The Nation-wide Swedish Family-Cancer Database

Updated Structure and Familial Rates

Kari Hemminki, Xinjun Li, Kamila Plna, Charlotta Granström and Pauli Vaittinen

From the Department of Biosciences, Karolinska Institute, Novum, Huddinge, Sweden

Correspondence to: Kari Hemminki, Department of Biosciences, Karolinska Institute, Novum, SE-141 57 Huddinge, Sweden. Tel: +46 8 608 9243. Fax: +46 8 608 1501. E-mail: Kari.Hemminki@cnt.ki.se

Acta Oncologica Vol. 40, No. 6, pp. 772–777, 2001

The Swedish Family-Cancer Database was expanded to include all Swedes born in 1932 and later (offspring) with their parents, totaling 10.2 million individuals. Cancer cases were retrieved from the Swedish Cancer Registry from the years 1958 to 1998, including over 1 million primary cancers and in situ tumors. Some 10% of offspring diagnosed with cancer lack any parental information. Incidence rates of cancers were similar in the database and in the Cancer Registry to age 70, but at higher ages the rates in the Database were lower, probably because of selection. The familial risk for all types of cancer in offspring was 1.73 when a parent had the same type of cancer. The familial rates were increased for all main cancer sites, except for the upper aerodigestive tract, stomach, liver, pancreas and bone marrow (leukemia). The rates were 7.47 for thyroid, 4.69 for testis, and over 2.00 for melanoma, ovary, prostate, skin, endocrine glands and endometrium.

Received 22 February 2001
Accepted 31 May 2001

Cancer is a multifactorial disease, caused by environmental and inherited events which transform a benign cell into a malignant cell through a sequence of molecular changes (1, 2). Most cancers are sporadic but some 1–5% are caused by single-gene, dominant traits (3). There are no estimates for the contribution of polygenic or recessive conditions on cancer but studies of twins suggest that such effects are important (4, 5). Familial aggregation of cancer may be due either to environmental factors shared by family members or to shared genes (6). Familial clustering has been an avenue to the understanding of cancer etiology and has been the basis for clinical decisions and counseling, as well as guiding the identification of cancer-related genes.

Traditionally, familial clustering of cancer has been studied in the clinical setting where probands and their multiple affected family members have been identified (3, 7). This approach has also been productive in terms of understanding cancer genetics. Many forms of cancer in which a single gene poses a high risk have been identified. The disadvantages include difficulties in obtaining large numbers of cases and in securing unbiased risk estimates. Estimation of risks at sites other than the index site has also been cumbersome. Of the 4 700 dominant and 2 800 recessive human genetic traits known in the early 1990s (8), some 440 were single-gene traits in which cancer was a complication; many of them were extremely rare, with only a few identified families world-wide (9). Clinical observation probably works for dominant diseases when these pose a risk of between 10 and 100, or more, above the population rate of the disease. For recessive conditions, clinical observation is less sensitive, and most of the results on recessive conditions have come from isolated populations with high rates of consanguineous marriage. Formal epidemiological studies have had little impact on defining new familial traits. Analytical epidemiological studies have been used for either hypothesis generating or quantifying the known risks. The Utah Population Database has also been used in quantifying known risks (10) or in gene identification. Studies of twins offer a third alternative for the genetic epidemiology of cancer. Dissection of heritable and environmental components is possible in such studies and the risk estimates should be robust. However, because of the rareness of twinning, even the large twin registers allow for risk estimation at the most common sites only (5). Few population-based studies for the identification of recessive cancers have been undertaken, which probably explains the current lack of knowledge in this area (11).

Statistics Sweden has maintained a nation-wide family register since 1995 and this has been a unique source for genetic epidemiological studies. We combined this register with the Swedish Cancer Registry to form the Swedish Family-Cancer Database for the first time, in 1996. In the beginning of 2001, after some 70 papers had been published or submitted, the fourth version of the Database

 ISSN 0284-186X

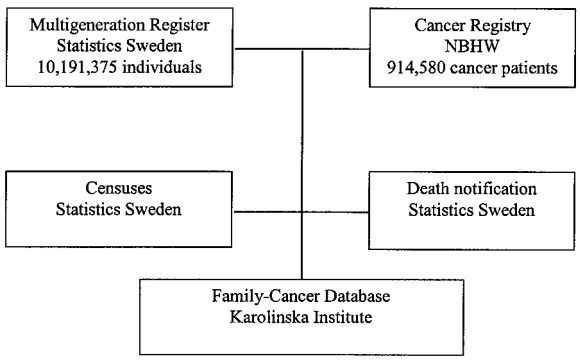

Fig. 1. the data linkage for the Family-Cancer Database.

was ready to be used. In the present article we describe some details of the structure of the Database and report the current familial risks for common cancers.

MATERIALS AND METHODS

Statistics Sweden maintains a 'Multigeneration Register' where children (offspring) born in Sweden in 1932 and later, are registered with their biological parents (those pleading parenthood at birth) as families. The data are complete, barring registration mistakes and inaccuracies, which mainly affect those born in the 1930s, and some groups of emigrants and immigrants. In addition, linkage to parents is incomplete in some groups of deceased offspring, as shown in the Results section of this article. This register was linked by the individually unique national registration number to the Cancer Registry from the years 1958–1998 (Fig. 1). Cancer registration is considered to be close to 100% currently (12). A four-digit diagnostic code in accordance with the 7th revision of the International Classification of Diseases (ICD-7) and subsequent ICD classifications is available. Cancers are also recorded according to the first or subsequent primary cancer, and cancer in situ. Additional linkages were carried out both to the national census data to obtain socioeconomic background data and to death notifications, for vital status determination. These were done at Statistics Sweden. Those who did not have a national registration number could not be linked. In the final matched records, these numbers were removed so that individuals cannot be identified in the Family-Cancer Database.

Familial risks were calculated for offspring whose parents had the same, concordant cancer, i.e., using parents as probands. In the Family-Cancer Database ageing offspring become parents in due course. Such individuals are considered independently, first as offspring, then as parents. Concordant cancers are extremely rare in three generations, and the dependence between the individuals has not been of concern in analysis. Standardized incidence ratios (SIRs) were calculated as the ratio of observed (O) to expected (E) number of cases. The expected numbers were calculated from 5-year age-, sex- and tumor-type-specific standard incidence rates (13). Confidence intervals (95% CI) were calculated assuming a Poisson distribution (13).

RESULTS

A scheme for linking the Multigeneration Register on over 10 million individuals with the Swedish Cancer Registry, national census data and death notifications is presented in Fig. 1. The linkage between the Multigeneration Register and Cancer Registry resulted in the identification of 914 580 patients with cancer. These included 602 663 first primary cancers among the parental (first) generation and 157 777 first primary cancers among the offspring (second) generation (Table 1). The father/son and mother/daughter lines in Table 1 indicate those offspring who have become parents in the course of follow-up. The number of multiple primaries was 67 178 (8.8% of the first cancers) and that of in situ cancers 216 567.

The longitudinal structure of the Family-Cancer Database is shown in Fig. 2. The Database contains offspring who were born in Sweden after 1931, with their parents registered at the time of birth. Some mothers and a larger number of fathers were born in the 1800s, the oldest father born in 1864. The annual accumulation of cancer cases to the parental and offspring generation in the Family-Cancer Database is shown in Fig. 3. The offspring generation has far fewer cases than the parental generation but the increase among the offspring generation is steep owing to advancing age. The sum of parental and offspring cancers is larger than the total sum of cases because older off-

Table 1

Number of cancer notifications for first and multiple primary invasive and in situ cancers in the Family-Cancer Database, 1958–1998

	First primary	Multiple primary	In situ	All
Father	318 878	30 533	27 308	376 719
Mother	283 785	28 299	64 318	376 402
Father/son	35 766	1 530	3 986	41 282
Mother/daughter	63 376	3 887	97 688	164 951
Son	27 485	1 056	1 710	30 251
Daughter	31 150	1 873	21 557	54 580
All	760 440	67 178	216 567	1 044 185

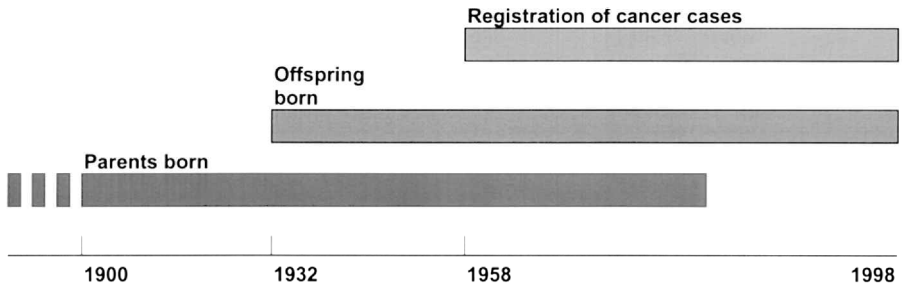

Fig. 2. Longitudinal structure of the populations and cancer registration in the Family-Cancer Database. The oldest parent is from 1864. The oldest offspring are from 1932. The Swedish Cancer Registry was started in 1958.

spring have become parents and contribute cases to both groups.

The linkage of offspring to their parents was partially incomplete among the deceased individuals, particularly among those who died before the 1990s. According to Table 2, 2.2% of the more than 7 million offspring had cancer, and 97.6% had a link to at least one parent. The linkage was 98.8% among those alive as of the end of 1998. Among the deceased 216 550 individuals, the linkage existed among 71.3%. There was a shortage of parental information on 15 560 deceased offspring, who had been diagnosed with cancer; this was 7.2% of the deceased offspring and implied that 9.8% of all offspring with cancer had no links to parents. Offspring who died before 1960 are missing from the Database.

The incidence of cancer in the Family-Cancer Database, generated by person-years calculated within the Database, was compared with the incidence in the Cancer Registry (Fig. 4). For all cancers, the incidences in the Database and the Cancer Registry were very similar up to about age 70 but thereafter the rate in the Database dropped in relation to the Cancer Registry (Fig. 4A). For pancreatic cancer, a fatal cancer, the same relationships held but the difference between the two sources was larger, particularly among the 80–84-year-olds (Fig. 4C). For squamous cell skin cancer (a cancer with a good prognosis) and breast cancer, the incidence rates did not differ much between the two databases (Fig. 4B and D). Of course, the data could have been analyzed by adjusting for birth cohort, period, region and other possible variables that might affect the incidence of cancer. However, the nearly superimposable age-incidence curves up to age 70 between the Cancer Registry and the Family-Cancer Database show that the rates from the two sources are fully comparable.

Familial risks were calculated for offspring by a concordant parental cancer, as defined by ICD7, for sites at which ≥ 15 cases were recorded/gender (Table 3). Among all offspring, all the familial risks were significantly increased, except for SIRs of the stomach, liver, pancreas and leukemia. Thyroid cancer showed the highest familial risk, at 7.47, followed by cancer of the testis (4.69), melanoma (2.57), ovary (2.34), prostate (2.20), endocrine glands (2.17) and endometrium (2.12). When both offspring and their parents were diagnosed before the age of 50, almost all SIRs were increased, thyroid exceeding a SIR of 20, and colon, rectum, endometrium, testis and endocrine glands exceeding a SIR of 5.

DISCUSSION

In 1995, Statistics Sweden created a family database, the 'Second Generation Register' (14). This initially included offspring born in Sweden in 1941 with their biological parents as families, 6 million individuals. It was expanded in 1998 to 9.6 million individuals, and a further current expansion amounting to over 10 million individuals born in Sweden at the end of 2000. The size of the population database is respectable in a country with a current population of 9 million, of which almost 10% are foreign-born. Any expansion to past generations would be laborious. This expansion in the year 2000 included offspring born after 1931, along with their parents, at which time the database was renamed as the 'Multigeneration Register'. We have linked the Second Generation Register to the Swedish Cancer Registry (started in 1958) to make the Family-Cancer Database in three earlier expanded versions

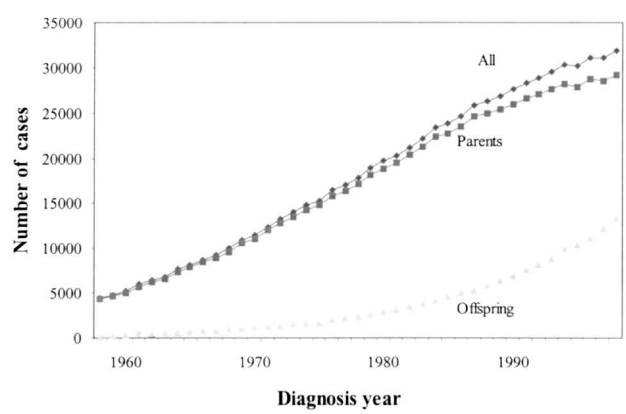

Fig. 3. Annual accumulation of cancer cases to the parental and offspring generations of the Family-Cancer Database. The sum of 'offspring' and 'parents' exceeds 'all' because some offspring have become parents and their cancers are included among parents' cancers.

Table 2

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

	Total no. of offspring	Offspring with cancer		Offspring linked to parent		Offspring with cancer not linked to parent	
	n	n	%	n	%	n	%
All offspring	7 017 638	157 779	2.2	6 846 587	97.6	22 271	0.3
Living offspring	6 801 088	102 153	1.5	6 692 209	98.4	6 711	0.1
Deceased offspring	216 550	55 626	25.7	154 378	71.3	15 560	7.2

in 1996, 1997 and 1999. The number of cancers in the second generation increased from 20 000 in 1996 to 92 000 in 1999; in the parental generation the increase was from 500 000 to 600 000 invasive cancers. The Family-Cancer Database is the largest population-based data set ever used for studies on familial cancer. In comparison, The Utah Population Database, successfully used in many cancer studies, has a different structure in that it contains more than two generations, and 42 000 cancers in 1994 (10). Registers can be linked in other counties as well, but this usually tends to be an ad hoc effort and the applications have remained limited. Such linkages are possible, for example, in the Nordic countries and in the UK (15–19).

Regarding the Family-Cancer Database, it is worth pointing out that the parents are registered at the time of birth of the child. Thus it is possible to track 'biological' parents, despite divorce and remarriage. The national personal registration number has now been deleted from the Database.

The present update of the Family-Cancer Database reached over 10 million individuals with over 1 million notifications of cancer—some three-quarters of all tumors in The Swedish Cancer Registry. The difference is mainly due to the segment of the population that had no children after 1931. By definition, they are not in the Family-Cancer Database but they would be recorded in the Cancer Registry if their cancers were diagnosed since 1958. Because we calculate all incidence rates based on the population in the Family-Cancer Database, the results are unbiased, as discussed later. Cancers in the second generation amounted to 155 000 and were far fewer than those in the parental generation, because of age distribution. The highest age in the offspring generation was 66 years, while in the parental generation there was no truncation. The controls that we have carried out have shown a good quality of data. The problem of linking deceased offspring to their parents, which has existed throughout the history of the Second Generation Register, has been partially

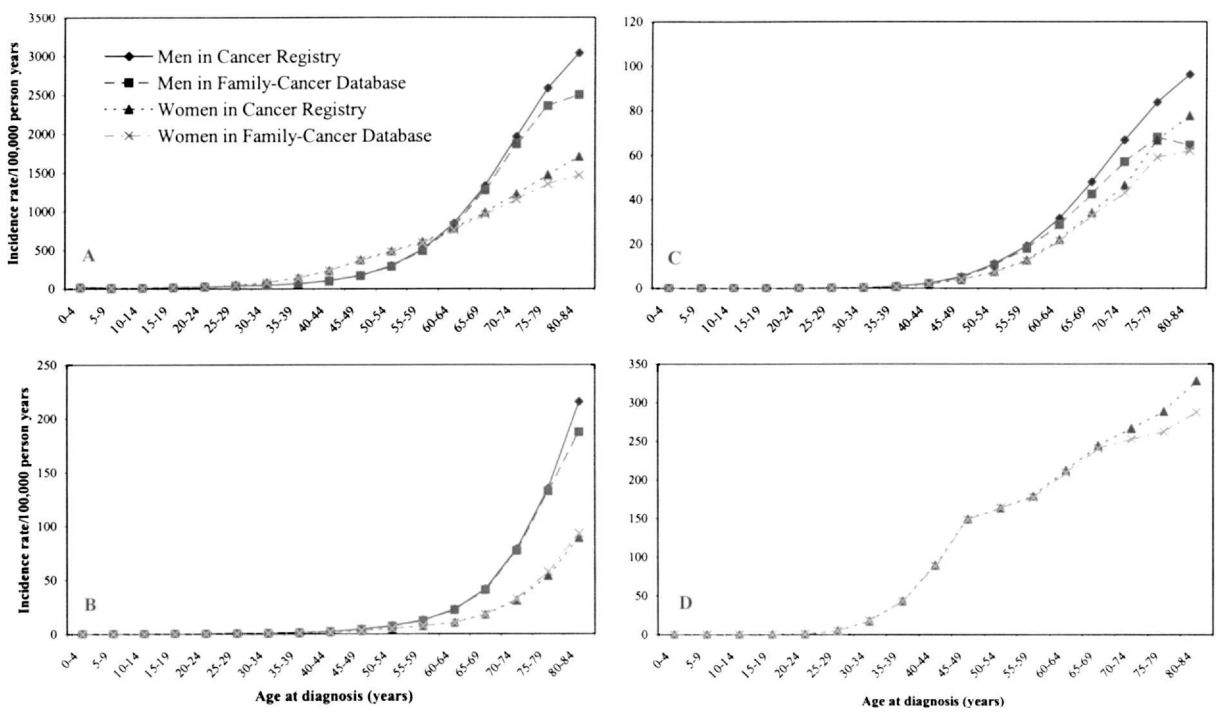

Fig. 4. Age-specific incidence rates of all (A), skin (B), pancreas (C) and female breast cancer (D) in the Family-Cancer Database and in the Cancer Registry, 1958–1998.

Table 3

SIR for concordant cancer in offspring by parental cancer

Cancer sites	Offspring at all ages					Parent's and offspring's age at diagnosis <50 years				
	O	E	SIR	95%CI		O	E	SIR	95%CI	
Upper aerodigestive tract	37	30.45	1.21	0.86	1.64		0.81			
Stomach	66	62.58	1.05	0.82	1.32	3	1.36	2.21	0.42	5.42
Colon	368	213.94	**1.72**	**1.55**	**1.90**	28	4.66	**6.01**	**3.99**	**8.44**
Rectum	107	69.63	**1.54**	**1.26**	**1.84**	8	0.94	**8.47**	**3.62**	**15.36**
Liver	31	31.82	0.97	0.66	1.35		0.31			
Pancreas	38	31.74	1.20	0.85	1.61		0.28			
Lung	305	218.81	**1.39**	**1.24**	**1.55**	4	2.41	1.66	0.43	3.69
Breast	2267	1344.06	**1.69**	**1.62**	**1.76**	247	97.93	**2.52**	**2.22**	**2.85**
Cervix	104	61.60	**1.69**	**1.38**	**2.03**	44	24.44	**1.80**	**1.31**	**2.37**
Endometrium	112	52.88	**2.12**	**1.74**	**2.53**	9	1.73	**5.21**	**2.36**	**9.17**
Ovary	135	57.77	**2.34**	**1.96**	**2.75**	22	5.28	**4.17**	**2.61**	**6.09**
Prostate	551	250.48	**2.20**	**2.02**	**2.39**		0.02			
Testis	15	3.20	**4.69**	**2.62**	**7.37**	12	1.95	**6.15**	**3.16**	**10.12**
Kidney	72	52.55	**1.37**	**1.07**	**1.70**	4	1.76	2.27	0.59	5.04
Urinary bladder	138	80.21	**1.72**	**1.45**	**2.02**	4	0.95	**4.21**	**1.09**	**9.34**
Melanoma	255	99.21	**2.57**	**2.26**	**2.90**	53	12.57	**4.22**	**3.16**	**5.43**
Skin	71	32.47	**2.19**	**1.71**	**2.73**		0.23			
Nervous system	173	124.27	**1.39**	**1.19**	**1.61**	50	22.19	**2.25**	**1.67**	**2.92**
Thyroid gland	71	9.51	**7.47**	**5.83**	**9.31**	42	1.96	**21.48**	**15.47**	**28.46**
Endocrine glands	64	29.51	**2.17**	**1.67**	**2.73**	17	2.29	**7.43**	**4.32**	**11.38**
Non-Hodgkin's lymphoma	87	60.21	**1.45**	**1.16**	**1.76**	1	2.97	0.34	0.00	1.32
Leukemia	66	59.03	1.12	0.86	1.40	5	6.07	0.82	0.26	1.70
Any sites	5198	3000.74	**1.73**	**1.69**	**1.78**	579	196.58	**2.95**	**2.71**	**3.19**

Bold type: 95%.
CI does not include 1.00.

resolved, as less than 10% of the deceased offspring with cancer now lack a link to a parent. The main problem has been with the introduction of the personal registration number, which took place in Sweden in 1947, to foreign-born individuals and to children in the early period. The proportion of offspring without links to parents in the current Database is so small that bias is unlikely in family studies.

Even the incidence rates of cancers in the Family-Cancer Database matched those in the Cancer Registry, except at ages over 70 years. The reason for the deviation in rates among the elderly is likely to be either population selection or a preferential loss of old individuals diagnosed with cancer. Population selection would be made on the basis of the inclusion criterion of individuals to the Database: they had to be parents to children born after 1931. A reproductive population is healthier than the total population, sometimes referred to as the 'healthy reproducer effect', possibly accounting for the deficit of cancers among the old (20). The possibility that elderly individuals with a cancer diagnosis were lost cannot be excluded because some offspring lacked parental information, and death from cancer could be one reason for the missing data.

The family data on the size and quality of the Swedish Family-Cancer Database can have numerous applications (14). We showed one application here, in giving estimates on familial risks among offspring by parental cancers. The familial rate for all cancers in offspring was 1.73 when a parent had the same kind of cancer. The familial risks were increased for all main cancer sites, with the exception of the upper aerodigestive tract, stomach, liver, pancreas and bone marrow (leukemia). The rates were 7.47 for thyroid, 4.69 for testis, and over 2.00 for melanoma, ovary, prostate, skin (squamous cell carcinoma), endocrine glands and endometrium. The results are in line with those from an earlier version of the Database, concerning a 0–53-year-old population (21), compared to the present 0–66-year-old population.

Familial risks can be calculated by any other probands or groups of probands. For example, analysis by sibling without an affected parent would be consistent with a recessive mode of inheritance or early childhood environmental effect, rarely studied in genetic epidemiology of cancer (11, 22). As another alternative, the cancer status of both parents can be considered, which may provide information on genetic and environmental interactions, also an area hardly touched upon outside the Family-Cancer Database (23, 24). For some of the common cancer sites, the number of affected individuals is so large, e.g. 2 267 offspring presenting with familial breast cancer, that many

types of definitions of probands can be used. The data on spouses and other non-biological family relationships allow an assessment to be made of the environmental effects on cancer (6, 25). The Database includes a follow-up period of 41 years, during which time two generations of families will have completed their propagation of children. Thus it is possible to assess the effects of parental age and family characteristics on offspring cancer risks, including the effects of family size and birth order (26). To some extent, it is possible to collect biological specimens of tumor samples from patients identified from the Database through hospitals. However, whether the Database can be used for molecular studies among a large number of patients remains to be seen.

ACKNOWLEDGEMENTS

Statistics Sweden and the Swedish Cancer Registry are thanked for their registers and willingness for collaboration. The study was supported by the Swedish Cancer Society and the Wallenberg Consortium North.

REFERENCES

1. Hanahan D, Weinberg R. The hallmarks of cancer. Cell 2000; 100: 57–70.
2. Hemminki K, Mutanen P. Genetic epidemiology of multistage carcinogenesis. Mut Res 2001; 473: 11–21.
3. Lynch H, Fusaro R, Lynch J. Hereditary cancer in adults. Cancer Detect Prev 1995; 19: 219–33.
4. Ahlbom A, Lichtenstein P, Malmström H, Feychting M, Hemminki K, Pedersen NL. Cancer in twins: genetic and nongenetic familial risk factors. J Natl Cancer Inst 1997; 89: 287–93.
5. Lichtenstein P, Holm N, Verkasalo P, et al. Environmental and heritable factors in the causation of cancer. N Engl J Med 2000; 343: 78–85.
6. Hemminki K, Dong C, Vaittinen P. Cancer risks to spouses and offspring in the Family-Cancer Database. Genet Epidemiol 2001; 20: 247–57.
7. Easton D. The inherited component of cancer. Br Med Bull 1994; 50: 527–35.
8. Vogel F, Motulsky A. Human genetics: problems and approaches. Heidelberg: Springer, 1996.
9. Mulvihill J, Davis S, Fromkin K. The catalogue of human genes predisposing to neoplasia. In: Weber W, Narid S, Mulvihill J, eds. Familial cancer management. Boca Raton: CRC Press, 1996: 203–37.
10. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst 1994; 86: 1600–7.
11. Hemminki K, Vaittinen P, Dong C, Easton D. Sibling risks in cancer: clues to recessive or X-linked genes? Br J Cancer 2001; 84: 388–91.
12. Centre for Epidemiology. Cancer incidence in Sweden 1997. Stockholm: 1999.
13. Esteve J, Benhamou E, Raymond L. Statistical methods in cancer research. Lyon: IARC, 1994.
14. Hemminki K. Genetic epidemiology: science and ethics on familial cancers. Acta Oncol 2001; 40: 439–44.
15. Olsen JH, Boice JD, Seersholm N, Bautz A, Fraumeni JJF. Cancer in the parents of children with cancer. N Engl J Med 1995; 333: 1594–9.
16. Carstensen B, Soll-Johanning H, Villadsen E, Söndergaard J, Lynge E. Familial aggregation of colorectal cancer in the general population. Int J Cancer 1996; 68: 428–35.
17. Sankila R, Olsen JH, Anderson H, et al. Risk of cancer among offspring of childhood-cancer survivors. N Engl J Med 1998; 338: 1339–44.
18. Peto J, Easton D, Matthews F, Ford D, Swerdlow A. Cancer mortality in relatives of women with breast cancer: the OPCS study. Int J Cancer 1996; 65: 275–83.
19. Easton D, Matthews F, Ford D, Swerdlow A, Peto J. Cancer mortality in relatives of women with ovarian cancer: the OPCS study. Int J Cancer 1996; 65: 284–94.
20. Hemminki K, Kyyrönen P. Parental age and risk of sporadic and familial cancer in offspring: implications for germ cell mutagenesis. Epidemiology 1999; 10: 747–51.
21. Hemminki K, Vaittinen P, Kyyrönen P. Age-specific familial risks in common cancers of the offspring. Int J Cancer 1998; 78: 172–5.
22. Dong C, Hemminki K. Modification of cancer risks in offspring by sibling and parental cancers from 2 112 616 nuclear families. Int J Cancer 2001; 91: 144–5023.
23. Hemminki K, Vaittinen P, Kyyrönen P. Modification of cancer risk in offspring by parental cancer. Cancer Causes Control 1999; 10: 125–9.
24. Plna K, Hemminki K. Re: High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutation-positive melanoma families. J Natl Cancer Inst 2001; 93: 323–4.
25. Hemminki K, Lönnstedt I, Vaittinen P, Lichtenstein P. Estimation of genetic and environmental components in colorectal and lung cancer and melanoma. Genet Epidemiol 2001; 20: 107–16.
26. Hemminki K, Mutanen P. Birth order, family size, and the risk of cancer in young and middle-aged adults. Br J Cancer 2001; 84: 1466–71.