Journal of Genetic Counseling (© 2007)
DOI: 10.1007/s10897-006-9067-y

Original Research

Accuracy of Self-Reported Personal History of Cancer in an Outpatient Breast Center

Francisco J. Dominguez,[1] Christine Lawrence,[2] Elkan F. Halpern,[3] Brian Drohan,[2] Georges Grinstein,[2] Dalliah M. Black,[1] Barbara L. Smith,[1] Michele A. Gadd,[1] Michele Specht,[1] Daniel B. Kopans,[3] Richard H. Moore,[3] Sherwood S. Hughes,[3] Constance A. Roche,[1] and Kevin S. Hughes[1,4]

The self-reporting of cancer history is becoming increasingly important, as it frequently guides medical decision-making. We studied the accuracy of personal cancer history using a self-administered questionnaire, comparing the results with the Tumor Registry at our institution. Among 39,662 records, we identified 3614 women with a single cancer in the Tumor Registry who reported none or one cancer on their questionnaire. The sensitivity in self-reporting cancers was 85.7%, ranging from 92.1% for breast cancer to 42.9% for leukemia. The accuracy for breast cancer and Hodgkin's Lymphoma was significantly better than other cancers ($p = 0.00027$, CI: 1.4–3.88). Analysis of patient's characteristics showed that Caucasians reported breast cancer more accurately than Asian/Pacific Islanders ($p = 0.008$), and those with Jewish ancestry more accurately than non-Jewish ($p = 0.0435$). These results will help us to improve data collection and thus improve medical decision-making.

KEY WORDS sensitivity; neoplasm; data collection; data reporting; questionnaires.

INTRODUCTION

In the absence of a universal medical record, clinicians are dependent on patient recollection and reporting to identify personal and family history. This information is particularly important in cancer patients, as it will frequently guide the medical decision-making and management options. Patients are often asked to complete self-administered questionnaires in physician's offices and other care facilities in effort to increase efficiency. The data collected is used to guide treatment decisions and referrals for consultation, often without substantial verification. Thus the accuracy of self-reported information has significant implications for health care. If reporting is incomplete, or incorrect, patients may not receive appropriate medical care, or referral to a Risk Clinic or genetic counselor may be overlooked. Prior studies have analyzed the accuracy of self-reported medical history (Balamouth *et al.*, 1965; Madow, 1973; Harlow and Linet *et al.*,1989; Kehoe *et al.*, 1994; Bergmann *et al.*, 2004; Paganini-Hill and Chao, 1993), and self reported cancer family history (Kerber and Slattery, 1997; Love *et al.*, 1985; Murff *et al.*, 2004; Eerola *et al.*, 2000; Garbers *et al.*, 2001; Anton-Culvert *et al.*, 1996; Parent *et al.*, 1995; Parent *et al.*, 1997; Ziogas and Anton-Culver, 2003), but few studies have analyzed the accuracy of self-reported data on personal cancer history (Desai *et al.*, 2001;

[1]Surgical Oncology Division, Massachusetts General Hospital, Harvard Medical School, 32 Fruit street YAW-7, Boston, Massachusetts, USA.
[2]Department of Computer Science, University of Massachusetts Lowell, Lowell, Massachusetts, USA.
[3]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 32 Fruit street YAW-7, Boston, Massachusetts USA.
[4]Correspondence should be directed to, Kevin S. Hughes, Surgical Oncology Division, Massachusetts General Hospital, Harvard Medical School, 32 Fruit street YAW-7, Boston, Massachusetts, USA; e-mail: kshughes@partners.org.

1059-7700/07 © 2007 National Society of Genetic Counselors, Inc.

Manjer *et al.*, 2004; Parikh-Patel *et al.*, 2003; Schrijvers *et al.*, 1994; Bergmann *et al.*, 1998). These studies used mailed questionnaires or in-person interviews and reported variable agreement for personal cancer history reporting.

This study was designed to determine the accuracy of reporting a personal cancer history using a self-administered scan form questionnaire or Tablet PC system, comparing results with the tumor registry data at our institution. This study is unique in that it deals with a highly efficient, low cost approach using a scan form in a mammography population. If the accuracy is acceptable, it opens the possibility of collecting substantial numbers of family histories with minimal effort. This information will help improve data collection and accuracy and consequently help provide better patient care. In addition, this information has implications for studies based on self-administered questionnaires.

PATIENTS AND METHODS

Women presenting to the Avon Comprehensive Breast Evaluation Center at the Massachusetts General Hospital from May of 2003 through July of 2005 completed a self-administered questionnaire that gathered family history and personal risk information using a scannable form or a tablet computer system. Under an IRB approved protocol, we reviewed this database. Patients selected which cancers she has had from a list that includes breast, bladder/kidney, colorectal, invasive cervical, leukemia, lung, Hodgkin's lymphoma, non-Hodgkin's lymphoma, melanoma, ovarian, thyroid, uterine, and other. The data was uploaded into a database and the information was available to clinicians for patient care the same day. A total of 39,662 records from 29,636 individuals were linked by medical record number with the hospital tumor registry datasets. We only used data from the most recent scan form and included only women who self-reported none or a single cancer and had only a single cancer filed in the tumor registry. Women who reported "other cancer" or reported a cancer but were not listed in the tumor registry were excluded from the overall analysis,but are reported in Table I.We determined the sensitivity and false-negative rate of cancer reporting. Specificity and false positive rate was not studied, as it was not possible to confirm or deny cancers reported by patients whose data is not in the tumor registry (The MGH Tumor Registry only records cancers diagnosed or treated at MGH, thus we could not confirm or deny information not in the Registry). In addition, we analyzed the differences in the accuracy of cancer reporting by age at study, time since diagnosis, ethnicity and genetic background among the group of patients with breast cancer.

Significance of differences was accomplished using two sided Chi-square and Fisher exact probability tests as appropriate. P values represented statistical significance at $p < 0.05$.

RESULTS

Among the 29,636 individuals, 3614 women had a single indexed cancer in the Tumor Registry and reported none or only one cancer on the questionnaire. Breast cancer (72.6%), colorectal cancer (5.9%), uterine cancer (4.6%) and ovarian cancer (3.6%) were the most frequently reported cancers in these patients (Table I).

For the following analyses, we excluded 'Other', as we felt this was not a fair or representative answer. The overall sensitivity in self-reporting cancers was 85.7%, with the highest rate for breast cancer (92.1%), and the lowest for leukemia (42.9%), with a false negative rate ranging from 7.9% (breast) to 57.1% (leukemia) (Table I). The sensitivity for breast cancer reporting was significantly better as compared with all the other cancers except Hodgkin's Lymphoma. ($p < 0.001$, CI: 1.4, 3.88).

Analysis of age at study, time since diagnosis, ethnicity and Jewish or Hispanic background compared with the accuracy of cancer self-reporting among the group of patients who reported breast cancer is presented in Table II. The sensitivity of cancer reporting was similar among different age groups. The analysis of ethnicity showed that Asian/Pacific Islanders reported their breast cancer significantly less frequently than Caucasians ($p = 0.008$), but otherwise there was not a statistically significant difference between the accuracy of cancer reporting among Caucasians/White, African American/Black, Caribbean/West Indians and American Indians. The study of the Jewish or Hispanic background showed that patients with Jewish ancestry reported their breast cancer significantly more often than non-Jewish ($p = 0.044$), and the patients reporting a Spanish origin had a similar rate of cancer reporting when compared with non-Spanish patients. There was some under-reporting of breast cancer in the year of diagnosis, but good reporting thereafter.

Table I. Sensitivity of Patient-reported Personal Cancer History

Cancer	Individuals	%	Sensitivity (%)	False negative (%)	OR	95% CI	*p* value
Breast	2624	72.6	92.1	7.9	0.42	0.26, 0.71	< 0.001
Hodgkin's Lymphoma	24	0.6	83.3	16.7	0.98	0.22, 3.35	1.0
Ovarian *	130	3.6	83.1	16.9	1.00		
Thyroid	86	2.4	81.4	18.6	1.12	0.55, 2.28	0.75
Lung	90	2.5	77.8	22.2	1.40	0.71, 2.76	0.32
Bladder/Kidney	85	2.4	76.5	23.5	1.51	0.77, 2.98	0.25
Melanoma	51	1.4	74.5	25.5	1.68	0.77, 3.66	0.19
Uterine	167	4.6	61.7	38.3	3.05	1.75, 5.31	< 0.001
Colorectal	213	5.9	61.0	39.0	3.13	1.84, 5.35	< 0.001
Non-Hodgkin's Lymphoma	64	1.8	60.9	39.1	3.15	1.59, 6.21	< 0.001
Invasive Cervical	73	2.0	45.2	54.8	5.95	3.11, 11.40	< 0.001
Leukemia	7	0.2	42.9	57.1	6.55	1.37, 31.32	0.024
Total	3614	100	85.7				

*Reference group.

Table II. Analysis of Age, Ethnicity and Jewish or Hispanic Background Related with Cancer Self-report

Characteristic	Individuals	Sensitivity (%)	OR	95% CI	P value
Age at Study					
≤ 44[a]	244	91.0	1.00		
45–64	1426	93.1	1.33	0.78, 2.18	0.25
65–74	545	92.7	1.25	0.69, 2.21	0.42
75–84	349	88.8	0.79	0.43, 1.41	0.39
≥ 85	60	88.3	0.75	0.29, 2.19	0.53
Background					
Caucasian or White[a]	2211	95.2	1.00		
Asian or Pacific Islander	47	85.1	0.29	0.13, 0.79	0.008
African American or Black	40	92.5	0.63	0.19, 3.23	0.44
Caribbean or West Indian	18	94.4	0.59	0.13, 36.46	0.59
American Indian/Aleutian/Eskimo	5	100.0			
No response/Unknown/Other	303	70,6	0.12	0.09, 0.17	< 0.001
Jewish ancestry					
Yes	173	98.3	3.16	1.04, 15.74	0.044
No[a]	2061	94.7	1.00		
No response/Unknown	390	75.6	0.17	0.13, 0.24	< 0.001
Spanish Origin					
Yes	62	91.9	0.6	0.24, 1.97	0.25
No[a]	2208	95.0	1.00		
No response/Unknown	354	74.3	0.15	0.11, 0.21	< 0.001

[a]Reference group.

There were 303 individuals who failed to provide a response for background, 390 who failed to provide a response for the presence or absence of Jewish heritage, and 354 that failed to provide a response for Spanish origin. These individuals were significantly more likely to also fail to report their cancer history with a two sided p value < 0.001 in all cases (data not shown).

DISCUSSION

These results show the sensitivity of personal cancer history reporting to be about 85.7% for all cancers and 92.1% for breast cancer. The high concordance we obtained using a self-administered questionnaire is comparable to or better than the results obtained by other investigators using the more labor intense method of personal interviews (Desai *et al.*, 2001; Manjer *et al.*, 2004; Bergmann *et al.*, 1998). On the other hand, the high rates of breast and ovarian cancer reporting are expected results among a group of women presenting to a breast center for cancer screening or follow up, in a large academic medical center. This high agreement is extremely important for the health provider, as important management decisions, particularly for patients in whom a hereditary cancer syndrome is suspected,

are frequently based on personal and family history information.

These findings have also helped us to improve the way we collect information, as most cancers with sensitivities under 70%, have in common ambiguous or confusing labels on the questionnaire. Some of these labels may require a higher degree of medical sophistication than can be expected. An individual diagnosed with 'colon' or 'rectal' cancer might not recognize 'colorectal cancer', while 'uterine' may be a confusing term for an individual who was told she had 'endometrial' cancer. 'Non-Hodgkin's lymphoma' and 'invasive cervical' might be too complicated. This information provides us with 2 approaches to improving our questionnaire, both of which are being pursued. We have revised the questionnaire to simplify and clarify labels that are too technical or medical in nature. For example, we have changed 'invasive cervical cancer' to 'cervical cancer', and changed 'colorectal cancer' to 'colon or rectal cancer'. In addition, we are exploring the possibility of uploading data from our Tumor Registry prior to the patient's visit, and then asking the patient to verify this information, rather than asking her cancer history de novo. We anticipate that these 2 modifications will increase the accuracy of the data collected.

Response to ethnicity questions and questions about cancer history might be considered sensitive. Differences observed may represent non-responders lack of trust in the health care system, but alternatively, non-responders might not understand how to use the questionnaire or may be annoyed by the questionnaire and refuse participation. Thus, patients who do not answer sensitive questions will be asked by the computer why they are not answering. Those unable to answer will be offered help. Those refusing to answer will be encouraged to answer in order to improve the quality of their health care.

Understanding the current level of sensitivity of self-administered questionnaires will help clinicians to use this data more selectively and cautiously. In addition, improving our methods of obtaining self-reported data will increase the accuracy and reliability of this information. Understanding the limitations of self-reported data, and developing methods to increase accuracy, is extremely important for health providers, as identification, decision-making and management for patients with a high-risk of hereditary cancer syndromes are frequently based on this information.

ACKNOWLEDGEMENTS

We want to thank Carol Venuti, Administrative Director, Dana Farber/Harvard Cancer Center Registry Services, for her help with the Tumor Registry data set.

REFERENCES

Anton-Culver, H., Kurosaki, T., Taylor, T. H., Gildea, M., Brunner, D., & Bringman, D. (1996). Validation of family history of breast cancer and identification of the BRCA1 and other syndromes using a population-based cancer registry. *Genet Epidemiol*, *13*, 193–205.

Balamuth, E. (1965). Health interview responses compared with medical records. Vital and health statistics, series 2, no 7. Rockville, MD: National Center for Health Statistics.

Bergmann, M. M., Calle, E. E., Mervis, C. A., Miracle-McxMahill, H. L., Thun, M. J., & Heath, C. W. (1998). Validity of self reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol*, *147*, 556–562.

Bergmann, M., Jacobs, E. J., Hoffmann, K., *et al.* (2004) Agreement of self-reported medical history: Comparison of an in-person interview with a self-administered questionnaire. *Eur J Epidemiol*, *19*, 411–416.

Desai, M. M., Bruce, M. L., Desai, R. A., & Druss, B. G. (2001). Validation of self reported cancer history: A comparison of health interview data and cancer registry records. *Am J Epidemiol*, *153*, 299–306.

Eerola, H., Blomqvist, C., Pukkala, E., Pyrhonen, S., & Nevanlinna, H. (2000). Familial breast cancer in southern Finland: how prevalent are breast cancer families and can we trust the family history reported by patients? *Eur J Cancer*, *36*, 1143–1148.

Garbers, V., Toniolo, P. G., & Taioli, E. (2001). Changes in self-reported family history of breast cancer with change in case-control status. *Eur J Epidemiol*, *17*, 517–520.

Harlow, S. D., & Linet, M. S. (1989). Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol*, *129*, 233–248.

Kehoe, R., Wu, S. Y., Leske, M. C., & Chylack, L. T. (1994). Comparing self reported and physician reported medical history. *Am J Epidemiol*, *139*, 813–818.

Kerber, R. A., & Slattery, M. L. (1997). Comparison of self-reported and database-linked family history of cancer data in a case-control study. *Am J Epidemiol*, *146*, 244–248.

Love, R. R., Evans, A. M., & Josten, D. M. (1985). The accuracy of patient reports of a family history of cancer. *J Chron Dis*, *38*, 289–293.

Madow, W. G. (1973). Net differences in interview data on chronic conditions and information derived from medical records. Vital and health statistics, series 2, no 57. Rockville, MD: National Center for Health Statistics.

Manjer, J., Merlo, J., & Berglund, G. (2004). Validity of self-reported information on cancer: Determinants of under and over-reporting. *Eur J Epidemiol*, *19*, 239–247.

Murff, H. J., Spiegel, D. R., & Syngal, S. (2004). Does this patient have a family history of cancer? An evidence based analysis of the accuracy of family cancer history. *JAMA*, *292*, 1480–1489.

Paganini-Hill, A., & Chao, A. (1993). Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. *Am J Epidemiol*, *138*, 101–106.

Parent, M. E., Ghadirian, P., Lacroix, A., & Perret, C. (1995). Accuracy of reports of familial breast cancer in a case control series. *Epidemiol*, *6*, 184–186.

Parent, M. E., Ghadirian, P., Lacroix, A., & Perret, C. (1997). The reliability of recollections of family history: implications for the medical provider. *J Cancer Educ Summer*, *12*(2), 114–120.

Parikh-Patel, A., & Allen, M., Wright, W. E. and the California Teachers Study Steering Committee (2003) Validation of Self-reported Cancers in the California Teachers Study. *Am J Epidemiol*, *157*, 539–545.

Schrijvers, C. T. M., Stronks, K., Van de Mheen, D. H., Coeberg, J. W., & Mackenbach, J. P. (1994). Validation of cancer prevalence data from records from a postal survey by comparison with cancer registry records. *Am J Epidemiol*, *139*, 408–414.

Ziogas, A., & Anton-Culver, H. (2003). Validation of family history data in cancer family registries. *Am J Prev Med*, *24*, 190–198.