Self and Registry Reported Cancer in a Population-Based Longitudinal Study

Barbara E. K. Klein, MD, MPH, **Kristine E. Lee, MS**, **Scot E. Moss, MA**, **Amy Trentham-Dietz, PhD**, and **Ronald Klein, MD, MPH**

University of Wisconsin School of Medicine and Public Health, Department of Ophthalmology & Visual Sciences (B.E.K.K., K.E.L., S.E.M., R.K.) and Department of Population Health Sciences (A.T.D.), Madison, Wisconsin, USA

Abstract

Purpose—To evaluate the concordance of cancer diagnosis from self and registry reports.

Methods—Self-reported diagnosis information from participants in a cohort study was compared with linkage data from the Wisconsin Cancer Reporting System.

Results—Overall, there was good agreement between self and registry reported cancers, with 90% of all matches being considered an exact match. Concordance varied by cancer site; agreement was excellent for breast (85.4%) and prostate (78.9%) cancers.

Conclusions—While self-reported cancer diagnoses for some cancers such as breast and prostate cancer is an important source of information and may be a reliable substitute when registry data are incomplete or not available, a combination of self and registry reports with mortality information may yield the most accurate information about cancer for purposes of health care planning and conducting epidemiologic studies.

INTRODUCTION

Population-based data for cancers are needed to estimate prevalence and to project incidence of cancers for estimating health care burden, to plan for care, to investigate possible causal associations, and to examine whether cancer may be a risk factor or indicator of other conditions. Self-reported information on cancer may be the only source of data in circumstances when cancer registry data are not available. In addition, self-reports may supplement deficits in registry data even when such data exist.

Other investigators have evaluated the concordance of self-reported cancer and cancer registry data. Parikh-Patel and colleagues used information from a mail survey[1] and both Desai and colleagues[2] and Simpson and colleagues[3] used data from in-person interviews to compare to registry information. In the study by Desai et al., the odds ratio (OR) for false negative reporting of cancer history was 10.00 (95% CI 1.65, 60.45) for nonwhites compared to whites. In addition, the odds ratios for false negative reporting increased with increasing age category; the ORs were 3.04 (0.76, 12.16), 2.29 (0.82, 6.42), 4.18 (1.46, 11.97), 10.06 (2.55, 39.73) for age groups 18-44, 65-74, 75-84, and ≥85 respectively compared to those 45-64 years of age. In the study by Simpson et al., specifics of false

Correspondence to: Barbara E. K. Klein, MD, MPH University of Wisconsin-Madison School of Medicine and Public Health Department of Ophthalmology and Visual Sciences 610 North Walnut St., 4th Floor WARF Madison, WI 53726-2336 Phone: (608) 263-0276 Fax: (608) 263-0279 kleinb@epi.ophth.wisc.edu.

negative rates of reporting by ethnicity and age were not given. However, the authors write that kappa scores used to compare self-reported condition with physician record of the medical conditions did decrease with age, but for diseases with excellent agreement (of which cancer was one) age did not affect agreement. Specific effects of race/ethnicity on agreement were not given.

Incidence of cancers by self-report was obtained from volunteers in the American Cancer Society's Cancer Prevention Study II and was compared against incidence obtained from cancer registries.[4] The investigators found that some registries under-ascertained cancer cases. Longitudinal data from population-based cohorts may be especially useful in evaluating the consistency of self-reports over time. Also, the number of cancer cases is likely to increase with increasing age of the population under study, and this is likely to improve estimates of the relationship of age to specific cancer sites from both sources of data. We examined self-reported data across four 5-year intervals from the Beaver Dam Eye Study population and compared the concordance between self versus registry reported diagnoses as well as the effects of demographic characteristics on the method of reporting.

MATERIALS AND METHODS

Participant Recruitment and Data Collection

The purpose of the Beaver Dam Eye Study was to collect information on the prevalence and incidence of age-related cataract, macular degeneration, and diabetic retinopathy. Participants underwent a baseline examination and survivors were invited to 3 follow-up examinations at 5-year intervals. Information pertaining to overall health and lifestyle, quality of life, environmental, and medical exposures was collected at each 5-year interval. Participants unable to come for an examination were interviewed over the phone. The tenets of the Declaration of Helsinki were followed, institutional review board approval was granted by the human subjects committee of the University of Wisconsin, and informed consent was obtained from each subject. Five thousand nine hundred twenty-four eligible individuals 43-84 years of age and living in the city and township of Beaver Dam, Wisconsin, were identified and invited for a baseline examination between March 1, 1988, and September 14, 1990.[5] Examinations were completed on 4,926 persons (83%). Ninety-nine percent of the population was non-Hispanic white.

Of 4,542 surviving participants from the baseline examination, 3,684 (81.1%) participated in the 5-year follow-up examination in 1993 to 1995 (Figure 1) and an additional 263 completed phone interviews. Similarly, 2937 participated in the 10-year follow-up examination in 1998-2000 and 2351 participated in the 15-year follow-up examination in 2003-2005. Comparisons of participants and non-participants at each examination phase have been detailed elsewhere.[6-9]

A medical questionnaire was administered at each study examination and phone interview. Each participant was asked, "Have you ever been told that you had cancer, skin cancer or malignancy of any type?" Responses were "no," "yes," and "don't know." If the response was "yes," they were asked, "How old were you when you were first told you had cancer?" and "What kind of cancer(s) was it (were they)?" Year of cancer diagnosis was calculated from age when only one cancer was reported. For additional cancers first reported during the follow-up visits, the year of diagnosis was assumed to occur midway during the examination interval (i.e., 2 years prior to current exam year).

Registry Reports and Match Definitions

The Wisconsin Cancer Reporting System (WCRS) was established in 1976.[10] Cancer reporting data are considered adequately complete beginning in 1980. Hospitals, physicians,

and clinics report cancer cases to the WCRS, in the Division of Public Health, Bureau of Health Information and Policy, Wisconsin Department of Health Services. Hospitals must report cases within six months of initial diagnosis or first admission following a diagnosis elsewhere. Clinics and physicians must report cases within three months of initial diagnosis or contact. All tumors with malignant cell types are reportable except basal and squamous cell carcinomas of the skin. Registry records for years covered in this study are estimated to be 95% complete according to the national standards for statewide completeness (all cancers combined) as calculated by the Centers for Disease Control and Prevention's National Program of Cancer Registries.

The WCRS conducted the data linkage and returned coded information for analysis using a random identification number to protect the confidentiality of the cancer registry data and comply with registry policies for data release. Several automated linkages were conducted based on seven combinations of exact matches by Social Security number, name (first, last, maiden, and initials) and/or date of birth. Then, registry staff manually reviewed inexact matches additionally using middle name, address, gender, and race/ethnicity.

We compared codes for cancer sites provided by the registry (categorized according to the International Classification of Diseases for Oncology11) with self-reports of cancer (categorized according to the International Classification of Diseases 912 codes). Code site groups corresponded to those used by the Surveillance, Epidemiology and End Results program (SEER)13 with a few exceptions: corpus uteri and uterus not otherwise specified were combined; vagina, vulva and other female genital organs were combined as other female genital organs; and all leukemias were combined. A self-report was considered to match the registry report if the cancer site was the same in both reports, or if the diagnosis year was within one year and the reports were of cancers that were anatomically proximate or similar in type for lymphomas, myeloma, leukemias, and other hematopoietic cancers (for example, "other oral" matched with "larynx"). A participant could have more than one cancer site/type and be counted multiple times.

Statistical Analysis

SAS (SAS Institute Inc., Cary, NC) was used for all analyses. Descriptive statistics (N, %, kappa) were carried out using PROC FREQ. Cohen's kappa coefficient with 95% confidence interval was calculated to provide a measure of correlation between the self and registry reports while accounting for the agreement expected by chance. Due to small numbers, the kappa statistic was calculated only for the three most commonly reported cancer sites. For concordance analysis, we distinguished between matches that were exact and those that matched using the broader criteria for proximate site or year of diagnosis ("any match").

To examine whether age, education or income were associated with accuracy of reported cancers, generalized estimating equations were used to include each distinct cancer while controlling for the correlation that resulted from multiple cancers for a person. We ran two separate models: one for exact site match between self and registry report, and another for any match between self and registry report.

Population for Analysis

Since registry data were considered complete from 1980 to 2003 at the time of the data linkage, all cancers (self-reported N=167 and registry N=57) outside those years were excluded from analysis (Figure 2). While the registry included melanoma of the skin, the self-reported cancers did not consistently distinguish melanoma from other skin cancers, so any skin cancer, including melanoma, was excluded from analysis (self-reported N=644, registry N=25). The registry data included many cancers that occurred after the last

examination for cohort participants. Those cancers were also excluded from these analyses (N=461).

There were 3,585 of the 4,926 persons included in the Beaver Dam Eye Study for whom there was no self-report of cancer and 685 persons where the only reported cancer was excluded from analysis. Among the 656 persons with self-reported non-skin cancer with diagnosis reported between 1980 and 2003, there were 720 unique cancers. 1,104 persons had at least one cancer that was reported from the registry; however, only 636 had non-skin cancers between 1980 and 2003 that also occurred prior to the last possible self-report year for that person. Among the 636 persons with registry reported cancer, there were 683 unique cancers. In total, there were 766 people reporting 867 unique cancers from either source (Figure 2). This group is used for all analyses.

RESULTS

Overall, 62% of the 867 cancers were reported by both self-report and registry, 21% were self-reported only and 17% were registry reported only. While 720 non-skin cancers were self-reported over the course of the 20-years of follow-up, only 272 diagnoses were reported at the baseline examination. An additional 155 cancers were initially reported at the first follow-up, 157 at the second and 136 at the third. Half of the 720 (N=355) self-reported cancers were diagnosed in persons who returned for follow-up visits. Of these 355 cancers, 24 (7.0%) were reported differently at later visits and 72 (20.0%) were not reported at later visits. These inconsistent self-reports of cancers were most common for oral, large bowel, and female genital cancers.

There was good agreement between self and registry reported cancers, with 90% of all matches being considered an exact match. Table 1 displays matches for different cancers identified in the population. As expected, large bowel cancer was commonly reported for men and women combined. While only 55% of large bowel cancers were defined as exact matches (i.e., both self-reports and registry reports list rectal cancer), 75% matched by broader criteria (e.g., self-report list colon cancer while registry lists rectal cancer). The kappa statistic for large bowel cancer was 0.74 (95% CI 0.67, 0.81) There were enough large bowel cancers to look at agreement for each specific site. Among the 122 reports of colon cancer, 104 were self-report. Of these, 55.8% were an exact match to the registry (kappa statistic was 0.60 [95% CI 0.52, 0.69]).

Similarly, of the 182 reported breast cancers, 159 were self-report with 84.3% also reported in the registry. There are no other cancers that could be considered under the any match (proximate site). The kappa statistic for breast cancer was 0.81 (95% CI 0.76, 0.85). Ninety-eight percent of the male genital cancers were prostate cancer. Of the 176 prostate cancers listed by self-report or in registry records, 147 were self-reported with 84% matched exactly in the registry with a kappa statistic of 0.77 (95% CI 0.71, 0.83).

For most cancers, agreement between registry and self-report for exact cancer match was similar to any match. Among the cancers or cancer groups with at least 10 persons with a self-report, breast, prostate, kidney cancers and lymphomas had the highest exact (as well as any) percent match between the two sources of report. Oral/throat and large bowel showed large improvements in agreement percentage when any match was considered. Oral/throat cancer had the poorest exact match percentage (25% of the 12 self-reported cancers were an exact match) but high "any" match percentage (83.3%). The latter finding is driven by 4 "other oral cavity" cancers (larynx coded in registry) and 3 "gum/other mouth" (other oral/throat cancer coded in the registry). Bone/connective cancers had the poorest match percentage (exact as well as any); many of these cancers were not reported to the registry

before 1992 due to the absence of clinic reporting prior to that time. However, when this type of cancer was reported in the registry, there was usually a corresponding self-report.

Age, sex, education and income were not significantly associated with concordance between self and registry reported cancer (Table 2) for either kind of match. When considering exact matches, gender was significant (P=0.05), with women slightly more likely to have a match than men.

DISCUSSION

Concordance of self and registry reported information on non-skin cancers is good in this population-based longitudinal study. Concordance was not significantly associated with age, gender, education or income for "any match" in our population, although gender was significant (if only barely so) for "exact match." Younger age has been reported to be associated with greater concordance of similar data in other studies,[1-4] as was non-Hispanic ethnicity[1,2] and greater education.[3,4] Our finding with regard to age was compatible with significant findings in the other studies. The lack of significant effect of education and income may reflect the relatively homogeneous nature of our population. An obvious failing of self-reported information was that tissue-specific characteristics (e.g., cell type and characteristics of the tumor) were lacking while cancer registries had full descriptions of cancers that include 3 categories of information: site (coded from ICD-O11), morphology (e.g., cell type; especially melanoma, lymphoma, leukemia, and myeloma), and behavior (e.g., malignant, secondary, and in-situ). These characteristics are important with regard to prognosis and etiology.

Self-reported data may miss some cancers, especially if there is only a small window for which self-reported data are available. Some cancers may affect participation in a study, thus would not be self-reported. Only 4% of self-reported cancers were lung cancer, which was lower than might be expected. However, a high percentage of the 481 registry reported cancers that occurred after our last participant visit were lung cancers. Of the 97 persons with registry reported lung cancers that were reported after we last saw the participant, 90% had lung cancer listed as the primary cause of death. For such fatal cancers, supplementation of self-reported data with death certificate information would be useful. However, for cancers such as breast, only 60% of the 20 breast cancers reported after the last participant visit were listed on the death certificate.

A limitation of registry data is that they only capture cancers diagnosed and treated in a particular geographic region. In Beaver Dam, many individuals spend winters out of state. The WCRS receives cancer reports directly from registries in 19 other states; however, to the extent that cancer diagnosis and treatment occurred exclusively outside of those 19 states, Wisconsin would not capture those data in its cancer registry. For studies that have many subjects in multiple states, obtaining registry data for each state may not be feasible, and even if obtained, there may not be comparable completeness and accuracy between state registries. Additionally, the small number of people reporting rare cancers constrained our ability to estimate comparability of self versus registry reporting for many types of cancer.

The registry reported data that we used from WCRS may have limited our results in two ways. Not all cancers were biopsied or removed in hospital facilities and may not have been read by hospital-based pathologists. Our data may have been affected to the extent that pathologists at different medical facilities may differ in their accuracy of reporting. Some cancers that did not require hospitalization and were identified in clinics were not required to be reported prior to 1992. It is likely that this led to underreporting of such cancers that occurred before 1992.

In summary, good comparability was found between self and registry reported cancers. Self-report is an important source of information on these conditions and may be the only source when registry data are unavailable or incomplete. Physicians often rely on patient self-report of personal and family histories of cancer (and other conditions). The accuracy of the self-report varies by type of cancer and often improves when cancer groups (such as cancer of throat, mouth, larynx combined or large bowel) are considered. These results are relevant for both clinical practice and research. If specific site information is needed, such as a distinction between colon and rectal cancer, then registry data would be preferred. For fatal cancers, inclusion of mortality information would greatly improve estimates of these cancers. As use of electronic medical record data increases, linkage between records systems may reduce our reliance on patient self report and improve completeness of central cancer registries. At this time, a thoughtful combination of self and registry reports and mortality information may yield the most accurate information about cancer for purposes of ascertaining cancer histories, health care planning, and conducting epidemiologic studies.

Acknowledgments

REFERENCES

1. Parikh-Patel A, Allen M, Wright WE. Validation of self-reported cancers in the California Teachers Study. Am J Epidemiol. 2003; 157:539–545. [PubMed: 12631544]
2. Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. Am J Epidemiol. 2001; 153:299–306. [PubMed: 11157418]
3. Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM, Fried LP. Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. J Am Geriatr Soc. 2004; 52:123–127. [PubMed: 14687326]
4. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. Am J Epidemiol. 1998; 147:556–562. [PubMed: 9521182]
5. Klein, R.; Klein, BE. Manual of Operations. U.S. Department of Commerce; Springfield, VA: 1991. The Beaver Dam Eye Study. NTIS Accession No. PB91-149823
6. Klein R, Klein BE, Linton KL, De Mets DL. The Beaver Dam Eye Study: visual acuity. Ophthalmology. 1991; 98:1310–1315. [PubMed: 1923372]
7. Klein R, Klein BE, Lee KE. Changes in visual acuity in a population. The Beaver Dam Eye Study. Ophthalmology. 1996; 103:1169–1178. [PubMed: 8764783]
8. Klein R, Klein BE, Lee KE, Cruickshanks KJ, Chappell RJ. Changes in visual acuity in a population over a 10-year period: The Beaver Dam Eye Study. Ophthalmology. 2001; 108:1757–1766. [PubMed: 11581046]
9. Klein R, Klein BE, Lee KE, Cruickshanks KJ, Gangnon RE. Changes in visual acuity in a population over a 15-year period: the Beaver Dam Eye Study. Am J Ophthalmol. 2006; 142:539–549. [PubMed: 17011842]
10. Wisconsin State Legislature. Wisconsin State Legislature. Wisconsin State Legislature; Madison, WI: 2009. Wisconsin State Statute 255.04. Cancer reporting. serial onlineAvailable from:

http://nxt.legis.state.wi.us/nxt/gateway.dll?f=templates&fn=default.htm&d=stats&jd=ch. %20255Accessed January 29, 2009

11. International Classification of Diseases for Oncology. 3rd ed.. World Health Organization; Geneva: 2000.

12. Centers for Disease Control and Prevention. International Classification of Diseases, Ninth Edition (ICD-9). National Center for Health Statistics; Hyattsville, MD: 1979.

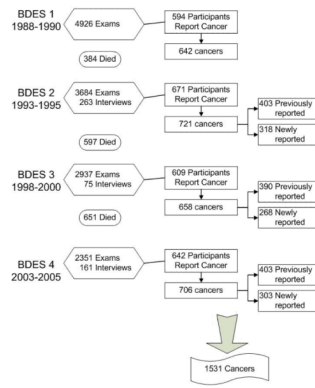13. SEER Program Coding and Staging Manual. National Cancer Institute, 2007; Bethesda, MD: 2007.

Figure 1.
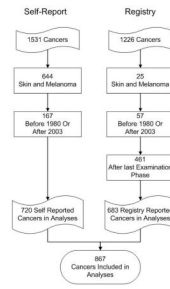Participants at each examination and numbers of self-reported cancer.

Figure 2.
Cancers included in analyses.

Table 1

Distribution of Matches Between Self-Reports and Registry Records Among 766 Adults with 867 Unique Non-Skin Cancer Diagnoses, 1980–2003.

Site of Cancer*	Either Source N	Self-Report			Registry		
		N	Exact Match %	Any Match %	N	Exact Match %	Any Match %
ORAL/THROAT	14	12	25.0	83.3	12	25.0	58.3
LARGE BOWEL	135	116	55.2	75.0	106	60.4	82.1
DIGESTIVE	28	18	50.0	61.1	21	42.9	57.1
LUNG/BRONCHUS	40	33	60.6	NA	27	74.1	NA
RESPIRATORY	15	7			13	38.5	76.9
BONE/CONNECTIVE	17	14	14.3	21.4	6		
BREAST†	182	159	84.3	NA	157	85.4	NA
FEMALE GENITAL	92	79	57.0	69.6	68	66.2	80.9
MALE GENITAL	179	147	81.6	83.0	154	77.9	79.2
URINARY	77	64	76.6	79.7	64	76.6	79.7
EYE/ORBIT	5	4			3		
BRAIN/CNS	8	1			8		
ENDOCRINE	6	4			4		
NON-HODGKIN LYMPHOMA	24	18	83.3	83.3	21	71.4	76.2
LEUKEMIA	19	14	57.1	57.1	13	61.5	69.2
MYELOMA	2	1			2		
OTHER HEMATOPOIETIC	3	3			0		
OTHER	6	3			4		

*
98% of the male genital cancers were prostate.

† BREAST site includes male breast, but none were observed in this population.

Table 2

Comparison of Participant Characteristics by Match Status for Self-Reports of Cancer with Registry Reports

Characteristic	Exact Match			Any Match		
	No (N=374)	Yes (N=471)	P-value[*]	No (N=320)	Yes (N=525)	P-value[*]
Age, years (mean ± SD)	70.71 ± 10.7	70.75 ± 9.6	0.14	71.08 ± 10.9	70.52 ± 9.6	0.63
Gender, % male	51	48	0.05	51	48	0.23
Education, years (mean ± SD)	11.69 ± 2.9	11.87 ± 2.9	0.26	11.67 ± 2.9	11.87 ± 2.9	0.23
Income (%)			0.53			0.60
< $10,000	19.6	16.6		18.8	17.4	
$10,000-19,000	24.2	26.9		23.5	27.1	
$20,000-29,000	19.6	20.4		19.5	20.4	
$30,000-44,000	15.3	16.9		16.3	16.2	
≥ $45,000	21.2	19.2		22.0	18.9	

[*]
P-value from generalized estimating equation model with all factors included.