Improving Cancer Incidence Estimates for American Indians in Minnesota

ABSTRACT

Objectives. The purpose of this study was to estimate cancer incidence for American Indians in Minnesota.

Methods. Indian Health Service enrollment data were linked to the Minnesota tumor registry to identify cancers among American Indians in Minnesota. Incidence rates for the 5 most common cancers in this population, estimated after the linkage, were compared with rates estimated before the linkage and with rates for the total population of Minnesota.

Results. The linkage identified 302 cancer cases not previously identified as occurring among American Indians in Minnesota. Postlinkage estimates suggested that incidence rates for prostate and colorectal cancer are similar to those for the total population of Minnesota, but that rates of lung and cervical cancer are significantly higher. Breast cancer rates are slightly lower than those for the total population of Minnesota but more than twice as high as previous estimates for American Indians.

Conclusions. The postlinkage estimates suggest different priorities for cancer education, prevention, and control than might be assumed from either prelinkage estimates or previously published data, and underscore the importance of using accurate and specific data for setting these priorities. (*Am J Public Health.* 1999;89:1673–1677)

Melissa R. Partin, PhD, Stephen J. Rith-Najarian, MD, Jonathan S. Slater, PhD, Jane E. Korn, MD, MPH, Nathaniel Cobb, MD, and John T. Soler, MPH

Although less than 1% of the US population identified themselves as American Indian in the 1990 US Census, this group represents more than 300 tribal groups distinct in culture, history, and health behaviors.[1] Because of the great diversity and small size of the American Indian population, accurately estimating the cancer burden and concomitant cancer prevention and control needs for a specific region or tribe is a difficult task. The task is complicated by 3 limitations that plague the data sources most often used to estimate cancer incidence in American Indian populations[1]: limited generalizability,[2] systematic underestimation, and inconsistent racial classification.[3]

With regard to limited generalizability, most existing data on the incidence of cancer among American Indians come from the Surveillance, Epidemiology, and End Results (SEER) tumor registries. Funded by the National Cancer Institute, these registries provide some of the richest periodic information about cancer incidence, stage of diagnosis, and survival that is available for US populations. However, the most recent race-specific data available from SEER relate only to American Indians living in New Mexico and Alaska.[2] Because previous studies have documented substantial variation in cancer incidence rates across regions and tribes of American Indians,[3–4] generalizing the SEER data to American Indians living outside of New Mexico and Alaska is problematic.

With regard to systematic underestimation, the Indian Health Service (IHS) provides regional cancer mortality estimates annually,[5] and cancer incidence rates based on hospital discharge data for the years 1980 through 1987 were recently published for American Indian and Alaska Native populations that used IHS services.[4] Although the rates published in this report provide more detail than previously available about regional and tribal differences in cancer incidence, these rates may substantially underestimate cancer incidence rates for some anatomic sites of cancer by excluding cancer diagnoses that did not require hospitalization or were made outside of the IHS network. Furthermore, because of probable variation by cancer site in the degree to which hospital discharge data underestimate cancer occurrence, these data are of questionable utility for assigning relative priorities to cancer prevention and control efforts directed toward cancers at different sites.

With regard to inconsistent racial classification, one alternative to relying on SEER or IHS data for identifying cases of cancer among American Indians in a specific region is to use data from state or regional tumor registries. However, because previous research has shown that American Indians are often misclassified as being of other races in health-related data sources,[6–13] this approach can also lead to underestimation of the incidence of cancer in this population if measures are not taken to more completely and accurately identify individuals of American Indian race in these data sets.[6,12] One solution to this problem is to link tumor registry data with data such as IHS enrollment rosters, which more reliably identify American Indians. Such linkages have been accom-

Melissa R. Partin, Jonathan S. Slater, Jane E. Korn, and John T. Soler are with the Minnesota Department of Health, Minneapolis; Dr. Partin is also with the Minneapolis VA Medical Center. Stephen J. Rith-Najarian is with the Indian Health Service, Albuquerque, NM. Nathaniel Cobb is with the Indian Health Service, Bemidji, Minn.

Requests for reprints should be sent to Melissa R. Partin, PhD, Center for Chronic Disease Outcomes Research (2E/152), Minneapolis VA Medical Center, 1 Veterans Dr, Minneapolis, MN 55417 (e-mail: melissa.partin@med.va.gov).

This article was accepted June 4, 1999.

plished in at least 2 states.[6,12] Because these data are regionally specific and identify cancers at different sites with equal accuracy, they are better suited to setting priorities for cancer education, prevention, and control policies in these regions than are either IHS hospital discharge or SEER data. The present study used this linkage approach to improve ascertainment of cases of cancer among American Indians in Minnesota.

Methods

Estimating Incident Cases of Cancer

The Minnesota Cancer Surveillance System (MCSS) has collected information about all Minnesota residents with newly diagnosed cancers since January 1988. Minnesota law requires that all confirmed malignant tumors diagnosed in Minnesota residents be reported to the MCSS. Completeness of case ascertainment in the MCSS has been estimated at 99.6%, and only 0.3% of records lack information on the primary surveillance elements of age and date at diagnosis, sex, county of residence, primary site, and histology.[14] However, information on race is missing from approximately 20% of MCSS records.

To improve the ascertainment of American Indian race, we linked the MCSS, using the procedures described below, to IHS rosters of individuals enrolled in the Bemidji area (which includes service units in Minnesota, Wisconsin, and Michigan). The enrollment rosters used for this study included active, inactive, and deceased IHS enrollees. Individuals in the enrollment rosters with residences outside of Minnesota were not excluded from the rosters for linkage purposes, since it is possible that they had moved to Minnesota without notifying the IHS of their change in residence. Non-American Indians enrolled in the IHS were eliminated from consideration in the counts of incident cases of cancer.

MCSS data for the period from January 1, 1988, through December 31, 1993, were linked to Bemidji area IHS enrollment rosters (cumulative through December 1995) by means of a probabilistic record linkage system. The linkage system is based on Bayesian analysis and decision theory[15] and is modeled (with corrections) on the approach provided in the SEER Data Management System User's Guide.[16] The system involves making pairwise comparisons, on the basis of last name, first name, middle initial, date of birth, and social security number, between every record in the 2 data sets that are being linked. Comparisons of the data sets are made separately for males and females. After both common faults in data (such as missing values and inconsistencies in the spelling of names) and the frequency with which certain names, birth dates, and other data occur in the data sets being linked are taken into account, each comparison is assigned a weight that reflects the extent of agreement in the fields being compared. On the basis of these weights, each comparison is declared a match, a nonmatch, or an undecided "gray zone." Decisions about gray zones are adjudicated manually by comparing additional information about specific cases, such as address, aliases, and telephone numbers.

The MCSS record linkage system has been tested with high-quality data sets in which all of the key identifiers have valid (i.e., nonmissing) data. In these tests, sensitivity for the MCSS record linkage system was found to be 98.2% and specificity was greater than 99.9%. Data sets with missing values tend to produce lower sensitivity and specificity. MCSS data for the key fields compared in the linkage were highly complete. In the IHS data set, however, approximately 9% of records were missing social security numbers, and approximately 0.01% were missing dates of birth. To increase the specificity and sensitivity of the MCSS–IHS linkage, we chose thresholds that resulted in a larger number of gray-zone cases than in the tests referred to above and conducted with more complete data sets.

MCSS cases that either matched the Bemidji area IHS enrollee roster or were originally identified in the MCSS as involving persons of American Indian race were counted as American Indian cancer cases when incidence rates were being calculated. Cases coded as American Indian in the MCSS but not appearing in the IHS roster were included in the incidence estimates for American Indians, since not all American Indians living in Minnesota are registered with local IHS offices.

Estimating the Population at Risk

The number of American Indians in Minnesota at risk for developing cancer during the period examined was calculated by summing estimates of the number of self-identified American Indians living in Minnesota provided by the US Bureau of the Census for the years 1988 through 1993. Estimates for 1990 were derived directly from the 1990 US Census. Estimates for 1991 through 1993 by means of linear regression techniques were projected with the most current 10 years of Indian birth and death data provided by the National Center for Health Statistics.[17] Estimates for 1988 and 1989 were derived from a smoothing technique that produced a gradual transition in population estimates between the 1980 and 1990 US Census years.[17]

Computation and Comparison of Rates

Average annual cancer incidence rates for the period from 1988 through 1993, directly age-standardized to the 1970 US total population, were calculated separately for males and females. The 95% confidence interval (CI) for each incidence rate was estimated with methods described by Armitage and Berry.[18]

The rates of cancer estimated for American Indians were compared with the rates for the total population of Minnesota through use of the standardized incidence ratio (SIR). Each SIR was calculated by taking the ratio of the observed number of cancer cases among American Indians in Minnesota to the number that would be expected if American Indians had the same age-specific cancer incidence rates as the general population of Minnesota. Byar's formula for Poisson-distributed observations[19] was used to obtain 95% CIs for each SIR. Because of the small size of the population on which these estimates were based, incidence estimates and SIRs are reported only for the 5 most common cancers occurring among American Indians in Minnesota.

Results

Identification of Cases of Cancer Among American Indians in Minnesota

Table 1 shows the number and percentage of cases of cancer among American Indians in Minnesota that were identified by combining IHS and MCSS data according to the source of identification. Combining these data revealed a total of 329 cases of cancer among male and 466 cases of cancer among female American Indians. Linking the MCSS to the IHS enrollment rosters led to the identification of 118 more cases of cancer among male and 184 more cases of cancer among female American Indians than would have been identified through the use of MCSS race data alone (see Table 1, row 2). Inclusion of cancer cases classified as occurring in American Indians in the MCSS but not matching the enrollment rosters for the Bemidji area IHS added 64 cases among male and 120 cases among female American Indians to the numerator of our incidence estimates (see Table 1, row 3).

Cancer Incidence Estimates

Table 2 shows estimates of cancer incidence rates derived before and after the linkage for the 5 most common cancers among American Indians in Minnesota. These can-

TABLE 1—Number and Percentage of American Indian Cancer Cases Identified Through the Linkage of Data From the Indian Health Service (IHS) and the Minnesota Cancer Surveillance System (MCSS), by Source of Identification

Source of identification	Male		Female	
	No.	%	No.	%
In IHS enrollment file and coded as American Indian in the MCSS	147	44.7	162	34.8
In IHS enrollment file but not coded as American Indian in the MCSS	118	35.9	184	39.5
Not in IHS enrollment file but coded as American Indian in the MCSS	64	19.5	120	25.8
Total	329	100	466	100

TABLE 2—Selected Age-Standardized[a] Invasive Cancer Incidence Rates per 100 000, Before and After Linkage of Data From Indian Health Service and Minnesota Cancer Surveillance System, 1988–1993

Cancer Site	Minnesota Total Population, 1988–1993		Minnesota American Indians 1988–1993, Before Linkage		Minnesota American Indians 1988–1993, After Linkage	
	Males	Females	Males	Females	Males	Females
Prostate	150.9	...[c]	61.9	...[c]	111.9	...[c]
	(148.8, 153.0)	...[c]	(42.6, 81.3)	...[c]	(85.8, 138.0)	...[c]
Lung/bronchus	65.6	33.1	59.3	36.0	97.1	55.5
	(64.2, 67.1)	(32.2, 34.1)	(64.4, 109.9)	(21.7, 50.4)	(73.2, 121.0)	(37.7, 73.2)
Colon/rectum	57.2	39.1	37.9	26.1	56.5	37.6
	(55.4, 59.0)	(37.8, 40.4)	(23.2, 52.6)	(13.7, 38.6)	(38.4, 74.6)	(22.9, 52.3)
Breast	...[c]	109.5	...[c]	44.5	...[c]	82.9
	...[c]	(107.7, 111.2)	...[c]	(29.5, 59.6)	...[c]	(61.8, 104.0)
Cervix	...[c]	7.5	...[c]	19.2	...[c]	28.8
	...[c]	(7.0, 7.9)	...[c]	(10.2, 28.4)	...[c]	(17.5, 40.2)
All sites	464.7	329.0	279.2	211.2	441.6	338.9
	(460.9, 468.4)	(326.1, 332.0)	(239.3, 319.1)	(177.7, 244.8)	(391.2, 491.9)	(296.3, 382.5)

[a]Rates are standardized to the 1970 US population.
[b]95% confidence intervals are shown in parentheses.
[c]Rate not calculated: fewer than 25 cases.

cers, ranked according to the magnitudes of their age-standardized incidence rates, were prostate, lung, breast, colorectal, and cervical cancers. For comparison purposes, rates calculated for the total population of Minnesota are also included in Table 2.

Comparing incidence rates obtained from MCSS data alone (before the linkage) with incidence rates obtained after supplementing the MCSS race data with information on IHS membership (after the linkage) gives a sense of how our estimates were affected by the more complete and accurate identification of American Indians. For each site of the 5 most common cancers among American Indians in Minnesota, the postlinkage estimates suggest much higher incidence rates than did the prelinkage estimates. Most notably, the postlinkage estimates for prostate and breast cancer incidence are almost twice as great as the prelinkage estimates, and these differences are statistically significant.

The SIRs in Table 3 offer 2 comparisons of cancer incidence in the American Indian and total populations of Minnesota: the comparison based on the prelinkage data and the comparison based on the postlinkage data. The prelinkage SIRs suggest that American Indians in Minnesota have anywhere from slightly to significantly lower cancer incidence rates for all but cervical cancer. For prostate, colorectal, breast, and all sites combined, however, the added cases identified by the linkage attenuate these differences. Most notably, the breast cancer incidence rate for American Indians increases from 62% below the rate for the general population before the linkage to 33% below that for the general population after the linkage, and the prostate cancer incidence rate increases from 52% below that for the general population to 14% below that for the general population after the linkage. With lung cancer, the prelinkage SIRs suggest that its rate among American Indians in Minnesota is no different from that for the general population. The postlinkage estimates, however, suggest that the rate among American Indians is 50% to 60% higher.

Discussion

The postlinkage estimates from this study paint a somewhat different picture of relative cancer burden among American Indian residents of Minnesota than might be assumed from either the prelinkage estimates or other, previously published estimates of cancer incidence in American Indian populations. The increased cervical cancer incidence rates found among American Indians in Minnesota in this study highlight a well-recognized need for prevention and control services.

If, however, either the prelinkage data or data from other published studies[2,4] were used for setting cancer education, prevention, and control priorities for the American Indian communities in Minnesota, lung, colorectal, prostate, and breast cancer may not have been considered topics of concern. By contrast, the postlinkage estimates suggest that public education and/or prevention and control efforts for these cancers are certainly warranted. For instance, although the value of mammography for American Indian women has been debated,[4,20–22] our study found postlinkage incidence rates of breast cancer among American Indians comparable to the rates for the general US population estimated 20 years ago, when mass mammography screening was first widely endorsed.[23] This strongly suggests

TABLE 3—Selected Standardized Incidence Ratios[a]: Incidence Rates for American Indians in Minnesota[b] Compared With Those for the General Population of Minnesota, 1988–1993

Cancer Site	Before the Linkage[b,c]		After the Linkage[b,c]	
	Males	Females	Males	Females
Prostate	0.48	...[d]	0.86	...[d]
	(0.34, 0.65)	...[d]	(0.67, 1.08)	...[d]
Lung/bronchus	0.96	0.97	1.58	1.51
	(0.68, 1.31)	(0.63, 1.43)	(1.22, 2.01)	(1.08, 2.07)
Colon/rectum	0.75	0.63	1.08	0.92
	(0.49, 1.09)	(0.37, 0.99)	(0.77, 1.48)	(0.61, 1.32)
Breast	...[d]	0.38	...[d]	0.67
	...[d]	(0.27, 0.53)	...[d]	(0.52, 0.85)
Cervix	...[d]	2.06	...[d]	2.94
	...[d]	(1.27, 3.14)	...[d]	(1.98, 4.19)
All sites	0.64	0.58	1.01	0.91
	(0.56, 0.74)	(0.50, 0.67)	(0.90, 1.13)	(0.81, 1.03)

[a]Rates are standardized to the 1970 US population.
[b]Based on cancer incidence rates calculated before and after the linkage of data from the Indian Health Service and the Minnesota Cancer Surveillance System.
[c]95% confidence intervals are shown in parentheses.
[d]Ratio not calculated: fewer than 25 cases.

that the incidence of breast cancer is certainly high enough among American Indians in Minnesota to warrant promoting the use of mammography in this population.

Similarly, the finding that lung cancer is the second most common cancer among both male and female American Indian residents of Minnesota suggests that developing effective cigarette smoking prevention and cessation programs, particularly those targeting youths, should be a top priority for cancer prevention and control efforts in the American Indian communities of Minnesota. The finding that their incidence rates of colorectal cancer are similar to those for the general population suggests that American Indians in Minnesota stand to benefit as much as the general population by adhering to the colorectal cancer screening guidelines recommended by the US Preventive Services Task Force.[24] Finally, the finding that prostate cancer is the most common cancer among American Indian men in Minnesota suggests that information about the benefits and limitations of screening and treatment for prostate cancer may soon be in high demand by this population.

Although the estimates derived in this study provide more accurate and specific information than those derived in previous studies, this study's estimates may still underrepresent the incidence of cancer among American Indians in Minnesota for the following reasons. First, because not all American Indian residents of Minnesota are registered with the IHS, some American Indians not identified as such in the MCSS may have been missed in our numerator counts. Second, because the means for identifying American Indians in the denominator of our formula (self-identification in the US Census) are less restrictive than the means for identifying American Indians in the numerator (enrollment in the Bemidji Area IHS or medical record identification), our estimates are probably smaller than those that would have been obtained if the numerator and denominator data had been derived from the same identification procedures. Third, although our linkage procedures did not automatically classify individuals with missing social security numbers as nonmatches, we might have identified more matches if the IHS data had had fewer records missing on this item.

Conclusion

The data obtained from the linkage used in our study are specific to American Indians in Minnesota, identify cancers for all anatomic sites with equal accuracy, and correct for potential inaccuracies in the classification of American Indians in the cancer registry data used in the study. Consequently, the linkage data provide better guidance for setting cancer education, prevention, and control priorities for American Indians in Minnesota than do either previously published or prelinkage estimates. Unlike other estimates of cancer incidence in American Indian populations,[2,4] the estimates from our linkage study suggest that lung cancer incidence rates are significantly higher among American Indians in Minnesota than in the general population, and that prostate and colorectal cancer incidence rates are similar to those in the general population. The increased cervical cancer rates and lower breast cancer rates among American Indian women in Minnesota revealed by the linkage are consistent with expectations based on prior research.[3,4,11] However, the postlinkage estimates reveal a greater excess of cervical cancer than do the prelinkage estimates and suggest that breast cancer incidence rates are more similar to the rates for the general population than previously assumed. These findings suggest different priorities for cancer education, prevention, and control than have been inferred from previous studies[4,20–22] and underscore the importance of using accurate and specific data for making local health-related policy decisions. □

Contributors

M. R. Partin, N. Cobb, J. S. Slater, J. E. Korn, and S. J. Rith-Najarian contributed to the conception and design of the study. J. T. Soler and M. R. Partin conducted the analysis, and all authors contributed to the interpretation of the analysis. M. R. Partin wrote the first draft of the manuscript, and all authors contributed to subsequent revisions.

Acknowledgments

This study was supported by Cooperative Agreement U57/CCU 506748–06 from the Centers for Disease Control and Prevention.

References

1. Ericksen EP. Problems in sampling the Native American and Alaska Native Populations. In: Sandefur GD, Rindfuss RR, Cohen B, eds. *Changing Numbers, Changing Needs: American Indian Demography and Public Health*. Washington, DC: National Academy Press; 1996:113–132.
2. Miller BA, Kolonel LN, Bernstein L, et al., eds. *Racial/Ethnic Patterns of Cancer in the United States 1988–1992*. Bethesda, Md: National Cancer Institute; 1996. NIH publication 96-4104.
3. Mahoney MC, Michalek AM. A meta-analysis of cancer incidence in United States and Canadian Native populations. *Int J Epidemiol*. 1991; 20:323–327.
4. Nutting PA, Freeman WL, Risser DR, et al. Cancer incidence among American Indian and Alaska Natives, 1980 through 1987. *Am J Public Health*. 1993;83:1589–1598.
5. Cobb N, Paisano RE. *Cancer Mortality Among American Indians and Alaska Natives in the United States: Regional Differences in Indian Health, 1989–1993*. Rockville, Md: Indian Health Service, US Dept of Health and Human Services; 1997. IHS publication 97-615-23.
6. Sugarman JR, Holiday M, Ross A, Castorina J, Hui Y. Improving American Indian cancer data in the Washington State cancer registry using linkages with the Indian Health Service and tribal records. *Cancer*. 1996;78:1564–1568.
7. Frost F, Tollestrup K, Ross A, Sabotta E, Kimball E. Correctness of racial coding of Ameri-

can Indians and Alaska Natives on the Washington State death certificate. *Am J Prev Med.* 1994;10:290–294.
8. Sugarman JR, Soderberg R, Gordon JE, Rivara FP. Racial misclassification of American Indians: its effect on injury rates in Oregon, 1989 through 1990. *Am J Public Health.* 1993;83: 681–684.
9. Watson CC, Bennett T, Reed FW, McBroom WH, Helgerson SD. Classification of American Indian race on birth and infant death certificates—California and Montana. *MMWR Morb Mortal Wkly Rep.* 1993;42:220–223.
10. Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in US infants. *JAMA.* 1992;267: 259–263.
11. Bleed DM, Risser DR, Sperry S, Hellhake D, Helgerson SD. Cancer incidence and survival among American Indians registered for Indian Health Service care in Montana, 1982–1987. *J Natl Cancer Inst.* 1992;84:1500–1505.
12. Frost F, Taylor V, Fries E. Racial misclassification of Native Americans in a Surveillance, Epidemiology, and End Results cancer registry. *J Natl Cancer Inst.* 1992;84:957–962.
13. Frost F, Shy KK. Racial differences between linked birth and infant death records in Washington State. *Am J Public Health.* 1980;70: 974–976.
14. Minnesota Department of Health. *The Occurrence of Cancer in Minnesota 1988–1994: Incidence, Mortality, Trends.* Minneapolis: Minnesota Cancer Surveillance System Biennial Report, May 1997.
15. Felligi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc.* 1969;2:172–178.
16. Nelson D. *SEER Data Management System II Linkage System Statistical User's Guide.* Bethesda, Md: National Cancer Institute; 1982.
17. Indian Health Service. *Trends in Indian Health—1995.* Rockville, Md: Indian Health Service, US Dept of Health and Human Services; 1995.
18. Armitage P, Berry G. *Statistical Methods in Medical Research.* Oxford, England: Blackwell Scientific Publications; 1987:117, 401.
19. Breslow NE, Day NE. *Statistical Methods in Cancer Research, Volume II—The Design and Analysis of Cohort Studies.* New York, NY: Oxford University Press; 1987:69–71.
20. Nutting PA, Calonge BN, Iverson DC, Green LA. The danger of applying uniform clinical policies across populations: the case of breast cancer in American Indians. *Am J Public Health.* 1994;84:1631–1636.
21. Funk K. Breast cancer mortality and the use of mammography within the IHS. *IHS Provider.* 1986;11:136–139.
22. Partin MR, Korn JE, Slater JS. Questionable data and preconceptions: reconsidering the value of mammography for American Indian women. *Am J Public Health.* 1997;87:1100–1102.
23. Miller BA, Ries LAG, Hankey BF, et al., eds. *SEER Cancer Statistics Review: 1973–1990.* Bethesda, Md: National Cancer Institute; 1993. NIH publication 93-2789.
24. US Preventive Services Task Force. *Guide to Clinical Preventive Services.* 2nd ed. Baltimore, Md: Williams & Wilkins; 1996.