**(1): p. 111-120.
47. Nagore, E., et al., *Smoking, sun exposure, number of nevi and previous neoplasias are risk factors for melanoma in older patients (60 years and over)*. J Eur Acad Dermatol, 2010. **24**(1): p. 50-57.
48. Webb, P.M., et al., *Relation between Infection with Helicobacter-Pylori and Living-Conditions in Childhood - Evidence for Person-to-Person Transmission in Early-Life*. Brit Med J, 1994. **308**(6931): p. 750-753.
49. Gao, L., et al., *Sibship size, Helicobacter pylori infection and chronic atrophic gastritis: a population-based study among 9444 older adults from Germany*. Int J Epidemiol, 2010. **39**(1): p. 129-34.
50. Hemminki, K., et al., *The Swedish Family-Cancer Database 2009: prospects for histology-specific and immigrant studies*. Int J Cancer, 2010. **126**(10): p. 2259-67.
51. Michels, K.B., et al., *Birthweight as a risk factor for breast cancer*. Lancet, 1996. **348**(9041): p. 1542-1546.
52. Xu, X.H., et al., *Birth Weight as a Risk Factor for Breast Cancer: A Meta-Analysis of 18 Epidemiological Studies*. J Womens Health, 2009. **18**(8): p. 1169-1178.

53. Richiardi, L., et al., *Birth order, sibship size, and risk for germ-cell testicular cancer*. *Epidemiology*, 2004. **15**(3): p. 323-329.
54. Hemminki, K. and P. Kyyronen, *Parental age and risk of sporadic and familial cancer in offspring: Implications for germ cell mutagenesis*. *Epidemiology*, 1999. **10**(6): p. 747-751.
55. Merrill, R.M., et al., *Cancer risk associated with early and late maternal age at first birth*. *Gynecologic oncology*, 2005. **96**(3): p. 583-593.
56. Kharazmi, E., et al., *Familial risk of early and late onset cancer: nationwide prospective cohort study*. *Bmj*, 2012. **345**: p. e8076.
57. Brandt, A., et al., *Age of onset in familial cancer*. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 2008. **19**(12): p. 2084-8.
58. Kharazmi, E., et al., *Effect of multiplicity, laterality, and age at onset of breast cancer on familial risk of breast cancer: a nationwide prospective cohort study*. *Breast Cancer Res Treat*, 2014. **144**(1): p. 185-92.
59. Collaborative Group on Hormonal Factors in Breast, C., *Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease*. *Lancet*, 2001. **358**(9291): p. 1389-99.
60. Bermejo, J.L. and K. Hemminki, *Familial risk of cancer shortly after diagnosis of the first familial tumor*. *J Natl Cancer Inst*, 2005. **97**(21): p. 1575-9.
61. Weiss, K.M. and J.D. Terwilliger, *How many diseases does it take to map a gene with SNPs?* *Nat Genet*, 2000. **26**(2): p. 151-157.
62. Vogelstein, B. and K.W. Kinzler, *The genetic basis of human cancer*. 2nd ed2002, New York: McGraw-Hill, Medical Pub. Division. xv, 821 p.
63. Houlston, R.S. and J. Peto, *The future of association studies of common cancers*. *Human genetics*, 2003. **112**(4): p. 434-5.
64. Isaacs, S.D., et al., *Risk of cancer in relatives of prostate cancer probands*. *J Natl Cancer Inst*, 1995. **87**(13): p. 991-6.
65. Lindstrom, L.S., et al., *Familial concordance in cancer survival: a Swedish population-based study*. *The Lancet. Oncology*, 2007. **8**(11): p. 1001-6.
66. Li, J., et al., *2q36.3 is associated with prognosis for oestrogen receptor-negative breast cancer patients treated with chemotherapy*. *Nat Commun*, 2014. **5**: p. 4051.
67. Lindstrom, L.S., et al., *Prognostic information of a previously diagnosed sister is an independent prognosticator for a newly diagnosed sister with breast cancer*. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 2014. **25**(10): p. 1966-72.
68. Antoniou, A.C., et al., *A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population*. *Nat Genet*, 2010. **42**(10): p. 885-92.
69. Haiman, C.A., et al., *A common variant at the TERT-CLPTMIL locus is associated with estrogen receptor-negative breast cancer*. *Nat Genet*, 2011. **43**(12): p. 1210-U61.

70. Broeks, A., et al., *Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium*. Hum Mol Genet, 2011. **20**(16): p. 3289-303.
71. Stacey, S.N., et al., *Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer*. Nat Genet, 2008. **40**(6): p. 703-6.
72. Stacey, S.N., et al., *Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer*. Nat Genet, 2007. **39**(7): p. 865-9.
73. Michailidou, K., et al., *Large-scale genotyping identifies 41 new loci associated with breast cancer risk*. Nat Genet, 2013. **45**(4): p. 353-61, 361e1-2.
74. Apetoh, L., et al., *Molecular interactions between dying tumor cells and the innate immune system determine the efficacy of conventional anticancer therapies*. Cancer research, 2008. **68**(11): p. 4026-4030.
75. Denkert, C., et al., *Tumor-Associated Lymphocytes As an Independent Predictor of Response to Neoadjuvant Chemotherapy in Breast Cancer*. Journal of Clinical Oncology, 2010. **28**(1): p. 105-113.
76. Loi, S., et al., *Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98*. J Clin Oncol, 2013. **31**(7): p. 860-7.
77. Gustafsson, P.E., M. Persson, and A. Hammarstrom, *Socio-economic disadvantage and body mass over the life course in women and men: results from the Northern Swedish Cohort*. Eur J Public Health, 2012. **22**(3): p. 322-7.
78. Lagerlund, M., et al., *Socio-economic factors and breast cancer survival--a population-based cohort study (Sweden)*. Cancer causes & control : CCC, 2005. **16**(4): p. 419-30.
79. Ji, J., et al., *Survival in ovarian cancer patients by histology and family history*. Acta oncologica, 2008. **47**(6): p. 1133-9.
80. Zweemer, R.P., et al., *Survival analysis in familial ovarian cancer, a case control study*. European journal of obstetrics, gynecology, and reproductive biology, 2001. **98**(2): p. 219-23.
81. Egbers, L., et al., *The prognostic value of family history among patients with urinary bladder cancer*. Int J Cancer, 2015. **136**(5): p. 1117-24.
82. Ji, J., et al., *Survival in bladder and renal cell cancers is familial*. Journal of the American Society of Nephrology : JASN, 2008. **19**(5): p. 985-91.
83. Ekblom, A., *The Swedish Multi-generation Register*. Methods Mol Biol, 2011. **675**: p. 215-20.
84. Ludvigsson, J.F., et al., *The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research*. Eur J Epidemiol, 2009. **24**(11): p. 659-67.
85. Socialstyrelsen, *Swedish Cancer Registry*.
86. Bhargava, R. and B.I. Dalal, *Two steps forward, one step back: 4th WHO classification of myeloid neoplasms (2008)*. Indian journal of pathology & microbiology, 2010. **53**(3): p. 391-4.

87. Harris, N.L., et al., *The World Health Organization classification of hematological malignancies report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November 1997*. *Modern Pathol*, 2000. **13**(2): p. 193-207.
88. Sobin, L.H., et al., *TNM classification of malignant tumours*. 7th ed2010, Chichester, West Sussex, UK ; Hoboken, NJ: Wiley-Blackwell. xx, 309 p.
89. Center for Epidemiology, *Cancer incidence in Sweden 2011, 2012*, The National Board of Health and Welfare: Stockholm.
90. Mattsson, B. and A. Wallgren, *Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978*. *Acta Radiol Oncol*, 1984. **23**(5): p. 305-13.
91. Barlow, L., et al., *The completeness of the Swedish Cancer Register: a sample survey for year 1998*. *Acta Oncol*, 2009. **48**(1): p. 27-33.
92. Socialstyrelsen. *Cause of Death Register*.
93. Johansson, L.A., C. Bjorkenstam, and R. Westerling, *Unexplained differences between hospital and mortality data indicated mistakes in death certification: an investigation of 1,094 deaths in Sweden during 1995*. *J Clin Epidemiol*, 2009. **62**(11): p. 1202-9.
94. Johansson, L.A. and R. Westerling, *Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics*. *Int J Epidemiol*, 2000. **29**(3): p. 495-502.
95. Nystrom, L., et al., *Determination of Cause of Death among Breast-Cancer Cases in the Swedish Randomized Mammography Screening Trials - a Comparison between Official Statistics and Validation by an End-Point Committee*. *Acta oncologica*, 1995. **34**(2): p. 145-152.
96. Johannesson, L., *The total population register of statistics Sweden new possibilities and better quality*. International Conference on Politics and Information Systems: Technologies and Applications, Vol 2, 2004: p. 6-11.
97. Verhage, M., *Population and Housing censuses based on a dwelling register and registration on dwellings*, 2010, Statistics Sweden.
98. Czene, K., P. Lichtenstein, and K. Hemminki, *Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database*. *Int J Cancer*, 2002. **99**(2): p. 260-6.
99. Clayton, D.G., *Model for Association in Bivariate Life Tables and Its Application in Epidemiological-Studies of Familial Tendency in Chronic Disease Incidence*. *Biometrika*, 1978. **65**(1): p. 141-151.
100. Efron, B. and R. Tibshirani, *An introduction to the bootstrap*1998, Boca Raton ; London: Chapman & Hall/CRC. xvi, 436.
101. Pfeiffer, R.M., et al., *Methods for testing familial aggregation of diseases in population-based samples: application to Hodgkin lymphoma in Swedish registry data*. *Ann Hum Genet*, 2004. **68**(Pt 5): p. 498-508.
102. Hughes, D. *What Does '5-Year Survival' in Breast Cancer Really Mean?* [cited 2015 2015-03-31]; Available from: <http://www.cancertherapyadvisor.com/breast-cancer-five-year-survival-patient-treatment/article/405934/2/>.

103. Capocaccia, R., G. Gatta, and L. Dal Maso, *Life expectancy of colon, breast and testicular cancer patients. An analysis of US-SEER population-based data*. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 2015.
104. Sellick, G.S., et al., *A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia*. *Blood*, 2007. **110**(9): p. 3326-33.
105. Paltiel, O., et al., *The incidence of lymphoma in first-degree relatives of patients with Hodgkin disease and non-Hodgkin lymphoma: results and limitations of a registry-linked study*. *Cancer*, 2000. **88**(10): p. 2357-66.
106. Olson, J.M. and H.J. Cordell, *Ascertainment bias in the estimation of sibling genetic risk parameters*. *Genetic epidemiology*, 2000. **18**(3): p. 217-35.
107. Leu, M., M. Reilly, and K. Czene, *Evaluation of bias in familial risk estimates: a study of common cancers using Swedish population-based registers*. *J Natl Cancer Inst*, 2008. **100**(18): p. 1318-25.
108. Leu, M., K. Czene, and M. Reilly, *The impact of truncation and missing family links in population-based registers on familial risk estimates*. *Am J Epidemiol*, 2007. **166**(12): p. 1461-7.
109. Goldin, L.R., et al., *Familial aggregation of Hodgkin lymphoma and related tumors*. *Cancer*, 2004. **100**(9): p. 1902-8.
110. Hemminki, K. and X. Li, *Familial risk of cancer by site and histopathology*. *Int J Cancer*, 2003. **103**(1): p. 105-9.
111. Collett, D., *Modelling survival data in medical research*. 2nd ed. Chapman & Hall/CRC texts in statistical science series 2003, Boca Raton, Fla. ; London: Chapman & Hall/CRC. 391 p.
112. Cox, D.R., *Regression Models and Life-Tables*. *J R Stat Soc B*, 1972. **34**(2): p. 187-+.
113. Schoenfeld, D., *Partial Residuals for the Proportional Hazards Regression-Model*. *Biometrika*, 1982. **69**(1): p. 239-241.
114. Lee, M., et al., *A unified model for estimating and testing familial aggregation*. *Stat Med*, 2013. **32**(30): p. 5353-65.
115. Marubini, E. and M.G. Valsecchi, *Analysing survival data from clinical trials and observational studies*. *Statistics in practice 1995*, Chichester ; New York: John Wiley & Sons. xvi, 414 p.
116. Rebora, P., Salim, A., Reilly, M., *bshazard: A Flexible Tool for Nonparametric Smoothing of the Hazard Function*. *The R Journal*, 2014. **6/2**: p. 114-22.
117. Rebora, P., K. Czene, and M. Reilly, *Timing of familial breast cancer in sisters*. *J Natl Cancer Inst*, 2008. **100**(10): p. 721-7.
118. R Development Core Team, *R: A language and environment for statistical computing*, 2015, R Foundation for Statistical Computing: Vienna, Austria.
119. Royston, P. and M.K. Parmar, *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects*. *Stat Med*, 2002. **21**(15): p. 2175-97.

120. Smith, P.L., *Splines as a Useful and Convenient Statistical Tool*. Am Stat, 1979. **33**(2): p. 57-62.
121. Durrleman, S. and R. Simon, *Flexible regression models with cubic splines*. Stat Med, 1989. **8**(5): p. 551-61.
122. Royston, P. and P. Lambert, *Flexible parametric survival analysis using stata : beyond the Cox model* 2011, College Station, Tex.: Stata. 344 p.
123. Lambert, P.C. and P. Royston, *Further development of flexible parametric models for survival analysis*. Stata J, 2009. **9**(2): p. 265-290.
124. Stata Corporation., *Stata base reference manual*. Release 13. ed 2013, College Station, TX: Stata Corporation.
125. Hosmer, D.W. and S. Lemeshow, *Applied logistic regression*. Wiley series in probability and mathematical statistics. Applied probability and statistics 1989, New York ; Chichester: Wiley. ix, 307 p.
126. Goldin, L.R., et al., *Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia*. Haematologica, 2009. **94**(5): p. 647-53.
127. Hemminki, K., P. Vaitinen, and P. Kyyronen, *Age-specific familial risks in common cancers of the offspring*. Int J Cancer, 1998. **78**(2): p. 172-5.
128. Hemminki, K. and P. Mutanen, *Parental cancer as a risk factor for nine common childhood malignancies*. Brit J Cancer, 2001. **84**(7): p. 990-993.
129. Conter, V., et al., *Long-term results of the Italian Association of Pediatric Hematology and Oncology (AIEOP) Studies 82, 87, 88, 91 and 95 for childhood acute lymphoblastic leukemia*. Leukemia, 2010. **24**(2): p. 255-64.
130. Talback, M., et al., *Cancer patient survival in Sweden at the beginning of the third millennium - predictions using period analysis*. Cancer Cause Control, 2004. **15**(9): p. 967-976.
131. Bhayat, F., et al., *The incidence of and mortality from leukaemias in the UK: a general population-based study*. BMC cancer, 2009. **9**: p. 252.
132. Goldin, L.R., et al., *Common occurrence of monoclonal B-cell lymphocytosis among members of high-risk CLL families*. British journal of haematology, 2010. **151**(2): p. 152-8.
133. Tegg, E.M., et al., *Evidence for a common genetic aetiology in high-risk families with multiple haematological malignancy subtypes*. British journal of haematology, 2010. **150**(4): p. 456-62.
134. Crowther-Swanepoel, D., et al., *Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk*. Nat Genet, 2010. **42**(2): p. 132-U59.
135. Crowther-Swanepoel, D., et al., *Verification that common variation at 2q37.1, 6p25.3, 11q24.1, 15q23, and 19q13.32 influences chronic lymphocytic leukaemia risk*. British journal of haematology, 2010. **150**(4): p. 473-479.
136. Crowther-Swanepoel, D., et al., *Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk*. British journal of haematology, 2011. **154**(2): p. 229-233.

137. Slager, S.L., et al., *Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL*. Blood, 2011. **117**(6): p. 1911-1916.
138. Goldin, L.R. and S.L. Slager, *Familial CLL: genes and environment*. Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program, 2007: p. 339-45.
139. Rebora, P., et al., *High risks of familial chronic lymphatic leukemia for specific relatives: signposts for genetic discovery?* Leukemia, 2012. **26**(11): p. 2419-21.
140. Trent, S. and W. Davies, *The influence of sex-linked genetic mechanisms on attention and impulsivity*. Biol Psychol, 2012. **89**(1): p. 1-13.
141. Czene, K., H.O. Adami, and E.T. Chang, *Sex- and kindred-specific familial risk of non-Hodgkin's lymphoma*. Cancer Epidemiol Biomarkers Prev, 2007. **16**(11): p. 2496-9.
142. Goldin, L.R., et al., *Familial aggregation and heterogeneity of non-Hodgkin lymphoma in population-based samples*. Cancer Epidemiol Biomarkers Prev, 2005. **14**(10): p. 2402-6.
143. McHale, S.M., A.C. Crouter, and C.J. Tucker, *Family context and gender role socialization in middle childhood: comparing girls to boys and sisters to brothers*. Child development, 1999. **70**(4): p. 990-1004.
144. Rust, J., et al., *The role of brothers and sisters in the gender development of preschool children*. Journal of experimental child psychology, 2000. **77**(4): p. 292-303.
145. Hemminki, K., J. Sundquist, and J.L. Bermejo, *Familial risks for cancer as the basis for evidence-based clinical referral and counseling*. Oncologist, 2008. **13**(3): p. 239-247.
146. Hemminki, K. and B. Chen, *Familial risks for colorectal cancer show evidence on recessive inheritance*. International Journal of Cancer, 2005. **115**(5): p. 835-838.
147. Johns, L.E. and R.S. Houlston, *A systematic review and meta-analysis of familial colorectal cancer risk*. Am J Gastroenterol, 2001. **96**(10): p. 2992-3003.
148. Kicinski, M., J. Vangronsveld, and T.S. Nawrot, *An Epidemiological Reappraisal of the Familial Aggregation of Prostate Cancer: A Meta-Analysis*. Plos One, 2011. **6**(10).
149. Hemminki, K., I. Lonnstedt, and P. Vaittinen, *A population-based study of familial cutaneous melanoma*. Melanoma Res, 2001. **11**(2): p. 133-140.
150. Jonsson, H., et al., *Uptake of prostate-specific antigen testing for early prostate cancer detection in Sweden*. Int J Cancer, 2011. **129**(8): p. 1881-8.
151. Haukka, J., et al., *Trends in Breast Cancer Mortality in Sweden before and after Implementation of Mammography Screening*. Plos One, 2011. **6**(9).
152. Boardman, L.A., et al., *Colorectal cancer risks in relatives of young-onset cases: Is risk the same across all first-degree relatives?* Clin Gastroenterol H, 2007. **5**(10): p. 1195-1198.
153. Ericson, K., et al., *Familial risk of tumors associated with hereditary non-polyposis colorectal cancer: a Swedish population-based study*. Scand J Gastroentero, 2004. **39**(12): p. 1259-1265.

154. Bermejo, J.L., C. Eng, and K. Hemminki, *Cancer characteristics in Swedish families fulfilling criteria for hereditary nonpolyposis colorectal cancer*. *Gastroenterology*, 2005. **129**(6): p. 1889-1899.
155. Rosato, V., et al., *Risk factors for young-onset colorectal cancer*. *Cancer Cause Control*, 2013. **24**(2): p. 335-341.
156. Choi, D.H., et al., *Incidence of BRCA1 and BRCA2 mutations in young Korean breast cancer patients*. *J Clin Oncol*, 2004. **22**(9): p. 1638-45.
157. de Sanjose, S., et al., *Prevalence of BRCA1 and BRCA2 germline mutations in young breast cancer patients: A population-based study*. *International Journal of Cancer*, 2003. **106**(4): p. 588-593.
158. Draisma, G., et al., *Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context*. *J Natl Cancer Inst*, 2009. **101**(6): p. 374-383.
159. Shapiro, S., J.D. Goldberg, and Hutchiso.Gb, *Lead Time in Breast-Cancer Detection and Implications for Periodicity of Screening*. *Am J Epidemiol*, 1974. **100**(5): p. 357-366.
160. Autier, P., et al., *Mammography Screening and Breast Cancer Mortality in Sweden*. *J Natl Cancer Inst*, 2012. **104**(14): p. 1080-1093.
161. Allen-Brady, K., et al., *Lobular breast cancer: excess familiarity observed in the Utah Population Database*. *Int J Cancer*, 2005. **117**(4): p. 655-61.
162. Kotsopoulos, J., et al., *Risk factors for ductal and lobular breast cancer: results from the nurses' health study*. *Breast cancer research : BCR*, 2010. **12**(6): p. R106.
163. Phipps, A.I., et al., *Risk factors for ductal, lobular, and mixed ductal-lobular breast cancer in a screening population*. *Cancer Epidemiol Biomarkers Prev*, 2010. **19**(6): p. 1643-54.
164. Ellis, I.O., et al., *Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up*. *Histopathology*, 1992. **20**(6): p. 479-89.
165. Lee, M., et al., *Patterns of changing cancer risks with time since diagnosis of a sibling*. *International Journal of Cancer*, 2015. **136**(8): p. 1948-1956.
166. Tournier, I., et al., *Germline mutations of inhibins in early-onset ovarian epithelial tumors*. *Hum Mutat*, 2014. **35**(3): p. 294-7.
167. Lorenzo Bermejo, J., R. Rawal, and K. Hemminki, *Familial association of specific histologic types of ovarian malignancy with other malignancies*. *Cancer*, 2004. **100**(7): p. 1507-14.
168. Hemminki, K., J. Sundquist, and A. Brandt, *Incidence and mortality in epithelial ovarian cancer by family history of any cancer*. *Cancer*, 2011. **117**(17): p. 3972-80.
169. Jones, S.C., et al., *Australian women's awareness of ovarian cancer symptoms, risk and protective factors, and estimates of own risk*. *Cancer causes & control : CCC*, 2010. **21**(12): p. 2231-9.
170. Howlander N, N.A., Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA



- (eds), *SEER Cancer Statistics Review, 1975-2011*, 2014, National Cancer Institute: Bethesda, MD.
171. Bjorkholm, M., et al., *No familial aggregation in chronic myeloid leukemia*. *Blood*, 2013. **122**(3): p. 460-1.
  172. Domingo-Domenech, E., et al., *Family clustering of blood cancers as a risk factor for lymphoid neoplasms*. *Haematologica*, 2005. **90**(3): p. 416-8.
  173. Goldin, L.R., et al., *Familial aggregation of acute myeloid leukemia and myelodysplastic syndromes*. *J Clin Oncol*, 2012. **30**(2): p. 179-83.
  174. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome*. *Nature*, 2009. **458**(7239): p. 719-24.
  175. Hemminki, K., et al., *Familial risks for common diseases: etiologic clues and guidance to gene identification*. *Mutation research*, 2008. **658**(3): p. 247-58.
  176. Sandin, S., et al., *The familial risk of autism*. *Jama*, 2014. **311**(17): p. 1770-7.
  177. Chen, R., et al., *Parental cancer diagnosis and child mortality--a population-based cohort study in Sweden*. *Cancer epidemiology*, 2015. **39**(1): p. 79-85.
  178. Malati, T., *Tumour markers: An overview*. *Indian journal of clinical biochemistry : IJCB*, 2007. **22**(2): p. 17-31.
  179. Hopper, J.L., D.T. Bishop, and D.F. Easton, *Population-based family studies in genetic epidemiology*. *Lancet*, 2005. **366**(9494): p. 1397-406.