

UC Irvine

UC Irvine Previously Published Works

Title

Analysis of Co-Aggregation of Cancer Based on Registry Data

Permalink

<https://escholarship.org/uc/item/49d437x4>

Journal

Public Health Genomics, 9(2)

ISSN

1662-4246

Authors

Matthews, Abigail G
Betensky, Rebecca A
Anton-Culver, Hoda
[et al.](#)

Publication Date

2006

DOI

10.1159/000091485

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Analysis of Co-Aggregation of Cancer Based on Registry Data

Abigail G. Matthews^{a,b} Rebecca A. Betensky^{a,b} Hoda Anton-Culver^c
Deborah Bowen^d Constance Griffin^e Claudine Isaacs^f Carol Kasten^g
Geraldine Mineau^h Susan Nayfield^g Joellen Schildkrautⁱ Louise Strong^j
Barbara Weber^k Dianne M. Finkelstein^{a,b}

^aMassachusetts General Hospital Biostatistics Center, and ^bDepartment of Biostatistics, Harvard School of Public Health, Boston, Mass., ^cUniversity of California, Irvine, Calif., ^dFred Hutchinson Cancer Research Center, Seattle, Wash., ^eJohns Hopkins University, Baltimore, Md., ^fGeorgetown University Lombardi Cancer Center, Washington, D.C., ^gNational Institutes of Health, Bethesda, Md., ^hUniversity of Utah, Salt Lake City, Utah, ⁱDuke University Medical Center, Durham, N.C., ^jUniversity of Texas MD Anderson Cancer Center, Houston, Tex., and ^kUniversity of Pennsylvania, Philadelphia, Pa., USA

Key Words

Familial aggregation · Association · Family study ·
Cancer Genetics Network

Abstract

Objective: An exploratory analysis of co-aggregation of cancers using registry-based data. **Methods:** We utilized sibships from over 18,000 families who had been recruited to the NCI-sponsored multi-institutional Cancer Genetics Network. The analysis assesses co-aggregation at the individual and family level and adjusts for ascertainment. **Results:** We found statistically significant familial co-aggregation of lung cancer with pancreatic (adjusted $p < 0.001$), prostate (adjusted $p < 0.003$), and colorectal cancers (adjusted $p = 0.004$). In addition, we found significant familial co-aggregation of pancreatic and colorectal cancers (adjusted $p = 0.018$), and co-aggregation of hematopoietic and (non-ovarian) gynecologic cancers (adjusted $p = 0.01$). **Conclusion:** This analysis identified familial aggregation of cancers for which a genetic component has yet to be established.

Copyright © 2006 S. Karger AG, Basel

Introduction

The Cancer Genetics Network (CGN) is a multi-institutional consortium that was developed by the National Cancer Institute as a resource for epidemiological and translational research into the genetic basis of cancer susceptibility. Since 1999, over 18,000 individuals (probands) have consented to participate in the CGN, and complete family and medical history has been collected on each participant and resides in a Core Registry database. This Core database is maintained and updated regularly both to retain contact and communication with CGN participants to invite them to participate in translational research (such as cancer screening and psychosocial research) studies, and to provide a resource for hypothesis-generating database studies of the genetic basis of cancer. One natural study that can be readily performed utilizing such registry data that includes family cancer history is an analysis of co-aggregation of disease in individuals and families. Such an identification of cancers that co-aggregate can be useful for understanding the etiology of disease. In addition, this knowledge can lead to a more focused screening for earlier detection of disease, often resulting in improved survival.

There have been many reports in the literature on evidence of cancers that aggregate in families. Reviews of the literature on familial aggregation of breast, ovarian and colorectal cancers are given in Hoffman et al. [1], Berchuck et al. [2], and Bonaïti-Pellié [3], respectively. Narod [4] and Kerber et al. [5] reported that prostate cancer also aggregates within families. More recent literature has reported on familial aggregation of pancreatic [5, 6], hematopoietic [5, 7], and lung cancers [5, 8]. Co-aggregation of pairs of distinct cancers has also been reported in the literature: colorectal cancer is known to co-aggregate with breast and ovarian cancers [9–11], and several studies have shown that breast and ovarian cancers cluster within families and within individuals [12, 13], primarily due to mutations in BRCA1/2 [11, 14]. Studies have also suggested that breast and ovarian cancers each co-aggregate with other gynecologic cancers, but none of these results was statistically significant [11, 15, 16].

Studies of co-aggregation of multiple less prevalent cancers require a large database of family medical history of disease such as the Cancer Genetics Network has developed. This paper is a report of the analysis of the CGN Registry that was undertaken to explore for novel evidence of cancers that co-aggregate at the individual- and family-level.

Materials and Methods

Study Population

The Cancer Genetics Network is a multi-site NCI-sponsored research consortium that recruits participants at each of eight clinical sites. Recruitment was population-based at four institutions (11,628 families) and based on clinic-, physician- or self-referral at four centers (6,253 families). The four centers with population-based subjects used hospital or public registries such as Surveillance, Epidemiology and End Results (SEER) [17] to contact and enrol patients and their family members. The participation rate in the population-based cancer registry centers was commonly between 70 and 90%. In the four clinic-based centers, physicians and other health care professionals directly referred patients to CGN Centers; the participation rate in the clinic-based centers was 45–90%. Anton-Culver et al. [18] give a more detailed description of the CGN Registry, and of the specific ascertainment schemes that were utilized. Detailed family history information on up to five generations was obtained through mailed questionnaires and telephone interviews. For unaffected relatives, information was obtained on date of birth, gender, vital status, and date of death, if applicable. For cancer-affected family members, information on the type of cancer and date of diagnosis was also collected. The disease statuses of the probands were confirmed from medical records, but those of their family members were not. Follow-up of probands is done annually to update changes in cancer status for

probands and their family members. The largest categories of cancer in probands were breast, prostate or multiple primary sites.

For the purpose of this analysis, disease sites were combined into categories as in DeVita et al. [19]: breast (female cases only), ovarian, prostate, colorectal, non-ovarian gynecologic, pancreatic, hematopoietic (primarily bone marrow) and lung. Gynecologic cancers consist mainly of cervical and uterine/endometrial cancers. Males were included in the single disease analyses of non-gender-specific cancers and analyses involving prostate cancer. Similarly, women were excluded from any analysis involving prostate cancer. The CGN participants analyzed in this paper consist of over 65,000 siblings (including all probands) who were recruited prior to January 2003.

Statistical Methods

For the analysis of multiple cancers, it is necessary to choose a method that appropriately captures the association between diseases and adequately adjusts for ascertainment. Thus, it would not be appropriate to use a simple odds ratio to identify cancers that cluster in families because this approach does not adjust for the co-aggregation of both diseases when assessing the degree of aggregation of each disease individually. For example, a simple odds ratio approach is not able to address whether ovarian cancer aggregates in families beyond its co-aggregation with breast cancer.

Hudson et al. [20] proposed a family predictive model that provides a method to adjust for all possible relationships between two diseases within families and within individuals. In addition, this method appropriately adjusts for the fact that some families are not population-based. In the simple case that individuals are homogeneous, the family predictive model specifies the log-odds of disease as a linear function of the number of relatives with disease. Familial aggregation is tested by assessing the departure of the regression coefficient from zero. This model can be extended to include individual-level covariates and pair-level predictors. This model is not applicable to data with widely varying family sizes [21, 22], and thus we restricted the analyses to sibships consisting of between two and five members. Only sibships were used in order to ensure approximate environment and age matching. For the analysis of aggregation of the female- (male-) specific cancers only sisterhoods (brotherhoods) were used. The model for aggregation of lung cancer included a covariate indicating whether the proband had ever smoked.

The analysis of multiple distinct cancers (say, A and B) that co-aggregate in families used the multivariate family predictive model of Hudson et al. [20]. The simplest form of the model specifies the log-odds of disease A as a linear function of an individual's disease B status, the number of their siblings with disease A, and the number of their siblings with disease B. For example, the log-odds of lung cancer for an individual is a linear function of his/her colorectal cancer status, the number of siblings with lung cancer, and the number of siblings with colorectal cancer. The coefficients of the model used for this analysis capture: (1) co-aggregation of colorectal and lung cancers within individuals; (2) aggregation of lung cancer within families; (3) aggregation of colorectal cancer in families, and (4) co-aggregation of colorectal and lung cancers within families.

Logistic regression underlies the family predictive model. Let $y_{k,j}$ denote the disease k ($k = A, B$) status of the j th individual, $s_{k,j}$ denote the number of their siblings with disease k , and p_k denote

the probability of disease k conditional on all other cancer outcomes in the family. Then, the simplest multivariate family predictive model implies the following logistic regression equations for the conditional log-odds of each disease:

$$\begin{aligned} \text{logit}[p_A(j)] &= \alpha_A + \delta y_{B,j} + \gamma_A s_{A,j} + \gamma_{AB} s_{B,j} \\ \text{logit}[p_B(j)] &= \alpha_B + \delta y_{A,j} + \gamma_B s_{B,j} + \gamma_{AB} s_{A,j} \end{aligned} \quad (1)$$

The parameters in this model have conditional interpretations: α_A (α_B) is the log-odds of disease A (B) given no other disease A (B) in the family, δ is the log-odds ratio for co-aggregation of diseases A and B within individuals, γ_{AB} is the log-odds ratio for co-aggregation of diseases A and B between family members, and γ_A (γ_B) is the log-odds ratio for aggregation of disease A (B). We note that the estimates of both levels of co-aggregation derived from the model are not useful because this basic application of the family predictive model treats the diseases as exchangeable with respect to co-aggregation. For example, at the individual (as well as family) level, the increase in the risk of lung cancer associated with having colorectal cancer is assumed to be of the same magnitude as the increase in the risk of colorectal cancer associated with having lung cancer. Although this simplifying assumption may not be valid for all diseases, especially in the case of uncommon diseases, the data are typically too sparse for a more complex model. Although the parameter estimates from these analyses may not be appropriate for prediction, they do form the basis for valid tests of association and thus we will focus only on the statistical inference about the co-aggregation of cancers that is provided by these methods.

The CGN Registry includes families recruited due to a personal or family history of cancer. To account for this ascertainment, we treated the proband's disease status as fixed by design. Thus probands enter our logistic regression models only as covariates and not as outcomes; they contribute to the number of relatives with disease. As this does not completely remove bias due to self-referral, we repeated the multivariate analyses that showed evidence of co-aggregation using only the population-based CGN families.

The methodology of generalized estimating equations (GEEs) [23] is used to adjust for the correlation among family members. A two-sided significance level of 0.05 was used in all tests of co-aggregation, with adjustment for multiple comparisons through control of the false discovery rate, as in Storey et al. [24]. Specifically, a q value cut-off of 0.05 is used, so that all associations with q values less than 0.05 are called significant and that on average, 5% of the truly null associations will be called significant. We refer to the q value as an adjusted p value throughout this paper. For the case of aggregation of individual cancers, there were not enough comparisons to use Storey et al. [24], and thus adjustment for multiple comparisons was done using Bonferroni [25] (also referred to as adjusted p values).

Results

There were 12,263 families used in this analysis. There were 1,159 colorectal cancers, 450 lung cancers, 185 hematopoietic cancers, and 149 pancreatic cancers. For female cancers, our analysis was based on 9,749 sis-

Table 1. Aggregation of individual cancers

Cancer	Cases per sibship	Number of sibships (%)	Adjusted p value
Hematopoietic	0	12,082 (98.5)	0.024
	1	177 (1.4)	
	2	4 (0.0)	
	3 or more	0 (0.0)	
Lung	0	11,849 (96.6)	<0.001
	1	381 (3.1)	
	2	30 (0.2)	
	3 or more	0 (0.0)	
Pancreatic	0	12,116 (98.8)	0.832
	1	145 (1.2)	
	2	2 (0.0)	
	3 or more	0 (0.0)	
Prostate*	0	5,505 (68.2)	<0.001
	1	2,006 (24.9)	
	2	456 (5.7)	
	3 or more	105 (1.3)	
Colorectal	0	11,179 (91.2)	<0.001
	1	1,014 (8.3)	
	2	66 (0.5)	
	3 or more	4 (0.0)	
Breast*	0	4,871 (50.0)	<0.001
	1	3,911 (40.0)	
	2	853 (8.8)	
	3 or more	114 (1.2)	
Ovarian*	0	9,093 (93.3)	0.184
	1	636 (6.5)	
	2	19 (0.2)	
	3 or more	1 (0.0)	
Gynecologic* (non-ovarian)	0	9,212 (94.5)	<0.001
	1	506 (5.2)	
	2	28 (0.3)	
	3 or more	3 (0.0)	

* Single-sex sibships only.

terhoods containing 5,972 cases of breast cancer, 677 of ovarian cancer and 571 cases of non-ovarian gynecologic cancer. For prostate cancer, the analysis was based on 8,072 brotherhoods with 3,264 cases of prostate cancer.

Familial Aggregation of Individual Cancers

Evaluation of familial aggregation of a single disease is driven by the number of families with two or more cases of disease. Table 1 gives the distribution of the num-

Table 2. Co-aggregation of two different cancers within an individual

Number of individuals with both cancers (adjusted p value)	Lung*	Pancreatic*	GYN**	Prostate***	Breast**	Ovarian**	Colorectal*
Hematopoietic*	1 (p = 0.28)	0 [†]	5 (p = 0.025)	14 (p = 0.08)	9 (p = 0.041)	3 (p = 0.28)	11 (p = 0.28)
Lung*	–	2 (p = 0.22)	7 (p = 0.22)	23 (p = 0.06)	35 (p = 0.041)	8 (p = 0.22)	20 (p = 0.28)
Pancreatic*	–	–	2 [†]	1 [†]	2 (p = 0.10)	1 [†]	1 (p = 0.30)
Gynecologic**	–	–	–	–	150 (p = 0.22)	31 (p < 0.001)	7 (p = 0.08)
Prostate***	–	–	–	–	–	–	100 (p = 0.28)
Breast**	–	–	–	–	–	151 (p < 0.001)	104 (p = 0.26)
Ovarian**	–	–	–	–	–	–	12 (p = 0.26)

* 39,572 individuals; ** 27,334 females; *** 22,300 males.

[†] GEEs cannot be performed – cancers are too rare.

Table 3. Co-aggregation of multiple cancers within a sibship

% families with at least one case of each disease (adjusted p value)	Lung	Pancreatic	GYN*	Prostate	Breast	Ovarian	Colorectal
Hematopoietic	0.1% (p = 0.28)	0.0% [†]	0.1% (p = 0.01)	0.9% (p = 0.26)	0.4% (p = 0.21)	0.1% (p = 0.28)	0.2% (p = 0.28)
Lung	–	0.1% (p < 0.001)	0.2% (p = 0.21)	2.4% (p < 0.0033)	1.3% (p = 0.21)	0.3% (p = 0.21)	0.5% (p = 0.004)
Pancreatic	–	–	0.1% [†]	0.5% [†]	0.2% (p = 0.08)	0.0% [†]	0.1% (p = 0.018)
Gynecologic*	–	–	–	–	3.7% (p = 0.12)	0.6% (p = 0.28)	0.9% (p = 0.003)
Prostate	–	–	–	–	–	–	4.2% (p = 0.004)
Breast	–	–	–	–	–	3.8% (p = 0.025)	2.5% (p < 0.003)
Ovarian	–	–	–	--	–	–	0.5% (p = 0.07)

* Non-ovarian.

[†] GEEs cannot be performed – cancers are too rare.

ber of cancers within sibships as well as the results of the family predictive models for each cancer individually. Our results confirmed the results published in earlier papers reporting familial aggregation of breast cancer [1, 5], colorectal cancer [3, 5], prostate cancer [4, 5], hematopoietic cancer [5, 7], and lung cancer [5, 8].

Familial Co-Aggregation of Distinct Cancers

In considering familial co-aggregation of two distinct cancers within families and within individuals, the number of families and individuals with at least one case of each disease drives co-aggregation. Table 2 gives the number of individuals with two (or more) cancers in a pair-wise fashion as well as the adjusted p values from the multivariate family predictive models assessing co-aggregation of cancers. These results for co-aggregation at the family-level are given in table 3. Note that these models

are different from those in table 1 in which only one cancer was considered at a time.

Our results confirmed those published in earlier papers reporting familial co-aggregation of breast and ovarian cancers [12, 13], and co-aggregation of colorectal and prostate cancers [5]. We also confirmed co-aggregation at the individual-level of breast and ovarian cancer [12, 13], as well as ovarian and non-ovarian gynecologic cancers [11, 16]. In addition, we identified associations that, to our knowledge, have not yet been reported. We found evidence of familial co-aggregation of lung cancer with pancreatic cancer (adjusted p < 0.001), prostate cancer (adjusted p < 0.003), and colorectal cancer (adjusted p = 0.004). We also found that hematopoietic and non-ovarian gynecologic cancers cluster together at the individual level (adjusted p = 0.025) and at the family level (adjusted p = 0.010). In addition, we found that pancreatic can-

cer co-aggregates in families with colorectal cancer (adjusted p value = 0.018). At the individual level, both hematopoietic and lung cancers co-aggregate negatively with breast cancer (adjusted p = 0.047 for both).

Discussion

The analysis of the CGN Registry revealed evidence of familial aggregation of cancers for which a genetic component has yet to be established: lung cancer [26] and hematopoietic cancer [27, 28]. It would be useful to further study the genetic and/or environmental factors responsible for the familial clustering of these cancers.

There are several limitations to studying disease aggregation using data collected in a family registry such as the CGN. First, there was a potential for misreporting of disease history. As the disease status of probands was confirmed by the CGN sites, but not those of their family members, there may be errors in the reporting of family history. For example deep organ cancers are often misclassified, and metastatic sites are sometimes reported as primary cancers when cancer history is not confirmed from medical records [29, 30]. This misreporting of metastatic sites as primary cancers may partially explain the familial co-aggregation we detected of lung and colorectal cancers since the lung is a frequent site of metastasis from the colon [31]. If there is greater misclassification of disease for unaffected probands compared to those who report a cancer history, then there may be an additional source of bias. As we analyzed only siblings, misreporting of family history should be minimized.

Second, information on the genotype of the proband and relatives, and the behavioral history (such as smoking history) of relatives was not recorded. We were therefore unable to remove families with known genetic syndromes. Also, we were unable to properly adjust for behavioral factors such as smoking. This presents a limitation to the interpretation of co-aggregation of two cancers where smoking is a risk factor for both diseases. This may explain the finding of familial co-aggregation of lung and pancreatic cancers, since it is possible that smoking behavior clusters in siblings. In absence of this, the proband's smoking status must be viewed as surrogate information.

Third, our analyses assumed that all families were sampled because of the disease status of the proband. In actuality, the ascertainment was more complex. For example, some probands referred themselves to the Network. It is not known why these probands chose to par-

ticipate; it may be due to a family (not a personal) history of cancer. In this case a familial association may be induced by the ascertainment scheme, and should be appropriately included in the analysis. To overcome the potential ascertainment biases, we performed a confirmatory re-analysis using the population-based subset of Registry families. Cancers were quite rare in this subset, but we were able to confirm virtually all our reported significant co-aggregation of diseases in individuals, as well as familial co-aggregation of lung and prostate cancers, and colorectal and prostate cancers. The evidence for co-aggregation of colorectal with other cancers was not statistically significant in this subgroup (partially due to reduced power), and there were insufficient numbers of cases in the population-based subset to perform some analyses, including lung cancer with pancreatic cancer, and hematopoietic cancer with non-ovarian gynecologic cancers. Our current research focuses on methods that can handle more complex ascertainment schemes, such as that exemplified by the heterogeneous CGN recruitment.

Lastly, evaluation of co-aggregation of cancers within individuals is complicated by competing risks [32], that is, an individual may die of lung cancer before developing another cancer. This is especially problematic when dealing with cancers with high mortality rates, such as ovarian, lung, pancreatic and hematopoietic cancers [17]. This would tend to diminish the evidence of aggregation. One approach to decrease the effects of competing risks is to adjust for age. This would also adjust for individuals who never had cancer before study participation including those who died before developing disease. Our current research also focuses on developing methods of extending the family predictive model to account for the ages at disease onset and censoring.

Despite these limitations, our analysis revealed several interesting disease associations, which could be useful in guiding future research into the responsible genetic and environmental factors. While there is clearly a chance of spurious associations, there is also a strong likelihood that unexplained multicancer phenotypes of variable penetrance do exist and that defining specific patterns will prove very important in linking them to the one or more genes that define each of these subsets. The CGN resource family history of cancer, and consent for future contact for research studies of these diseases should be viewed as a rich source available to the scientific community for cancer genetics research. Further information on this resource is available on the web [33, 34].

Acknowledgements

We wish to thank the editors and reviewers as well as Dr. Daniel Haber for useful comments. This research was supported by the National Cancer Institute through cooperative agreements with Carolina-Georgia Center of the Cancer Genetics Network, Lombardi Cancer Center Cancer Genetics Network, Mid-Atlantic Cancer Genetics Network, Northwest Cancer Genetics Network, Rocky

Mountain Cancer Genetics Coalition, Texas Cancer Genetics Consortium, University of Pennsylvania Cancer Genetics Network, and University of California, Irvine-University of California, San Diego Cancer Genetics Network Center, and PIs, and Informatics Technology Groups of Massachusetts General Hospital, Boston, University of California, Irvine, Yale University Informatics Technology Group, and PIs, as well as NIH grants CA 74302 and CA 75971.

References

- Hofmann W, Schlag PM: BRCA1 and BRCA2 – breast cancer susceptibility genes. *J Cancer Res Clin Oncol* 2000;126:487–496.
- Berchuck A, Carney M, Lancaster JM, Marks J, Fitreal AP: Familial breast-ovarian cancer syndromes: BRCA1 and BRCA2. *Clin Obstet Gynecol* 1998;41:157–166.
- Bonaïti-Pellié C: Genetic risk factors in colorectal cancer. *Eur J Cancer Prev* 1999;8(suppl 1): 27–32.
- Narod S: Genetic epidemiology of prostate cancer. *Biochem Biophys Acta* 1998;1423:F1–F13.
- Kerber RA, O'Brien E: A cohort study of cancer risk in relation to family histories of cancer in the Utah Population Database. *Cancer* 2005;103:1906–1915. Epub 2005 March 18.
- Hruban RH, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJA, Falatko F, Yeo CJ, Kern SE: Familial pancreatic cancer. *Ann Oncol* 1999;10(suppl 4):69–73.
- Horwitz M: The genetics of familial leukemia. *Leukemia* 1997;11:1347–1359.
- Bromen K, Pohlabein H, Jahn I, Ahrens W, Jöckel KH: Aggregation of lung cancer in families: Results from a population-based case-control study in Germany. *Am J Epidemiol* 2000;152:497–505.
- Nelson CL, Sellers TA, Rish SS, Potter JD, McGovern PG, Kushi LH: Familial clustering of colorectal, breast, uterine, and ovarian cancers as assessed by family history. *Genet Epidemiol* 1993;10:235–244.
- Vasen HFA, Wijnen JT, Menko FH, Kleibouker JH, Taal BG, Griffioen G, Nagengast FM, Meijers-Heijboer EH, Bertario L, Varesco L, Bisgaard M-L, Mohr J, Fodde R, Khan PM: Cancer risk in families with hereditary non-polyposis colorectal cancer diagnosed by mutation analysis. *Gastroenterology* 1996;110: 1020–1027.
- The Breast Cancer Linkage Consortium: Cancer risks in BRCA2 mutation carriers. *J Natl Cancer Inst* 1999;91:1310–1316.
- Schildkraut JM, Risch N, Thompson WD: Evaluating genetic association among ovarian, breast and endometrial cancer: evidence for a breast/ovarian cancer syndrome. *Am J Hum Genet* 1989;45:521–529.
- Sutcliffe S, Pharoah PDP, Eason DF, Ponder BAJ, UKCCCR Familial Ovarian Cancer Study Group: Ovarian and breast cancer risks to women in families with two or more cases of ovarian cancer. *Int J Cancer* 2000;87:110–117.
- Ford D, Easton DF, Bishop T, Narod SA, Goldgar DE, Breast Cancer Linkage Consortium: Risk of cancer in BRCA1-mutation carriers. *Lancet* 1994;343:692–695.
- 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;92:144–150.
- Fearon ER: Human cancer syndromes: clues to the origin and nature of cancer. *Science* 1997; 278:1043–1050.
- Gloekler Ries LA, Koasry CL, Hankey BF, Miller BA, HARRAS A, Edwards BK (eds): SEER Cancer Statistics Review, 1973–1994. Bethesda, National Cancer Institute, 1997.
- Anton-Culver H, Ziogas A, Finkelstein D, Griffin C, Hanson J, Isaacs C, Kasten-Sportes C, Mineau G, Nadkarni P, Potter JD, Rimer B, Schildkraut J, Strong L, Weber B, Winn D, Hiatt R, Nayfield S: The cancer genetics network: recruitment results and pilot studies. *Community Genet* 2003;6:171–177.
- De Vita VT, Hellman S, Rosenberg SA (eds): *Cancer: Principles and Practice*, ed 5. Philadelphia, Lippincott-Raven, 1997.
- Hudson J, Laird N, Betensky R: Multivariate logistic regression for familial aggregation of two disorders. I. Development of models and methods. *Am J Epidemiol* 2001;153:500–505.
- Cox DR, Wermuth N: A note on the quadratic exponential binary distribution. *Biometrika* 1994;81:403–408.
- Betensky RA, Whittemore AS: An analysis of correlated multivariate binary data: Application to familial cancers of the ovary and breast. *Appl Statist* 1996;45:411–429.
- Liang KY, Zeger SL: Longitudinal data analysis using generalized linear models. *Biometrika* 1986;83:13–22.
- Storey JD, Tibshirani R: Statistical significance for genomewide studies. *Proc Natl Acad Sci* 2003;100:9440–9445.
- Pagano M, Gauvreau K: *Principles of Biostatistics*, ed 2. New York, Duxbury Thomson Learning, 2000.
- Kiyohara C, Otsu A, Shirakawa T, Fukada S, Hopkin JM: Genetic polymorphisms and lung cancer susceptibility: a review. *Lung Cancer* 2002;37:241–256.
- Bevan S, Catovsky D, Marossy A, Matutes E, Popat S, Antonovic P, Bell A, Berrebi A, Gaminara EJ, Quabeck K, Ribeiro I, Mauro FR, Stark P, Sykes H, Van Dongen J, Wimperis J, Wright S, Yuille MR, Houlston RS: Linkage analysis for ATM in familial B cell chronic lymphocytic leukaemia. *Leukemia* 1999;82: 775–781.
- Bevan S, Catovsky D, Matutes E, Antonovic P, Auger MJ, Ben-Bassat I, Bell A, Berrebi A, Gaminara EJ, Junior ME, Mauro FR, Quabeck K, Rassam SMB, Reid C, Ribeiro I, Stark P, Van Dongen JJM, Wimperis J, Wright S, Marossy A, Yuille MR, Houlston RS: Linkage analysis for major histocompatibility complex-related genetic susceptibility in familial chronic lymphocytic leukemia. *Blood* 2000;96: 3982–3984.
- Colditz GA, Martin P, Stamfer MJ, Willett WC, Sampson L, Rosner B, Hennekens CH, Speizer FE: Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986;123:894–900.
- Mitchell RJ, Brewster D, Campbell H, Porteous MEM, Wyllie AH, Bird CC, Dunlop MG: Accuracy of reporting family history of colorectal cancer. *Gut* 2004;53:291–295.
- Vogelsang H, Haas S, Hierholz C, Berger U, Siewert JR, Präuer H: Factors influencing survival after resection of pulmonary metastases from colorectal cancer. *Br J Surgery* 2004;91: 1066–1071.
- Rothman KJ, Greenland S: *Modern Epidemiology*, ed 2. Philadelphia, Lippincott Williams & Wilkins, 1998.
- Cancer Genetics Network Statistical Coordinating Center: <http://hedwig.mgh.harvard.edu/cgnpub/>
- National Cancer Institute Cancer Genetics Network Site: <http://epi.grants.cancer.gov/CGN/>