Availability and Accuracy of Cancer and Smoking Data Obtained from Next of Kin for Decedents in a Retrospective Cohort Study

Kyle Steenland, PhD; and Teresa Schnorr, PhD

We report on the availability and accuracy of next-of-kin data on 372 decedents gathered in the conduct of a retrospective cancer incidence study. Two mailed questionnaires were followed by phone calls for nonrespondents. The overall response rate for the next of kin of decedents was 59%, markedly less than the 79% response rate for living cohort members. Validity of data provided by next of kin regarding cancer incidence and smoking habits of the decedents was assessed via a comparison with medical records, which were considered accurate for the purposes of this comparison. For the 117 men for whom we had both types of records, next of kin were generally able to report accurately both whether the decedent ever had cancer (91% agreement) and the cancer type (84% agreement). Next-of-kin data generally agreed with medical record data for number of years smoked or for number of years since quitting, whereas next of kin underreported the amount smoked by decedents. The difficulty in contacting next of kin and their frequent inability to provide detailed and accurate information on smoking may make it difficult to adjust for confounding by smoking in nested case-control studies.

Epidemiologists frequently conduct record-based cohort mortality studies. Investigators may then contact the cohort directly (including the next of kin of cohort decedents) to determine disease incidence rather than mortality. Investigators may also seek direct contact with the cohort (or a sample of it) in order to collect further data on unmeasured covariates (eg, smoking, diet).

From the National Institute for Occupational Safety and Health, 4676 Columbia Pkwy, Cincinnati, OH 45226 (address correspondence to Dr Steenland).

0096-1736/88/3004-0348$02.00/0

Although contacting those members of a cohort who are still alive generally yields a fairly high response rate and presumably accurate information,[1] the decedents in such studies pose a more difficult problem.[2] There are two major problems encountered. First, it is more difficult to contact the next of kin of decedents than to contact live cohort members. Second, the data provided by next of kin may not be as accurate.

The example discussed here stems from a record-based cohort mortality study of steelworkers exposed to acid mist.[3] After the mortality study was completed, we sought to contact all cohort members to collect information on cancer incidence and smoking habits, as part of a cancer incidence study (K. Steenland et al, unpublished data, 1987). For decedents we sought interviews with next of kin. We also sought medical records (admissions, discharges, autopsy, pathology) for the decedents.

The cohort was composed of 1,156 men who had been employed as steelworkers at three Midwestern plants. All men in this cohort had worked for at least 6 months in a steel mill, of which at least one day was between 1940 and 1965. By 1986, 373 (32%) of these men had died.

There are several questions of interest to epidemiologists which we have attempted to answer in this paper.

(1) To what degree is it possible to contact next of kin of decedents, including decedents who may have died a long time ago? How does the response rate for next of kin compare with that of live cohort members?

(2) For what percentage of decedents can medical records be obtained? Can a complete medical history be determined from these records? How often are smoking data available from these medical records?

(3) How much data on smoking and cancer incidence are available from next of kin?

(4) How well do the data on smoking and cancer incidence provided by next of kin compare with the same data available from medical records?

Methods

Interviewing the Next of Kin of Decedents and Live Cohort Members

Last known addresses for all cohort members (alive and dead) were available from personnel records from the prior cohort mortality study, but in many cases these addresses dated back many years. A better source of valid addresses for the next of kin of decedents was the death certificate. We also attempted to obtain more up-to-date addresses for next of kin via the Health Care Financing Administration (HCFA) (which pays benefits to widows who can be identified because their computerized record includes the Social Security number of their decreased husband). We also sought addresses for next of kin via the Internal Revenue Service (IRS) (which has addresses by Social Security number for all individuals who have recently paid taxes). Widows may continue to pay taxes using their deceased husband's Social Security number. Finally, for those next of kin for whom no valid addresses could be obtained, we sought telephone numbers from directory assistance in the city of last known residence. It should be noted that some of these data sources (IRS, HCFA) may not be available to other investigators outside the government.

Initial attempts to obtain interviews from next of kin were made by certified mail. To all next of kin for whom we had addresses we sent a questionnaire about the decedent's smoking habits and history of cancer incidence. This self-administered questionnaire took about 15 minutes to complete. If no response was obtained from the first mailing, yet the address was accurate (based on the return receipt), we sent a second letter. In case of nonresponse to both letters, we sought phone numbers via directory assistance. We then sought to administer the same questionnaire via phone.

The work of contacting next of kin by telephone was performed under contract by a company experienced in telephone interview techniques, including quality control. In addition to all those who did not respond to the mailing, we also followed up by telephone those whose mailed interview was deemed incomplete. The same sources and methods were used to trace living and dead cohort members, except for the use of additional information for decedents from death certificates.

Obtaining Medical Records for Decedents

The National Institutes for Occupational Safety and Health, under the 1970 legislation creating the agency, has legal access to medical records needed for research (Occupational Safety and Health Act, Public Law 91-596, 91st Congress, Dec. 29, 1970, Sections 20 and 8b). For the purpose of determining laryngeal cancer incidence, we sought medical records (for all 360 decedents for whom we had death certificates) from all hospitals and doctors listed on the certificate. For those who died at home, we wrote to major hospitals in their local areas. Coroners' reports were sought when coroners had signed death certificates. Medical records that were requested included all admissions, pathology reports, discharge reports, and autopsies.

Next-of-kin *v* Medical Records Data

We sought to compare information on cancer incidence and smoking history provided by the next of kin with the same data provided by the medical history. For cancer incidence, if the medical record included a lifetime medical history covering all major illnesses, the medical record was presumed to be accurate. For both smoking and medical history assessments, we only used medical record data which was recorded during the last year of a man's life.

Self-reported smoking histories are generally considered accurate in comparison to data provided by the next of kin. Smoking information on the medical history is self-reported, provided directly by the patient upon admission. Hence, whenever adequate smoking information was available from the medical record, we have considered that smoking information to be accurate in comparison with smoking data provided by next of kin. Subjects were classified as current, former, or never smokers. Subjects described as "not a smoker" or "nonsmoker" on medical records were excluded because they could not be classified. No smoking information on the medical record was used if that information was provided by someone other than the patient (eg, the patient was comatose at admission).

Each next of kin was asked "To your knowledge, did the deceased ever have cancer?" If yes, "What type (site) of cancer?" All types of cancer reported by the next of kin were recorded. All medical records were reviewed to determine whether adequate information existed to determine lifetime cancer incidence. Those with adequate information were then categorized by cancer type. Both next of kin and medical records could have listed more than one type of cancer for the decedent. Next of kin were not specifically asked to list primary as opposed to metastatic sites, and sometimes reported more than one site. Medical records may have listed more than primary site. If any cancer site reported by the next of kin matched any primary cancer site as reported in the medical records, we considered the records to be in agreement. Analyses consisted of calculating the extent of agreement between next of kin responses and medical records for the decedent ever having had cancer, as well as for cancer type. Sensitivity, specificity, and κ statistics were calculated to assess the degree of matching between sources. κ statistics reflect the level of agreement between two sources.[4]

Regarding smoking, each next of kin was asked to classify the decedent by cigarette smoking status (never, former, or current, evaluated at the year of

death). Next of kin also reported on duration of smoking (years), amount smoked (packs per day, divided into half-pack categories, eg, less than 1 was 1 half-pack, 1 to 2 was 1.5). Years since quitting were asked for former smokers. Similar data were abstracted from medical records to the extent possible. We measured agreement for smoking status by a κ statistic, whereas detailed smoking data (continuous) were assessed via pairwise correlations and paired t test.[4]

A number of decedents had quit smoking upon diagnosis of disease, often shortly before death. Whereas next-of-kin data reflected smoking status at death, medical record data might pre-date death by up to 1 year. To avoid confusion regarding former smokers, we arbitrarily defined as a current smoker any man who was smoking at any point within 2 years of death.

Results

Response Rate for Interviews

Table 1 indicates the response rate, as well as the type of interview conducted, for both the next of kin of decedents and for living cohort members. As expected, the response rate for the former (59%) was considerably lower than for the latter (79%). The principal problems for next of kin were the inability to identify any living next of kin or the inability to locate any valid address for them. This occurred for 32.8% of the decedents (n = 122). In contrast, we were unable to locate valid addresses for only 13.5% of the living cohort members.

For the decedents, 60% of the next-of-kin respondents were spouses, 18% were children, 10% were siblings, and 12% were others. Based on a random sample of 11% of the next of kin, the next of kin had known the decedent for an average of 38 years. Hence, on the average, the next of kin should have been able to provide such data as lifetime smoking histories or cancer incidence.

The response rate for next of kin increased by decade of death of the decedent. The response was 40% for those who died before 1960, 47% for 1960 to 1969 deaths, 66% for 1970 to 1979 deaths, and 72% for deaths after 1980. The average year of death was 1971.

Ability of Next of Kin to Provide Information

The next of kin generally indicated that they were able to tell us whether the decedent had cancer (96.3%).

TABLE 1
Interview Response Rate for Cohort, by Vital Status

	Live	Deceased (Next of Kin)
Total sought	783	373
Death certificates obtained	–	360 (96.5%)
Completed questionnaire	621 (79.3%)	220 (59.0%)
Mailed interview	480 (61.3%)	146 (39.1%)
Phone interview	141 (18.0%)	74 (19.9%)
No response	162 (20.7%)	153 (41.0%)
No address found	106 (13.5%)	122 (32.7%)
Some address found	56 (7.2%)	31 (8.3%)

Almost all next of kin were able to answer whether the decedent had ever smoked cigarettes (98.2%), whether the decedent had been a cigarette smoker at time of death (94.8%), and whether he had ever smoked a pipe or cigar (97.7%). However, fewer next-of-kin respondents could answer more detailed questions about cigarette smoking. The percentages of respondents who provided information about amount smoked and number of years smoked were 80.0% and 70.3%, respectively. For former smokers, 71.8% of next-of-kin respondents provided information on the year when the decedent quit smoking.

In general, among those next of kin who did respond, spouses were able to provide more information than other types of next of kin. Spouses represented about two thirds of those who were able to provide information on most variables, but only about one half of those who were not able to provide it.

Data Obtained Via Medical Records

Seventy-five percent of the decedents had died in hospitals. The other 25% had either died at home (17.5%) or died on the way to the hospital (7.5%).

A hospital record was obtained for 68.9% of all decedents. Lack of any hospital record was the result of our inability to identify a hospital where the deceased had been treated (17.5%) or the failure of a hospital to maintain past records (13.6%, especially for those dead-on-arrival).

We also sought records from all doctors (or coroners) who signed the death certificates. These sources provided records for only 10% of all decedents. Combining all sources of records, we obtained some type of information on 73.6% of the decedents.

The quality of records varied considerably. The cause of death was available for 67.8% of the decedents, which represented almost all men for whom some type of record could be obtained. However, an adequate lifetime medical history (listing all major illnesses) was available for only 41.7% of the decedents, representing 56.6% of those for whom some medical record was available. Medical records provided smoking data adequate to classify a man as a current, former, or never smoker for only 24.2% (n = 87) of all decedents. Detailed smoking information was even less available (20.6% for amount smoked, 7.8% for duration smoked).

Medical records were more likely to be available for men who died recently. We obtained some medical record for 58% of decedents who died prior to 1960, 63% for those who died in 1960 to 1969, 75% for those who died in 1970 to 1979, and 90% for deaths after 1979.

Cancer Incidence: Next-of-kin *v* Medical Records Data

Table 2 shows that the next of kin correctly reported a history of cancer for 49 of the 59 (83.1%) decedents with a positive medical record cancer history. Underreporting was more common than overreporting. Seven

decedents were identified as having no history of cancer from the interview, when, in fact, the medical records indicated such a history (three lung, two prostate, one leukemia, and one bladder). Only three next of kin reported a positive cancer history which was not substantiated by the medical records. In these three cases, the next of kin reported that the decedent had a positive history of prostate, pancreatic, and unknown type of cancer.

Table 3 shows that agreement between the next of kin responses and medical records for specific cancer sites was generally good ($\kappa = 0.80$, $P < .0001$). Misclassification of sites tended to occur when the next of kin could not report the cancer type (n = 6). Misclassification may have also occurred if the next of kin reported a metastatic rather than a primary site. For example, in one case, a lung cancer was reported as a "lymph node" cancer and in two other instances, bladder cancers cases were reported as prostatic cancer.

We reanalyzed the data according to type of next of kin (spouse *v* all others). Although both spouses and other types of next of kin were about equally accurate in reporting whether the decedent had ever had cancer (κ 0.83 *v* 0.80), the spouses agreed with medical record data more than other types of next of kin when reporting cancer type (κ 0.84 *v* 0.69).

TABLE 2
Correspondence Between Next-of-kin Response and Medical Records on Whether the Decedent Ever Had Cancer*

Next-of-kin Response	Medical Records		
	Yes	No	Total
Yes	49	3	**52**
No	7	51	**58**
Don't know	3	4	**7**
Total	**59**	**58**	**117**

* Sensitivity = 49/56 = .875; specificity = 51/54 = .944.

Cigarette Smoking: Next-of-kin *v* Medical Records Data

Table 4 indicates the level of agreement between both sources regarding smoking status. Overall, for the 54 men for whom we had both sources of data, 87% of the next of kin reported the same smoking status which have been self-reported by the decedents in medical records. More detailed smoking information was infrequently available from medical records, further reducing the sample size for such comparisons, making firm conclusions difficult. Data from next of kin correlated fairly well with medical record data for number of years smoked ($r = .60$, $P = .07$, n = 10), and number of years since quitting for former smokers ($r = .63$, $P = .04$, n = 11). Correlations were worse for amount smoked ($r = .24$, $P = .13$, n = 39). No significant difference (using paired *t* tests) were observed between the two sources (medical records *v* next of kin) for the average number of years smoked (mean 35.3 *v* 38.7, $P = .45$ for paired test), or for the average number of years since quitting for former smokers (mean 9.6 *v* 10.4, $P = .69$ for paired test). However, medical records indicated that the decedents smoked significantly more packs per day than reported by next of kin (1.6 *v* 1.3, $P = .02$ for paired test).

Discussion

Our attempts to locate all cohort members (n = 1,156) for the cancer incidence study were quite extensive. We used death certificates, HCFA, telephone directory assistance, personnel records, IRS) to initially locate all men. We then mailed interviews (twice if necessary),

TABLE 3
Medical Records *v* Next-of-kin Responses for Specific Cancer Type*

Next-of-kin Response	Medical Records													
	Throat/ Larynx	Stomach	Colorectal	Liver	Pancreas	Lung	Prostate	Bladder	Brain	Lymph Nodes	Leukemia	Reticulum Cell Sarcoma	Unknown Type	Total
Throat/Larynx	5													**5**
Stomach		1												**1**
Colorectal			4											**4**
Liver				1										**1**
Pancreas					4									**4**
Lung						18								**18**
Prostate							1	2						**3**
Bladder								2						**2**
Brain									1					**1**
Lymph nodes						1								**1**
Leukemia											3			**3**
Reticulum cell sarcoma														**0**
Unknown type	3		1									1	1	**5**
Total	**8**	**1**	**5**	**1**	**4**	**19**	**1**	**4**	**1**	**0**	**3**	**1**	**1**	**49**

* $\kappa = 0.80$, $P < .0001$, "unknown type" for both next of kin and medical records means that it was known that the subject had cancer but that type could not be determined.

TABLE 4
Correspondence Between Next-of-kin Response and Medical Records for Cigarette Smoking Status of the Decedent*

Next-of-kin Response	Medical Records			Total
	Never	Current	Former	
Never	0	0	3	3
Current	0	31	1	32
Former	1	2	16	19
Total	**1**	**33**	**20**	**54**

* $\kappa = 0.747, P < .0001$

and subsequently called nonrespondents. The entire effort took over a year. We did the mailings ourselves, and further tracing of nonrespondents and phone call follow-up interviews were done by a contractor, at a cost of $35,000. With all these efforts, we were able to obtain interviews with next of kin for only 59% of decedents and 79% of living persons.

Considering the cost of the clerical time (6 months) as well as the cost of the contractor, each interview (next of kin or live) cost approximately $50. Obtaining the medical records for decedents (clerical time) also cost about $50 per medical record obtained.

Our success in obtaining both next-of-kin interviews and medical record data was greatly increased for more recent deaths. Thirty-nine percent of the deaths occurred prior to 1970. For these early deaths the response rate for next of kin was 45% *v* 68% for later deaths.

The three plants in the study were in or near large cities (Chicago, Pittsburgh, and Cincinnati), and cohort members in such cities may be more mobile than workers in smaller towns. Other investigators may have better success in locating next of kin in cohorts in which deaths are more recent and workers have lived in smaller towns. For example, investigators at the University of Pittsburgh have reported that, in two nested case-control studies, they were able to obtain information from next of kin for 70% of their decedents (n = 303), compared with our 59%.[2] The decedents in these two studies had died somewhat later than those in our cohort, and more had lived in smaller towns. In other work currently being conducted by one of us (K. S.), we have sought information on smoking from the next of kin of 2,750 decedents who died in 1982 to 1983 throughout the United States. In this work, in which deaths were quite recent, our response rate was 81%.

The question then arises whether the data from next of kin are reliable. Our data on cancer incidence indicates that next of kin are fairly good informants regarding cancer incidence of decedents. In a study of cancer incidence, next of kin might be relied upon to identify cases of interest, and then medical records be sought to confirm only these cases.

There are a number of studies in the literature regarding the availability and accuracy of smoking data for decedents available from next of kin.[2,5-8] Pickle et al[6] discuss only availability, whereas the other four studies also assess accuracy.

In our study, next of kin were usually able to provide broad information on the smoking habits of the decedents. However, next of kin were less frequently able to provide detailed information on these variables. Although our sample size was small, we found that, when medical records were used as the standard for comparison, next of kin were fairly accurate in their reporting of overall cigarette smoking status of the decedent (never, former, or current). The result agrees well with previous reports.

In our study, data from next of kin on the number of years smoked and the number of years since quitting for former smokers corresponded well with data reported by the decedent himself on medical records, although sample sizes for these comparisons were quite small. In contrast, next of kin underreported the amount the decedent had smoked. Lerchen and Samet[8] also reported high correlation between spouses and decedents for duration of smoking, and less correlation for amount smoked. Data from Rogot and Reid,[5] in contrast to ours, showed that next of kin overreported the amount smoked by decedents.

The fact that interviews with next of kin may be obtained for only 60% to 70% of decedents, that detailed smoking information is available from perhaps only 70% to 80% of such interviews, and that such detailed information may not be accurate, has implications for investigators attempting to conduct nested case-control studies to control for smoking (or other variables) within occupational cohorts. Given that the case subjects are usually all dead in such studies, investigators may be unable to obtain data on confounders for an important number of cases, reducing power for any analysis which requires data on confounders. If the case subjects are all dead while many control subjects are still alive, the fact that most missing data occurs among decedents *v* live study subjects raises questions of possible bias. Even with the interviews which are obtained, missing and inaccurate data on detailed smoking habits may mean that investigators can only classify case or control subjects into categories of never, former, and current smokers, without regard to duration or amount smoked or time since quitting.

The combined effect of losing a number of case subjects and control subjects, and the restriction to very broad smoking categories, may mean that nested case-control studies which are conducted to control for confounding by smoking are not particularly effective. Several authors have provided examples of the degree to which incomplete or inaccurate data on confounders may cause formal analytic methods of control over those confounders to be relatively inaccurate.[9-12]

In light of the above, it might be preferable to use whatever data can be gathered on smoking in the cohort to make a broad adjustment for the confounding effects of smoking, such as suggested by Axelson,[13] rather than using more formal methods. Alternatively, investigators may choose to collect no further data on smoking and rely on indirect adjustments.[14] Data from several authors indicate that, despite the theoretical potential for confounding, in practice, correction for confounding by smoking rarely changes results.[2,15]

Acknowledgments

The authors acknowledge the efforts of Betty Walpole and Mary Hogan in obtaining medical records, and the services of John Ward at Westat in obtaining interviews.

References

1. Petitti D, Friedman G, Kahn W: Accuracy of information on smoking habits provided on self-administered research questionnaires. *Am J Public Health* 1981;71:308–311.

2. Marsh G, Sachs D, Callahan C, et al: Direct methods of obtaining information on tobacco use in occupational studies. *Am J Ind Med* 1988;13:71–103.

3. Beaumont J, Leveton H, Knox K, et al: Lung cancer mortality in workers exposed to sulfuric acid mist and other acid mists in steel pickling operations. *JNCI* 1988;79:911–921.

4. Fleiss J: *Statistical Methods of Rates and Proportions*, Wiley and Sons, New York, 1973.

5. Rogot E, Reid D: The validity of data from next-of-kin in studies of mortality among migrants. *Int J Epidemiol* 1975;4:51–54.

6. Pickle L, Brown L, Blot W: Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol* 1983;118:99–108.

7. Pershagen G, Axelson O: A validation of questionnaire information on occupational exposure and smoking. *Scand J Work Environ Health* 1982;8:24–28.

8. Lerchen M, Samet J: An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol* 1984;123:481–489.

9. Checkoway H, Waldman G: Assessing the possible extent of confounding in occupational case-referent studies. *Scand J Work Environ Health* 1985;11:131–133.

10. Kupper L: Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. *Am J Epidemiol* 1984;120:643–648.

11. Greenland S: The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112:564–569.

12. Axelson O, Steenland K: Indirect methods of assessing the effects of tobacco use in occupational studies. *Am J Ind Med* 1988;13:105–118.

13. Axelson O: Aspects on confounding in occupational health epidemiology. *Scand J Work Environ Health* 1978;4:85–89.

14. Steenland K, Beaumont J, Halperin W: Methods of control for smoking in occupational cohort mortality studies. *Scand J Work Environ Health* 1984;10:143–149.

15. Blair A, Hoar S, Walrath J: Comparison of crude and smoking-adjusted standardized mortality ratios. *J Occup Med* 1985;27:881–884.

Tokyo Woes

Suddenly, thousands of people want a personal relationship with the world's No. 3 trade power. In US colleges, the study of Japanese is the fastest-growing of all foreign languages. Enrollment in courses at the Japan Society in New York has more than tripled in the past five years. . . . Says Carl Kay, president of Japanese Language Services, . . . "With an MBA or a law degree and a mastery of Japanese, you can write your own ticket."

So what's the downside? . . . It is a wildly difficult language, a sort of moon tongue with no Western reference points. English, with its many Latin roots, is a doorway to a family of other languages—Italian, French, and so forth. Japanese is a doorway to Mongolian.

First, take Japanese on the page. More than 90% of the people of Japan can perform the casually brilliant act of reading it. Their written language is constructed of three wholly different sets of written characters: kanji, hiragana, and katakana.

Kanji are Chinese ideographs, which the Japanese began to appropriate in the sixth century, transforming most with new meanings. Today, you need to know about 2,000 kanji to be able to read the morning paper. There are thousands more. Even the character for "one" has seven different pronunciations, depending on how it's used.

Hiragana is a Japanese alphabet used to spell words for objects and concepts that are native to Japan.

Katakana is another Japanese alphabet, used to spell words for objects and concepts imported to Japan from foreign lands.

In printed Japanese, all three sets of symbols are tossed together, higgledy-piggledy, and the pages are read from back to front.

—From "Tokyo Woes" by Pope Brock in "Office Politics," *GQ* 1987;57(4):165.