

*Short Communication*How Accurately Does the Medical Record Capture Maternal History of Cancer?¹Gail Husson² and Lisa J. Herrinton

Division of Research, Kaiser Permanente, Oakland, California 94611

Abstract

We sought to assess the reliability of information regarding the maternal history of cancer by comparing the medical records of 214 women with breast cancer, ages 26–59 years and diagnosed in 1974–1995, and of their controls with the medical records of their mothers. Subjects were members of Kaiser Permanente, Northern California, selected for a study of early-life predictors of breast cancer. For any type of cancer identified in the mother's medical record, the proportion noted in the daughter's medical record at least 6 months before the daughter's diagnosis or reference date was 56% among cases and 32% among controls. The odds ratio for the association of maternal cancer history with breast cancer risk was 2.1 using the maternal record and 3.5 using the subject's record. For a maternal history of breast cancer, the proportion noted in the subject's record was 79% among cases and 57% among controls, and the odds ratios were 4.0 and 6.5, respectively. We believe that the case-control difference in missing information was attributable to higher utilization of breast cancer screening among cases. This study illustrates the need to assess the impact of screening differences on the ascertainment of information from the medical records.

Introduction

The family history of cancer is an important covariate in etiological studies of breast and other cancers and can be obtained from the medical record if: (a) the study subject was knowledgeable of her family history; (b) she reported her family history to her health care provider; and (c) the provider recorded the information in the record. Previous estimates of the reliability of the family history of disease have compared information recorded in the medical record with self-reported information obtained during an interview (1, 2). As part of a case-control study of early-life predictors of breast cancer, we obtained information not only from the subjects' but also from the mothers' medical records. Thus, we were able to estimate the accuracy of medical-record-based maternal cancer history using the maternal record as the standard.

Methods

The data were collected for a case-control study designed to assess the relationship of childhood height, weight, and other characteristics with risk of breast cancer. Cases with breast cancer and their controls were selected from the membership of Kaiser Permanente, a large health maintenance organization operating in Northern California since 1946. Women whose cancer was diagnosed during the period 1974–1995 and who had joined Kaiser Permanente when ≤ 12 years old were identified through the Kaiser Permanente Regional Cancer Registry, which reports directly to the Northern California Cancer Registry, a participant in the SEER³ program. Controls were sampled from the health-plan membership after frequency-matching on age and birth year, and they were assigned a reference date so that the distribution of reference dates in the controls corresponded to the distribution of diagnosis dates in the cases. The controls were required to have been members of Kaiser Permanente on their reference date. Of the 214 cases and 214 controls included in the case-control study, 38 were excluded from the present analysis because 9 (6 cases, 3 controls) were adopted and 29 (19 cases, 10 controls) did not have a maternal medical record. Thus, 189 cases and 201 controls were included in the present analysis. The study was restricted to women who had first joined Kaiser Permanente at age 12 years or younger because the primary hypothesis of the study concerned early-life characteristics. Hence, maternal records were generally available in the health plan. This restriction and the fact that the earliest possible year of entry into the health plan was 1946 resulted in the upper age limit on cases and controls being 59 years. Trained medical record analysts obtained information on height, weight, family history, reproductive factors, hormone use, and history of benign breast disease recorded before the diagnosis/reference date from the subjects' medical records. We also obtained information from the parental medical records, including parental height, weight, and personal history of cancer.

The medical record analyst ascertained from the subject's chart a "history of cancer in the mother" from notes made 6 months or longer before the case's diagnosis date or the corresponding reference date in the control. Diagnosis year and cancer site were recorded if available. If there was no mention of family history in the medical record, the subject's maternal history was coded as negative. The medical record analyst ascertained from the mother's medical record the mother's history of cancer before the diagnosis or reference date. This information could have come from self-reported information of a personal history of cancer or from records of the cancer diagnosis or treatment. Again, no mention of a personal history was coded as negative, and

Received 10/25/99; revised 3/29/00; accepted 5/8/00.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Supported by National Cancer Institute Grant K07 CA 70969.

² To whom requests for reprints should be addressed, at Division of Research, 3505 Broadway Avenue, Oakland, CA 94611. E-mail: gXH@dor.kaiser.org.

³ The abbreviations used are: SEER, Surveillance, Epidemiology, and End Results; DCIS, ductal carcinoma *in situ*.

information on diagnosis date and cancer site were recorded if available. The mother's cancer history was coded as "unknown" if more than 10 years had elapsed since the last chart entry before the diagnosis date or reference date in controls, or if minimal information was recorded in the medical record as determined by the medical record analyst.

Family history information from the subject's record was cross-tabulated with personal history information from the mother's record separately for cases and controls, and the odds ratios for the two sources of family history information were computed and compared to estimate the bias, and the κ statistic was computed to summarize the level of agreement (3). In further analyses, we examined whether case-control differences could be explained by differences in the distributions of: (a) the year of the mother's diagnosis; (b) the mother's age at diagnosis; or (c) the daughter's reference date. We hypothesized that an earlier diagnosis of breast cancer in the mother would have allowed greater time for the information to be transmitted to the daughter and to the provider and, in regard to reference date, that providers may have been more thorough in taking family histories in more recent years. These hypotheses were tested using the Wilcoxon rank-sum test (4) in the 91 subjects whose mothers had a history of cancer.

Results

The women included in the study were relatively young with one-half of them being diagnosed in their 20s and 30s and one-half being diagnosed during the earlier years of the study, 1974–1990. There were striking differences in missing information between cases and controls on some variables including marital status, alcohol use, reproductive history, and menopausal status (Table 1).

Fifty-seven (30%) of the mothers of cases and 34 (17%) of the mothers of controls had a history of cancer noted in their medical record. The proportion of these that were noted in the subject's (daughter's) medical record was 56% among cases and 32% among controls. The odds ratio of 2.1 for the association of any maternal cancer history with risk of breast cancer calculated using the maternal medical record changed to 3.5 when the subject's medical record was used as the source of information. κ a measure of the precision of the subject's medical record relative to the maternal medical record, was 0.59 for cases and 0.42 for controls. In addition, 128 case mothers had no recorded history of cancer in their records, but 4% of these were recorded as having had a history in the subject's record. For the 162 control mothers, this proportion was 1% (Table 2). We examined the case-control difference in the mother's age at diagnosis and the year the mother was diagnosed with cancer to evaluate whether or not the mothers of cases had earlier diagnoses. The mean age of the cancer diagnosis among mothers of cases was 58 years while the mean age among mothers of controls was 61 years ($P = 0.213$). For both cases and controls, the median year of the mother's diagnosis was 1978 and the median reference year was 1992.

We also examined reports specifically of breast cancer because this was the focus of our case-control study. Twenty-four (13%) mothers of cases and seven (3%) mothers of controls had a history of breast cancer noted in their medical record. The proportion of these that were noted on the subject's medical record was 79% among cases and 57% among controls (Table 3). This case-control difference in ascertainment changed the odds ratio for the association of

Table 1 Demographic and reproductive characteristics of Kaiser Permanente members with incident breast cancer diagnosed 1973–1995 and controls

Cases and controls were ≤ 12 years old at first entry into Kaiser Permanente during 1946–1975.

Characteristic	Cases (%) $n = 189$	Controls (%) $n = 201$
Age at reference date, yr		
26–35	43 (22.8)	51 (25.4)
36–40	56 (29.6)	54 (26.9)
41–45	55 (29.1)	60 (29.9)
46–59	35 (18.5)	36 (17.9)
Year of reference date		
1974–1987	48 (25.4)	54 (26.9)
1988–1990	36 (19.0)	37 (18.4)
1991–1993	48 (25.4)	56 (27.9)
1994–1995	57 (30.2)	54 (26.9)
Ethnicity		
White	141 (74.6)	155 (77.1)
African-American	32 (16.9)	33 (16.4)
Other	16 (8.5)	13 (6.5)
Marital status		
Single	27 (14.3)	33 (16.4)
Married	113 (59.8)	85 (42.3)
Other	36 (19.0)	18 (9.0)
Unknown	13 (6.9)	65 (32.3)
Alcohol use		
Nondrinker	43 (22.8)	42 (20.9)
Drinker	128 (67.7)	119 (59.2)
Unknown	18 (9.5)	40 (19.9)
Parity		
Nulliparous	53 (28.0)	62 (30.8)
1	42 (22.2)	36 (17.9)
2	62 (32.8)	48 (23.9)
≥ 3	32 (16.9)	37 (18.4)
Unknown	0 (0.0)	18 (9.0)
Age at first full-term birth, yr		
Nulliparous	53 (28.0)	62 (30.8)
15–19	37 (19.6)	25 (12.4)
20–24	38 (20.1)	51 (25.4)
25–29	35 (18.5)	20 (10.0)
30–42	20 (10.6)	17 (8.5)
Unknown	6 (3.2)	26 (12.9)
Age at menarche, yr		
9–11	38 (20.1)	33 (16.4)
12	48 (25.4)	45 (22.2)
13	42 (22.2)	37 (18.4)
14–18	22 (11.6)	27 (13.4)
Unknown	39 (20.6)	59 (29.4)
Menopausal status		
Premenopausal	148 (78.3)	132 (66.0)
Postmenopausal	29 (15.3)	18 (9.0)
Unknown	12 (6.3)	50 (25.0)
Mother's history breast cancer ^a		
Yes	38 (20.1)	15 (7.5)
No	150 (79.4)	183 (91.0)
Unknown	1 (0.5)	3 (1.5)
Mother's history of any cancer ^a		
Yes	57 (30.2)	34 (16.9)
No	128 (67.7)	162 (80.6)
Unknown	4 (2.1)	5 (2.5)

^a Those for whom there was no mention of personal history of breast cancer in the mother's record are coded "No." Those for whom more than 10 years had elapsed since last chart entry before the diagnosis date or reference date are coded as "Unknown."

maternal history of breast cancer with risk of breast cancer from 4.0 using the maternal record to 6.5 using the subject's record. The κ was 0.71 for cases and 0.66 for controls. The mean age of the mother's breast cancer diagnosis was 57

Table 2 Mother's history of cancer as recorded in the subject's and mother's medical records:^a Kaiser Permanente Northern California, breast cancer cases and controls, diagnosed with incident breast cancer 1974–1995, ages ≤ 12 years at entry into Kaiser Permanente since 1946

	Mother's cancer history recorded in the mother's medical record ^b			
	Cases (%)		Controls (%)	
	No <i>n</i> = 128	Yes <i>n</i> = 56	No <i>n</i> = 162	Yes <i>n</i> = 33
Mother's cancer history recorded in the subject's medical record				
No	123 (96)	24 (43)	158 (99)	22 (67)
Yes	5 (4)	32 (57)	2 (1)	11 (33)

^a OR, 2.1 (1.46–2.74) using mother's medical record; OR, 3.5 (2.98–4.01) using subject's medical record.

^b Unknowns excluded.

years for cases, compared with 63 years for controls ($P = 0.149$). The median for the year of mother's diagnosis was 1981 for cases and 1983 for controls ($P = 0.831$). The median for daughter's reference year was 1993 for cases and 1990 for controls ($P = 0.340$).

Discussion

We sought to estimate the quality of the medical record to determine a woman's maternal history of cancer by comparing the subject's and mother's medical records. The primary limitation of this study that should be kept in mind when interpreting the results pertains to the generalizability of the sample. Our sample comprised younger women, who were long-term members of a health maintenance organization and who were diagnosed with breast cancer during 1974–1995. The age distribution of the women at diagnosis may differ from other studies using the medical record to ascertain family history. Also, having nearly complete access to the medical records that cover such a long period of a woman's life is unusual.

With respect to missing information on covariates such as marital status, alcohol use, and presumably family history, we believe, through indirect evidence, that the case-control difference can be attributed to earlier identification of breast cancer among frequent users, who would have undergone more frequent breast cancer screening. Frequent users would also have more opportunity to provide information on the covariates. Evidence for this comes from the comparison of the proportion of DCIS in our sample relative to the proportion in the entire health plan because DCIS is diagnosed through screening. In our sample of women diagnosed during 1974–1995, 20% of the cases had a diagnosis of DCIS compared with 15% among cases region-wide in Kaiser Permanente Northern California during 1990–1995.⁴ In other words, we believe that, had the rate of screening been higher among the controls, more of them would have been identified as cases, and more of them would have been identified as having a maternal history of cancer and had their marital status and alcohol status recorded. We believe that differences in the recording of information because of knowledge of the subject's cancer status could not explain

⁴ G. Husson, unpublished data.

Table 3 Maternal history of breast cancer as recorded in the subject's and mother's medical records^a

	Mother's history of breast cancer recorded in the mother's medical record ^b			
	Cases (%)		Controls (%)	
	No <i>n</i> = 160	Yes <i>n</i> = 24	No <i>n</i> = 183	Yes <i>n</i> = 6
Mother's breast cancer recorded in the subject's medical record				
No	153 (96)	4 (17)	183 (100)	2 (33)
Yes	7 (4)	19 (83)	0 (0)	4 (67)

^a OR, 4.0 (3.42–4.58) using mother's medical record; OR, 6.4 (5.92–6.88) using subject's medical record.

^b Unknowns excluded.

the case-control difference, because we ascertained family history only if it was noted at least 6 months before the reference date. In addition, we re-abstracted charts to determine to what extent missing information was attributable to error by the medical record abstractor, but we found that the abstractor's error rate was very low and did not account for the case-control difference. In our sample, it is not necessary to be a user in order to be a case, because non-utilizing cases are noted in the Kaiser Permanente cancer registry through feedback from the local SEER registry. Thus, it was not utilization *per se* that differed between cases and controls, but it was utilization that increased the probability that a woman's breast cancer was diagnosed (*i.e.*, breast cancer screening).

In our sample of women ages 26–59 years, 30% of cases and 17% of controls had a maternal cancer history. At best, only 60% of all of the cancers were documented in the daughter's medical record, although for breast cancer, the percentage was higher. We believe that the case-control difference in the completeness of information in the subjects' medical records was attributable to greater utilization, especially breast cancer screening, among cases. Although we observed a case-control difference in the quantity of missing data, there was only a small case-control difference in the precision of the information.

We are not aware of any other studies comparing subjects' medical records with the medical records of their mothers. However, there have been comparisons of subjects' self-reported family history with their own medical record or the medical record of their relatives. Kerber and Slattery (5) compared interview information with genealogical information among 125 Utah residents with colon cancer, ages 30–79 years at diagnosis during 1992–1995, and 206 matched controls. The genealogy record had been linked to the Utah Cancer Registry. They observed a κ of 0.73 among cases and 0.58 among controls. Two other studies of breast cancer compared the subjects' interview information with the subjects' medical record and noted somewhat higher levels of agreement than we observed, as would be expected (1, 2).

As the medical record becomes more automated, it will be used more frequently in epidemiological studies. Knowledge of its limitations is crucial in both the design and interpretation of studies. This study allows researchers who obtain data from medical records to estimate the degree of misclassification resulting from missing information on family history. It should be kept in mind that the amount of missing information in our

population may be different from that in other populations because these women are long-time members of the health maintenance organization. This study also emphasizes the need to assess the impact of past screening differences and health plan utilization differences on exposure ascertainment from the medical record in analytical studies of cancers for which screening is available.

References

1. Cook, L. S., White, E., Schwartz, S. M., McKnight, B., Daling, J. R., and Weiss, N. S. A population-based study of contralateral breast cancer following a first primary breast cancer (Washington, United States). *Cancer Causes Control*, 7: 382–390, 1996.
2. Horwitz, R. I. Comparison of epidemiologic data from multiple sources. *J. Chronic Dis.*, 39: 889–896, 1986.
3. Stokes, M. E., Davis, C. S., and Koch, G. G. *Categorical Data Analysis Using the SAS System*, pp. 98–100. Cary, NC: SAS Institute Inc., 1995.
4. Cody, R. P., and Smith, J. K. *Applied Statistics and the SAS Programming Language*, Ed. pp. 128–130. Englewood Cliffs, NJ: Prentice-Hall Inc., 1991.
5. Kerber, R. A., and Slattery, M. L. Comparison of self-reported and database-linked family history of cancer data in a case-control study. *Am. J. Epidemiol.*, 146: 244–248, 1997.