
Validation of Family History Data in Cancer Family Registries

Argyrios Ziogas, PhD, Hoda Anton-Culver, PhD

Background: Although family history information on cancer is used to infer risk of the disease in population-based, case-control, cohort, or family-based studies, little information is available on the accuracy of a proband's report. In this study, we sought to determine the validity of the reporting of family history of cancer by probands in population-based and clinic-based family registries of breast, ovarian, and colorectal cancers.

Methods: To assess the accuracy of probands' reported family history of cancer in their relatives, we compared the family history from the personal interview of each proband to a reference standard that included pathology reports, self-reports, or death certificates on the relatives. Our study included 1111 families that accounted for 3222 relatives who were verified. To account for within-family correlations in the responses, we used a generalized estimating equation approach.

Results: The probability of agreement between the proband-reported cancer status in a relative with the reference standard varied by cancer site and by degree of relationship to the proband. This probability for first-degree relatives was 95.4% (95% confidence interval [CI]=92.6–98.3) for female breast cancer; 83.3% (95% CI=72.8–93.8) for ovarian cancer; 89.7% (95% CI=85.4–94.0) for colorectal cancer; and 79.3% (95% CI=70.0–88.6) for prostate cancer.

Conclusions: We found high reliability of probands' reporting on most cancer sites when they reported on first-degree relatives and moderate reliability for their reporting on second- and third-degree relatives. Overreporting of cancer was rare (2.4%). Race or ethnicity and gender of the proband did not influence the accuracy of reporting. However, degree of relationship to the proband, type of cancer, age at diagnosis of the proband, and source of ascertainment of probands were statistically significant predictors of accuracy of reporting. (Am J Prev Med 2003;24(2):190–198) © 2003 American Journal of Preventive Medicine

Introduction

Family history of cancer in first-degree relatives has been shown to be a risk factor associated with increased risk of developing cancer^{1–3} and has been used to identify individuals for genetic and molecular studies. From these studies, estimates on the proportion of hereditary cancer because of susceptibility genes have been made. In addition, families with a history of cancer in multiple generations have been useful in cloning susceptibility genes.^{4,5} Prediction models are usually based on explicit family histories of breast and ovarian cancers or summaries of the number of affected first- and second-degree relatives with breast or ovarian cancer.^{6–9} These models are used by physicians and genetic counselors interested

in helping women to understand their risk of breast and ovarian cancers and the preventive options available to them.

In the present study, we sought to validate the reporting of family history of cancer by cancer-affected probands in population-based and clinic-based family registries of breast, ovarian, and colorectal cancers. The objectives of the study were (1) to systematically evaluate the consistency of cancer-affected proband-reported information on cancer in their first-, second-, and third-degree relatives; (2) to determine the positive and negative predictive values as well as the probabilities of agreement between the proband-reported cancer status in a relative with the reference standard for various cancer sites; and (3) to determine the effect of the characteristics of the proband and the characteristics of the proband's relatives on the probability of agreement between information given by the proband's interview and the reference standard (pathology, self-reporting, and death certificate).

From the Epidemiology Division, Department of Medicine, University of California, Irvine, California

Address correspondence and reprint requests to: Hoda Anton-Culver, PhD, Epidemiology Division, Department of Medicine, University of California Irvine, 224 Irvine Hall, Irvine CA 92697-7550. E-mail: hantoncu@uci.edu.

Methods

This validation study was conducted in the course of the creation of the family registries of probands for breast, ovarian, and colorectal cancers at the University of California at Irvine (UCI), in Orange County, California. According to the U.S. Bureau of the Census, roughly 1% of Americans reside in Orange County, California (2.7 million, estimated as of July 1, 1999). Approximately 54% of Orange County residents are white, 2% African American, 30% Hispanic, and 14% Asian.

The cancer family registries used for this study included both population-based and clinic-based probands.^{10,11} Population-based proband ascertainment included all breast cancer cases diagnosed in Orange County, California, during the 1-year period beginning March 1, 1994 (UCI IRB no. HS91-137); all ovarian cancer cases diagnosed during the 2-year period beginning March 1, 1994 (UCI IRB no. HS91-137); and a weighted sample of familial colorectal cancer cases diagnosed during the 2-year period beginning March 1, 1994 (UCI IRB no. 93*257). In particular, for colorectal cancer probands, the population sampling was done with a stratification to increase the proportion of patients who were familial or younger than 65 years and, at the same time, to maintain the population-based status. Clinic-based proband ascertainment included a number of families enrolled in the family registries as referrals from Orange County physicians. At least annually, the UCI cancer surveillance program contacts community, clinic, and hospital physicians to let them know about new and ongoing genetic epidemiology studies and to encourage them to refer their high-risk probands to enroll in these studies. The clinic-ascertained group is included in this study because of the importance of generating data that would be helpful in clinical settings. We do recognize, however, that the data from this group are generalizable only to subjects who have a strong family history of cancer and are considered to be at high cancer risk. The characteristics of this group were different from the population-based group, because this group was referred to us because of their cancer familiarity. The same protocol was used for interviews and verification of cancer in relatives for clinic-based and population-based probands.

Data on approximately 1200 breast cancer, 300 ovarian cancer, and 1200 colorectal cancer families were available in the family registries. A description of the family registries has been given elsewhere.^{10,11} All probands included in this study were cancer affected. In the current study, we included all families in which there was at least one relative, affected or unaffected, with a method of verification available. Our study included 670 breast cancer families, 123 ovarian cancer families, and 318 colorectal cancer families.

Relatives were classified as first-degree, which included parents, siblings, and children of the proband; second-degree, which included grandparents, aunts, uncles, half-siblings, nieces, and nephews; and third-degree, which included first cousins and grandchildren. A proband was considered to be "familial" with respect to a specified cancer if at least one first-degree relative (parent or sibling) in addition to the proband was diagnosed with the same cancer as the proband.

Data Collection

After initial contact with the proband's physician(s), a description of the study and an invitation to participate were mailed to the proband. This mailing was followed by a telephone interview to obtain family history information (including first-, second-, and third-degree relatives). Interviewers entered the family history data into the Genetics Registry Information System (GRIS) database through a set of computer screens that capture demographic information, tumor information, and protocol status for probands and family members. The system has functions to generate merge files for creating personalized letters and mailing labels. After the telephone interview, a verification report and pedigree were produced from GRIS that showed the family history information reported by the proband during the interview. These reports were mailed to the proband so that he or she could complete data items not known during the interview and verify all information collected. The verification table included dates of birth and death, types of cancer, dates of diagnoses, hospital at diagnosis, and other relevant information. Release-of-information forms also were sent to the proband for approval and signature in order to obtain medical records of those relatives who were deceased. After obtaining permission from the proband, the interviewers contacted the proband's living relatives who had cancer and asked them to sign their own release-of-information forms.

Family history verification is a dynamic process and can continue over a long period. Therefore, in the analysis presented here, we selected July 2000 as a cutoff date. Statistical analysis included all relatives for whom verification was obtained using the methods described in the Verification Methods section. The relatives included in these analyses were those reported by the probands to have had cancer or to have died, as well as unaffected siblings or cousins of the proband. All families for whom we completed at least a level of verification on at least one relative were included.

Verification Methods

Malignancies reported on the relatives of the proband were verified by one of the following methods: (1) obtaining pathology reports, tumor tissue samples, or clinical records, which we will refer to throughout this report as "pathology"; (2) obtaining "self-reports" from affected and unaffected relatives of probands through a structured questionnaire and personal interviews; or (3) obtaining death certificates on deceased relatives. We obtained an authorization form from the proband or next of kin to obtain release of medical information (pathology reports and tumor tissue samples), to confirm malignancies, or to obtain death certificates. For a sample of living relatives who were reported by the probands to be unaffected, we obtained permission from the probands to contact these relatives to verify that they did not have cancer. For deceased unaffected relatives, we obtained death certificates and verified the probands' reporting. To assess the accuracy of reported family history of cancer in the relatives of probands, we compared the reported family history from the personal interview of the proband with a reference standard that included pathology reports or self-reports from the relative or death certificates on deceased

relatives. For relatives with more than one source of verification information of their cancer status, we chose pathology over both the self-report and death certificate. In some cases, self-reports were obtained before death; in these cases, self-reports were chosen over death certificates. We used the date the proband returned the verification table as a reference date. The pathology report, self-report, and death certificate were adjusted according to this date.

Statistical Analysis

We calculated the positive and negative predictive values (PPV and NPV) on the cancer status of relatives by using the family history data reported through a personal interview of the proband and parallel data on the same relatives by using a reference standard as described in the Introduction. We also determined (1) the probability of agreement between the proband-reported cancer in a relative given the cancer status of that relative, as determined by the reference standard (probability of agreement of cancer [PAC] in a relative), and (2) the probability of agreement between the proband-reported absence of cancer in a relative given that the data from the reference standard indicated that this relative had no cancer (probability of agreement of no cancer [PANC] in a relative). For the above analyses, we used data collected from the proband during the personal interview, and we considered pathology, self-report, and death certificate as the reference standard. To account for the within-family correlations on the responses, we used a generalized estimating equation (GEE) approach with a log link and an exchangeable correlation structure.¹² This approach accounts for the variable number of relatives per proband, allows for the correlation among responses within families, and provides appropriate confidence intervals (CIs) of the estimates. To account for the sampling of our study, we used logistic regression models,¹³ in which the predictor variable was the reporting of family history of cancer from the proband and the response was the reference standard. We calculated the results for PPV and NPV as well as the PAC and PANC and their 95% CIs for specific cancers, stratified by the degree of relationship to the proband.

We investigated the effect of characteristics of the proband and the relatives, such as age at diagnosis, race or ethnicity, gender, and type of cancer on the probability of disagreement. Because the false-negative rate was different from the false-positive rate, we used separate regression models for the false positives and the false negatives. To account for the within-family correlations on the misclassification rate, we used a GEE approach with a log link and an exchangeable correlation structure. For these models, we considered the response variable to be simple agreement between the proband's interview and the reference standard. For each combination of family history of cancer (breast, prostate, colorectal, lung, and all cancers combined), we fitted separate models on the response described for the false negatives. Similarly, we fitted a model of all cancers combined among the false positives. In all models, we included a number of predictors and used the deviate scores to determine the best fitting model.

Results

Population Under Study

In the present study, we included families of 1111 probands; 670 (60.3%) were breast cancer probands, 123 (11.1%) were ovarian cancer probands, and 318 (28.6%) were colorectal cancer probands. Table 1 shows the characteristics of the probands with respect to race or ethnicity, ascertainment source, gender, age at diagnosis, and proband-reported family histories of breast, ovarian, and colorectal cancers. The majority of probands were non-Hispanic white (1022 [92.0%]); the primary source of ascertainment was population-based (1042 [93.8%]); and the gender was predominantly female (939 [84.5%]). The distribution of age at diagnosis of the probands with breast, ovarian, or colorectal cancer shows that 312 probands (28.0%) were diagnosed at an age younger than 50 years, 545 probands (49.1%) were diagnosed between the ages of 50 and 69 years, and 254 probands (22.9%) were diagnosed at age 70 years or older. The mean age at diagnosis for the probands was 56.6 years old (SD=13.3).

Family Size

The distribution of relatives by degree of relationship to the proband across the three groups of probands is shown in Table 2. On the average, there were 7.1 first-degree relatives, 19.4 second-degree relatives, and 14.3 third-degree relatives per family. On the average, we obtained verification on 1.8 first-degree relatives, 2.4 second-degree relatives, and 1.6 third-degree relatives per family, with a verification rate of 25.4%, 12.4%, and 10.7%, respectively. An average of 2.0 first-degree relatives, 2.6 second-degree relatives, and 2.2 third-degree relatives per family were diagnosed with cancer. We obtained verification on 1.4 first-degree relatives, 1.7 second-degree relatives, and 1.4 third-degree relatives per family diagnosed with cancer, with a verification rate of 70.4%, 65.1%, and 61.3%, respectively.

Verification Methods

In total we included 3222 relatives (1320 [41%] men and 1902 [59%] women) for whom we had a form of verification; of these, 1692 (52.5%) were first-degree relatives, 1214 (37.7%) were second-degree relatives, and 320 (9.9%) were third-degree relatives. We were able to obtain 474 pathology reports, 777 self-reports, and 2142 death certificates. We had multiple forms of verification from some relatives; specifically, we obtained both pathology reports and death certificates for 85 relatives, both pathology and self-reports for 75 relatives, and both self-reports and death certificates for 7 relatives.

Among the 85 relatives for whom we obtained pathology reports and death certificates, 52 (61.2%) were

Table 1. Characteristics of probands participating in the validation study

Characteristic	Breast		Ovarian		Colorectal		Total	
	n=670	(60.3%)	n=123	(11.1%)	n=318	(28.6%)	N=1111	
Race/ethnicity								
White	626	(93.4%)	113	(91.9%)	283	(90.0%)	1022	(92.0%)
African American	2	(0.3%)	0	(0.0%)	6	(1.9%)	8	(0.7%)
Hispanic	30	(4.5%)	7	(5.7%)	15	(4.7%)	52	(4.7%)
Asian	9	(1.3%)	3	(2.4%)	10	(3.1%)	22	(2.0%)
Other	3	(0.5%)	0	(0.0%)	4	(1.3%)	7	(0.6%)
Ascertainment								
Population-based	638	(95.2%)	118	(95.9%)	286	(89.9%)	1042	(93.8%)
Clinic-based	32	(4.8%)	5	(4.1%)	32	(10.1%)	69	(6.2%)
Gender								
Male	8	(1.1%)	—	—	164	(51.5%)	172	(15.5%)
Female	662	(98.9%)	123	(100.0%)	154	(48.4%)	939	(84.5%)
Age at diagnosis (years)								
<40	62	(9.3%)	17	(13.8%)	17	(5.3%)	96	(8.6%)
40–49	156	(23.3%)	29	(23.6%)	31	(9.7%)	216	(19.4%)
50–59	177	(26.4%)	29	(23.6%)	75	(26.6%)	281	(25.3%)
60–69	132	(19.7%)	25	(20.3%)	107	(33.7%)	264	(23.8%)
70+	143	(21.3%)	23	(18.7%)	88	(27.7%)	254	(22.9%)
Family history^a								
Breast	228	(34.0%)	32	(26.0%)	64	(20.1%)	324	(29.2%)
Ovarian	32	(4.9%)	12	(11.1%)	18	(5.7%)	62	(5.6%)
Colorectal	88	(13.1%)	12	(11.1%)	200	(62.9%)	300	(27.0%)

^aFamily history is based on first-degree relatives (father, mother, brother, and sister).

first-degree, 25 (29.4%) were second-degree, and eight (9.4%) were third-degree relatives of probands. Among the 85 relatives, in 68 (80.0%) pathology reports and death certificates were in complete agreement for each cancer site; in the remaining 17, there was disagreement between the pathology report and the death certificate. Among those 17 relatives, 12 (14.1%) had no cancer reported on the death certificate; for five relatives (5.9%), the metastatic cancer site rather than the primary site was reported on the death certificate. Among the 75

relatives for whom we obtained both a pathology report and self-report, 62 (82.7%) were first-degree, seven (9.3%) were second-degree, and six (8.0%) were third-degree relatives. In this group, 69 (92.0%) had complete agreement between the self-report and pathology for each cancer site.

In Table 3, first-degree relatives are characterized as positive or negative for their own history of a number of different cancer sites, as reported by the proband in the interview, compared with the reference standard. Table 3 also shows estimates of PPV and NPV as well as PAC

Table 2. Distribution of relatives in families with at least one relative participating in the validation study

Relative	Breast				Ovarian				Colorectal			
	N	Mean	(SD)	[range] ^a	N	Mean	(SD)	[range]	N	Mean	(SD)	[range]
Distributions of relatives (affected and unaffected) from all families												
1st degree	670	6.8	(2.9)	[2–23]	123	6.8	(2.9)	[3–17]	318	7.7	(2.4)	[2–24]
2nd degree	670	18.9	(9.2)	[11–70]	122	18.1	(9.8)	[6–72]	318	21.0	(10.5)	[4–71]
3rd degree	620	14.7	(11.9)	[1–74]	107	15.3	(13.6)	[1–78]	280	13.5	(12.4)	[1–91]
Distributions of verified relatives (affected and unaffected) from all families												
1st degree	566	1.8	(1.3)	[1–9]	107	1.8	(1.2)	[1–7]	285	1.7	(1.2)	[1–9]
2nd degree	325	2.3	(1.7)	[1–12]	60	3.2	(2.4)	[1–11]	114	2.3	(1.8)	[1–10]
3rd degree	153	1.5	(1.0)	[1–7]	25	1.8	(1.0)	[1–5]	27	1.6	(1.1)	[1–5]
Distributions of cancer status in relatives from all families												
1st degree	525	1.9	(1.2)	[1–8]	79	1.9	(1.0)	[1–5]	304	2.2	(1.3)	[1–8]
2nd degree	555	2.7	(1.8)	[1–11]	95	2.7	(1.7)	[1–8]	224	2.3	(1.5)	[1–11]
3rd degree	314	2.3	(1.8)	[1–11]	49	2.6	(2.3)	[1–9]	119	1.9	(1.1)	[1–6]
Distributions of cancer status in relatives with verified cancer												
1st degree	369	1.4	(0.7)	[1–5]	53	1.5	(0.8)	[1–4]	249	1.4	(0.9)	[1–6]
2nd degree	286	1.7	(1.1)	[1–6]	44	1.8	(1.2)	[1–5]	86	1.6	(0.9)	[1–5]
3rd degree	114	1.4	(0.8)	[1–5]	20	1.4	(0.7)	[1–3]	25	1.2	(0.5)	[1–3]

^aN is the number of probands (families); Mean (SD) [Range] refer to the number of relatives per family.

Table 3. Estimates of PAC, PANC, PPV, and NPV in first-degree relatives using data from proband interviews and reference standard

Type of cancer	Interview/reference standard				PAC (95% CI)	PANC (95% CI)	PPV (95% CI)	NPV (95% CI)
	^a -/-	+/-	-/+	+/+				
Females (n^b=1070)								
Breast	850	23	9	188	95.4 (92.6–98.3)	97.4 (96.4–98.4)	89.1 (84.1–93.0)	98.9 (98.0–99.5)
Cervix	1049	3	9	9	50.0 (31.8–68.2)	99.7 (99.4–100)	75.0 (42.8–94.5)	99.1 (98.4–99.6)
Endometrium	1035	17	8	10	55.6 (34.2–76.9)	98.4 (97.9–98.9)	37.0 (19.4–57.6)	99.2 (98.5–99.7)
Ovary	1017	11	7	35	83.3 (72.8–93.8)	98.9 (98.4–99.5)	76.1 (61.2–87.4)	99.3 (98.6–99.7)
Males (n=622)								
Prostate	557	7	12	46	79.3 (70.0–88.6)	98.8 (98.4–99.5)	86.7 (74.7–94.5)	97.9 (96.3–98.9)
Females and Males (N=1692)								
Esophagus	1680	3	2	7	77.8 (52.9–100)	99.8 (99.7–100)	70.0 (34.7–93.3)	99.9 (99.6–99.9)
Stomach	1644	12	12	24	66.7 (53.1–80.2)	99.3 (98.9–99.6)	66.7 (49.0–81.4)	99.3 (98.7–99.6)
Colorectum	1454	44	20	174	89.7 (85.4–94.0)	97.1 (96.3–97.8)	79.8 (73.9–84.9)	98.6 (97.9–99.2)
Pancreas	1657	7	4	24	85.7 (73.5–97.9)	99.6 (99.3–99.9)	77.4 (58.9–90.4)	99.8 (99.4–99.9)
Liver	1669	12	3	8	72.7 (59.4–86.1)	99.3 (99.1–99.5)	40.0 (19.1–63.9)	99.8 (99.5–100)
Lung	1588	9	15	80	84.2 (77.4–91.0)	99.4 (99.1–99.8)	89.9 (81.7–95.3)	99.1 (98.5–99.5)
Melanoma	1661	8	5	18	78.3 (62.4–94.2)	99.5 (99.3–99.8)	69.2 (48.2–85.7)	99.7 (99.3–99.9)
Bladder	1667	2	11	12	52.2 (36.5–67.9)	99.9 (99.7–100)	85.7 (57.2–98.2)	99.3 (98.8–99.7)
Brain	1674	6	0	12	100 (73.5–100)	99.6 (99.2–99.9)	66.7 (41.0–86.7)	100 (99.7–100)
Thyroid	1684	2	1	5	83.3 (75.6–91.1)	99.9 (99.8–99.9)	71.4 (29.0–96.3)	99.9 (99.7–100)
Lymphoma	1666	5	3	18	85.7 (71.6–99.8)	99.7 (99.5–99.9)	78.3 (56.3–92.5)	99.8 (99.5–100)
Leukemia	1663	2	6	21	77.8 (63.8–91.7)	99.9 (99.7–100)	91.3 (72.0–98.9)	99.6 (99.2–99.9)

^aMinus (-), absence of cancer; plus (+), presence of cancer.

^bn = Number of relatives.

CI, confidence interval; NPV, negative predictive value; PAC, probability of agreement of cancer; PANC, probability of agreement of no cancer; PPV, positive predictive value.

and PANC with corresponding 95% CI. The PPV and PAC measurements were higher than 75%, with many of them higher than 90% for breast, ovary, prostate, colorectal, pancreas, and lung cancers, as well as lymphoma and leukemia. Cancers of the female pelvic organs, e.g., cervix and endometrium, as well as bladder cancer were among the cancer types with the lowest PPV and PAC. In general, the NPV and PANC were more than 97.1% for all cancer sites. Although there were differences in these estimates across the cancer sites by the type of first-degree relative (e.g., parents vs siblings), as reported by the proband, these differences were not statistically significant. The estimated PPV and PAC on cancer status reported by probands were higher for siblings compared with parents for breast,

colorectal, and ovarian cancers, and higher for parents compared with siblings for prostate and lung cancers.

Estimates of PPV and PAC as well as NPV and PANC with corresponding 95% CIs for second-degree relatives are shown in Table 4. In general, the PPV and PAC of data reported by the proband for second-degree relatives were lower compared with those reported for first-degree relatives (Table 3), with notably large differences for cancer sites such as ovary, prostate, colorectum, and pancreas. The PPV and PAC for data reported by the proband on third-degree relatives were lower for all cancer sites, with the exception of cancers of the brain (71.4%), pancreas (71.4%), and female breast (69.8%), and for leukemia (72.7%) (data are not shown).

Table 4. Estimates of PAC, PANC, PPV, and NPV in second-degree relatives using data from proband interviews and reference standard

Type of cancer	Interview/reference standard				PAC (95% CI)	PANC (95% CI)	PPV (95% CI)	NPV (95% CI)
	^a -/-	+/-	-/+	+/+				
Females (<i>n</i>^b=633)								
Breast	496	12	22	103	82.4 (76.6–88.2)	97.6 (96.4–98.9)	89.6 (82.5–94.5)	95.8 (93.6–97.3)
Cervix	615	5	8	5	38.5 (16.6–60.3)	99.2 (98.7–99.7)	50.0 (18.7–81.3)	98.7 (97.5–99.4)
Endometrium	614	11	5	3	37.5 (3.2–71.8)	98.2 (97.8–98.7)	21.4 (4.6–50.8)	99.2 (98.1–99.7)
Ovary	590	9	19	15	44.1 (30.9–57.3)	98.5 (97.8–99.2)	62.7 (40.6–81.2)	96.8 (95.2–98.1)
Males (<i>n</i>=581)								
Prostate	526	10	15	30	66.7 (55.1–78.2)	98.1 (97.2–99.1)	75.0 (58.8–87.3)	97.2 (95.5–98.4)
Females and Males (N=1214)								
Stomach	1142	21	23	28	54.9 (43.0–66.8)	98.2 (97.6–98.8)	57.1 (42.2–71.2)	98.0 (97.1–98.7)
Colorectum	1106	18	38	52	57.8 (49.5–66.0)	98.4 (97.8–99.0)	74.3 (62.4–84.0)	96.7 (95.5–97.6)
Pancreas	1181	7	18	8	30.8 (16.9–44.7)	99.4 (99.1–99.7)	53.3 (26.6–78.7)	98.5 (97.6–99.1)
Lung	1123	11	33	47	58.8 (49.7–67.8)	99.0 (98.5–99.6)	81.0 (68.6–90.1)	97.1 (96.0–98.0)
Bladder	1191	1	13	9	40.9 (27.0–54.8)	99.9 (99.8–100)	90.0 (55.5–99.7)	98.9 (98.2–99.4)
Kidney	1199	3	9	3	25.0 (3.8–46.2)	99.8 (99.7–100)	50.0 (11.8–88.2)	99.3 (98.6–99.6)
Brain	1191	5	7	11	61.1 (42.0–80.2)	99.6 (99.3–99.9)	68.8 (41.3–89.0)	99.4 (98.9–99.8)
Lymphoma	1193	5	8	8	50.0 (29.7–70.3)	99.6 (99.3–99.9)	61.5 (31.6–86.1)	99.3 (98.7–99.7)
Leukemia	1185	1	5	23	82.1 (69.3–95.0)	99.9 (99.8–100)	95.8 (78.9–99.9)	99.6 (99.0–99.9)

^aMinus (–), absence of cancer; plus (+), presence of cancer.

^b*n* = Number of relatives.

CI, confidence interval; NPV, negative predictive value; PAC, probability of agreement of cancer; PANC, probability of agreement of no cancer; PPV, positive predictive value.

Predictors of Accuracy of Reported Family History of Cancer

Another objective of our analysis was to determine which of the proband's characteristics affect the disagreement between the family history data given by the proband and that of the reference standard. The results presented in Table 5 are restricted on the false-positive rate for all cancers combined and on the false-negative rate for all cancers combined, female breast cancer, prostate cancer, lung cancer, and colorectal cancer. After we accounted for familial correlations within a family, we found that male probands were more likely to overreport (i.e., report cases that were not true) all cancers combined compared with female probands ($p=0.0021$), and probands from clinic-based ascertainment were more likely to overreport all cancers combined compared with population-based probands ($p=0.0136$). In addition, probands ascertained by either mode were more likely to overreport cancer in first-degree relatives compared with second-degree relatives

($p=0.0161$). No statistically significant associations were observed between the false-positive rate and age of diagnosis, race or ethnicity, or family history of breast, ovarian, or colorectal cancer.

Younger probands were more likely to report family history data with a lower false-negative rate (Table 5), particularly for female breast cancer ($p=0.0008$), prostate cancer ($p=0.002$), and colorectal cancer ($p=0.027$). In general, with the exception of breast cancer, clinic-based ascertained probands were more accurate compared with population-based ascertained probands. Furthermore, clinic-ascertained probands reported false-negative rates that were significantly lower than those reported by population-ascertained probands for all cancers combined ($p=0.0217$), lung cancer ($p=0.038$), and colorectal cancer ($p=0.023$). Breast cancer probands were significantly more accurate than colorectal cancer probands in reporting breast cancer in their first-degree relatives. Finally, reporting by non-white probands was consistently more accurate than

Table 5. False-positive and false-negative rates by probands' and relatives' characteristics

Characteristics	False-positive rate (%)	False-negative rate (%)				
		All cancer	All cancer	Breast ^a	Prostate	Lung
Probands' characteristics						
Gender						
Male	6.1	9.8	33.2	25.0	37.5	15.2
Female	1.9*	14.1	10.9	28.7	29.1	27.9
Age at diagnosis (years)						
<50	3.1	13.9	7.1	17.1	35.5	19.5
50–59	1.6	12.9	7.1	20.0	30.2	30.4*
60–69	2.7	13.6	16.9	41.2	31.7	21.4
70+	1.9	13.2	22.2**	41.9**	29.6	22.6*
Race/ethnicity						
Non-Hispanic white	2.5	13.8	12.8	28.6	29.0	23.6
Nonwhite	1.0	8.4	8.3	25.0	41.2	23.1
Ascertainment						
Population-based	2.1	14.2	12.0	28.6	31.6	26.2
Clinic-based	6.2*	7.3*	15.7	25.0	7.7*	12.1*
Family history of breast						
No	2.7	14.1	22.8 ^b	24.1	32.3	20.9
Yes	1.9	12.4	19.5	38.2	26.6	30.2
Family history of ovarian						
No	2.3	13.4	12.4	28.2	29.5	23.5
Yes	3.9	13.5	13.8	30.0	40.0	23.8
Family history of colorectal						
No	1.8	15.1	11.2	23.2	30.6	51.3 ^b
Yes	4.0	9.6	18.1	41.9*	28.7	36.8
Type of proband						
Breast	2.3	14.3	9.6	28.8	27.8	36.5
Ovarian	0.0	17.8	10.0	28.0	33.3	15.6
Colorectal	4.8	9.9	29.6**	26.7	34.0	26.3
Relatives' characteristics						
Gender						
Male	2.2	13.8		28.3	30.5	22.1
Female	2.5	13.2	12.5		29.3	25.0
Age at diagnosis (years)						
<50		12.0	11.9		33.3	15.1
50–59		13.5	13.1	55.6	31.8	33.3
60–69		12.1	12.0	20.6*	30.8	25.9
70+		15.4	13.4	28.6*	27.4	22.1
Relationship to the proband						
1st degree	3.5	7.7	4.6	20.7	15.8	10.3
2nd degree	1.3*	16.6***	17.6***	33.3**	41.3***	42.2***
3rd degree	0.0	27.4***	30.2***	50.0*	46.4***	63.6***

*0.01 < *p* < 0.05.**0.001 < *p* < 0.01.****p* < 0.001.^aFemales only.^bRate based on second- and third-degree relatives.

reporting by white probands for all cancer sites, with the exception of lung cancer; however, these differences were not statistically significant.

The relationship of the relative to the proband was found to be the most consistent predictor of accuracy among the false negatives. Second- and third-degree relatives were more likely to be reported inaccurately by the probands for all cancers combined, female breast cancer, lung cancer, prostate cancer, and colorectal cancer (*p*<0.001). Gender of the relative and age at diagnosis of the relative were not significant predictors,

except in the case of prostate cancer, whereby the proband was more likely to be accurate for a relative diagnosed at an older age (*p*=0.010 for 60–69 years and *p*=0.023 for 70+ years) compared with a relative diagnosed at a younger age.

Discussion

Although family history information is collected in clinical and research settings and is used to infer risk of the disease in population-based, case-control, cohort,

or family-based studies, little information is available on the accuracy of a proband's reported family history.¹⁴⁻¹⁸ In the limited studies of validation of family history of cancer, most reports concentrate on the first-degree relatives of the proband, and most studies use case-control study designs. Inaccurate reporting of the disease status of relatives of the proband can result in biased estimates of familial aggregation and represent a major source of misclassification in genetic and epidemiologic studies. In fact, in case-control studies, nondifferential misclassification of the disease status of the relatives results in biased estimates of the odds ratio toward the null.¹⁹ Differential misclassification can result in upward or downward estimates of the odds ratio.²⁰

We report here data indicating that cancer-affected probands report their family history of cancer with high PPV (Tables 3 and 4). In addition, the probability of the proband reporting a cancer in a relative, given that the relative had that cancer (PAC), is more than 95% for breast cancer in first-degree relatives; the probability of a true absence of cancer (PANC) in a relative given that the proband also reported a negative cancer history for that relative is more than 95% for first- and second-degree relatives. However, we noted major differences in the quality of data by cancer site and degree of relationship of the relative to the proband. The validity of data for most types of cancer among first-degree relatives was similar in our study to that reported by Airewele et al.¹⁸ but higher than those reported by Kerber and Slattery.¹⁷ This finding might account for the way these studies were conducted compared with the present study. In both our study and that of Airewele et al.,¹⁸ multiple contacts were made with the families to test information reliability, whereas in Kerber and Slattery¹⁷ that situation was not true. Similar to other reports, we found poor validity of data reported on individual cancer sites of the female pelvic organs, such as cervix or endometrium, as well as bladder cancer among first-degree relatives.

Reported family history of cancer by probands was significantly more accurate for first-degree relatives than for second- and third-degree relatives. The variability between the different cancer sites was larger for second- and third-degree relatives compared with first-degree relatives. Our study clearly indicates that reporting of family history is more reliable for selected cancers. Some cancers, such as female breast cancer, prostate cancer, and leukemia, were reported accurately for first-, second- and third-degree relatives. In addition, the PPV was more than 70% for most of the cancers in first-degree relatives and remained more than 70% for many of the cancer sites, such as breast, prostate, colorectal, lung, and bladder, and for leukemia in second-degree relatives.

In our study, similar to that of Kerber and Slattery,¹⁷ younger probands were better able to report familial

female breast cancer, familial prostate cancer, and familial colorectal cancer compared with older probands. Male probands were more likely than female probands to overreport all cancers combined. In general, nonwhite probands reported more accurate family history of cancer than white probands. However, the difference in accuracy was not significant. Similarly, we found no statistically significant associations with respect to accuracy of the family history of probands and race or ethnicity of the probands. In addition to age at diagnosis of the proband, the best predictor of reporting accuracy was the ascertainment source of the proband.

Because genetic counseling is done in clinical settings based on reported family history of cancer, obtaining an accurate family history of cancer among relatives is important in predicting cancer risk. Even though the proportion of families who were clinic based in our study was 6.3%, accounting for 305 (9.5%) relatives, we observed statistically significant differences in the validity of the reported data with respect to false negatives and false positives (Table 5). Probands from clinic-based ascertainment sources compared with population-based sources were more likely to overreport (report more cases than were true) cancer in their relatives but were also likely to be more accurate in their reporting.

As noted in Subjects and Methods, the ascertainment of this clinic-based group is an established referral procedure that we generated over the past 10 years, whereby primary care physicians, hematology or oncology specialists, gynecologic oncologists, and surgical oncologists refer their patients who are at high risk of cancer because they may have multiple affected relatives. The differences observed between population-based and clinic-based probands might be due, in part, to the fact that the clinic-based probands were more informed and more motivated about their risk. However, because of the small proportion of clinic-based families in our validation study, we recommend that this study be replicated with a larger series of clinic-based ascertained probands.

Limitations

A limitation of our study was that it did not include unaffected probands, who might be compared in their reporting with affected probands. However, the purpose of this study was to estimate the accuracy of family history reporting by affected probands, and the results are generalizable to this group. Another limitation of our study was that the number of clinic-based ascertained probands was small. However, even with the small sample size, the estimate of accuracy of reporting among the clinic-based probands was higher than the population-based probands. Finally, not all families were included in the study. The analyses were done for

all families that had any verification as of July 2000, and families that did not have verification by that date were not included. It is possible that families with a positive family history for cancer tend to respond and complete their verification. We do not believe that this is a source of bias, because our analyses were based on individuals rather than families.

To our knowledge, this is the largest validation study of familial cancers and the first to use population-based cancer family registries. We found high reliability for most cancer sites among first-degree relatives and moderate reliability for second- and third-degree relatives. Overreporting of cancer was rare (2.4%), similar to the 2.9% reported by Airewele et al.¹⁸ Race or ethnicity and gender of the proband did not influence the accuracy of reporting. However, the source of ascertainment of probands was important.

The cancer registry data were collected under subcontract 050K-8710 with the U.S. Department of Health Services as part of its statewide cancer reporting program, mandated by the Health and Safety Code, Sections 103875 and 203885. The personal interview data were collected with support by the National Institutes of Health, National Cancer Institute grant CA-58860; the National Institutes of Health, National Cancer Institute grant 5 U24 CA78134-05, University of California at Irvine–University of California at San Diego (UCI–UCSD) Cancer Genetics Network Center; and the Lon V. Smith Foundation grant LVS-18840. The authors wish to acknowledge the staff members of the UCI Epidemiology Division for their assistance in data collection and data processing. We also acknowledge the statistical advice we received from Dr. Robert Elston.

References

1. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am J Hum Genet* 1991;48:232–42.
2. Whittemore AS, Wu AH, Kolonel LN, et al. Family history and prostate cancer risk in black, white, and Asian men in the United States and Canada. *Am J Epidemiol* 1995;141:32–40.
3. Isaacs SD, Kiemeny LA, Baffoe-Bonnie A, Beaty TH, Walsh PC. Risk of cancer in relatives of prostate cancer probands. *J Natl Cancer Inst* 1995;87:991–6.
4. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66–71.
5. Wooster R, Neuhausen SL, Quirk Y, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. *Science* 1994;265:2088–90.
6. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
7. Claus ER, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer: implications for risk prediction. *Cancer* 1994;73:643–51.
8. Couch FJ, DeShano ML, Blackwood A, et al. BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. *N Engl J Med* 1997;336:1409–15.
9. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet* 1998;62:145–58.
10. Ziogas A, Gildea M, Cohen P, et al. Cancer risk estimates for family members of a population-based family registry for breast cancer and ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2000;9:103–11.
11. Peel D, Ziogas A, Li F, et al. Characterization of hereditary nonpolyposis colorectal cancer families from a population-based series of cases. *J Natl Cancer Inst* 2000;92:1517–22.
12. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–30.
13. Smith JP, Hagdu A. Sensitivity and specificity for correlated observations. *Stat Med* 1992;11:1503–9.
14. Anton-Culver H, Kurosaki T, Taylor TH, Gildea M, Brunner D, Bringman D. Validation of family history of breast cancer and identification of the BRCA1 and other syndromes using a population-based cancer registry. *Genet Epidemiol* 1996;13:193–205.
15. Aitken J, Bain C, Ward M, Siskind V, MacLennan R. How accurate is self-reported family history of colorectal cancer? *Am J Epidemiol* 1995;141:863–71.
16. Novakovic B, Goldstein AM, Tucker MA. Validation of family history of cancer in deceased family members. *J Natl Cancer Inst* 1996;88:1492–3.
17. Kerber RA, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. *Am J Epidemiol* 1997;146:244–8.
18. Airewele G, Adatto P, Cunningham J, et al. Family history of cancer in patients with glioma: a validation study of accuracy. *J Natl Cancer Inst* 1998;90:543–4.
19. Rothman KJ. *Modern epidemiology*. Boston: Little, Brown, 1986:84–9.
20. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol* 1991;134:1233–44.